

Р. Р. БИККЕНИН, М. Н. ЧЕСНОКОВ

ТЕОРИЯ ЭЛЕКТРИЧЕСКОЙ СВЯЗИ

Рекомендовано

*Учебно-методическим объединением по образованию
в области телекоммуникаций в качестве учебного пособия
для студентов высших учебных заведений, обучающихся
по направлению подготовки «Телекоммуникации»*



Москва

Издательский центр «Академия»

2010

УДК 621.3.01(075.8)

ББК 32.88-01я73

Б603

Рецензенты:

д-р техн. наук, проф. кафедры «Системы радиосвязи» МТУСИ *П. Н. Сердюков*;

д-р техн. наук, проф. Военной академии связи им. С. М. Буденного *И. Н. Оков*;

д-р физ.-мат. наук, проф. Военно-морского института радиоэлектроники
им. А. С. Попова *Е. Ю. Бутырский*

Биккенин Р. Р.

Б603 Теория электрической связи : учеб. пособие для студ. высших учебных заведений / Р. Р. Биккенин, М. Н. Чесноков. — М. : Издательский центр «Академия», 2010. — 336 с.

ISBN 978-5-7695-6510-6

Приведены основы теории сигналов. Рассмотрены методы формирования сигналов, синтеза и анализа помехоустойчивости оптимальных устройств обработки и фильтрации в условиях случайных помех. Представлены модели различных каналов телекоммуникаций. Дан анализ основных проблем передачи информации.

Для студентов высших учебных заведений. Может быть полезно специалистам по защите информации.

УДК 621.3.01(075.8)

ББК 32.88-01я73

*Оригинал-макет данного издания является собственностью
Издательского центра «Академия», и его воспроизведение любым способом
без согласия правообладателя запрещается*

© Биккенин Р. Р., Чесноков М. Н., 2010

© Образовательно-издательский центр «Академия», 2010

ISBN 978-5-7695-6510-6 © Оформление. Издательский центр «Академия», 2010

65-74-73-101-75-67

ПРЕДИСЛОВИЕ

В 1924 г. в нашей стране появился один из первых учебников по радиотехнике [20], написанный профессором И. Г. Фрейманом, который руководил первой в России кафедрой радиотехники, образованной в 1917 г. в Петербургском электротехническом институте. В то время предмет «Радиотехника», преподававшийся специалистам связи, включал в себя следующие вопросы: формирование и усиление сигналов, методы и устройства модуляции, фильтрация и прием сигналов в условиях помех, применение антенн и многие другие, которые в настоящее время изучаются в рамках курса «Теория электрической связи» (ТЭС).

В конце 1940-х гг. в сознании специалистов по радиотехнике и связи произошел поистине революционный переворот, явившийся следствием публикации двух научных работ, сразу ставших фундаментальными. Эти работы и по настоящее время не утратили своего огромного значения. Одна из них — «Теория потенциальной помехоустойчивости» [10] была докторской диссертацией, которую в 1947 г. успешно защитил В. А. Котельников, впоследствии действительный член Академии наук СССР, лауреат многих престижных отечественных и международных премий за научные достижения. Автором другой работы — «Математическая теория связи» [22], опубликованной в 1948 г., является американский инженер и математик, профессор К. Э. Шеннон.

Эти выдающиеся ученые получили строгие математические доказательства и определили условия передачи по каналам связи определенного количества информации в заданное время и с высокой достоверностью. Они также обосновали потенциально достижимые характеристики систем связи, что определило магистральное направление их дальнейшего развития на многие годы.

Одновременно данные работы послужили основой формирования общего подхода к постановке учебно-

го курса ТЭС, который в настоящее время является базовым во всех высших учебных заведениях, где ведется подготовка специалистов по телекоммуникациям.

В данном учебном пособии систематизирован материал дисциплины «Теория электрической связи», охватывающий основные положения программ этой дисциплины, которая соответствует требованиям Государственного образовательного стандарта высшего профессионального образования по направлению 654400 «Телекоммуникации» подготовки инженеров. Также пособие может использоваться для подготовки бакалавров и магистров по направлению «Телекоммуникации» и, кроме того, при обучении студентов по смежным направлениям: «Комплексная защита объектов информатизации», «Радиотехника» и др.

В настоящем учебном пособии на достаточном для понимания и усвоения уровне математической строгости, но без лишнего формализма, представлены теоретические положения и принципы построения основных элементов в современных системах передачи информации, а также рассмотрены вопросы формирования сигналов посредством всевозможных методов модуляции. При этом особое внимание уделено дискретным видам сигналов. Представлены математические модели различных каналов телекоммуникаций. С использованием вероятностных методов выполнен синтез оптимальных приемных устройств (демодуляторов) и дана оценка их помехоустойчивости в условиях воздействия разнообразных помех. Рассмотрены основные понятия теории информации и помехоустойчивого кодирования, показаны потенциальные характеристики каналов связи по передаче определенного количества информации с учетом реально существующих ограничений. Приведены методы передачи и обработки непрерывных сигналов на основе оптимальной нелинейной фильтрации. Там, где это представлялось целесообразным, даны указания на возможность реализации рассмотренных теоретических положений в современных системах телекоммуникаций.

При изложении материала авторы использовали традиционные принципы и подходы, сформировавшиеся за время многолетней деятельности кафедры «Теоретические основы связи и радиотехники» СПбГУТ

им. профессора М. А. Бонч-Бруевича. При написании учебного пособия авторы также учитывали методы изложения ряда проблемных вопросов дисциплины «Теория электрической связи» смежных дисциплин на аналогичных кафедрах других высших учебных заведений.

Кроме того, в учебном пособии обобщен многолетний опыт преподавания авторами дисциплины «Теория электрической связи».

Ввиду ограниченности объема пособия остались не рассмотренными либо представлены конспективно отдельные вопросы учебного курса. Тем не менее авторы уверены в том, что настоящее пособие будет способствовать формированию целостного общего взгляда на процессы, происходящие в системах связи, правильному пониманию места и значимости теории связи в современных условиях, когда системы телекоммуникаций прочно вошли в жизнь всего мирового сообщества и пронизывают все сферы и стороны человеческой деятельности.

Авторы благодарят А. В. Евстратова, Р. К. Жумабаева, Б. С. Каменецкого, А. А. Соболева, А. В. Боровика за оказанную помощь по подготовке рукописи к изданию.

1.1. Понятия информации, сообщения, сигнала

В теории и технике электрической связи существует ряд основополагающих, фундаментальных понятий, к числу которых, в первую очередь, относятся понятия информации, сообщения и сигнала.

Рассмотрим эти понятия, используя рекомендации сборника научно-технической терминологии в области теории передачи информации, разработанного Академией наук СССР.

Под *информацией* (от лат. *informatio* — изложение, разъяснение) принято понимать сведения о событиях, явлениях, предметах, являющихся объектом ряда операций: передачи, распределения, преобразования, хранения или непосредственного использования.

Во всем многообразии окружающего нас мира мы постоянно сталкиваемся с информацией и процессами ее передачи и хранения. Так, органы чувств человека занимаются сбором информации об окружающем внешнем мире, а его нервная система передает эту информацию в головной мозг, который ее перерабатывает, а затем рассылает (распределяет) в виде «приказов», также являющихся информацией, по нервным волокнам (линиям связи) в мышцы.

Аналогично информация передается в любой организации, где совместно трудится множество людей в виде приказов, распоряжений и других указаний, т. е. без чего невозможна деятельность большого коллектива. Перечень подобных примеров можно продолжать и дальше. Однако и так ясно, что задачи сбора, передачи, преобразования информации очень важны в различных областях человеческой деятельности, в том числе в системах электросвязи (телекоммуникаций).

В ряде случаев понятие информации отождествляется с понятием *данные*, что находит широкое применение в системах цифровой связи.

В целом информацию можно трактовать как совокупность знаний человека об окружающем его мире.

Системы электросвязи предназначены для передачи информации от источника сообщений, находящегося в некоторой точке пространства, потребителю (получателю) сообщения, который находится в другой точке. Использованное в этом определении понятие сообщения наряду с понятием информации широко применяется в теории электрической связи.

Обычно под *сообщением* (message) понимают форму представления информации в целях ее хранения, обработки, преобразования или непосредственного применения. При этом используются различные знаки и символы, например определенные слова и фразы в человеческой речи, рисунки, математические знаки, виды колебаний и т. п.

Сообщения могут быть функциями времени, как, например, речь в телефонной связи, программа новостей, передаваемая по телевидению и др. Однако в ряде случаев сообщения не являются функциями времени, например текст телеграммы или неподвижное изображение для передачи средствами факсимильной связи. Различают также дискретные и непрерывные сообщения. К дискретным сообщениям относятся текст, цифровые данные, а к непрерывным — речь, телевизионное изображение, температура или давление при передаче телеметрических данных и т. д.

Фактически человек всегда имел дело не с абстрактной информацией, а с конкретными сообщениями, которые вырабатываются источниками в целях последующей передачи в системе электросвязи. Под *источником сообщений* (message source) принято понимать устройство, которое в каждый момент времени выбирает некоторое сообщение из множества (ансамбля) сообщений. Если имеется вероятностная модель, с помощью которой можно дать полное описание процесса появления сообщений на выходе источника, считают, что источник сообщений задан. Например, в качестве источника сообщений можно рассматривать оператора, работающего на телеграфном аппарате. При этом должны быть известны вероятности появления отдельных сообщений, буквенных сочетаний, слов, предложений. Иначе говоря, для любых $n = 1, 2, \dots$ и $i = 0, \pm 1, \pm 2, \dots$ и любой последовательности сообщений $(x^{i+1}, \dots, x^{i+n})$, выбираемых из множества сообщений X , определена вероятность $P(x^{i+1}, \dots, x^{i+n})$ появления этой последовательности.

Устройство, для которого предназначено сообщение, вырабатываемое источником, называется *получателем сообщений*. Получателем может быть человек-оператор или различные регистрирующие устройства, в том числе электронно-вычислительные машины (ЭВМ).

По виду источника и получателя сообщений принято различать системы электросвязи, осуществляющие передачу:

- акустических или звуковых сигналов (телефония, радиовещание);
- текста (телеграфия) и данных от ЭВМ;
- неподвижных изображений (факсимильная связь);
- подвижных изображений (телевидение);
- данных телеметрии, контроля (например, системы охранной, пожарной сигнализации и др.).

Для передачи сообщений на определенное расстояние используются различные материальные носители (бумага, магнитный диск и др.) или некоторый физический процесс (звуковые волны, электромагнитные волны и т. д.). В системах электросвязи для передачи сообщений применяются различные сигналы.

Сигналом (от *лат.* *signum* — знак) принято называть физический процесс, например в виде тока или напряжения, отображающий передаваемое сообщение. Сигнал всегда является функцией времени, если даже сообщение, которое он переносит, не описывается временной функцией. Сигнал в системах электрической связи в простейшем случае обозначается

$$u_i(t, A, \omega, \varphi), t_1 \leq t \leq t_2,$$

где i — номер сигнала; $t_2 - t_1 = T$ — интервал определения сигнала во времени; A, ω, φ — параметры, т. е. соответственно амплитуда, частота и фаза сигнала.

В зависимости от множества возможных значений параметров и области определения во времени различают следующие виды сигналов:

- непрерывный и по уровню, и во времени (аналоговый);
- непрерывный по уровню, но дискретный во времени;
- дискретный (квантованный) по уровню, но непрерывный во времени;
- цифровой, т. е. дискретный и по уровню, и во времени.

Примеры различных видов сигналов представлены на рис. 1.1.

Так, речевой сигнал является непрерывным и во времени, и по уровню, а датчик, определяющий значение температуры через каждые 5 мин, выдает сигналы непрерывные по значению (амплитуде), но дискретные во времени.

Передаваемое сообщение и соответствующий ему сигнал не должны быть детерминированными, т. е. заранее полностью известными и предсказуемыми. В этом случае передача сообщений не имеет никакого смысла, так как при отсутствии неопределенности значений сигналов получателю не будет доставляться новая информация. Только случайная величина или случайная функция может быть носителем информации. Обычно это реализуется по-

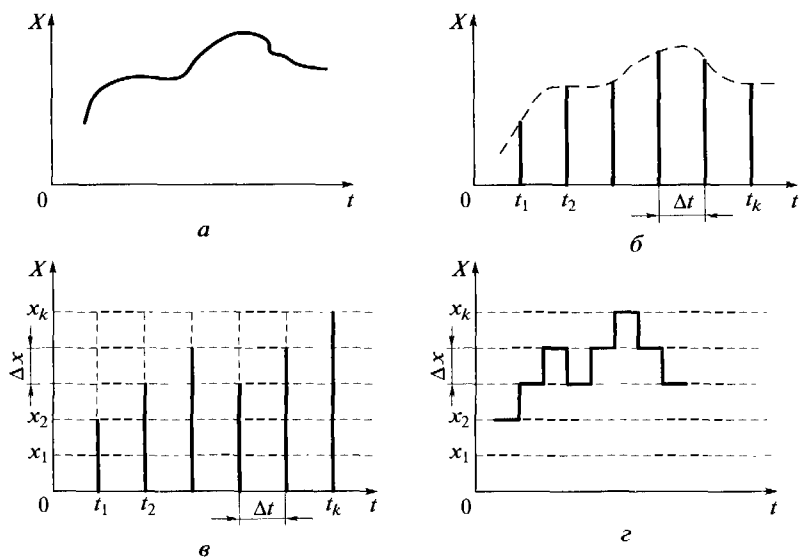


Рис. 1.1. Примеры основных видов сигналов:

a — непрерывный и по уровню, и во времени; *б* — непрерывный по уровню, но дискретный во времени; *в* — дискретный по уровню, но непрерывный во времени; *г* — дискретный и по уровню, и во времени

средством избрания из некоторого множества вариантов (реализаций) какого-то одного. При этом выбор осуществляется с некоторой вероятностью. Например, из множества значений температуры, выдаваемых датчиком, в текущий момент времени предпочтение отдается только одному из них.

Сообщения, сигналы их отображающие, а также помехи, искажающие сигнал, имеют случайный характер. Поэтому в теории электрической связи для объяснения ряда понятий широко используются теория вероятностей, а также теория случайных процессов и математической статистики. На их основе рассматриваются свойства сигналов, свойства среды их распространения, методы обработки сигналов и количество информации, передаваемой от источника сообщений к получателю.

В теории электрической связи сигнал принято отождествлять с объектом транспортирования. Следовательно, аппаратура связи по существу является техникой транспортирования или передачи сигналов по каналам телекоммуникаций.

Определим параметры сигнала, которые являются основными при его передаче. К числу таких параметров обычно относятся: длительность, динамический диапазон, ширина спектра.

Любой сигнал, являющийся функцией времени, имеет начало и конец. Следовательно, *длительность* (T) определяет интервал времени, в пределах которого сигнал существует.

Динамическим диапазоном (D) называется отношение наибольшего значения мгновенной мощности к ее наименьшему значению, при котором обеспечивается заданное качество передачи информации. Иногда под динамическим диапазоном понимается отношение мощностей сигнала и помехи. Динамический диапазон определяется в децибелах. В системах радиовещания отношение сигнал/шум составляет порядка 50... 60 дБ при передаче музыкальных программ и 30 дБ — при передаче речевых сигналов, а в системах телевидения это отношение равно 60 дБ.

Под *шириной спектра* (F) сигнала принято понимать диапазон (полосу) частот, в пределах которого сосредоточена его основная мощность. Спектр сигнала в принципе может быть неограниченным, однако его сознательно сокращают с учетом ограниченных возможностей техники связи.

Так, при телефонной связи речевой сигнал передают в полосе частот от 300 до 3400 Гц, т. е. ширина спектра сигнала в этом случае $F = 3,1$ кГц. Этого диапазона частот оказывается вполне достаточно для обеспечения разборчивости речи и узнаваемости абонентов по голосу.

При передаче телевизионного сигнала важнейшим требованием является четкость принимаемого изображения. При стандарте в 625 строк верхняя частота сигнала составляет примерно 6 МГц, т. е. спектр сигнала видеоизображения занимает значительно более широкую полосу частот, чем спектр сигнала звукового сопровождения.

При телеграфной связи ширина спектра сигнала, определяемая скоростью его передачи (телеграфирования), составляет $(1,5 \dots 3,0)v$, где v — скорость передачи, измеряемая в бодах и равная числу электрических посылок, передаваемых в 1 с. Обычно $v = 50$ Бод, тогда $F \approx 75$ Гц.

В заключение можно ввести общую характеристику: *объем сигнала* $V_c = F_c T_c D_c$, которая дает наиболее полное представление о возможностях сигнала как переносчика информации. Чем больше объем, тем большее количество информации может перенести сигнал, но с другой стороны тем труднее такой сигнал передавать по каналу с необходимым качеством.

1.2. Модель системы передачи информации

Рассмотрим структурную схему простейшей одноканальной системы передачи информации. Введем понятие канала связи.

Под *каналом связи* (communication link) в теории и технике электрической связи принято понимать совокупность различных средств, включая физическую среду, которая обеспечивает передачу сигналов от источника к получателю сообщений. Причем физической средой для передачи сигналов может быть кабель в проводной связи, атмосфера в наземной радиосвязи и т. д.

В самом общем виде структурная схема системы передачи информации показана на рис. 1.2.

На передающей стороне преобразование сообщения в сигнал осуществляется с помощью преобразователя. В телефонии для этой цели служит микрофон, который превращает акустические колебания в пропорционально изменяющееся электрическое напряжение. В телеграфии с помощью телеграфного аппарата (телетайпа) оператор заменяет последовательность знаков сообщения (букв, цифр) последовательностью двоичных кодовых символов (0 и 1). В телетайпе они преобразуются в электрические посылки постоянного тока. В телевидении при передаче изображения преобразователем является передающая телевизионная трубка.

Далее следует операция кодирования (coding), под которой понимают преобразование дискретного сообщения в последовательность кодовых символов, осуществляемое по определенному правилу. При этом каждому элементу сообщения присваивается определенная совокупность кодовых символов, называемая *кодовой комбинацией* (*кодовым словом*), а совокупность всех кодовых комбинаций называется *кодом*. Правило кодирования принято задавать кодовой таблицей, в которой каждому сообщению соответствует определенная кодовая комбинация. Понятие кодирования применимо только к дискретным сообщениям, поэтому

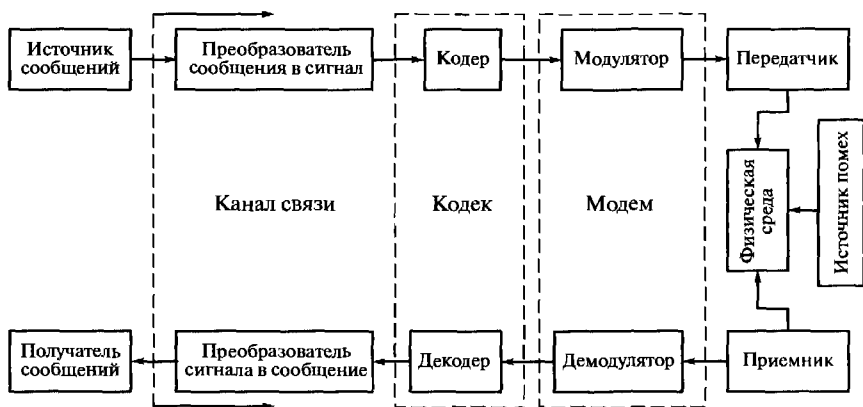


Рис. 1.2. Структурная схема системы передачи информации

чтобы закодировать речевое сообщение, являющееся аналоговым, его необходимо сначала представить в дискретной форме.

В телеграфии первичное кодирование осуществляется с помощью телетайпа, в котором каждая буква, каждая цифра и каждый служебный знак (точка, запятая, знак сложения и т. д.) кодируются первичным кодом. Например, это может быть международный телеграфный код № 2 (МТК-2), каждая комбинация которого содержит по пять двоичных символов. Число возможных комбинаций в этом коде составляет $2^5 = 32$. Этого вполне достаточно для кодирования всех букв русского алфавита, а для кодирования остальных знаков следует использовать регистровый принцип. В этом случае одна и та же комбинация применяется три раза: в русском, латинском и цифровом регистрах. Общее число разных знаков (букв, цифр и др.), применяемых в коде МТК-2, равно 84.

В 1963 г. появился код ASCII (American Standard Code for Information Interchange) — стандартный американский код для обмена информацией, разработанный для использования в телеграфной связи. При создании первых персональных ЭВМ фирма IBM приняла его в качестве стандарта для кодирования информации. Каждая комбинация данного кода, состоящая из семи двоичных символов, позволяла использовать 128 кодовых комбинаций. Несколько позже этот код был расширен и дополнен: его комбинации стали содержать по восемь двоичных разрядов, и число этих комбинаций возросло до 256. Благодаря этому его стали применять для кодирования информации не только на английском языке, но и на многих других языках мира. В настоящее время все текстовые сообщения, передаваемые в сети Интернет, кодируются с использованием только этого кода.

Коды МТК-2 и ASCII относятся к так называемым *равномерным* кодам, поскольку каждая комбинация в них содержит одно и то же число двоичных символов. Также существуют *неравномерные* коды, комбинации в которых имеют разную длину.

Типичным представителем неравномерных кодов является код Морзе, созданный в 1838 г. американским изобретателем и художником Самюэлем Морзе. В этом коде символ «1», соответствующий токовой посылке, называется точкой, а три единицы — тире. Символ «0» используется как разделительный знак внутри кодовой комбинации, а совокупность из трех нулей разделяет между собой кодовые комбинации. Данный код до сих пор применяется в системах слуховой телеграфной радиосвязи. В 2004 г. в коде Морзе появился символ @ и соответствующая ему кодовая комбинация, введенная Международным союзом электросвязи для удобства передачи адресов электронной почты.

К неравномерным относятся и широко известные коды Хаффмана и Шеннона — Фано. В них, как и в коде Морзе, сообщения,

встречающиеся чаще (с большей вероятностью), кодируются короткими кодовыми комбинациями, а сообщения, появляющиеся реже (с меньшей вероятностью), — более длинными кодовыми комбинациями. Это свойство позволяет устранять избыточность в источниках сообщений (т. е. производить «сжатие» информации). Такие коды называются *префиксными*, поскольку в своем составе они не имеют кодовых комбинаций, которые являются началом (префиксом) любых других. Данное свойство позволяет легко распознавать принимаемые сообщения. Код Шеннона—Фано более простой в построении, однако код Хаффмана несколько удобнее в практической реализации. Код Хаффмана используется в технике факсимильной связи и в компьютерных технологиях при создании файлов видеозображений в формате JPEG, а также для сжатия видеосигналов в телевизионной цифровой технике на основе стандарта MPEG.

Код Хаффмана, предложенный в 1952 г., можно построить следующим образом. Сначала все сообщения располагаются в порядке убывания вероятностей их появления. Затем сообщения с наименьшими вероятностями, стоящие внизу, объединяются в одно промежуточное (вспомогательное) сообщение, которому приписывается значение, равное сумме вероятностей сообщений, из которых оно составлено. Полученную таким образом точку называют узлом, а пути, ведущие в нее, обозначают кодовыми символами: 1 (верхний) и 0 (нижний). Затем из оставшихся сообщений с учетом промежуточного сообщения вновь находят пару с наименьшими вероятностями, которая аналогично объединяется в очередное промежуточное сообщение с значением вероятности, равным сумме вероятностей, входящих в эту пару сообщений. Во второй полученный узел также ведут два пути: единичный и нулевой. Далее процесс объединения продолжается рекурсивно до получения завершающего вспомогательного сообщения с суммарной вероятностью, равной единице. Эту последнюю полученную точку называют корнем кодового дерева, ветвями которого являются пути, приводящие в соответствующие узлы. Считывание кодовых символов производится в обратном направлении: от корня дерева к исходным сообщениям. В качестве кодовых комбинаций сообщений записываются последовательности двоичных символов, встречающиеся на каждом пути, соединяющем соседние узлы дерева.

Пример построения кодового дерева и полученные при этом комбинации для источника сообщений, создающего символы a, b, c, d, e, f , вероятности P_k (где $k = \overline{1,6}$) которых равны соответственно 0,33, 0,22, 0,13, 0,12, 0,11, 0,09, представлены на рис. 1.3, *а*. На рис. 1.3, *б* приведена таблица, поясняющая процесс объединения сообщений в промежуточные узлы.

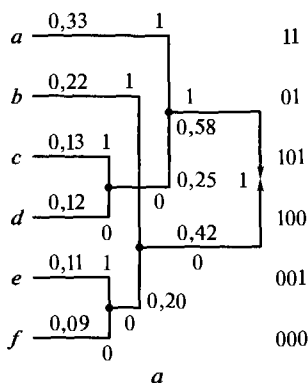


Рис. 1.3. Пример построения кодового дерева (а) и таблица шагов (б) при построении кода Хаффмана

Символ	P_k	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Шаг 5
a	0,33	0,33	0,33	0,33	0,58	1
b	0,22	0,22	0,22	0,42	0,42	
c	0,13	0,13	0,25	0,25		
d	0,12	0,12				
e	0,11	0,20	0,20			
f	0,09					

б

Среднюю длину комбинации кода Хаффмана можно найти с помощью соотношения

$$n_{\text{ср}} = \sum_{k=1}^N P_k n_k,$$

где P_k — вероятность появления k -го сообщения, содержащего n_k двоичных символов.

Для рассмотренного примера $n_{\text{ср}} = 2,35$. Если эти сообщения кодировать равномерным простым кодом, то каждая комбинация должна содержать по три двоичных символа, т. е. $n = 3$. Следовательно, выигрыш в длине кодовой комбинации в среднем составляет примерно 22 %.

Рассмотренные коды относятся к так называемым *первичным кодам*. Равномерные телеграфные коды, представленные ранее, называются также *простыми (примитивными)*, или *кодами без избыточности*. Это связано с тем, что искажение любого из символов комбинации приводит к образованию новой разрешенной комбинации, т. е. к ошибке, что выражается в регистрации буквы или цифры, отличающейся от переданного знака.

Существуют также коды, *корректирующие ошибки* (error correction), или *помехоустойчивые*, которые строятся таким образом, чтобы для передачи сообщений применялись не все возможные комбинации, а только часть из них, называемые *разрешенными*. Это позволяет обнаруживать и исправлять ошибки при искажениях некоторых символов. Корректирующие свойства таких кодов обеспечиваются целенаправленным введением в комбинации примитивных кодов дополнительных (избыточных) символов. Эта операция выполняется в кодирующем устройстве — *кодере*.

Примером одного из простейших равномерных корректирующих кодов является код с постоянным весом, т. е. с одинаковым числом единиц в любой из разрешенных кодовых комбинаций, общее число которых определяется соотношением $N = C_n^k = n! / [k!(n - k)!]$.

Наиболее известен код, в котором имеется $N = C_7^3 = 35$ разрешенных 7-элементных комбинаций, содержащих по три токовых и четыре бестоковых посылки. Изменение данного соотношения при передаче сообщений свидетельствует о появлении искажений. С помощью такого кода обнаруживаются одиночные и другие нечетные ошибки. При этом необнаруженными остаются искажения, называемые трансформацией, т. е. искажения, при которых единицы преобразуются в нулевые символы и одновременно нули преобразуются в единицы, но при этом сохраняется соотношение три единицы и четыре нуля.

В общем случае построение корректирующего кода, способного не только обнаруживать, но и исправлять возникающие ошибки, достаточно сложная задача, которая решается с использованием ряда разделов высшей алгебры.

Далее закодированный сигнал поступает в модулятор.

Модуляцией (modulation) называется преобразование исходного сигнала посредством изменения параметров сигнала-переносчика в соответствии с преобразуемым (модулируемым) сигналом. В качестве сигнала-переносчика информации применяется гармоническое высокочастотное колебание, импульсная последовательность или шумовой процесс.

При использовании в качестве сигнала-переносчика гармонического колебания $S(t) = U \cos(\omega t + \varphi)$ возможна реализация трех видов модуляции: амплитудной (АМ), частотной (ЧМ) и фазовой (ФМ). При использовании в качестве управляющего колебания закодированной последовательности двоичных кодовых символов получим дискретную (цифровую) модуляцию, которую принято называть *манипуляцией*.

Поясним сказанное с помощью рис. 1.4. При АМ символу «1» соответствует передача колебания на несущей частоте в течение

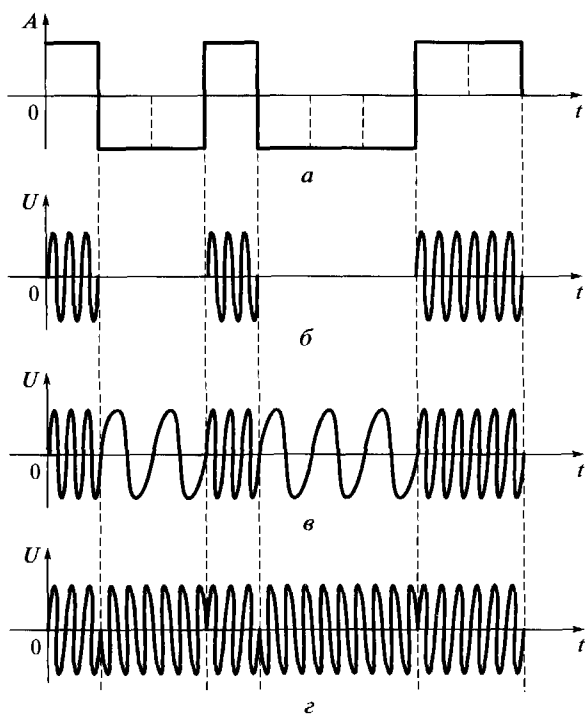


Рис. 1.4. Виды дискретной модуляции сигналов:

a — модулирующий сигнал; *б* — амплитудная модуляция; *в* — частотная модуляция; *г* — фазовая модуляция

времени T (длительность посылки), а символу «0» — отсутствие колебания (пауза). При ЧМ осуществляется поочередная передача колебаний с частотой f_1 , что соответствует передаче символа «1», и колебаний с частотой f_0 , что соответствует передаче символа «0». При двоичной ФМ происходит изменение фазы несущего колебания на 180° при каждой смене полярности в управляющей последовательности прямоугольных посылок.

Длительность T посылки управляющего сигнала позволяет определить техническую скорость передачи (скорость манипуляции), которую принято выражать числом посылок, передаваемых в секунду. Данная единица измерения скорости получила наименование бод (по имени французского изобретателя телеграфного аппарата и кода Ж.-М. Э. Бодо). Один бод соответствует передаче одной электрической посылки в течение одной секунды. Если длительность посылки задается в секундах, то скорость передачи $v = 1/T$, Бод.

Усиление модулированных сигналов по мощности и вывод их в линию реализует *передатчик (transmitter)*. В каналах радиосвязи на выходе передатчика включается антенна, которая осуществляет преобразование электрических сигналов в электромагнитные колебания и излучает их в окружающее пространство. Основными характеристиками современного передатчика являются диапазон применяемых частот, мощность и коэффициент полезного действия (КПД). В зависимости от свойств канала связи и предназначения передатчика его мощность может колебаться от долей до нескольких тысяч ватт. Для сравнения можно сказать о передатчиках сотовых телефонов и широкоэмитательных станций, ведущих радиотрансляцию на сотни и тысячи километров. Диапазоны частот, применяемые в настоящее время, имеют также весьма широкие границы: от сотен килогерц до тысяч мегагерц.

Поскольку отправитель и получатель сообщений в системе передачи информации находятся в различных точках пространства, то между передатчиком и приемником создается некоторая физическая среда. В системах проводной связи — это электрический или оптический кабель, а в системах радиосвязи — область естественного пространства, по которому распространяются электромагнитные волны (радиоволны). В процессе передачи сигнал ослабляется и может искажаться вследствие воздействия всевозможных помех.

Антенна приемника улавливает лишь незначительную долю энергии, которая излучается передающей антенной. Далее происходит усиление принятого колебания и выделение сигнала, несущего информацию, предназначенную конкретному получателю. Эти операции осуществляются в *приемнике (receiver)*. Основными характеристиками приемника являются диапазон применяемых частот, чувствительность — способность принимать весьма слабые сигналы на фоне помех, а также избирательность, под которой понимают способность выделять полезные сигналы из совокупности передаваемых колебаний и посторонних мешающих воздействий, отличающихся от принимаемого сигнала частотой.

Принятый сигнал поступает в демодулятор.

Демодуляция (demodulation) — это преобразование модулированного сигнала, искаженного помехами, в модулирующий сигнал. Иными словами, посредством демодуляции восстанавливается первичный сигнал, отображающий переданное сообщение. Далее этот сигнал поступает в *устройство преобразования сигнала в сообщение*.

В телефонной связи, радиовещании или при звуковом вещании таким устройством является громкоговоритель, в факсимильной связи — приемный факсимильный аппарат, а в телевидении приемная телевизионная трубка.

В системах передачи дискретных сообщений в процессе демодуляции элементы сигнала преобразуются в последовательность кодовых символов, которая поступает в декодер.

Декодирование (decoding) — это восстановление дискретного сообщения по выходному сигналу демодулятора, осуществляемое с учетом правила кодирования. Если на передаче был применен помехоустойчивый или корректирующий код, то на выходе декодера образуются кодовые комбинации первичного (простого) кода. Например, при передаче текстовых сообщений роль преобразователя сигнала в сообщение выполняет приемный буквопечатающий телеграфный аппарат (телетайп), с помощью которого и будет отпечатан текст телеграммы. В системах слуховой телеграфной радиосвязи в качестве преобразователя выступает человек, т. е. оператор, на слух определяющий, какой из сигналов кода Морзе («точка» или «тире») был передан. Оператор также выполняет операцию декодирования, записывая на бумаге текст переданного сообщения.

В системах буквопечатающей телеграфной связи определение сигнала выполняется автоматически с помощью специального устройства, в котором задается некоторое значение порога. Если принятый сигнал превысил пороговое значение, то выдается символ кода, например «1», а если не превысил, — выдается символ «0». В отдельных случаях могут применяться два пороговых значения: положительное и отрицательное. Тогда, если сигнал принимается с искажениями и его уровень оказывается в промежутке между этими порогами, никакого решения не выносится. Вместо сомнительной посылки сигнала вырабатывается особый символ, называемый *стиранием*. Введение такого третьего решения повышает вероятность правильного декодирования принятой кодовой комбинации за счет ее повторной передачи по запросу с приемной стороны.

Таким образом, в системах передачи дискретных сообщений решение о передаваемом сообщении принимается в два этапа. Первой решающей схемой в этом случае является демодулятор, а второй — декодер. В системах передачи аналоговых сообщений решение выносится сразу в демодуляторе. Иногда при передаче дискретных сообщений применяется процедура приема сообщений в целом. В этом случае одним устройством выполняется совместная операция демодуляции-декодирования, в результате чего входящий ряд сигналов сразу преобразуется в последовательность знаков (букв) сообщения.

Существует ошибочное мнение, что демодуляция и декодирование — это операции, обратные модуляции и кодированию, выполняемые с принятым сигналом. На самом деле в результате различных искажений и воздействия помех принятое колебание может существенно отличаться от переданного сигнала. Поэтому

данные операции являются наиболее сложными в системе передачи информации. Для принятия решения о переданном сообщении необходимо детально проанализировать принятый сигнал, для чего его подвергают различным преобразованиям, которые называются обработкой сигнала. Следовательно, одной из задач теории электрической связи является отыскание правил (процедур) оптимальной обработки сигнала, при которых решение о переданном сообщении является наиболее достоверным.

Завершая рассмотрение системы передачи информации, отметим, что качество обработки сигналов существенным образом зависит от точности *синхронизации* переданных и принятых сигналов. При этом различают следующие виды синхронизации: *тактовую* — установление границ посылок сигналов, *цикловую*, при которой следует различать границы кодовых комбинаций, *синхронизацию несущих частот* и др. Неточности синхронизации приводят к снижению достоверности приема информации. Сбой в работе системы синхронизации делает вообще невозможным правильный прием переданных сообщений. Подробно системы синхронизации рассматриваются в специальных курсах.

Проанализированная система передачи информации является одноканальной, т. е. она обеспечивает передачу информации от одного источника к одному получателю.

Существуют также многоканальные системы. Упрощенная схема одной из таких систем показана на рис. 1.5, в которой по одной общей линии связи обеспечивается обмен информацией между несколькими абонентами. В такой системе первичные сигналы, подлежащие передаче, преобразуются посредством модуляторов M_1, M_2, \dots, M_n в электрические сигналы $U_1(t), U_2(t), \dots, U_n(t)$, а затем объединяются в аппаратуре уплотнения. Полученный таким образом групповой (суммарный) сигнал $U_{\Sigma}(t)$ передается по линии связи. На приемной стороне колебание $Z(t) = U_{\Sigma}(t) + n(t)$, искаженное помехами, с помощью устройства разделения, основу которого составляют индивидуальные фильтры $\Phi_1, \Phi_2, \dots, \Phi_n$,



Рис. 1.5. Структурная схема многоканальной системы передачи информации

разделяется на сигналы $\hat{U}_1(t)$, $\hat{U}_2(t)$, ..., $\hat{U}_n(t)$, которые с помощью демодуляторов D_1 , D_2 , ..., D_n преобразуются в первичные сигналы $\hat{a}_1(t)$, ..., $\hat{a}_n(t)$. Для разделения сигналов обычно используется их различие по частоте, времени или форме.

В заключение отметим, что в современных системах передачи дискретных сообщений принято различать две группы относительно самостоятельных устройств: кодеки и модемы.

Кодек (сокращение словосочетания кодер–декодер) — устройство, в котором сообщение в процессе передачи преобразуется в код (кодер), а код в процессе приема преобразуется в сообщение (декодер). *Модем* (сокращение словосочетания модулятор–демодулятор) — устройство, преобразующее при передаче код в сигнал (модулятор), а при приеме сигнал в код (демодулятор). Обычно эти устройства выполняются в виде целостных узлов, через которые проходят цепи на передачу и на прием точно так же, как в телефонных аппаратах.

1.3. Помехи и искажения в каналах связи

В процессе прохождения по реальным каналам связи сигналы подвергаются искажениям, поэтому получаемые сообщения воспроизводятся с некоторыми ошибками. Эти ошибки обусловлены характеристиками тракта передачи, а также помехами, действующими на сигнал. Изменение характеристик тракта, как правило, имеет регулярный характер, и поэтому их можно в большинстве случаев устранить посредством соответствующей коррекции. Помехи же, действующие на сигнал, имеют случайный характер, т. е. они заранее неизвестны и потому их влияние нельзя полностью устранить.

Помехой принято называть любое случайное воздействие на сигнал, которое снижает достоверность воспроизведения передаваемых сообщений. Существующие помехи весьма разнообразны по своей природе и физическому воздействию.

В радиоканалах различают:

- атмосферные помехи, обусловленные грозовыми электрическими процессами. Наиболее вредное воздействие эти помехи оказывают в области длинных и средних волн. Первым обнаружил их негативное влияние изобретатель радио А. С. Попов;
- индустриальные помехи, возникающие из-за резких изменений тока в цепях электроустройств. Это помехи от электротранспорта, систем зажигания двигателей, медицинских установок, электродвигателей;
- помехи от посторонних радиостанций, возникающие вследствие плохой фильтрации гармоник сигнала, недостаточной ста-

бильности частот, нарушения регламента рабочих частот, нелинейности каналов, что приводит к образованию новых колебаний;

- космические помехи, обусловленные электромагнитными процессами, происходящими на Солнце, звездах и других внеземных объектах.

В каналах проводной связи основными видами помех являются импульсные шумы и прерывание связи. Импульсные шумы возникают при автоматической коммутации и вследствие перекрестных наводок. *Прерыванием связи* называется явление, при котором сигнал либо резко затухает, либо совсем пропадает, например из-за нарушения контактов при соединениях.

Все указанные помехи относятся к *внешним* помехам, однако имеются и *внутренние* помехи, возникающие в аппаратуре, например в усилителях и преобразователях частоты. Внутренние помехи обусловлены, главным образом, наличием тепловых шумов — хаотического движения носителей заряда (электронов) в проводниках. Эти помехи принципиально неустранимы.

В общем случае влияние помех на полезный сигнал можно представить в виде оператора

$$Z(t) = L[S(t), n(t)].$$

В зависимости от характера взаимодействия с сигналом помехи подразделяются на аддитивные и мультипликативные.

Аддитивной называется помеха, которая при образовании выходного сигнала представляется в виде слагаемого:

$$Z(t) = S(t) + n(t).$$

Мультипликативной называется помеха, которая при образовании выходного сигнала представляется в виде множителя входного сигнала:

$$Z(t) = S(t) K(t),$$

где $K(t)$ — некоторый случайный процесс.

Примером мультипликативной помехи являются замирания, заключающиеся в случайном изменении уровня и соответственно мощности сигнала из-за непостоянства условий распространения радиоволн. В проводных каналах мультипликативной помехой может быть прерывание связи, при котором сигнал в линии резко затухает.

К аддитивным помехам можно отнести все рассмотренные виды внешних и внутренних помех.

В реальных каналах имеются и аддитивные, и мультипликативные помехи, поэтому в них

$$Z(t) = S(t) K(t) + n(t).$$

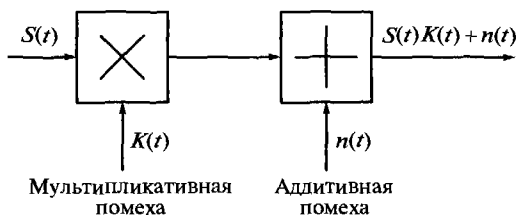


Рис. 1.6. Схема действия помех в линии связи

Схема действия помех в линии связи показана на рис. 1.6.

В заключение отметим, что между сигналом и помехой отсутствует принципиальное различие. Более того, они существуют как единое целое, хотя и противоположны по своему действию. Например, излучение передатчика радиостанции, являясь полезным для приемника того абонента, которому оно предназначено, одновременно может служить помехой для приемников тех абонентов, которым оно не предназначено.

1.4. Основные характеристики систем передачи информации

Как уже отмечалось, задачей любой системы передачи информации является доставка сообщения от источника (абонента) потребителю (другому абоненту), находящихся при этом в разных местах. Отличие принимаемого сообщения от переданного определяется степенью искажения соответствующего сигнала вследствие воздействия на него помех. Поэтому для оценки качества передачи и приема информации введено понятие помехоустойчивости — способности системы передачи информации противостоять вредному воздействию помех. При расчете помехоустойчивости используется количественная оценка степени соответствия принятого сообщения переданному с помощью величины, называемой *верностью*, или *достоверностью*.

При передаче дискретных сообщений степень соответствия обычно определяется *вероятностью ошибки*, обозначаемой символом P . Нахождение расчетных соотношений для определения вероятности ошибки в теории электрической связи является иногда сложнейшей задачей. Эти соотношения являются функциональными зависимостями P от отношения сигнал/помеха:

$$P = f(q = P_c / P_n),$$

где P_c — мощность сигнала; P_n — мощность помехи.

При передаче непрерывных сообщений степень соответствия (достоверность) определяется значением среднеквадратичного отклонения принятого сообщения от переданного:

$$\varepsilon^2 = \int_0^T [u(t) - \tilde{u}(t)]^2 dt,$$

где $u(t)$, $\tilde{u}(t)$ — соответственно переданное и принятое сообщения.

Показатель достоверности — вероятность ошибки — определяет качество передачи информации. Однако для оценки качества применяется также и другой важнейший показатель — скорость передачи информации.

Ранее была введена единица технической скорости — бод. Наряду с ней при расчете скорости передачи информации от двоичного источника применяется единица $R_{\text{и}} = 1/T$, бит/с, где T — длительность посылки. Если код не является двоичным, т. е. его основание $m > 2$, то $R_{\text{и}} = \log_2 m / T$.

Наконец, как и для сигнала, введем понятие емкости канала связи. *Емкостью канала* ($V_{\text{к}}$) называется произведение трех параметров: времени ($T_{\text{к}}$), в течение которого ведется передача, полосы пропускания ($F_{\text{к}}$) и динамического диапазона ($D_{\text{к}}$). При проектировании системы передачи информации должно выполняться условие $V_{\text{с}} \leq V_{\text{к}}$. В простейших системах необходимо обеспечить $T_{\text{с}} \approx T_{\text{к}}$, $F_{\text{с}} \approx F_{\text{к}}$ и $D_{\text{с}} \approx D_{\text{к}}$. При невыполнении этих условий появляются искажения.

Контрольные вопросы

1. Дать определение понятий информации, сообщения, сигнала. Привести классификацию и примеры сообщений.
2. Дать определения источника и получателя сообщений. Перечислить виды систем передачи сообщений в соответствии с видом источника и получателя.
3. Дать определение сигнала. Привести простейшую классификацию сигналов. Указать, применяют ли в системах телекоммуникаций для передачи информации детерминированные сигналы и пояснить почему?
4. Дать определение понятий объема сигнала и объема канала. Привести условие передачи информации без искажений.
5. Начертить простейшую схему системы передачи информации, сформулировать определение канала связи, пояснить понятия кодирования и модуляции.
6. Назвать простейшие первичные коды, применяемые в системах телекоммуникаций. Пояснить различие равномерного и неравномерного кодов. Указать, для чего предназначены корректирующие коды.

7. Пояснить, что понимают под демодуляцией и декодированием и в чем состоит их отличие от кодирования и модуляции. В чем заключается смысл двухэтапной процедуры приема сообщений?

8. Пояснить, что понимают под помехоустойчивостью и достоверностью приема дискретных и непрерывных сообщений и как численно они оцениваются.

9. Что в канале связи является помехой? Назвать основные виды помех в каналах радио- и проводной связи и определить, какие помехи являются аддитивными и мультипликативными.

Глава 2

МОДЕЛИ СИГНАЛОВ

2.1. Классификация сигналов

В теории электрической связи при описании сигналов и помех возникает задача поиска математических моделей, которые с необходимой степенью точности отображали бы реальные процессы в каналах передачи информации. Поскольку сигналы являются электрическими колебаниями, изменяющимися во времени, то их базовой математической моделью должна быть временная функция, т. е. функция, значения которой изменяются во времени.

Различают сигналы детерминированные и случайные.

Детерминированным называется сигнал, мгновенные значения которого в любой момент времени можно предсказать с вероятностью, равной единице.

Например, при $\xi \rightarrow 0$ сигнал $v(t, \xi)$ переходит в единичную функцию включения (рис. 2.1):

$$\sigma(t) = \lim_{\xi \rightarrow 0} v(t, \xi) = \begin{cases} 1, & t > 0; \\ 0, & t < 0. \end{cases}$$

Функция включения описывает процесс перехода некоторого физического объекта из нулевого состояния в единичное. Переход осуществляется по линейному закону за время $t = 2\xi$. При $\xi \rightarrow 0$ в пределе переход производится мгновенно. Данная математическая модель получила название функции Хевисайда.

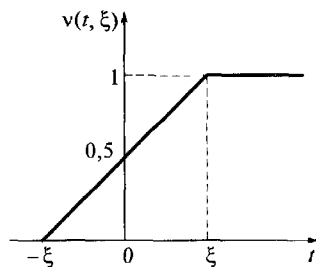


Рис. 2.1. Функция Хевисайда (функция включения)

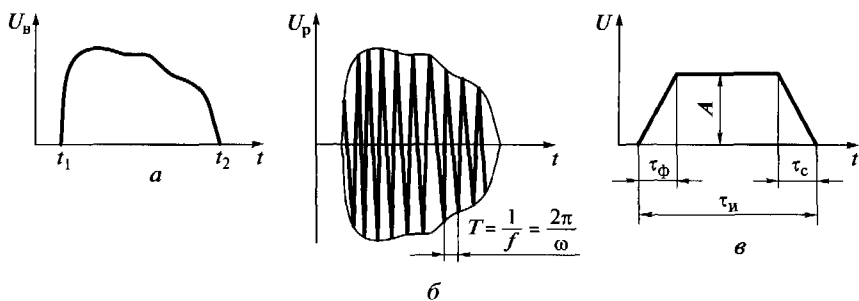


Рис. 2.2. Виды импульсных сигналов и их параметры:

a — видеоимпульс; *б* — радиоимпульс; *в* — основные параметры импульса

Важный класс сигналов представляют собой *импульсные* сигналы, под которыми понимают колебания, существующие только на конечном отрезке (в интервале) времени. При этом различают видеоимпульсы и радиоимпульсы (рис. 2.2).

Видеоимпульсу $U_{в}(t) \geq 0$ соответствует радиоимпульс $U_{р}(t) = U_{в}(t) \cos(\omega t + \varphi)$, где $U_{в}(t)$ — огибающая радиоимпульса; $\cos(\omega t + \varphi)$ — высокочастотное заполнение; ω , φ — соответственно частота и фаза сигнала, имеющие произвольные значения. В ТЭС видеоимпульсы принято задавать амплитудой (A), длительностью импульса ($\tau_{и}$), длительностью фронта ($\tau_{ф}$) и длительностью спада ($\tau_{с}$). Происхождение термина *видеоимпульс* объясняется тем, что данные колебания впервые стали применяться в технике телевидения.

Например, импульсный сигнал прямоугольной формы в пределе при $\xi \rightarrow 0$ переходит в δ -функцию (рис. 2.3, 2.4):

$$\delta(t) = \lim_{\xi \rightarrow 0} \eta(t, \xi) = \begin{cases} \infty, & t = 0; \\ 0, & t \neq 0. \end{cases}$$

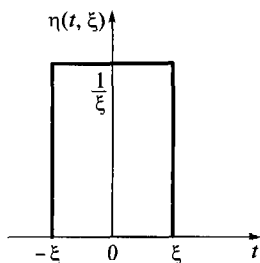


Рис. 2.3. Прямоугольный видеоимпульс



Рис. 2.4. Определение δ -функции

Дельта-функция (функция Дирака) везде равна нулю за исключением точки $t = 0$, в которой

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

Дельта-функция обладает важным свойством, называемым *фильтрующим*:

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

Физический смысл фильтрующего свойства заключается в том, что интеграл по времени от произведения непрерывной функции и дельта-функции равен значению этой непрерывной функции в точке, где сосредоточен δ -импульс.

Соответствующая схема измерения мгновенных значений аналогового сигнала в точке t_0 показана на рис. 2.5. Точность измерения $S(t_0)$ зависит от ширины прямоугольного видеоимпульса, который приближенно представляет собой δ -функцию, широко применяемую в теории электрической связи.

Детерминированные сигналы можно подразделить на периодические и непериодические. Рассмотренные ранее сигналы являются непериодическими.

Периодическим называется повторяющийся во времени сигнал, для которого справедливо условие

$$S(t) = S(t \pm kT),$$

где k — любое целое число; T — период, являющийся конечным отрезком.

Например, периодическим является гармонический сигнал $S(t) = U \cos(\omega t + \varphi)$, $-\infty < t < \infty$. Здесь U , $\omega = 2\pi/T$; φ — соответственно постоянные амплитуда, угловая частота и начальная фаза. Такой сигнал широко применяется в качестве переносчика информации при формировании модулированных сигналов.

Случайными называются сигналы, значения которых заранее неизвестны и могут быть предсказаны только с некоторой вероятностью, меньшей единицы. Например, электрические колебания, соответствующие речевому сигналу, музыке или последова-

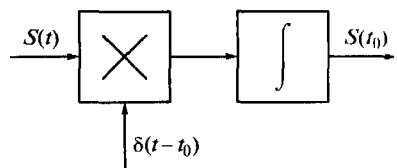


Рис. 2.5. Схема измерения мгновенных значений произвольного аналогового сигнала $S(t)$

тельности знаков телеграфного кода при передаче неповторяющегося текста.

Отметим, что любой сигнал, несущий в себе информацию, должен быть обязательно случайным. Приведенные ранее детерминированные сигналы, являющиеся полностью известными, информации не переносят.

Сигналы также подразделяются на одномерные и многомерные. *Одномерным* называется сигнал, который описывается одной функцией времени. Примером такого сигнала является напряжение на зажимах цепи или ток в ветви.

Многомерным (векторным) называется сигнал, образованный некоторым упорядоченным множеством одномерных сигналов $U(t) = \{U_1(t), U_2(t), \dots, U_N(t)\}$, где N — размерность сигнала. Примером такого сигнала может служить последовательность посылок на выходе кодера, т. е. кодовая комбинация, или система напряжений на зажимах многополюсника.

Наконец, сигналы могут быть *дискретными* и *непрерывными*, как во времени t , так и по уровню, о чем уже говорилось при рассмотрении сообщений.

2.2. Сигналы как элементы функциональных пространств

В ТЭС при решении задач возникают, в первую очередь, следующие вопросы.

1. Какой величиной можно охарактеризовать сигнал?
2. Каким образом любой сигнал можно представить в виде совокупности более простых элементов?
3. Как сравнивать между собой разные сигналы?

Эти и многие другие вопросы можно решить с помощью методов функционального анализа, основанных на использовании понятия пространств с различными свойствами.

В современной математике *пространство* (Π) определяется как множество каких-либо объектов, называемых его точками (это могут быть векторы, функции, состояния физической системы, геометрические фигуры и т. д.), причем между этими точками заданы некоторые соотношения, определяющие геометрию пространства. В соответствии с этими соотношениями различают линейные, метрические и нормированные пространства. Существуют также пространства, обладающие сразу несколькими свойствами. Например, евклидово пространство R_n представляет собой линейное метрическое пространство с размерностью n , где элементами являются векторы, и норма вводится через скалярное произведение этих векторов.

Дадим определения линейного пространства (ЛП), метрики и нормы.

Обозначим x, y, z точки или элементы пространства:

- 1) $x = \bar{x}, y = \bar{y}, z = \bar{z}$ — векторы (действительные или комплексные);
- 2) $x = x(t), y = y(t), z = z(t)$, — функции (действительные или комплексные);
- 3) другие элементы.

Линейным (L) называется пространство, в котором выполняются следующие правила сложения элементов и умножения их на скаляр α .

А. Для любых $x, y \in L$ однозначно определяется третий элемент, принадлежащий этому же пространству $z \in L$ и называемый суммой x и y , для которого справедливы следующие условия:

- 1) $x + y = y + x$ — коммутативность;
- 2) $x + (y + z) = (x + y) + z$ — ассоциативность;
- 3) $x + 0 = x$ для всех $x \in L$, где 0 — нулевой элемент;
- 4) $x + (-x) = 1$ для всех $x \in L$, где x — противоположный элемент, 1 — единичный элемент.

Б. Имеется множество элементов, называемых скалярами, и введена операция, называемая умножением на скаляр, при которой для любого α и любого $x \in L$ определен элемент $\alpha x \in L$, при этом:

- 1) $\alpha(\beta x) = (\alpha\beta)x$;
 - 2) $1x = x$;
 - 3) $(\alpha + \beta)x = \alpha x + \beta x$;
 - 4) $\alpha(x + y) = \alpha x + \alpha y$.
- } дистрибутивность

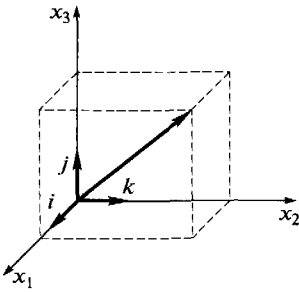
Например, прямая линия или совокупность действительных чисел с обычными арифметическими операциями сложения и умножения представляют собой линейное пространство R_1 , а множество всех аналоговых сигналов, отличных от нуля в интервале времени от t_1 до t_2 и равных нулю вне этого интервала, — линейное пространство L_2 .

Линейное пространство, как и обычное трехмерное пространство, имеет некоторую систему координат (СК). В качестве СК используется конечная или бесконечная совокупность элементов — векторов или функций $\{x_1, x_2, \dots, x_n\}$, принадлежащих пространству L и являющихся линейно независимым координатным

базисом, т. е. базисом, для которого равенство $\sum_{i=1}^n \alpha_i x_i = 0$ возмож-

но лишь в случае, если все $\alpha_i = 0$. Это равенство означает, что в линейно независимом множестве элемент является линейной комбинацией других элементов.

Рис. 2.6. Разложение вектора на составляющие в трехмерном пространстве



Любой элемент линейного пространства можно представить в виде разложения по координатному базису $\{x_i\}$:

$$x = \sum_{i=1}^n c_i x_i, \quad (2.1)$$

где c_i — проекции элемента x относительно выбранного базиса.

В случае когда элементами пространства являются векторы (действительные или комплексные), разложение (2.1) имеет вид

$$\bar{x} = \sum_{i=1}^n c_i \bar{x}_i. \quad (2.2)$$

Если же элементы линейного пространства представляют собой функции (действительные или комплексные), выражение (2.1) будет иметь вид

$$x(t) = \sum_{i=1}^n c_i x_i(t). \quad (2.3)$$

Число n называется размерностью линейного пространства. Любое множество n линейно независимых векторов в ЛП может служить его базисом, следовательно, ЛП имеет не один базис.

Например, вектор в обычном трехмерном пространстве можно записать следующим образом:

$$x = x_1 \bar{i} + x_2 \bar{j} + x_3 \bar{k},$$

где x_1, x_2, x_3 — проекции вектора \bar{x} на оси координат, называемые его координатами; $\bar{i}, \bar{j}, \bar{k}$ — взаимно-перпендикулярные единичные векторы (координатный базис, орты или оси координат); $x_1 \bar{i}, x_2 \bar{j}, x_3 \bar{k}$ — составляющие вектора \bar{x} (рис. 2.6).

Следующей ступенью в усовершенствовании структуры пространства является установление расстояния между его элементами, которое также называется **метрикой**. Иными словами, каждой паре элементов пространства ставится в соответствие

положительное число, которое трактуется как расстояние между элементами. В качестве расстояния используется функционал $d(x, y) = R$ — метрика, обладающая следующими свойствами:

- 1) $d(x, y) \geq 0$ и $d(x, y) = 0$ только в случае, если $x = y$;
- 2) $d(x, y) = d(y, x)$ — симметрия;
- 3) $d(x, y) < d(x, z) + d(z, y)$ — неравенство треугольника.

Эти свойства интуитивно понятны, так как известно:

- расстояние — это неотрицательная величина;
- расстояние от x до y равно расстоянию от y до x ;
- длина одной стороны треугольника не может превосходить сумму длин двух других его сторон.

Например, множество действительных чисел с расстоянием $d(x, y) = |x - y|$ является метрическим пространством.

Объединим геометрические свойства, характерные для метрических пространств, и алгебраические свойства линейных пространств. Это можно сделать, если определить действительное число, характеризующее «размер» элемента в ЛП. Такое число называется *нормой* элемента и обозначается $\|x\|$.

Норма должна удовлетворять следующим аксиомам:

- 1) $\|x\| \geq 0$;
- 2) $\|\alpha x\| = |\alpha| \|x\|$;
- 3) $\|x + y\| \leq \|x\| + \|y\|$.

С учетом аксиомы о наличии у каждого элемента противоположного ему элемента в качестве метрики можно выбрать величину $d(x, y) = \|x - y\|$. Тогда ЛП становится нормированным линейным пространством. Заметим, что если x — это вектор, то норма — это его длина, или расстояние от конца вектора до начала координат.

Последним шагом в развитии структуры пространства является введение дополнительной геометрической характеристики — скалярного произведения двух элементов (x, y) . Определим это произведение для комплексного линейного пространства. Тогда скалярное произведение для действительного пространства будет его частным случаем.

Скалярное произведение должно удовлетворять следующим условиям:

- 1) $(x, y) = (y, x)^*$, где звездочкой обозначена комплексно сопряженная величина;
- 2) $(\alpha x, y) = \alpha(x, y)$;
- 3) $(x_1 + x_2, y) = (x_1, y) + (x_2, y)$;
- 4) $(x, x) \geq 0$.

Важным следствием приведенного определения скалярного произведения является то, что величина $\|x\| = (x, x)^{\frac{1}{2}}$ является нормой в линейном пространстве.

С другой стороны, $d(x, y) = \|x - y\| = (x - y, x - y)^{\frac{1}{2}}$, т.е. скалярное произведение определяет норму, которая в свою очередь определяет метрику. Следовательно, при введении указанной метрики пространство со скалярным произведением становится метрическим.

Линейное метрическое пространство, в котором норма вводится с помощью скалярного произведения, называется **евклидовым пространством**, или **пространством Евклида**.

В заключение отметим одно важное свойство. Два элемента пространства x и y называются ортогональными только в том случае, если их скалярное произведение равно нулю, т.е. $(x, y) = 0$.

Рассмотрим пространства, применяемые в теории электрической связи.

Пространство Евклида (R_n). Элементами данного пространства являются либо вещественные векторы $\bar{x} = x_1, x_2, \dots, x_n$, либо комплексные векторы $\dot{\bar{x}} = \dot{x}_1, \dot{x}_2, \dots, \dot{x}_n$ с размерностью n . Соответственно в первом случае имеем вещественное пространство Евклида (ВПЕ), а во втором — комплексное пространство Евклида (КПЕ).

Разложение вектора по координатному базису в ВПЕ имеет вид (2.2).

Скалярное произведение векторов \bar{x} и \bar{y} в этом случае

$$(\bar{x}, \bar{y}) = \sum_{i=1}^n x_i y_i,$$

норма вектора определяется соотношением

$$\|\bar{x}\| = \sqrt{(\bar{x}, \bar{x})} = \sqrt{\sum_{i=1}^n x_i^2},$$

а метрика вычисляется как

$$d(\bar{x}, \bar{y}) = \|\bar{x} - \bar{y}\| = \sqrt{((\bar{x} - \bar{y}), (\bar{x} - \bar{y}))} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

В КПЕ скалярное произведение двух комплексных векторов $\dot{\bar{x}} = \dot{x}_1, \dot{x}_2, \dots, \dot{x}_n$ и $\dot{\bar{y}} = \dot{y}_1, \dot{y}_2, \dots, \dot{y}_n$ запишем в виде

$$(\dot{\bar{x}}, \dot{\bar{y}}) = \sum_{i=1}^n \dot{x}_i \dot{y}_i^*,$$

где * — обозначение комплексного сопряжения.

Норма в этом случае имеет вид

$$\|\dot{\vec{x}}\| = \sqrt{(\dot{\vec{x}}, \dot{\vec{x}})} = \sqrt{\sum_{i=1}^n \dot{x}_i \dot{x}_i^*},$$

а метрика определяется выражением

$$d(\dot{\vec{x}}, \dot{\vec{y}}) = \|\dot{\vec{x}} - \dot{\vec{y}}\| = \sqrt{((\dot{\vec{x}} - \dot{\vec{y}}), (\dot{\vec{x}} - \dot{\vec{y}}))} = \sqrt{\sum_{i=1}^n (\dot{x}_i - \dot{y}_i)(\dot{x}_i - \dot{y}_i)^*}.$$

Например, если дано КПЕ с размерностью

$$n = 1, \text{ а } \dot{\vec{x}} = \dot{x} = x_1 + j\hat{x}_1 \text{ и } \dot{\vec{y}} = \dot{y} = y_1 + j\hat{y}_1,$$

где $\hat{x}_1; \hat{y}_1$ — произвольные точки пространства, норму и метрику элемента в данном пространстве соответственно можно определить следующим образом:

$$\begin{aligned} \|\dot{x}_1\| &= \sqrt{(x_1 + j\hat{x}_1)(x_1 - j\hat{x}_1)} = \sqrt{x_1^2 + \hat{x}_1^2}; \\ d(\dot{x}_1, \dot{y}_1) &= \|\dot{x}_1 - \dot{y}_1\| = \sqrt{(x_1 - y_1)^2 + (\hat{x}_1 - \hat{y}_1)^2}. \end{aligned}$$

Пространство Гильберта. При стремлении n к бесконечности ВПЕ и КПЕ преобразуются соответственно в бесконечномерные вещественное пространство Гильберта (ВПГ) и комплексное пространство Гильберта (КПГ). Элементами этих пространств являются вещественные $x(t)$ или комплексные $\dot{x}(t)$ функции времени, которые могут быть заданы либо в конечном интервале времени (например, $t \in \left[-\frac{T}{2}, \frac{T}{2}\right]$ или $t \in [0, T]$), либо в бесконечном интервале времени: $t \in (-\infty, \infty)$. Пространства Гильберта при этом соответственно обозначаются $L_2(T)$ или $L_2(\infty)$.

Рассмотрим случай ограниченного интервала времени T . Тогда для ВПГ $L_2(T)$ скалярное произведение определяется в виде

$$(x(t), y(t)) = (x, y) = \int_{-T/2}^{T/2} x(t)y(t)dt, \quad (2.4)$$

норма

$$\|x(t)\| = \|x\| = \sqrt{\int_{-T/2}^{T/2} x^2(t)dt}, \quad (2.5)$$

а метрика (расстояние между точками ВПГ) задается соотношением

$$d(x(t), y(t)) = d(x, y) = \|x(t) - y(t)\| = \sqrt{\int_{-T/2}^{T/2} (x(t) - y(t))^2 dt}. \quad (2.6)$$

Пусть $x(t)$ — напряжение на резисторе с сопротивлением 1 Ом. Тогда $x^2(t)$ — мгновенная мощность, а квадрат нормы — это энергия, выделяемая на резисторе за время T .

Заметим, что если в ВПГ сигналы $x(t)$ и $y(t)$ ортогональны, то их скалярное произведение равно нулю:

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

Для КПГ характеристики пространства вводятся следующим образом.

Скалярное произведение

$$(\dot{x}(t), \dot{y}(t)) = (\dot{x}, \dot{y}) = \int_{-T/2}^{T/2} \dot{x}(t)\dot{y}^*(t)dt. \quad (2.7)$$

Если в качестве $\dot{x}(t)$ и $\dot{y}(t)$ рассматриваются комплексные сигналы, то при равенстве скалярного произведения этих сигналов нулю их считают ортогональными в усиленном смысле. Впервые это понятие ввел Л. М. Финк.

В каналах связи со случайной фазой применение сигналов, ортогональных в усиленном смысле, обеспечивает наибольшую помехоустойчивость (см. гл. 8). Это объясняется следующим свойством этих сигналов: при сдвиге фаз составляющих какого-либо сигнала на произвольный угол их ортогональность сохраняется.

Норма в КПГ определяется следующим образом:

$$\|\dot{x}(t)\| = \|\dot{x}\| = \sqrt{(\dot{x}, \dot{x})} = \sqrt{\int_{-T/2}^{T/2} \dot{x}(t)\dot{x}^*(t)dt} = \sqrt{\int_{-T/2}^{T/2} |\dot{x}(t)|^2 dt},$$

а метрика задается выражением

$$d(\dot{x}(t), \dot{y}(t)) = d(\dot{x}, \dot{y}) = \|\dot{x} - \dot{y}\| = \sqrt{\int_{-T/2}^{T/2} (\dot{x}(t) - \dot{y}(t))(\dot{x}^*(t) - \dot{y}^*(t))dt}.$$

Рассмотрим теперь характеристики пространства $L_2(\infty)$ в случае, когда $t \in (-\infty, \infty)$. Возьмем КПГ, в котором элементы пространства — это комплексные функции, заданные в бесконечном интервале времени. Очевидно, что соотношение характеристик ВПГ определяется соотношением характеристик КПГ, если перейти от комплексных функций к вещественным. Здесь возможны два варианта:

1) $\int_{-\infty}^{\infty} |\dot{x}(t)|^2 dt < \infty$ — это функции, интегрируемые с квадратом (норма и энергия ограничены);

2) $\int_{-\infty}^{\infty} |\dot{x}(t)|^2 dt = \infty$ — это функции, неинтегрируемые с квадратом (норма и энергия неограничены).

Второй вариант возможен, например, для периодического сигнала, у которого энергия бесконечна.

В первом варианте скалярное произведение, норма и метрика определяются выражениями (2.4), (2.5), (2.6), если в интегралах, входящих в эти выражения, конечные пределы $\left[-\frac{T}{2}, \frac{T}{2}\right]$ заменить на бесконечные $(-\infty, \infty)$.

Во втором варианте скалярное произведение определяется следующим образом:

$$(\dot{x}(t), \dot{y}(t)) = (\dot{x}, \dot{y}) = \lim_{T \rightarrow \infty} 1/T \int_{-T/2}^{T/2} \dot{x}(t)\dot{y}(t)^* dt,$$

а норма и метрика вводятся через скалярное произведение:

$$\|\dot{x}(t)\| = \|\dot{x}\| = \sqrt{(\dot{x}, \dot{x})} = \sqrt{\lim_{T \rightarrow \infty} 1/T \int_{-T/2}^{T/2} |\dot{x}(t)|^2 dt};$$

$$d(\dot{x}(t), \dot{y}(t)) = d(\dot{x}, \dot{y}).$$

Разложение произвольного сигнала $x(t)$ по базису пространства $L_2(t)$ или $L_2(\infty)$ имеет вид (2.3).

Пример 2.1. Вычислим норму радиоимпульса с прямоугольной формой огибающей (рис 2.7).

Импульс существует на отрезке времени $t \in (0, T)$ и описывается функцией $S(t) = U_c \cos(\omega_0 t + \varphi_0)$.

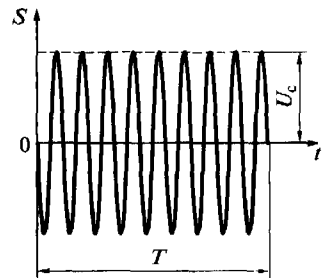


Рис. 2.7. Радиоимпульс с прямоугольной формой огибающей

В соответствии с выражением (2.5), учитывая пределы интегрирования $(0, T)$, запишем

$$\begin{aligned}\|S(t)\| &= \sqrt{U_c^2 \int_0^T \cos^2(\omega_0 t + \varphi_0) dt} = \sqrt{\frac{U_c^2}{\omega_0} \int_0^{\omega_0 T + \varphi_0} \cos^2 x dx} = \\ &= \sqrt{\frac{U_c^2}{4\omega_0} [2(\omega_0 T + \varphi_0) + \sin 2(\omega_0 T + \varphi_0)]}.\end{aligned}$$

Если длительность импульса намного больше периода высокочастотного колебания, т. е. $\omega_0 T \gg 1$, то независимо от значений ω_0 и φ_0

$$\|S(t)\| \approx \sqrt{\frac{U_c^2 T}{2}} = \sqrt{E}.$$

Пример 2.2. Рассмотрим геометрическую интерпретацию характеристик пространства $L_2(T)$ для четырехпозиционного сигнала с фазовой модуляцией (ФМ4), который можно задать соотношением

$$S_j(t) = \sqrt{\frac{2E}{T}} \cos(\omega_0 t + j \frac{\pi}{2}); \quad j = 0, 1, 2, 3; \quad t \in (-T/2, T/2).$$

Норма этого сигнала

$$\|S_j(t)\| = \sqrt{E},$$

где E — энергия сигнала, постоянная для всех значений j .

Представим сигнал $S_j(t)$ в виде

$$S_j(t) = \sqrt{\frac{2}{T}} \cos(\omega_0 t) \sqrt{E} \cos\left(\frac{\pi j}{2}\right) - \sqrt{\frac{2}{T}} \sin(\omega_0 t) \sqrt{E} \sin\left(\frac{\pi j}{2}\right).$$

Введем базисные функции:

$$x_1(t) = \sqrt{\frac{2}{T}} \cos(\omega_0 t); \quad x_2(t) = \sqrt{\frac{2}{T}} \sin(\omega_0 t).$$

Эти функции являются ортонормированными (т. е. их скалярное произведение равно нулю, а норма равна единице) и линейно независимыми. Им соответствуют взаимно-перпендикулярные оси, которые обозначим \bar{x}_1 и \bar{x}_2 . Разложение по координатному базису для сигнала $S_j(t)$ имеет вид

$$S_j(t) = c_{j1} x_1(t) + c_{j2} x_2(t).$$

Определим величины c_{j1}, c_{j2} для $j = 0, 1, 2, 3$:

$$c_{01} = \sqrt{E}; \quad c_{02} = 0; \quad c_{11} = 0; \quad c_{12} = -\sqrt{E}; \quad c_{21} = -\sqrt{E}; \\ c_{22} = 0; \quad c_{31} = 0; \quad c_{32} = \sqrt{E}.$$

Совокупность из двух чисел $\{c_{j1}, c_{j2}\}$ для $j = 0, 1, 2, 3$ задает вектор $\vec{S}_j = c_{j1}\vec{x}_1 + c_{j2}\vec{x}_2$ в двумерном пространстве (см. рис. 5.4, а).

Точки, соответствующие концам векторов, отмеченные на рис. 5.4, а называются сигнальными, а полученная фигура — соответственно сигнальным созвездием для сигнала ФМ4.

Расстояние между сигналами $S_k(t)$ и $S_l(t)$ определяется непосредственно как расстояние между сигнальными точками: $d_{02} = d_{13} = 2\sqrt{E}$; $d_{01} = d_{12} = d_{23} = d_{30} = \sqrt{2E}$.

Здесь и далее встречается запись амплитуды сигнала в виде $\sqrt{2E/T}$, тогда как в литературе обычно гармоническое колебание определяется выражением

$$S(t) = U_c \cos \omega t,$$

где U_c — максимальное значение амплитуды, которое в $\sqrt{2}$ раз больше ее эффективного значения, т. е. $U_c = \sqrt{2}U_{cp}$.

Если сигнал считать током или напряжением на нагрузке с сопротивлением 1 Ом, то его средняя мощность $P = U_{cp}^2$. Тогда $S(t) = \sqrt{2P} \cos \omega t$. Так как $P = E/T$ (где E — средняя энергия, T — длительность элемента сигнала) окончательно запишем

$$S(t) = \sqrt{2E/T} \cos \omega t.$$

2.3. Обобщенный ряд Фурье

Пусть в комплексном пространстве Гильберта $L_2(T)$, $t \in \left[-\frac{T}{2}, \frac{T}{2}\right]$,

задана полная система комплексных ортогональных функций $\{\psi_i(t)\}$, где $i = 1, 2, \dots, \infty$.

Система ортогональных функций называется полной, или замкнутой, если не существует непрерывной функции, не равной нулю тождественно и ортогональной ко всем функциям системы.

Функции являются попарно ортогональными, т. е. можно записать:

$$(\psi_i(t), \psi_j(t)) = \int_{-T/2}^{T/2} \psi_i(t)\psi_j^*(t) dt = \begin{cases} 0, & i \neq j; \\ h_i = \|\psi_i(t)\|^2, & i = j. \end{cases}$$

Если норма функций $\|\dot{\psi}_i(t)\| = 1$, то функции называются ортонормированными. Можно также показать, что функции $\dot{\psi}_i(t)$ являются линейно-независимыми и, следовательно, образуют координатный базис, при котором справедливо разложение (2.3). В этом случае говорят, что задан ортогональный (или ортонормированный) базис.

Пусть некоторый сигнал $S(t)$ является сигналом с интегрируемым квадратом, который можно представить в виде ряда

$$S(t) = \sum_{i=1}^{\infty} c_i \dot{\psi}_i(t), \quad (2.8)$$

где c_i — коэффициенты разложения в ортогональном базисе $\{\dot{\psi}_i(t)\}$.

Определим коэффициенты разложения c_j , для чего умножим обе части равенства на базисную функцию с произвольным номером j и произведем интегрирование по времени:

$$\int_{-T/2}^{T/2} S(t) \dot{\psi}_j^*(t) dt = \sum_{i=1}^{\infty} c_i \cdot \int_{-T/2}^{T/2} \dot{\psi}_i(t) \dot{\psi}_j^*(t) dt.$$

Поскольку базис является ортогональным, то в правой части равенства останется только один член суммы с номером $i = j$. Тогда

$$\dot{c}_j = \frac{1}{h_j} \int_{-T/2}^{T/2} S(t) \dot{\psi}_j(t) dt = \frac{1}{h_j} (S(t), \dot{\psi}_j(t)). \quad (2.9)$$

Иными словами, коэффициент разложения c_j равен скалярному произведению сигнала $S(t)$ и базисной функции $\dot{\psi}_i(t)$.

Выражение (2.8), в котором коэффициенты \dot{c}_i определяются равенством (2.9) при $i = j$, называется обобщенным рядом Фурье и ему можно дать геометрическую трактовку. Коэффициенты обобщенного ряда Фурье \dot{c}_i являются проекциями вектора сигнала $\dot{S}(t)$ на базисное направление, т. е. на ортогональные оси (единичные орты $\dot{\psi}_i(t)$).

В качестве ортогонального базиса в ТЭС широко применяется система тригонометрических (гармонических) функций $\sin k\omega_0 t$, $\cos k\omega_0 t$. (Подробнее это будет рассмотрено далее.) Кроме того, в качестве примера можно рассмотреть использование ортогональных функций Уолша, представляющих собой систему функций типа прямоугольных волн, первые четыре из которых показаны на рис 2.8. В общем случае функции Уолша могут быть построены на основе известных в математике матриц Адамара. Эти функции удобно задавать на отрезке $t \in [0, T]$, где они равны ± 1 , их принято обозначать $wal(k, \theta)$, где k — номер функции Уолша; $\theta = t/T$ — безразмерное время.

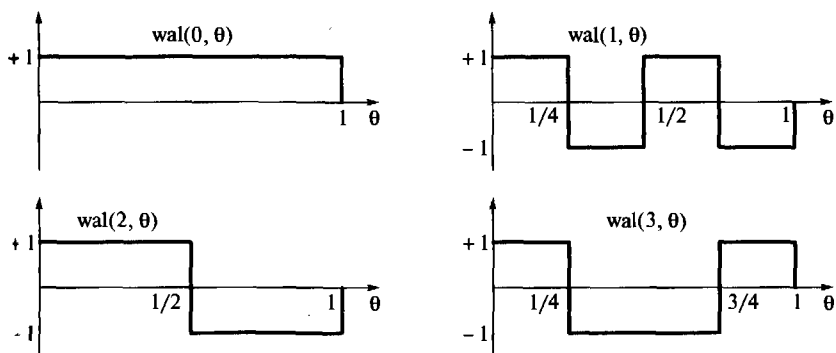


Рис. 2.8. Временные диаграммы первых четырех функций Уолша

Представление сигнала в виде обобщенного ряда Фурье на основе функций Уолша имеет вид

$$S(t) = \sum_{i=1}^{\infty} c_i wal(k, t/T).$$

Для проверки свойства ортогональности найдем скалярное произведение, например первой и второй функций Уолша:

$$\int_0^1 wal(1, \theta) wal(2, \theta) d\theta = \int_0^{1/4} 1 \cdot 1 d\theta + \int_{1/4}^{1/2} 1 \cdot (-1) d\theta + \int_{1/2}^{3/4} (-1) \cdot 1 d\theta + \int_{3/4}^1 (-1)(-1) d\theta = 0.$$

Помимо ортогональности функции Уолша обладают еще свойством *мультипликативности*, т. е. произведение любых двух функций Уолша также является функцией Уолша:

$$wal(i, \theta) wal(k, \theta) = wal(p, \theta),$$

где $p = i \oplus k$; \oplus — знак сложения по модулю 2.

Для такого сложения числа i и k представляются в двоичной форме и суммируются следующим образом:

$$0 \oplus 0 = 0; 1 \oplus 1 = 0; 1 \oplus 0 = 1; 0 \oplus 1 = 1.$$

Например: $i = 1 \Rightarrow 001$ (двоичн.), $k = 3 \Rightarrow 011$ (двоичн.), тогда $001 \oplus 011 = 010 \Rightarrow 2$ (десятичн.).

Иными словами, $wal(1, \theta) wal(3, \theta) = wal(2, \theta)$, в чем нетрудно убедиться.

К функциям Уолша можно применять логические операции, поэтому они находят широкое применение в устройствах форми-

рования и цифровой обработки сигналов на базе микропроцессоров. Сигналы на основе функций Уолша используются в цифровых многоканальных системах передачи информации. В настоящее время они также применяются в сотовой связи на основе стандарта CDMA и его модификаций.

2.4. Спектральное представление колебаний

2.4.1. Спектральное представление периодических колебаний

При формировании и обработке сигналов часто приходится иметь дело с колебаниями, описываемыми периодическими функциями вида $S(t) = S(t \pm kT)$, где $k = 1, 2, \dots$; T — период (рис. 2.9).

Зададим на отрезке $\left[-\frac{T}{2}, \frac{T}{2}\right]$ полную систему тригонометрических ортогональных функций:

$$\{\psi_k(t)\} = \begin{cases} 1, \cos \omega_1 t, \cos 2\omega_1 t, \dots, \cos k\omega_1 t; \\ \sin \omega_1 t, \sin 2\omega_1 t, \dots, \sin k\omega_1 t, \end{cases}$$

где $\omega_1 = 2\pi/T$ — основная частота.

Выбор такого базиса наиболее распространен, так как в результате его использования обеспечиваются:

- сравнительная простота формирования гармонических колебаний;
- инвариантность сигналов относительно их преобразований в линейных электрических цепях (ЛЭЦ). При прохождении через ЛЭЦ изменяются только амплитуда и начальная фаза составляющих.

В этом случае для сигнала $S(t)$ обобщенный ряд Фурье будет иметь вид

$$S(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\omega_1 t + b_k \sin k\omega_1 t). \quad (2.10)$$

Здесь коэффициенты разложения

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} S(t) \cos k\omega_1 t dt; \quad b_k = \frac{2}{T} \int_{-T/2}^{T/2} S(t) \sin k\omega_1 t dt; \quad (2.11)$$

$$a_0 = \frac{2}{T} \int_{-T/2}^{T/2} S(t) dt,$$

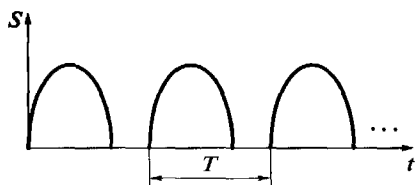


Рис. 2.9. Пример периодической функции

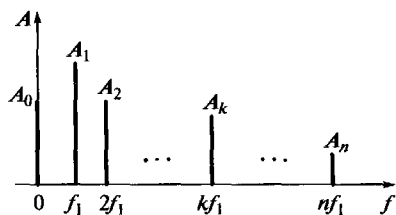


Рис. 2.10. Амплитудная спектральная диаграмма периодического сигнала

откуда видно, что периодический сигнал содержит постоянную составляющую и бесконечный набор гармонических колебаний с частотами $\omega_k = k\omega_1$ ($k = 1, 2, \dots$), которые являются кратными основной частоте.

Представим коэффициенты разложения в виде

$$a_k = A_k \cos \varphi_k; \quad b_k = A_k \sin \varphi_k,$$

где $A_k = \sqrt{a_k^2 + b_k^2}$ — амплитуда; $\varphi_k = -\operatorname{arctg} \frac{b_k}{a_k}$ — начальная фаза.

Тогда ряд Фурье можно записать в другой эквивалентной форме:

$$S(t) = \frac{A_0}{2} + \sum_{k=1}^{\infty} A_k \cos(k\omega_1 t - \varphi_k). \quad (2.12)$$

Выражение (2.12) можно интерпретировать графически как сумму гармоник с частотами, кратными основной частоте $f_1 = 1/T$, амплитудами A_k и начальными фазами φ_k . При этом различают амплитудный и фазовый спектры.

Амплитудный спектр показан на рис. 2.10.

Рассмотрим теперь комплексную форму записи сигнала в виде ряда Фурье. В этом случае в качестве базисных функций применяются комплексные экспоненты:

$$\{\psi_k(t)\} = \{\exp(jk\omega_1 t)\},$$

где $k = 0, \pm 1, \pm 2, \dots$; $j = \sqrt{-1}$ — мнимая единица.

Такая система также является полной.

Соответственно обобщенный ряд Фурье, известный как ряд Фурье в комплексной форме, приобретает следующий вид:

$$S(t) = \frac{1}{2} \sum_{k=-\infty}^{\infty} \dot{A}_k e^{jk\omega_1 t}. \quad (2.13)$$

Здесь комплексная амплитуда k -й гармоники

$$\dot{A}_k = \frac{2}{T} \int_{-T/2}^{T/2} S(t) e^{-jk\omega_1 t} dt \quad (2.14)$$

связана с A_k , a_k , b_k , φ_k следующими выражениями:

$$\begin{aligned}\dot{A}_k &= A_k e^{j\varphi_k} = a_k - jb_k; \\ \dot{A}_{-k} &= A_k e^{-j\varphi_k} = a_k + jb_k; \\ a_k &= \frac{1}{2}(\dot{A}_k + \dot{A}_{-k}); \quad b_k = j\frac{1}{2}(\dot{A}_k - \dot{A}_{-k}).\end{aligned}$$

Из этих выражений следует, что фазовый угол φ_k является нечетной функцией относительно k , т.е. относительно частоты, а модуль комплексной амплитуды A_k — четной.

Как видно из соотношения (2.13), суммирование распространено на положительные (k) и отрицательные ($-k$) частоты. Последние имеют фиктивный характер. Действительную составляющую $A_k \cos(k\omega_1 t + \varphi_k)$ сигнала с частотой $k\omega_1$ получают как сумму двух комплексных функций:

$$\frac{1}{2} A_k e^{j(k\omega_1 t + \varphi_k)} + \frac{1}{2} A_k e^{-j(k\omega_1 t + \varphi_k)} = A_k \cos(k\omega_1 t + \varphi_k),$$

т.е. происходит взаимное уничтожение мнимых частей.

Отметим также, что спектр периодического сигнала является дискретным и бесконечным. С увеличением периода следования сигналов разность частот между соседними гармониками уменьшается, т.е. спектр сгущается.

Из формул (2.11) и (2.14) видно, что изменение периода T сказывается на значениях амплитуд спектральных составляющих: с увеличением T амплитуды уменьшаются. Однако форма спектра амплитуд при этом сохраняется.

Пример 2.3. Определим спектр периодической последовательности прямоугольных импульсов с амплитудой E , длительностью t_n и периодом следования T .

Итак, пусть

$$S(t) = \begin{cases} E & \text{при } -kT - \frac{t_n}{2} \leq t \leq kT + \frac{t_n}{2}, \quad k = 0, \pm 1, \pm 2, \dots; \\ 0 & \text{при других } t. \end{cases}$$

По формуле (2.14) находим

$$\dot{A}_k = \frac{2}{T} \int_{-t_n/2}^{t_n/2} E e^{-jk\omega_1 t} dt = 2E \frac{t_n}{T} \frac{\sin \frac{k\omega_1 t_n}{2}}{\frac{k\omega_1 t_n}{2}}.$$

Комплексная амплитуда \dot{A}_k пропорциональна функции вида $\frac{\sin x}{x}$ и содержит только вещественную часть, т. е.

$$A_k = 2E \frac{t_n}{T} \left| \frac{\sin \frac{k \omega_1 t_n}{2}}{\frac{k \omega_1 t_n}{2}} \right|.$$

Начальные фазы гармоник следующие: $\varphi = 0$, в интервалах частот $f \in \left[n \frac{1}{t_n}, (n+1) \frac{1}{t_n} \right]$, где $n = 0, 2, 4, \dots$

$\varphi_k = \pm \pi$ в интервалах частот $f \in \left[n \frac{1}{t_n}, (n+1) \frac{1}{t_n} \right]$, $n = 1, 3, 5, \dots$

С учетом этих соотношений ряд Фурье (2.13) для последовательности прямоугольных импульсов можно записать в виде

$$S(t) = E \frac{t_n}{T} \left(1 + 2 \sum_{k=1}^{\infty} \frac{\sin \frac{k \omega_1 t_n}{2}}{\frac{k \omega_1 t_n}{2}} \cos k \omega_1 t \right). \quad (2.15)$$

Пример 2.4. Периодическая последовательность двухполярных импульсов $\pm E$ задана в виде

$$S_1(t) = \begin{cases} E & \text{при } kT - \frac{t_n}{2} \leq t \leq kT + \frac{t_n}{2}, k = 0, \pm 1, \pm 2, \dots; \\ -E & \text{при других } t. \end{cases}$$

Представим $S_1(t)$ как результат суммирования однополярных импульсов с амплитудой $2E$ и постоянной составляющей $-E$, т. е. получим выражение $S_1(t) = 2S(t) - E$.

Подставив в это выражение $S_1(t)$ в виде (2.15), получим ряд Фурье для $S_1(t)$:

$$S_1(t) = E \left(2 \frac{t_n}{T} - 1 + 4 \frac{t_n}{T} \sum_{k=1}^{\infty} \frac{\sin \frac{k \omega_1 t_n}{2}}{\frac{k \omega_1 t_n}{2}} \cos k \omega_1 t \right). \quad (2.16)$$

2.4.2. Спектральное представление непериодических колебаний

Рассмотренные ранее периодические сигналы на практике почти не встречаются. В подавляющем большинстве случаев в теории и технике связи приходится иметь дело с сигналами и помехами, которые по существу являются непериодическими и к которым аппарат рядов Фурье не применим. Поэтому вместо них используют интегралы Фурье. Такое представление получают посредством перехода от ряда Фурье при стремлении периода повторения сигнала к бесконечности, т. е. $T \rightarrow \infty$ или $f_1 = 1/T = \omega_1/(2\pi) \rightarrow 0$.

Рассмотрим вновь ряд Фурье в экспоненциальной форме:

$$S(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \dot{A}_k T e^{jk\omega_1 t} \Delta\omega,$$

где $\Delta\omega = \omega_1 = [k\omega_1 - (k-1)\omega_1]$ — разность между частотами соседних гармоник, $k = 1, 2, \dots$

Определим комплексную спектральную плотность в виде

$$S(j\omega) = \lim_{T \rightarrow \infty} \dot{A}_k T = \lim_{\Delta\omega \rightarrow 0} 2\pi \frac{\dot{A}_k}{\Delta\omega}.$$

При $\Delta\omega \rightarrow 0$ интервалы между соседними гармониками неограниченно сокращаются, поэтому сумму можно заменить интегралом:

$$S(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) e^{j\omega t} d\omega. \quad (2.17)$$

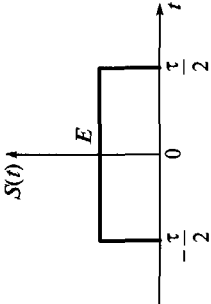
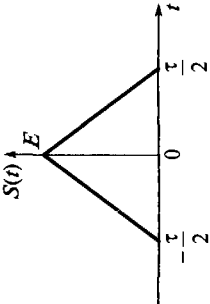
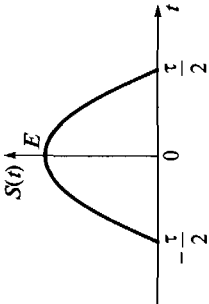
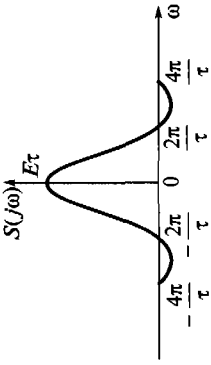
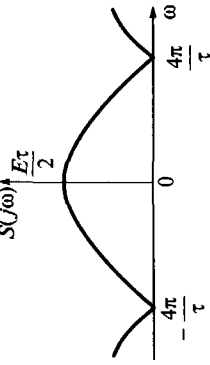
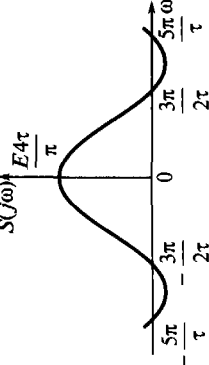
Для нахождения комплексной спектральной плотности $S(j\omega)$, или комплексного спектра, можно использовать выражение

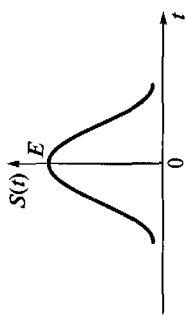
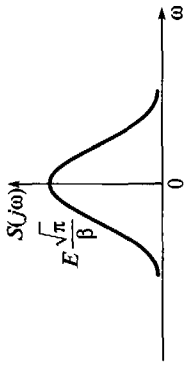
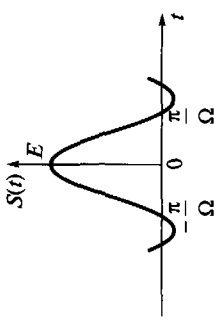
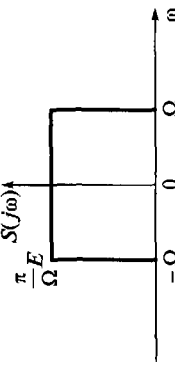
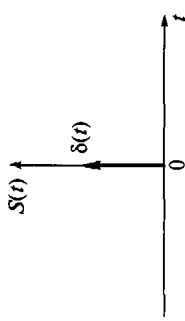
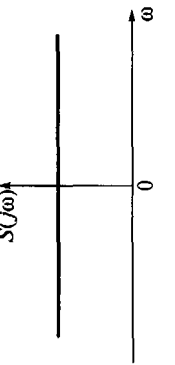
$$S(j\omega) = \int_{-\infty}^{\infty} S(t) e^{-j\omega t} dt, \quad (2.18)$$

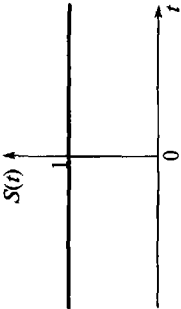
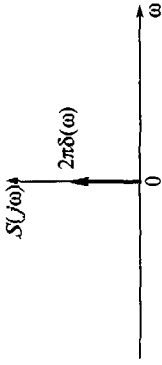
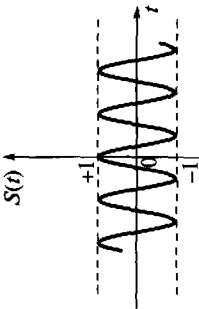
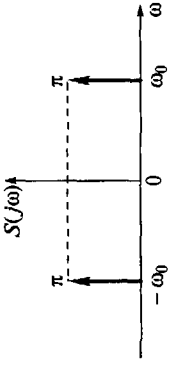
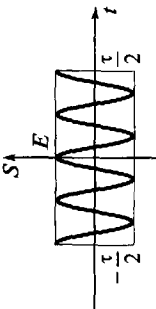
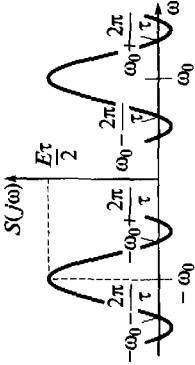
которое следует из формулы (2.14) для комплексной амплитуды и отличается от нее только наличием множителя $2/T$.

Выражения (2.17) и (2.18) имеют фундаментальное значение в ТЭС. Первое из них называется *обратным преобразованием Фурье* для сигнала $S(t)$, или *операцией синтеза*, поскольку с его помощью сигнал восстанавливается (синтезируется) из спектральных составляющих. Второе выражение называется *прямым преобразованием Фурье*, или *операцией анализа* сигнала на основе определения его спектральных составляющих. В символической записи соответствие между сигналом $S(t)$ и его преобразованием Фурье $S(j\omega)$ отображается следующим образом: $S(t) \rightleftharpoons S(j\omega)$.

Таблица 2.1

Сигнал $S(t)$	Спектральная плотность $S(\omega)$
$S_1(t) = \begin{cases} E, & t \leq \frac{\tau}{2}; \\ 0, & t > \frac{\tau}{2} \end{cases}$	
$S(t) = \begin{cases} E \left(1 - \frac{2 t }{\tau}\right), & t < \frac{\tau}{2}; \\ 0, & t > \frac{\tau}{2} \end{cases}$	
$S(t) = \begin{cases} E \cos \omega t, & t < \frac{\tau}{2}; \\ 0, & t > \frac{\tau}{2} \end{cases}$	
$E\tau \frac{\sin\left(\frac{\omega\tau}{2}\right)}{\omega\tau \frac{\tau}{2}}$	
$\frac{E\tau}{2} \frac{\sin^2\left(\frac{\omega\tau}{4}\right)}{\left(\frac{\omega\tau}{4}\right)^2}$	
$E \frac{4\tau}{\pi} \frac{\cos\left(\frac{\omega\tau}{2}\right)}{1 - \left(\frac{\omega\tau}{2}\right)^2}$	

Сигнал $S(t)$	Спектральная плотность $S(j\omega)$
$S(t) = E e^{-\beta^2 t^2}$	$E \sqrt{\frac{\pi}{\beta}} e^{-\left(\frac{\omega}{2\beta}\right)^2}$  
$S(t) = E \frac{\sin \frac{\Omega t}{2}}{\frac{\Omega t}{2}}$	$S(j\omega) = \begin{cases} \frac{\pi E}{\Omega}, & \omega \leq \Omega; \\ 0, & \omega > \Omega \end{cases}$  
$S(t) = \delta(t)$	$S(j\omega) = 1$  

$S(t) = 1$		$S(j\omega) = 2\pi\delta(\omega)$	
$S(t) = \cos \omega_0 t$		$S(j\omega) = \pi[\delta(\omega - \omega_0) + \delta(\omega + \omega_0)]$	
$S(t) = S_1(t) \cos \omega_0 t$ $S_1(t) = \begin{cases} E, & t \leq \frac{\tau}{2} \\ 0, & t > \frac{\tau}{2} \end{cases}$		$S(j\omega) = \frac{E\tau}{2} \frac{\sin\left(\frac{(\omega - \omega_0)\tau}{2}\right)}{(\omega - \omega_0)\tau} + \frac{E\tau}{2} \frac{\sin\left(\frac{(\omega + \omega_0)\tau}{2}\right)}{(\omega + \omega_0)\tau}$	

С учетом четности модуля $S(\omega)$ и нечетности фазы $\varphi(\omega)$ обратное преобразование Фурье (2.17) можно записать в виде

$$S(t) = \frac{1}{\pi} \int_0^{\infty} |S(j\omega)| \cos[\omega t + \varphi(\omega)] d\omega.$$

В соответствии с выражением (2.17) и последней формулой непериодическое колебание $S(t)$ можно рассматривать как сумму комплексных экспоненциальных составляющих $\frac{1}{2\pi} S(j\omega) d\omega e^{j\omega t}$ или сумму гармонических составляющих $\frac{1}{\pi} S(\omega) d\omega \cos[\omega t + \varphi(\omega)]$,

частоты которых располагаются бесконечно близко друг к другу. На основании этого можно говорить о непрерывном, или сплошном, спектре непериодического колебания.

Бесконечно малые амплитуды составляющих определяются выражением $A_m = c |S(j\omega)| d\omega$, где $c = \frac{1}{2\pi}$ для формулы (2.17) или $c = \frac{1}{\pi}$

с учетом четности модуля $S(\omega)$, откуда $|S(j\omega)| = \frac{1}{c} \frac{A_m}{d\omega}$, т.е. $|S(j\omega)|$ отражает спектральную (частотную) плотность амплитуд составляющих. Другими словами, $|S(j\omega)|$ — это амплитуда, отнесенная к бесконечно малой полосе частот $d\omega$.

Особенностью комплексного спектра является его распространение как в положительной, так и в отрицательной области частот. Следовательно, двум составляющим частот ω и $-\omega$ со спектральной плотностью $\frac{1}{2} |S(j\omega)|$ комплексного спектра соответствует одна гармоническая составляющая частоты ω со спектральной плотностью $|S(j\omega)|$.

Сопоставив формулы (2.18) и (2.14) для непериодического колебания и периодического колебания, полученного из исходного непериодического сигнала посредством его периодического продолжения, можно сделать вывод о том, что спектры таких колебаний совпадают по форме и отличаются лишь масштабом.

В табл. 2.1 приведены результаты расчетов $S(j\omega)$ для колебаний (импульсов) различной формы.

2.4.3. Распределение мощности и энергии в спектре колебания. Ширина спектра колебания

При обработке сигналов важно знать, какие составляющие колебания являются энергетически значимыми и какова ширина

спектра колебания. Ответ на эти вопросы дают соотношения Рэля — Парсеваля для периодических и непериодических колебаний.

Среднюю мощность периодического сигнала $S(t)$ в соответствии с выражением (2.10) можно представить в виде

$$P = \frac{1}{T} \int_0^T \left[\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos k\omega_1 t + b_k \sin k\omega_1 t) \right]^2 dt.$$

Возведя в квадрат выражения под знаком суммы, получим следующие виды

$$a_k^2 \cos^2 k\omega_1 t = a_k^2 \left(\frac{1}{2} + \frac{1}{2} \cos 2k\omega_1 t \right);$$

$$b_k^2 \sin^2 k\omega_1 t = b_k^2 \left(\frac{1}{2} - \frac{1}{2} \cos 2k\omega_1 t \right);$$

интегрирование которых дает $\frac{a_k^2 T}{2}$, $\frac{b_k^2 T}{2}$, а интегрирование произведений косинусов и синусов по свойству ортогональности функции — нуль.

Следовательно,

$$P = \left(\frac{a_0}{2} \right)^2 + \frac{1}{2} \sum_{k=1}^{\infty} (a_k^2 + b_k^2) = \left(\frac{A_0}{2} \right)^2 + \frac{1}{2} \sum_{k=1}^{\infty} A_k^2. \quad (2.19)$$

В математике выражение (2.19) называется равенством Рэля — Парсеваля, из которого следует, что средняя мощность периодического колебания равна сумме средних мощностей составляющих его гармоник.

Зависимость $P = f(\omega)$ называется спектром мощности периодического колебания.

Для непериодических колебаний равенство Рэля — Парсеваля имеет вид

$$W = \int_{-\infty}^{\infty} S^2(t) dt = \frac{1}{\pi} \int_0^{\infty} |S(j\omega)|^2 d\omega. \quad (2.20)$$

Левая часть равенства (2.20) определяет энергию колебания.

Следовательно, $\frac{|S(j\omega)|^2}{\pi} = G(\omega)$ — есть не что иное как энергия колебания, приходящаяся на один радиан полосы частот для текущего значения ω . Иными словами, $G(\omega)$ является спектральной плотностью энергии колебания $S(t)$ и характеризует распределение энергии в полосе частот колебания. Спектры мощности и

спектральные плотности энергии называются энергетическими спектрами.

Теоретически реальные колебания как ограниченные во времени имеют бесконечный спектр. Однако рассмотрение спектров различных колебаний (например, представленных в табл. 2.1) показывает, что спектральная плотность амплитуд уменьшается (монотонно или немонотонно) с ростом частоты.

Этот же вывод следует и из анализа равенств Рэля — Парсевалья (2.19) и (2.20), в соответствии с которыми $A_k \rightarrow 0$ при $k \rightarrow \infty$, а также $|S(j\omega)|^2 \rightarrow 0$ при $\omega \rightarrow \infty$. В противном случае ряд (2.19) был бы расходящимся (так как мощность бесконечна) или интеграл (2.20) был бы расходящимся (так как энергия бесконечна). Это позволяет условно говорить о занимаемой полосе частот, или ширине спектра колебания. Наиболее распространена так называемая энергетическая ширина спектра, определяемая как полоса частот ΔF , в пределах которой сосредоточена большая часть средней мощности (энергии) колебаний.

Строгих правил относительно процентного содержания мощности (энергии) в пределах занимаемой полосы частот часто не существует (обычно это 90...99 %). Соответственно, задавая значение $\delta = 0,9 \dots 0,99$, ширину полосы ΔF можно определить с использованием соотношений (2.19) или (2.20).

Так, например, для непериодических колебаний, используя выражение (2.20), можно записать

$$\delta W = \frac{1}{\pi} \int_0^{2\pi\Delta F} |S(j\omega)|^2 d\omega$$

Решив это интегральное уравнение (обычно численно), можно определить ΔF .

2.5. Связь между временными и спектральными характеристиками колебаний

Как видно из примеров, приведенных в табл. 2.1, теоретически сигналы ограниченной длительности имеют бесконечный спектр. Однако практически под шириной спектра принято понимать область частот ΔF , в пределах которой сосредоточено не менее 90 % всей энергии. При этом вводится также понятие эквивалентной длительности τ_e , под которой понимают интервал времени, в пределах которого сосредоточена основная энергия сигнала.

Использование понятия эффективной ширины спектра ΔF и эквивалентной длительности τ_e позволяет выявить зависимость

Таблица 2.2

Импульс	τ_3	ΔF	$\Delta F \tau_3$
Прямоугольный	$0,9\tau$	$0,811/\tau$	0,73
Треугольный	$0,541\tau$	$0,841/\tau$	0,46
Косинусоидальный	$0,596\tau$	$0,731/\tau$	0,43
Колокольный	$0,825 \cdot 1/\beta$	$0,26\beta$	0,22

между полосой частот, занимаемой импульсным колебанием, и формой колебания. Эту зависимость поясняет табл. 2.2, в которой приведены произведения $\Delta F \tau_3$ для различных форм импульсов (ΔF , τ_3 определяются для 90 % полной энергии импульса).

Видно, что произведение $\Delta F \tau_3$ для определенной формы импульса значение постоянное, т.е. $\Delta F \tau_3 = \text{const}$. Это выражение указывает на связь между шириной спектра и длительностью импульса: чем короче импульс, тем шире его спектр.

Второй важный вывод, следующий из табл. 2.2, заключается в том, что произведение $\Delta F \tau_3$ зависит от степени «гладкости» формы импульса, которая характеризуется наличием или отсутствием разрывов функции, описывающей импульс, и ее производных. Например, произведение $\Delta F \tau_3$ является наибольшим для импульса прямоугольной формы, имеющего разрывы как в описывающей функции, так и в ее производных. Наименьшее значение $\Delta F \tau_3$ характерно для импульса колокольной формы (так как функция, описывающая эту форму импульса, непрерывна, как и все ее производные).

Данное положение имеет фундаментальное значение в курсе ТЭС: с одной стороны, оно позволяет установить требования к ширине полосы пропускания приемных устройств (чем короче импульс, тем шире должна быть полоса пропускания), а с другой стороны, из него следует важный вывод о том, что форма используемых колебаний должна иметь высокую степень «гладкости», что широко используется на практике.

2.6. Теоремы о спектрах

Как следует из преобразования Фурье, существует взаимосвязь между действиями, совершаемыми с сигналами во временной области, и их представлениями в частотной области, и наоборот. Эта взаимосвязь формулируется в виде теорем о спектрах, применение которых оказывается весьма полезным на практике.

Таблица 2.3

Название теоремы	Временное представление	Спектральное представление
Теорема сложения	$S(t) = S_1(t) + S_2(t)$	$S(j\omega) = S_1(j\omega) + S_2(j\omega)$
Теорема временного сдвига	$S(t) = S_1(t \pm t_0)$	$S_1(j\omega) = S(j\omega)e^{\pm j\omega t_0}$
Теорема смещения (модуляции)	$S_1(t) = S(t)e^{\pm j\omega_0 t}$	$S_1(j\omega) = S(j(\omega \pm \omega_0))$
Теорема об изменении масштаба	$S_1(t) = S\left(\frac{t}{a}\right)$	$S_1(j\omega) = aS(ja\omega)$
Теорема о дифференцировании	$S_1(t) = \frac{dS(t)}{dt}$	$S_1(j\omega) = j\omega S(j\omega)$
Теорема об интегрировании	$S_1(t) = \int_{-\infty}^t S(t)dt$	$S_1(j\omega) = \left(\frac{1}{j\omega}\right)S(j\omega)$
Теорема о свертке	$S(t) = \int_{-\infty}^{\infty} S_1(\tau)S_2(t-\tau)d\tau = S_1(t)S_2(t)$	$S(j\omega) = S_1(j\omega)S_2(j\omega)$
Преобразования Фурье	$S_1(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega)e^{j\omega t} d\omega$ $S(j\omega) = \int_{-\infty}^{\infty} S(t)e^{-j\omega t} dt$	

В табл. 2.3 представлена формальная трактовка некоторых из этих теорем. В последней строке этой таблицы приведена пара преобразований Фурье. Для каждой теоремы показаны операции с сигналом во временной области и даны соответствующие спектральные представления.

Приведем краткие формулировки теорем о спектрах.

Теорема сложения спектров: спектр суммы колебаний равен сумме спектров слагаемых колебаний.

Теорема временного сдвига (запаздывания): при сдвиге колебания во времени (изменении начального момента отсчета времени) спектральная плотность амплитуд сохраняется постоянной, а спектр фаз изменяется на значение, пропорциональное частоте и времени сдвига с учетом его знака.

Теорема смещения (модуляции): умножение колебания $S(t)$ на $e^{j\omega_0 t}$ приводит к смещению его спектра на значение ω_0 .

Теорема об изменении масштаба: растяжение колебания во времени ($a > 1$) влечет за собой сжатие его частотного спектра и увеличение спектральной плотности амплитуд. Сжатие колебания во времени ($a < 1$) приводит к расширению его частотного спектра и уменьшению спектральной плотности амплитуд.

Теорема о свертке: свертка двух колебаний $S_1(t)$ и $S_2(t)$ соответствует перемножению их спектров.

Остальные приведенные в табл. 2.3 теоремы не требуют пояснения.

2.7. Понятие функции корреляции

Для непериодических сигналов $S_1(t)$ и $S_2(t)$ функция взаимной корреляции определяется следующим выражением:

$$B_{S_1, S_2}(\tau) = \int_{-\infty}^{\infty} S_1(t) S_2(t + \tau) dt,$$

где $\tau > 0$ — некий временной отрезок.

С учетом приведенных ранее свойств преобразования Фурье запишем:

$$S(t \pm \tau) \Leftrightarrow S(j\omega) e^{\pm j\omega\tau};$$

$$S(-t) \Leftrightarrow S^*(j\omega),$$

где $S^*(j\omega)$ — величина, комплексно-сопряженная с $S(j\omega)$.

Можно записать также следующее выражение:

$$B_{S_1, S_2}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_1^*(j\omega) S_2(j\omega) e^{j\omega\tau} d\omega,$$

из которого при $\tau = 0$ несложно получить

$$\int_{-\infty}^{\infty} S_1(t) S_2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_1^*(j\omega) S_2(j\omega) d\omega = (S_1, S_2).$$

Последнее выражение представляет собой *обобщенную формулу Рэлея—Парсеваля*, которая показывает, что скалярное произведение двух комплексных сигналов выражается через их спектральные плотности.

Введенная функция взаимной корреляции позволяет с единых позиций описывать различие сигналов по форме, а также их взаимное расположение на оси времени. Поясним это на примере.

Пусть сигналы $S_1(t)$ и $S_2(t)$ ортогональны, т. е.

$$\int_{-\infty}^{\infty} S_1(t)S_2(t)dt = 0.$$

При прохождении этих сигналов через различные устройства в некоторых случаях оказывается, что сигнал $S_2(t)$ будет сдвинут на время τ относительно сигнала $S_1(t)$, и наоборот. Поэтому взаимокорреляционную функцию можно считать мерой «устойчивости» ортогонального состояния при сдвигах сигналов во времени.

Помимо данной характеристики для количественной оценки степени отличия сигнала $S(t)$ и его же смещенной во времени копии $S(t + \tau)$ вводят автокорреляционную функцию, равную скалярному произведению сигнала и его копии.

В приведенном ранее выражении для $B_{S_1S_2}(\tau)$ заменим $S_1(t) = S_2(t) = S(t)$, тогда

$$B_S(\tau) = \int_{-\infty}^{\infty} S(t)S(t+\tau)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(j\omega)|^2 e^{j\omega\tau} d\omega.$$

При $\tau = 0$ получим равенство Парсеваля для непериодических сигналов:

$$B_S(0) = \int_{-\infty}^{\infty} S^2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(j\omega)|^2 d\omega = W.$$

Поясним физический смысл этого равенства: если считать $S(t)$ напряжением или током на нагрузке 1 Ом, то полная энергия сигнала E_S , развиваемая на этой нагрузке, будет равна энергии всех его частотных составляющих.

Отметим кратко свойства автокорреляционной функции.

1. Ее четность, т. е. $B_S(\tau) = B_S(-\tau)$. Заметим также, что функция взаимной корреляции не является четной относительно аргумента τ , т. е. $B_{S_1S_2}(\tau) \neq B_{S_1S_2}(-\tau)$.

2. При любом значении временного сдвига τ модуль автокорреляционной функции не превосходит энергии сигнала: $|B_S(\tau)| \leq B_S(0) = E_S$.

Автокорреляционная функция имеет вид симметричной кривой с центральным максимумом, который всегда положительный. В зависимости от вида сигнала $S(t)$ эта кривая может быть монотонно убывающей или колеблющейся.

2.8. Ряд Котельникова. Теорема отсчетов

Из рассмотренных ранее преобразований Фурье следует, что для восстановления сигнала по его спектру необходимо учитывать все составляющие с частотами, находящимися в интервале от нуля до бесконечности. Однако физически такая процедура нереальна. Кроме того, вклад составляющих при $\omega \rightarrow 0$ пренебрежимо мал, поскольку энергия сигнала является ограниченной. Наконец, любое устройство для передачи и приема сигналов имеет конечную ширину полосы пропускания. Поэтому обычно рассматриваются сигналы, спектральная плотность которых отличается от нуля в некоторой ограниченной полосе частот.

С другой стороны, используемые в системах телекоммуникаций линии радио- и проводной связи также имеют ограничения частотного ресурса. Следовательно, существует серьезная проблема: как в ограниченном диапазоне частот поместить все увеличивающееся множество сигналов, спектры которых состоят из компонент, сосредоточенных в некоторой сравнительно узкой полосе частот. Одним из возможных путей решения данной проблемы может быть передача не самих полосовых сигналов, а лишь их выборки или дискретных отсчетов, содержащих информацию обо всем сигнале. Однако при этом возникает естественный вопрос: можно ли это сделать и с каким временным интервалом должны производиться эти выборки, чтобы не потерять даже части передаваемой информации? Ответ на этот вопрос содержится в фундаментальной работе «О пропускной способности «эфира» и проволоки в электросвязи», которая появилась в 1933 г. и принадлежала 25-летнему инженеру В. А. Котельникову. Именно ему в этой работе удалось сделать то, что предопределило перспективы развития теории и техники телекоммуникаций на многие годы вперед.

Отметим, что похожие вопросы исследовали и другие специалисты и ученые. Так, в 1915 г. Е. Т. Уиттакер, занимаясь математической теорией интерполяции, доказал теорему, посвященную проблеме аппроксимации целых функций конечной степени. В математике это была одна из рядовых теорем, и к теории связи она не имела никакого отношения. В 1928 г. американец Найквист, решая задачу передачи телеграфных (дискретных) сообщений без искажений, обратил внимание на возможность взятия выборок сигнала через интервалы времени, обратно пропорциональные ширине его спектра. Однако решаемая им проблема отличалась от проблемы передачи аналоговых полосовых сигналов, рассматривавшихся В. А. Котельниковым, хотя здесь и имелись общие моменты. Наконец, в 1948 г. К. Э. Шеннон вновь доказал теорему, повторяющую результаты работы В. А. Котельникова, ко-

торию опубликовали весьма ограниченным тиражом, и она могла быть неизвестна за границей.

В настоящее время специалистам связи, радиотехники и других смежных областей науки и техники хорошо знакома *теорема отсчетов*, нередко именуемая в отечественной литературе теоремой Котельникова. В некоторых источниках ее также называют теоремой отсчетов (выборки) Уиттакера — Котельникова — Шеннона. На самом деле В. А. Котельников доказал семь теорем, две из которых являются основополагающими, а остальные их дополняют и конкретизируют. Их значение так велико, что по сути они являются основой современной теории и практики электрической связи, на которой построены многие системы телекоммуникаций.

Приведем формулировки основных теорем Котельникова.

Теорема 1. Любую функцию $S(t)$, состоящую из частот от 0 до F_B периодов в секунду, можно представить рядом

$$S(t) = \sum_{k=-\infty}^{\infty} S(k\Delta t) \frac{\sin 2\pi F_B(t - k\Delta t)}{2\pi F_B(t - k\Delta t)}, \quad (2.21)$$

где k — целые числа; $S(k\Delta t)$ — постоянные, зависящие от $S(t)$; F_B — верхняя частота спектра.

И наоборот, любая функция $S(t)$, представленная этим рядом, состоит лишь из частот от 0 до F_B периодов в секунду.

Выражение (2.21) принято называть рядом Котельникова. Если обратиться к обобщенному ряду Фурье (2.8), то легко увидеть, что в выражении (2.21) базисными по существу являются отсчетные функции

$$\psi_k(t) = \frac{\sin 2\pi F_B(t - k\Delta t)}{2\pi F_B(t - k\Delta t)},$$

представленные на рис. 2.11, а коэффициенты разложения по данному базису $c_k = S(k\Delta t)$.

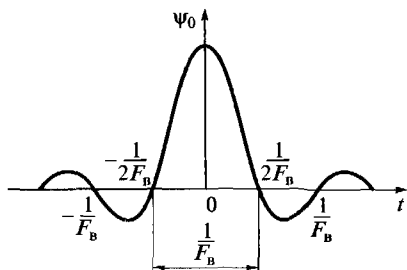


Рис. 2.11. График функции отсчетов

Теорема 2. Любую функцию (сигнал) $S(t)$, состоящую из частот от 0 до F_b периодов в секунду, можно непрерывно передавать с любой точностью с помощью чисел, следующих друг за другом через интервалы времени $\Delta t = 1/(2F_b)$ секунд.

Отметим, что промежутки времени, через которые берутся отсчеты, получили название *интервалов Найквиста*. Доказательство данной теоремы В.А. Котельников построил на следующих рассуждениях. Если измерять величину $S(t)$ при $t = n/(2F_b)$ где n — целое число, то можно записать

$$S\left(\frac{n}{2F_b}\right) = S(n\Delta t)\omega_b;$$

$$\omega_b = 2\pi F_b.$$

Все члены ряда (2.21) для данного значения t обращаются в нули. Исключение составляет член с номером $k = n$, который можно вычислить путем раскрытия неопределенности. Так как он равен величине $S(n\Delta t)\omega_b$, значит, через каждую $1/(2F_b)$ -ю секунду можно узнавать очередное значение $S(k\Delta t)$. Если эти значения передавать по очереди через интервалы времени Δt , то по ним можно согласно соотношению (2.21) восстановить функцию $S(t)$ с любой степенью точности.

Вторая теорема Котельникова вошла во многие учебники под названием теоремы отсчетов.

Приведем теперь основные свойства функции отсчетов.

1. В момент времени $t = 0$ функция отсчетов имеет максимальное значение, равное единице. Это можно легко показать. Пусть $k = 0$, тогда $\psi_0(t) = \frac{\sin 2\pi F_b t}{2\pi F_b t}$. Раскроем неопределенность по правилу Лопиталя, обозначив $2\pi F_b t = x$.

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{(\sin x)'}{x'} = \lim_{x \rightarrow 0} \frac{\cos x}{1} = 1.$$

2. В моменты времени $t = \pm k\Delta t = k/(2F_b)$, где $k = 1, 2, \dots$, отсчетные функции обращаются в нуль.

3. Ширина главного лепестка функции на нулевом уровне равна $1/F_b$.

4. Отсчетные функции являются ортогональными в бесконечно большом интервале времени, т. е.

$$(\psi_i(t), \psi_k(t)) = \int_{-\infty}^{\infty} \psi_i(t) \psi_k(t) dt = \begin{cases} \frac{1}{2F_b}, & i = k; \\ 0, & i \neq k. \end{cases}$$

Рассмотрим теперь процесс восстановления непрерывного сигнала $S(t)$ по его дискретным отсчетам. Из теоремы отсчетов следует, что для передачи по каналу связи сигнала $S(t)$ с ограниченным спектром необходимо выполнить следующие операции.

1. Взять отсчеты мгновенных значений сигнала $S(k\Delta t)$ через интервалы времени $\Delta t = k / (2F_B)$, где $k = 1, 2, \dots$, т.е. найти величины $S = \left(\frac{1}{2F_B} \right)$, $S = \left(\frac{2}{2F_B} \right)$, $S = \left(\frac{3}{2F_B} \right)$, ...

2. Передать по каналу найденные отсчеты любым из возможных методов.

3. На приемной стороне восстановить переданные отсчеты и сформировать короткие импульсы с длительностью $\tau_n \ll \Delta t = 1 / (2F_B)$ и амплитудами $S(k\Delta t)$.

4. Сформировать функции отсчетов $\psi_k(t)$, $k = 1, 2, \dots$, как показано на рис. 2.12.

5. Произвести суммирование найденных функций и получить в результате сигнал $\tilde{S}(t)$, который будет пропорционален (или равен) переданному сигналу $S(t)$.

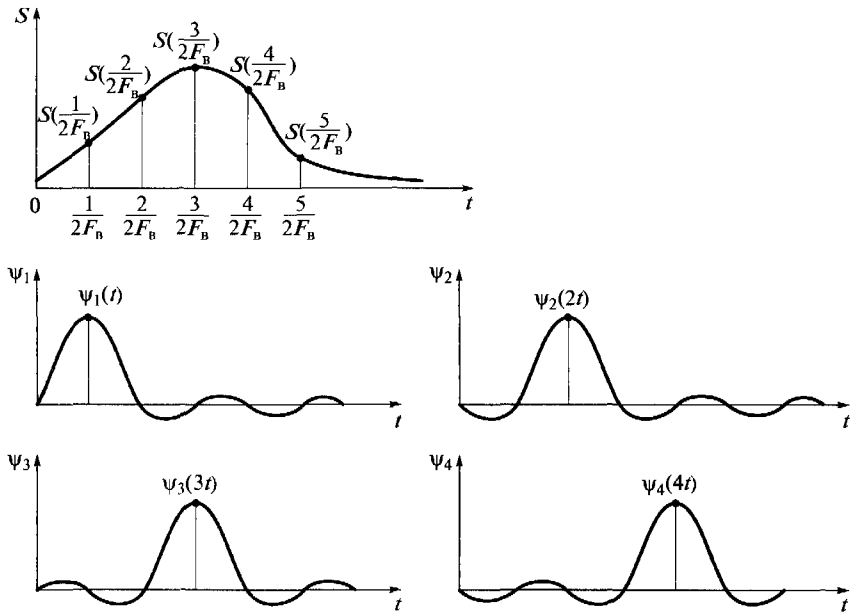


Рис. 2.12. Формирование отсчетов аналогового сигнала в дискретные моменты времени

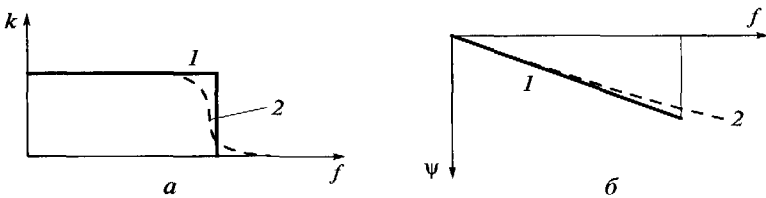


Рис. 2.13. Амплитудно-частотная (а) и фазо-частотная (б) характеристики фильтров, формирующих функции отсчетов:

1 — идеального; 2 — реального

Для получения отсчетных функций обычно применяется фильтр нижних частот с шириной полосы пропускания, равной F_b , на вход которого следует подать короткий импульс с длительностью τ_n и амплитудой $S(k\Delta t)$. Если фильтр считать идеальным, а на его вход подавать дельта-импульс $\delta(t)$, то отсчетная функция на выходе не будет иметь искажений, так как АЧХ $k(f)$ фильтра равномерная, а ФЧХ $\varphi(f)$ — линейная (рис. 2.13):

$$k(f) = \begin{cases} 1, & 0 \leq f \leq F_b, \\ 0, & f > F_b; \end{cases}$$

$$\varphi(f) = -2\pi f \tau_n.$$

На самом деле при восстановлении сигнала возникают погрешности. Рассмотрим кратко причины их возникновения на практике.

1. Характеристики реальных фильтров $k(f)$ и $\varphi(f)$ отличаются от идеальных (см. кривые 2 на рис. 2.13), что приводит к отклонениям реальных отсчетных функций от идеальных, а следовательно, к появлению некоторых неточностей восстановления непрерывного сигнала $S(t)$.

2. Для восстановления сигнала по его отсчетным функциям необходимо просуммировать бесконечное множество членов ряда Котельникова (2.21). Однако реальные сигналы $S(t)$ имеют ограниченные спектры и рассматриваются в конечном интервале времени T . В связи с этим точное разложение приходится заменять приближенным, т.е. при котором также суммируется конечное число членов ряда:

$$\hat{S}(t) = \sum_{k=-n/2}^{n/2} S(k\Delta t) \frac{\sin 2\pi F_b(t - k\Delta t)}{2\pi F_b(t - k\Delta t)}.$$

Число отсчетов, определяющее $\hat{S}(t)$, при $\Delta t = 1/(2F_b)$ составляет $n = T/\Delta t + 1 = 2F_b T + 1$ (обычно $2F_b T \gg 1$), поэтому $n = 2F_b T$.

Параметр $n = 2F_b T$, иногда обозначаемый символом B и называемый *базой сигнала*, играет важную роль в ТЭС при рассмотрении сложных сигналов.

Ясно, что погрешность при восстановлении сигнала будет тем больше, чем меньше число слагаемых учитывается при суммировании.

3. Спектры реальных сигналов не равны нулю за пределами граничной частоты. Основная доля энергии сигналов приходится на частоты от нуля до F_b , но небольшая доля этой энергии может быть и выше граничной частоты. Значение относительной среднеквадратичной погрешности можно определить следующим выражением:

$$\frac{\varepsilon^2(t)}{S^2(t)} = \frac{\int_{F_b}^{\infty} |S(jF)|^2 df}{\int_0^{\infty} |S(jF)|^2 df} = \frac{\Delta E}{E_S},$$

где ΔE — часть энергии, которая выходит за пределы полосы частот $[0, F_b]$ и не учитывается при восстановлении сигнала; E_S — полная энергия сигнала.

При этом оказывается, что погрешность за счет отбрасываемой части спектра сигнала становится тем больше, чем «медленнее» убывает спектр за пределами граничной частоты F_b .

В заключение отметим основные области применения теоремы отсчетов.

1. Упрощение представления сигналов или помех. Вместо непрерывных функций они задаются наборами отсчетных значений, и при этом не происходит потерь информации, содержащейся в непрерывном сигнале.

2. Передача непрерывных сообщений по каналам связи с помощью дискретных или цифровых сигналов.

3. Цифровая обработка непрерывных сигналов.

Примеры использования теоремы отсчетов будут рассмотрены далее.

2.9. Дискретное преобразование Фурье

До сих пор мы рассматривали аналоговые (непрерывные) сигналы. Однако в настоящее время при формировании и обработке сигналов $S(t)$ зачастую используются выборки $S(t_k)$, которые возникают при дискретизации непрерывного сигнала. В результате получают последовательность $\dots S_{-k}, \dots, S_{-1}, S_0, S_1, \dots, S_k, \dots$ отсчетных значений сигнала $S(t)$ в точках $\dots t_{-k}, \dots, t_{-1}, t_0, t_1, \dots, t_k$ соответственно. Ясно, что для таких сигналов неприменим аппарат преобразований, основанный на интеграле Фурье. Для них существует особый математический аппарат, называемый дискретным преобразованием Фурье (ДПФ).

Рассмотрим N -мерное комплексное евклидово пространство, в котором можно ввести следующую систему базисных функций, называемых *базисом Фурье* (элементами базиса являются векторы):

$$\psi_0 = (1, 1, 1, \dots, 1);$$

$$\psi_1 = (1, e^{j2\pi/N}, e^{j4\pi/N}, \dots, e^{j2(N-1)\pi/N});$$

$$\psi_2 = (1, e^{j4\pi/N}, e^{j8\pi/N}, \dots, e^{j4(N-1)\pi/N});$$

$$\psi_{n-1} = (1, e^{j2(N-1)\pi/N}, e^{j4(N-1)\pi/N}, \dots, e^{j2(N-1)^2\pi/N}).$$

Число $\exp(j2\pi nk/N)$ является k -й составляющей n -го базисного вектора. При любых n и k данное число представляет собой значение корня N -й степени из единицы в поле комплексных чисел.

Скалярное произведение базисных функций

$$(\psi_m, \psi_l) = \sum_{k=0}^{N-1} e^{j2\pi(m-l)k/N} = \begin{cases} N, & m=l; \\ 0, & m \neq l. \end{cases}$$

Здесь сумма обращается в нуль при $m \neq l$, поскольку все слагаемые являются комплексными числами с единичным модулем и линейно нарастающим аргументом.

Базис Фурье ортогонален, но не нормирован к единице, так как $\|\psi_m\|^2 = N$, $m=0, N-1$.

Найдем коэффициенты разложения некоторого сигнала S по базисным функциям:

$$S = \sum_{k=0}^{N-1} c_k \psi_k.$$

Для этого умножим обе части данного равенства скалярно на базисный вектор ψ с номером n :

$$(S, \psi_n) = \sum_{k=0}^{N-1} c_k (\psi_k, \psi_n).$$

Поскольку базисные функции являются ортогональными, в правой части этого уравнения отлично от нуля одно слагаемое с номером $k = n$. Поэтому

$$(S, \psi_n) = c_n \|\psi_n\|^2 = c_n N.$$

Тогда

$$c_n = \frac{1}{N} \sum_{k=0}^{N-1} S_k e^{-j2\pi nk/N}. \quad (2.22)$$

Выражение (2.22) определяет последовательность коэффициентов, образующих дискретное преобразование Фурье.

Если известны коэффициенты разложения, т. е. комплексные амплитуды c_n , при $n=0, N-1$, можно найти отсчетные значения сигнала S_k , выполнив обратное дискретное преобразование Фурье (ОДПФ):

$$S_k = \sum_{n=0}^{N-1} c_n e^{j2\pi nk/N}. \quad (2.23)$$

Выражения (2.22) и (2.23) являются по существу дискретными аналогами обычной пары преобразований Фурье для непрерывных сигналов.

Отметим некоторые свойства ДПФ.

1. Дискретное преобразование Фурье является линейным преобразованием, т. е. сумме (разности) сигналов соответствует сумма (разность) их ДПФ.

2. Число коэффициентов разложения c_n , где $n=0, N-1$, в ДПФ равно числу отсчетов сигнала, т. е. N . При $n=N$ коэффициент $c_N = c_0$.

3. Коэффициент c_0 (постоянная составляющая) представляет собой среднее значение всех отсчетов сигнала:

$$c_0 = \frac{1}{N} \sum_{k=0}^{N-1} S_k.$$

4. При четном N из формулы (2.22) получим

$$c_{\frac{N}{2}} = \frac{1}{N} \sum_{k=0}^{N-1} S_k (-1)^k.$$

5. Если отсчетные значения сигнала S_k — вещественные числа, то коэффициенты разложения ДПФ, номера которых расположены симметрично относительно $N/2$, образуют комплексно-сопряженные пары:

$$c_{\frac{N}{2}+n} = c_{\frac{N}{2}-n}^*, \quad n=1, 2, \dots, \frac{N}{2}-1.$$

Коэффициенты $c_{N/2+1}, \dots, c_{N-1}$ соответствуют отрицательным частотам. Поэтому при изучении амплитудного спектра они не дают дополнительной информации.

В заключение отметим следующее. При вычислении N коэффициентов ДПФ необходимо выполнить N^2 операций умножения. Если обрабатываются большие массивы, т. е. когда $N \geq 1000$, применять ДПФ и ОДПФ в реальном масштабе становится затруднительным из-за ограниченного быстродействия ЭВМ. В по-

следнее время разработаны алгоритмы быстрых преобразований Фурье (БПФ), при которых число операций умножения сокращается примерно до $N \log N$. Такие операции в курсе ТЭС не изучаются, они являются предметом рассмотрения в специальных курсах и при детальном изучении вопросов цифровой обработки сигналов.

2.10. Аналитический сигнал

2.10.1. Общие сведения

Пусть имеется гармоническое колебание

$$S(t) = S_m \cos(\omega_0 t + \varphi) = S_m \cos \psi(t), \quad (2.24)$$

где S_m , ω_0 , φ — соответственно амплитуда, частота и начальная фаза сигнала; $\psi(t)$ — текущая фаза сигнала.

Этот же сигнал можно представить в виде

$$\begin{aligned} S(t) &= S_m \cos \psi(t) = \operatorname{Re} \{ \hat{S}(t) \} = \operatorname{Re} \{ S_m e^{j\psi(t)} \} = \\ &= \operatorname{Re} \{ S_m \cos(\omega t + \varphi) + j S_m \sin(\omega t + \varphi) \} = \operatorname{Re} \{ S(t) + j \hat{S}(t) \}, \end{aligned}$$

где $\hat{S}(t) = S_m \sin(\omega t + \varphi)$ — сопряженный сигнал, полученный из исходного сигнала поворотом его фазы на $-\pi/2$.

На комплексной плоскости такой сигнал S изображается в виде вектора, как показано на рис. 2.14.

Негармонические сигналы подобно сигналу (2.24) можно представить в виде процесса с изменяющейся амплитудой (оггибающей) $S_m(t)$ и полной фазой $\psi(t)$, т.е. $S(t) = S_m(t) \cos \psi(t)$. Однако такое представление в общем случае является неоднозначным.

Действительно, пусть задан сигнал $S(t)$. Выбрав для него произвольную функцию $S_1(t)$ и считая $S_{m1}(t) = \sqrt{S^2(t) + S_1^2(t)}$, а $\psi_1(t) = \arctg \frac{S_1(t)}{S(t)}$, получим $S(t) = S_{m1}(t) \cos \psi_1(t)$.

Выбрав затем другую функцию $S_2(t)$, можно получить другой набор «амплитуд» и фаз: $\{S_{m2}(t), \psi_2(t)\}$ и т.д.

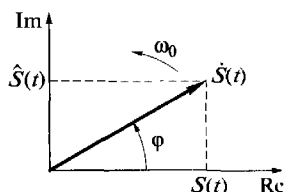


Рис. 2.14. Геометрическое представление комплексного сигнала

Для того чтобы представление было однозначным, как в случае гармонического сигнала, сопряженный сигнал должен быть получен из исходного сигнала посредством поворота всех его гармонических составляющих на $-\pi/2$.

Рассмотрим теперь сигнал без постоянной составляющей, представленный в виде ряда

$$S(t) = \sum_{k=1}^{\infty} a_k \cos \omega_k t + b_k \sin \omega_k t; \quad \omega_k = k\omega_0; \quad \omega_0 = \frac{2\pi}{T}.$$

Определим для него сопряженный сигнал из исходного сигнала посредством поворота всех его составляющих на $-\pi/2$:

$$\widehat{S}(t) = \sum_{k=1}^{\infty} a_k \sin \omega_k t - b_k \cos \omega_k t.$$

Тогда комплексный сигнал будет иметь вид

$$\dot{S}(t) = S(t) + j\widehat{S}(t), \quad (2.25)$$

а его реальная часть

$$S(t) = \operatorname{Re}\{\dot{S}(t)\}.$$

Отметим, что сопряженный сигнал $\widehat{S}(t)$ можно получить из исходного, не прибегая к спектральным представлениям, а используя интегральное преобразование Гильберта:

$$\widehat{S}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{S(\tau)}{t - \tau} d\tau. \quad (2.26)$$

Исходный сигнал $S(t)$ получим из сопряженного сигнала с помощью обратного преобразования Гильберта:

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

Функция $\frac{1}{t - \tau}$, называемая ядром преобразования Гильберта, имеет разрыв при $t = \tau$, поэтому интегралы следует понимать в смысле их главного значения, например:

$$\widehat{S}(t) = \frac{1}{\pi} \lim_{\xi \rightarrow 0} \left[\int_{-\infty}^{t-\xi} \frac{S(\tau)}{t - \tau} d\tau + \int_{t+\xi}^{\infty} \frac{S(\tau)}{t - \tau} d\tau \right].$$

Часто применяется символическая запись преобразований Гильберта:

$$\widehat{S}(t) = H[S(t)]; \quad S(t) = H^{-1}[\widehat{S}(t)].$$

Нетрудно увидеть, что прямое преобразование Гильберта эквивалентно прохождению сигнала $S(t)$ через фильтр с импульсной характеристикой $g(t) = \frac{1}{\pi t}$, а обратное преобразование Гильберта эквивалентно прохождению сопряженного сигнала $\hat{S}(t)$ через фильтр, импульсная характеристика которого $g(t) = -\frac{1}{\pi t}$.

Действительно, можно записать

$$\hat{S}(t) = S(t) * g(t) = \int_{-\infty}^{\infty} S(\tau)g(t - \tau)d\tau.$$

Подставив в это выражение импульсную характеристику вида $g(t - \tau) = \frac{1}{\pi(t - \tau)}$, получим формулу (2.26).

Дадим теперь определение рассмотренного сигнала.

Комплексный сигнал, полученный на основе преобразования Гильберта, называется *аналитическим* и записывается в виде выражения (2.25), где исходный сигнал есть реальная часть аналитического сигнала. Заметим, что выражение (2.25), в котором $S(t)$ и $\hat{S}(t)$ связаны между собой преобразованиями Гильберта, во-первых, позволяет получить однозначное представление вида (2.24), а во-вторых, обуславливает ряд важных свойств сигнала $\hat{S}(t)$, из-за которых он получил название «аналитический».

Приведем (без доказательства) лишь важнейшие свойства аналитического сигнала, используемые в теории связи.

1. Преобразования Гильберта являются линейными. Так, для прямого преобразования Гильберта это свойство можно записать в виде

$$H[a_1S_1(t) + a_2S_2(t)] = a_1H[S_1(t)] + a_2H[S_2(t)],$$

причем при любых постоянных a_1 и a_2 . Справедливость этого свойства следует непосредственно из выражений (2.26) и (2.27).

2. Преобразования Гильберта от постоянной величины тождественно равны нулю, т. е.

$$H[C] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{C}{t - \tau} d\tau = 0; \quad H^{-1}[C] = 0.$$

Это свойство следует из того факта, что ядро преобразования Гильберта есть нечетная функция аргумента τ относительно точки $t = \tau$, следовательно, интеграл от нечетной функции $\frac{c}{t - \tau}$ в пределах $(-\infty, \infty)$ равен нулю.

3. Если при каком-нибудь $t = \tau$ исходный сигнал достигает экстремума (максимума или минимума), то в окрестности этой точки сопряженный сигнал проходит через нуль. Сказанное иллюстрирует рис. 2.15, где совмещены графики $S(\tau)$ (рис. 2.15, а) и ядра преобразования $\frac{1}{t-\tau} = f(\tau)$ (рис. 2.15, б) в точке t , где функция имеет максимум. Результат преобразования Гильберта $\hat{S}(\tau)$ показан на рис. 2.15, в.

Нетрудно увидеть, что функция $f(\tau) = S(\tau) \frac{1}{t-\tau}$ является нечетной функцией аргумента τ , а значит, интеграл от нее в симметричных пределах $(-\infty, \infty)$ будет равен нулю.

4. Преобразование Гильберта от гармонических сигналов имеет вид

$$H[\cos \omega t] = \sin \omega t \operatorname{sgn}(\omega);$$

$$H[\sin \omega t] = -\cos \omega t \operatorname{sgn}(\omega),$$

где $\operatorname{sgn}(\omega) = \begin{cases} +1 & \text{при } \omega \geq 0; \\ -1 & \text{при } \omega < 0. \end{cases}$

Очевидно, что для положительных частот

$$H[\cos \omega t] = \sin \omega t;$$

$$H[\sin \omega t] = -\cos \omega t.$$

Доказательство каждого из указанных свойств следует из анализа сведений, приведенных в данном подразделе.

5. Сдвиг фаз всех составляющих действительного сигнала на угол φ соответствует умножению аналитического сигнала на $e^{j\varphi}$, т. е. аналитический сигнал после поворота фаз $\hat{S}_\varphi(t) = \hat{S}(t)e^{j\varphi}$, откуда легко вычислить и действительный сигнал:

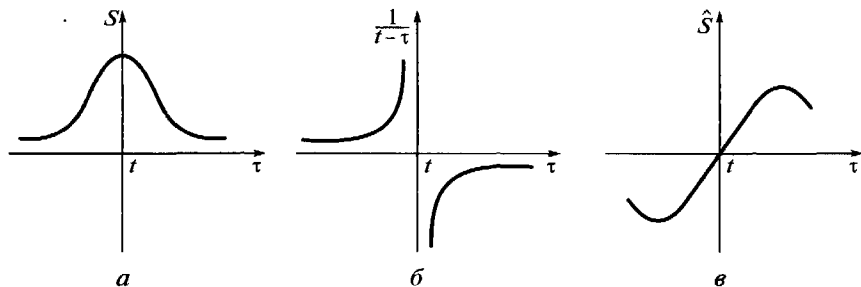


Рис. 2.15. Пояснение свойств преобразований Гильберта:

а — исходный сигнал; б — ядро преобразования; в — сопряженный сигнал

$$S_{\varphi}(t) = \operatorname{Re}\{\dot{S}_{\varphi}(t)\} = \operatorname{Re}\{\dot{S}(t)e^{j\varphi}\} = \operatorname{Re}\{[S(t) + j\widehat{S}(t)]e^{j\varphi}\} = \\ = S(t)\cos\varphi - \widehat{S}(t)\sin\varphi.$$

Использование понятия аналитического сигнала для определения формы действительного сигнала после поворота фаз всех его спектральных составляющих на один и тот же угол φ существенно облегчает задачу нахождения действительного сигнала. В противном случае для этого было бы необходимо с помощью преобразования Фурье найти комплексную спектральную плотность, произвести смещение фаз и затем проделать обратное преобразование Фурье.

6. Сдвиг частот всех составляющих сигнала на некоторую величину f_0 при $f > 0$ или $f < 0$ (преобразование частоты сигнала, причем само изменение частоты f_0 может быть как положительным, так и отрицательным) соответствует умножению аналитического сигнала $\dot{S}(t)$ на множитель $e^{j2\pi f_0 t}$, т.е.

$$\dot{S}_{f_0}(t) = \dot{S}(t)e^{j2\pi f_0 t},$$

откуда легко найти и действительный сигнал:

$$S_{f_0}(t) = \operatorname{Re}\{[S(t) + j\widehat{S}(t)]e^{j2\pi f_0 t}\} = S(t)\cos 2\pi f_0 t - \widehat{S}(t)\sin 2\pi f_0 t.$$

Без использования понятия аналитического сигнала решить эту задачу также было бы весьма сложно.

7. В спектре аналитического сигнала содержатся только положительные частоты. Спектр, полученный посредством преобразования Фурье, имеет вид

$$S_{\dot{s}}(\omega) = S_S(\omega) + jS_{\widehat{s}}(\omega) = \begin{cases} 2S_S(\omega), & \omega \geq 0; \\ 0, & \omega < 0. \end{cases}$$

Аналогично в спектре комплексно-сопряженного аналитического сигнала

$$\dot{S}^*(t) = S(t) - j\widehat{S}(t)$$

содержатся только отрицательные частоты:

$$S_{\dot{s}^*}(\omega) = \begin{cases} 2S_S(\omega), & \omega < 0; \\ 0, & \omega \geq 0. \end{cases}$$

Данные соотношения вытекают из формулы Эйлера.

8. Произведение аналитического сигнала $\dot{S}(t)$ и сопряженного с ним аналитического сигнала $\dot{S}^*(t)$ равно квадрату огибающей исходного действительного сигнала $S(t)$:

$$\dot{S}(t)\dot{S}^*(t) = [S(t) + j\hat{S}(t)][S(t) - j\hat{S}(t)] = S^2(t) + \hat{S}^2(t) = S_m^2(t).$$

Таким образом, модуль аналитического сигнала $\dot{S}(t)$ равен огибающей сигнала, т. е. $|\dot{S}(t)| = S_m(t)$.

2.10.2. Огибающая, мгновенная фаза и мгновенная частота сигнала

Комплексный сигнал $\dot{S}(t) = S(t) + j\hat{S}(t)$, как известно, можно представить в экспоненциальной форме:

$$\dot{S}(t) = S_m(t)e^{j\psi(t)}, \quad (2.28)$$

откуда следует, что

$$S(t) = S_m(t) \cos \psi(t); \quad (2.29)$$

$$\hat{S}(t) = S_m(t) \sin \psi(t). \quad (2.30)$$

Решая два последних уравнения относительно $S_m(t)$ и $\psi(t)$, найдем

$$S_m(t) = |\dot{S}(t)| = \sqrt{S^2(t) + \hat{S}^2(t)}; \quad (2.31)$$

$$\psi(t) = \arg \dot{S}(t) = \operatorname{arctg} \frac{\hat{S}(t)}{S(t)}. \quad (2.32)$$

Величина $S_m(t)$ в этих выражениях называется мгновенной амплитудой, или огибающей, сигнала, а величина $\psi(t)$ — мгновенной фазой сигнала. Производная от мгновенной фазы во времени (если она существует), называется мгновенной круговой частотой сигнала:

$$\omega(t) = \frac{d\psi(t)}{dt}. \quad (2.33)$$

Из формулы (2.31) следует, что $S_m(t) \geq S(t)$, причем равенство имеет место при тех значениях t , для которых $S(t) > 0$. Легко убедиться, что в этих точках производная огибающей совпадает с производной сигнала, т. е. $S_m(t) = S(t)$ (откуда и название — огибающая сигнала).

2.10.3. Узкополосные сигналы

В радиотехнике и ТЭС широко применяются так называемые узкополосные сигналы, которые являются полосовыми со спек-

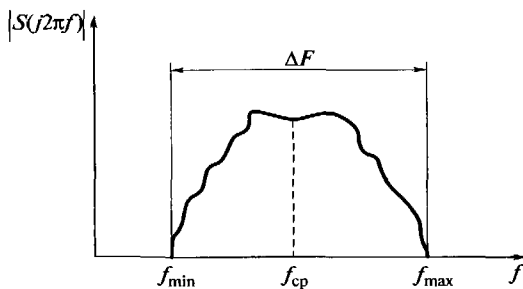


Рис. 2.16. Спектр полосового сигнала

тром, показанным на рис. 2.16, но ширина их спектра значительно меньше средней частоты, т. е. $\Delta F \ll f_{cp}$, где $f_{cp} = \frac{f_{max} + f_{min}}{2}$, а f_{max} и f_{min} — соответственно средняя, максимальная и минимальная частоты спектра сигнала.

Для узкополосных сигналов (и помех) представления (2.28) и (2.29) особенно удобны, так как в этом случае огибающая и мгновенная частота оказываются медленно изменяющимися функциями по сравнению с $\cos \psi(t)$ и, следовательно, по сравнению с самим сигналом $S(t)$. При этом формулу (2.29) удобно записать следующим образом:

$$S(t) = S_m(t) \cos[\omega_{cp} t + \theta(t)], \quad (2.34)$$

причем $\theta(t) = \psi(t) - \omega_{cp} t$; $\dot{\omega}(t) = \omega_{cp} + \frac{d\theta(t)}{dt}$.

Возможно и преобразование соотношения (2.28) вида

$$S(t) = \operatorname{Re}\{S_m(t) e^{j[\omega_{cp} t + \theta(t)]}\} = \operatorname{Re}\{\dot{S}_m(t) e^{j\omega_{cp} t}\}. \quad (2.35)$$

Здесь

$$\dot{S}_m(t) = S_m(t) e^{j\theta(t)} \quad (2.36)$$

представляет собой функцию времени, называемую комплексной огибающей сигнала $S(t)$. Модуль этой функции является обычной огибающей, а аргумент — мгновенной начальной фазой $\theta(t)$.

Комплексную огибающую можно также представить в виде

$$\dot{S}_m(t) = S_m(t) \cos \theta(t) + j S_m(t) \sin \theta(t) = S_c(t) + j S_s(t). \quad (2.37)$$

Здесь действительные функции времени $S_c(t)$ и $S_s(t)$ являются квадратурными составляющими комплексной огибающей или низкочастотными квадратурными составляющими. С их помощью сигнал можно представить в виде суммы:

$$\begin{aligned}
 S(t) &= S_m(t) \cos[\omega_{cp}(t) + \theta(t)] = \\
 &= S_c(t) \cos \omega_{cp} t - S_s(t) \sin \omega_{cp} t,
 \end{aligned}
 \tag{2.38}$$

что следует из выражений (2.34) и (2.37).

Учитывая «медленность» изменения функций $S_c(t)$ и $S_s(t)$ по сравнению с $\cos \omega_{cp} t$ и $\sin \omega_{cp} t$ из выражений (2.38) можно получить сопряженный сигнал:

$$\hat{S}(t) = S_c(t) \sin \omega_{cp} t + S_s(t) \cos \omega_{cp} t.$$

Подставив выражения (2.38) и (2.39) в формулу (2.31), нетрудно убедиться, что $S_m(t)$ — огибающая сигнала.

Схема, изображенная на рис. 2.17, иллюстрирует процесс формирования низкочастотных квадратурных составляющих сигнала.

Обратим особое внимание на следующее: нельзя путать понятия спектральных составляющих и мгновенной частоты, так как в первом случае частоты, входящие в спектр, не зависят от времени, а во втором — мгновенная частота есть функция времени, которая определяет скорость изменения фазы. Спектр сигнала можно измерить с помощью прибора — спектроанализатора, который выполняет приближенное преобразование Фурье. Мгновенная частота измеряется частотным детектором, работа которого будет рассмотрена далее, но по существу он реализует выражение (2.33).

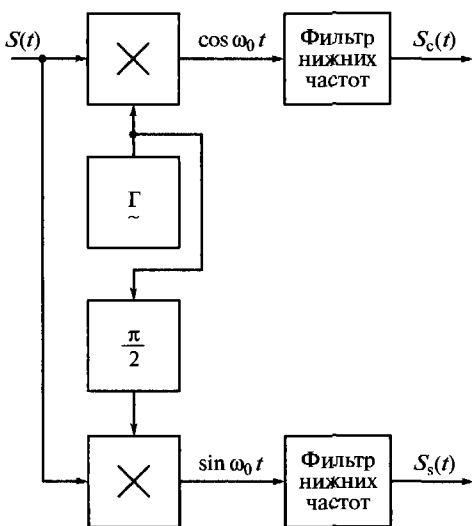


Рис. 2.17. Схема формирования квадратурных составляющих узкополосного сигнала

Контрольные вопросы

1. Какие сигналы являются периодическими и непериодическими, одномерными и многомерными, детерминированными и случайными?
2. В чем заключается различие между видео- и радиоимпульсами?
3. Что представляет собой дельта-функция и в чем заключается смысл ее фильтрующего свойства?
4. Назвать аксиоматические положения, которые определяют понятие линейного пространства.
5. Что представляет собой координатный базис в функциональном пространстве?
6. Дать определения понятий, характеризующих функциональное пространство (норма, метрика, скалярное произведение).
7. Какие функциональные пространства носят имена Евклида и Гильберта и какими параметрами они определяются?
8. Что понимают под обобщенным рядом Фурье, как определяются его коэффициенты и какой смысл они имеют?
9. Что представляют собой функции Уолша и каковы их основные свойства?
10. Записать ряд Фурье на основе гармонических функций и пояснить, почему эти функции применяются чаще других. Указать вид сигналов, для которых справедливы также ряды Фурье.
11. Записать ряд Фурье в комплексной форме. Пояснить смысл входящих в него величин.
12. Записать преобразования Фурье и пояснить их смысл. Определить, что понимают под комплексной спектральной плотностью.
13. Для каких сигналов применяют дискретное преобразование Фурье и какими свойствами оно обладает?
14. Пояснить смысл функций авто- и взаимной корреляции и как они связаны со спектром сигнала.
15. Написать ряд Котельникова и пояснить, что в нем является базисными функциями.
16. Пояснить, как осуществить восстановление аналогового сигнала по дискретным выборкам и указать через какие промежутки времени они берутся.
17. Записать соотношения для огибающей, фазы и частоты аналитического сигнала и пояснить их смысл.

Глава 3

НЕЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ В КАНАЛАХ ЭЛЕКТРОСВЯЗИ

3.1. Общие сведения о преобразованиях сигналов

При рассмотрении общей структурной схемы системы электросвязи было установлено, что в процессе передачи и приема сигнал подвергается различным преобразованиям. Эти преобразования принято подразделять на *линейные* и *нелинейные*.

К линейным преобразованиям применяется принцип суперпозиции, согласно которому гармонический сигнал, проходя через линейную систему, остается неизменным по форме, а изменяются только его амплитуда и начальная фаза.

К линейным преобразованиям относятся:

- усиление и ослабление сигнала;
- фильтрация;
- временные задержки, приводящие к фазовым сдвигам;
- интегрирование и дифференцирование сигналов.

В результате таких преобразований на выходе линейных устройств не появляются колебания на новой частоте. Вместе с тем из анализа системы передачи информации очевидна необходимость преобразования спектра сигналов посредством модуляции, что объясняется физическими особенностями каналов связи.

Так, по радиоканалу практически невозможно обеспечить передачу сигналов в низкочастотном диапазоне, в котором сосредоточен спектр звуковых частот. В связи с этим возникает задача переноса спектра сигнала с сохранением в нем информации об исходном передаваемом сообщении. Эта процедура, как было показано ранее, называется модуляцией. При этом на приемной стороне необходима обратная процедура, называемая демодуляцией. Таким образом, мы установили неизбежность применения нелинейных преобразований, к числу которых относятся:

- процедуры модуляции и демодуляции, связанные с трансформацией (переносом) спектра передаваемого сигнала и обе-

спечивающие появление новых частотных составляющих, отсутствовавших на входе устройства преобразования;

- преобразование частоты, под которым понимают превращение узкополосного сигнала $S_{\text{вх}} = U(t) \cos [\omega_0 t + \varphi(t)]$ на входе преобразователя в новый узкополосный сигнал $S_{\text{вых}}(t) = kU(t) \cos [\omega_{\text{пр}} t + \varphi(t) + \varphi_0]$ на его выходе с теми же законами изменения огибающей и начальной фазы, но сосредоточенный по спектру около частоты $f_{\text{пр}}$ (k и φ_0 — константы). При этом различают преобразование частоты вверх ($f_{\text{пр}} > f_0$) и преобразование частоты вниз ($f_{\text{пр}} < f_0$);

- умножение и деление частоты. В этом случае входной сигнал $S_{\text{вх}}(t) = U_c \cos [\omega_0 t + \varphi_0]$ превращается в выходной сигнал $S_{\text{вых}}(t) = kU_c \cos [(\omega_0 m/n)t + \varphi_1]$, где m и n — целые числа.

Отметим, что процедуру преобразования частоты применяют в приемных устройствах, а процедуры перемножения и деления — при генерировании необходимой сетки частот в передающих и приемных устройствах.

Указанные процедуры реализуются в нелинейных преобразователях и линейных параметрических цепях. К нелинейным преобразователям относятся полупроводниковые приборы (диоды, биполярные, полевые транзисторы) и электронные приборы, работающие с достаточно большими входными сигналами.

Нелинейные преобразователи подразделяются следующим образом:

- *нелинейные безынерционные преобразователи* (резистивные нелинейные элементы), в которых отклик $i(t_0)$ в момент времени t_0 зависит только от значения входного сигнала $u(t_0)$ в этот же момент времени. Зависимость выходного тока от входного напряжения $i(u)$ называется вольт-амперной характеристикой;

- *нелинейные инерционные преобразователи*, в которых наблюдается зависимость выходного тока в момент времени t_0 от значений входного напряжения в другой момент времени t . Такой эффект наблюдается в случае, когда время прохождения носителей заряда через прибор и период колебаний входного сигнала примерно одинаковые.

Цепи, имеющие в своем составе хотя бы один элемент, параметр которого зависит от времени, называются *параметрическими*. Например, угольный микрофон с проводимостью, изменяющейся под действием звукового давления. К параметрическим элементам относятся также резисторы $R(t)$, конденсаторы $C(t)$ и индуктивности $L(t)$.

В параметрических цепях выполняется принцип суперпозиции. Также для них характерно возникновение новых спектральных составляющих, поэтому параметрические цепи применяются для реализации процедур модуляции и демодуляции.

3.2. Методы аппроксимации характеристик нелинейных преобразователей сигналов

3.2.1. Общие сведения

Ранее отмечалось, что нелинейный преобразователь может быть описан с помощью вольт-амперной характеристики, которую, как правило, получают экспериментальным путем и затем представляют в виде графической зависимости тока от напряжения. Однако использовать вольт-амперную характеристику в графической форме неудобно. Поэтому возникает задача аппроксимации, под которой понимают приближенное представление нелинейной характеристики математическими методами.

Аппроксимирующая функция должна удовлетворять следующим требованиям:

- быть простой, но допускать последующую математическую обработку;
- достаточно точно отображать экспериментально полученную характеристику.

Исходя из указанных требований в ТЭС применяются следующие методы аппроксимации вольт-амперных характеристик:

- полиномиальная (степенная);
- кусочно-линейная;
- аппроксимация трансцендентными функциями (экспоненциальными, тригонометрическими и др.).

3.2.2. Полиномиальная аппроксимация

Данный метод аппроксимации удобно применять при рассмотрении принципов действия многих нелинейных преобразователей: модуляторов, демодуляторов, генераторов и других при воздействии на них одного либо нескольких гармонических колебаний.

Степенная аппроксимация заключается в записи вольт-амперной характеристики $i = f(u)$ в виде полинома (многочлена) n -й степени:

$$i = \sum_{k=0}^n a_k U^k = a_0 + a_1 U + a_2 U^2 + \dots + a_n U^n.$$

С помощью такого полинома можно произвести аппроксимацию с любой степенью точности. При этом точность будет выше, если использовать полиномы высоких порядков. Однако это неудобно для анализа.

На практике применяют следующие полиномы:

- первой степени ($k = 1$) (рис. 3.1)

$$i = a_0 + a_1 u;$$

- второй степени ($k = 2$)

$$i = a_0 + a_1 u + a_2 u^2;$$

- укороченный полином третьей степени ($k = 3, a_2 = 0$)

$$i = a_0 + a_1 u + a_3 u^3;$$

- укороченный полином пятой степени ($k = 5, a_2 = a_4 = 0$)

$$i = a_0 + a_1 u + a_3 u^3 + a_5 u^5.$$

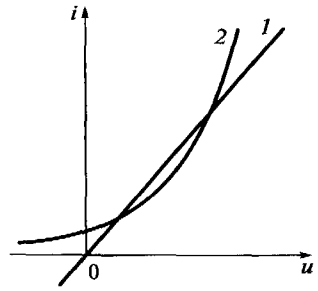


Рис. 3.1. Аппроксимация вольт-амперной характеристики нелинейного преобразователя прямой линией (1) и параболой (2)

Отметим, что аппроксимация многочленом первой степени, которая приводит к получению прямой линии 1 (см. рис. 3.1), не позволяет изучать нелинейные преобразования, например перенос спектра сигнала на другую частоту. Данный вид аппроксимации применяется только при изучении линейных процессов, например усиления. Полином второй степени (квадратичная парабола 2) уже позволяет изучать нелинейные процессы, но только при воздействии слабых сигналов.

В общем случае для использования степенной аппроксимации необходимо знать коэффициенты a_k полинома, которые обычно определяют с помощью *метода выбранных точек*. Иными словами, коэффициенты a_k находят из условия равенства значений ординат аппроксимированной и действительной характеристик в выбранных точках.

Для аппроксимации полинома n -й степени в пределах интервала ΔU , задаваемого диапазоном изменения напряжения u , выбирают $n + 1$ значений напряжения u_1, \dots, u_{n+1} и определяют значения токов i_1, \dots, i_{n+1} .

В простейшем случае значения u_k находят делением интервала ΔU на n равных частей Δ , как показано на рис. 3.2. При $n = 4$ значение $\Delta = (u_5 - u_1) / 4$. Далее составляется следующая система уравнений:

$$i_1 = a_0 + a_1 u_1 + \dots + a_n u_1^n;$$

$$i_2 = a_0 + a_1 u_2 + \dots + a_n u_2^n;$$

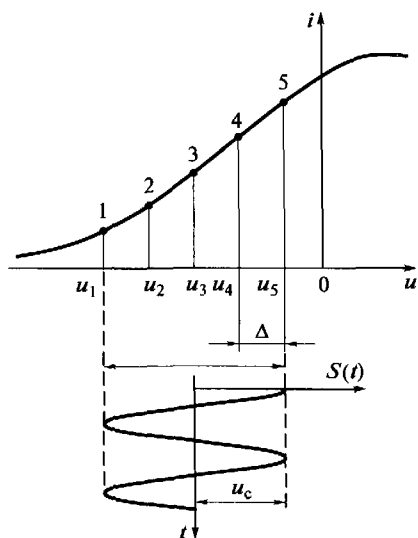
$$\vdots$$

$$i_k = a_0 + a_1 u_k + \dots + a_n u_k^n;$$

$$\vdots$$

$$i_{n+1} = a_0 + a_1 u_{n+1} + \dots + a_n u_{n+1}^n.$$

Рис. 3.2. Определение ординат нелинейной характеристики методом выбранных точек



Здесь i_k и u_k являются известными значениями.

Решение данной системы уравнений определяют коэффициенты a_k . Если значение $u = 0$ располагается внутри интервала аппроксимации ΔU , то коэффициент a_0 находят как значение тока при $u = 0$.

При вычислении коэффициентов a_k иногда предъявляется дополнительное требование равенства не только ординат действительной и аппроксимированной характеристик, но и их производных.

Следует иметь в виду, что аппроксимированная характеристика может резко отличаться от действительной за пределами интервала аппроксимации ΔU . Поэтому степенная аппроксимация широко применяется при анализе работы нелинейных устройств, на которые поступают относительно небольшие сигналы. Использовать этот метод при больших отклонениях мгновенных значений входного сигнала от рабочей точки нецелесообразно по причине ухудшения точности.

3.2.3. Кусочно-линейная аппроксимация

Данный метод основан на приближенной замене реальной плавно изменяющейся вольт-амперной характеристики $i = f(u)$ отрезками прямых линий с различными наклонами.

На рис. 3.3 показана аппроксимированная характеристика, содержащая два линейных участка.

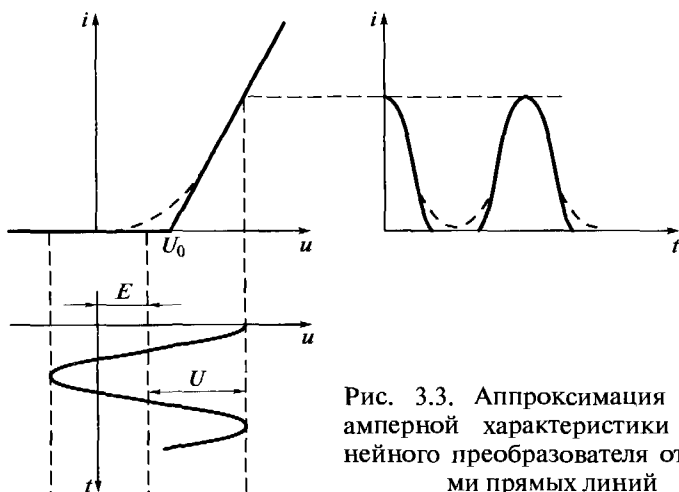


Рис. 3.3. Аппроксимация вольт-амперной характеристики нелинейного преобразователя отрезками прямых линий

Математически эту аппроксимированную характеристику можно записать в виде

$$i(U) = \begin{cases} 0, & U < U_0; \\ S(U - U_0), & U \geq U_0, \end{cases}$$

где U_0 — напряжение отсечки; S — крутизна характеристики, имеющая размерность проводимости (См или А/В).

На рис. 3.3 методом проекций построены импульсы тока, получаемые при воздействии гармонического колебания с большой амплитудой. При разложении импульсов в ряд Фурье постоянная составляющая и амплитуды нескольких первых гармоник также близки друг к другу. Иными словами, при больших сигналах кусочно-линейная аппроксимация дает достаточную точность расчетов. При малых сигналах точность падает, и результаты могут быть неверными.

Таким образом, метод кусочно-линейной аппроксимации обычно применяется при анализе процессов нелинейных преобразований в случае больших амплитуд входных сигналов.

3.2.4. Аппроксимация трансцендентными функциями

В данном методе в качестве аппроксимирующих функций используются экспоненты либо их суммы (но не более чем из двух слагаемых), гиперболические, тригонометрические и некоторые

другие функции. Чаще всего применяется показательная аппроксимация. Так, в частности, характеристику полупроводникового диода можно аппроксимировать экспонентой:

$$i = I_0(e^{\alpha U} - 1),$$

где I_0 — обратный ток насыщения; α — постоянная, характеризующая температурный потенциал.

Данное выражение хорошо определяет начальный участок характеристики. Иными словами, показательная аппроксимация достаточно точна при небольших амплитудах входных сигналов. В противном случае погрешность расчетов оказывается значительной.

Для определения пригодности этого метода аппроксимации для расчета применяют так называемое приведение к линейному виду, сущность которого заключается в следующем. Сначала представленное ранее выражение для тока i логарифмируется:

$$\alpha U = \ln(i/I_0 + 1).$$

Далее по реальной вольт-амперной характеристике строится зависимость

$$\ln(i/I_0 + 1) = f(u).$$

Затем в диапазоне напряжений ΔU проверяется степень отличия этой характеристики от прямой линии. Если это отличие оказывается небольшим, то исходную характеристику нелинейного элемента (полупроводникового диода) можно аппроксимировать экспонентой.

Нелинейные зависимости более сложного вида можно аппроксимировать суммой двух трансцендентных функций. Например, характеристика туннельного диода описывается выражением

$$i = A U e^{-\alpha U} + D (e^{\beta U} - 1),$$

в котором первое слагаемое определяет туннельный ток, а второе — диффузионный.

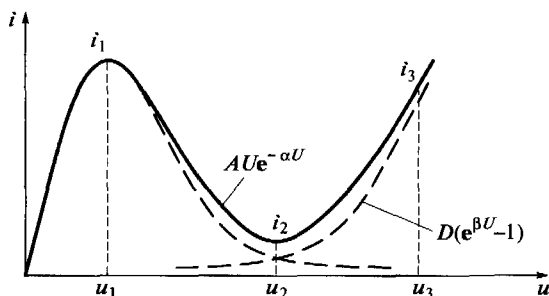


Рис. 3.4. Пример аппроксимации характеристики туннельного диода трансцендентными функциями

Графическое изображение этой характеристики приведено на рис. 3.4, где сплошной линией показан суммарный ток, а штриховой — его компоненты.

3.3. Спектральный анализ колебаний в нелинейных преобразователях

3.3.1. Общие сведения

При изучении процессов нелинейных преобразований в первую очередь приходится решать задачу нахождения спектра колебаний на выходе преобразователей. Данная задача формулируется следующим образом.

Имеется безынерционный нелинейный преобразователь, характеристика которого аппроксимируется зависимостью $i = f(u)$. На вход этого преобразователя поступает так называемое полигармоническое колебание вида

$$u(t) = \sum_{k=0}^n U_k \cos(\omega_k t + \varphi_k).$$

В частном случае это может быть моногармоническое колебание вида

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

Необходимо определить спектр тока i на выходе преобразователя. Такая задача получила название *спектральный анализ*. Для решения этой задачи можно использовать аппарат рассмотренных ранее рядов Фурье. Однако такой метод определения спектра отклика оказывается весьма трудоемким, поэтому на практике применяются специальные методы спектрального анализа, связанные с рассмотренными ранее методами аппроксимации характеристик нелинейных преобразователей. При этом наибольшее распространение получили:

- метод кратных дуг;
- метод формул трех и пяти ординат;
- метод функций Бесселя от мнимого аргумента;
- метод угла отсечки.

3.3.2. Метод кратных дуг

Данный метод является основным при использовании полиномиальной аппроксимации. Его удобно применять при анализе нелинейных преобразований в процессе модуляции, демодуляции, преобразования и деления частоты.

Пусть вольт-амперная характеристика нелинейного резистивного элемента аппроксимирована многочленом n -й степени:

$$i = a_0 + a_1 u + a_2 u^2 + \dots + a_n u^n.$$

На вход преобразователя подается гармоническое колебание вида

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

Выполнив соответствующие подстановки, получим

$$i = a_0 + a_1 U \cos(\omega t + \varphi) + a_2 U^2 \cos^2(\omega t + \varphi) + \dots + a_n U^n \cos^n(\omega t + \varphi).$$

Воспользуемся следующими известными формулами:

$$\cos^2 x = \frac{1}{2}(1 + \cos 2x);$$

$$\cos^3 x = \frac{1}{4}(3 + \cos 3x);$$

$$\cos^4 x = \frac{1}{8}(3 + 4 \cos 2x + \cos 4x);$$

$$\cos^5 x = \frac{1}{16}(10 \cos x + 5 \cos 3x + \cos 5x),$$

позволяющими степени косинусов (синусов) заменить тригонометрическими формулами кратных аргументов, отсюда и происходит название данного метода.

Предположим $x = \omega t + \varphi$, тогда, выполнив очевидные подстановки, получим:

$$i = I_0 + I_1 \cos(\omega t + \varphi) + I_2 \cos 2(\omega t + \varphi) + \dots + I_n \cos n(\omega t + \varphi).$$

Здесь

$$I_0 = a_0 + a_2 U^2 / 2 + 3a_4 U^4 / 8 + \dots$$

$$I_1 = a_1 U + 3a_3 U^3 / 4 + 5a_5 U^5 / 8 + \dots$$

$$I_2 = a_2 U^2 / 2 + a_4 U^4 / 4 + \dots$$

$$I_3 = a_3 U^3 / 4 + 5a_5 U^5 / 16 + \dots$$

При произвольном номере гармоники общее выражение для тока имеет вид

$$I_n = \sum_{k=0}^{\infty} \frac{(2k+n)!}{2^{2k+n-1} k!(k+n)!} a_{2k+n} U^{2k+n}.$$

Спектр амплитуд тока на выходе нелинейного преобразователя при воздействии одного гармонического колебания показан на рис. 3.5, а.

Из сказанного можно сделать следующие выводы.

1. Выходной спектр нелинейного преобразователя при воздействии гармонического сигнала является линейчатым. При этом составляющие сигнала имеют частоты, кратные частоте входного сигнала. Наивысший номер составляющей спектра равен степени аппроксимирующего полинома.

2. Постоянная составляющая и амплитуды четных гармоник определяются только четными степенями напряжения.

3. Значение текущей фазы k -й гармоники больше значения фазы входного сигнала в k раз, т.е. $x_k = \omega_k t + \varphi_k = k(\omega t + \varphi)$, откуда $\varphi_k = k\varphi$.

Ранее отмечалось, что нелинейные преобразования вызывают появление новых спектральных составляющих, которых не было на входе. Данный эффект проявляется наиболее ярко, если на вход преобразователя подается колебание, являющееся суммой нескольких гармоник с различными частотами.

Пусть на нелинейный резистивный элемент поступает так называемое бигармоническое колебание вида

$$u(t) = U_1 \cos \omega_1 t + U_2 \cos \omega_2 t.$$

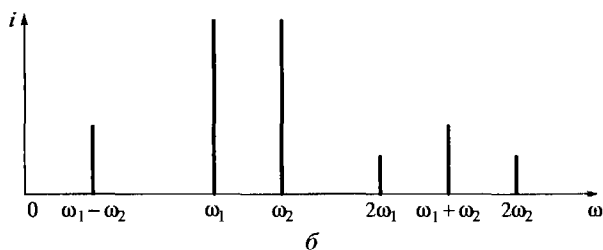
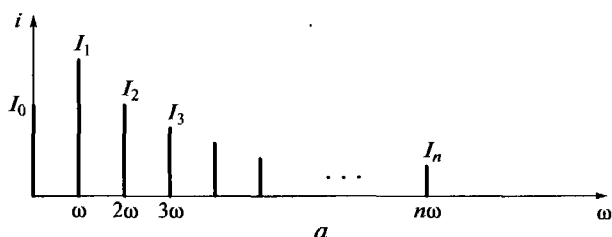


Рис. 3.5. Амплитудные спектры тока на выходе нелинейного преобразователя при воздействии одного (а) и двух (б) гармонических колебаний

Для упрощения анализа рассмотрим случай, когда вольт-амперная характеристика описывается многочленом 2-й степени (т.е. слабо нелинейный режим):

$$i(u) = a_0 + a_1 U + a_2 U^2.$$

Выполним соответствующие подстановки:

$$\begin{aligned} i(u) &= a_0 + a_1 U_1 \cos \omega_1 t + a_1 U_2 \cos \omega_2 t + a_2 U_1^2 \cos^2 \omega_1 t + \\ &+ 2a_2 U_1 U_2 \cos \omega_1 t \cos \omega_2 t + a_2 U_2^2 \cos^2 \omega_2 t = a_0 + a_2 (U_1^2 + U_2^2) / 2 + \\ &+ a_1 U_1 \cos \omega_1 t + a_1 U_2 \cos \omega_2 t + a_2 U_1^2 \cos 2\omega_1 t / 2 + a_2 U_2^2 \cos 2\omega_2 t / 2 + \\ &+ a_2 U_1 U_2 \cos (\omega_1 + \omega_2) t + a_2 U_1 U_2 \cos (\omega_1 - \omega_2) t. \end{aligned}$$

Из этого выражения видно, что в выходном колебании содержатся составляющие, которые имелись в спектре входного колебания, а также появились и новые гармоники. Иными словами, на выходе имеются постоянная составляющая и первые, вторые гармоники входных сигналов. Принципиально новым является появление двух комбинационных колебаний с частотами $\omega_1 + \omega_2$ и $\omega_1 - \omega_2$. Амплитуды этих колебаний, равные $a_2 U_1 U_2$, в одинаковой степени зависят от амплитуд каждого из входных сигналов. Комбинационные колебания обращаются в нуль, если на входе устройства отсутствует любой из двух входных сигналов.

Спектр амплитуд тока для рассмотренного случая показан на рис. 3.5, б.

3.3.3. Метод трех и пяти ординат

Данный метод применяют, как правило, при графических расчетах для оценки нелинейных искажений, возникающих в модуляторах, усилителях и других устройствах. Его отличительной особенностью является то, что в нем не требуется осуществлять аппроксимацию вольт-амперной характеристики нелинейного элемента.

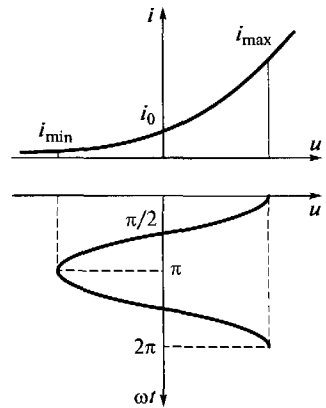
Метод формул трех ординат позволяет определить значения постоянной составляющей и амплитуды первых двух гармоник тока в следующем виде:

$$i = I_0 + I_1 \cos \omega t + I_2 \cos^2 \omega t.$$

Пусть характеристика нелинейного преобразователя задана графически (рис. 3.6).

Выберем на графике три ординаты и потребуем, чтобы значения тока в этих точках совпадали с его действительными значениями. Иными словами, возьмем следующие точки:

Рис. 3.6. Определение значений гармоник тока посредством выбора ординат характеристики нелинейного преобразователя



$$i_{\max} \quad \text{при } \omega t = 0, \quad u = U;$$

$$i_0 \quad \text{при } \omega t = \pi/2, \quad u = 0;$$

$$i_{\min} \quad \text{при } \omega t = \pi, \quad u = -U.$$

Подставив выбранные значения в формулу для тока, получим систему из трех уравнений:

$$i_{\max} = I_0 + I_1 + I_2;$$

$$i_0 = I_0 - I_2;$$

$$i_{\min} = I_0 - I_1 + I_2.$$

Решив эту систему уравнений относительно I_0, I_1, I_2 , получим:

$$I_0 = \frac{i_{\max} + i_{\min} + 2i_0}{4}; \quad I_1 = \frac{i_{\max} - i_{\min}}{2}; \quad I_2 = \frac{i_{\max} + i_{\min} - 2i_0}{4}.$$

Аналогично используют формулу пяти ординат для определения значений тока первых четырех гармоник. В этом случае точность расчетов будет выше. Однако в целом точность нахождения амплитуд гармоник с использованием данного метода невысока: ошибка растет с увеличением амплитуды подводимого напряжения.

3.3.4. Метод функций Бесселя

Данный метод применяется при анализе работы демодуляторов и преобразователей частоты в случае, когда вольт-амперная характеристика аппроксимируется экспоненциальной функцией.

Пусть имеется нелинейный преобразователь в виде полупроводникового диода, характеристика которого аппроксимирована выражением

$$i = A(e^{au} - 1).$$

На его вход подается напряжение

$$u(t) = E + U \cos \omega t.$$

Подставив это напряжение в аппроксимирующее выражение, получим

$$i = A(e^{\alpha E} e^{\alpha U \cos \omega t} - 1).$$

Это выражение можно представить в виде ряда Фурье, для чего необходимо найти коэффициенты разложения, используя выражения из теории функций Бесселя:

$$e^{x \cos \varphi} = I_0(x) + 2 \sum_{n=1}^{\infty} I_n(x) \cos n\varphi;$$

$$e^{x \sin \varphi} = I_0(x) + 2I_1(x) \sin \varphi + 2I_2(x) \cos 2\varphi +$$

$$+ 2I_3(x) \sin 3\varphi + 2I_4(x) \cos 4\varphi + \dots,$$

где $I_n(x)$ — модифицированная функция Бесселя n -го порядка от аргумента x .

При этом $I_0(0) = 1$, $I_1(0) = I_2(0) = \dots = 0$, $x = \alpha U$, $\varphi = \omega t$.

Тогда разложение в ряд Фурье будет иметь следующий вид:

$$i = A(e^{\alpha E} I_0(\alpha U) - 1) + 2Ae^{\alpha E} I_1(\alpha U) \cos \omega t +$$

$$+ 2Ae^{\alpha E} I_2(\alpha U) \cos 2\omega t + 2Ae^{\alpha E} I_3(\alpha U) \cos 3\omega t \dots$$

Из этого разложения видно, что постоянная составляющая тока

$$I_0 = A[e^{\alpha E} I_0(\alpha U) - 1],$$

амплитуда первой гармоники

$$I_1 = 2 A e^{\alpha E} I_1(\alpha U),$$

амплитуда n -й гармоники

$$I_n = 2Ae^{\alpha E} I_n(\alpha U).$$

Расчеты по данным выражениям показывают, что с увеличением номера гармоники ее амплитуда уменьшается. При выполнении расчетов можно использовать подробные таблицы функций Бесселя, приведенные в специальных справочниках.

3.3.5. Метод угла отсечки

Метод угла отсечки применяют при кусочно-линейной аппроксимации вольт-амперных характеристик. Он весьма эффективен для расчетов умножителей частоты, усилителей и генераторов, собранных на полупроводниковых приборах и лампах.

Пусть имеется нелинейный преобразователь, вольт-амперная характеристика которого аппроксимирована соотношением

$$i = \begin{cases} S(u - U_0), & u \geq U_0; \\ 0, & u < U_0, \end{cases}$$

где S — крутизна характеристики; U_0 — напряжение отсечки.

На рис. 3.7 эта характеристика представляет собой две прямые линии.

Рассмотрим воздействие напряжения

$$u = E + U \cos \omega t,$$

где E — напряжение смещения, которое определяет рабочую точку.

Из рис. 3.7 видно, что нелинейный элемент работает с отсечкой, т.е. часть входного напряжения, которая не заштрихована, не участвует в создании тока. Получаемые при этом импульсы тока характеризуются двумя величинами: высотой i_{\max} и шириной, т.е. углом отсечки, который обозначен θ .

Углом отсечки называется часть периода колебания с частотой ω , в течение которого или ток изменяется от максимального значения до нуля, или входное напряжение изменяется от максимального U до U_0 . Угол отсечки $\theta = \omega t$ и измеряется в градусах или в радианах. Следовательно, можно записать

$$E + U \cos \theta = U_0,$$

откуда

$$\cos \theta = \frac{U_0 - E}{U}.$$

Угол отсечки может принимать значения от нуля (ток не проходит) до π , что соответствует линейному режиму работы преобразователя. Если напряжение смещения E равно напряжению от-

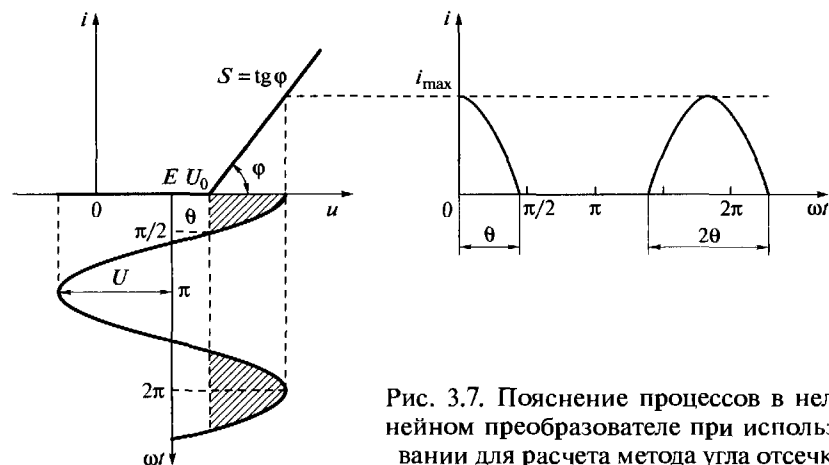


Рис. 3.7. Пояснение процессов в нелинейном преобразователе при использовании для расчета метода угла отсечки

сечки, то $\theta = \pi/2$, т.е. в этом случае проходят только положительные полупериоды входного сигнала.

Определим значение выходного тока. Для чего подставим выражение входного напряжения u в соотношение, которым аппроксимирована вольт-амперная характеристика:

$$i = Su - SU_0 = SE + SU \cos \omega t - SU_0.$$

Поскольку при $\omega t = \theta$ ток равен нулю ($i = 0$), можно записать

$$0 = SE + SU \cos \theta - SU_0.$$

Вычитанием второго выражения из первого получим

$$i = SU(\cos \omega t - \cos \theta).$$

При $\omega t = 0$ выходной ток имеет максимальное значение ($i = i_{\max}$), т.е. $i_{\max} = SU(1 - \cos \theta)$.

Полученная графическим построением периодическая последовательность импульсов тока является четной функцией, поэтому ее можно представить в виде ряда Фурье, в котором содержатся постоянная составляющая и косинусоидальные гармоники:

$$i = I_0 + I_1 \cos \omega t + I_2 \cos^2 \omega t + \dots$$

Постоянную составляющую найдем из соотношения

$$I_0 = \frac{1}{2\pi} \int_{-\theta}^{\theta} i(\omega t) d(\omega t) = \frac{1}{2\pi} \int_{-\theta}^{\theta} SU(\cos \omega t - \cos \theta) d(\omega t) = SU \gamma_0(\theta),$$

где $\gamma_0(\theta) = \frac{1}{\pi}(\sin \theta - \theta \cos \theta)$ — коэффициент постоянной составляющей.

Соответственно амплитуда первой гармоники

$$I_1 = \frac{1}{2\pi} \int_{-\theta}^{\theta} i(\omega t) \cos \omega t d(\omega t) = SU \gamma_1(\theta),$$

где $\gamma_1(\theta) = \frac{1}{\pi}(\theta - \sin \theta \cos \theta)$.

Аналогично определяется амплитуда n -й гармоники:

$$I_n = SU \gamma_n(\theta) = \frac{1}{2\pi} \int_{-\theta}^{\theta} SU(\cos \omega t - \cos \theta) \cos n\omega t d(\omega t),$$

где $\gamma_n(\theta) = \frac{2 \sin n\theta \cos \theta - n \cos n\theta \sin \theta}{\pi n(n^2 - 1)}$; $n = 2, 3, 4, \dots$

Иногда при расчете удобнее использовать нормированные коэффициенты гармоник (нормированные относительно значения максимального тока i_{\max}):

$$\alpha_n(\theta) = I_n / i_{\max} = \gamma_n(\theta) / (1 - \cos\theta).$$

Эти коэффициенты являются функциями только угла отсечки θ , поэтому для них имеются справочные графики и специальные таблицы, которые удобно применять в расчетах нелинейных преобразователей. Заметим, что для наиболее часто применяемого режима при $\theta = \pi/2$

$$\gamma_0(\pi/2) = \alpha_0(\pi/2) = 1/\pi; \quad \gamma_1(\pi/2) = \alpha_1(\pi/2) = 1/2.$$

Анализ зависимостей $\alpha_n(\theta)$ позволяет сделать следующие выводы.

1. Амплитуды гармоник спектра сигнала имеют максимум при оптимальных углах отсечки, которые определяются соотношением $\theta \approx 120^\circ/n$. Например, амплитуда второй гармоники максимальна при $\theta \approx 60^\circ$, 3-й гармоники — при $\theta \approx 40^\circ$ и т.д. Данное обстоятельство позволяет выбирать оптимальную рабочую точку при некоторых видах нелинейных преобразований, например при умножении частоты.

2. Амплитуды всех нечетных гармоник (за исключением первой) равны нулю при $\theta = 90^\circ$, что избавляет от нежелательных гармонических составляющих.

Порядок операций при использовании метода угла отсечки следующий.

1. Для выбранной рабочей точки по известным значениям амплитуды сигнала U , напряжения смещения E , напряжения отсечки U_0 и крутизны ВАХ S вычисляется угол отсечки:

$$\cos\theta = \frac{U_0 - E}{U}.$$

2. Для найденного угла отсечки θ определяется максимальное значение тока по формуле

$$i_{\max} = SU(1 - \cos\theta).$$

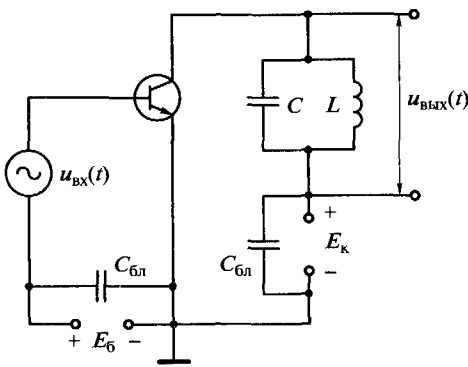
3. По графикам или таблицам определяются коэффициенты $\alpha_n(\theta)$.

4. Вычисляются амплитуды гармонических составляющих $I_n = \alpha_n(\theta) i_{\max}$.

3.4. Нелинейное усиление и умножение частоты

Задача усиления колебаний в общем виде формулируется следующим образом. На вход некоторого преобразователя поступает сигнал $S_1(t)$. На выходе необходимо получить сигнал $S_2(t) =$

Рис. 3.8. Нелинейная схема усиления (умножения) частоты



$= kS_1(t - \tau)$ при $k > 1$, который повторяет форму входного сигнала, но усиливается за счет энергии местного источника. Время запаздывания τ при этом не должно искажать форму сигнала.

Рассмотрим в качестве примера схему резонансного усилителя на транзисторе, контур на выходе которого настроен на частоту входного сигнала (рис. 3.8). На вход усилителя подается сигнал $u_{вх} = U_{вх} \cos \omega t$. Для вольт-амперной характеристики используем кусочно-линейную аппроксимацию.

Обратимся сначала к линейному усилению. Угол отсечки в этом случае $\theta = 180^\circ$. Рис. 3.9 соответствует этому режиму работы усилителя, так как входной сигнал не выходит за пределы линейного участка характеристики $i_k(u_{б-э})$. Поэтому ток на выходе усилителя изменяется по тому же закону, что и входной сигнал. Выходной

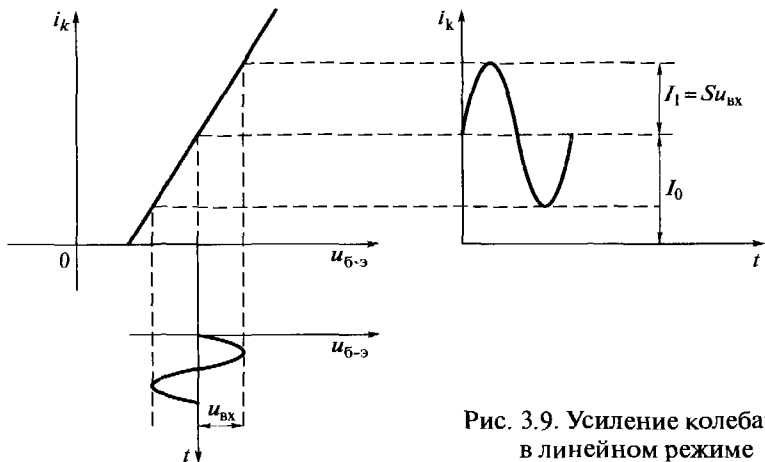


Рис. 3.9. Усиление колебаний в линейном режиме

ток содержит постоянную составляющую I_0 и переменную составляющую с амплитудой I_1 и частотой ω .

Выходное напряжение в контуре на выходе усилителя $u_{\text{вых}} = U_k \cos \omega t$, где $U_k = SR_3 U_{\text{вх}}$; S — крутизна вольт-амперной характеристики; R_3 — эквивалентное сопротивление контура. При $R_3 \gg 1/S$ получим коэффициент усиления по напряжению $k = U_k / U_{\text{вх}} = SR_3 \gg 1$.

Мощность, расходуемая источником энергии, $P_0 = I_0 E_k$.

Полезная мощность переменного тока в нагрузке $P_1 = 0,5 I_1 U_k$.

Тогда КПД усилителя $\eta = P_1 / P_0$, или $\eta = \frac{1}{2} \frac{I_1 U_k}{I_0 E_k}$.

В линейном режиме всегда $I_1 \leq I_0$, а амплитуда напряжения в контуре на выходе усилителя U_k не может быть больше напряжения источника энергии E_k , т.е. $U_k \leq E_k$. Если принять $I_1 = I_0$, а $U_k = E_k$, то в самом благоприятном случае коэффициент полезного действия не превысит значения $\eta = 0,5$.

Рассмотрим теперь нелинейный режим работы усилителя, который иллюстрирует рис. 3.10. В данном случае угол отсечки $\theta < 180^\circ$, отношение $I_1 / I_0 = \gamma_1 / \gamma_0 > 1$. Например, при $\theta = 60^\circ$ отношение $\gamma_1 / \gamma_0 \approx 1,8$ (из справочных таблиц или графиков). Тогда при $U_k = E_k$ коэффициент полезного действия $\eta \approx 0,8 \dots 0,9$. Угол отсечки менее 60° нежелателен, поскольку при этом существенно уменьшается γ_1 , а следовательно, и полезная мощность P_1 .

Из сказанного можно сделать следующий вывод: в нелинейных режимах усиления можно получить гораздо большие значения КПД, чем в линейных. Правда, при этом возникают нелинейные

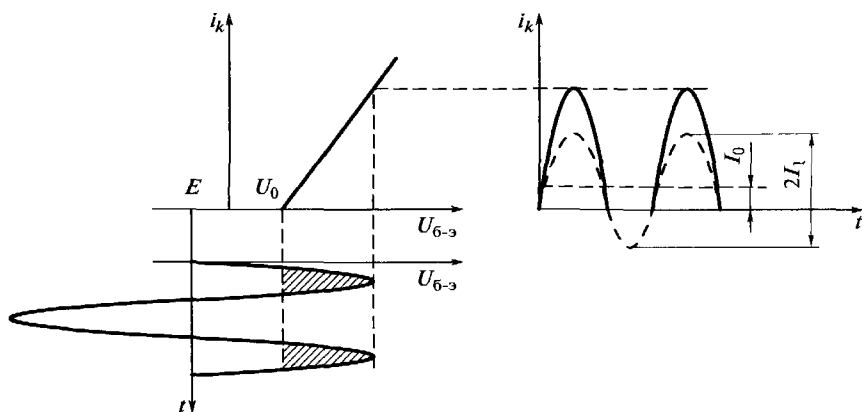


Рис. 3.10. Усиление колебаний в нелинейном режиме

искажения сигнала, связанные с появлением дополнительных гармонических составляющих.

Рассмотрим теперь умножение частоты. Под умножением частоты принято понимать формирование на выходе устройства колебания, частота которого в целое число раз больше частоты входного сигнала. Если на входе действует сигнал $u = U \cos \omega_0 t$, то на выходе должна быть частота колебания $\omega_{\text{вых}} = n \omega_0$. Принципиальным фактом здесь является наличие нелинейного преобразователя, поскольку умножение частоты связано с появлением новых спектральных составляющих, которых не было на входе. Реализовать умножитель частоты применением только линейных элементов невозможно.

Пусть нелинейный преобразователь аппроксимирован многочленом n -й степени и на его вход подается гармоническое колебание. Ранее было показано, что в этом случае выходной ток

$$i = a_0 + a_1 U \cos \omega t + a_2 U^2 \cos^2 \omega t + a_3 U^3 \cos^3 \omega t + \dots$$

После разложения данного соотношения по методу формул кратных дуг нетрудно увидеть, что выходной сигнал содержит вторую, третью и другие гармоники. Поэтому необходимую гармонику можно выделить с помощью фильтра, включаемого на выходе нелинейного преобразователя.

В качестве нелинейного преобразователя при умножении частоты можно использовать рассмотренный ранее нелинейный усилитель, только теперь в качестве нагрузки следует взять колебательный контур, настроенный на частоту n -й гармоники. На вход такого умножителя частоты подается колебание большой амплитуды, что приводит к нелинейным эффектам и позволяет использовать при расчете кусочно-линейную аппроксимацию и метод угла отсечки.

Ранее было показано, что амплитуда n -й гармоники $I_n = \alpha_n(\theta) i_{\text{max}}$. Если значение i_{max} задано, то коэффициент $\alpha_n(\theta)$ будет максимален при оптимальном угле отсечки θ , который выбирается из справочных таблиц или графиков. Причем максимум коэффициента $\alpha_n(\theta)$ имеет место, если $\theta_{\text{опт}} \approx 120^\circ / n$. Так, для удвоения частоты оптимальный угол отсечки должен быть равен 60° , а для утроения — $\theta \approx 40^\circ$.

Расчеты коэффициентов $\alpha_n(\theta)$ показывают, что они быстро уменьшаются с ростом номера гармоник, а следовательно, уменьшаются и амплитуды гармоник. Поэтому данную схему применяют только для умножения частоты в небольшое число раз, т. е. в 2—3 раза.

Для умножения частоты в большее число раз следует преобразовать гармоническое колебание в периодическую последователь-

ность прямоугольных импульсов. Известно, что в спектральном составе таких импульсов имеются колебания всех частот, кратных частоте их следования. Поэтому требуемая гармоника далее выделяется с помощью узкополосного фильтра.

3.5. Преобразование частоты сигнала

Под преобразованием частоты понимают перемещение спектра сигнала по шкале частот без изменения соотношений между его компонентами.

Иными словами, огибающая и фаза выходного сигнала должны повторять аналогичные параметры входного сигнала, т. е. если $S_{\text{вх}}(t) = U_c(t)\cos[\omega t + \varphi(t)]$, то $S_{\text{вых}}(t) = kU_c(t)\cos[\omega_0 t + \varphi(t)]$.

Основными устройствами, где применяется преобразование частоты, являются радиоприемные устройства, в которых осуществляется перенос спектра в область более низких частот в целях повышения эффективности фильтрации помех. Также преобразование частоты производится и во входных каскадах ретрансляторов радиорелейных линий связи для обеспечения фильтрации помех, после чего в их выходных каскадах реализуется восстановление передаваемого сигнала посредством преобразования его частоты до значения входного сигнала.

Преобразование частоты осуществляется в нелинейных или параметрических цепях.

Рассмотрим преобразование частоты с помощью параметрической цепи, ток в которой определяется выражением

$$i = Au_2u_1.$$

Данному соотношению соответствует схема, содержащая параметрический резистивный элемент с проводимостью $g = Au_2$, изменяющейся пропорционально сигналу u_2 (рис. 3.11).

Здесь могут применяться нелинейные элементы (полупроводниковые диоды, транзисторы), которые при воздействии сильного сигнала изменяют свои параметры. В качестве такого сигнала используют гармоническое колебание, получаемое от отдельного

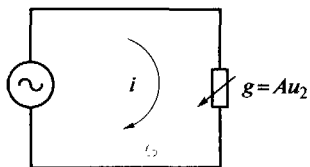


Рис. 3.11. Схема параметрической цепи с резистивным элементом

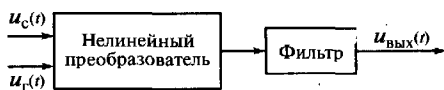


Рис. 3.12. Функциональная схема преобразователя частоты

генератора. (В радиоприемниках такой генератор называется гетеродином.)

Пусть на вход преобразователя, функциональная схема которого показана на рис. 3.12, поступает сигнал $u_c(t) = U_c \cos \omega_c t$, и гетеродин формирует гармоническое колебание $u_r(t) = U_r \cos \omega_r t$.

Подставив эти выражения в формулу тока, получим

$$i = AU_c U_r \cos \omega_c t \cos \omega_r t = \frac{1}{2} AU_c U_r \cos(\omega_c - \omega_r)t + \frac{1}{2} AU_c U_r \cos(\omega_c + \omega_r)t.$$

Для выделения любого из этих колебаний на выходе преобразователя частоты включается полосовой фильтр, настроенный либо на разностную, либо на суммарную частоту. В первом случае выполняется преобразование вниз, а во втором — преобразование вверх (рис. 3.13).

Отдельно остановимся на случае, когда частота колебания гетеродина больше частоты входного сигнала, т.е. $\omega_c - \omega_r < 0$ или $\omega_r > \omega_c$. Отрицательное значение частоты физического смысла не имеет, поэтому данное колебание, поскольку косинус — четная функция, представляют в следующем виде:

$$\frac{1}{2} AU_c U_r \cos(\omega_c - \omega_r)t = \frac{1}{2} AU_c U_r \cos(\omega_r - \omega_c)t.$$

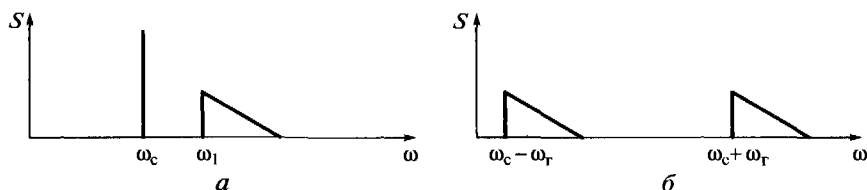


Рис. 3.13. Спектры сигналов:
а — исходного; б — преобразованных вверх и вниз

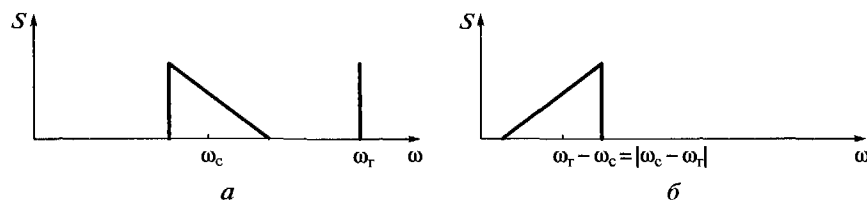


Рис. 3.14. Инверсия спектра в процессе преобразования частоты

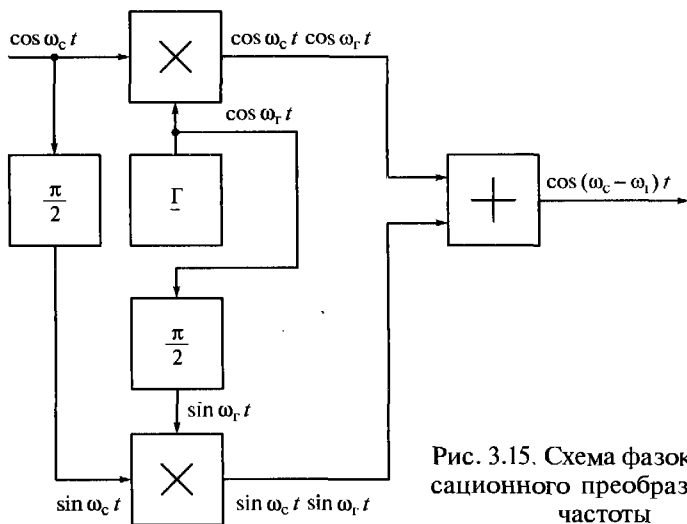


Рис. 3.15. Схема фазокомпенсационного преобразователя частоты

Результирующий спектр является зеркальным отображением спектра исходного сигнала, т.е. происходит инверсия спектра в процессе преобразований частоты (рис. 3.14).

Рассмотрим также метод преобразования частоты, соответствующий случаю, когда частотные составляющие на выходе нельзя разделить с помощью фильтра. Этот метод основан на соотношении

$$\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta.$$

Схема соответствующего преобразователя показана на рис. 3.15. Здесь фазовращатель — элемент, поворачивающий фазу на 90° ($\pi/2$), фактически реализует преобразование Гильберта. Такая схема называется фазокомпенсационным преобразователем, а в некоторых источниках ее называют оптимальным преобразователем.

3.6. Перемножение сигналов

Задача перемножения сигналов формулируется следующим образом: на входе нелинейного устройства действуют два сигнала — $S_1(t)$ и $S_2(t)$. При этом на выходе необходимо получить сигнал $S(t) = S_1(t)S_2(t)$. Операция перемножения сигналов применяется практически во всех нелинейных преобразованиях. Например, она использовалась в процессе преобразования частоты сигналов. Кроме того, на ее основе выполняются процедуры модуляции и демодуляции сигналов.

Пусть имеется нелинейный преобразователь, вольт-амперная характеристика которого аппроксимирована многочленом 2-й степени:

$$i(u) = a_0 + a_1 u + a_2 u^2.$$

На вход данного устройства поступает колебание, содержащее два гармонических сигнала с различными частотами (ω_1 и ω_2), т. е. можно записать

$$u(t) = S_1(t) + S_2(t) = U_1 \cos \omega_1 t + U_2 \cos \omega_2 t.$$

Подставив это выражение в предыдущее, получим

$$i(u) = a_0 + a_1 U_1 \cos \omega_1 t + a_1 U_2 \cos \omega_2 t + a_2 U_1^2 \cos^2 \omega_1 t + \\ + a_2 U_2^2 \cos^2 \omega_2 t + 2a_2 U_1 U_2 \cos \omega_1 t \cos \omega_2 t,$$

откуда видно, что отклик устройства содержит множество различных составляющих, в том числе и требуемое произведение. Совершенно очевидно, что все лишние составляющие должны быть устранены, для чего применяют балансные устройства и фильтры.

Схема балансного перемножителя в эквивалентном виде изображена на рис. 3.16. Выходное напряжение в данной схеме пропорционально разности токов $i_1 - i_2$, где

$$i_1 = a_0 + a_1(u_1 + u_2) + a_2(u_1 + u_2)^2;$$

$$i_2 = a_0 + a_1(u_1 - u_2) + a_2(u_1 - u_2)^2.$$

Таким образом, можно записать

$$u_{\text{ВЫХ}} = R(2a_1 u_2 + 4a_2 u_1 u_2),$$

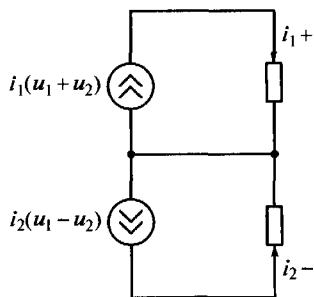


Рис. 3.16. Эквивалентная схема балансного перемножителя сигналов

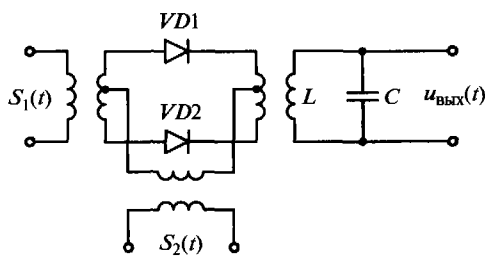


Рис. 3.17. Схема балансного диодного перемножителя сигналов

т. е. видно, что здесь имеется необходимое произведение двух входных сигналов.

Схема балансного перемножителя, реализованная на диодах, изображена на рис. 3.17. На выходе такой схемы включен фильтр для выделения необходимой составляющей произведения $u_1 u_2$.

Возможны и другие схемы перемножителей.

Контрольные вопросы

1. Назвать и пояснить методы нелинейных преобразований в каналах телекоммуникаций. Указать, какие для этого применяются электрические цепи.

2. В каких целях и какие методы аппроксимации нелинейных преобразователей применяются в электросвязи?

3. В чем заключается сущность степенной (полиномиальной) аппроксимации и как она связана с методом спектрального анализа на основе тригонометрических формул кратного аргумента?

4. Пояснить метод угла отсечки и назвать область его применения.

5. Пояснить, что представляют собой линейное и нелинейное усиления и почему в ряде случаев нелинейное усиление предпочтительнее.

6. В чем заключается смысл преобразования частоты и в каких случаях возникает инверсия спектра преобразуемого сигнала?

7. Что представляет собой и для чего применяется перемножение сигналов?

Глава 4

МОДУЛЯЦИЯ И ДЕМОДУЛЯЦИЯ СИГНАЛОВ

4.1. Общие сведения о модулированных сигналах

Как уже отмечалось, передача сигнала на большие расстояния производится высокочастотными колебаниями, т.е. гармоническими узкополосными квазигармоническими колебаниями вида

$$S(t) = A(t) \cos[\omega_c t + \varphi(t)].$$

Несущая частота сигнала ω_c выбирается с учетом:

- дальности связи;
- особенностей распространения электромагнитных колебаний;
- мощности передатчика и др.

В любом случае несущая частота ω_c должна быть значительно больше наивысшей частоты Ω первичного сигнала, который отображает сообщение, т.е. $\Omega \ll \omega_c$. Поэтому в одном периоде модулирующего сигнала $T = 2\pi/\Omega$ укладываются сотни, тысячи и более периодов высокочастотного колебания $T_c = 2\pi/\omega_c$.

Основные методы модуляции, которые давно известны, вытекают из квазигармонического представления узкополосных сигналов. При непрерывном первичном сигнале различают следующие методы аналоговой модуляции:

- линейные, к которым относят амплитудную модуляцию (АМ), балансную (БМ) и однополосную (ОМ);
- нелинейные, к которым относят угловую модуляцию.

Угловая модуляция называется частотной (ЧМ), если в соответствии со значениями модулирующего сигнала происходит изменение мгновенной частоты несущего колебания. Угловая модуляция называется фазовой (ФМ), если по закону передаваемого сообщения изменяется мгновенная фаза сигнала-переносчика.

Наряду с аналоговой модуляцией в технике электросвязи (в системах телекоммуникаций) применяются следующие виды дискретной модуляции с гармоническим несущим колебанием:

- дискретная амплитудная модуляция (ДАМ);

- дискретная частотная модуляция (ДЧМ);
- дискретная фазовая модуляция (ДФМ);
- относительная фазовая модуляция (ОФМ);
- квадратурная амплитудная модуляция (КАМ);
- амплитудно-фазовая модуляция (АФМ);
- дискретная частотная модуляция с непрерывной фазой (ДЧМНФ);
- дискретная частотная модуляция с минимальным сдвигом (ММС).

В общем случае при дискретной модуляции модулирующий сигнал может иметь несколько градаций $M = 2, 3, 4, \dots$. В случае если число градаций $M = 2$, модулирующим сигналом являются двоичные послылки (0, 1). Такую дискретную модуляцию называют манипуляцией.

Наряду с гармоническими колебаниями в качестве переносчика информации используются периодические импульсные последовательности. В этом случае выделяют следующие импульсные методы модуляции:

- амплитудно-импульсная модуляция (АИМ), при которой в соответствии с модулирующим сигналом изменяется амплитуда периодического импульсного процесса;
- широтно-импульсная модуляция (ШИМ), при которой изменяется длительность импульсов;
- времяимпульсная модуляция (ВИМ), также называемая фазоимпульсной модуляцией (ФИМ), при которой пропорционально модулирующему колебанию изменяется временное положение импульсов относительно их исходного состояния (тактовых точек). В этом случае время иногда называют фазой.

Наконец, в системах телекоммуникаций применяют аналогоцифровую модуляцию, при которой входной сигнал модулятора является непрерывным, а на выходе образуется двоичная дискретная последовательность импульсов. При этом различают:

- импульсно-кодovou модуляцию (ИКМ), при которой модулированный сигнал представляет собой двоичную запись квантованных по уровню значений исходного сигнала;
- дельта-модуляцию (ДМ), при которой определяется разность значений сигнала в текущий и предыдущий моменты времени, после чего эта разность квантуется на два уровня: +1, если последующее значение больше предыдущего, т.е. разность положительная, и -1, если последующее значение меньше предыдущего, т.е. разность отрицательная.

Безусловно, в настоящее время известны и другие методы модуляции и соответственно модулированные сигналы. Однако в виду ограниченности объема книги рассмотреть их все не представляется возможным.

4.2. Амплитудная модуляция

Амплитудной модуляцией называется изменение амплитуды несущего сигнала в соответствии с модулированным колебанием. Например, имеем высокочастотное несущее колебание $U_0 \cos \omega_0 t$ и первичный сигнал $A(t) = U_0 + ax(t)$, где U_0 — постоянная составляющая. Результирующий амплитудно-модулированный сигнал получим на основе перемножения несущего колебания и первичного сигнала:

$$S(t) = [U_0 + ax(t)] \cos \omega_0 t.$$

Пусть $x(t)$ является гармоническим колебанием с частотой Ω , т. е. $x(t) = X \cos \Omega t$. Тогда $S(t) = (U_0 + aX \cos \Omega t) \cos \omega_0 t$. Здесь $x(t)$ — медленно меняющаяся во времени функция по сравнению с высокочастотным колебанием ω_0 , т. е. $\Omega \ll \omega_0$.

Введем следующее обозначение: $U_\Omega = aX$ — максимальное приращение амплитуды огибающей.

Временные диаграммы, иллюстрирующие процесс амплитудной модуляции тональным колебанием, показаны на рис. 4.1.

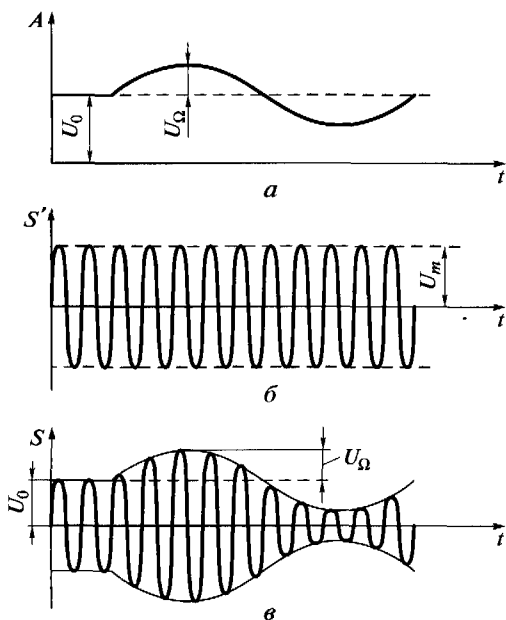


Рис. 4.1. Временные диаграммы, иллюстрирующие амплитудную модуляцию:

a — первичный сигнал; *б* — высокочастотное несущее колебание; *в* — модулированный сигнал

Коэффициентом модуляции называется отношение амплитуды U_{Ω} огибающей к амплитуде U_0 несущего колебания, т.е. $m = U_{\Omega}/U_0$. Обычно $0 < m < 1$.

Глубиной модуляции называется коэффициент модуляции, выраженный в процентах. Следовательно, можно записать

$$S(t) = U_0(1 + m \cos \Omega t) \cos \omega_0 t.$$

Раскроем данное выражение, что позволит определить спектр АМ-сигнала:

$$S(t) = U_0 \cos \omega_0 t + \frac{m}{2} U_0 \cos(\omega_0 + \Omega)t + \frac{m}{2} U_0 \cos(\omega_0 - \Omega)t.$$

Из этого выражения видно, что АМ-колебание, спектр которого при модуляции одним гармоническим сигналом изображен на рис. 4.2, содержит три составляющие:

- колебание несущей частоты ω_0 с амплитудой U_0 ;
- колебания верхней боковой частоты $\omega_0 + \Omega$ с амплитудой $U_b = mU_0/2$;
- колебания нижней боковой частоты $\omega_0 - \Omega$ с $U_n = mU_0/2$.

Из сказанного можно сделать следующие выводы.

1. Ширина спектра равна удвоенной частоте модуляции $\Delta\omega = 2\Omega$.
2. Амплитуда несущего колебания при модуляции не изменяется, а амплитуды колебаний боковых частот пропорциональны глубине модуляции, т.е. амплитуде модулирующего сигнала.
3. При $m = 1$ амплитуды колебаний боковых частот равны половине амплитуды несущего колебания, т.е. $0,5U_0$. При $m = 0$ боковые частоты отсутствуют, что соответствует немодулированному колебанию.

На практике однотоновые АМ-сигналы используются крайне редко. Более реален случай, когда низкочастотный модулированный сигнал имеет сложный спектральный состав:

$$A(t) = \sum_{k=1}^n X_k \cos(\Omega_k t + \varphi_k).$$

Здесь частоты Ω_k образуют упорядоченную возрастающую последовательность $\Omega_1 < \Omega_2 < \dots < \Omega_n$, а амплитуды X_k и фазы φ_k — произвольные.

В этом случае для АМ-сигнала можно записать следующее аналитическое соотношение:

$$S(t) = U_0 \left[1 + \sum_{k=1}^n m_k \cos(\Omega_k t + \varphi_k) \right] \times \cos(\omega_0 t + \varphi_0),$$

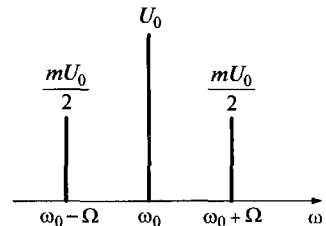


Рис. 4.2. Спектр колебаний при амплитудной модуляции одним низкочастотным гармоническим сигналом

где m_k — парциальные коэффициенты модуляции, представляющие собой коэффициенты модуляции соответствующих компонентов первичного сигнала.

Спектральное разложение производится так же, как и для однотонового АМ-сигнала:

$$S(t) = U_0 \cos(\omega_0 t + \varphi_0) + \sum_{k=1}^n \frac{m_k}{2} U_0 \cos[(\omega_0 + \Omega_k)t + \varphi_0 + \varphi_k] + \\ + \sum_{k=0}^n \frac{m_k}{2} U_0 \cos[(\omega_0 - \Omega_k)t + (\varphi_0 - \varphi_k)].$$

Из этого разложения видно, что в спектре кроме несущего колебания содержатся группы верхних и нижних боковых колебаний. При этом спектр верхних боковых колебаний является копией спектра модулирующего сигнала, сдвинутой в область высоких частот на значение ω_0 , а спектр нижних боковых колебаний располагается зеркально относительно ω_0 .

Спектры исходного полосового сигнала и амплитудно-модулированного сигнала показаны на рис. 4.3.

Определим мощность АМ-колебания, для чего рассмотрим вновь случай модуляции одной гармоникой. Будем считать, что $\omega_0 \gg \Omega$. В этом случае амплитуда $U(t) = U_0(1 + m \cos \Omega t)$ за период высокочастотного колебания практически не изменяется, по-

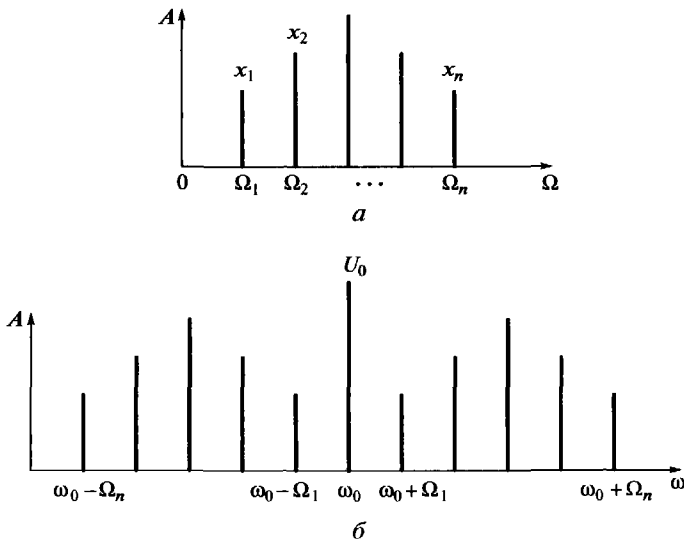


Рис. 4.3. Спектры исходного полосового (а) и амплитудно-модулированного (б) сигналов

этому среднюю мощность, выделяемую на сопротивлении 1 Ом в течение этого времени, запишем в виде

$$P_{\text{ср}\omega} = U^2(t)/2 = U_0^2(1 + m \cos \Omega t)^2/2 = P_0(1 + m \cos \Omega t)^2 = \\ = P_0(1 + 2m \cos \Omega t + m^2/2 + m^2 \cos 2\Omega t/2),$$

где $P_0 = U_0^2/2$ — мощность несущего колебания.

Из этой формулы видно, что, если $m \approx 1$, при $\Omega t = 0$ мощность $P_{\text{ср}} = P_0(1 + m)^2 \approx 4P_0$, а при $\Omega t = \pi$ мощность $P_{\text{ср}} = P_0(1 - m)^2 \approx 0$.

Таким образом, при 100%-й модуляции, когда $m = 1$, мощность АМ-колебания изменяется в пределах $0 < P_{\text{ср}} < 4P_0$.

Найдем теперь среднее значение мощности за период низкой частоты. В этом случае средняя мощность всего АМ-колебания есть сумма мощностей несущей частоты и двух боковых частот — нижней и верхней, следовательно, при сопротивлении 1 Ом нагрузки средняя мощность несущей частоты

$$P_0 = \frac{U_0^2}{2},$$

а каждая из боковых составляющих имеет мощность

$$P_i = P_{\text{н.б}} = \frac{1}{2} \left(\frac{m}{2} U_0 \right)^2 = \frac{m^2}{4} P_0.$$

Теперь несложно получить общую мощность АМ-сигнала за период колебания низкой частоты Ω :

$$P_{\text{ср}} = P_0 + P_{\text{в.б}} + P_{\text{н.б}} = P_0 \left(1 + \frac{m^2}{2} \right).$$

Из этой формулы видно, что при 100%-й модуляции 66,6 % всей мощности, излучаемой передатчиком, затрачивается на передачу несущей частоты и только 33,3 % мощности приходится на оба колебания боковых частот, которые как раз и содержат полезную информацию.

Следовательно, для более эффективного использования мощности передатчика целесообразно передавать модулированный сигнал без колебания несущей частоты. Кроме того, для уменьшения ширины спектра, занимаемого сигналом, желательно передавать только одну из боковых полос, поскольку оба боковых колебания содержат одну и ту же информацию.

4.3. Балансная модуляция

Как уже отмечалось, значительная доля мощности АМ-сигнала сосредоточена в несущем колебании, которое не несет никакой

полезной информации и в процессе модуляции не изменяется. Следовательно, для более эффективного использования мощности передатчика можно формировать модулированные сигналы с подавленным несущим колебанием.

Вид гармонической модуляции, в результате которой спектр сигнала содержит только две боковые полосы, где сосредоточена полезная информация, и не имеет колебания на несущей частоте называется *балансной модуляцией*. В отличие от АМ-сигнала при модуляции одной гармоникой представление БМ-сигнала имеет вид

$$S(t) = U_0 m \cos \Omega t \cos(\omega_0 t + \varphi_0) = \\ = \frac{U_0 m}{2} \cos[(\omega_0 + \Omega)t + \varphi_0] + \frac{U_0 m}{2} \cos[(\omega_0 - \Omega)t + \varphi_0],$$

т. е. здесь имеет место перемножение двух сигналов: модулирующего и несущего. Получаемые при этом колебания можно трактовать как биения двух гармонических сигналов с одинаковыми амплитудами $U_0 m/2$ и частотами, равными верхней и нижней боковой частотами.

В случае когда модулирующий сигнал содержит множество гармоник, БМ-сигнал принимает вид

$$S(t) = \frac{U_0}{2} \sum_{k=0}^n m_k \cos[(\omega_0 + \Omega_k)t + \varphi_0] + \frac{U_0}{2} \sum_{k=0}^n m_k \cos[(\omega_0 - \Omega_k)t + \varphi_0],$$

откуда видно, что здесь, как и при АМ-модуляции, имеются две симметричные группы верхних и нижних частот.

На рис. 4.4, *а* показан спектр сигнала при балансной модуляции одним низкочастотным гармоническим колебанием, а на рис. 4.4, *б* — его осциллограмма. Осциллограмма БМ-сигнала показывает наличие высокочастотного заполнения, однако здесь нет колебания несущей частоты, поскольку при переходе огибающей через нуль фаза высокочастотного заполнения изменяется скачком на 180° . Следовательно, если такой БМ-сигнал поступит

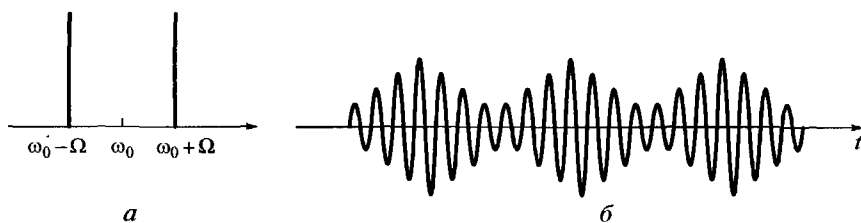


Рис. 4.4. Спектр (*а*) и осциллограмма (*б*) сигнала при балансной модуляции одним низкочастотным гармоническим колебанием

в колебательный контур, настроенный на частоту ω_0 , то колебания, возникающие в нем в текущий момент, будут погашены колебаниями последующего периода. Таким образом, выходной эффект контура оказывается практически минимальным.

Для формирования БМ-сигнала может применяться схема, представленная на рис. 4.5, в которой перемножение сигналов производится на основе рассмотренного метода.

В заключение отметим следующее. При балансной модуляции мощность передатчика используется эффективнее, чем при амплитудной модуляции, так как в этом случае нет затрат энергии на излучение несущей частоты. Однако БМ широкого применения не находит из-за сложности восстановления несущей частоты на приеме.

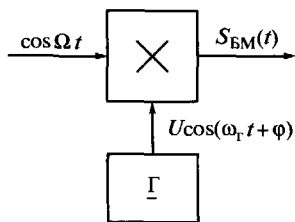


Рис. 4.5. Структурная схема балансного модулятора

4.4. Однополосная модуляция

Однополосной модуляцией называется вид гармонической модуляции, при которой энергия полезного сигнала сосредоточена только в одной из боковых полос: верхней или нижней. ОМ-сигнал можно записать как частный случай АМ-сигнала (без несущей и без одной из боковых полос). Если используется верхняя боковая полоса, то при модуляции одной гармоникой ОМ-сигнал можно записать в виде

$$S(t) = \frac{U_0 m}{2} \cos[(\omega_0 + \Omega)t + \varphi_0] = \\ = \frac{U_0 m}{2} [\cos \Omega t \cos(\omega_0 t + \varphi_0) - \sin \Omega t \sin(\omega_0 t + \varphi_0)].$$

При использовании нижней боковой полосы с частотой $\omega_0 - \Omega$ знак «+» перед Ω изменяется на «-».

В случае модуляции полосовым сигналом ОМ-сигнал имеет вид

$$S(t) = \frac{U_0 m}{2} \sum_{k=1}^n m_k \cos[(\omega_0 \pm \Omega_k)t + \varphi_0] = \\ = \frac{U_0}{2} \sum_{k=1}^n m_k \cos \Omega_k t \cos(\omega_0 t + \varphi_0) \mp \frac{U_0}{2} \sum_{k=1}^n m_k \sin \Omega_k t \sin(\omega_0 t + \varphi_0).$$

Однополосную модуляцию можно объяснить, используя понятие аналитического или гильбертового сигнала. Ранее уже упоминалось, что перенос частот (транспонирование спектра с сохранением его формы) можно трактовать как умножение гильбертового сигнала на $e^{j\omega_0 t}$, где ω_0 — несущая частота. Пусть первичный (низкочастотный) сигнал $\hat{x}(t) = x(t) + j\hat{x}(t)$, где \hat{x} — сигнал, сопряженный с $x(t)$.

Тогда

$$S(t) = \text{Re}\{\hat{x}(t)e^{j\omega_0 t}\} = \text{Re}\{x(t)e^{j\omega_0 t} + j\hat{x}(t)e^{j\omega_0 t}\} = \\ = \hat{x}(t)\cos\omega_0 t - \hat{x}(t)\sin\omega_0 t.$$

Приняв $x(t) = \frac{U_0 m}{2} \cos\Omega t$, увидим, что данное выражение со-

впадает с выражением для ОМ-сигнала с верхней боковой полосой. ОМ-сигнал с нижней боковой полосой можно получить, заменив сигнал $\hat{x}(t)$ на сопряженный с ним сигнал $\hat{x}^*(t)$. Следовательно, однополосная модуляция по существу является преобразованием частоты, т.е. сдвигом спектра первичного сигнала в область более высоких частот.

ОМ-сигнал можно сформировать с помощью схемы, показанной на рис. 4.6, которая аналогична схеме БМ-сигнала, только на ее выходе включен фильтр для выделения либо верхней боковой полосы, либо нижней. Для формирования ОМ-сигнала можно использовать и фазокомпенсационный метод преобразования частоты, рассмотренный ранее.

В заключение необходимо отметить следующее:

1. При однополосной модуляции в отличие от амплитудной изменяются и амплитуда, и фаза несущей частицы.

2. Однополосная модуляция находит широкое применение в технике электрической связи, так как обеспечивает эффективное использование мощности передатчика, а также излучаемой полосы частот, которая вдвое меньше, чем при амплитудной и балансной модуляциях.

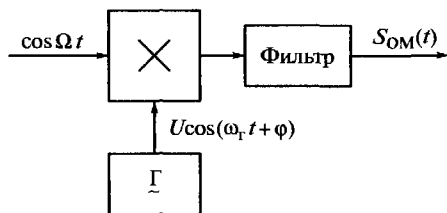


Рис. 4.6. Схема формирования сигнала с однополосной модуляцией

4.5. Формирование сигналов с амплитудной модуляцией

Амплитудным модулятором принято называть устройство, которое при подаче на его входы гармонического несущего колебания $U_n(t) = U_0 \cos \omega_0 t$ и низкочастотного модулирующего сигнала $U_{\text{мод}}(t) = U_m \cos \Omega t$ создает на выходе АМ-сигнал.

Амплитудные модуляторы реализуются на основе преобразования спектра суммы двух сигналов с помощью нелинейного безынерционного элемента. Схема простейшего амплитудного модулятора показана на рис. 4.7. В нем в качестве нелинейного элемента применяется диод, вольт-амперная характеристика которого аппроксимируется полиномом второй степени:

$$i = a_0 + a_1 u + a_2 u^2.$$

В схеме действует сумма двух напряжений

$$u = U_0 \cos \omega_0 t + U_m \cos \Omega t.$$

Выполнив необходимые подстановки, получим

$$i = a_0 + a_1(U_0 \cos \omega_0 t + U_m \cos \Omega t) + a_2(U_0^2 \cos^2 \omega_0 t + 2U_0 U_m \cos \Omega t \cos \omega_0 t + U_m^2 \cos^2 \Omega t).$$

Напомним, что $\cos^2 \alpha = \frac{1}{2}(1 + \cos 2\alpha)$, а $\cos \alpha \cos \beta = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$.

Следовательно, в спектре выходного тока содержатся первые и вторые гармоники частот ω_0 и Ω , а также их комбинационные частоты $\omega_0 \pm \Omega$. Для получения АМ-сигнала необходимо выделить колебания с частотами ω_0 , $(\omega_0 + \Omega)$, $(\omega_0 - \Omega)$, что реализуется с помощью колебательного контура, настроенного на частоту ω_0 . В этом случае на выходе схемы получают АМ-сигнал вида

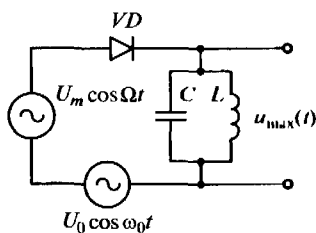
$$u_{\text{АМ}}(t) = i R_{\text{экв}} = a_1 R_{\text{экв}} U_0 \left(1 + \frac{2a_2 U_m}{a_1} \cos \Omega t\right) \cos \omega_0 t$$

или

$$u_{\text{АМ}}(t) = U_m (1 + m \cos \Omega t) \cos \omega_0 t,$$

где $U_{\text{АМ}} = a_1 R_{\text{экв}} U_0$; $m = 2 \frac{a_2}{a_1} U_m$; $R_{\text{экв}}$ — эквивалентное сопротивление контура.

Рис. 4.7. Схема простейшего амплитудного модулятора



Отметим, что на практике в качестве нелинейных элементов обычно применяют не диоды, а транзисторы или лампы (триоды и пентоды). При этом несущее колебание подается во входную цепь нелинейного элемента. Модулирующий сигнал в транзисторах подается в цепь базы или коллектора, а в электронных лампах — в цепь сетки (сеточная модуляция). Анализ процесса модуляции в этом случае производят на основе метода угла отсечки, который был рассмотрен ранее.

4.6. Методы демодуляции

4.6.1. Амплитудная демодуляция

Напомним, что под демодуляцией, или детектированием, понимают преобразование модулированного сигнала в первичный модулирующий сигнал. Данный процесс реализуется в нелинейных или параметрических устройствах, поскольку он связан с получением низкочастотных колебаний на основе высокочастотного сигнала.

Иными словами, когда на вход демодулятора подается АМ-сигнал вида $u_{\text{АМ}}(t) = U_0(1 + m \cos \Omega t) \cos \omega_0 t$, на выходе необходимо получить низкочастотный сигнал $U_{\Omega}(t) = U_m \cos \Omega t$, пропорциональный передаваемому сообщению.

Для демодуляции АМ-сигнала можно применить безынерционный нелинейный преобразователь, на выходе которого включен фильтр, пропускающий только низкочастотные составляющие спектра.

Обратимся к схеме простейшего диодного детектора (рис. 4.8). При этом рассмотрим отдельно два режима: слабых и сильных сигналов.

Режим слабых сигналов. Данный режим называют квадратичным детектированием, так как в этом случае вольт-амперную характеристику диода представляют полиномом второй степени:

$$i(u) = a_1 u + a_2 u^2.$$

Подставив выражение АМ-сигнала в этот полином, получим:

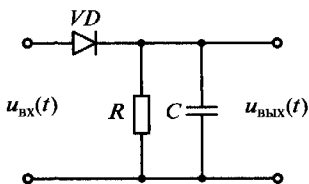


Рис. 4.8. Схема диодного детектора амплитудно-модулированного сигнала

$$\begin{aligned}
 i(u) &= a_1 U_0 (1 + m \cos \Omega t) \cos \omega_0 t + a_2 U_0^2 (1 + m \cos \Omega t)^2 \cos^2 \omega_0 t = \\
 &= a_1 U_0 (1 + m \cos \Omega t) \cos \omega_0 t + \frac{a_2 U_0^2}{2} (1 + m \cos \Omega t)^2 + \\
 &\quad + \frac{a_2 U_0^2}{2} (1 + m \cos \Omega t)^2 \cos 2\omega_0 t.
 \end{aligned}$$

Поскольку на выходе нелинейного преобразователя установлен низкочастотный RC -фильтр, то высокочастотные составляющие можно исключить из рассмотрения. Следовательно, результатом детектирования будем считать колебание вида

$$\begin{aligned}
 i(u) &= \frac{a_2 U_0^2}{2} (1 + m \cos \Omega t)^2 = \frac{a_2 U_0^2}{2} + \frac{a_2 U_0^2}{2} m^2 \cos^2 \Omega t + \\
 &+ a_2 U_0^2 m \cos \Omega t = \frac{a_2 U_0^2}{2} + \frac{a_2 U_0^2 m^2}{4} + \frac{a_2 U_0^2 m^2}{4} \cos 2\Omega t + \\
 &\quad + a_2 U_0^2 m \cos \Omega t.
 \end{aligned}$$

Постоянные составляющие (первые два слагаемых в этом выражении) также легко отфильтровываются разделительным конденсатором, поэтому на выходе демодулятора получим сигнал вида

$$u(t) = i(u)R = a_2 U_0^2 R m \cos \Omega t + \frac{1}{4} a_2 U_0^2 R m^2 \cos 2\Omega t.$$

Данный выходной сигнал содержит полезную составляющую $a_2 U_0^2 R m \cos \Omega t$, которая повторяет закон модулирующего сигнала. Однако здесь также появилось колебание с удвоенной частотой модуляции, которого не было при передаче сигнала. Данное слагаемое, являющееся следствием квадратичности вольт-амперной характеристики диода, представляет собой нелинейное искажение, которое оценивается коэффициентом гармоник (коэффициентом нелинейных искажений)

$$k = \frac{\frac{a_2 U_0^2 m^2}{4}}{a_2 U_0^2 m} = \frac{m}{4}.$$

Легко увидеть, что при 100%-й модуляции коэффициент гармоник достигает 25 %, т.е. оказывается чрезвычайно большим, что свидетельствует о значительных нелинейных искажениях даже в случае однотоновой модуляции.

Если передается сложный полосовой сигнал, который содержит в спектре большое число частот, то при детектировании возникает значительное количество гармоник и комбинационных

составляющих, которые при глубокой модуляции оказывают очень сильное влияние на разборчивость этого сигнала. Поэтому для качественной демодуляции телефонных сигналов или музыки квадратичное детектирование не применяется. Обычно в таких случаях входное колебание усиливают, а затем осуществляют демодуляцию в режиме большого сигнала.

Режим большого сигнала. Вновь рассмотрим диодный детектор, на выходе которого включен фильтр в виде параллельной RC -цепи. При этом параметры фильтра выбираются из условия $1/(\omega_0 C) \ll R \ll 1/(\Omega C)$, так как только в этом случае он будет способен подавлять высокочастотные спектральные составляющие. Кроме того, для нормальной работы демодулятора необходимо выбрать большое сопротивление нагрузки R , которое должно значительно превышать сопротивление диода R_d в прямом направлении, т. е. должно выполняться условие $R \gg R_d$.

Воспользуемся методом кусочно-линейной аппроксимации для описания вольт-амперной характеристики диода и запишем следующее соотношение для тока:

$$i(U) = \begin{cases} SU, & U \geq 0; \\ 0, & U < 0. \end{cases}$$

Для упрощения будем считать, что в отличие от предыдущего режима здесь напряжение смещения $U_0 = 0$.

Пусть на вход демодулятора поступает обычное гармоническое колебание вида

$$U(t) = U \cos \omega_0 t; \quad \omega_0 = 2\pi / T_0.$$

При этом сначала происходит зарядка конденсатора через открытый диод, а затем его разрядка через сопротивление нагрузки, причем с учетом заданного ранее соотношения сопротивлений диода и нагрузки разрядка конденсатора происходит медленнее, чем зарядка. Поэтому выходной сигнал будет представлять собой пилообразную кривую с относительно малой высотой зубцов, а средний уровень выходного напряжения U_0 окажется близким к амплитуде входного сигнала. В результате большую часть времени диод будет заперт, поскольку выходное напряжение приложено к нему в обратном направлении и служит для него напряжением смещения U_0 .

Пренебрегая пульсациями выходного напряжения, будем считать, что оно примерно постоянно, т. е. $U_{\text{вых}} = -U_0 = -I_0 R$, где I_0 — постоянная составляющая выходного тока. Следовательно, ток может протекать только если $U_{\text{вх}} > U_{\text{вых}}$, что иллюстрирует рис. 4.9. В этом случае, когда огибающая амплитуды входного сигнала претерпевает изменения, сигнал на выходе демодулятора будет синхронно повторять эти колебания.

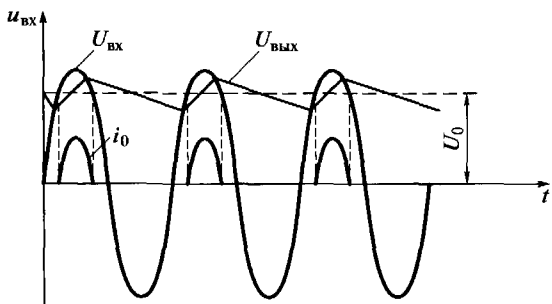


Рис. 4.9. Пояснение процесса детектирования амплитудно-модулированного сигнала

Рассмотрим теперь рис. 4.10, иллюстрирующий определение выходного тока с учетом кусочно-линейной аппроксимации и влияния напряжения смещения U_0 . Здесь при больших сопротивлениях R нагрузки диод будет работать с небольшими углами отсечки θ , т. е. ток будет протекать только в течение части периода, соответствующей заштрихованной области сигнала.

Найдем угол отсечки θ , для чего запишем выражение для напряжения на диоде:

$$u = U \cos \omega_0 t - I_0 R.$$

Тогда при $u = 0$ угол отсечки θ , равный значению $\omega_0 t$, можно определить в виде

$$\cos \theta = \frac{I_0 R}{U}. \quad (4.1)$$

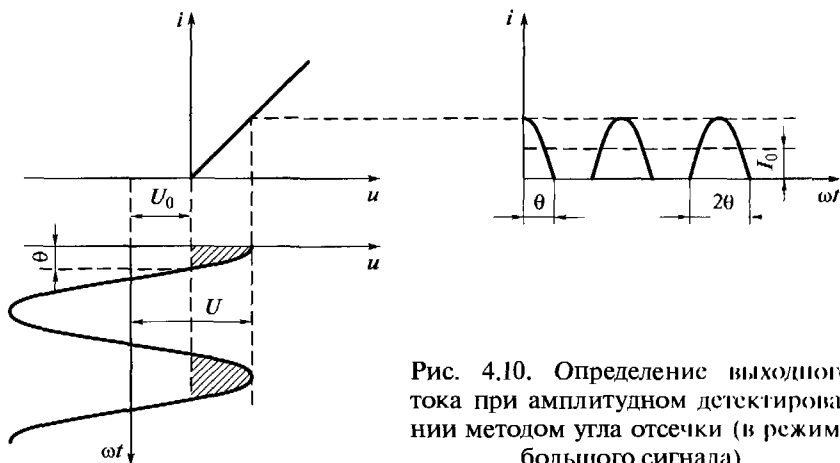


Рис. 4.10. Определение выходного тока при амплитудном детектировании методом угла отсечки (в режиме большого сигнала)

В соответствии с методом угла отсечки постоянная составляющая тока

$$I_0 = SU \gamma_0(\theta) = SU \frac{1}{\pi} (\sin \theta - \theta \cos \theta).$$

Подставив данное выражение в формулу (4.1), получим

$$\cos \theta = (SR / \pi)(\sin \theta - \theta \cos \theta)$$

или

$$\operatorname{tg} \theta - \theta = \frac{\pi}{SR}.$$

В полученные выражения амплитуда сигнала не входит. Следовательно, угол отсечки не зависит от параметров демодулируемого сигнала, а определяется исключительно параметрами схемы: крутизной характеристики диода S и сопротивлением нагрузки R . Тогда согласно формуле (4.1) выходной сигнал (ток I_0) будет прямо пропорционален входному колебанию. Иными словами, детектор обладает линейной характеристикой, и процесс демодуляции осуществляется практически без искажений. Рассмотренный метод демодуляции принято называть линейным детектированием. Однако при этом следует помнить, что линейный детектор является нелинейным устройством, которое работает с отсечкой тока.

Диодные детекторы считаются квадратичными (в режиме слабого сигнала) при амплитудах входных сигналов $U \leq 0,1 \dots 0,2$ В и линейными (в режиме большого сигнала) при $U > 0,5 \dots 1$ В.

4.6.2. Детектирование сигналов с балансной и однополосной модуляцией

Ранее было показано, что сигналы с балансной модуляцией в отличие от АМ-сигналов не содержат колебаний на несущей частоте. Поэтому при их демодуляции необходимо использовать особый вид преобразования, при котором в спектр БМ-сигнала вводится недостающее колебание, формируемое с помощью отдельного генератора. Такой метод приема сигнала был предложен в 1930-е гг. и получил название *синхронного детектирования*. Его автором является советский ученый Е. Г. Момот.

Для реализации данного метода принимаемый сигнал $S(t)$ перемножается с колебанием $A(t)$, которое вырабатывается генератором приемника (гетеродином). Затем полученный сигнал пропускается через фильтр нижних частот. Для перемножения при-

меняется нелинейное устройство, аналогичное балансному преобразователю, рассмотренному ранее. Фактически схема демодулятора идентична схеме модулятора БМ-сигнала (см. рис. 4.5), и отличается только тем, что при передаче на вход перемножающего устройства поступает низкочастотное модулирующее колебание, а на приемной стороне получают прошедший по каналу связи высокочастотный радиосигнал. Кроме того, на выходе демодулятора имеется фильтр нижних частот.

Рассмотрим подробнее процедуру детектирования БМ-сигнала. При этом влиянием помех для упрощения анализа здесь пренебрегаем.

Пусть на вход демодулятора поступает сигнал $S(t) = b(t) \cos(\omega_c t + \varphi_c)$, где $b(t) = U_c \cos \Omega t$. Предположим сначала, что местное колебание $A(t) = \cos(\omega_r t + \varphi_r)$, формируемое гетеродином, отличается по частоте и фазе от принимаемого сигнала, т.е. $\omega_r \neq \omega_c$, $\varphi_c \neq \varphi_r$. Тогда после перемножения получим

$$\begin{aligned} S(t)A(t) &= b(t) \cos(\omega_c t + \varphi_c) \cos(\omega_r t + \varphi_r) = \\ &= 0,5b(t) \{ \cos[(\omega_c - \omega_r)t + \varphi_c - \varphi_r] + \cos[(\omega_c + \omega_r)t + \varphi_c + \varphi_r] \}. \end{aligned}$$

В случае когда удастся обеспечить равенство частот сигнала и гетеродина ($\omega_r = \omega_c$) на выходе фильтра остается компонента

$$S(t)A(t) = 0,5b(t) \cos \Delta\varphi = 0,5U_c \cos \Omega \cos \Delta\varphi,$$

где $\Delta\varphi = \varphi_c - \varphi_r$ — разность фаз сигнала и колебания, вырабатываемого гетеродином.

Иными словами, с помощью фильтра можно выделить основной низкочастотный сигнал, содержащий передаваемое сообщение, и отфильтровать колебание с удвоенной частотой сигнала $2\omega_c = \omega_c + \omega_r$ при $\omega_r = \omega_c$.

В этом случае передаваемый сигнал $b(t) = U_c \cos \Omega t$ правильно воспроизводится при приеме, однако его амплитуда зависит от косинуса фазовой ошибки $\Delta\varphi$. Так, при $\Delta\varphi = 60^\circ$ амплитуда полезного сигнала уменьшается вдвое, а при $\Delta\varphi = 90^\circ$ переданное сообщение полностью теряется, т.е. его правильный прием становится невозможным. Максимум амплитуды сигнала достигается при равенстве фазовой ошибки нулю ($\Delta\varphi = 0$), что соответствует точной синхронизации частоты и фазы местного гетеродина относительно принимаемого сигнала. Обеспечение этого и является самым сложным при реализации синхронного детектирования, поскольку БМ-сигнал не содержит несущего колебания и его необходимо воспроизвести с высокой степенью точности для выполнения демодуляции.

Для решения данной задачи существует три метода:

- передача несущей частоты по отдельному каналу;

- восстановление несущего колебания из спектральных составляющих боковых полос посредством нелинейного преобразования;

- передача пилот-сигнала, т. е. несущего колебания со значительно уменьшенной амплитудой.

Восстановленное при приеме недостающее колебание далее используется в системе автоподстройки для поддержания частоты и фазы гетеродина в необходимых пределах. Однако в любом случае схема приемника БМ-сигналов получается весьма сложной. Зачастую синхронное детектирование просто нереализуемо вследствие значительных флюктуаций параметров сигнала в среде распространения радиоволн.

Рассмотрим теперь процедуру демодуляции однополосных сигналов, при практической реализации которой можно использовать такую же функциональную схему, как и в случае балансной модуляции. Пусть на вход демодулятора поступает ОМ-сигнал $S(t) = U_c \cos[(\omega_c + \Omega)t + \varphi_c]$ с верхней боковой полосой. Тогда на выходе перемножающего устройства получим

$$\begin{aligned} S(t)A(t) &= U_c \cos[(\omega_c + \Omega)t + \varphi_c] \cos(\omega_r t + \varphi_r) = \\ &= 0,5 U_c \{ \cos[(\omega_c + \Omega - \omega_r)t + \varphi_c - \varphi_r] + \cos[(\omega_c + \Omega + \omega_r)t + \varphi_c + \varphi_r] \}. \end{aligned}$$

Если частоты принимаемого сигнала и гетеродина равны ($\omega_r = \omega_c$), то на выходе фильтра получим колебание

$$S(t)A(t) = 0,5 U_c \cos(\Omega t + \Delta\varphi); \Delta\varphi = \varphi_c - \varphi_r,$$

откуда видно, что сигнал в результате демодуляции оказался перемещенным на свое исходное место на оси частот, а колебание с удвоенной частотой через фильтр не проходит. Вместе с тем в результирующем сигнале появился дополнительный сдвиг фаз, который оказывается эквивалентным временной задержке в процессе передачи сигнала. Небольшое значение $\Delta\varphi$ не оказывает существенного влияния на речевое сообщение, т. е. на слух оно не воспринимается, и им можно пренебречь.

Кроме того, в отличие от балансной модуляции при детектировании ОМ-сигналов требования к точности частоты также не являются очень жесткими. Так, несовпадение частот сигнала и гетеродина в 1...2 Гц при передаче речи и музыки на слух совершенно не заметны, но неточность в 3...10 Гц уже не позволяет осуществлять качественную передачу музыки. Для телефонного сигнала расхождение частот на 50...100 Гц и немного больше не приводит к значительным искажениям сигнала, т. е. хотя при этом и появляются хрипы, разборчивость речи еще сохраняется. С учетом сказанного отметим, что однополосная модуляция находит широкое применение в различных системах многоканальной свя-

зи и в системах коротковолновой радиосвязи, где частотный ресурс является крайне ограниченным.

4.7. Угловая модуляция

Пусть имеется гармоническое высокочастотное колебание $S(t) = U_c \cos(\omega_c t + \varphi_c)$, которое можно записать в виде $S(t) = U_c \cos \psi(t)$, где $\psi(t) = \omega_c t + \varphi_c$ — полная мгновенная фаза, определяющая текущее значение фазового угла. Отсюда вытекает следующее определение: вид гармонической модуляции, при которой под воздействием управляющего сигнала изменяется параметр $\psi(t)$ модулируемого колебания, а его амплитуда при этом сохраняется неизменной, называется *угловой модуляцией*. Такая модуляция реализуется в двух вариантах: как *фазовая* и как *частотная*. В первом варианте изменениям подвергается фаза несущего колебания, а во втором — по закону управляющего сигнала изменяется частота сигнала — переносчика информации.

Рассмотрим сначала фазовую модуляцию. Пусть управляющим (модулирующим) является гармонический сигнал

$$x(t) = U_\Omega \sin \Omega t.$$

Тогда мгновенная (полная) фаза исходного несущего колебания будет иметь вид

$$\psi(t) = \omega_c t + \varphi_c + a U_\Omega \sin \Omega t,$$

где φ_c — начальная фаза высокочастотного колебания; a — коэффициент пропорциональности.

В данном выражении первые два слагаемых определяют фазу исходного немодулированного колебания, а третье слагаемое показывает изменение фазы колебания в результате модуляции. Будем называть *индексом модуляции* максимальное отклонение фазы модулированного колебания от фазы немодулированного колебания:

$$M = \Delta \varphi_{\max} = a U_\Omega.$$

Индекс модуляции, как следует из данного выражения, пропорционален амплитуде U_Ω модулирующего сигнала и играет такую же роль, как и коэффициент модуляции в выражениях для АМ-сигналов.

С учетом введенных обозначений ФМ-сигнал примет вид

$$S(t) = U_c \cos(\omega_c t + \varphi_c + M \sin \Omega t),$$

откуда его мгновенную частоту можно определить как произвольную от фазы:

$$\omega(t) = \frac{d\psi(t)}{dt} = \omega_c + M\Omega \cos \Omega t.$$

Нетрудно заметить, что ФМ-сигнал в различные моменты времени имеет разные значения мгновенной частоты, которые отличаются от частоты несущего колебания ω_c на значения $\Delta\omega = M\Omega \cos \Omega t$. Следовательно, данный сигнал можно рассматривать и как колебание, модулированное по частоте. Наибольшее отклонение частоты ω от значения несущей частоты ω_c получило название *девиации частоты*:

$$\Delta\omega_d = M\Omega.$$

Перейдем теперь к частотной модуляции, при которой, как отмечалось ранее, изменяется мгновенная частота сигнала в соответствии с изменениями управляющего низкочастотного колебания:

$$\omega = \omega_c + ax(t),$$

где a — коэффициент пропорциональности; $x(t) = U_\Omega \cos \Omega t$ — гармонический низкочастотный сигнал.

Так же, как и при фазовой модуляции, здесь можно ввести понятие девиации частоты, которое в данном случае описывается выражением

$$\Delta\omega_d = aU_\Omega$$

и характеризует максимальное изменение частоты относительно ее исходного значения ω_c . Тогда мгновенную частоту можно записать в виде

$$\omega(t) = \omega_c + \Delta\omega_d \cos \Omega t.$$

Нетрудно определить и фазу колебания. Поскольку частота характеризует скорость изменения фазы, то ее значение можно найти, проинтегрировав последнее выражение:

$$\psi(t) = \int_0^t \omega dt = \omega_c t + \frac{\Delta\omega_d}{\Omega} \sin \Omega t + \varphi_c.$$

Таким образом получим аналитическое выражение ЧМ-сигнала в виде

$$S(t) = U_c \cos \left(\omega_c t + \varphi_c + \frac{\Delta\omega_d}{\Omega} \sin \Omega t \right).$$

Здесь слагаемое $(\Delta\omega_d/\Omega) \sin \Omega t$ в составе аргумента косинуса характеризует изменения мгновенной фазы в процессе частотной модуляции. Следовательно, ЧМ-сигнал можно отождествить с ФМ-сигналом, у которого индекс модуляции

$$M = \Delta\omega_d / \Omega.$$

Тогда окончательно выражение для ЧМ-сигнала примет вид

$$S(t) = U_c \cos(\omega_c t + \varphi_c + M \sin \Omega t),$$

т.е. оно практически совпадает с выражением для фазовой модуляции.

Из сказанного следует, что фазовая и частотная модуляции имеют много общего и характеризуются одними и теми же параметрами: индексом модуляции и девиацией частоты. Осциллограммы этих видов сигналов при модуляции посредством гармонического колебания практически не отличаются. Однако наряду с отмеченным сходством частотная и фазовая модуляции имеют следующие принципиальные различия:

- при фазовой модуляции индекс M пропорционален амплитуде низкочастотного колебания и не зависит от частоты Ω , а девиация $\Delta\omega_d$, наоборот, связана с частотой модулирующего сигнала прямой пропорциональной зависимостью;
- при частотной модуляции девиация частоты $\Delta\omega_d$ зависит только от амплитуды модулирующего колебания U_Ω и не связана с его частотой. Индекс модуляции в этом случае обратно пропорционален низкой частоте управляющего сигнала Ω .

Зависимости параметров фазовой и частотной видов модуляций от частоты управляющего колебания представлены на рис. 4.11, причем в обоих случаях считается, что модулирующий сигнал имеет одинаковую амплитуду, а частота Ω изменяется в некоторых заданных пределах.

Отметим, что различия между частотной и фазовой модуляциями проявляются наиболее заметно, если в качестве управляющего используется сложный полосовой сигнал с большим числом спектральных составляющих на разных частотах.

Определим среднюю мощность сигнала при угловой модуляции. Поскольку $\Omega \ll \omega$, можно считать, что в пределах периода

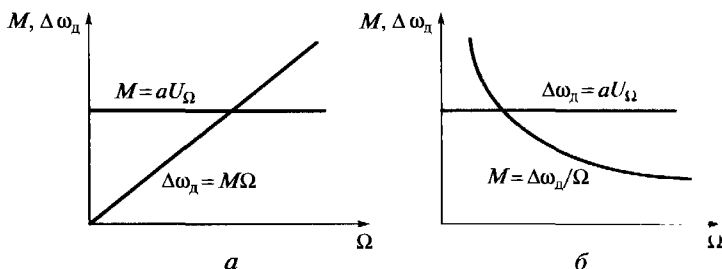


Рис. 4.11. Зависимости параметров фазовой (а) и частотной (б) модуляций от частоты управляющего низкочастотного колебания

высокой частоты $T = 2\pi / \omega$ колебания являются гармоническими с неизменной амплитудой. Средняя мощность таких сигналов на нагрузке с сопротивлением 1 Ом

$$P_{\text{ср}} = \frac{1}{T} \int_0^T S^2(t) dt = \frac{1}{2} U_c^2.$$

Точно таким же будет значение мощности и в другие периоды T , а также и в течение всего периода колебания низкой частоты Ω . Следовательно, при угловой модуляции средняя мощность будет такой же, как и при отсутствии модуляции, произойдет лишь ее перераспределение между спектральными составляющими.

Теперь найдем спектр сигнала при угловой модуляции. Обычно рассматривают *узкополосную* и *широкополосную* угловую модуляции. В первом случае считается, что индекс модуляции $M < 0,5$ рад, а во втором — $M > 0,5$ рад. Чаще всего в системах телекоммуникаций применяется широкополосная частотная модуляция с индексом $M \gg 1$, так как она более помехоустойчивая.

Рассмотрим сначала узкополосную угловую модуляцию, обратившись вновь к общему соотношению, которое имеет один и тот же вид при обоих вариантах модуляции — ФМ и ЧМ. Для удобства преобразований будем считать $\varphi_c = 0$ и, воспользовавшись известным тригонометрическим выражением, запишем

$$S(t) = U_c \cos(M \sin \Omega t) \cos \omega_c t - U_c \sin(M \sin \Omega t) \sin \omega_c t.$$

Приняв $M \ll 1$, можно использовать следующие приближенные равенства:

$$\sin(M \sin \Omega t) \approx M \sin \Omega t; \quad \cos(M \sin \Omega t) \approx 1.$$

В результате получим

$$\begin{aligned} S(t) &= U_c \cos \omega_c t - U_c M \sin \Omega t \sin \omega_c t = \\ &= U_c \cos \omega_c t + \frac{U_c M}{2} \cos(\omega_c + \Omega)t - \frac{U_c M}{2} \cos(\omega_c - \Omega)t. \end{aligned}$$

Из этого выражения видно, что спектр сигнала при узкополосной угловой модуляции похож на спектр простейшего АМ-сигнала, показанного на рис 4.2. В данном спектре также содержится несущее колебание и две боковые составляющие, частоты которых равны $\omega_c + \Omega$ и $\omega_c - \Omega$. Индекс модуляции M в данном случае играет ту же роль, что и коэффициент амплитудной модуляции m . Ширина спектра при узкополосной угловой модуляции также равна удвоенной частоте модуляции. Однако наблюдается существенное отличие ее от АМ-сигнала: нижняя боковая составляющая имеет дополнительный сдвиг фазы 180° . Спектральная диаграмма сигнала при узкополосной угловой модуляции в случае если $M \ll 1$ имеет вид, изображенный на рис. 4.12.

Рассмотрим широкополосную угловую модуляцию. Теперь, когда $M > 1$, последнее выражение, которое похоже на выражение для АМ-сигнала, уже не может использоваться для нахождения спектра угловой модуляции. Поэтому будем исходить из того, что функции $\cos(M \sin \Omega t)$ и $\sin(M \sin \Omega t)$ являются периодическими и их можно разложить в ряды Фурье. Причем первая из этих функций является четной, а вторая — нечетной.

Из теории функций Бесселя известно, что ряды Фурье для данных функций имеют следующий вид:

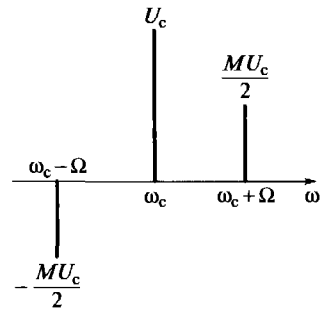


Рис. 4.12. Спектральная диаграмма сигнала с узкополосной угловой модуляцией

$$\begin{aligned} \cos(M \sin \Omega t) &= J_0(M) + 2J_2(M) \cos 2\Omega t + 2J_4(M) \cos 4\Omega t + \dots = \\ &= J_0(M) + 2 \sum_{n=1}^{\infty} J_{2n}(M) \cos 2n\Omega t; \end{aligned}$$

$$\begin{aligned} \sin(M \sin \Omega t) &= 2J_1(M) \sin \Omega t + 2J_3(M) \sin 3\Omega t + \dots = \\ &= 2 \sum_{n=0}^{\infty} J_{2n+1}(M) [\sin(2n+1)\Omega t], \end{aligned}$$

где $J_n(M) = \sum_{m=1}^{\infty} \frac{(-1)^m (M/2)^{2m+n}}{m!(m+n)!}$ — функция Бесселя первого рода n -го порядка от аргумента M .

Для этих функций, как уже отмечалось, в специальной литературе имеются подробные справочные таблицы и графики, а по виду они похожи на затухающие гармонические колебания.

Подставим эти выражения в формулу для сигнала с угловой модуляцией:

$$\begin{aligned} S(t) &= U_c [J_0(M) + 2 \sum_{n=1}^{\infty} J_{2n}(M) \cos 2n\Omega t] \cos \omega_c t - \\ &- U_c \{ 2 \sum_{n=0}^{\infty} J_{2n+1}(M) \sin[(2n+1)\Omega t] \} \sin \omega_c t. \end{aligned}$$

Наконец, после несложных преобразований окончательно получим выражение:

$$\begin{aligned} S(t) &= U_c [J_0(M) \cos \omega_c t + J_1(M) \cos(\omega_c + \Omega)t - J_1(M) \cos(\omega_c - \Omega)t + \\ &+ J_2(M) \cos(\omega_c + 2\Omega)t + J_2(M) \cos(\omega_c - 2\Omega)t + \\ &+ J_3(M) \cos(\omega_c + 3\Omega)t - J_3(M) \cos(\omega_c - 3\Omega)t + \dots], \end{aligned}$$

откуда видно, что спектр сигналов с угловой модуляцией является дискретным. Состоит такой спектр из несущего колебания с частотой ω_c и амплитудой $U_c J_0(M)$ и бесконечного числа симметрично расположенных боковых составляющих с частотами $\omega_c \pm n\Omega$ и амплитудами $U_c J_n(M)$. В отличие от АМ-сигналов при угловой модуляции амплитуда несущего колебания может принимать разные значения вследствие колебательного характера функций Бесселя. В некоторых случаях несущее колебание вообще отсутствует. Исходя из значений функций Бесселя, это может происходить, например при индексах модуляции $M = 2, 4, 5, 6$ и др. Кроме того, отдельные боковые составляющие также могут обращаться в нуль в зависимости от выбора индекса модуляции и соответствующего значения функции Бесселя.

В целом с увеличением индекса модуляции полоса частот, занимаемая сигналом, увеличивается, поэтому теоретически спектр сигналов с угловой модуляцией является бесконечно широким. На практике ширину спектра ограничивают значением $\Delta\omega \approx 2\Delta\omega_d$, где $\Delta\omega_d$ — девиация частоты.

Индекс модуляции при этом составляет порядка десятков или сотен единиц. В таких случаях ФМ- и ЧМ-сигналы оказываются широкополосными и могут использоваться только в тех диапазонах, которые обладают большим частотным ресурсом. В первую очередь, в диапазоне ультракоротких волн (УКВ), т. е. метровых и более коротких (частоты порядка десятков и сотен мегагерц).

В завершение рассмотрим, как влияют параметры модулирующего колебания $x(t) = U_\Omega \cos \Omega t$ на спектры сигналов с фазовой и частотной модуляцией, для чего используем приближенное выражение для ширины спектра.

Проведенный анализ показывает:

- изменения амплитуды U_Ω модулирующего сигнала приводят к одинаковым трансформациям ширины спектра в обоих случаях. С возрастанием амплитуды низкочастотного колебания происходит пропорциональное увеличение индекса модуляции. В результате спектры ФМ- и ЧМ-сигналов расширяются за счет увеличения числа спектральных составляющих;

- изменения частоты Ω модулирующего колебания оказывают различное влияние на ширину спектров сигналов с угловой модуляцией. Так, при фазовой модуляции при любых изменениях низкой частоты индекс модуляции остается тем же, поэтому число спектральных составляющих также не изменяется. Однако ширина спектра этого сигнала будет изменяться в соответствии с пропорциональным изменением частоты Ω . Иными словами, интервалы между частотными составляющими спектра могут либо уменьшаться, либо увеличиваться, что и влияет на трансформацию спектра. При частотной модуляции с уменьшением частоты Ω

управляющего колебания индекс модуляции M увеличивается, что приводит к росту числа спектральных составляющих. В результате ширина спектра практически не изменится из-за сокращения частотного интервала между гармониками спектра.

Широкий спектр сигнала при угловой модуляции обеспечивает помехоустойчивость и качество передачи сообщений, превосходящие аналогичные показатели при амплитудной модуляции. По этой причине сигналы с частотной модуляцией в настоящее время широко применяются для высококачественного радиовещания в метровом диапазоне волн. Фазовая модуляция для передачи аналоговых сообщений, как правило, не используется по причине более сложной реализации, однако она находит применение в системах передачи дискретных сообщений.

4.8. Формирование и детектирование сигналов с угловой модуляцией

В настоящее время все известные методы формирования сигналов с угловой модуляцией принято подразделять на прямые и косвенные. Поскольку, как уже отмечалось, частотная и фазовая модуляции имеют много общего, то, как правило, методы формирования сигналов угловой модуляции рассматривают на примере частотной модуляции.

Сущность прямых методов частотной модуляции заключается в изменении частоты генератора гармонических колебаний посредством непосредственного воздействия на параметры его колебательного контура. Так как частота колебаний генератора, имеющего в своем составе колебательный контур, определяется резонансной частотой этого контура $\omega_p = \omega_c = 1/\sqrt{L_k C_k}$, то эту частоту можно изменять, изменяя емкость C_k или индуктивность L_k в соответствии с законом модулирующего (управляющего) колебания. Более удобным на практике является воздействие на емкость колебательного контура, что обычно реализуется подключением дополнительной емкости параллельно конденсатору.

В современных схемах генераторов, реализуемых на транзисторах, для управления частотой применяются полупроводниковые диоды — *варикапы*, особенностью которых является их способность изменять емкость $p-n$ -перехода под воздействием внешнего управляющего колебания. Если внешнее напряжение между p - и n -областями данного прибора отсутствует, то в нем возникает потенциальный барьер и внутреннее электрическое поле. При подаче на диод напряжения в обратном направлении высота этого барьера увеличивается. Внешнее напряжение будет способство

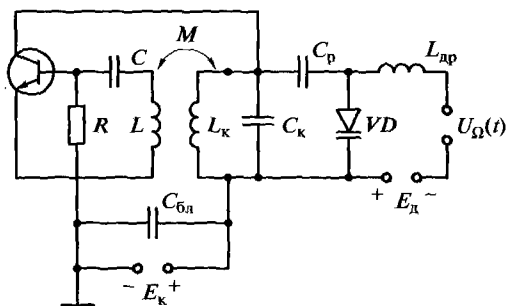


Рис. 4.13. Схема устройства формирования сигналов с угловой модуляцией на основе варикапа:

$C_{\text{бл}}$ — блокировочный конденсатор

вать смещению носителей заряда (электронов) в глубь n -области, поэтому произойдет расширение области $p-n$ -перехода, которая и будет представлять собой простейший конденсатор, где обкладками являются границы этой области. Если внешнее напряжение имеет переменный характер, то границы области будут изменяться, а следовательно, будет изменяться и емкость варикапа.

На рис. 4.13 представлена схема генератора, который вырабатывает гармоническое высокочастотное колебание вида $u(t) = U \cos \omega_c t$. Параллельно контуру генератора подключен варикап VD , на который подается постоянное напряжение смещения $E_{\text{д}}$, обеспечивающее запирающие диода, а также модулирующий сигнал $U_{\Omega}(t)$. В соответствии с изменениями модулирующего колебания изменяются параметры диода, что приводит к пропорциональному изменению частоты генератора. Конденсатор C_p большой емкости служит для разделения цепей питания генератора и варикапа по постоянному току. Дроссель с индуктивностью $L_{\text{др}}$ преграждает путь колебаниям высокой частоты в источник модулирующего сигнала.

Введем следующие обозначения: C_0 — средняя емкость варикапа при отсутствии модулирующего колебания; $C_{\text{к}} = C + C_0$ — средняя емкость колебательного контура; ΔC — изменение емкости варикапа, пропорциональное колебаниям управляющего напряжения $U_{\Omega}(t)$.

Тогда частоту генерируемых колебаний можно записать в виде

$$\begin{aligned} \omega &= \omega_p + \Delta\omega = 1/\sqrt{L_{\text{к}}(C_{\text{к}} + \Delta C)} = 1/\sqrt{L_{\text{к}}C_{\text{к}}(1 + \Delta C/C_{\text{к}})} = \\ &= \omega_p / \sqrt{1 + \Delta C/C_{\text{к}}}, \end{aligned}$$

где $\omega_p = 1/\sqrt{L_k C_k}$ — средняя частота колебаний генератора; $\Delta\omega$ — отклонение частоты от ее среднего значения, т. е. девиация частоты.

Разделив это выражение на ω_k , получим

$$1 + \Delta\omega / \omega_p = 1 / \sqrt{1 + \Delta C / C_k},$$

откуда

$$\frac{\Delta C}{C_k} = - \frac{2\Delta\omega / \omega_p + (\Delta\omega / \omega_p)^2}{(1 + \Delta\omega / \omega_p)^2}.$$

Из этой формулы видно, что относительное изменение емкости колебательного контура связано нелинейной зависимостью с относительным изменением частоты генератора. На практике при передаче речевых сообщений и музыки в диапазоне ультракоротких волн значение $\Delta\omega / \omega_p \ll 1$. Следовательно, последнее выражение упрощается и принимает вид

$$\Delta C / C_k \approx -2\Delta\omega / \omega_p,$$

откуда следует, что изменение частоты колебаний будет происходить пропорционально изменению емкости варикапа ΔC . Знак « \leftarrow » в формуле показывает, что увеличение емкости контура приводит к уменьшению частоты колебаний. Также из формулы видно, что изменения частоты будут происходить пропорционально изменениям модулирующего низкочастотного сигнала, поскольку емкость варикапа напрямую зависит от управляющего напряжения $U_{\Omega}(t)$.

Отметим, что рассмотренное устройство, дополненное несложным элементом — дифференцирующей цепочкой, можно также использовать для формирования сигналов с фазовой модуляцией. Через эту цепочку следует пропускать управляющий сигнал непосредственно перед процессом модуляции. Далее полученное колебание используют для осуществления частотной модуляции. Поскольку частота является производной от фазы колебания, а дифференцирование — линейная операция, то на выходе рассматриваемого устройства получают фазомодулированное колебание, параметры которого пропорциональны исходному низкочастотному колебанию.

Среди косвенных методов формирования сигналов с угловой модуляцией наибольшее распространение получил метод Армстронга, американского радиоинженера, предложенный в 1930-х гг. Устройство, реализующее этот метод, схематично представлено на рис. 4.14. Принцип построения такого устройства следует из выражения для сигнала с угловой модуляцией при значении ин-

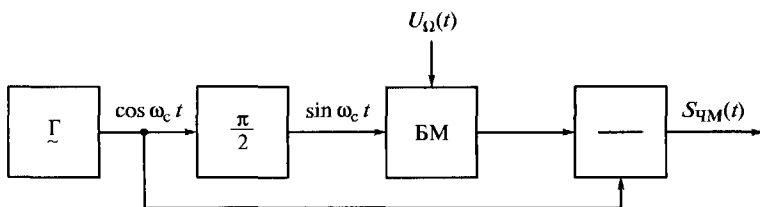


Рис. 4.14. Схема устройства формирования сигналов с угловой модуляцией методом Армстронга

декса модуляции $M \ll 1$, которое, как уже отмечалось, имеет вид

$$S(t) = U \cos \omega_c t - UM \sin \Omega t \sin \omega_c t.$$

В этом выражении второе слагаемое является по существу сигналом балансной модуляции.

На балансный модулятор (БМ) подается высокочастотное колебание от генератора после поворота его фазы на 90° . На второй вход модулятора поступает управляющий сигнал $U_\Omega(t)$. Результирующим на выходе устройства будет сигнал угловой модуляции. Правда, получить здесь можно только колебания с малыми индексами модуляции. Поэтому в некоторых случаях на выходе этого устройства включается умножитель частоты, с помощью которого сигнал с малым значением девиации частоты преобразуется в колебание с большим значением девиации.

Для детектирования сигналов угловой модуляции, в частности с частотной модуляцией, известно несколько методов. Одним из широко распространенных является метод, основанный на процедуре преобразования ЧМ-колебаний в амплитудно-модулированный сигнал и последующем детектировании его с помощью амплитудного детектора. Такая процедура заключается в следующем. На первом этапе частотно-модулированный сигнал пропускается через амплитудный ограничитель в целях устранения нежелательных изменений огибающей, которые появляются под воздействием помех в канале радиосвязи. Если исключить эту процедуру, то качество демодуляции резко ухудшится. На втором этапе ЧМ-сигнал преобразуется в амплитудно-модулированное колебание с помощью схемы с расстроенным колебательным контуром, после чего осуществляется процесс непосредственной демодуляции этого сигнала.

Схема детектора частотно-модулированных сигналов показана на рис. 4.15. В точной схеме колебательный контур, являющийся нагрузкой, настраивается на резонансную частоту $\omega_p \neq \omega_c$, которая отличается от частоты несущего колебания. Через контур про-

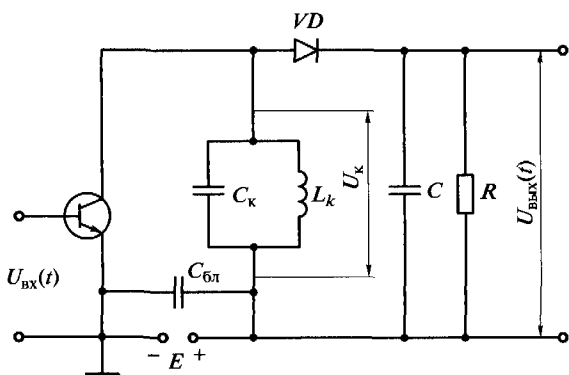


Рис. 4.15. Схема детектора частотно-модулированного сигнала на основе расстроенного колебательного контура

текает ток (проходит ЧМ-колебание) с постоянной амплитудой, но изменяющейся частотой $\omega(t) = \omega_c + \Delta\omega(t)$, среднее значение которой отличается от частоты настройки контура. Вследствие этого амплитуда выходного напряжения колебательного контура будет повторять все изменения частоты входного сигнала. Иными словами, закон изменения амплитуды высокочастотного колебания $U_k(t)$ будет соответствовать закону изменения частоты $\Delta\omega(t)$, что иллюстрирует рис. 4.16. Далее изменения этого колебания с помощью амплитудного детектора преобразуются в низкочастотный сигнал, который выделяется в нагрузку, содержащей конденсатор C и резистор R (см. рис. 4.15).

Рассмотренная схема обладает двумя особенностями. Во-первых, здесь необходима настройка колебательного контура на

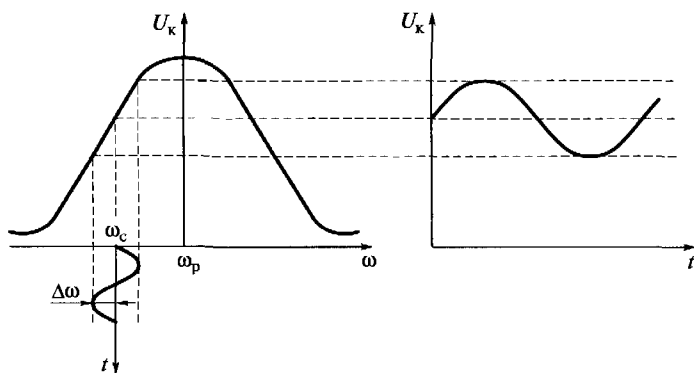


Рис. 4.16. Пояснение процесса преобразования частотно-модулированного колебания в амплитудно-модулированный сигнал

частоту, отличающуюся от частоты несущего колебания ω_c . Во-вторых, линейный участок резонансной кривой является ограниченным, поэтому при демодуляции в схеме могут возникать нелинейные искажения. Для компенсации данных искажений обычно применяют второй колебательный контур и еще один амплитудный детектор. В этом случае выходное напряжение определяется как разность напряжений, получаемых на выходах каждого из детекторов. Схема с двумя детекторами получила название дискриминатора с расстроенными контурами. При этом один из контуров настраивается на частоту, большую, чем несущая частота сигнала ω_c , а другой — на частоту меньшую ω_c .

В заключение отметим, что подобные частотные демодуляторы могут также применяться для детектирования фазомодулированных колебаний. Для этого выходное напряжение детектора в отличие от процедуры модуляции следует пропустить через интегрирующее устройство.

Контрольные вопросы

1. Перечислить методы модуляции при гармоническом переносчике сообщений.
2. Пояснить сущность амплитудной модуляции и вид спектра амплитудно-модулированного колебания при модуляции одним гармоническим колебанием.
3. Определить мощность амплитудно-модулированного сигнала за период колебаний высокой и низкой частот.
4. В чем заключается смысл балансной и однополосной гармонических модуляций?
5. В чем состоит отличие квадратичного детектирования амплитудно-модулированного сигнала от режима сильного сигнала?
6. Пояснить, почему угловая модуляция реализуется как фазовая и частотная и как определить индекс модуляции и девиацию частоты в каждом из этих случаев.
7. Объяснить сущность узкополосной угловой модуляции и показать спектр сигнала для этого случая.
8. Пояснить сущность широкополосной угловой модуляции и перечислить, чем определяется ширина спектра сигналов в этом случае.
9. Как формируются и детектируются сигналы с угловой модуляцией?

Глава 5

ОСНОВНЫЕ ВИДЫ ДИСКРЕТНОЙ МОДУЛЯЦИИ СИГНАЛОВ

5.1. Дискретная амплитудная модуляция

Аналоговые методы модуляции, рассмотренные ранее, широко применяются в технике телекоммуникаций, но наряду с ними существуют и другие способы формирования сигналов для применения в современных системах передачи информации. Особо перспективными в настоящее время являются дискретные сигналы, представляющие собой дальнейшее развитие классических видов сигналов, модулированных по амплитуде, частоте или фазе.

Дискретной называется модуляция, при которой сигнал на выходе модулятора является дискретным. В случае когда и на входе модулятора сигнал имеет дискретную форму, говорят о *цифровой модуляции*, или *манипуляции*.

Все дискретные виды модуляции реализуются таким образом, что число значений модулирующего (управляющего) сигнала является конечным, т. е. $m = 1, 2, \dots, M$, в отличие от аналоговых видов модуляции, в которых число градаций модулирующего сигнала бесконечно. В частном случае, когда $m = 2$, управляющим является двоичный сигнал, реализации которого принимают значения 1 и 0, соответствующие появлению токовой и бестоковой посылки. Амплитуда токовой посылки при этом остается неизменной в течение ее длительности.

Остановимся несколько подробнее на дискретной амплитудной модуляции (ДАМ), которая в зарубежных источниках именуется Amplitude Shift Keying — ASK. В этом случае сигнал можно представить математической моделью

$$S_m(t) = \operatorname{Re}[A_m b(t) e^{i(\omega_c t + \varphi)}] = A_m b(t) \cos(\omega_0 t + \varphi); \quad (5.1)$$
$$0 \leq t \leq T; m = \overline{1, M},$$

где $A_m = \dot{A}_m$ — комплексная амплитуда сигнала, которая может принимать действительные положительные или отрицательные

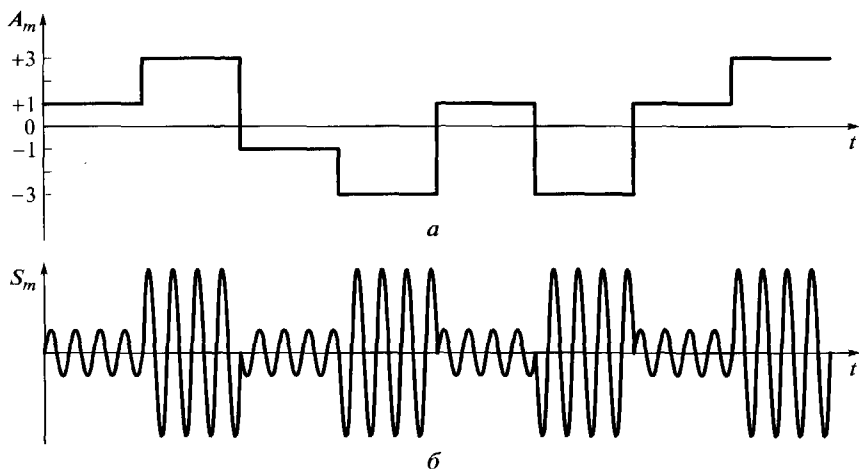


Рис. 5.1. Временные диаграммы при дискретной амплитудной модуляции:
a — управляющий сигнал; *б* — модулированный сигнал

значения, число которых $M = 2^k$, $k = 1, 2, \dots$; $b(t)$ — управляющий вещественный сигнал, обычно имеющий форму прямоугольного видеоимпульса; $e^{i\omega_0 t}$ — комплексная несущая; φ — начальная фаза сигнала.

Дискретные значения амплитуды можно определить с помощью выражения

$$A_m = (2m - 1 - M)d,$$

причем здесь $2d$ — расстояние между двумя соседними амплитудами сигнала.

На рис. 5.1 представлены управляющий и модулированный сигналы для случая, когда $M = 4$. Каждой из $m = 1, \dots, M$ амплитуд можно приписать определенную комбинацию, содержащую k двоичных символов. Следовательно, с помощью дискретной амплитудной модуляции можно передать сразу несколько символов, объединенных в некоторый блок (кодовое слово), что позволяет увеличить скорость передачи информации в k раз.

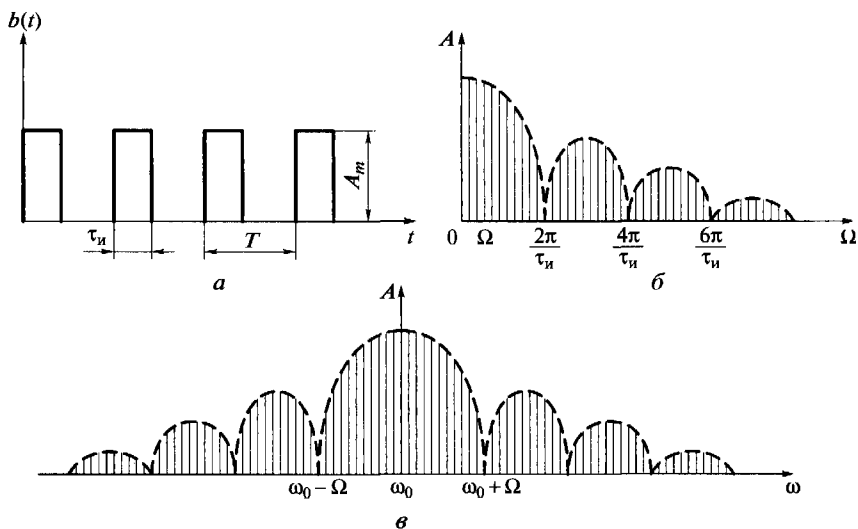
Определим спектр ДАМ для частного случая: $M = 2$, $A_0 = 0$, $A_1 = A$. Модулирующий сигнал $b(t)$, представляющий собой последовательность прямоугольных импульсов с единичной амплитудой, длительностью t_n и периодом следования T , можно представить в виде ряда Фурье (2.15). Подставив выражение (2.15), где $E = 1$, в математическую модель (5.1) и обозначив скважность импульсной последовательности $q = T/t_n$, для сигнала двоичной дискретной амплитудной модуляции получим выражение

$$S(t) = \frac{A}{q} \cos(\omega_0 t + \varphi) + \frac{A}{q} \sum_{k=1}^{\infty} \frac{\sin \frac{k\pi}{q}}{k\pi} \{ \cos[(\omega_0 - k\Omega)t + \varphi] + \cos[(\omega_0 + k\Omega)t + \varphi] \}. \quad (5.2)$$

На рис. 5.2 показаны управляющая последовательность прямоугольных импульсов, спектр последовательности видеоимпульсов и спектр сигнала вида (5.2) при коэффициенте модуляции $m = 1$ и скважности $q > 1$. По существу эти спектры аналогичны, но результирующий спектр смещен относительно исходного значения ω_0 .

Анализ сигналов дискретной амплитудной модуляции показывает:

- в спектре такого сигнала так же, как и в спектре сигнала с аналоговой амплитудной модуляцией, содержится колебание на несущей частоте и гармонические колебания в двух боковых полосах, иными словами, спектр является симметричным относительно несущего колебания с частотой ω_0 ;
- огибающая спектра такого сигнала определяется формой модулирующего дискретного сигнала, в частности при прямоугольных видеоимпульсах она имеет форму графика функции $\sin x/x$;



5.2. Пояснение процесса дискретной амплитудной модуляции:

a — управляющая последовательность прямоугольных импульсов; b — спектр последовательности видеоимпульсов; $в$ — спектр модулированного сигнала

- при скважности такого сигнала $q = 2$ основная доля энергии сосредоточена в полосе частот $\Delta f \approx (6 \dots 10)F_m$, где $F_m = \Omega / (2\pi)$ — частота модуляции. При этом 90 % энергии сосредоточено в полосе частот до $6F_m$ и более 97 % — в полосе до $10F_m$.

Как будет показано далее, ДАМ-сигналы обладают самой низкой помехоустойчивостью среди всех сигналов с дискретными видами модуляции, и в этом заключается их недостаток. Кроме того, в спектре этих сигналов содержатся две боковые полосы, поэтому для их передачи необходима полоса частот, которая должна быть вдвое больше, чем полоса для передачи низкочастотного сигнала. Следовательно, по аналогии с однополосной аналоговой модуляцией можно использовать однополосную дискретную амплитудную модуляцию. Для такого сигнала в качестве математической модели используется следующее выражение:

$$S(t) = \text{Re}\{A_m[b(t) \pm \hat{i}b(t)]e^{i\omega t}\}; \quad m = \overline{1, M},$$

где $\hat{b}(t)$ — преобразование Гильберта от сигнала $b(t)$, который, как и в предыдущем случае, является модулирующей последовательностью прямоугольных видеоимпульсов.

Знак «+» здесь соответствует формированию сигнала на верхней боковой полосе, а знак «-» отождествляется с нижней боковой полосой.

Данное выражение по существу идентично выражению для однополосной аналоговой модуляции.

В заключение можно отметить, что формирование и демодуляция сигналов с дискретной амплитудной модуляцией осуществляются с помощью методов и схем, рассмотренных ранее применительно к аналоговой амплитудной модуляции.

5.2. Дискретная фазовая модуляция

Дискретная фазовая модуляция — ДФМ (Phase Shift Keying — PSK) в настоящее время является одним из наиболее широко применяемых видов модуляции сигналов. Математическая модель сигнала в этом случае имеет вид

$$S_m(t) = \text{Re}\left[U_c e^{i\{\omega t + 2\pi(m-1)/M\}}\right] = U_c \cos\left(\omega t + \frac{2\pi(m-1)}{M}\right);$$

$$t \in [0, T],$$

где U_c — огибающая, имеющая прямоугольную форму; $m = \overline{1, M}$; M — число возможных вариантов фазы сигнала; $k = 1, 2, \dots$

Заметим, что при ДФМ каждое значение из $M = 2^k$ фаз по аналогии с ДАМ принято отождествлять с некоторой комбинацией,

которая содержит k двоичных элементов. Таким образом, здесь можно сразу передать не только одну посылку, но и несколько символов, объединенных в блок (кодовое слово). Это также позволяет увеличить скорость передачи информации в k раз.

В частном случае, когда $k = 1$, $M = 2$, посылки ДФМ-сигнала можно записать в следующем виде:

$$S_0(t) = U_c \cos(\omega t + \varphi) \text{ — соответствует символу «0»};$$

$$S_1(t) = U_c \cos(\omega t + \pi + \varphi) \text{ — соответствует символу «1»},$$

где φ — начальная фаза высокочастотного колебания.

Легко заметить, что здесь $S_1(t) = -S_0(t)$. Нередко данный сигнал называют фазоманипулированным (ФМ). В иностранной литературе термину *фазовая манипуляция* соответствует термин *binary PSK (BPSK)*. Временная диаграмма этого сигнала представлена на рис. 1.4, *г*. Здесь при каждой очередной перемене полярности управляющей последовательности прямоугольных импульсов происходит скачкообразная смена информационной фазы, которая принимает значения либо 0, либо 180° .

В функциональном пространстве ФМ-сигнал, реализации которого имеют равные значения энергии, могут быть представлены векторами одинаковой длины. При этом векторы, выходящие из общей начальной точки, изображаются отрезками, расположенными на одной прямой линии, но направленными в разные стороны, что соответствует сдвигу между ними на угол 180° или π .

В дальнейшем будет показано, что именно такие сигналы среди всех известных сигналов с дискретными видами модуляции обладают наибольшей помехоустойчивостью, так как в векторном пространстве расстояние между концами векторов, соответствующих ФМ-сигналам, оказывается наибольшим. Следовательно, в этом случае сложнее при воздействии помех перепутать векторы сигналов.

Фазоманипулированные сигналы можно сформировать с помощью схемы, представленной на рис. 5.3, включающей в себя генератор высокочастотного гармонического колебания, соединенный с одним ключом непосредственно, а с другим — через фазовращатель на 180° . Первый из ключей открывается управляющими посылками «0», а второй — посылками «1». В результате на выходе схемы образуются радиоимпульсы, фазы которых скачкообразно изменяются в моменты смены полярностей в управляющей (модулирующей) последовательности.

Рассмотренный ФМ-сигнал, где $M = 2$, $k = \log_2 M = 1$, называют также однократным. Число вариантов фаз, как отмечалось ранее, здесь равно двум, при этом каждое значение фазы соответствует передаче одного из двоичных символов: «0» или «1». Су-

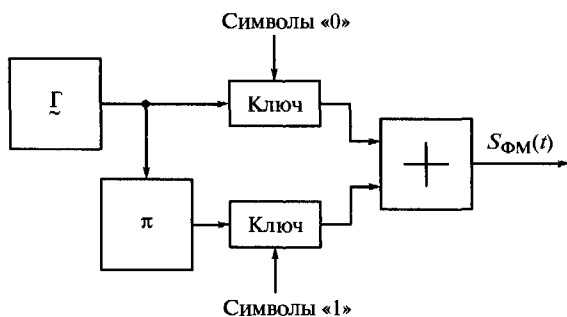


Рис. 5.3. Схема формирования сигнала с дискретной фазовой модуляцией

существуют также многократные (многопозиционные) ФМ-сигналы. Так, при $k = 2$, $M = 4$ дискретная фазовая модуляция называется двухкратной, а при $k = 3$, $M = 8$ — трехкратной.

При многопозиционной ДФМ, как уже говорилось, передаваемая двоичная последовательность символов сообщения разделяется на отдельные блоки, каждый из которых содержит k элементов. Далее всякому блоку символов ставится в соответствие реализация сигнала со своим значением фазы. Чаще всего для этой цели используется двоичный код Грея.

Код Грея относится к классу безызбыточных отраженных, или рефлексных (от *лат. reflexio* — отражение), кодов. Такое название эти коды получили из-за симметричности комбинаций в кодовой таблице относительно некоторой воображаемой линии (оси симметрии или отражения), которая как бы делит эту таблицу с кодовыми блоками на две равные части. Симметрия проявляется в том, что в каждой из этих половин имеются равноотстоящие от середины кодовые комбинации, в которых полностью совпадает некоторое количество кодовых символов. Сказанное иллюстрирует табл. 5.1, в которой представлены трехэлементные комбинации простого безызбыточного кода и кода Грея.

Таблица 5.1

Десятичное число	Простой код	Код Грея
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110

Десятичное число	Простой код	Код Грея
5	101	111
6	110	101
7	111	100

Отметим следующие особенности кода Грея.

1. Любые две кодовые комбинации, расположенные рядом в таблице, отличаются друг от друга только одним символом (разрядом). Следовательно, при сложении по модулю два 2 двух соседних комбинаций получается единица. Это свойство можно использовать для проверки верности принятых кодовых блоков.

2. Младшие ($n - 1$) разряды кодовых комбинаций, расположенных симметрично относительно оси отражения, являются идентичными. Отличие проявляется только в старшем n -м разряде: по одну сторону оси здесь находится символ «1», а по другую — символ «0». Данная ось симметрии, называемая главной осью, в n -элементном коде Грея проходит между комбинациями, которые в двоичном представлении соответствуют числам $2^{n-1} - 1$ и 2^{n-1} . В частности, в рассматриваемой таблице главная ось находится между 3-й и 4-й комбинациями. Таким образом получают следующие пары симметричных кодовых комбинаций: 0—7; 1—6; 2—5; 3—4. В этих комбинациях совпадают 2 и 3-й символы, а 1-е различаются. Помимо главной оси в коде Грея имеются еще несколько осей симметрии. Так, в рассматриваемой таблице есть еще две оси: между 1-м и 2-м, а также между 5-м и 6-м символами. В этом случае совпадают младшие и старшие разряды, а в средних — имеется различие.

Формирование комбинаций кода Грея на основе простого двоичного кода производится следующим образом:

1) над двоичным числом (комбинацией простого кода) записывается это же число, но со сдвигом вправо на один разряд. При этом самый младший разряд отбрасывается;

2) исходное и сдвинутое числа поразрядно складываются по модулю два. Полученный при этом результат соответствует блоку кода Грея.

Для обратного перевода кода Грея в простой двоичный код необходимо выполнить следующее:

1) цифру самого старшего разряда оставить без изменений;

2) каждую последующую цифру инвертировать столько раз, сколько единиц ей предшествует в комбинации кода Грея.

Многопозиционные сигналы ДФМ удобно представлять в виде векторов или точек в векторном линейном пространстве, поскольку

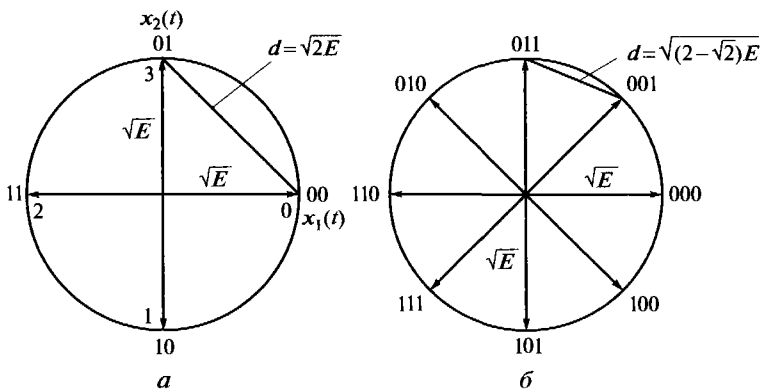


Рис. 5.4. Представление четырехпозиционного (а) и восьмипозиционного (б) сигналов с дискретной фазовой модуляцией векторами в функциональном пространстве: $x_1(t), x_2(t)$ — координатный базис; 0... 3 — номера сигнальных точек; 00... 11; 000... 1111 — комбинации в коде Грея

ку при $M > 2$ сигнал имеет достаточно сложный вид, и изображать его графически как функцию времени оказывается чрезвычайно неудобным. На рис. 5.4 показаны векторы, соответствующие четырехпозиционному (двухкратному) и восьмипозиционному (трехкратному) ДФМ-сигналам.

В общем случае сдвиг фаз между соседними сигналами, определяемый выражением $\Delta\varphi = 2\pi/M$ (где $M = 2^k, k = 1, 2, \dots$), обуславливает ориентацию векторов относительно друг друга. Показанные построения выполняются таким образом, чтобы все сигнальные точки, т.е. концы векторов, располагались на окружности, радиус R которой (длина вектора-сигнала) зависит от энергии E посылки сигнала, т.е. $R = \sqrt{E}$.

Расстояние между сигнальными точками представляет собой евклидово расстояние. Наибольшее значение евклидовой метрики имеет однократная двоичная фазовая манипуляция. В этом случае $d_2 = 2\sqrt{E}$, при двухкратной ДФМ евклидово расстояние $d_4 = \sqrt{2E} \approx 1,41\sqrt{E}$, а при трехкратной — $d_8 = \sqrt{(2 - \sqrt{2})E} \approx 0,765\sqrt{E}$. Уменьшение евклидова расстояния приводит к снижению помехоустойчивости, что обусловлено ухудшением процедуры различения векторов-сигналов при воздействии помех, поскольку суммарный вектор сигнала и аддитивной помехи может изменить свою ориентацию относительно начального положения. Как уже отмечалось, наилучшей различимостью обладают двоичные ФМ-сигналы, помехоустойчивость которых называется потенциальной.

Определим теперь спектр дискретного фазомодулированного сигнала. И сделаем это для частного случая — двоичного фазо-

манипулированного (ФМ) сигнала, поскольку проделать подобную процедуру применительно к многопозиционной ДФМ оказывается сложным.

Двоичный ФМ-сигнал можно записать в следующем виде:

$$S_{\text{ФМ}}(t) = U_m S_{\Omega}(t) \cos(\omega t + \varphi). \quad (5.3)$$

В этом выражении $S_{\Omega}(t)$ — модулирующая последовательность, представляющая собой периодическую последовательность двухполярных (± 1) прямоугольных импульсов с длительностью t_n и периодом T . Обозначим $\omega = \omega_0$ частоту несущего колебания.

Используем разложение в ряд Фурье периодической последовательности двухполярных ($\pm E$) прямоугольных импульсов в виде (2.16), которое было получено ранее. Задав в этом выражение $E = 1$, подставим его в формулу для $S_{\text{ФМ}}(t)$. Тогда для двоичного ФМ-сигнала спектральное представление

$$\begin{aligned} S_{\text{ФМ}}(t) = & U_m \left(2 \frac{t_n}{T} - 1 \right) \cos(\omega_0 t + \varphi) + \\ & + 2 \frac{t_n}{T} U_m \sum_{k=1}^{\infty} \frac{\sin \frac{k\Omega t_n}{2}}{\frac{k\Omega t_n}{2}} \{ \cos[(\omega_0 + k\Omega)t + \varphi] + \\ & + \cos[(\omega_0 - k\Omega)t + \varphi] \}. \end{aligned} \quad (5.4)$$

Наиболее интересен случай синхронной передачи, когда в качестве модулирующего применяется двухполярный сигнал (см. рис. 1.4, *a*), при котором все посылки имеют одинаковую длительность (при этом $\frac{t_n}{T} = \frac{1}{2}$).

Для этого случая из предыдущего выражения получим

$$\begin{aligned} S_{\text{ФМ}}(t) = & \\ = & U_m \sum_{k=1}^{\infty} \frac{\sin \frac{k\Omega t_n}{2}}{\frac{k\Omega t_n}{2}} \{ \cos[(\omega_0 + k\Omega)t + \varphi] + \cos[(\omega_0 - k\Omega)t + \varphi] \}. \end{aligned} \quad (5.5)$$

Из формулы (5.5) видно, что при двоичной фазовой манипуляции, когда информационная фаза в процессе передачи сообщений принимает значения 0 или 180° , в спектре сигнала отсутствует колебание на несущей частоте. Этот спектр становится похожим на спектр сигналов с балансной амплитудной модуляцией, где также нет несущего колебания, что приводит к серьезным трудностям в практической реализации систем передачи информации с ФМ-сигналами.

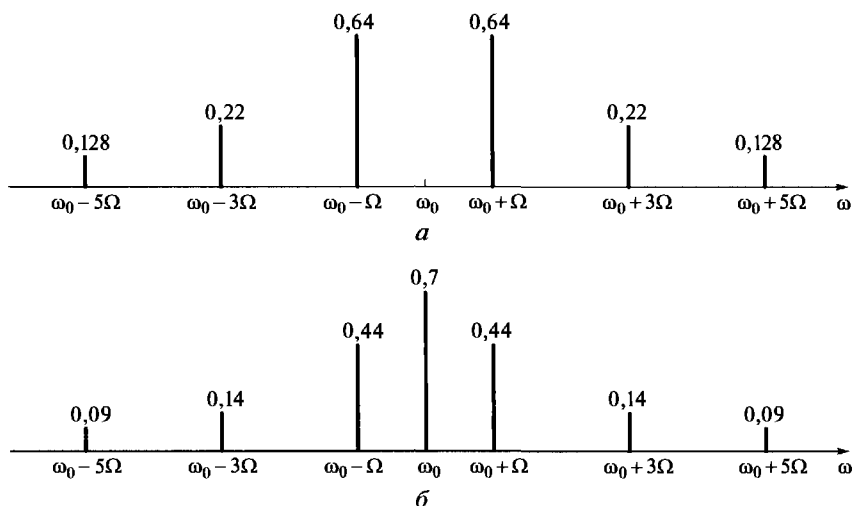


Рис. 5.5. Зависимости спектров сигналов с дискретной фазовой модуляцией от манипуляционных сдвигов фаз 180° (а) и 90° (б)

В других случаях, когда манипуляционная фаза принимает иные значения, например $\pi/2$ или $\pi/4$, спектр ДФМ-сигнала так же, как спектр ДАМ-сигнала, будет содержать несущее колебание и боковые составляющие, которые отстоят друг от друга на частотные интервалы, кратные частоте манипуляции. Сказанное иллюстрирует рис. 5.5.

Для приема ДФМ-сигналов может применяться нелинейный преобразователь, который реализуется, например на основе балансной схемы перемножения сигналов, приведенной на рис. 3.17. В общем случае процедура демодуляции ФМ-сигналов сводится к двум операциям:

- перемножению входного колебания, являющегося смесью сигнала и помех, с опорным сигналом, который вырабатывается генератором (гетеродином) в приемнике;
- выделению необходимой составляющей с помощью фильтра.

Для удобства дальнейших преобразований ФМ-сигнал можно представить в следующем виде:

$$S(t) = U_c \cos\left(\omega_c t + \varphi_c \pm \frac{\pi}{2}\right),$$

и пусть при этом гетеродином вырабатывается колебание вида

$$S_r(t) = U_r \cos(\omega_r t + \varphi_r).$$

Тогда в результате перемножения получим выражение

$$\begin{aligned} S(t) \times S_c(t) &= U_c \cos\left(\omega_c t + \varphi_c \pm \frac{\pi}{2}\right) U_r \cos(\omega_r t + \varphi_r) = \\ &= \frac{U_c U_r}{2} \cos\left[(\omega_c + \omega_r)t + (\varphi_c + \varphi_r) \pm \frac{\pi}{2}\right] + \\ &+ \frac{U_c U_r}{2} \cos\left[(\omega_c - \omega_r)t + (\varphi_c - \varphi_r) \pm \frac{\pi}{2}\right]. \end{aligned}$$

В случае равенства частот сигнала и гетеродина ($\omega_c = \omega_r$) получим

$$S(t) \times S_r(t) = \frac{U_c U_r}{2} \cos\left[2\omega_c t + (\varphi_c + \varphi_r) \pm \frac{\pi}{2}\right] + \frac{U_c U_r}{2} \cos\left[(\varphi_c - \varphi_r) \pm \frac{\pi}{2}\right].$$

Колебание удвоенной частоты после перемножения отфильтровывается. Поэтому на выходе демодулятора появляется сигнал, амплитуда которого зависит от разности фаз принимаемого сигнала и гетеродина. (В целях некоторого упрощения влияние помех здесь не учитывается.)

Пусть передается посылка, соответствующая символу «0», фаза сигнала в этом случае равна $\varphi_c + \pi/2$. Пусть также обеспечивается синфазность колебаний сигнала и гетеродина. Тогда фаза гетеродина $\varphi_r = \varphi_c + \pi/2$. В этом случае разность фаз обоих колебаний равна нулю, и на выходе демодулятора получим

$$S(t) \times S_r(t) = \frac{U_c U_r}{2} \cos 0 = \frac{U_c U_r}{2} = A.$$

Пусть теперь на входе демодулятора появляется колебание с фазой, равной $\varphi_c - \pi/2$, что соответствует передаче символа «1», но при этом сохраняется синфазность двух колебаний и неизменным остается прежнее значение фазы гетеродина $\varphi_r = \varphi_c + \pi/2$. Тогда разность фаз $\varphi_c - \varphi_r$ окажется равной 180° , и выходной сигнал изменит свою полярность на противоположную:

$$S(t) \times S_r(t) = \frac{U_c U_r}{2} \cos 180^\circ = -\frac{U_c U_r}{2} = -A.$$

Таким образом, в результате данной процедуры передававшиеся посылки сигнала на приеме будут восстанавливаться. Однако основная сложность практической реализации сигналов дискретной фазовой модуляции, и в частности фазоманипулированных сигналов, заключается в формировании опорного колебания гетеродина. Это колебание, в том числе по частоте и начальной фазе, должно совпадать с аналогичными параметрами принима-

емого сигнала, который в процессе передачи по каналу связи подвергается воздействию случайных помех. Поскольку, как было показано ранее, в спектре ФМ-сигнала отсутствует колебание на несущей частоте, подчас возникают непреодолимые сложности в создании опорного колебания гетеродина, удовлетворяющего указанным требованиям, что в ряде случаев является препятствием для практического внедрения этих сигналов.

5.3. Квадратурная амплитудная модуляция

Принцип квадратурной амплитудной модуляции (КАМ) непосредственно вытекает из квадратурного представления сигнала, смысл которого заключается в том, что любое гармоническое колебание с произвольной фазой можно записать в виде линейной комбинации двух колебаний: по законам функций синуса и косинуса. Это следует из следующих тригонометрических равенств:

$$\sin(\omega t \pm \varphi) = \cos \varphi \sin \omega t \pm \sin \varphi \cos \omega t;$$

$$\cos(\omega t \pm \varphi) = \cos \varphi \cos \omega t \mp \sin \varphi \sin \omega t.$$

Здесь $\cos \varphi$ и $\sin \varphi$ — коэффициенты разложения, а $\cos \omega t$ и $\sin \omega t$ — базисные функции, которые имеют по отношению друг к другу сдвиг $\pi/2$ или 90° , т. е. они находятся в квадратуре, откуда происходит и название данного метода модуляции. Обычно колебание $\sin \omega t$ называют синфазной составляющей (*I*-сигналами), а колебание $\cos \omega t$ — квадратурной составляющей (*Q*-сигналами).

В иностранной литературе квадратурная амплитудная модуляция — Quadrature Amplitude Shift Keying (QASK).

Сущность КАМ заключается в том, что в каждом из квадратурных каналов производится дискретная амплитудная модуляция несущих колебаний $\cos \omega t$ и $\sin \omega t$ с помощью двух независимых управляющих последовательностей сигналов. При этом некоторому уровню амплитуды в одном канале может соответствовать другой уровень амплитуды во втором канале. Результирующий сигнал представляет собой сумму этих колебаний. Таким образом, два независимых сообщения одновременно будут передаваться в одной общей среде.

В частном случае в качестве управляющих сигналов возможно использование последовательности прямоугольных видеоимпульсов с амплитудой ± 1 . Тогда квадратурная амплитудная модуляция фактически будет совпадать с двухкратной (четырёхпозиционной) дискретной фазовой модуляцией, рассмотренной ранее, при которой сигнальные векторы в пространстве сдвинуты относительно друг друга на $\pi/2$ или 90° (см. рис. 5.4, а).

В общем случае КАМ-сигнал можно представить следующий математической моделью:

$$S_m(t) = \text{Re}[(A_{mc} + jA_{ms})b(t)e^{i(\omega t + \varphi)}] = \\ = A_{mc}b(t)\cos\varphi\cos\omega t - A_{ms}b(t)\sin\varphi\sin\omega t,$$

где A_{mc} , A_{ms} — информационные амплитуды двух квадратурных несущих колебаний; $b(t)$ — отображение формы импульса модулирующего сигнала; $m = \overline{1, M}$; $M = 2^k$, $k = 1, 2, \dots$ — число возможных вариантов амплитуд сигнала.

При $M = 2$, т.е. когда на любой из двух квадратурных несущих используется двухуровневая ДАМ, каждый из двух потоков сообщений может находиться в одном из двух состояний ± 1 , а суммарный поток — в одном из $2 \times 2 = 4$ состояний, что соответствует четырехуровневой КАМ (КАМ-4).

При $M = 4$ реализуется четыре уровня дискретной амплитудной модуляции, и A_{mc} , A_{ms} могут принимать значения ± 1 , ± 3 , а объединенный поток символов будет находиться в одном из $4 \times 4 = 16$ состояний, что соответствует КАМ-16. При этом чем больше число состояний, тем выше оказывается скорость передачи информации, поскольку на каждое состояние приходится большее число передаваемых информационных символов.

Схемы формирователя (модулятора) и демодулятора КАМ-сигналов соответственно представлены на рис. 5.6. Процесс демоду-

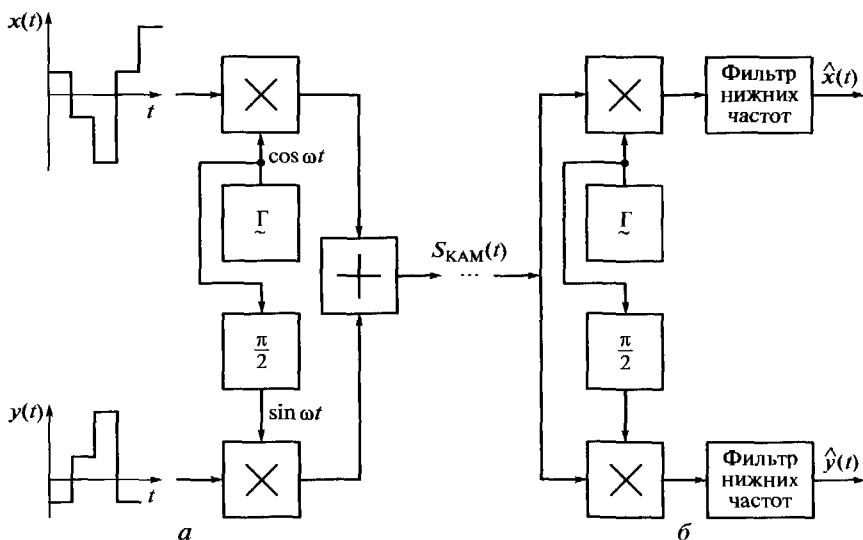


Рис. 5.6. Схемы формирователя (а) и демодулятора (б) сигнала квадратурной амплитудной модуляции

ляции этих сигналов по существу совпадает с процессом получения низкочастотных квадратурных составляющих, который рассматривался в подразд. 2.10.

В общем случае из-за большой сложности КАМ-сигналы практически невозможно изобразить графически временными зависимостями. По аналогии с ДФМ-сигналами их принято представлять точками в пространстве, только в прямоугольной системе координат. Совокупность всех точек (сигналов) называется сигнальным созвездием.

На рис. 5.7 показаны представления сигналов КАМ-4 и КАМ-16. В первом случае созвездие содержит четыре сигнальные точки, каждая из которых соответствует передаче двух двоичных символов. Во втором созвездии содержится 16 точек, на каждую из которых приходится по четыре двоичных символа. При построении сигнальных созвездий КАМ принято использовать код Грея. Поэтому при воздействии помех переход из любой сигнальной точки в соседнюю область может приводить к ошибочному приему только одного двоичного символа.

По сравнению с ДФМ, где сигнальные точки располагаются на окружности, квадратурная амплитудная модуляция позволяет более эффективно использовать пространство сигналов. Это приводит к увеличению расстояния между соседними точками, улучшает различимость сигналов и повышает помехоустойчивость передачи дискретных сообщений.

Как уже говорилось, КАМ реализуется посредством суммирования двух ортогональных (квадратурных) колебаний. Следовательно, в одной полосе частот одновременно размещаются два идентичных, но независимых друг от друга ДАМ-сигнала. Поэто-

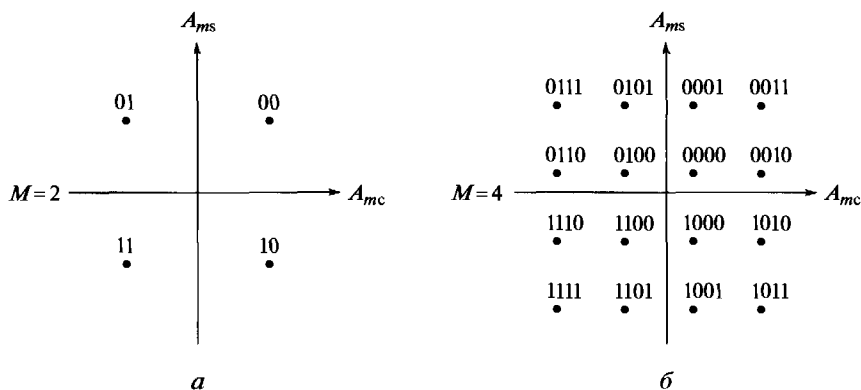


Рис. 5.7. Представления сигналов КАМ-4 (а) и КАМ-16 (б) двухмерного пространства

му ширина спектра КАМ равняется ширине спектра одного сигнала с дискретной амплитудной модуляцией.

5.4. Амплитудно-фазовая модуляция

Возрастание потоков информации, передаваемой в различных каналах телекоммуникаций, способствовало поиску и реализации новых эффективных методов формирования сигналов. В частности, к этим методам относится многопозиционная дискретная фазовая модуляция, а также квадратурная амплитудная модуляция, повышающая рациональность использования отводимой полосы частот. Одним из направлений дальнейшего повышения эффективности передачи информации является объединение двух указанных методов модуляции, т.е. использование амплитудно-фазовой модуляции (АФМ).

Амплитудно-фазовой модуляцией (Amplitude Phase Keying — АРК) называется метод гармонической модуляции, в результате которой сигнал формируется на основе комбинирования амплитуд и фазовых сдвигов несущего колебания. АФМ-сигналы, как и сигналы, рассмотренные ранее, задаются с помощью базисных функций в виде отрезков гармонических колебаний. В частности, базисными могут быть функции вида

$$\psi_1(t) = \sqrt{\frac{2E}{T}} \cos \omega t; \quad \psi_2(t) = \sqrt{\frac{2E}{T}} \sin \omega t; \quad t \in [0, T],$$

где $E = \|S\|^2 = \int_0^T S^2(t) dt$ — энергия сигнала; T — длительность посылки сигнала.

Один из простейших АФМ-сигналов, имеющий четыре реализации, в этом базисе представляется следующими выражениями:

$$S_1(t) = \sqrt{\frac{2E}{T}} \cos \omega t;$$

$$S_2(t) = \sqrt{\frac{2E}{T}} \cos(\omega t + 2\pi/3);$$

$$S_3(t) = \sqrt{\frac{2E}{T}} \cos(\omega t + 4\pi/3);$$

$$S_4(t) = 0.$$

На рис. 5.8, а АФМ-сигнал представлен точками 1...4, три из которых равномерно распределены на окружности с радиусом

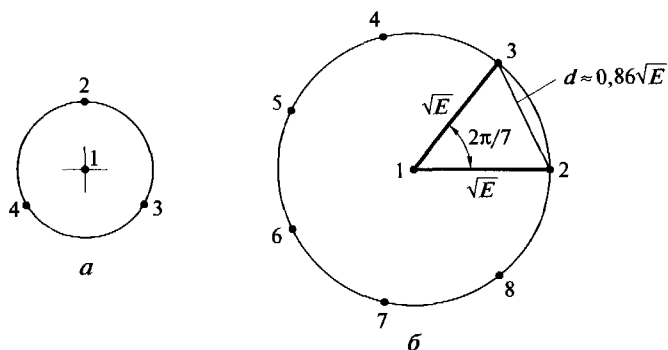


Рис. 5.8. Представление сигналов амплитудно-фазовой модуляции точками функционального двухмерного пространства:
a — АФМ-4; *б* — АФМ-8

$R = \sqrt{E}$, а четвертая расположена в центре этой окружности (в начале координат). Данный сигнал обозначается как (1, 3).

Восьмипозиционный АФМ-сигнал, где в сигнальном созвездии одна из точек 1...8 также расположена в центре окружности, а семь — на окружности с радиусом $R = \sqrt{E}$, показан на рис. 5.8, б. Расстояние между точками на окружности можно найти с помощью теоремы синусов для треугольника:

$$d = \frac{\sqrt{E} \sin(2\pi/7)}{\cos(\pi/7)} \approx 0,86\sqrt{E}.$$

Данная конструкция оказывается эффективнее конструкции восьмипозиционного ДФМ-сигнала, у которого, как было показано ранее, расстояние между сигнальными точками $d_8 = 0,765\sqrt{E}$.

Вместе с тем АФМ-сигналы имеют существенный недостаток, который заключается в том, что различные их реализации имеют неодинаковые значения энергии, это несколько усложняет практическую реализацию демодуляторов АФМ-сигналов. Однако, несмотря на отмеченный недостаток, такие сигналы применяются в системах цифровой связи, в частности в спутниковой, поскольку с увеличением числа позиций сигналы с дискретной фазовой модуляцией становятся менее эффективными по сравнению с АФМ-сигналами.

Сигнал с амплитудно-фазовой модуляцией получают в результате суммирования нескольких образованных независимо друг от друга ФМ-сигналов с соответственно подобранными амплитудами и фазами. Для этого колебание несущей частоты подают на два одинаковых фазовых модулятора, каждый из которых управ-

ляется различными последовательностями прямоугольных видеоимпульсов, и выходные колебания этих модуляторов складываются. При этом уровень одного из них ослабляется в два раза с помощью аттенюатора, что и соответствует процедуре амплитудной модуляции. В результате сигнальное созвездие будет представлять собой результат суммирования сигнальных векторов каждого из каналов.

Прием АФМ-сигналов осуществляется с помощью демодуляторов, реализуемых на основе комбинирования амплитудных и фазовых детекторов, принцип работы которых был рассмотрен ранее.

5.5. Дискретная частотная модуляция

Сигнал дискретной частотной модуляции — ДЧМ Frequency Shift Keying — FSK) в общем случае представляет собой последовательность радиоимпульсов (посылок), которые передаются на различных частотах f_1, f_2, \dots, f_M .

ДЧМ-сигнал формируется посредством сдвига частоты несущего колебания f_0 на значение $f_n = \frac{1}{2} \Delta f I_n$, где $\Delta f = 1/T$ — разность частот между смежными сигналами, при которой обеспечивается их ортогональность; $I_n = \pm 1, \pm 3, \dots, \pm(m-1)$. Переход от одной частоты к другой производится по закону модулирующей дискретной последовательности, показанной на рис. 5.9, а. Модулированный ДЧМ-сигнал для случая $M > 2$ в частотно-временной области (в виде частотно-временной матрицы) представлен на рис. 5.9, б.

Для осуществления ДЧМ-модуляции используют $M = 2^k$ отдельных генераторов, настроенных на заданные частоты. При этом для передачи в канал выбирается одна из M частот, соответствующая k -символьному блоку информации.

При использовании посылок одинаковой энергии, равной E , ДЧМ-сигнал можно представить выражением

$$S_m(t) = \operatorname{Re}\{\dot{S}_m(t)e^{j\omega t}\} = \sqrt{2E/T} \cos(\omega t + m\Delta\omega t);$$

$$m = \overline{1, M}; t \in [0, T]; \Delta\omega = 2\pi\Delta f,$$

где $\dot{S}_m(t) = \sqrt{2E/T} e^{jm\Delta\omega t}$ — эквивалентный низкочастотный сигнал.

В частном случае, когда $M = 2$, т.е. когда число реализаций равно двум, ДЧМ-сигнал можно записать в следующем виде:

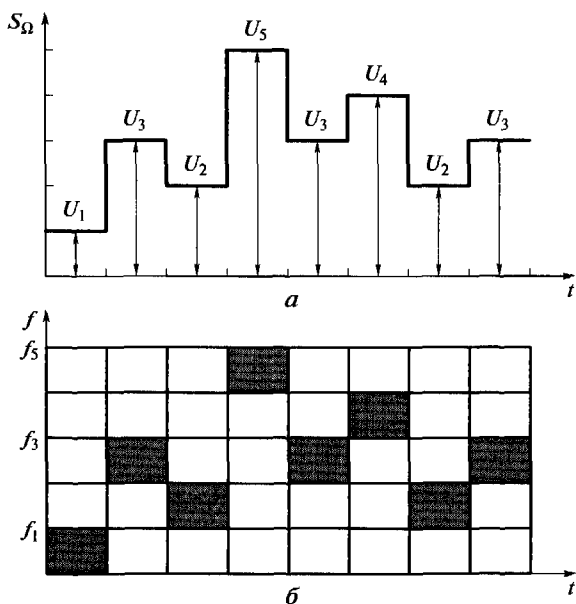


Рис. 5.9. Пояснения принципа дискретной частотной модуляции: a — управляющая последовательность; b — частотно-временная матрица

$$S_0(t) = \sqrt{2E/T} \cos(\omega_0 t + \varphi_0);$$

$$S_1(t) = \sqrt{2E/T} \cos(\omega_1 t + \varphi_1),$$

где $\Delta\omega = (\omega_1 - \omega_0)$ — разнос частот; $\omega_1 > \omega_0$.

Девияция частоты

$$\Delta\omega_d = \Delta\omega / 2 = \frac{\omega_1 - \omega_0}{2}.$$

Дискретная частотная модуляция характеризуется еще одним параметром — индексом частотной модуляции, под которым, как и ранее, понимается отношение девиации частоты $\Delta\omega_d$ к частоте Ω модулирующего сигнала или средней частоте следования посылок, т. е.

$$m_f = \Delta\omega_d / \Omega_m,$$

где $\Omega = \frac{2\pi}{2T} = \pi/T$; T — длительность посылки сигнала.

Определим спектр ДЧМ-сигнала при $M = 2$. Частота сигнала в этом случае

$$\omega = \omega_c + E_c \Delta\omega_d.$$

Здесь средняя частота сигнала $\omega_c = \frac{\omega_1 + \omega_0}{2}$, а двухполярный модулирующий сигнал

$$E_c = \begin{cases} +1, & t \in [(k-1)T, kT]; \\ -1, & t \in [kT, (k+1)T], \end{cases}$$

где $k = 1, 2, \dots$

Тогда ДЧМ-сигнал можно записать в виде

$$S(t) = \sqrt{2E/T} \cos \Psi(t).$$

Здесь

$$\Psi(t) = \int_0^t (\omega_c + E_c \Delta \omega_d) dt = \int_0^t (\omega_c + \Delta \omega_d \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{\sin(2k-1)\Omega t}{2k-1}) dt.$$

После преобразования получим выражение

$$\begin{aligned} S(t) = \sqrt{2E/T} \frac{2}{\pi} & \left\{ \frac{1}{m_f} \sin\left(\frac{\pi}{2} m_f\right) \cos \omega_c t + \right. \\ & + \frac{m_f}{m_f^2 - 1} \cos\left(\frac{\pi}{2} m_f\right) [\cos(\omega_c - \Omega)t - \cos(\omega_c + \Omega)t] - \\ & - \frac{m_f}{m_f^2 - 2} \sin\left(\frac{\pi}{2} m_f\right) [\cos(\omega_c - 2\Omega)t + \cos(\omega_c + 2\Omega)t] - \\ & \left. - \frac{m_f}{m_f^2 - 3^2} \cos\left(\frac{\pi}{2} m_f\right) [\cos(\omega_c - 3\Omega)t + \cos(\omega_c + 3\Omega)t] + \dots \right\}. \end{aligned}$$

В качестве примера на рис. 5.10 представлен спектр ДЧМ-сигнала при $m_f = 2$.

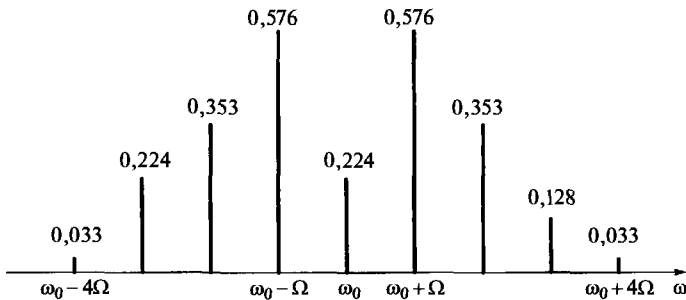


Рис. 5.10. Спектр сигнала с дискретной частотной модуляцией при индексе модуляции $m_f = 2$

Анализируя спектр ДЧМ-сигнала, можно отметить следующее.

1. С увеличением индекса частотной модуляции m_f амплитуда несущего колебания уменьшается.

2. При индексе модуляции, близком к единице ($m_f \approx 1$), основная часть мощности сигнала заключена в несущей частоте и боковых составляющих на частотах $\omega_c + \Omega$ и $\omega_c - \Omega$.

3. Ширина спектра ДЧМ-сигнала примерно в два раза превышает ширину спектров сигналов ДАМ и ДФМ.

В заключение отметим, что демодуляция сигналов дискретной частотной модуляции осуществляется на основе тех же схем, что и при аналоговой частотной модуляции.

5.6. Дискретная частотная модуляция с непрерывной фазой

Ранее был рассмотрен метод дискретной частотной модуляции, основанный на подключении к входу канала одного из нескольких генераторов частот. В этом случае выбор конкретного генератора определяется некоторой комбинацией символов модулирующей последовательности. Причем в момент переключения может возникать разрыв фазы, так как при подключении очередного генератора фазы вырабатываемых им колебаний имеют любые произвольные значения.

Для данного метода, иногда называемого методом с разрывом фазы, характерны следующие недостатки.

1. Значительные переходные процессы, которые приводят к искажениям сигнала.

2. Появление дополнительных частотных составляющих, которые находятся за пределами основного спектра сигнала, что вызывает увеличение полосы частот для его передачи.

3. Большой уровень внутрисистемных помех в многоканальных системах передачи информации.

Для устранения указанных недостатков применяется дискретная частотная модуляция без разрыва фазы, или с непрерывной фазой — ДЧМНФ (Continuous phase — FSK). При ДЧМНФ информационный сигнал модулирует всего одну несущую, которая изменяется непрерывно в результате подключения дополнительных емкостей к колебательному контуру генератора гармонической несущей.

В общем случае управляющий (модулирующий) сигнал имеет вид дискретной последовательности прямоугольных импульсов, амплитуды которых принимают значения на множестве $I_n = \pm 1, \pm 3, \dots, \pm(M-1)$, где M — четное число.

Для упрощения задачи рассмотрим случай, когда $I_n = \pm 1$.

В этом случае модулированный сигнал можно представить в виде

$$S(t) = \sqrt{2E/T} \cos(\omega t + \Psi(t) + \varphi_0); \quad t \in [0, T],$$

где $\Psi(t)$ — изменяющаяся во времени фаза несущей; φ_0 — начальная фаза сигнала.

Фаза сигнала, который передается в интервале времени $(n-1)T \leq t \leq nT$, где $n = 1, 2, \dots$, имеет вид

$$\Psi(t) = \pi m_f \sum_{k=1}^{n-1} I_k + \frac{\pi m_f I_n (t - (n-1)T)}{T},$$

где $m_f = \frac{\Delta\omega T}{2\pi} = \Delta f T$ — индекс частотной модуляции; $\Delta f = \frac{\Delta\omega}{2\pi}$ — разность между частотами, соответствующими передаче определенного информационного символа; $I_n = \pm 1$ — последовательность информационных символов; T — длительность информационной посылки.

В этом выражении первое слагаемое соответствует значению фазы, накопленному от предыдущих информационных символов, переданных до момента $t = (n-1)T$. Второе слагаемое характеризует текущее изменение фазы во время передачи очередной посылки до момента $t = nT$.

При ДЧМНФ набег, или приращение, фазы за время передачи очередной посылки I_n составляет

$$\Delta\varphi = \frac{\pi m_f I_n (nT - (n-1)T)}{T} = \pm \pi m_f.$$

Очень часто для ДЧМНФ-сигналов строят графические зависимости возможных законов изменения фазы несущего колеба-

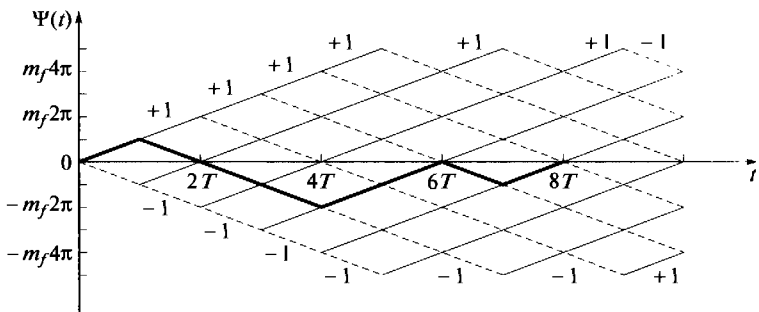


Рис. 5.11. Варианты фазовых траекторий при двоичной ДЧМНФ

ния. Такие зависимости связывают с возможными значениями информационных посылок и называют фазовыми траекториями (фазовыми деревьями). Так, для ДЧМНФ с двоичными символами $I_n = \pm 1$ фазовые траектории имеют вид, представленный на рис. 5.11, где сплошные линии соответствуют передаче символа «+1», а штриховые — передаче символа «-1». Фазовое дерево является кусочно-линейным вследствие того, что модулирующие (управляющие) импульсы выбираются прямоугольными.

Пример фазовой траектории, генерируемой последовательно-стью вида (+1, -1, -1, -1, +1, +1, -1, +1) также показан на рис. 5.11.

В заключение отметим, что в отличие от методов модуляции, рассмотренных ранее, ДЧМНФ-сигналы относятся к сигналам, имеющим память.

5.7. Дискретная частотная модуляция с минимальным сдвигом

Дискретной частотной модуляцией с минимальным сдвигом (ММС) называют частный случай модуляции с непрерывной фазой, при которой индекс модуляции $m_f = 0,5$. В иностранной литературе — Minimum (frequency) Shift Keying — MSK. Данный вид модуляции появился в начале 1970-х гг.

Фаза несущего колебания ММС-сигнала в интервале времени $nt \leq t \leq (n+1)T$ принимает вид

$$\begin{aligned} \Psi(t) &= \frac{\pi}{2} \sum_{k=1}^{n-1} I_k + \frac{\pi}{2T} I_n (t - (n-1)T) = \\ &= \frac{\pi I_n}{2T} t + \frac{\pi}{2} \sum_{k=1}^{n-1} I_k - (n-1) \frac{\pi}{2} I_n = \frac{\pi I_n}{2T} t + \varphi_n. \end{aligned}$$

Тогда ММС-сигнал можно представить выражением

$$S(t) = \sqrt{2E/T} \cos \left(\omega_c t + \frac{\pi I_n}{2T} t + \varphi_n \right); \quad \omega_c = 2\pi f_c.$$

С учетом того что $I_n = \pm 1$, набег фазы за время посылки $t = T$ составит $\Delta\varphi = \pm \frac{\pi}{2}$. При передаче символа «+1» значение $\Delta\varphi = 90^\circ$,

а при передаче символа «-1» значение $\Delta\varphi = -90^\circ$.

ММС-сигнал можно также записать и в другом виде:

$$S(t) = \sqrt{2E/T} \cos \left(2\pi \left(f_c + \frac{I_n}{4T} \right) t + \varphi_n \right).$$

Введем обозначения частот:

$$f_0 = f_c + \frac{1}{4T}; f_1 = f_c - \frac{1}{4T}.$$

Тогда

$$S_i(t) = \sqrt{2E/T} \cos(2\pi f_i t + \varphi_n); i = 0, 1.$$

Данное выражение соответствует прямому методу получения ММС-сигнала, так как частота каждой посылки в интервале T определяется непосредственно передаваемым в текущий момент времени символом: «0» или «1» (с учетом замены символа «0» на +1, а символа «1» на -1).

Возможен также квадратурный метод формирования ММС-сигналов, который определяется следующим представлением:

$$\begin{aligned} S(t) &= \sqrt{2E/T} \cos\left(\omega_c t + \frac{\pi I_n}{2T} t + \varphi_n\right) = \\ &= \sqrt{2E/T} \cos\left(I_n \frac{\pi t}{2T}\right) \cos(\omega_c t + \varphi_n) + \\ &+ \sqrt{2E/T} \sin\left(I_n \frac{\pi t}{2T}\right) \sin(\omega_c t + \varphi_n); t \in [0, T]. \end{aligned}$$

Схема формирователя ММС-сигнала, реализующая данное представление, приведена на рис. 5.12. Здесь на вход устройства поступает двухполярный (в виде посылок ± 1) управляющий видеосигнал, который с помощью коммутатора поочередно с частото-

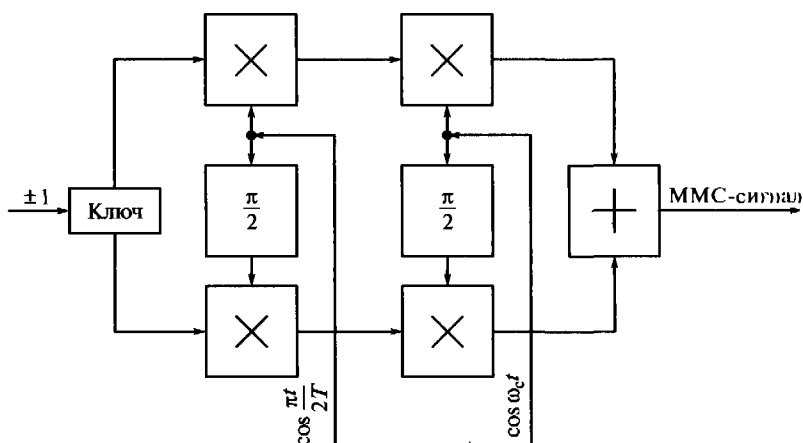


Рис. 5.12. Схема формирователя ММС-сигнала

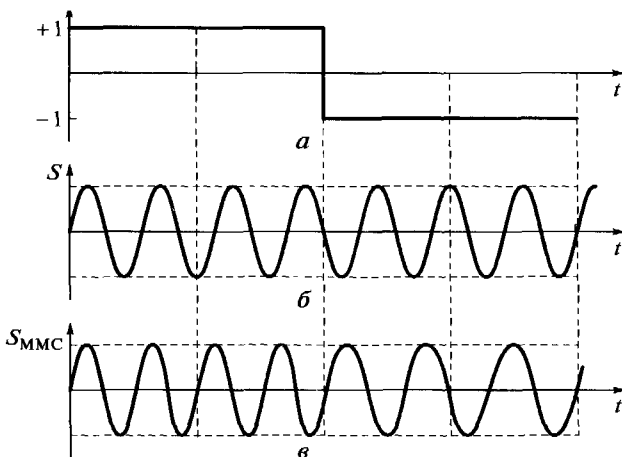


Рис. 5.13. Временное представление ММС-сигнала:

a — управляющий сигнал; *б* — несущее колебание; *в* — модулированный сигнал

той $1/(4T)$ подключается к каналам, где формируются квадратурные составляющие. При этом на выходе из сумматора получают ММС-сигнал.

Пример такого сигнала представлен на рис. 5.13.

Ранее было показано, что ДЧМНФ-сигналы, к которым относятся и ММС-сигналы, можно описывать с помощью фазовых деревьев. Однако это не всегда удобно, поскольку они растут с увеличением времени и становятся чрезвычайно громоздкими. Поэтому в целях упрощения вместо дерева применяют структуру, называемую фазовой решеткой. Такая решетка строится с учетом однозначности фазы лишь в интервале от 0 до 2π , или (что эквивалентно) от $\varphi = -\pi$ до $\varphi = \pi$. При построении решетки показывают финальные значения фазы, которые имеет сигнал только в моменты времени, кратные длительности посылки $t = nT$. В частности, для ММС-сигнала финальные состояния фазы $\varphi = \left\{ 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2} \right\}$.

Используя приведенное ранее фазовое дерево, несложно построить фазовую решетку для ММС-сигнала с индексом модуляции $m_f = 0,5$ (рис. 5.14).

Для преобразования дерева в решетку достаточно определить значения фазы по модулю 2π в области $(0, 2\pi)$. После этого необходимо провести все возможные пути, ведущие в точки с заданным значением фазы. Так, например, чтобы из точки 0 попасть в точку $-\pi/2$ при $t = 3T$, по кодовому дереву необходимо

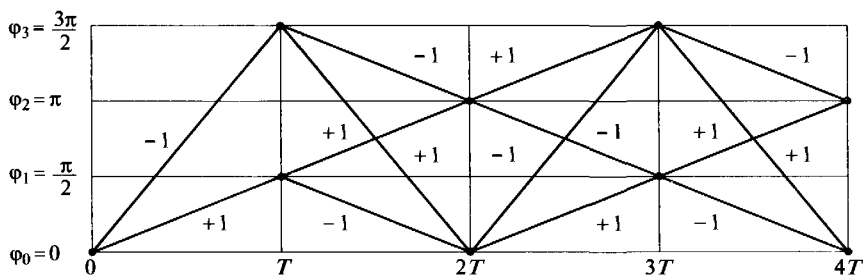


Рис. 5.14. Фазовая решетка состояний двоичного ММС-сигнала

пройти либо путь $-1, -1, +1$, либо путь $+1, -1, -1$. На решетке с учетом приведения фазы по модулю 2π финальная точка $-\pi/2$ будет соответствовать значению фазы $3\pi/2$. Первый путь при $t = 3T$ проходит от 0 через точки $3\pi/2$ ($-\pi/2$), π ($-\pi$) и на последнем этапе через отрезок $+1$ в заданную точку $3\pi/2$ ($-\pi/2$). Второй путь также начинается в нулевой точке и через точки $\pi/2, 0$, отрезок -1 приводит в точку $3\pi/2$ ($-\pi/2$). (Здесь в скобках указаны значения фаз дерева, которые приведены по модулю 2π к значениям фаз решетки.)

Рассмотрим теперь метод приема ММС-сигналов. Приемник таких сигналов может быть реализован на основе последовательного соединения двухканальной схемы синхронного приема, где обрабатываются сигналы суммарной или разностной частоты $f_c \pm \frac{1}{4T}$, и устройства, которое отыскивает путь по решетке на

основе минимального евклидова расстояния от принятого сигнала. Поиск может быть осуществлен на основе алгоритма Витерби.

На принципах работы первого из устройств обработки сигнала останавливаться не будем. Подробнее такие схемы оптимального приема будут рассмотрены несколько позднее. Отметим только, что задачей этой части демодулятора является вычисление метрик (евклидова расстояния между принятым и ожидаемым сигналами):

$$\Delta_n(I) = \Delta_{n-1}(I) + \int_{nT}^{(n+1)T} [Z_n(t) - S_{in}(t)]^2 dt; \quad i = \overline{1, M}; \quad n = 1, 2, \dots$$

Здесь первое слагаемое $\Delta_{n-1}(I)$ является метрикой путей (ветвей) до момента времени $t = nT$, а второе слагаемое представляет собой дополнительный прирост метрики, который вносит сигнал $I = \pm 1$ в интервале $t \in [nT, (n+1)T]$, что соответствует объему алфавита ММС-сигналов при $M = 2$. Решетка сигнала, как уже отмечалось ранее, имеет четыре фазовых состояния.

5.8. Обработка ММС-сигналов на основе алгоритма Витерби

Данный алгоритм предложил в 1967 г. американский ученый А. Витерби. Алгоритм предназначен для определения методом динамического программирования кратчайшего пути в ориентированном графе, т. е. в решетке.

Рассмотрим алгоритм Витерби на примере сигналов с относительным (дифференциальным) кодированием, которые являются более простыми, чем ММС-сигналы, но также обладают памятью.

На рис. 5.15 представлены два вида сигналов. Первый из этих видов — последовательность двухполярных прямоугольных импульсов, в которой произведена замена информационных посылок 1 и 0 соответственно на +1 и -1. Второй вид сигналов отличается от первого тем, что здесь переход от одного уровня амплитуды к другому производится только в случае передачи посылки 1. При передаче символа 0 полярность посылки сохраняется и на следующем шаге. Такой вид преобразования называется относительным, или дифференциальным, кодированием. Математически это преобразование описывается выражением

$$b_k = a_k \oplus b_{k-1},$$

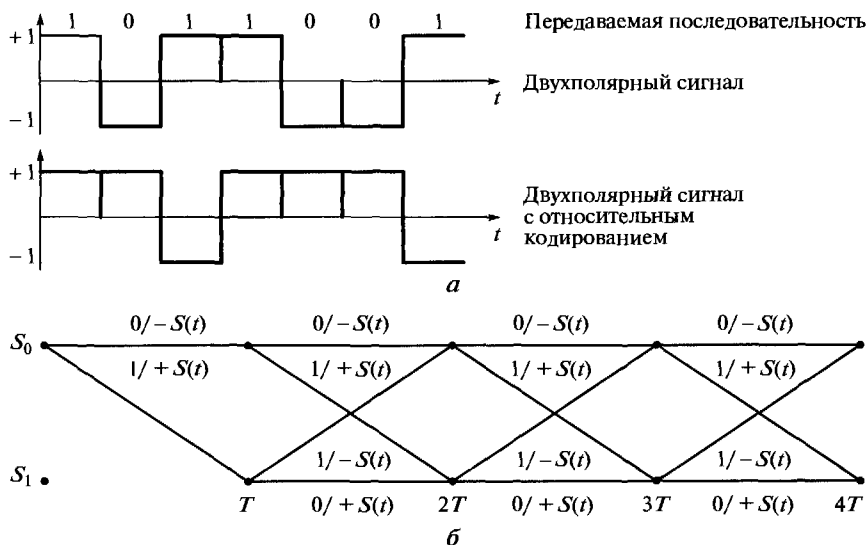


Рис. 5.15. Сигнал с относительным кодированием (а) и его фазовая решетка (б)

где b_k — перекодированная двоичная последовательность; a_k — двоичная информационная последовательность, подлежащая кодированию; \oplus — знак сложения по модулю два, которое производится по правилу $0 \oplus 0 = 0$, $0 \oplus 1 = 1$, $1 \oplus 0 = 1$, $1 \oplus 1 = 0$.

Таким образом, если $b_k = 1$, то передается посылка $+1$, если $b_k = 0$, то передается импульс с амплитудой -1 . Заметим, что относительное кодирование вводит в сигнал память.

Так же, как и ММС-сигналы, процесс передачи на основе дифференциального кодирования, с одной стороны, можно описать с помощью решетки, которая выглядит в этом случае несколько проще, что облегчает рассмотрение алгоритма Витерби. С другой стороны, относительное кодирование имеет самостоятельное значение и будет рассмотрено далее.

В решетке для дифференциального кодирования, представленной на рис. 5.15, б, возможны два сигнала, соответствующие сигнальным точкам $S_0 = -S_1$ или $S_1 = 0$, $S_1 = 1$. Состояния системы отмечены в решетке точками (узлами), а в переходах (на путях) между ними показаны поступающие информационные символы и передаваемые сигналы. Память (L) сигналов с дифференциальным кодированием равна одному символу, т.е. $L = 1$. Легко увидеть, что структура решетки регулярно повторяется после двух начальных переходов. Так, в точке $t = 2T$ имеется два пути, входящих в каждый узел, и два пути, которые выходят из каждого узла.

Перейдем теперь непосредственно к алгоритму Витерби. Допустим, что процесс поиска начинается с состояния S_0 . В точке $t = T$ на выходе приемника имеем сигнал $Z_1(t) = S_1(t) + n_1(t)$, а в точке $t = 2T$ — смесь сигнала и помехи $Z_2(t) = S_2(t) + n_2(t)$, где $n(t)$ — аддитивная помеха. Пути, входящие в точку S_0 при $t = 2T$, соответствуют посылкам $0, 0$ и $1, 1$ или сигналам $-S, -S$ и $S, -S$. Пути, входящие в узел S_1 при $t = 2T$, соответствуют посылкам $0, 1$ и $1, 0$ или сигналам $-S, S$ и S, S .

Для точки S_0 при $t = 2T$ вычислим значения метрик евклидова расстояния:

$$\begin{aligned}\Delta_0(0, 0) &= (Z_1 + S)^2 + (Z_2 + S)^2; \\ \Delta_0(1, 1) &= (Z_1 - S)^2 + (Z_2 + S)^2.\end{aligned}$$

Эти метрики сравниваются, после чего путь, имеющий большее значение метрики отбрасывается, а путь с меньшим значением метрики запоминается, и его называют выбранным.

Аналогично вычисляются метрики для двух путей, которые входят в точку S_1 в момент $t = 2T$:

$$\begin{aligned}\Delta_0(0, 1) &= (Z_1 + S)^2 + (Z_2 - S)^2; \\ \Delta_0(1, 0) &= (Z_1 - S)^2 + (Z_2 - S)^2.\end{aligned}$$

Далее эти метрики сравниваются, и путь с большим значением метрики отбрасывается, а с меньшим — запоминается. Таким образом, в момент времени $t = 2T$ сохраняются два пути, один из которых соответствует сигналу S_0 , а другой — сигналу S_1 .

При приеме сигнала $S_3(t)$ в момент $t = 3T$ вновь рассчитаем метрики двух путей, входящих в состояние S_0 . Допустим, что при $t = 2T$ выбранными являются пути $(0, 0)$ в состоянии S_0 и $(0, 1)$ — в состоянии S_1 . Тогда две метрики для путей, входящих в S_0 , запишем в виде

$$\Delta_0(0, 0, 0) = \Delta_0(0, 0) + (Z_3 + S)^2;$$

$$\Delta_0(0, 1, 1) = \Delta_1(0, 1) + (Z_3 + S)^2.$$

После сравнения этих метрик выбранным будем считать путь с меньшим значением метрики. Аналогично метрики двух путей, входящих в узел, соответствующий сигналу S_1 в момент $t = 3T$, запишем в виде

$$\Delta_1(0, 0, 1) = \Delta_0(0, 0) + (Z_3 - S)^2;$$

$$\Delta_1(0, 1, 0) = \Delta_1(0, 1) + (Z_3 - S)^2.$$

И снова отбросим один из путей с большей метрикой.

Далее процесс будет продолжаться до тех пор, пока очередные сигналы поступают на вход приемника. По алгоритму Витерби вычисляются две метрики для двух вариантов путей, входящих в узел на каждом шаге поиска по решетке. Затем исключается один из путей на каждом узле. Два выбранных пути далее позволяют продвигаться вперед до следующего состояния $t = nT$. В итоге число возможных вариантов путей в процессе поиска по решетке сокращается в два раза на каждом шаге.

На практике считается вполне достаточным число шагов поиска по решетке, равное $5L$, где L — память (в нашем случае $L = 1$). После этого на основании мажоритарного правила выносится решение о переданном информационном символе. Иными словами, принятым считается тот символ, сумма метрик путей для которого оказалась наименьшей (либо наибольшей).

Преимущество алгоритма Витерби состоит в значительном сокращении числа операций в процессе приема двоичной последовательности длиной n символов. В общем случае для решения такой задачи необходимо перебрать 2^n вариантов. При больших значениях n это приводит к необозримым количествам вычислений. Алгоритм Витерби уменьшает число последовательностей при поиске по решетке за счет отсеечения тех из них, которые не имеют минимального значения суммы метрик на всех предыдущих шагах.

Алгоритм Витерби находит в настоящее время широкое применение в технике электрической связи не только для обработки ММС-сигналов, но и для декодирования сверточных кодов. Отметим также, что ММС-сигналы реализуются в системах сотовой связи, использующих стандарт GSM.

5.9. Амплитудно-импульсная модуляция

Ранее отмечалось, что помимо гармонических колебаний в качестве переносчиков информации в системах телекоммуникаций применяются периодические импульсные последовательности. Рассмотрение импульсных методов модуляции обычно начинают с амплитудно-импульсной модуляции (АИМ), которая так же, как и другие виды импульсной модуляции, основана на дискретизации во времени, т. е. на замене непрерывного сигнала последовательностью его мгновенных значений (отсчетов), взятых в дискретные моменты времени. Согласно теореме отсчетов интервалы между соседними выборками $\Delta t = 1 / (2F_b)$, где F_b — верхняя граничная частота полосового сигнала.

Для описания дискретизации во времени используем так называемую импульсную функцию дискретизации a_d , которая представляет собой периодическую последовательность δ -функций, следующих друг за другом через интервалы времени Δt :

$$a_d(t) = \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t).$$

Математическую модель дискретизации сигнала $S(t)$ во временной области можно представить как результат умножения этого сигнала на функцию a_d :

$$S_d(t) = \int_{-\infty}^{\infty} S(t)a_d(t)dt = \sum_{k=-\infty}^{\infty} S(k\Delta t)\delta(t - k\Delta t),$$

где $S(k\Delta t)$ — выборка (отсчет) непрерывного сигнала $S(t)$ в момент времени $t = k\Delta t$.

Таким образом получим периодическую последовательность δ -импульсов, умноженных на значения сигнала в моменты взятия отсчетов (см. рис. 2.5). На самом деле реализовать импульсную функцию дискретизации на основе δ -импульсов невозможно, поэтому при взятии отсчетов обычно используется периодическая последовательность коротких прямоугольных импульсов с длительностью τ , которые удовлетворяют условию $\tau / \Delta t \ll 1$. В результате получим амплитудно-модулированную последовательность импульсов, т. е. АИМ-сигнал.

На практике на один из входов импульсного модулятора подается аналоговый низкочастотный сигнал $u_{\Omega}(t)$, а на другой его вход поступают импульсы длительностью τ с интервалом повторения Δt . На выходе модулятора появляется последовательность импульсов, амплитуда и площадь каждого из которых пропорциональна отсчетному значению амплитуды аналогового сигнала. Таким образом, АИМ-сигнал можно получить посредством модуляции по амплитуде периодической последовательности импульсов. При этом остальные их параметры остаются неизменными.

Известны два вида амплитудно-импульсной модуляции: АИМ-1 и АИМ-2 (рис. 5.16). В первом случае вершина импульса следует за изменениями модулирующего напряжения, а во втором — вершина импульса сохраняется плоской.

На практике чаще применяются сигналы АИМ-2 с плоской вершиной импульсов, равной сигналу $u_{\Omega}(t)$ в момент времени, соответствующий середине текущего импульса. Выравниванием вершины импульса обеспечивается более высокая помехоустойчивость сигнала АИМ-2 по сравнению с сигналом АИМ-1.

Рассмотрим теперь спектральное представление АИМ-сигнала. Периодическую импульсную последовательность можно записать в виде

$$S_0(t) = A_0 \left(1 + 2 \sum_{k=1}^{\infty} \frac{\sin k\pi/q}{k\pi/q} \cos k\Omega_0 t \right),$$

где $A_0 = A/q$; A — амплитуда импульсов; $q = T/\tau$ — скважность импульсов; $\Omega_0 = 2\pi/T$ — частота следования импульсов.

Тогда АИМ-сигнал будет иметь вид

$$S_0(t) = A_0 [1 + m u_{\Omega}(t)] \left(1 + 2 \sum_{k=1}^{\infty} \frac{\sin k\pi/q}{k\pi/q} \cos k\Omega_0 t \right),$$

где m — коэффициент модуляции; $u_{\Omega}(t)$ — низкочастотный модулирующий сигнал.

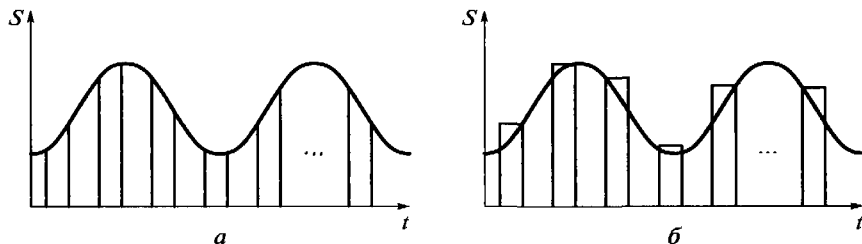


Рис. 5.16. Временное представление сигналов в виде АИМ-1 (а) и АИМ-2 (б)

Спектр АИМ-сигнала представляет собой как бы многократно повторенный спектр обычной амплитудной модуляции, в котором роль несущих частот выполняют гармоники частоты следования импульсов. Передача АИМ-сигнала будет осуществляться без искажений при условии, что частота повторения импульсов

$$F = \frac{1}{T} \geq 2F_{\text{в}}, \text{ что вытекает из теоремы отсчетов.}$$

Для реализации передачи по каналу связи АИМ-сигнал формируется в два этапа:

1) по закону низкочастотного модулирующего колебания изменяется амплитуда периодической последовательности импульсов постоянного тока (видеоимпульсов) небольшой длительности;

2) посредством видеоимпульсов модулируется высокочастотное гармоническое колебание. Для этого применяется, как правило, амплитудная модуляция, но может использоваться и любой другой из аналоговых видов модуляции. В результате осуществляется перенос спектра модулированных видеоимпульсов на частоту несущего колебания. Полученный при этом сигнал принято обозначать АИМ-АМ.

Демодуляция таких сигналов осуществляется на основе методов, рассмотренных ранее.

В заключение отметим, что АИМ-сигналы обладают невысокой помехоустойчивостью.

5.10. Широтно-импульсная модуляция

Широтно-импульсной модуляцией (ШИМ) называется такой вид модуляции, при котором информационным параметром является длительность импульсов. Различают одностороннюю ШИМ, при которой изменение длительности посылки происходит за счет перемещения одного из фронтов импульсов, и двухстороннюю ШИМ, при которой длительность посылки изменяется в результате одновременного перемещения обоих фронтов импульсов. Для практической реализации более удобной является односторонняя широтно-импульсная модуляция.

По аналогии с видами амплитудно-импульсной модуляции существует два вида широтно-импульсной модуляции: ШИМ-1 и ШИМ-2. Сущность этих видов ШИМ легко понять с помощью рис. 5.17, из которого видно, что для получения импульсной последовательности, модулированной по длительности, необходимо сравнить непрерывный сигнал $u_{\Omega}(t)$ с периодической последовательностью пилообразных импульсов $u_{\text{п}}(t)$. При этом начало «зуб-

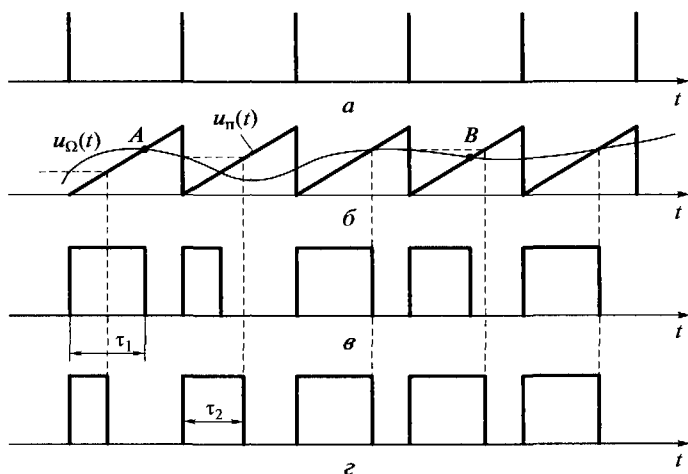


Рис. 5.17. Временное представление сигналов с широтно-импульсной модуляцией:

a — стробирующие импульсы; *б* — аналоговый сигнал и пилообразные импульсы; *в* — импульсы ШИМ-1; *г* — импульсы ШИМ-2

ца пилы» соответствует переднему фронту импульса ШИМ, а задний фронт импульса формируется по точкам пересечения *A* и *B*, которые соответствуют равенству мгновенных значений сигналов $u_{\Omega}(t)$ и $u_n(t)$. Таким образом, длительность импульса ШИМ-1 равна соответственно τ_1 , а ШИМ-2 — τ_2 .

Рассмотрим теперь спектр ШИМ-сигнала. Пусть прямоугольные немодулированные импульсы имеют длительность τ_n . В качестве начала времени отсчета примем середину этих импульсов. Период повторения видеоимпульсов $T = \frac{2\pi}{\Omega_0}$. При этом любой из импульсов можно представить в виде

$$u(t) = \begin{cases} A_0, & t \in [t_1, t_2]; \\ 0, & t_1 > t, t > t_2, \end{cases}$$

где t_1 — момент начала импульса; t_2 — момент окончания импульса; $t_2 - t_1 = \tau_n$.

Для модуляции используем гармоническое колебание вида $u_{\Omega}(t) = \sin \Omega t$, в соответствии с которым изменяется временное положение только переднего фронта видеоимпульсов на значение $\Delta\tau \leq \tau_n$. Тогда положение фронтов при модуляции можно описать выражениями

$$t_1(t) = -0,5\tau_n - \Delta\tau \sin \Omega t; \quad t_2(t) = 0,5\tau_n.$$

С учетом этих выражений представим промодулированную последовательность видеоимпульсов в виде ряда Фурье:

$$S(t) = \frac{A_0\tau_n}{T} + \frac{A_0\Delta\tau}{T} \sin \Omega t + \frac{A_0}{\pi} \sum_{k=1}^{\infty} \frac{1}{k} \times \\ \times \left[\sin k\Omega_0 \left(t + \frac{\tau_n}{2} + \Delta\tau \sin \Omega t \right) - \sin k\Omega_0 \left(t - \frac{\tau_n}{2} \right) \right].$$

Используем известное из теории бесселевых функций соотношение

$$\sin(x + y \sin z) = \sum_{n=-\infty}^{\infty} I_n(y) \sin(x + nz),$$

где $I_n(y)$ — функция Бесселя 1-го рода n -го порядка.

Тогда выражение для ШИМ-сигнала можно записать в виде:

$$S(t) = A_0 \left\{ \frac{1}{q} + \frac{\Delta\tau}{q\tau_n} \sin \Omega t - \sum_{k=1}^{\infty} \frac{1}{\pi k} \sin k\Omega_0 \left(t - \frac{\tau_n}{2} \right) + \right. \\ \left. + \sum_{k=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{\pi k} I_n \left(\frac{2k\pi \Delta\tau}{q\tau_n} \right) \sin \left[(k\Omega_0 + n\Omega)t + \frac{\pi k}{q} \right] \right\}.$$

Анализируя спектр ШИМ-сигнала, можно отметить следующее.

1. Спектр ШИМ-сигнала содержит следующие составляющие: постоянную и компоненты с частотой Ω , с частотами $k\Omega_0$ и с частотами $k\Omega_0 \pm n\Omega$, причем число составляющих $k\Omega_0$ и $k\Omega_0 \pm n\Omega$ бесконечно велико.

2. Спектр ШИМ-сигнала оказывается сложнее спектра АИМ-сигнала при одном и том же законе модуляции. Они могут быть похожи только при малых значениях девиации импульсов, т. е. когда $\Delta\tau \ll \tau_n$.

3. Амплитуды боковых частот ШИМ-сигнала определяются значениями функций Бесселя n -го порядка в зависимости от отношения $\Delta\tau/\tau_n$, которое называется коэффициентом глубины модуляции длительности импульсов. Предельное значение $\Delta\tau/\tau_n = 1$.

Процедура демодуляции ШИМ-сигналов осуществляется в два этапа. Сначала ШИМ-сигнал преобразуют в АИМ-сигнал, а затем с помощью фильтра выделяют модулирующий сигнал. В этом случае удастся избежать искажений, которые могут возникать за счет составляющих с частотами $k\Omega_0 - n\Omega$ и значительными амплитудами.

5.11. Временная (фазовая) импульсная модуляция

Временной импульсной модуляцией (ВИМ) называется вид модуляции, при которой информационным параметром является временной сдвиг импульса относительно его среднего положения. При этом значение сдвига определяется законом моделирующего колебания.

Временную импульсную модуляцию иногда называют фазоимпульсной модуляцией (ФИМ). Ее сущность поясняет рис. 5.18.

ВИМ-сигналы формируются на основе сравнения модулирующего колебания $u(t)$ с последовательностью пилообразных импульсов $u_{н}(t)$, поэтому по аналогии с ШИМ при практической реализации различают сигналы ВИМ-1 и ВИМ-2.

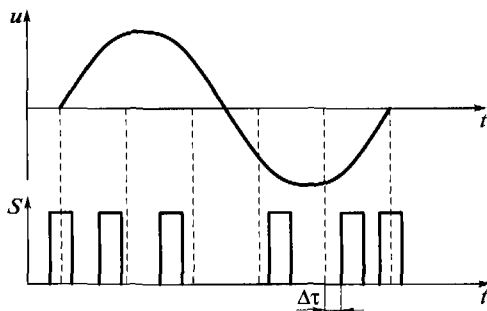


Рис. 5.18. Пояснение принципа временной импульсной модуляции

Данные сигналы обладают более высокой помехоустойчивостью в сравнении с рассмотренными видами импульсной модуляции, поскольку в случае ВИМ возможен выбор оптимальной полосы пропускания приемника вследствие постоянства длительности импульсов, что нереализуемо, в частности, при ШИМ.

По аналогии с предыдущими рассуждениями можно получить спектр ВИМ-сигналов. При этом оказывается, что спектр ВИМ-сигналов значительно сложнее не только спектра АИМ-сигналов, но и спектра ШИМ-сигналов.

В заключение отметим, что временная импульсная модуляция применяется в основном в многоканальных системах космической телеметрии для передачи аналоговой и дискретной информации. Кроме того, ВИМ-сигналы могут использоваться в оптических каналах связи.

5.12. Импульсно-кодовая модуляция

Импульсно-кодовая модуляция предназначена для передачи непрерывных сообщений в цифровой форме. В ее основу положены процедуры дискретизации сигналов по времени, квантование дискретных выборок по уровню и последующее их кодирование двоичным кодом.

В общем виде структурная схема системы передачи аналоговых (непрерывных) сообщений представлена на рис. 5.19. В данной схеме процессы, происходящие в аналого-цифровом преобразователе (АЦП), называются импульсно-кодовой модуляцией (ИКМ). Процедура кодирования здесь обычно представляет собой отображение десятичного номера уровня квантования в двоичной форме с помощью простого безыбыточного кода. Вместе с тем в системах с ИКМ широко применяются рассмотренные ранее рефлексные коды Грея.

Процедуру формирования ИКМ-сигнала иллюстрирует рис. 5.20.

Полученный на выходе АЦП (см. рис. 5.19) ИКМ-сигнал в передатчике преобразуется в последовательность радиоимпульсов, которые далее поступают в линию связи. На приемной стороне сигнал после демодуляции в приемнике поступает в цифроаналоговый преобразователь (ЦАП), где происходит обратное преобразование (восстановление) непрерывного сообщения. Для этого в состав ЦАП входит декодирующее устройство, с помощью которого кодовые комбинации превращаются в квантованную последовательность выборок (отсчетов). Затем с помощью сглаживающего фильтра по квантованным значениям происходит восстановление непрерывного сообщения.

Отметим, что процедура квантования при передаче сопровождается округлением мгновенных значений непрерывного со-



Рис. 5.19. Структурная схема системы цифровой передачи аналоговых сообщений на основе импульсно-кодовой модуляции

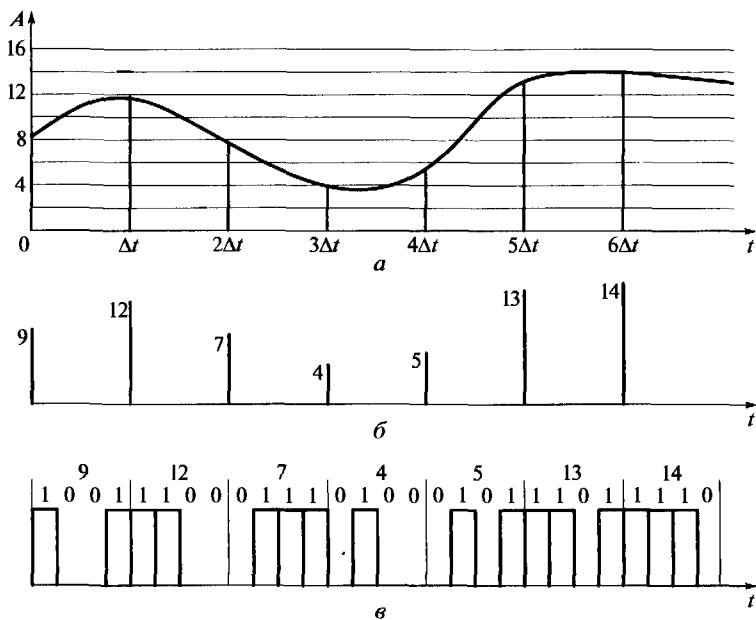


Рис. 5.20. Пояснение принципа формирования ИКМ-сигнала:

a — передаваемое сообщение; *б* — отсчеты дискретизации; *в* — кодированный сигнал

общения до ближайшего дискретного уровня. При этом возникает погрешность, называемая ошибкой (шумом) квантования. Шум квантования не связан с помехами в канале связи и определяется исключительно выбором числа уровней квантования. Увеличением числа уровней ошибку квантования можно уменьшить, однако при этом возрастает число кодовых символов, приходящихся на одну квантованную выборку. Это, в свою очередь, приводит к сокращению длительности символов и расширению спектра сигнала.

В соответствии с международными нормами для передачи стандартного телефонного сигнала с полосой частот $\Delta f = 3,1$ кГц предусматриваются частота дискретизации $f_d = 8$ кГц и число уровней квантования $L = 256$. Следовательно, на один отсчет приходится $n = 8$ двоичных разрядов ($n = \log_2 L$). При таком способе преобразования сигнала скорость передачи информации в канале $V = n f_d = 8 \cdot 8 \cdot 10^3 = 64 \cdot 10^3$ бит/с.

Для передачи сигналов радиовещательных станций с высоким качеством при использовании ИКМ должны обеспечиваться частота дискретизации $f_d = 32$ кГц и длина кодовой комбинации $n = 12$. В этом случае скорость передачи $V \approx 380$ бит/с.

При реализации ИКМ различают квантование двух видов: равномерное и неравномерное.

Вид характеристики квантователя с равномерным шагом показан на рис. 5.21. По горизонтальной оси здесь отложены значения входных величин, а по вертикальной — значения выходных дискретных величин. При равномерном квантовании размеры ступенек одинаковые, т. е. шаг квантования имеет постоянное значение:

$$\Delta = u_{\max} / 2_{n-1},$$

где u_{\max} — максимальное значение передаваемого аналогового сигнала.

Для повышения качества передачи сигналов при ИКМ применяют неравномерное квантование. В этом случае высокие уровни входных сигналов квантуются с большим шагом, а низкие уровни — с меньшим шагом. Такой подход позволяет гораздо лучше различать слабые уровни сигнала.

Для получения неравномерного квантования при передаче применяют нелинейное преобразование, называемое компрессией, в результате которого динамический диапазон сигнала «сжимается». Затем сигнал подвергается равномерному квантованию. Совокупность двух процедур компрессии и равномерного квантования оказывается эквивалентной неравномерному квантованию.

На приемной стороне осуществляют обратное преобразование, которое называется экспандированием, т. е. «растяжением» динамического диапазона.

Два нелинейных преобразования на передающей и приемной сторонах называют компандированием. Аналитически характеристику компрессора обычно аппроксимируют μ -соотношением:

$$b = f(a) = \frac{\lg(1 + \mu|a|)}{\lg(1 + \mu)} \operatorname{sign}(a); \quad -1 \leq a \leq 1; \quad -1 \leq b \leq 1;$$

$$\operatorname{sign}(a) = \begin{cases} 1, & a > 0; \\ 0, & a = 0; \\ -1, & a < 0. \end{cases}$$

Значение μ определяет глубину компрессии сигнала. Обычно рекомендуют $\mu = 100$ и 255 .

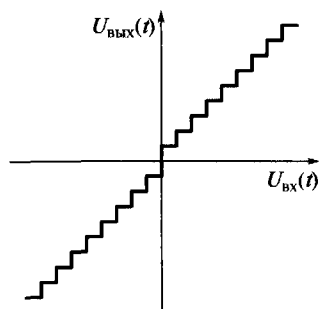


Рис. 5.21. Характеристика квантователя с равномерным шагом

В заключение отметим, что ИКМ-сигнал так же, как и другие рассмотренные ранее виды импульсной модуляции, перед передачей в канал связи подвергается вторичной модуляции. Для этого может использоваться любой из видов модуляции с гармоническим переносчиком, однако чаще применяются фазовые методы дискретной модуляции.

5.13. Дифференциальная импульсно-кодовая модуляция

В рассмотренной системе передачи информации с импульсно-кодовой модуляцией каждая квантованная выборка сигнала кодируется независимо от всех прочих отсчетов. Вместе с тем анализ речевых сигналов показывает, что соседние отсчеты имеют корреляционные взаимосвязи. Так, при частоте дискретизации $f_d = 8$ кГц коэффициент корреляции составляет 0,85 и более. Следовательно, естественную избыточность речи можно использовать для сокращения длины кодовой комбинации при кодировании номера квантованной выборки, что, в конечном счете, приводит к экономии полосы занимаемых частот. Одним из методов, позволяющих реализовать данную задачу, является кодирование с предсказанием (рис. 5.22).

Идея этого метода заключается в следующем: последовательность коррелированных отсчетов $b(k)$ передаваемого сигнала поступает на вход вычитающего устройства, на другой вход которого подается сигнал предсказания $b_{\text{пр}}(k)$, сформированный по предыдущим выборкам. В результате вычитания образуется сигнал ошибки предсказания $\varepsilon(k)$. Далее при использовании ИКМ этот сигнал подвергается процедурам квантования и кодирования, а затем поступает в тракт передачи. На приемной стороне посредством обратной процедуры переданный отсчет сигнала восстанавливается, для чего принятое значение ошибки складывается с предсказанным сигналом ошибки, который вырабатывается



Рис. 5.22. Принцип построения системы передачи информации с предсказанием

в приемнике. В результате получают истинное значение сигнала.

Недостатком рассмотренного метода передачи является эффект размножения ошибок. Это связано с тем, что погрешность восстановления сигнала при дифференциальной ИКМ определяется не только одним отсчетом, но и погрешностью ряда предыдущих выборок. Для устранения этого явления нередко применяют периодическую передачу полного значения сигнала.

В заключение отметим, что учет корреляционных связей передаваемого сигнала при дифференциальной ИКМ позволяет сократить длину кодовой комбинации на один-два разряда, что приводит к существенному снижению скорости передачи при сохранении прежнего качества восстановления сигнала. В некоторых случаях скорость снижается до $32 \cdot 10^3$ бит/с, что позволяет удвоить число каналов телекоммуникаций в типовых цифровых трактах, где скорость передачи равна $64 \cdot 10^3$ бит/с.

5.14. Дельта-модуляция

Под дельта-модуляцией (ДМ) понимают особый случай дифференциальной импульсно-кодовой модуляции, при которой разность текущей и предыдущей выборок передаваемого сигнала квантуется на два уровня. При этом разность отсчетов может принимать положительное или отрицательное значение, что указывает либо на увеличение, либо на уменьшение сигнала за время между двумя соседними выборками.

При дельта-модуляции в отличие от ИКМ каждый отсчет сигнала передается всего лишь одной двоичной посылкой, что требует более высокой частоты дискретизации, которая в этом случае будет равна скорости передачи информации.

В общем виде структурная схема системы с ДМ показана на рис. 5.23. В этой схеме квантованные отсчеты передаваемого сообщения после суммирования (интегрирования) поступают по цепи обратной связи на вычитающее устройство. Эти отсчеты можно представить в следующем виде:

$$u^*(t_n) = \Delta u \sum_{k=1}^{n-1} l_k; \quad u(t_0) = 0.$$

Здесь

$$l_k = \begin{cases} +1, & u(t_k) - u^*(t_{k-1}) \geq 0; \\ -1, & u(t_k) - u^*(t_{k-1}) < 0, \end{cases}$$

а Δu — шаг квантования.

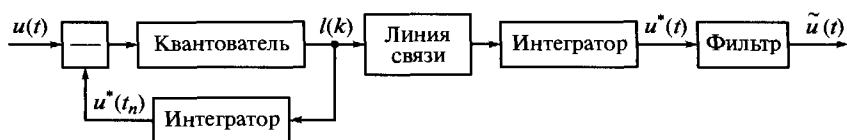


Рис. 5.23. Структурная схема системы передачи аналоговых сообщений на основе дельта-модуляции

В вычитающем устройстве происходит сравнение дискретного отсчета входного сигнала с предыдущим отсчетом, полученным в результате суммирования. На выходе квантователя в этом случае будут формироваться значения $l_k = \pm 1$, которые далее поступают в канал связи. Одновременно импульсы $\Delta u l_k$ подаются на сумматор (интегратор) для формирования очередного квантованного отсчета, как указывалось ранее. На выходе интегратора квантованный сигнал $u^*(t)$ имеет вид ступенчатой функции. Каждый импульс $+1$ увеличивает, а каждый импульс -1 уменьшает ступенчатую функцию на один шаг квантования, т. е. при дельта-модуляции соседние значения данной функции отличаются друг от друга на один шаг квантования.

На приемной стороне процедура обработки принятого сигнала реализуется с помощью интегратора, на выходе которого (при отсутствии помех) вновь формируется ступенчатое напряжение $u^*(t)$. После сглаживания этого напряжения с помощью фильтра получают сигнал $\tilde{u}(t)$, достаточно близкий к переданному сигналу $u(t)$.

Разность сигналов $\tilde{u}(t) - u(t)$ называется шумом квантования. Значение этого шума можно уменьшить за счет увеличения частоты дискретизации и уменьшения шага квантования Δu . Однако чрезвычайно малый шаг квантования приводит к перегрузкам. Для устранения этого недостатка применяют адаптивную дельта-модуляцию, при которой шаг квантования изменяется в зависимости от текущего значения передаваемого сигнала.

В заключение можно отметить следующее. Сигнал ИКМ имеет большую частоту взятия отсчетов по сравнению с сигналом ДМ. При этом на одну выборку при ДМ приходится один передаваемый импульс, а при ИКМ — несколько, число которых зависит от числа уровней квантования.

Анализ показывает, что при одинаковой достоверности приема при обоих видах модуляции частота следования импульсов примерно одинаковая. Следовательно, обе указанные системы сигналов занимают примерно одинаковую полосу частот. При этом системы с ДМ оказываются значительно проще в реализации. Отметим также, что систему передачи информации с дельта-

модуляцией предложил в 1949 г. наш соотечественник инженер Л. А. Коробков. За границей она появилась только в 1952 г.

Контрольные вопросы

1. Что представляют собой сигналы с дискретной амплитудной модуляцией?
2. Поясните сущность дискретной фазовой модуляции и в чем проявляется особенность спектра.
3. Поясните, как формируются сигналы с квадратурной амплитудной модуляцией и что такое сигнальное созвездие.
4. Какими преимуществами обладают сигналы с дискретной частотной модуляцией с непрерывной фазой и как они формируются?
5. Что представляют собой сигналы с минимальным частотным сдвигом, в чем смысл прямого и квадратурного методов их формирования и как можно обрабатывать такие сигналы?
6. Поясните сущность алгоритма Витерби и расскажите, где он применяется.
7. Дайте сравнительную оценку импульсно-кодовой модуляции и дельта-модуляции и поясните, какие преимущества дает учет корреляционных связей дискретных выборок аналоговых сигналов.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ СЛУЧАЙНЫХ СИГНАЛОВ И ПОМЕХ**6.1. Методы задания случайных процессов**

Передача сообщений по каналам связи осуществляется во времени, поэтому и сигналы $S_i(t)$ (где $i = 1, 2, \dots$), посредством которых передается информация, являются временными функциями. Однако если бы эти функции $S_i(t)$ были все время неизменными (детерминированными), т. е. заранее известными, то они не могли бы использоваться для передачи сообщений. Только случайные функции или случайные процессы пригодны для решения этой задачи.

Под *случайным (стохастическим) процессом* (сигналом) $S(t)$ понимают множество временных функций $S_i(t)$ (где $i = 1, 2, \dots$), которые в любой момент времени могут принимать различные значения с заданным распределением вероятностей. Функции $S_i(t)$ называются реализациями случайного процесса, а совокупность возможных реализаций — ансамблем. В дальнейшем под сигналом $S(t)$ будем понимать напряжение или ток, развиваемый на некоторой нагрузке, а интервал времени наблюдения сигнала (процесса) обозначим T .

Реализации не являются случайными функциями. Тем не менее предсказать, какова будет реализация процесса в каждом конкретном опыте, невозможно. Факт наличия в случайном сигнале различных реализаций позволяет ему быть переносчиком информации, следовательно, необходимо установить соответствия между сообщениями и различными реализациями сигнала. Тогда при приеме по принятой реализации можно было бы абсолютно точно восстановить переданное сообщение. В действительности это невозможно, поскольку в любом канале связи присутствуют помехи, которые также являются случайными процессами, а следовательно, разрушителями информации. Поэтому восстановление переданных сообщений можно осуществлять только с некоторой вероятностью.

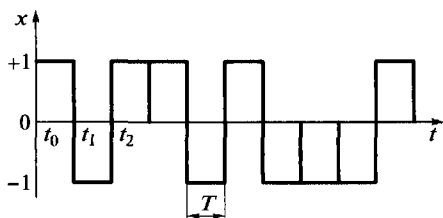


Рис. 6.1. Случайный телеграфный сигнал

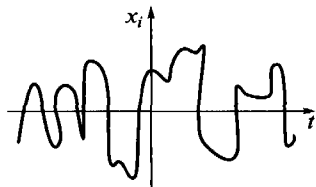


Рис. 6.2. Реализация случайного процесса

Случайные процессы в некоторых случаях могут задаваться на дискретном множестве значений моментов времени: t_1, t_2, \dots . Эти процессы являются дискретными во времени случайными последовательностями. Пример такой последовательности показан на рис. 6.1, из которого видно, что данный процесс $X(t_k)$ задается таким образом, что с вероятностью P он принимает значение $+1$, а с вероятностью $1 - P$ — значение -1 , которые остаются постоянными в интервалах времени $t \in [(k - 1)T, kT]$, где $k = 0, 1, 2, \dots$

Значительно чаще случайные процессы задаются непрерывно во времени по всей оси $-\infty < t < +\infty$ (рис. 6.2).

Реальные сигналы бывают, как правило, ограничены по времени и задаются в интервале $t \in [0, T]$, поэтому их называют финитными. Число реализаций случайного процесса может быть конечным, счетным (бесконечным, но при этом все реализации могут быть пронумерованы) и несчетным.

Случайный процесс задается с помощью совокупности его реализаций (рис. 6.3).

Любой случайный процесс можно определить лишь в вероятностном смысле. Поэтому говорят, что случайный процесс задан, если указано правило нахождения его вероятностных характеристик для любого момента времени в пределах области определения. При счетном или конечном числе реализаций их можно просто перечислить и указать вероятности появлений.

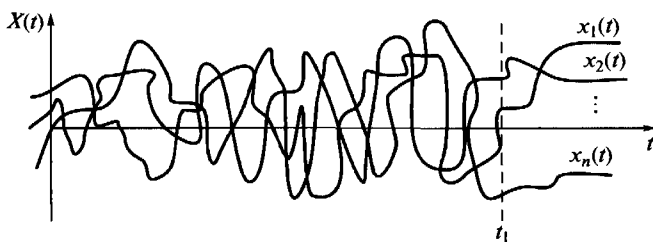


Рис. 6.3. Задание случайного процесса совокупностью его реализаций

Например, пусть случайный сигнал $S(t)$, заданный в интервале времени $t \in [0, T]$, имеет реализации

$$S_0(t) = U_c \cos \omega_0 t;$$

$$S_1(t) = U_c \cos \omega_1 t$$

с вероятностями $P(S_0) = P_0$ и $P(S_1) = P_1 = 1 - P_0$, где U_c , ω_0 , ω_1 , P_0 — некоторые постоянные величины. Нетрудно увидеть, что это простейший пример случайного процесса с конечным числом реализаций, т. е. сигнала, информационным параметром которого является частота ω . Следовательно, это сигнал с частотной манипуляцией.

Рассмотрим некоторый случайный процесс $X(t)$. Зафиксировав момент времени t_1 , получим сечения случайного процесса $X(t_1) = x_1$, представляющие собой случайную величину (непрерывную или дискретную).

Пусть задан некоторый уровень x , тогда можно определить вероятность того, что в момент времени $t = t_1$ значения случайного процесса не превысят значения x :

$$F_1(x, t) = P\{X(t) \leq x\}.$$

Данная функция называется одномерной интегральной функцией распределения (ИФР) вероятностей случайного процесса. Если ИФР является дифференцируемой функцией, то, взяв от нее производную, можно получить одномерную плотность вероятностей (ПВ):

$$\omega_1(x, t) = \frac{\partial F_1(x, t)}{\partial x}.$$

Для случайных непрерывных процессов ПВ определяется отношением

$$\omega_1(x, t) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{|\Delta x|} \geq 0,$$

в числителе которого находится вероятность того, что случайная величина $X(t)$ принимает значение в интервале $x \leq X \leq x + \Delta x$.

Очевидно также, что

$$F_1(x, t) = \int_{-\infty}^x \omega_1(x_1, t) dx_1.$$

На рис. 6.4 показаны типовые графики одномерных плотности вероятностей и интегральной функции распределения.

Найденные функции являются простейшими характеристиками случайного процесса, определяющие этот процесс только в отдельные фиксированные моменты времени.

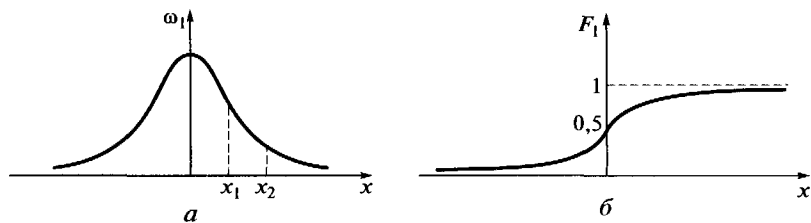


Рис. 6.4. Графики одномерной плотности вероятностей (а) и одномерной интегральной функции распределения вероятностей (б)

Более полно случайный процесс характеризует двухмерная интегральная функция распределения $F_2(x_1, t_1; x_2, t_2)$, определяющая вероятность того, что в момент времени t_1 функция $X(t_1)$ будет не больше x_1 , а в момент времени t_2 функция $X(t_2)$ не превысит уровень x_2 :

$$F_2(x_1, t_1; x_2, t_2) = P\{X(t_1) \leq x_1; X(t_2) \leq x_2\}.$$

При условии существования частной производной второго порядка можно найти двухмерную плотность распределения вероятностей:

$$\frac{\partial^2 F_2(x_1, t_1; x_2, t_2)}{\partial x_1 \partial x_2} = \omega_2(x_1, t_1; x_2, t_2).$$

В общем случае определяют n -мерную ИФР:

$$F_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\},$$

которая выражает свойства случайного процесса в произвольных сечениях.

Аналогично определяется n -мерная ПВ:

$$\omega_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{\partial^n F_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{\partial x_1 \dots \partial x_n}.$$

В частном случае, если значения случайного процесса $X(t)$ в любой момент времени являются независимыми, то n -мерная плотность вероятностей примет вид

$$\omega_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \prod_{k=1}^n \omega_1(x_k, t_k).$$

Нахождение n -мерных ПВ и ИФР — весьма трудная задача, которую удается решить далеко не всегда, поэтому для большинства практических задач гораздо удобнее использовать более простые и наглядные числовые характеристики случайного процесса.

Математическое ожидание (момент первого порядка)

$$M\{X(t)\} = \int_{-\infty}^{\infty} x\omega_1(x, t)dx,$$

где $\omega_1(x, t)$ — одномерная плотность распределения вероятностей случайного процесса.

Математическое ожидание является некоторой средней функцией, около которой группируются возможные реализации случайного процесса.

Дисперсия (центральный момент второго порядка)

$$D\{X(t)\} = M\{X^2(t)\} - [M\{X(t)\}]^2 = \int_{-\infty}^{\infty} [x - M\{X(t)\}]^2 \omega_1(x, t)dx.$$

Дисперсия также является функцией времени и характеризует меру разброса реализаций относительно среднего значения случайного процесса, т. е. вокруг его математического ожидания.

Математическое ожидание и дисперсия характеризуют случайный процесс только в отдельные моменты времени. Поэтому применяют также функцию корреляции, которая характеризует статистическую связь между сечениями $X(t_1)$ и $X(t_2)$ случайного процесса в моменты времени t_1 и t_2 и определяет среднюю скорость его изменения во времени.

Функцию корреляции определяют как математическое ожидание произведения двух сечений центрированного случайного процесса (центрированным процессом называется разность между случайным процессом и его математическим ожиданием):

$$B(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - M\{X(t_1)\}][x_2 - M\{X(t_2)\}] \omega_2(x_1, x_2; t_1, t_2) dx_1 dx_2.$$

Данная функция называется автокорреляционной. Можно также определить взаимную корреляционную функцию двух случайных процессов $X(t)$ и $Y(t)$:

$$B_{XY}(t_1, t_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - M\{X(t_1)\}][y_2 - M\{Y(t_2)\}] \times \\ \times \omega_2(x_1, y_2; t_1, t_2) dx_1 dy_2,$$

где $\omega_2(x_1, y_2; t_1, t_2)$ — совместная плотность вероятностей сечений процессов $X(t)$ и $Y(t)$ соответственно в моменты t_1 и t_2 .

В ряде случаев рассматривают нормированную корреляционную функцию

$$R(t_1, t_2) = \frac{B(t_1, t_2)}{[D\{X(t_1)\}D\{X(t_2)\}]^{\frac{1}{2}}}.$$

Если при увеличении интервала времени $\tau = t_2 - t_1$ функция корреляции убывает, это свидетельствует об ослаблении статистической связи между мгновенными значениями случайного процесса. В случае если сечения $X(t_1)$ и $X(t_2)$ независимы, функция корреляции между этими сечениями равна нулю.

6.2. Стационарность и эргодичность случайных процессов

При передаче информации по каналам связи условия приема в течение длительного времени могут оставаться постоянными. Неизменность этих условий порождает случайные процессы, протекающие однородно во времени, т.е. такие процессы, свойства которых сохраняются, в том числе при изменении начала отсчета времени. Вместе с тем в ряде случаев эти условия могут не выполняться. В связи с этим в теории связи случайные процессы подразделяются на стационарные и нестационарные.

Строго стационарным, т.е. *стационарным в узком смысле*, называется случайный процесс, плотность распределения вероятностей которого не изменяется при произвольном сдвиге во времени начала отсчета, т.е. для любых значений $-\infty < t < \infty$:

$$\omega_n(x_1, \dots, x_n; t_1, \dots, t_n) = \omega_n(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau),$$

где τ — произвольный интервал времени.

Отсюда следует независимость (инвариантность) числовых характеристик или моментных функций такого случайного процесса: математического ожидания и дисперсии от сдвига во времени.

Стационарным в широком смысле называется процесс, математическое ожидание и дисперсия которого не зависят от времени, а функция корреляции зависит только от значения τ . Из данного определения следует, что функция корреляции такого процесса — это четная функция от τ , т.е. $B(\tau) = B(-\tau)$.

Кроме того, можно показать, что $B(\tau) \leq B(0) = \sigma^2$, где σ^2 — дисперсия стационарного процесса. Для этого следует найти математическое ожидание квадрата разности значений стационарного процесса $X(t)$ в моменты времени t и $t + \tau$:

$$\begin{aligned} M\{[X(t) - X(t + \tau)]^2\} &= M\{X^2(t) + X^2(t + \tau) - 2X(t)X(t + \tau)\} = \\ &= M\{X^2(t)\} + M\{X^2(t + \tau)\} - 2M\{X(t)X(t + \tau)\} = \\ &= 2[D\{X(t)\} - B(\tau)]. \end{aligned}$$

Поскольку квадрат любого числа всегда не отрицательный, т.е. выражение $2[D\{X(t)\} - B(\tau)] \geq 0$, то можно утверждать, что дисперсия всегда положительная, т.е. $D\{X(t)\} \geq 0$.

Очевидно, что случайный процесс стационарный в узком смысле является стационарным и в широком смысле. Однако процесс стационарный в широком смысле не является стационарным в узком смысле.

Для случайных процессов стационарных в широком смысле введем важное свойство — *эргодичность*, определяющее, что усреднение одной реализации $x(t)$ процесса по времени при условии, что это время T достаточно велико, приводит примерно к такому же результату, как усреднение реализаций случайного процесса $X(t)$ по множеству (ансамблю). В этом случае математическое ожидание случайного процесса характеризует постоянную составляющую, а дисперсия — мощность переменной составляющей.

Свойство эргодичности имеет важное практическое значение, так как позволяет при изучении статистических свойств случайного процесса рассматривать только одну реализацию, но за достаточно длинный промежуток времени. Например, при изучении характеристик шумов на выходах одинаковых усилителей можно исследовать выходной шум только одного из них.

При эргодическом процессе справедливы следующие соотношения для математического ожидания, дисперсии и корреляционной функции:

$$M\{X(t)\} = \frac{1}{T} \int_0^T X(t) dt;$$

$$D\{X(t)\} = \frac{1}{T} \int_0^T \left[X(t) - \frac{1}{T} \int_0^T X(t) dt \right]^2 dt;$$

$$B(\tau) = \frac{1}{T} \int_0^T X(t) X(t + \tau) dt.$$

Корреляционная функция $B(\tau)$ характеризует степень статистических связей между двумя сечениями случайного процесса, разделенными интервалом τ . Поэтому с увеличением τ связи ослабевают и в пределе $\tau \rightarrow \infty$ сечения становятся статистически независимыми. В этом случае корреляционная функция $B(\tau) \rightarrow 0$.

Рассмотрим нормированную функцию корреляции (коэффициент корреляции)

$$R(\tau) = \frac{B(\tau)}{D\{X(t)\}}.$$

Для нее справедливы следующие соотношения:

$$R(\tau) = R(-\tau); R(0) = 1; |R(\tau)| \leq 1.$$

Введем также два определения интервала корреляции.

1. Интервал корреляции определяется как промежуток времени, при превышении которого ($\tau > \tau_k$) коэффициент корреляции $R(\tau)$ становится менее заданного значения (0,05 ... 0,10).

2. Интервал корреляции $\tau_k = \int_0^{\infty} |R(\tau)| d\tau$ определяется как основание прямоугольника высотой $R(\tau) = 1$ и с площадью, равной площади, ограниченной кривой $|R(\tau)|$ и осями координат (рис. 6.5).

В заключение найдем моментные функции случайного процесса, например случайного телеграфного сигнала $X(t)$, показанного на рис. 6.1.

Математическое ожидание и дисперсия этого сигнала определяются достаточно просто:

$$M\{X(t)\} = (+1)\frac{1}{2} + (-1)\frac{1}{2} = 0;$$

$$D\{X(t)\} = (+1)^2\frac{1}{2} + (-1)^2\frac{1}{2} = 1.$$

Для нахождения функции корреляции рассмотрим сечения, взятые в моменты времени t_i и t_{i+1} . Введем обозначение $\tau = t_{i+1} - t_i$ и определим математическое ожидание произведения $X(t_i)X(t_{i+1})$. Если $\tau > T$, где T — длительность одного импульса, то сечения могут попадать в интервалы времени с различными значениями их номеров k и тогда $R(\tau) = 0$, так как в этом случае с равными вероятностями произведение $X(t_i)X(t_{i+1})$ равно ± 1 и математическое ожидание равно 0.

При $\tau < T$ интересен случай попадания обоих сечений интервалы времени с одинаковыми номерами. (Случай попадания сечений в разные интервалы, который также может иметь место, рассматривался ранее.) При $\tau < T$ математическое ожидание произ-

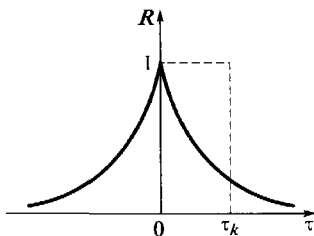


Рис. 6.5. Определение интервала корреляции методом равномерного прямоугольника

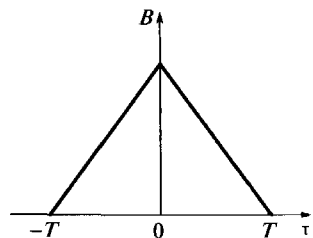


Рис. 6.6. Функция корреляции случайного телеграфного сигнала

ведения $X(t_i)X(t_{i+1})$ равно вероятности того, что оба сечения находятся в одном интервале, что имеет место, если одно из сечений отстоит от момента смены полярностей не более чем на $T - |\tau|$. Вероятность этого события $1 - |\tau|/T$. Таким образом,

$$B(\tau) = \begin{cases} 1 - \frac{|\tau|}{T}, & |\tau| \leq T; \\ 0, & |\tau| > T. \end{cases}$$

Данное соотношение иллюстрирует график, приведенный на рис. 6.6.

6.3. Спектральные характеристики случайных процессов

Для изучения спектров детерминированных сигналов в теории связи широко применяется аппарат преобразования Фурье, в соответствии с которым любой сигнал $S(t)$ можно определить через комплексную спектральную плотность (КСП) частоты $S(j\omega)$:

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

при этом КСП

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

Однако вероятностный характер случайных процессов не позволяет использовать КСП для их спектрального анализа, так как эта характеристика одной конкретной реализации сигнала, которая не описывает весь случайный сигнал, т.е. весь его ансамбль в целом.

Значительно более удобной характеристикой случайного сигнала является спектральная плотность мощности. Поэтому рассмотрим ее подробнее. Будем по-прежнему считать, что случайный сигнал является эргодическим. Тогда, используя равенство Парсеваля, определим спектральную плотность мощности, называемую также энергетическим спектром, с помощью следующего соотношения для энергии случайного сигнала $S(t)$:

$$E = \int_{-\infty}^{\infty} S^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(j\omega)|^2 d\omega.$$

Определим среднюю мощность сигнала при $T \rightarrow \infty$:

$$P = \lim_{T \rightarrow \infty} \frac{E}{T} = \lim_{T \rightarrow \infty} \left(\frac{1}{2\pi} \right) \int_{-\infty}^{\infty} \frac{|S(j\omega)|^2}{T} d\omega.$$

Необходимость перехода к пределу объясняется тем, что в противном случае мощность процесса будет зависеть от длительности его реализаций. Нас же интересуют реализации процесса в среднем.

Функция вида

$$G(\omega) = \lim_{T \rightarrow \infty} \frac{|S(i\omega)|^2}{T}$$

называется спектральной плотностью мощности (СПМ). Данное определение СПМ случайного процесса справедливо только для эргодических процессов, так как оно характеризует распределение мощности по частоте только одной реализации. Чтобы определить СПМ для совокупности реализаций, необходимо провести усреднение по ансамблю всех возможных реализаций случайного процесса.

Физический смысл спектральной плотности мощности заключается в том, что она является неслучайной функцией частоты и определяет распределение энергии случайного процесса по оси частот. При этом не учитываются индивидуальные свойства процесса такие, как фазовые соотношения между реализациями.

Спектральную плотность мощности, или энергетический спектр, можно определить и с помощью теоремы Винера-Хинчина, согласно которой она является преобразованием Фурье от корреляционной функции:

$$G(\omega) = \int_{-\infty}^{\infty} B(\tau) e^{-j\omega\tau} d\tau.$$

На основании обратного преобразования Фурье запишем

$$B(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{j\omega\tau} d\omega.$$

Приняв $\tau = 0$, найдем среднюю мощность, или дисперсию, случайного процесса:

$$\sigma^2 = B(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) d\omega = \int_{-\infty}^{\infty} G(f) df,$$

где $\omega = 2\pi f$.

Учитывая свойства четности корреляционной функции, запишем

$$G(f) = 2 \int_0^{\infty} B(\tau) \cos(2\pi f \tau) d\tau;$$

$$B(\tau) = 2 \int_0^{\infty} G(f) \cos(2\pi f \tau) df.$$

Это другая запись соотношений Винера-Хинчина, которые следуют из соответствующих равенств:

$$e^{j\omega\tau} = \cos \omega\tau + j \sin \omega\tau;$$

$$e^{-j\omega\tau} = \cos \omega\tau - j \sin \omega\tau.$$

Отрицательных частот не существует, поэтому вводят одностороннюю спектральную плотность $G_0(f)$, справедливую для положительных частот, причем

$$G_0(f) = \begin{cases} 2G(f) & \text{для } f \geq 0; \\ 0 & \text{для } f < 0. \end{cases}$$

Подставив $G(f)$ из этого выражения в соотношения Винера-Хинчина, получим

$$G_0(f) = 4 \int_0^{\infty} B(\tau) \cos(2\pi f \tau) d\tau;$$

$$B(\tau) = \int_0^{\infty} G_0(2\pi f) \cos(2\pi f \tau) d\tau.$$

Эти соотношения устанавливают связь между $B(\tau)$ и $G_0(f)$.

В практических приложениях различные случайные процессы часто сравниваются между собой по эффективной ширине энергетического спектра, которая определяется отношением площади под кривой энергетического спектра к максимальному значению спектральной плотности мощности:

$$\Delta\omega = 2\pi F_s = \frac{1}{G_{\max}(\omega)} \int_0^{\infty} G(\omega) d\omega.$$

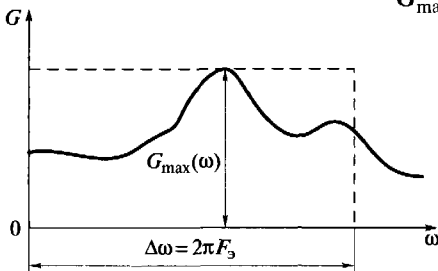


Рис. 6.7. Определение эффективной ширины энергетического спектра методом равновеликого прямоугольника

Определение эффективной ширины спектра поясняет рис. 6.7.

По аналогии с интервалом корреляции, который был определен ранее, эффективную ширину энергетического спектра можно трактовать как основание эквивалентного прямоугольника с высотой $G_{\max}(\omega)$ и площадью, равной площади фигуры, ограниченной кривой СПМ и осями координат.

Так как энергетический спектр и функция корреляции случайного процесса являются парой преобразований Фурье, эффективная ширина энергетического спектра и интервал корреляции (ширина $B(\tau)$) связаны следующим соотношением:

$$F_3 \tau_k \approx 1,$$

откуда видно, что чем «шире» спектр случайного процесса, тем «уже» его функция корреляции, и наоборот.

Рассмотрим теперь несколько примеров.

Пример 6.1. Найдем энергетический спектр случайного телеграфного сигнала, корреляционная функция которого была определена ранее:

$$\begin{aligned} G(f) &= 2 \int_0^{\infty} B(\tau) \cos(2\pi f \tau) d\tau = 2 \int_0^T \left(1 - \frac{\tau}{T}\right) \cos(2\pi f \tau) d\tau = \\ &= 2 \int_0^T \cos(2\pi f \tau) d\tau - \frac{2}{T} \int_0^T \tau \cos(2\pi f \tau) d\tau = \frac{2 \sin(2\pi f T)}{2\pi f} - \\ &\quad - \frac{2}{T} \left[\frac{\cos(2\pi f \tau)}{(2\pi f)^2} \Big|_0^T + \frac{\tau \sin(2\pi f \tau)}{2\pi f} \Big|_0^T \right] = \frac{\sin(2\pi f T)}{\pi f} - \\ &\quad - \frac{2}{T (2\pi f)^2} (1 - \cos(2\pi f T)) - \frac{\sin(2\pi f T)}{\pi f} = T \frac{\sin^2(\pi f T)}{(\pi f T)^2}. \end{aligned}$$

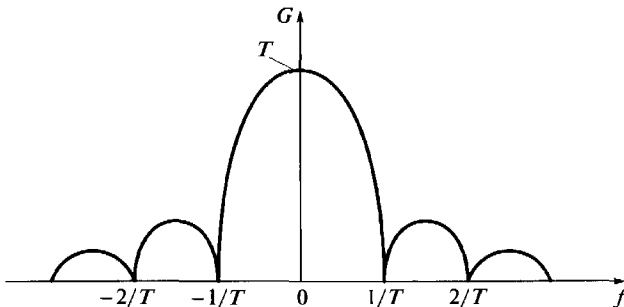


Рис. 6.8. Энергетический спектр случайного телеграфного сигнала

Здесь был использован табличный интеграл

$$\int x \cos(ax) dx = \frac{1}{a^2} \cos ax + \frac{x}{a} \sin ax.$$

Полученный результат иллюстрирует рис. 6.8, из которого видно, что спектр является непрерывной функцией частоты. При этом значительная часть энергии сигнала сосредоточена в области частот $|f| < 1/T$. За пределами этой области энергетический спектр везде, кроме частот, кратных значению $1/T$, также отличен от нуля, хотя доля энергии там весьма мала.

Пример 6.2. *Белым шумом* называется случайный процесс $n(t)$, у которого спектральная плотность мощности равномерна в бесконечной полосе частот и равна $N_0/2$.

Так как СПМ (или энергетический спектр) является четной функцией и ее можно задавать только для положительных частот, то N_0 будем называть односторонней спектральной плотностью белого шума (рис. 6.9).

Функция корреляции белого шума имеет вид

$$B(\tau) = \frac{N_0}{2} \delta(\tau),$$

где $\delta(\tau)$ — дельта-функция, обладающая следующими свойствами:

$$\delta(\tau) = \begin{cases} \infty, & \tau = 0; \\ 0, & \tau \neq 0; \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(\tau) d\tau = 1.$$

Корреляционная функция белого шума показана на рис. 6.10.

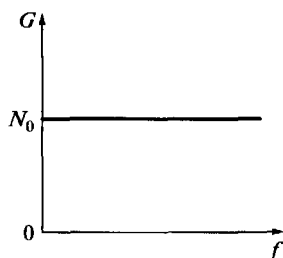


Рис. 6.9. Односторонняя спектральная плотность белого шума

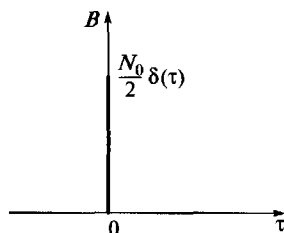


Рис. 6.10. Функция корреляции белого шума

Определение корреляционной функции $B(\tau)$ белого шума и дельта-функции означает статистическую независимость (некоррелированность) его сечений в любые, сколь угодно близкие моменты времени. Следовательно, белый шум можно называть абсолютно случайным процессом. По существу он является аналогом белого света в оптике, а также математической идеализацией.

Тем не менее на практике довольно часто встречаются случаи, когда реальный спектр помех можно аппроксимировать такой моделью. Примером подобной помехи может служить тепловой шум в резисторах или других проводниках. Спектральная плотность теплового шума является практически равномерной в полосе частот и примерно равной $6 \cdot 10^{12}$ Гц. Ясно, что эта полоса значительно шире, чем полосы пропускания реальных радиоприемников.

Пример 6.3. *Квазибелым шумом* называется случайный стационарный процесс с равномерной спектральной плотностью мощности в полосе, ограниченной сверху частотой F_B . Дисперсия квазирandomного шума

$$\sigma^2 = B(0) = N_0 F_B.$$

Функция корреляции определяется выражением

$$B(\tau) = \int_0^{F_B} N_0 \cos(2\pi f \tau) df = N_0 F_B \frac{\sin(2\pi F_B \tau)}{2\pi F_B \tau}.$$

График корреляционной функции квазирandomного шума показан на рис. 6.11, из которого видно, что $B(\tau)$ равна нулю при значениях τ , кратных $1/(2F_B)$. Иными словами, сечения процесса, разделенные интервалом времени $k/(2F_B)$, где k — целое число, не коррелированы между собой.

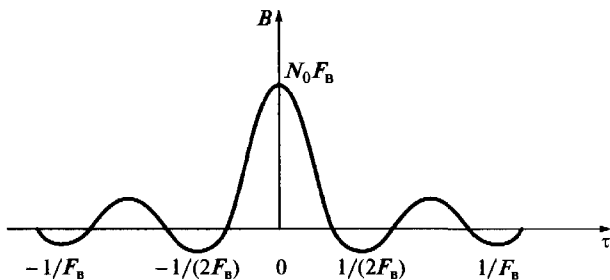


Рис. 6.11. Функция корреляции низкочастотного квазирandomного шума

6.4. Случайные процессы, определяемые одномерной и двухмерной плотностями вероятностей

Одним из примеров случайных процессов, которые можно полностью определить одномерной плотностью вероятности, является белый шум. Как уже отмечалось, белый шум — это абсолютно случайный процесс. Иными словами, в этом процессе значения x_1, x_2, \dots, x_n , взятые в моменты времени t_1, t_2, \dots, t_n , статистически независимы друг от друга, как бы близко они ни находились по отношению друг к другу. Следовательно, двухмерная плотность вероятностей белого шума равна произведению одномерных плотностей:

$$\omega(x_1, x_2, t_1, t_2) = \omega(x_1, t_1) \omega(x_2, t_2).$$

В общем случае для этого случайного процесса n -мерная плотность вероятностей определяется выражением

$$\omega(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \prod_{k=1}^n \omega(x_k, t_k),$$

т. е. зависит только от его одномерных плотностей.

Случайный гауссовский процесс для произвольного времени также можно задать одномерной плотностью вероятностей:

$$\omega_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right],$$

где σ^2 и m — дисперсия (мощность) случайного процесса и математическое ожидание, в общем случае зависящие от t .

Графическая зависимость данной плотности вероятностей показана на рис. 6.4.

Гауссовский процесс в ряде случаев можно описать двухмерной плотностью вероятностей:

$$\omega_2(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2\sigma_2^2(1-R^2)}} \exp\left\{-\frac{1}{2(1-R^2)} \left[\frac{(x_1-m_1)^2}{\sigma_1^2} - 2R \frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2} \right]\right\},$$

где σ_1^2, σ_2^2 и m_1, m_2 — соответственно дисперсии и математические ожидания сечений x_1 и x_2 случайного процесса; R — коэффициент корреляции между сечениями x_1 и x_2 .

В частном случае при $R = 0$ сечения x_1 и x_2 некоррелированы, и тогда

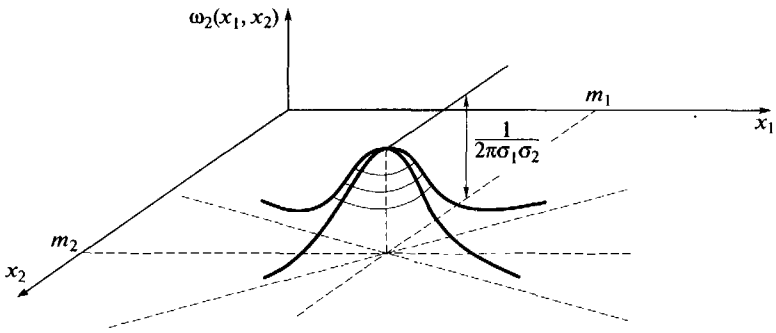


Рис. 6.12. Двухмерная плотность вероятностей случайного гауссовского процесса

$$\omega_2(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2\sigma_2^2}} \exp \left\{ -\frac{(x_1 - m_1)^2}{2\sigma_1^2} - \frac{(x_2 - m_2)^2}{2\sigma_2^2} \right\}.$$

Для наглядности на рис. 6.12 показана поверхность, представляющая собой двухмерную плотность вероятностей независимых случайных гауссовских величин с равными дисперсиями.

В самом общем случае можно получить n -мерную гауссовскую плотность распределения вероятностей.

Обратим внимание на следующие свойства гауссовских процессов.

1. Для гауссовского процесса понятия стационарности в широком и узком смысле совпадают, поскольку их плотность вероятностей не зависит от сдвига начала отсчета по оси времени.

2. В результате линейных преобразований гауссовского процесса (например, при фильтрации сигналов) получается также гауссовский процесс.

3. При выполнении нелинейных преобразований свойства гауссовского процесса утрачиваются. Например, при перемножении двух гауссовских процессов получается негауссовский процесс.

4. Если негауссовский случайный процесс с интервалом корреляции τ_k воздействует на инерционную линейную систему интегрирующего типа с постоянной времени τ_c , то процесс, получаемый на выходе этой системы, приближается к гауссовскому. Приближение тем точнее, чем сильнее становится неравенство $\tau_c \gg \tau_k$.

Отметим, что случайные гауссовские процессы играют исключительно важную роль в теории связи. Так, из центральной предельной теоремы, которая доказывается в курсе теории вероятностей, следует, что сумма большого числа разных случайных

процессов стремится к гауссовскому процессу, несмотря на то что их исходное распределение может существенно отличаться от гауссовского закона.

Например, процесс, называемый белым шумом, который обусловлен тепловым движением электронов, образующих ток, представляет собой сумму большого числа независимых случайных возмущений и согласно этой теореме будет гауссовским. Кроме того, гауссовское распределение обладает рядом замечательных свойств, отмеченных ранее, которые значительно упрощают многие сложные расчеты в теории связи. Благодаря этому теория связи и ее практические приложения получили значительное развитие.

В теории связи наряду с гауссовскими большую роль играют случайные марковские процессы, или процессы без последействия.

Марковским называется случайный процесс, условная плотность вероятностей которого в момент времени t_n зависит только от последнего значения λ_{n-1} в момент времени t_{n-1} и не зависит от более ранних значений:

$$\omega(\lambda_n, t_n | \lambda_1, \dots, \lambda_n; t_1, \dots, t_n) = \omega(\lambda_n, t_n | \lambda_{n-1}, t_{n-1}).$$

Условные плотности вероятностей марковских процессов называются также плотностями вероятностей перехода, или просто переходными плотностями.

Зафиксируем моменты времени $t_i < t_j < t_k$. Тогда в качестве определения марковского процесса можно принять следующее соотношение:

$$\begin{aligned} & P\{\lambda(t_i) \leq \lambda_i, \lambda(t_k) \leq \lambda_k | \lambda(t_j) = \lambda_j\} = \\ & = P\{\lambda(t_i) \leq \lambda_i | \lambda(t_j) = \lambda_j\} = P\{\lambda(t_k) \leq \lambda_k | \lambda(t_j) = \lambda_j\}. \end{aligned}$$

Данная запись, имеющая симметричный вид относительно времени, означает, что если зафиксировано состояние случайного процесса в текущий момент t_j , то будущее его состояние (при t_k) не зависит от прошлого состояния (при t_i). Поэтому марковские процессы называются процессами без последействия.

В общем виде n -мерную плотность вероятностей марковского процесса, дающую наиболее полное его описание, можно представить выражением

$$\omega_n(\lambda_1, \lambda_2, \dots, \lambda_n; t_1, t_2, \dots, t_n) = \omega(\lambda_1, t_1) \prod_{k=1}^{n-1} \omega(\lambda_{k+1}, t_{k+1} | \lambda_k, t_k).$$

Это означает, что любое n -мерное распределение марковского процесса определяется через начальную одномерную плотность $\omega(\lambda_1, t_1)$ и вероятности перехода.

Условная плотность вероятности непрерывных марковских процессов обладает следующими свойствами:

- она всегда положительная, т. е. $\omega(\lambda_{k+1}, t_{k+1} | \lambda_k, t_k) \geq 0$, и нормирована к единице, т. е.

$$\int_{-\infty}^{\infty} \omega(\lambda_{k+1}, t_{k+1} | \lambda_k, t_k) d\lambda_k = 1;$$

- если интервал между сечениями в моменты времени t_{k+1} и t_k уменьшается, то условная плотность стремится к дельта-функции, т. е.

$$\lim_{t_{k+1} \rightarrow t_k} \omega(\lambda_{k+1}, t_{k+1} | \lambda_k, t_k) = \delta(t_{k+1} - t_k);$$

- удовлетворяет уравнению Колмогорова — Чэпмена, называемого также уравнением Смоуловского.

Пусть имеется марковский процесс и выбраны его значения в моменты времени $t_0 < t' < t$ (рис. 6.13). Тогда можно записать уравнение, позволяющее находить плотность вероятности перехода при известном значении процесса в промежуточном сечении:

$$\omega(\lambda, t | \lambda_0, t_0) = \int_{-\infty}^{\infty} \omega(\lambda, t | \lambda', t') \omega(\lambda', t' | \lambda_0, t_0) d\lambda'.$$

В теории марковских непрерывных процессов большое значение имеет так называемое уравнение Фоккера—Планка—Колмогорова, или диффузионное уравнение. Свое название оно получило в 1930-е гг., когда впервые использовалось при решении задач теплопроводности и диффузии.

Пусть задан момент времени $t_0 < t$. В этом случае условная вероятность перехода удовлетворяет дифференциальному уравнению в частных производных:

$$\begin{aligned} \frac{\partial \omega(\lambda, t | \lambda_0, t_0)}{\partial t} = & - \frac{\partial}{\partial \lambda} [k_1(\lambda, t) \omega(\lambda, t | \lambda_0, t_0)] + \\ & + \frac{1}{2} \frac{\partial^2}{\partial \lambda^2} [k_2(\lambda, t) \omega(\lambda, t | \lambda_0, t_0)], \end{aligned}$$

где $k_1(\lambda, t)$ — коэффициент сноса, характеризующий среднюю скорость изменения случайного процесса; $k_2(\lambda, t)$ — коэффициент диффузии, определяющий среднюю скорость изменения дисперсии случайного процесса.

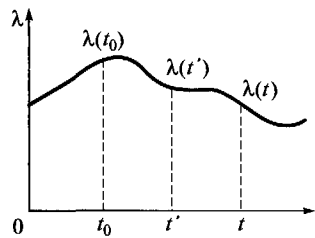


Рис. 6.13. Определение уравнения Колмогорова — Чэпмена для случайного марковского процесса

В зависимости от вида этих коэффициентов марковский процесс может иметь различное распределение вероятностей.

Данное уравнение справедливо при условии, что в начальный момент времени t_0 плотность вероятностей представляет собой дельта-функцию:

$$\omega(\lambda_0, t_0) = \delta(\lambda - \lambda_0).$$

В отдельных случаях марковский процесс одновременно может быть и гауссовским. Для этого необходимо и достаточно, чтобы коэффициент корреляции был экспоненциальным:

$$R(\tau) = e^{-\alpha|\tau|},$$

где α — некоторое постоянное значение.

В теории связи марковские процессы находят широкое применение. Так, они могут служить математической моделью сообщений, сигналов, помех. Кроме того, их применяют при решении ряда задач оптимальной фильтрации, т.е. их использование удобно при нестабильностях частоты и фазы колебаний, включая доплеровские сдвиги частоты, а также при случайных изменениях амплитуд сигналов в каналах связи.

6.5. Узкополосные случайные процессы

В радиотехнике и теории связи важную роль играют случайные узкополосные процессы. Такие процессы возникают при прохождении сигналов и помех через сравнительно узкополосные фильтры. Рассмотрим подробнее характеристики данных процессов.

Назовем узкополосным такой случайный процесс, у которого спектральная плотность отлична от нуля только в сравнительно узкой полосе частот $\Delta f = |f - f_0| \ll f_0$, где f_0 — средняя частота спектра.

Для описания этого процесса удобно применять его представление в квазигармонической форме:

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

где $A(t)$, $\varphi(t)$ — амплитуда (огibaющая) и начальная фаза, являющиеся случайными медленно изменяющимися по сравнению с $\cos\omega_0 t$ функциями; $\omega_0 = 2\pi f_0$.

Графическое изображение такого процесса имеет вид колебания, промодулированного по амплитуде и фазе.

Раскроем приведенное представление:

$$x(t) = A_c(t)\cos\omega_0 t - A_s(t)\sin\omega_0 t,$$

где $A_c(t) = A(t) \cos \varphi(t)$, $A_s(t) = A(t) \sin \varphi(t)$ — квадратурные компоненты сигнала, которые также являются случайными процессами, но медленно изменяются относительно периода несущей.

При таком представлении обработка сигнала (через обработку низкочастотных квадратурных составляющих) реализуется технически проще. Устройства для такой обработки сигнала будут рассмотрены далее.

Используем вновь понятие комплексного сигнала и запишем приведенное ранее представление в виде

$$\dot{x}(t) = x(t) + j\dot{x}(t) = \dot{A}(t)e^{j\omega_0 t}.$$

Здесь сопряженный процесс

$$\begin{aligned}\tilde{x}(t) &= \text{Im}[\dot{A}(t)e^{j\omega_0 t}] = A_c(t) \sin \omega_0 t + A_s(t) \cos \omega_0 t; \\ x(t) &= \text{Re}[\dot{A}(t)e^{j\omega_0 t}],\end{aligned}$$

т. е. действительный сигнал равен вещественной части своего комплексного представления.

Комплексная огибающая (амплитуда) сигнала — $\dot{A}(t) = A(t)e^{j\varphi(t)}$.

Огибающую и фазу узкополосного сигнала можно представить следующими выражениями:

$$\begin{aligned}A(t) &= |\dot{A}(t)| = \sqrt{x^2(t) + \tilde{x}^2} = \sqrt{A_c^2(t) + A_s^2(t)}; \\ \varphi(t) &= \text{arctg} \frac{A_s(t)}{A_c(t)}.\end{aligned}$$

Модуль комплексной огибающей является амплитудой действительного узкополосного процесса, а фаза комплексной огибающей совпадает с фазой действительного процесса.

Введем также понятие мгновенной частоты узкополосного процесса:

$$\begin{aligned}\omega(t) &= \omega_0 + \frac{d\varphi}{dt} = \omega_0 + \frac{d}{dt} \left[\text{arctg} \frac{A_s(t)}{A_c(t)} \right] = \\ &= \omega_0 + \frac{A_s'(t)A_c(t) - A_c'(t)A_s(t)}{A_c^2(t) + A_s^2(t)}.\end{aligned}$$

Заметим, что $\omega(t)$ является случайной функцией времени и не представляет собой круговую частоту колебания в привычном для всех понимании. Для квазигармонического представления случайного процесса мгновенная частота характеризует скорость изменения фазы.

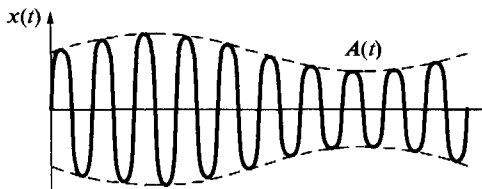


Рис. 6.14. Пример квазигармонического случайного процесса

На рис. 6.14 представлен характерный вид узкополосного процесса.

Отметим также, что случайный узкополосный процесс $x(t)$ и сопряженный с ним процесс $\tilde{x}(t)$ связаны друг с другом с помощью преобразований Гильберта:

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

В частном случае, как отмечалось ранее,

$$H[\cos \omega_0 t] = \cos\left(\omega_0 t - \frac{\pi}{2}\right) = \sin \omega_0 t;$$

$$H[\sin \omega_0 t] = \sin\left(\omega_0 t - \frac{\pi}{2}\right) = -\cos \omega_0 t.$$

Найдем теперь корреляционную функцию случайного узкополосного процесса. Для определенности рассмотрим процесс, спектральная плотность мощности которого имеет вид, показанный на рис. 6.15, т.е. $G_0(f) = N_0$ в пределах некоторой полосы частот:

$$\begin{aligned} B(\tau) &= \int_{f_0-F}^{f_0+F} G_0(f) \cos(2\pi f\tau) df = \frac{N_0}{2\pi\tau} \sin(2\pi f\tau) \Big|_{f_0-F}^{f_0+F} = \\ &= \frac{N_0}{2\pi\tau} [\sin 2\pi(f_0 + F)\tau - \sin 2\pi(f_0 - F)\tau]. \end{aligned}$$

Используя выражение

$$\sin \alpha - \sin \beta = 2 \cos\left(\frac{\alpha + \beta}{2}\right) \sin\left(\frac{\alpha - \beta}{2}\right),$$

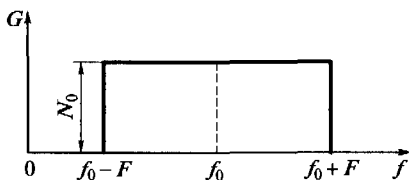


Рис. 6.15. Спектральная плотность мощности случайного узкополосного процесса (высокочастотного квазигармонического шума)

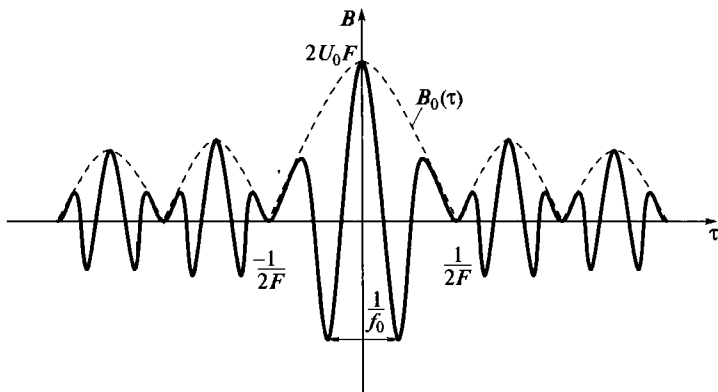


Рис. 6.16. Функция корреляции случайного узкополосного процесса (высокочастотного квазизелого шума)

получим

$$B(\tau) = 2N_0F \frac{\sin 2\pi F\tau}{2\pi F\tau} \cos 2\pi f_0\tau$$

или

$$B(t) = B_0(\tau) \cos 2\pi f_0\tau,$$

где $B_0(\tau)$ — медленно изменяющаяся по сравнению с $\cos 2\pi f_0\tau$ функция.

Графическое изображение функции $B(\tau)$ представлено на рис. 6.16.

Наконец, укажем два свойства квадратурных составляющих $A_c(t)$ и $A_s(t)$, которые, как уже отмечалось, также являются случайными процессами.

1. Если исходный узкополосный процесс $x(t)$ является гауссовским, то его квадратурные составляющие также подчиняются гауссовскому распределению.

Данное свойство можно доказать следующим образом. Для узкополосного гауссовского процесса вида

$$x(t) = A(t) \cos[\omega_0 t + \varphi(t)] = A_c(t) \cos(\omega_0 t) - A_s(t) \sin(\omega_0 t)$$

с помощью преобразования Гильберта, являющегося линейным, можно найти сопряженный узкополосный процесс:

$$\tilde{x}(t) = A(t) \sin[\omega_0 t + \varphi(t)] = A_c(t) \sin(\omega_0 t) + A_s(t) \cos(\omega_0 t).$$

Вследствие свойства линейности преобразования процесс $\tilde{x}(t)$ также будет гауссовским. Несложно теперь показать, что квадра-

турные составляющие связаны с $x(t)$ и $\bar{x}(t)$ с помощью линейных зависимостей:

$$\begin{aligned} A_c(t) &= x(t) \cos(\omega_0 t) + \bar{x}(t) \sin(\omega_0 t); \\ A_s(t) &= -x(t) \sin(\omega_0 t) + \bar{x}(t) \cos(\omega_0 t), \end{aligned}$$

поэтому $A_c(t)$ и $A_s(t)$ также имеют гауссовское распределение вероятностей.

2. Пусть имеется узкополосный гауссовский стационарный процесс $x(t)$ с нулевыми средними значениями и функцией корреляции

$$B(\tau) = \sigma^2 \rho(\tau) \cos 2\pi f_0 \tau,$$

где σ^2 — дисперсия процесса; $\rho(\tau)$ — медленно изменяющаяся по сравнению с $\cos 2\pi f_0 \tau$ функция.

Тогда составляющие $A_c(t)$ и $A_s(t)$ также являются гауссовскими стационарными процессами. При этом они независимы, их математические ожидания равны нулю и автокорреляционные функции имеют одинаковые значения $B(\tau) = \sigma^2 \rho(\tau)$. Отсюда следует и равенство их дисперсий, так как $D[x(t)] = B(0)$.

6.6. Огибающая и фаза квазигармонического случайного процесса

Определим плотность вероятностей огибающей $A(t)$ и фазы $\varphi(t)$ узкополосного случайного процесса. Для этого рассмотрим совместную плотность вероятностей $\omega(A_c, A_s)$ квадратурных составляющих $A_c(t)$ и $A_s(t)$ в одном сечении. Считая процесс гауссовским, а его квадратурные составляющие независимыми, запишем

$$\omega(A_c, A_s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{A_c^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{A_s^2}{2\sigma^2}} = \frac{1}{2\pi\sigma^2} e^{-\frac{A_c^2 + A_s^2}{2\sigma^2}}.$$

Найдем сначала интегральную функцию распределения рассматриваемого процесса, для чего проинтегрируем последнее выражение. Это удобно сделать, если перейти к полярной системе координат. Следовательно, можно записать

$$\begin{aligned} A_c(t) &= \rho(t) \cos \varphi(t); \quad A_s(t) = \rho(t) \sin \varphi(t); \\ dA_c(t) dA_s(t) &= \rho d\rho d\varphi. \end{aligned}$$

Тогда

$$F(A, \varphi) = \iint_V \frac{1}{2\pi\sigma^2} e^{-\frac{A_c^2 + A_s^2}{2\sigma^2}} dA_c dA_s =$$

$$= \int_0^A \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho d\rho \int_0^\varphi d\varphi = \left[1 - \exp\left(-\frac{A^2}{2\sigma^2}\right)\right] \frac{\varphi}{2\pi}.$$

Из полученного выражения видно, что двухмерная функция распределения оказалась равной произведению двух одномерных функций:

$$F(A) = 1 - \exp\left(-\frac{A^2}{2\sigma^2}\right), \quad A \geq 0;$$

$$F(\varphi) = \frac{\varphi}{2\pi}, \quad \varphi \in [0, 2\pi].$$

Это означает, что огибающая $A(t)$ и фаза $\varphi(t)$ случайного узкополосного процесса в одном сечении статистически независимы, при этом фаза является равномерно распределенной по всей области возможных значений от 0 до 2π .

Теперь определим плотность вероятностей огибающей как производную от соответствующей функции распределения:

$$\omega(A) = \frac{dF(A)}{dA} = \frac{A}{\sigma^2} e^{-\frac{A^2}{2\sigma^2}}, \quad A \geq 0$$

или

$$\omega(z) = z e^{-\frac{z^2}{2}},$$

где

$$z = \frac{A}{\sigma} \geq 0.$$

Таким образом получено известное распределение Рэлея, графическое изображение которого представлено на рис. 6.17.

Максимум кривой данного распределения, соответствующий значению огибающей $A = \sqrt{\sigma^2}$, составляет $\omega(A) = \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2}}$. Если

дисперсия процесса увеличивается, то максимум кривой сдвигается вправо и уменьшается. При этом в отличие от гауссовского распределения переменная здесь может принимать только положительные значения.

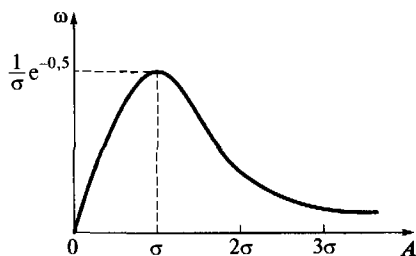


Рис. 6.17. Плотность вероятностей огибающей квазигармонического случайного гауссовского процесса

Математическое ожидание огибающей процесса

$$M\{A(t)\} = \int_0^{\infty} A\omega(A) dA = \sqrt{\sigma^2 \frac{\pi}{2}}.$$

Дисперсия, т. е. средняя мощность флуктуаций,

$$D\{A(t)\} = M\{A^2\} - (M\{A\})^2 = 2\sigma^2 - \frac{\pi}{2}\sigma^2 = \frac{4-\pi}{2}\sigma^2.$$

Аналогично определим плотность вероятностей фазы (рис. 6.18):

$$\omega(\varphi) = \frac{dF(\varphi)}{d\varphi} = \frac{1}{2\pi}, \quad \varphi \in [0, 2\pi],$$

откуда видно, что фаза квазигармонического процесса равномерно распределена во всей области ее определения.

Рассмотрим теперь более сложный случай, когда имеется сумма узкополосного гауссовского процесса и гармонического сигнала:

$$\begin{aligned} u(t) &= n(t) + s(t) = A(t) \cos[\omega_0 t + \varphi(t)] + U_c \cos \omega_0 t = \\ &= [A_c(t) + U_c] \cos \omega_0 t - A_s(t) \sin \omega_0 t. \end{aligned}$$

Здесь по-прежнему $A_c(t) = A(t) \cos \varphi(t)$, а $A_s(t) = A(t) \sin \varphi(t)$.

Тогда

$$u(t) = \rho(t) \cos[\omega_0 t + \varphi(t)],$$

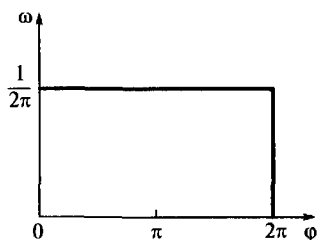


Рис. 6.18. Плотность вероятностей фазы квазигармонического случайного гауссовского процесса

где $\rho(t) = \sqrt{[A_c(t) + U_c]^2 + A_s^2(t)}$ — огибающая; $\varphi(t) = \arctg \frac{A_s(t)}{A_c(t) + U_c}$ — фаза смеси сигнала и шума.

Заметим, что процесс $u(t)$ в отличие от предыдущего случая не является стационарным, поскольку его математическое ожидание $M\{u(t)\} = U_c \cos \omega_0 t$ зависит от времени. Вместе с тем квадратурные составляющие $A_c(t)$ и $A_s(t)$ являются гауссовскими стационарными процессами. При этом $M\{A_c(t)\} = U_c$, $M\{A_s(t)\} = 0$, а дисперсии их равны σ^2 .

Введем обозначения $x(t) = A_c(t) + U_c$, $y(t) = A_s(t)$ и рассмотрим совместную плотность вероятностей квадратурных составляющих этого процесса:

$$\omega(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-U_c)^2 + y^2}{2\sigma^2}}.$$

Перейдем, как и ранее, к полярным координатам:

$$x = \rho \cos \varphi; \quad y = \rho \sin \varphi; \quad dx dy = \rho d\rho d\varphi.$$

Тогда совместную плотность можно записать в виде

$$\omega(\rho, \varphi) = \frac{\rho}{2\pi\sigma^2} \exp\left(-\frac{\rho^2 + U_c^2 - 2U_c \rho \cos \varphi}{2\sigma^2}\right).$$

Из этого выражения видно, что в отличие от предыдущего случая здесь огибающая и фаза совокупности регулярного сигнала и случайного узкополосного процесса зависят друг от друга, на что указывает величина $\rho \cos \varphi$.

Чтобы найти одномерную плотность распределения огибающей, проинтегрируем полученное выражение по всем значениям фазы, которая изменяется от 0 до 2π :

$$\omega(\rho) = \frac{\rho}{2\pi\sigma^2} e^{-\frac{\rho^2 + U_c^2}{2\sigma^2}} \int_0^{2\pi} e^{\frac{U_c \rho \cos \varphi}{\sigma^2}} d\varphi.$$

В этом выражении интеграл представляет собой модифицированную функцию Бесселя нулевого порядка:

$$I_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{z \cos \theta} d\theta,$$

для которой, как отмечалось ранее, имеются специальные справочные таблицы.

При $z = 0$ функция $I_0(z) = 1$, затем с увеличением $|z|$ она монотонно возрастает.

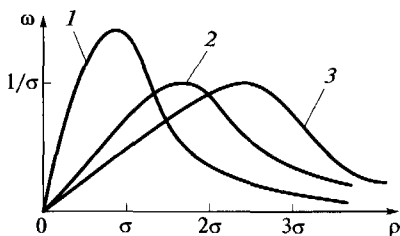


Рис. 6.19. Плотность распределения огибающей суммы гармонического сигнала и случайного узкополосного гауссовского процесса (распределение Райса)

1 — при $U_c = 0$; 2, 3 — при $U_c \neq 0$

Таким образом,

$$\omega(\rho) = \frac{\rho}{\sigma^2} e^{-\frac{\rho^2 + U_c^2}{2\sigma^2}} I_0\left(\frac{U_c \rho}{\sigma^2}\right).$$

Полученное распределение вероятностей называется распределением Райса, или обобщенным распределением Рэлея. Если в этом выражении принять $U_c = 0$, $I_0(0) = 1$, т.е. данное выражение будет представлять собой распределение Рэлея, рассмотренное ранее.

На рис. 6.19 показана плотность распределения огибающей для случаев, когда $U_c \neq 0$ и $U_c = 0$. Из рис. 6.19 видно, что с увеличением значения огибающей сигнала максимум кривой распределения будет смещаться вправо, т.е. в сторону больших значений ρ . При этом форма кривой приближается к кривой гауссовского распределения. В случае когда $U_c / \sqrt{\sigma^2} \gg 1$, распределение Райса в окрестности точки $\rho = U_c$, приблизительно совпадающей со средним значением огибающей, стремится к гауссовскому распределению с математическим ожиданием $m = U_c$ и дисперсией σ^2 .

Для определения плотности распределения фазы необходимо проинтегрировать совместную плотность $\omega(x, y)$ по всем значениям огибающей:

$$\omega(\varphi) = \int_0^{\infty} \omega(\rho, \varphi) \rho d\rho.$$

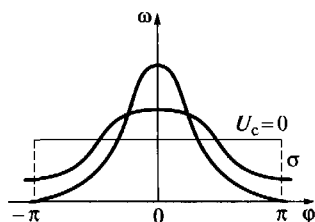
В результате получим

$$\omega(\varphi) = \frac{1}{2\pi} e^{-\frac{U_c^2}{2\sigma^2}} + \frac{U_c \cos \varphi}{2\sigma\sqrt{2\pi}} F\left(\frac{U_c \cos \varphi}{\sigma}\right) e^{-\frac{U_c^2}{2\sigma^2} \sin^2 \varphi},$$

$$0 \leq \varphi \leq 2\pi,$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ — интегральная функция гауссовского распределения.

Рис. 6.20. Семейство кривых распределения фаз суммы гармонического сигнала и случайного узкополосного гауссовского процесса



Если сигнал равен нулю ($U_c = 0$), то закон распределения фазы является равномерным:

$$\omega(\varphi) = 1 / (2\pi),$$

что соответствует предыдущему рассмотренному случаю с рэле-евским распределением.

На рис. 6.20 представлено семейство кривых распределения фаз. При большом отношении U_c/σ (сигнал/шум) фаза суммарных колебаний изменяется в небольших пределах относительно фазы сигнала. При уменьшении амплитуды сигнала (или увеличении уровня шумов) колебания фазы возрастают, приближаясь к предельному равновероятному (равномерному) распределению.

В заключение отметим, что здесь фаза сигнала была принята равной нулю. Однако в теории связи доказывается, что все приведенные рассуждения сохраняют свою справедливость и в том случае, если начальная фаза сигнала не равна нулю, а может принимать любое случайное значение.

6.7. Некоторые модели источников сообщений

В качестве модели источника дискретных сообщений, например данных на выходе ЭВМ, в теории связи принято использовать модель случайного телеграфного сигнала. В этом случае, как уже отмечалось, процесс с вероятностью $P = 1/2$ может принимать в дискретных точках, кратных T , значения ± 1 , которые сохраняются постоянными в интервале времени T . При этом значения случайного процесса в очередных интервалах принимаются независимо от того, какими они были ранее. Такой процесс можно отнести к марковским процессам, точнее к частному (вырожденному) случаю марковских цепей с двумя состояниями. Поскольку числовые характеристики и энергетический спектр случайного телеграфного сигнала уже были найдены ранее, вновь к этому возвращаться не будем.



Рис. 6.21. Динамическая система, формирующая модель речевого сигнала

Перейдем к модели телефонного, или речевого, сообщения. Под речевым сообщением принято понимать электрические колебания, которые появляются на выходе электроакустического преобразователя при воздействии на него акустической волны. Преобразователем в данном случае является микрофон вместе с фильтром. С их помощью формируется первичный телефонный сигнал в заданной полосе частот. Как правило, эта полоса, равная 3,1 кГц, располагается в пределах 300... 3400 Гц. Это, как уже отмечалось, стандартный телефонный канал, или канал тональной частоты (ТЧ).

Смысловое содержание акустической речевой волны априорно неизвестно, поэтому телефонное сообщение следует рассматривать как случайный процесс, реализации которого наблюдаются на выходе источника (электроакустического преобразователя). Для описания модели речевого сигнала в теории связи используют марковские процессы, формируемые с помощью динамической системы в виде последовательно соединенных RC -фильтров нижних и верхних частот (рис. 6.21).

Если на вход такой системы подается стационарный белый гауссовский шум $n(t)$ со спектральной плотностью N_0 , нулевым средним значением и функцией корреляции $\frac{N_0}{2} \delta(\tau)$, то случайный процесс $b(t)$ на выходе можно описать системой стохастических дифференциальных уравнений

$$\begin{aligned} \frac{dx_1}{dt} &= -\alpha_1 x_1(t) + \alpha_1 n(t); \\ \frac{dx_2}{dt} &= -\alpha_2 x_2(t) + k_0 [-\alpha_1 x_1(t) + \alpha_1 n(t)], \end{aligned}$$

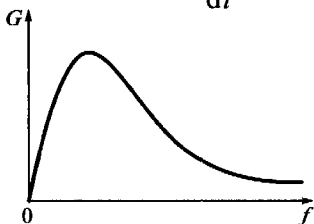


Рис. 6.22. Спектральная плотность мощности случайного процесса на выходе динамической системы

где $x_1(t)$ — напряжение на конденсаторе C_1 ; $x_2(t) = b(t)$ — напряжение на резисторе R_2 ; k_0 — коэффициент передачи устройства развязки; $\alpha_1 = 1/(R_1 C_1)$, $\alpha_2 = 1/(R_2 C_2)$.

В результате решения данной системы уравнений получим спектральную плотность мощности и функцию корреляции данного случайного процесса :

$$G(f) = \frac{\alpha_1^2 N_0 k_0 \omega^2}{2(\omega^2 + \alpha_1^2)(\omega^2 + \alpha_2^2)}; \quad \omega = 2\pi f;$$

$$B(\tau) = \frac{\alpha_1^2 N_0 k_0^2}{4(\alpha_1^2 - \alpha_2^2)} [\alpha_1 e^{-\alpha_1 |\tau|} - \alpha_2 e^{-\alpha_2 |\tau|}].$$

В частном случае при $\alpha_1 = \alpha_2$ график спектральной плотности мощности имеет вид, показанный на рис. 6.22.

Контрольные вопросы

1. Какие процессы называются случайными (стохастическими) и что понимается под реализацией (ансамблем реализаций) случайного процесса?
2. Как задаются случайные процессы?
3. Что понимается под математическим ожиданием, дисперсией и функцией корреляции случайного процесса?
4. Что представляют собой стационарность и эргодичность случайного процесса?
5. Как связаны между собой спектральная плотность мощности и функция корреляции?
6. Что понимается под белым шумом и чем от него отличается квазибелый шум?
7. Какие случайные процессы являются марковскими и какой смысл имеют коэффициенты сноса и диффузии, входящие в уравнение Колмогорова?
8. Как определяются огибающая, фаза, частота и квадратурные составляющие случайного узкополосного процесса?
9. Поясните законы распределения огибающей и фазы случайного квазигармонического процесса.

7.1. Классификация каналов

Напомним, что под каналом связи понимают совокупность технических средств и среды распространения, предназначенной для передачи сигналов электросвязи от одного абонента к другому в определенной полосе частот и с заданной скоростью. В курсе ТЭС в составе канала связи рассматривают последовательное соединение функциональных блоков, при котором обеспечивается передача информации с наибольшей эффективностью. В зависимости от решаемых задач в качестве канала телекоммуникаций понимают различную совокупность функциональных блоков или узлов (кодирующих устройств, модуляторов и т. д.).

Каналы телекоммуникаций обычно классифицируются по различным признакам.

В зависимости от видов передаваемых сообщений различают каналы телеграфной связи, факсимильной связи, телефонной связи, звукового вещания, видеотелефонной связи, передачи данных, телевидения и телеметрии.

По виду среды распространения сигналов различают следующие *каналы радиосвязи*: прямой радиовидимости, ионосферной радиосвязи, тропосферной радиосвязи, радиорелейной связи, метеорной радиосвязи и космической (спутниковой) связи, а также следующие каналы *проводной связи*: воздушные линии, кабельные линии, волоконно-оптические линии, каналы оптико-электронные и каналы гидроакустической связи.

По характеру преобразований сигналов в процессе передачи различают каналы линейные и нелинейные.

Весьма важным признаком классификации каналов телекоммуникаций является диапазон используемых в них частот. Например, по каналам, образованным с помощью симметричных кабельных линий, передаются сигналы в полосе, ограниченной сверху несколькими сотнями килогерц. В каналах с коаксиальными-

Таблица 7.1

Диапазон волн	Длина волны	Наименование диапазона частот	Частота колебаний
Мириаметровые (сверхдлинные — СДВ)	100 ... 10 км	Очень низкие частоты (ОНЧ)	3 ... 30 кГц
Километровые (длинные — ДВ)	10 ... 1 км	Низкие частоты (НЧ)	30 ... 300 кГц
Гектометровые (средние — СВ)	1000 ... 100 м	Средние частоты (СЧ)	300 ... 3000 кГц
Декаметровые (короткие — КВ)	100 ... 10 м	Высокие частоты (ВЧ)	3 ... 30 МГц
Ультракороткие: метровые (МВ)	10 ... 1 м	Очень высокие частоты (ОВЧ)	30 ... 300 МГц
дециметровые (ДМВ)	100 ... 10 см	Ультравысокие частоты (УВЧ)	300 ... 3000 МГц
сантиметровые (СМВ)	10 ... 1 см	Сверхвысокие частоты (СВЧ)	3 ... 30 ГГц
миллиметровые (ММВ)	10 ... 1 мм	Крайневысокие частоты (КВЧ)	30 ... 300 ГГц
децимиллиметровые (ДММВ)	1 ... 0,1 мм	Гипервысокие частоты (ГВЧ)	300 ... 3000 ГГц

ми кабелями, являющимися основой сетей магистральной связи, диапазон используемых частот достигает сотен мегагерц.

Наиболее широкий диапазон частот применяется в каналах радиосвязи, его границы начинаются примерно на частоте $3 \cdot 10^1$ Гц и простираются до $3 \cdot 10^{12}$ Гц. Свойства и характеристики каналов радиосвязи во многом определяются диапазоном радиоволн.

Разделение электромагнитных волн на диапазоны представлено в табл. 7.1.

В настоящее время, благодаря широкому внедрению квантовых генераторов и лазеров, интенсивно применяются каналы оптического диапазона, в котором передача информации осуществляется посредством световых волн. Здесь используются частоты порядка 10^{14} ГГц (длина волн около 1 мкм).

Рассмотрим простейшую структурную схему системы передачи информации, представленную на рис. 7.1.

В этой схеме для теории электрической связи представляет интерес классификация каналов телекоммуникаций по характеру сигналов на входе и выходе. Обычно различают:

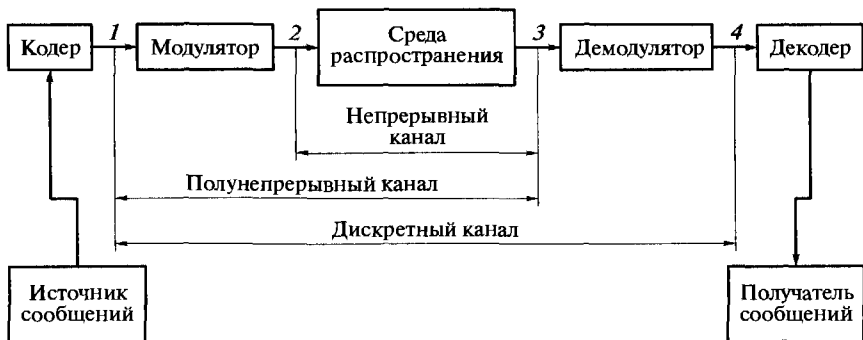


Рис. 7.1. Структурная схема системы передачи информации:
 1... 4 — точки схемы

- *непрерывные* каналы, на входе и выходе которых сигналы аналоговые (непрерывные по уровням). Это каналы между точками 2 и 3, т. е. между выходом модулятора и входом демодулятора;
- *дискретные* каналы, на выход которых поступают дискретные по уровням сигналы, и с выхода также снимаются дискретные сигналы. Это каналы между точками 1 и 4;
- *дискретно-непрерывные*, или полунепрерывные, каналы, т. е. дискретные каналы со стороны входа и непрерывные со стороны выхода между точками 1 и 3.

Из рис. 7.1 видно, что любые дискретные и полунепрерывные каналы включают в себя непрерывный канал, и свойства этого непрерывного канала существенно влияют на модели дискретного и полунепрерывного каналов. Очень часто простой математической модели непрерывного канала может соответствовать достаточно сложная модель дискретного канала.

Отметим также, что дискретность и непрерывность канала не связаны с видом передаваемых сообщений. В теории связи, как было показано ранее, имеются случаи передачи дискретных сообщений по непрерывному каналу, а также непрерывных сообщений по дискретному каналу.

7.2. Модели непрерывных каналов

7.2.1. Общие сведения

Любой канал связи можно задать с помощью математической модели, если известно множество сигналов на его входе, а также определен случайный процесс (сигнал) на его выходе.

Как уже отмечалось, случайный процесс также считается заданным, если известно его распределение вероятностей.

Однако дать точное математическое описание реальных каналов — весьма сложная задача. Поэтому, как правило, принимают несколько упрощенные модели, которые тем не менее учитывают все самое существенное, что влияет на процесс передачи информации. Второстепенные детали, не оказывающие значительного влияния, в этом случае отбрасываются.

7.2.2. Канал с аддитивным гауссовским шумом

Это одна из самых простейших моделей. Сигнал на выходе такого канала имеет вид

$$z(t) = \mu u(t - \tau) + n(t),$$

где $u(t)$ — входной сигнал; μ и τ — известные постоянные соответственно коэффициент передачи канала и время запаздывания сигнала; $n(t)$ — случайная аддитивная помеха, являющаяся гауссовским белым шумом с нулевым средним значением и односторонней спектральной плотностью $G_0(f) = N_0$.

Под $n(t)$ в ряде случаев понимают также квазибелый шум с равномерной спектральной плотностью мощности в полосе, ограниченной некоторой частотой.

Можно несколько усложнить данную модель, если коэффициент передачи μ и время запаздывания τ считать известными функциями времени:

$$z(t) = \mu(t) u [t - \tau(t)] + n(t).$$

Данная модель удовлетворительно описывает большинство проводных каналов, а также каналы радиосвязи прямой видимости.

7.2.3. Канал с неопределенной фазой сигнала и аддитивным шумом

Обратимся вновь к случайным узкополосным процессам и их квазигармоническому представлению. В этом случае сигнал на выходе канала можно представить выражением

$$z(t) = \mu[u(t) \cos \theta - \tilde{u}(t) \sin \theta] + n(t),$$

где $\tilde{u}(t)$ — сопряженный канал, получаемый посредством преобразования Гильберта от $u(t)$; $\theta = \omega_0 \tau$ — случайная фаза, как правило, распределенная равномерно в интервале от 0 до 2π .

Данная модель может применяться для описания тех же каналов, что и предыдущая, но при условии колебаний фазы. Такие колебания вызываются небольшими изменениями протяженности канала, свойств среды, через которую проходит сигнал, а также нестабильностью опорных (задающих) генераторов в передатчиках и приемниках.

7.2.4. Однолучевой гауссовский канал с флуктуациями амплитуд и фаз сигнала (с общими замираниями)

Данный канал можно описать тем же выражением, что и канал с определенной фазой сигнала и аддитивным шумом. Однако в этом случае множитель μ и фаза θ являются случайными процессами. Следовательно, случайными являются квадратурные составляющие:

$$X(t) = \mu(t) \cos \theta(t); \quad Y(t) = \mu(t) \sin \theta(t).$$

Тогда сигнал на выходе канала можно записать в виде

$$z(t) = \mu(t)[u(t) \cos \theta(t) - \hat{u}(t) \sin \theta(t)] + n(t).$$

Если считать квадратурные составляющие случайными гауссовскими процессами, то огибающая сигнала при прохождении по такому каналу будет иметь распределение Райса или Рэлея. При этом фаза может быть равномерно распределенной или иметь более сложный характер зависимости. Обычно ее считают в таких каналах равномерно распределенной в интервале от 0 до 2π .

Данные каналы получили название каналов с общими рэлеевскими или райсовскими замираниями, поскольку соотношения между всеми составляющими спектра сигнала не изменяются при передаче. Условием применимости такой модели является узкополосность сигналов.

Модель однолучевого канала с общими замираниями достаточно хорошо описывает многие каналы радиосвязи в различных диапазонах волн.

7.2.5. Многолучевой гауссовский канал с частотно-селективными замираниями

Данная модель канала является обобщением предыдущей модели и соответствует ситуации, когда в точку приема приходит несколько лучей сигнала, имеющих различное время τ_k задержки (запаздывания) k -го луча.

Тогда сигнал на выходе канала можно записать в виде

$$z(t) = \sum_{k=1}^n \mu_k(t) [u(t - \tau_k) \cos \theta_k(t) - \hat{u}(t - \tau_k) \sin \theta_k(t)] + n(t),$$

где n — число лучей в канале.

В рассматриваемой модели канала условие $\Delta t \ll 1/F_c$ обычно не выполняется. Здесь F_c — верхняя граничная частота спектра сигнала, а Δt — разность хода лучей (средняя разность времени распространения между лучами). Поэтому на разных участках спектра сигнала процессы замираний отличаются друг от друга, следовательно, могут наблюдаться значительные искажения формы сигнала.

Данная модель наряду с предыдущей широко применяется для отображения процессов передачи сигналов во многих реальных каналах радиосвязи.

7.2.6. Канал со сложной аддитивной помехой

Данный канал описывается любой из рассмотренных ранее моделей, но при этом дополнительно к ним добавляются помехи других видов. Иными словами, в таких каналах кроме шумовой гауссовской помехи с нулевым математическим ожиданием и спектральной плотностью $G(f) = N_0$ могут действовать сосредоточенные по спектру или времени помехи. Эти помехи являются аддитивными, поскольку они складываются с передаваемым сигналом в отличие от рассмотренных ранее замираний сигналов, называемых мультипликативной помехой.

Сосредоточенные по спектру помехи образуются сигналами посторонних радиостанций, излучением высокочастотных генераторов (промышленных, медицинских) или создаются преднамеренно в целях нарушения связи. В общем случае эти помехи являются узкополосными модулированными квазигармоническими колебаниями. Их распределение, как правило, такое же, как у полезных сигналов. Наибольшее влияние на качество связи эти помехи оказывают в диапазоне коротких (декаметровых) радиоволн.

К импульсным относятся помехи в виде коротких одиночных импульсов, промежутки времени между которыми значительно превышают время затухания сигнала от предыдущих посылок. Это атмосферные и индустриальные помехи, создаваемые различной электрической аппаратурой, например системой зажигания двигателя автомобиля, рентгеновскими установками, электрическими двигателями и др.

Вероятностное распределение этих помех описывается распределением амплитуд импульсов, которые подчиняются логарифмически нормальному закону:

$$\omega(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}},$$

где m и σ^2 — математическое ожидание и дисперсия рассматриваемого случайного процесса.

Для описания вероятностных характеристик интервалов времени между импульсами применяется распределение Пуассона, согласно которому случайная величина λ принимает бесконечное счетное множество возможных значений с вероятностями

$$P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

В наибольшей степени импульсные помехи оказывают воздействие на качество передачи информации в проводных каналах, а также в каналах радиосвязи в диапазоне сверхдлинных волн.

7.2.7. Канал с межсимвольной интерференцией

Данная модель соответствует случаю передачи информации с большой скоростью по каналу с ограниченной полосой пропускания. В связи с этим переходные процессы от предыдущих и последующих элементов сигнала способствуют появлению помех — межсимвольной интерференции (МСИ), приводящей к искажению текущего символа.

Данная модель широко применяется для описания процессов передачи информации в каналах проводной связи и радиосвязи. В радиоканалах причиной межсимвольной интерференции чаще всего является многолучевое распространение радиоволн. В каналах проводной связи межсимвольная интерференция вызывается нелинейностью фазочастотной характеристики канала и ограниченностью его полосы пропускания.

Например, пусть передается последовательность посылок с длительностью T , которые соответствуют цепочке из символов $b_{-Q}, b_{-Q-1}, \dots, b_{-1}, b_0, \dots, b_{D-1}, b_D$. Обозначим через $S_r(t)$ отклик линейного канала на посылку, соответствующую символу b_r . При этом $rT \leq t \leq (Q + r + 1)T$, где $Q = \left\lfloor \frac{\Delta\tau}{T} \right\rfloor$ — относительная память канала, равная целой части данного отношения; $\Delta\tau$ — время рассеяния (длительность переходного процесса в канале).

Тогда принимаемое колебание $z(t)$ в интервале анализа $T_a = (D + 1)T$, где $D \geq Q$ — задержка в числе символов принятия решения о передаваемом символе b_0 , будет иметь вид

$$z(t) = s_0(t) + g_{\text{МСИ}}(t) + n(t),$$

где $s_0(t)$ — сигнал, соответствующий передаваемому символу b_0 ; $n(t)$ — аддитивный гауссовский шум в канале.

Сигнал МСИ, образованный посылками, переданными до и после анализируемого сигнала

$$g_{\text{МСИ}}(t) = g_1(t) + g_2(t),$$

где $g_1(t)$ и $g_2(t)$ — посылки, переданные соответственно до и после анализируемого сигнала.

Сигнал, определяющий остаточный сигнал МСИ, образованный символами, переданными ранее символа b_0 ,

$$g_1(t) = \sum_{r=-Q}^{-1} S_r(t - rT).$$

Сигнал, определяющий сигнал МСИ, который образован символами, переданными после символа b_0 ,

$$g_2(t) = \sum_{r=1}^D S_r(t - rT).$$

Обычно чем больше скорость передачи символов $1/T$ в канале с ограниченной полосой, тем большее число соседних посылок с сигналом $S_0(t)$ будет определять сигнал $g_{\text{МСИ}}(t)$. В частном случае при $T_a = T$ ($D = 0$) сигнал $g_2(t) = 0$.

Отметим, что при значительных уровнях шумов в канале с МСИ предельное качество приема сигналов можно обеспечить только при оптимальных методах обработки. При этом, если передаваемый сигнал помимо МСИ подвергается другим искажениям, то рассмотренная модель существенно усложняется.

7.3. Модели дискретных каналов связи

7.3.1. Общие сведения

Ранее отмечалось, что в состав дискретного канала всегда входит непрерывный канал. Следовательно, от модели непрерывного канала можно перейти к модели дискретного канала, которая по существу является отображением тех процессов, которые происходят при передаче информации в непрерывном канале. Однако это зачастую значительно усложняет модель канала, поэтому

при анализе нередко используются упрощенные модели дискретного канала, при задании которых свойства входящих в них непрерывных каналов не учитываются.

Будем полагать, что модель дискретного канала задана, если известно множество входных сигналов, а также распределение условных вероятностей сигналов на выходе. В данном случае входными сигналами считают любую из множества последовательностей $B^{[n]}$, а выходными сигналами — последовательности $\hat{B}^{[n]}$, каждая из которых содержит n кодовых символов. В частном случае кодовые символы являются двоичными («0» и «1»). В общем случае число различных кодовых символов, или основание кода, обозначается m . Тогда все последовательности длиной n , число которых равно m^n , образуют n -мерное конечное векторное пространство. В этом пространстве задается операция сложения в виде поразрядного суммирования по модулю m , а также умножение на скаляр.

В случае когда $m = 2$, имеем векторное пространство, называемое пространством Хемминга. Скалярное произведение в этом случае запишем в виде

$$(x, y) = \sum_{k=1}^n x_k y_k,$$

где Σ — знак суммирования в обычном смысле.

Норма двоичного вектора представляется выражением

$$\|x\| = \sqrt{\sum_{k=1}^n x_k^2} = \sqrt{\sum_{k=1}^n x_k}$$

и определяется числом содержащихся в нем единиц. Норму такого вектора называют весом вектора и обычно обозначают w . Расстояние между векторами в пространстве Хемминга равно норме их разности:

$$d(x, y) = \|x - y\| = \sum_{k=1}^n (x_k - y_k) = \sum_{k=1}^n (x_k \oplus y_k).$$

Здесь знак \oplus означает сложение по модулю два (по mod 2), которое, как отмечалось ранее, выполняется по следующим правилам: $0 \oplus 0 = 0$; $1 \oplus 1 = 0$; $0 \oplus 1 = 1$; $1 \oplus 0 = 1$. Заметим, что сложение и вычитание по mod 2 эквивалентны.

Например:

$$\begin{array}{r} x = 1001011 \\ y = 0101101 \\ \hline x \oplus y = 1100110 \end{array}$$

В пространстве Хемминга расстояние между двоичными векторами равно числу позиций (элементов), которыми различаются эти векторы. В данном примере $d(x, y) = 4$.

Введем понятие вектора ошибок, под которым принято понимать поэлементную разность между принятым и переданным векторами. Иными словами, при передаче по дискретному каналу вектор сигнала складывается с вектором ошибок:

$$\tilde{B}^{[n]} = B^{[n]} + E^{[n]},$$

где $E^{[n]}$ — вектор ошибок, который в дискретном канале играет такую же роль, как и помеха в непрерывном канале.

В двоичном канале при $m = 2$ смысл вектора ошибок понимается наиболее просто. Каждая единица в этом векторе указывает на то, что именно на данной позиции в векторе сигнала произошла ошибка. Нулевая позиция в векторе ошибок свидетельствует о правильном приеме соответствующего элемента сигнала, т.е. переход от непрерывного канала к дискретному соответствует преобразованию любой помехи в поток ошибок. При этом различные модели дискретного канала будут описываться разными распределениями вероятностей векторов ошибок.

Рассмотрим теперь наиболее часто встречающиеся простейшие модели дискретных каналов.

7.3.2. Дискретный симметричный канал без памяти

Пусть $B = (b_1, \dots, b_k, \dots, b_n)$ — последовательность символов на входе канала, а $\tilde{B} = (\tilde{b}_1, \dots, \tilde{b}_k, \dots, \tilde{b}_n)$ — последовательность символов на выходе, где $k = 1, n$. Тогда вероятность появления символа \tilde{b}_k при условии, что b_k задано, определяется условной, или переходной, вероятностью $P(\tilde{b}_k | b_k)$.

Дискретный канал называется каналом без памяти, если каждый символ на его выходе зависит только от соответствующего символа на входе, а условная вероятность выходной последовательности $\tilde{B} = (\tilde{b}_1, \dots, \tilde{b}_n)$ при заданной входной последовательности $B = (b_1, \dots, b_n)$ определяется равенством

$$P(\tilde{B} | B) = \prod_{k=1}^n P(\tilde{b}_k | b_k).$$

Пусть ошибки любого из символов возникают независимо с вероятностью P , а правильный прием происходит с вероятностью

$(1 - P)$. Тогда условную вероятность приема символа b_j при передаче b_i запишем в виде

$$P(b_j | b_i) = \begin{cases} \frac{P}{m-1}, & j \neq i; \\ 1 - P, & j = i. \end{cases}$$

Будем далее рассматривать передачу двоичных символов «0» и «1». Вероятность появления любого n -мерного вектора ошибок, содержащего t единиц, которые соответствуют фактическим ошибкам, определяется выражением

$$P(E^{[n]}) = P^t(1 - P)^{n-t}.$$

В n -мерном векторе t ошибок могут занимать любые позиции, поэтому несложно определить вероятность появления t ошибок, расположенных произвольно в пределах последовательности длиной n ,

$$P(t) = C_n^t P^t (1 - P)^{n-t},$$

где $C_n^t = \frac{n!}{t!(n-t)!}$ — число различных сочетаний t ошибок (биномиальный коэффициент).

Для случая, когда переходная вероятность $P \ll 1$, вероятность ошибочного приема последовательности длиной n определяется выражением

$$P_{\text{ош}} = \sum_{t=1}^n C_n^t P^t (1 - P)^{n-t} = 1 - (1 - P)^n \approx nP.$$

Рассмотренную модель дискретного канала можно изобразить графически в виде переходных вероятностей (рис. 7.2). Данная модель дискретного канала является по сути аналогом рассмотренного ранее непрерывного канала с аддитивным белым шумом, в котором отсутствуют замирания.

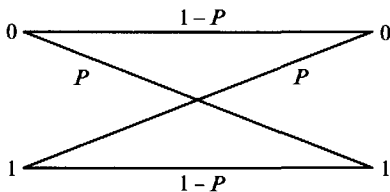


Рис. 7.2. Граф переходных вероятностей двоичного симметричного канала

7.3.3. Двоичный симметричный канал со стираниями

Данная модель канала является обобщением предыдущей рассмотренной модели в случае, когда на выходе появляется дополнительный третий символ (третье решение), если не удастся достоверно опознать переданный элемент сигнала.

Для получения такой модели на выходе канала вводят двухпороговое устройство со значением порога $|J|$. Тогда сигнал, принятый в этом канале, будет иметь вид

$$\tilde{z} = \begin{cases} +1, & z > J; \\ *, & -J \leq z \leq J; \\ -1, & z < -J. \end{cases}$$

Данный канал называется каналом со стиранием в нулевой зоне, поскольку здесь правильное решение может быть, если значение порога превышено при фактически переданном символе, а ошибочное решение — если превышение порога соответствует непередававшемуся элементу. Наконец, когда значение выходного сигнала z оказывается в интервале $-J \leq z \leq J$, производится стирание этого символа. Вероятности этих событий соответственно составляют $1 - P_0 - P_c$, P_0 и P_c , где P_0 — вероятность ошибки, а P_c — вероятность стирания символа.

Рассматриваемый канал можно описать матрицей переходных вероятностей:

$$\begin{bmatrix} 1 - P_0 - P_c & P_0 & P_c \\ P_0 & 1 - P_0 - P_c & P_c \end{bmatrix}$$

и соответственно графом, показанным на рис. 7.3.

В отличие от предыдущей модели здесь произведена замена 0 на +1 и 1 на -1.

Для рассматриваемой модели вероятность ошибки можно рассчитать по формуле

$$P_0 = \int_{-\infty}^{-J} \omega_0(z) dz = \int_J^{\infty} \omega_1(z) dz,$$

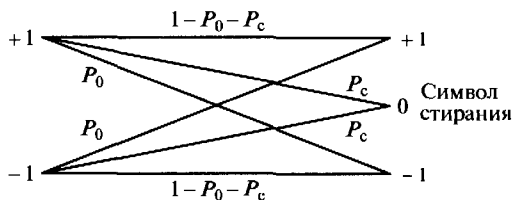


Рис. 7.3. Граф переходных вероятностей двоичного канала со стираниями

а вероятность стирания по формуле

$$P_c = \int_{-J}^J \omega_0(z) dz = \int_{-J}^J \omega_1(z) dz,$$

где $\omega_0(z)$ и $\omega_1(z)$ — плотность распределения выходного сигнала канала при передаче соответственно символа «0»; и символа «1».

По аналогии с предыдущим каналом можно рассчитать вероятность того, что в последовательности символов длиной n произошло t ошибок и τ стираний, расположенных произвольно:

$$P(t, \tau) = C_n^t C_{n-t}^\tau P_0^t P_c^\tau (1 - P_0 - P_c)^{n-t-\tau}.$$

В заключение отметим, что за счет введения стирания или отказа от принятия решения удастся значительно снизить вероятность ошибки, правда за счет возрастания вероятности стирания.

7.3.4. Дискретный несимметричный канал без памяти

В данном канале ошибки возникают независимо друг от друга, но вероятности их возникновения являются не одинаковыми, т. е. $P(\tilde{1}|0) \neq P(\tilde{0}|1)$. Иными словами, вероятность приема символа «1» при передаче символа «0» не равна вероятности приема символа «0» при передаче символа «1». Переходные вероятности в этом канале принимают вид, представленный на рис. 7.4.

Причины данного явления кроются в процессах, происходящих в реальных каналах связи, однако в ряде случаев это может быть следствием плохой регулировки аппаратуры.

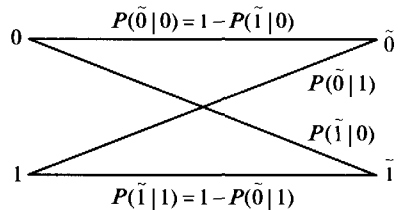


Рис. 7.4. Граф переходных вероятностей двоичного несимметричного канала

7.3.5. Дискретный канал с памятью

Канал передачи дискретной информации принято описывать моделью канала с памятью, т.е. моделью в которой символ выходной последовательности зависит статистически от соответствующего текущего символа, а также от предыдущих входных и выходных вероятностей появления символов.

Данную модель можно задать с помощью матрицы вероятностей

$$\begin{bmatrix} 1 - P_0 & P_0 \\ P_1 & 1 - P_1 \end{bmatrix},$$

где $1 - P_0$ — условная вероятность правильного приема $(k + 1)$ -го символа, если предыдущий элемент принят также верно; P_0 — условная вероятность ошибочного приема $(k + 1)$ -го символа, если предыдущий символ был принят верно; P_1 — условная вероятность ошибочного приема $(k + 1)$ -го символа при ошибочно принятом предыдущем символе; $1 - P_1$ — условная вероятность правильного приема $(k + 1)$ -го символа при неверно принятом предыдущем элементе.

Средняя (безусловная) вероятность ошибки в этом канале находится из уравнения

$$P = P_1 P + P_0(1 - P),$$

откуда

$$P = \frac{P_0}{1 + P_0 - P_1}.$$

Данная модель при своей простоте не всегда точно воспроизводит свойства реальных каналов, поэтому разработано достаточно много других более точных, но и более сложных моделей, которые здесь не рассматриваются.

Отметим только, что память в канале вызывается различными причинами. Так, одной из них являются замирания в радиоканале. В проводных каналах причиной памяти принято считать коммутационные помехи, возникающие при переключениях, что выводит канал из строя на некоторое время. Иногда причиной памяти могут быть особенности методов модуляции и демодуляции. Например, использование относительной фазовой модуляции приводит к сдвиганию ошибок, что подробнее будет рассмотрено далее.

7.4. Линейные преобразования случайных сигналов в каналах связи

При передаче информации по каналам связи принимаемые сигналы отличаются от переданных вследствие воздействия помех. Кроме того, как отмечалось ранее, различие сигналов обусловлено наличием в каналах связи линейных и нелинейных устройств. В связи с этим при анализе и синтезе элементов канала передачи информации приходится решать задачи по определению параметров случайных сигналов и помех на выходе канала. В большинстве случаев решение подобных задач сводится к нахождению корреляционной функции (временному анализу) или к нахождению спектральной плотности мощности, т. е. энергетического спектра (частотному анализу). При этом параметры элементов, через которые проходят сигналы, считаются заданными.

В системах передачи информации наиболее часто встречаются следующие линейные преобразования помехи или аддитивной смеси сигнала и помехи: фильтрация; интегрирование и дифференцирование; умножение на известную функцию времени; временная задержка.

В качестве примера рассмотрим подробнее прохождение через линейный элемент случайного гауссовского процесса с нулевым математическим ожиданием и дисперсией σ^2 .

Напомним, что любой линейный элемент с постоянными параметрами характеризуется передаточной функцией $K(j\omega)$, а также импульсной реакцией $g(t)$, под которой понимают полученное на его выходе колебание при подаче на вход единичного импульса (дельта-функции).

Комплексная передаточная функция

$$K(j\omega) = K(\omega) \exp[j\varphi(\omega)],$$

где $K(\omega)$ — амплитудно-частотная характеристика; $\varphi(\omega)$ — фазочастотная характеристика.

Передаточная функция и импульсная реакция в этом случае являются связанными друг с другом преобразованиями Фурье:

$$K(j\omega) = \int_{-\infty}^{\infty} g(t) e^{-j\omega t} dt;$$
$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K(j\omega) e^{j\omega t} d\omega.$$

Если на вход линейного элемента поступает случайный стационарный сигнал $X(t)$, то выходной сигнал $Y(t)$ можно найти на основе известного интеграла Дюамеля (интеграла свертки):

$$Y(t) = \int_0^{\infty} g(\tau)X(t-\tau)d\tau.$$

Здесь нижний предел равен нулю, так как для физически реализуемых элементов $g(t) = 0$ при $t < 0$.

Напомним также, что линейные преобразования гауссовского случайного процесса не изменяют вида закона его распределения, т. е. процесс на выходе остается гауссовским, а изменяются только его параметры (дисперсия, корреляционная функция, спектральная плотность мощности).

Определим сначала функцию корреляции выходного процесса:

$$\begin{aligned} B_Y(t_1, t_2) &= M\{Y(t_1)Y(t_2)\} = \\ &= M\left\{\int_0^{\infty} g(\tau_1)X(t_1-\tau_1)d\tau_1 \int_0^{\infty} g(\tau_2)X(t_2-\tau_2)d\tau_2\right\} = \\ &= \int_0^{\infty} \int_0^{\infty} M\{X(t_1-\tau_1)X(t_2-\tau_2)\}g(\tau_1)g(\tau_2)d\tau_1d\tau_2 = \\ &= \int_0^{\infty} \int_0^{\infty} B_X(\tau+\tau_1-\tau_2)g(\tau_1)g(\tau_2)d\tau_1d\tau_2, \end{aligned}$$

где $\tau = t_2 - t_1$; $B_X(\tau + \tau_1 - \tau_2)$ — функция корреляции входного случайного процесса.

Из приведенной формулы видно, что функция корреляции на выходе линейного элемента не зависит от моментов t_1 и t_2 по отдельности, а определяется значением их разности τ . Поэтому если на входе линейного устройства действует стационарный процесс, то и на его выходе это свойство сохраняется.

Перейдем теперь к частотному анализу и установим взаимосвязь между энергетическими спектрами на входе и выходе линейного элемента, для чего используем теорему Винера-Хинчина:

$$\begin{aligned} G_Y(\omega) &= \int_{-\infty}^{\infty} B_Y(\tau)e^{-j\omega\tau}d\tau = \\ &= \int_{-\infty}^{\infty} \left[\int_0^{\infty} \int_0^{\infty} B_X(\tau+\tau_1-\tau_2)g(\tau_1)g(\tau_2)d\tau_1d\tau_2 \right] e^{-j\omega\tau}d\tau = \\ &= \int_{-\infty}^{\infty} B_X(\tau+\tau_1-\tau_2)e^{-j\omega(\tau+\tau_1-\tau_2)}d\tau \int_0^{\infty} g(\tau_1)e^{j\omega\tau_1}d\tau_1 \int_0^{\infty} g(\tau_2)e^{-j\omega\tau_2}d\tau_2. \end{aligned}$$

Легко увидеть, что в этом выражении первый из трех интегралов представляет собой энергетический спектр на входе линей-

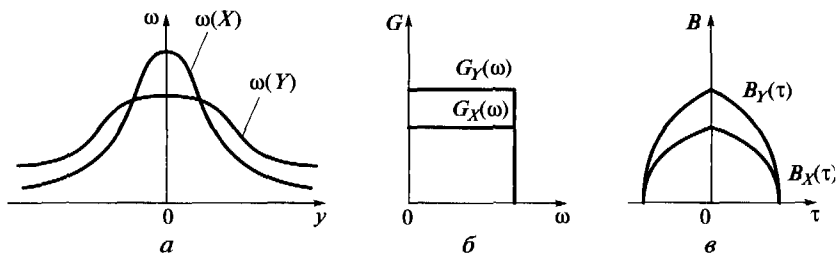


Рис. 7.5. Линейные преобразования характеристик случайного процесса: *а* — плотности распределения вероятностей; *б* — энергетического спектра; *в* — функции корреляции

ного элемента, а остальные интегралы — произведения передаточной функции и сопряженной с ней величины:

$$G_Y(\omega) = G_X(\omega)K(-j\omega)K(j\omega) = G_X(\omega)|K(j\omega)|^2,$$

где $|K(j\omega)|^2$ — квадрат модуля передаточной функции.

Используя полученные выражения для функции корреляции и энергетического спектра, определим характеристики случайного стационарного процесса на выходе безынерционного линейного элемента с передаточной функцией $K(j\omega) = K$.

Пусть процесс является гауссовским с плотностью распределения

$$\omega(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

Тогда несложно найти энергетический спектр и дисперсию:

$$G_Y(\omega) = K^2 G_X(\omega); D\{Y(t)\} = K^2 D\{X(t)\} = K^2 \sigma^2.$$

Выражение для дисперсии следует из того, что

$$D\{Y\} = B_Y(0) = \int_{-\infty}^{\infty} G_Y(f) df.$$

Плотность вероятностей и функцию корреляции можно описать следующими выражениями:

$$\omega(y) = \frac{1}{\sqrt{2\pi K^2 \sigma^2}} \exp\left(-\frac{y^2}{2K^2 \sigma^2}\right); B_Y(\tau) = K^2 B_X(\tau).$$

Данные выражения иллюстрирует рис. 7.5, из которого видно, что характер распределения случайного процесса на выходе сохраняется, но изменяются его параметры.

7.5. Нелинейные преобразования случайных сигналов в каналах связи

Реальные каналы связи не всегда можно описать линейными моделями, так как они содержат элементы, со сложными вольт-амперными характеристиками, т. е. полупроводниковые диоды и транзисторы. Кроме того, нелинейность возникает в некоторых радио- и оптических каналах при повышенной мощности передатчиков.

Построение и анализ моделей нелинейных каналов связи в общем случае является достаточно сложной задачей. Отдельные частные случаи рассматривались ранее применительно к разным видам модуляции и демодуляции сигналов, поэтому здесь ограничимся анализом модели, в которой сигналы на выходе $Y(t)$ и входе $X(t)$ канала связаны выражением $Y(t) = \varphi[X(t)]$.

В ТЭС, как правило, преобразование $X \rightarrow Y$ является однозначным (т. е. отсутствует гистерезис), а обратное преобразование $Y \rightarrow X$ не всегда однозначное, как например квадратичная цепь с характеристикой $Y = kX^2$.

При нелинейных преобразованиях спектр случайного процесса на выходе претерпевает значительные изменения. Изменяется также и вид закона распределения при прохождении случайным сигналом нелинейных элементов канала.

Пусть на входе нелинейной системы имеется смесь сигнала и аддитивного шума:

$$X(t) = S(t) + n(t),$$

которая занимает сравнительно узкую полосу частот F_c .

В этом случае на выходе системы будут присутствовать составляющие трех видов:

- результаты биений частотных составляющих сигнала ($c \times c$);
- результаты биений составляющих шума ($ш \times ш$);
- результаты биений сигнала и шума ($c \times ш$).

Обычно все эти сигналы разделить на выходе системы не удастся, в чем и заключается основная трудность данного анализа.

В случае если известна характеристика нелинейной системы $Y = \varphi(X)$, а также двумерная плотность распределения $\omega(x_1, x_2, t_1, t_2)$ случайного процесса на входе, для определения характеристик процесса на выходе $Y(t)$ используется следующий алгоритм.

Сначала находится математическое ожидание

$$M\{Y(t)\} = M\{\varphi[X(t)]\} = \int_{-\infty}^{\infty} \varphi(x)\omega(x, t)dx,$$

а затем с его помощью определяется корреляционная функция

$$B_Y(t, \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\varphi(x_1(t)) - M\{Y(t)\}][\varphi(x_2(t + \tau)) - M\{Y(t + \tau)\}] \times \\ \times \omega(x_1, x_2, t, t + \tau) dx_1 dx_2,$$

После чего на основании теоремы Винера-Хинчина с помощью преобразования Фурье по найденной функции корреляции вычисляется спектральная плотность мощности.

В общем случае нахождение закона распределения выходного процесса и его характеристик при нелинейных преобразованиях является весьма сложной задачей, поэтому здесь подробно не рассматривается. Отметим только, что для узкополосных процессов при их квазигармоническом представлении в некоторых случаях задача несколько упрощается.

7.6. Случайные преобразования сигналов в каналах связи

Ранее были рассмотрены детерминированные преобразования сигналов при прохождении их по каналам связи. Вместе с тем в отдельных звеньях канала, в частности в среде распространения, сигнал также подвергается случайным преобразованиям. В простейшем случае это может происходить под воздействием аддитивных помех. В других ситуациях к этому добавляются случайные изменения параметров канала, которые могут приводить к необратимым искажениям передаваемых сигналов.

Остановимся подробнее на моделях каналов, с которыми приходится встречаться на практике.

Простейшим является случайный канал, описываемый уравнением

$$Y(t) = \mu(t)X[t - \tau(t)],$$

где $X(t)$ и $Y(t)$ — случайные процессы соответственно на входе и выходе канала; $\mu(t)$ — случайный коэффициент передачи; $\tau(t)$ — флуктуирующее время задержки сигнала.

Будем считать, что на входе канала имеется узкополосный процесс, который удобно представить в квазигармонической форме:

$$X(t) = A(t) \cos[\omega_0 t + \Phi(t)],$$

где $A(t)$ и $\Phi(t)$ — медленно изменяющиеся функции.

Пусть временная задержка τ достаточно мала. Тогда можно считать, что $A(t - \tau) \approx A(t)$, $\Phi(t - \tau) \approx \Phi(t)$.

Выходной сигнал в этом случае запишем в виде

$$Y(t) = \mu(t)A(t - \tau) \cos [\omega_0(t - \tau) + \Phi(t - \tau)] \approx \\ \approx \mu(t)A(t) \cos [\omega_0 t + \Phi(t) - \theta] = \mu(t) \cos \theta X(t) - \mu(t) \sin \theta \tilde{X}(t),$$

где $\theta = \omega_0 \tau$ — фазовый сдвиг в канале; $\tilde{X}(t)$ — процесс, сопряженный с $X(t)$.

Из этого выражения видно, что временная задержка сигнала τ приводит к сдвигу фаз в канале. При этом даже при сравнительно небольших колебаниях времени задержки фазовый сдвиг θ может изменяться в достаточно больших пределах вследствие больших значений $\omega_0 = 2\pi f_0$. Для этого достаточно выполнение условия $\Delta\tau \gg 1/f_0$, где $\Delta\tau$ — среднеквадратическое отклонение временной задержки, f_0 — средняя частота спектра сигнала. В радиоканалах это условие обычно выполняется.

Более сложным является случай многолучевого распространения сигнала по каналу, при котором в точку приема поступает сигнал, пришедший по нескольким параллельным путям. При этом значения коэффициента передачи и времени задержки сигнала у разных лучей распространения различные и колеблющиеся в небольших пределах (рис. 7.6).

Данная ситуация возникает при распространении сигналов в неоднородной среде с отражением от различных неоднородностей. Приход нескольких лучей с разным временем запаздывания и сложение их в одной точке вызывает явление замирания, которое выражается в случайных изменениях передаточной функции канала:

$$\dot{K}(t, j\omega) = \int_{-\infty}^{\infty} G(t, \tau) e^{-j\omega\tau} d\tau,$$

где $G(t, \tau)$ — импульсная реакция канала.

Определим распределение вероятностей комплексной передаточной функции $\dot{K}(t, j\omega)$ и, в первую очередь, ее модуля K , для чего запишем:

$$\dot{K}(t, j\omega) = Ke^{j\theta} = X(t, \omega) + jY(t, \omega),$$

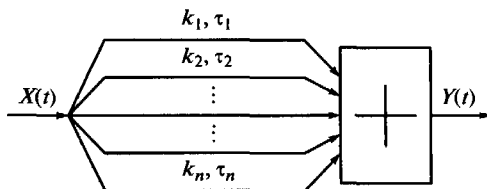


Рис. 7.6. Многолучевое распространение сигнала

где $K = |\dot{K}|$ и θ — соответственно модуль и аргумент передаточной функции, являющиеся случайными функциями t и ω ; $X = K \cos \theta$, $Y = K \sin \theta$ — квадратурные составляющие.

С другой стороны, согласно принятой модели многолучевой передачи сигналов можно записать

$$\dot{K}(t, j\omega) = \sum_{k=1}^n K_k e^{j\theta_k} = \sum_{k=1}^n K_k \cos \theta_k + j \sum_{k=1}^n K_k \sin \theta_k,$$

откуда

$$X(t, \omega) = \sum_{k=1}^n K_k \cos \theta_k; \quad Y(t, \omega) = - \sum_{k=1}^n K_k \sin \theta_k.$$

Величины X и Y образуются как результат сложения большого числа слабо коррелированных величин с ограниченными дисперсиями. Следовательно, согласно центральной предельной теореме их можно считать распределенными по гауссовскому закону.

Пусть все величины K_k имеют один порядок, а фазовые сдвиги являются достаточно большими. Тогда дисперсии квадратурных составляющих X и Y будут равны σ^2 , а математические ожидания $m_X = m_Y = 0$. В этом случае модуль коэффициента передачи K подчиняется рэлеевскому распределению:

$$\omega(K) = \frac{K}{\sigma^2} \exp\left(-\frac{K^2}{2\sigma^2}\right), \quad K \geq 0.$$

В этом случае предполагается, что фаза θ имеет равномерное распределение в интервале от 0 до 2π , а дисперсия σ^2 квадратурных составляющих равна средней мощности принимаемого сигнала.

Кроме рассмотренной ситуации в каналах радиосвязи возможны случаи, при которых в одном из лучей распространения коэффициент передачи оказывается значительно больше, чем в других. В результате в точку приема может приходиться регулярный, т. е. незамирающий луч.

Тогда модуль коэффициента передачи будет подчиняться обобщенному распределению Рэля или распределению Райса:

$$\omega(K) = \frac{K}{\sigma^2} \exp\left(-\frac{K^2}{2\sigma^2} - q^2\right) I_0\left(\frac{K\sqrt{2q}}{\sigma}\right), \quad K \geq 0,$$

где $q^2 = \frac{m_X^2 + m_Y^2}{2\sigma^2}$ — отношение средних мощностей регулярной и флуктуирующей составляющих; $I_0(x)$ — модифицированная функция Бесселя нулевого порядка.

Подобные модели замираний, основанные на многолучевом распространении сигналов, характерны для большинства радиоканалов, а также для некоторых других каналов связи, в том числе и для проводных.

Контрольные вопросы

1. По каким параметрам классифицируются каналы телекоммуникаций?
2. Как описываются следующие модели каналов: с постоянными параметрами и белым шумом, с неизвестной фазой, с замираниями сигналов?
3. Что представляет собой пространство Хэмминга и как определяются его параметры?
4. Опишите модели следующих дискретных каналов связи: двоичного симметричного без памяти, со стираниями символов, с памятью и несимметричного.
5. В чем заключается сущность временного и частотного анализа при линейных преобразованиях в каналах связи?
6. К каким последствиям приводят случайные преобразования в каналах телекоммуникаций?

ЭЛЕМЕНТЫ ТЕОРИИ ПОМЕХОУСТОЙЧИВОГО ПРИЕМА ДИСКРЕТНЫХ СООБЩЕНИЙ

8.1. Постановка задачи синтеза оптимального приемника

В процессе передачи по каналам связи, как уже отмечалось, сигналы подвергаются непредсказуемым искажениям после чего правильное восстановление сообщений может быть затруднено. Однако при этом приемник можно построить таким образом, чтобы он обеспечивал максимум возможной верности приема, т. е. минимум средней вероятности ошибки.

Такой подход к построению приемника, впервые предложенный академиком В.А. Котельниковым в 1946 — 1947 гг., лежит в основе современной теории помехоустойчивой передачи информации. *Идеальным приемником* он назвал устройство, которое при заданных условиях обеспечивает максимум вероятности правильного приема сигналов или минимум средней вероятности ошибки. Помехоустойчивость такого приемника (демодулятора) принято считать потенциальной.

Смысл приведенного определения заключается в том, что принципиально невозможно превысить значение помехоустойчивости такого приемника, хотя и можно к ней приблизиться. Противный случай будет противоречить данному определению.

Критерием идеального наблюдателя (критерием Котельникова) принято называть такой критерий, согласно которому качество приемников оценивается значением вероятности правильного приема элемента сигнала.

С использованием данного критерия рассмотрим общий подход к задаче синтеза оптимального приемника.

Обычно под синтезом оптимального приемника принято понимать нахождение оптимальной в соответствии с некоторым критерием качества приема процедуры обработки наблюдаемого случайного процесса $z(t)$, который является смесью полезного сигнала и помех. Далее на основе такого синтеза создается устройство, реализующее найденную процедуру обработки.

Будем полагать, что сообщения α_i (где $i=1, M$) на выходе источника появляются с вероятностями $P(\alpha_i)$, причем $\sum_{i=1}^M P(\alpha_i) = 1$.

Соответствующие им сигналы $S_i(t)$ полностью известны на приемной стороне, поскольку считается, что сигналы передаются по каналу с постоянными параметрами, когда коэффициент передачи μ и время задержки τ в канале известны.

Колебания на входе приемника запишем в виде

$$z(t) = S_i(t) + n(t); 0 \leq t \leq T, \quad (8.1)$$

где $n(t)$ — аддитивная гауссовская помеха.

Получатель не знает, какой вариант сигнала передавался, т. е. ему неизвестен индекс i . Источником ошибок при принятии решения в приемнике является шум $n(t)$. Задача состоит в синтезе структуры оптимального приемника, который при полностью известных параметрах сигнала называется когерентным.

Будем считать, что применяемые сигналы $S_i(t)$ являются финитными, и их длительность равна T . Предположим также, что известны границы тактового интервала, на котором происходит обработка принимаемого колебания, т. е. в системе обеспечена надежная тактовая синхронизация.

Найдем, какое правило решения будет оптимальным, если критерием оптимальности является критерий идеального наблюдателя. Другими словами, найдем правило решения, обеспечивающее минимум вероятности ошибки.

Пусть приемник по результатам наблюдения принимаемой смеси вычисляет апостериорную вероятность $P\{\alpha_i | z(t)\}$, т. е. условную вероятность того, что поступившим на вход канала символом является α_i , при условии, что на вход приемника поступит реализация $z(t)$.

Пусть в наблюдаемой смеси $z(t)$ содержится сигнал $S_i(t)$, т. е. передается символ α_i . Приемник принимает решение β_e ($e=1, M$), которое может быть верным, если $\beta_e = \alpha_i$, и ошибочным, когда $\beta_e \neq \alpha_i$. При этом в приемнике формируется апостериорная вероятность $P\{\alpha_i | z(t)\}$, которая в данном случае является вероятностью правильного приема. Тогда вероятность ошибочного решения

$$P_{\text{ош}}(z(t), \beta_i) = 1 - P\{\alpha_i | z(t)\}. \quad (8.2)$$

Сравним между собой два приемника, из которых один выбирает решение $\beta_i = \alpha_i$, а другой при той же реализации $z(t)$ — $\beta_j = \alpha_j$.

Из формулы (8.2) следует, что

$$P_{\text{ош}}(z(t), \beta_i) < P_{\text{ош}}(z(t), \beta_j), \text{ если } P\{\alpha_i | z(t)\} > P\{\alpha_j | z(t)\},$$

т. е. вероятность ошибки меньше у того приемника, который выбирает решение, соответствующее большей апостериорной вероятности. Следовательно, приемник, выбирающий решение, соответствующее максимальной апостериорной вероятности

$$P\{\alpha_i | z(t)\} = \max_{j=1, M} P\{\alpha_j | z(t)\},$$

обеспечивает меньшую вероятность ошибок, чем любой другой приемник, т. е. является оптимальным по критерию идеального наблюдателя.

Иными словами, символ α_i регистрируется при выполнении следующей системы неравенств:

$$P\{\alpha_i | z(t)\} > P\{\alpha_j | z(t)\}; \quad j = \overline{1, M}; \quad \forall j \neq i, \quad (8.3)$$

которую сокращенно можно записать в виде

$$\alpha_i = \arg\{\max_j P\{\alpha_j | z(t)\}\}; \quad j = \overline{1, M}, \quad (8.4)$$

т. е. в качестве решения $\beta_i = \alpha_i$ принимается такое решение, для которого апостериорная вероятность максимальна.

При передаче двоичных сигналов данное правило решения принимает вид

$$P\{0 | z(t)\} > P\{1 | z(t)\}.$$

Если это неравенство выполняется, то регистрируется символ «0», а в противном случае — «1».

Обозначение \arg означает выбор аргумента функции $P\{\alpha_j | z(t)\}$.

Получим сначала алгоритм оптимального приема при поступлении в приемник дискретных отсчетов принимаемой реализации.

Приемник, работающий по такому алгоритму, назовем цифровым. Он интересен сам по себе в связи с широким распространением цифровой обработки сигнала. Далее из оптимального цифрового приемника получим оптимальный аналоговый приемник, т. е. приемник, обрабатывающий непрерывную наблюдаемую реализацию.

Итак, пусть на вход приемника поступают отсчеты смеси сигнала и помехи:

$$z(t_k) = S_i(t_k) + n(t_k); \quad k = \overline{1, n},$$

где n — число отсчетов на длительности посылки сигнала.

При этом интервалы между отсчетами

$$\Delta t = t_k - t_{k-1} = \frac{1}{2F_c} = \frac{T}{n},$$

где F_c — ширина спектра сигнала.

Следовательно, выборки делаются на основе теоремы отсчетов.

Алгоритм работы оптимального цифрового приемника можно получить из формулы (8.3) или (8.4), выполнив подстановку $P(\alpha_i / z_1, \dots, z_n)$, под которой понимается условная апостериорная вероятность того, что действительно передавался символ α_i , если в приемник поступила совокупность отсчетов z_1, \dots, z_n .

Тогда для цифрового приемника получим следующие правила решения:

$$P(\alpha_i / z_1, \dots, z_n) \geq P(\alpha_j / z_1, \dots, z_n); j = \overline{1, M}; \forall j \neq i; \quad (8.5)$$

$$\alpha_i = \arg\{\max_j P(\alpha_j / z_1, \dots, z_n)\}. \quad (8.6)$$

Пусть выражение

$$\omega_n(z_1, \dots, z_n / \alpha_i) = \omega_n\{[z_1 - S_i(t_1), \dots, z_n - S_i(t_n)] / S_i\} \quad (8.7)$$

представляет собой n -мерную условную плотность распределения вероятностей отсчетов принятой реализации в точках t_k ($k = 1, 2, \dots, n$), которая определяется реализацией сигнала и помехами в канале связи.

Введем также безусловную n -мерную плотность вероятностей отсчетов принятой реализации, которую можно определить по формуле полной вероятности

$$\omega_n(z_1, \dots, z_n) = \sum_{i=1}^M P(\alpha_i) \omega_n\{z_1, \dots, z_n / \alpha_i\}. \quad (8.8)$$

Тогда условную вероятность $P(\alpha_i / z_1, \dots, z_n)$ можно найти с использованием известной формулы Байеса:

$$P(\alpha_i / z_1, \dots, z_n) = \frac{P(\alpha_i) \omega_n(z_1, \dots, z_n / \alpha_i)}{\omega_n(z_1, \dots, z_n)}. \quad (8.9)$$

Определим теперь отношение правдоподобия, т. е. отношение n -мерных плотностей вероятности

$$\Lambda_i(z_1, \dots, z_n) = \frac{\omega_n(z_1, \dots, z_n / \alpha_i)}{\omega_n(z_1, \dots, z_n)}. \quad (8.10)$$

Это отношение показывает, на сколько i -я гипотеза правдоподобней, чем нулевая.

Нулевая гипотеза соответствует случаю, когда на входе приемника нет сигнала, а присутствует только помеха. С учетом отношения (8.10) из формулы (8.9) получим

$$P(\alpha_j / z_1, \dots, z_n) = P(\alpha_j) \Lambda_j(z_1, \dots, z_n). \quad (8.11)$$

Подставив выражение (8.11) в формулу (8.5) получим выражение, соответствующее алгоритму оптимального цифрового приемника, реализующего критерий идеального наблюдателя:

$$\frac{\Lambda_i(z_1, \dots, z_n)}{\Lambda_j(z_1, \dots, z_n)} \geq \frac{P(\alpha_j)}{P(\alpha_i)}; \quad j = \overline{1, M}; \quad \forall j \neq i. \quad (8.12)$$

В случае когда все сообщения равновероятны, т. е. $P(\alpha_i) = 1/M$, данное правило решения упрощается и приобретает следующий вид:

$$\Lambda_i(z_1, \dots, z_n) \geq \Lambda_j(z_1, \dots, z_n); \quad j = \overline{1, M}; \quad \forall j \neq i. \quad (8.13)$$

Найдем теперь правило решения при обработке непрерывной реализации $z(t)$ без ее дискретизации. Для этого необходимо, чтобы число отсчетов n на отрезке длительности T стремилось к бесконечности. При этом функционал отношения правдоподобия i -й гипотезы по отношению к нулевой гипотезе запишем в виде

$$\Lambda_i\{z(t)\} = \lim_{n \rightarrow \infty} \Lambda_i(z_1, \dots, z_n). \quad (8.14)$$

Правило решения для оптимального приемника по критерию идеального наблюдателя в случае непрерывного наблюдения, следующее из соотношения (8.12), будет иметь вид

$$\frac{\Lambda_i\{z(t)\}}{\Lambda_j\{z(t)\}} \geq \frac{P(\alpha_j)}{P(\alpha_i)}; \quad j = \overline{1, M}; \quad \forall j \neq i. \quad (8.15)$$

В свою очередь, правило максимального правдоподобия (8.13) для случая непрерывного наблюдения можно записать в виде

$$\Lambda_i\{z(t)\} \geq \Lambda_j\{z(t)\}; \quad j = \overline{1, M}; \quad \forall j \neq i. \quad (8.16)$$

Правила решений (8.12) и (8.13) являются основой для синтеза оптимальных цифровых приемников, работающих при дискретном наблюдении.

Правила решений (8.15) и (8.16) позволяют определить структуры оптимальных приемников при непрерывном наблюдении.

Рассмотренная постановка задачи синтеза относится к классу так называемых задач различения сигналов, т. е. в этом случае определяют, какой конкретно из двух или нескольких вариантов сигнала присутствует в наблюдаемом колебании, являющемся смесью сигнала и помех. Именно такие задачи в достаточно общей постановке являются предметом рассмотрения в теории связи.

По отношению к рассмотренным задачам несколько иными являются задачи по обнаружению сигнала на фоне помех. В этом

случае предполагается, что получателю неизвестен факт наличия или отсутствия сигнала в принятом колебании, следовательно, по результатам наблюдений необходимо принять соответствующее решение. В такой постановке подобные задачи рассматриваются в системах передачи защищенной информации, например в системах пожарной или иной охранной сигнализации, а также в системах радио- и гидролокации.

8.2. Статистические критерии оптимального приема сигналов

Рассмотренный выше критерий идеального наблюдателя (критерий Котельникова) представляется наиболее целесообразным для применения в теории и технике электрической связи.

Вместе с тем этот критерий имеет следующие недостатки:

- необходимость наличия априорной информации о вероятностях передачи какого-либо сообщения, которая не всегда известна;
- равновероятность появления любой из ошибок.

В ряде случаев на второй недостаток можно не обращать внимания. Например, при передаче буквенно-цифрового текста любая ошибка является одинаково нежелательной. Однако в других случаях разные ошибки могут приводить к различным последствиям.

Например, в системах автоматической пожарной сигнализации значительно опаснее пропустить сигнал о пожаре, чем сыграть ложную тревогу при отсутствии пожара на самом деле. Такого рода последствия в военной сфере характерны для систем радио- и гидролокации, т. е. значительно больше проблем создает пропуск воздушной или подводной цели, которая может применить оружие по объекту, чем объявление ложной тревоги, которая не повлечет за собой серьезных последствий — гибели людей, техники и др.

В ситуациях, когда невозможно определить априорную вероятность передачи отдельных сообщений, а последствия возникновения различных ошибок разные, применяют критерий Неймана — Пирсона. В соответствии с этим критерием приемник считается оптимальным, если при заданной вероятности ложной тревоги $P_{л.т}$ (ошибочное обнаружение сигнала, когда он фактически отсутствует) будет обеспечена минимальная вероятность пропуска сигнала $P_{пр}$.

Представим непрерывные реализации $z(t)$ в интервале наблюдений $(0, T)$ в виде вектора координат $z = (z_1, \dots, z_n)$ при разложении по координатному базису.

Пусть $\omega_n(z|1)$ и $\omega_n(z|0)$ — n -мерные условные плотности распределения координат вектора при условии, что в принимаемой

смеси содержится или отсутствует сигнал. В соответствии с этим все пространство принимаемых решений можно разделить различными способами на две области: область решения V_0 об отсутствии сигнала и область решения V_1 о его наличии. При этом всегда найдется оптимальный способ разделения пространства, который обеспечит равенство вероятности ложной тревоги некоторому наперед заданному значению ε , т. е.

$$P_{л.т} = \int_{V_1} \omega(z|0) dz = \varepsilon,$$

где $\omega_n(z|0)$ — плотность распределения помехи, так как символ «0» соответствует в данном случае отсутствию сигнала.

Иными словами, вероятность ложной тревоги зависит от вероятностных характеристик помехи и выбора области V_1 о наличии сигнала. С другой стороны, задание этой области определяет вероятность правильного обнаружения и соответственно пропуска сигнала, которая должна быть минимальной:

$$P_{пр} = \int_{V_0} \omega(z|1) dz = 1 - \int_{V_1} \omega(z|1) dz,$$

где $\int_{V_1} \omega(z|1) dz$ — вероятность правильного обнаружения сигнала.

Здесь и далее в данном подразделе интегрирование осуществляется по векторному аргументу, а область интегрирования и интегралы — многомерные.

Если вероятность $P_{л.т}$ известна, то минимум вероятности пропуска сигнала $P_{пр}$ достигается при выполнении неравенства

$$\omega(z|1) / \omega(z|0) > \lambda,$$

где λ — значение порога, определяемое заданным значением вероятности ложной тревоги $P_{л.т} = \varepsilon$.

Заметим, что в отличие от критерия идеального наблюдателя критерий Неймана — Пирсона не требует знания априорных вероятностей передаваемых сообщений.

Наиболее общим критерием, при котором устраняется второй недостаток критерия идеального наблюдателя, является байесовский критерий минимума среднего риска. Его сущность заключается в том, что любой паре переданного символа α_i и принятого символа α_j ($i \neq j$) приписывают некоторое числовое значение $L(\alpha_i, \alpha_j)$, называемое *потерей*, или *стоимостью*. Чем более нежелательна ошибка, тем большую стоимость она имеет. Правильному приему в этом случае будет соответствовать нулевая потеря.

Введем понятие условного риска, т. е. условное математическое ожидание значения потери при передаче некоторого символа α_i :

$$R_i = \sum_{j=1}^M P(\tilde{\alpha}_j | \alpha_i) L(\alpha_i, \alpha_j) = \sum_{j=1}^M L(\alpha_i, \alpha_j) \int_{A_j} \omega(z | \alpha_i) dz,$$

где $\omega(z | \alpha_i)$ — условная плотность распределения колебания $z(t)$ на входе приемника при передаче символа α_i ; A_j — область принятия решения о появлении колебания $z(t)$ в приемнике, если передавался символ α_j .

При усреднении R_i по всем символам α_i (где $i = \overline{1, M}$) получим значение среднего риска

$$R_{\text{ср}} = \sum_{i=1}^M \sum_{j=1}^M P(\alpha_i) L(\alpha_i, \alpha_j) \int_{A_j} \omega(z | \alpha_i) dz.$$

Согласно рассматриваемому критерию оптимальным будет приемник, обеспечивающий наименьшее значение среднего риска $R_{\text{ср}}$. Приемник, в котором достигается этот минимум, называется байесовским. В этом случае потери, обусловленные ошибками в принятии решений, будут сведены к минимальным. Ограниченность применения этого критерия обуславливается, во-первых, необходимостью знания априорных вероятностей передаваемых символов, а, во-вторых, некоторой неопределенностью при назначении стоимостей потерь.

В частном случае, когда ошибки одинаково нежелательны, критерий минимального среднего риска совпадает с критерием идеального наблюдателя. Байесовский приемник преобразуется в этом случае в идеальный приемник по критерию Котельникова.

Можно показать, что в принципе все критерии связаны с нахождением отношения правдоподобия и отличаются лишь значением порога B , при превышении которого принимается решение о передаче определенного символа сообщения α_i :

$$\Lambda_{ij} = \frac{\Lambda_i}{\Lambda_j} > B.$$

В заключение отметим, что большинство систем передачи дискретных сообщений обладает примерно одинаковыми априорными вероятностями передаваемых сообщений. Кроме того, в этих системах ошибки являются одинаково нежелательными. Поэтому далее всегда будем применять критерий Котельникова, который в таких случаях приводит к правилу максимального правдоподобия.

8.3. Синтез оптимального когерентного приемника в условиях аддитивного шума

На основе правил решений (8.13) и (8.16), оптимальных по максимуму правдоподобия, получим структуры оптимального циф-

рового и аналогового приемников в случае приема сигналов на фоне аддитивного белого гауссовского шума.

Рассмотрим сначала синтез структуры цифрового приемника. Пусть на вход приемника поступают дискретные отсчеты принимаемой смеси $z(t_k)$, равные $S_i(t_k) + n(t_k)$. Дискретные отсчеты $n(t_k)$ представляют собой отсчеты белого гауссовского шума $n(t)$ с нулевым средним значением и односторонней спектральной плотностью N_0 . Дисперсия дискретной случайной величины $n(t_k)$ определяется выражением $\sigma^2 = N_0 F_c$, где F_c — ширина спектра сигнала. С учетом того что интервал дискретизации $\Delta t = \frac{1}{2F_c}$, дисперсия $\sigma^2 = \frac{N_0}{2\Delta t}$.

Отсчеты белого гауссовского шума некоррелированы и независимы, поэтому n -мерная плотность распределения отсчетов представляет собой произведение одномерных плотностей:

$$\begin{aligned} \omega_n(z_1, \dots, z_n) &= \omega_n(z_1, \dots, z_n/0) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2(t_k)}{2\sigma^2}} = \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^n z^2(t_k)\right\}. \end{aligned} \quad (8.17)$$

Поскольку разности $z(t_k) - S_i(t_k) = n(t_k)$ являются отсчетами белого гауссовского шума, то n -мерная условная плотность распределения с учетом независимости различных отсчетов определяется выражением

$$\begin{aligned} \omega_n(z_1, \dots, z_n / \alpha_i) &= \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[z(t_k) - S_i(t_k)]^2}{2\sigma^2}} = \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^n [z(t_k) - S_i(t_k)]^2\right\}. \end{aligned} \quad (8.18)$$

Подставив выражения (8.17) и (8.18) в отношение правдоподобия (8.10), получим

$$\begin{aligned} \Lambda_j(z_1, \dots, z_n) &= \\ &= \exp\left\{\frac{1}{2\sigma^2} \sum_{k=1}^n z^2(t_k)\right\} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^n [z(t_k) - S_i(t_k)]^2\right\}. \end{aligned} \quad (8.19)$$

Наконец, подставив выражение (8.19) в правило решения (8.13), после логарифмирования по натуральному основанию найдем процедуру обработки сигналов оптимальным цифровым приемником:

$$\sum_{k=1}^n [z(t_k) - S_i(t_k)]^2 < \sum_{k=1}^n [z(t_k) - S_j(t_k)]^2. \quad (8.20)$$

Согласно соотношению (8.20) принимается решение о том, что в принимаемой аддитивной смеси содержится сигнал $S_i(t)$, т. е. регистрируется символ α_i , для которого сумма квадратов разности отсчетов наблюдаемого колебания $z(t_k)$ и сигнала $S_i(t_k)$ будет наименьшей среди всех остальных. Физический смысл работы приемника ясен: если в принимаемой смеси содержится сигнал $S_i(t)$, то разности $z(t_k) - S_i(t_k)$ будут обусловлены шумом и сумма их квадратов с большой вероятностью будет меньше соответствующей суммы квадратов других разностей $z(t_k) - S_j(t_k)$.

Получим теперь структуру оптимального приемника при непрерывном наблюдении. Ясно, что максимальная (потенциальная) помехоустойчивость оптимального приемника будет реализована при обработке сигнала в течение всей посылки с длительностью T , а не в дискретных отсчетах. Осуществим в выражении (8.19) предельный переход при $\Delta t \rightarrow 0$. Число отсчетов (сечений) при этом будет стремиться к бесконечности, сумма преобразуется в интеграл, а функция отношения правдоподобия трансформируется в функционал отношения правдоподобия:

$$\Lambda_i[z(t)] = \exp \left\{ \frac{1}{N_0} \int_0^T z^2(t) dt \right\} \exp \left\{ -\frac{1}{N_0} \int_0^T [z(t) - S_i(t)]^2 dt \right\}. \quad (8.21)$$

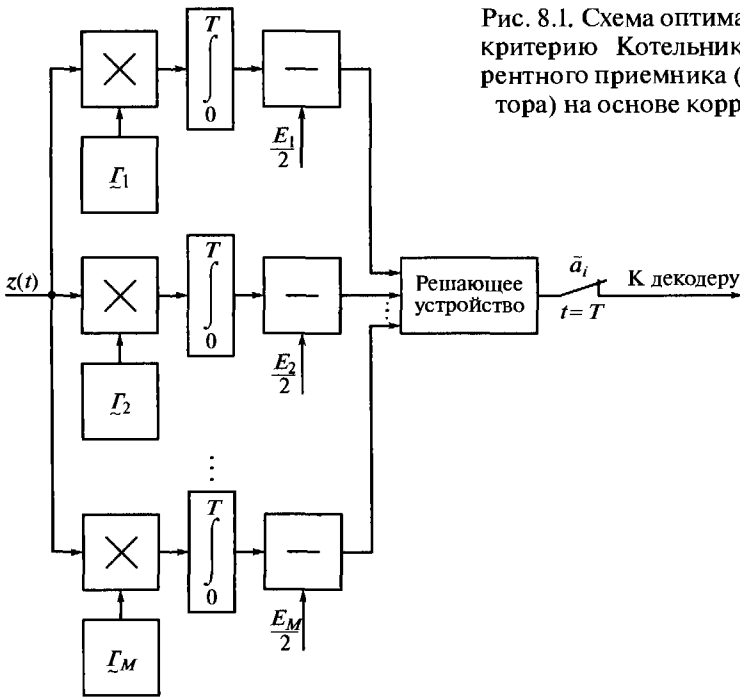
Заметим, что формула (8.21) поясняет физический смысл функционала отношения правдоподобия: он является функцией от функции, т. е. принимает то или иное числовое значение в зависимости от того, какая функция в него входит. Так, функционал (8.21) принимает максимальное значение, если в принимаемой смеси содержится сигнал $S_i(t)$.

Подставив функционал (8.21) в правило решения (8.16), после логарифмирования получим следующие соотношения для процедуры обработки принимаемой смеси:

$$\int_0^T [z(t) - S_i(t)]^2 dt < \int_0^T [z(t) - S_j(t)]^2 dt; \quad i, j = \overline{1, M}; \quad \forall j \neq i. \quad (8.22)$$

Согласно соотношениям (8.22) символ α_i регистрируется, если интеграл в левой части наименьший. В данном соотношении левая и правая части являются расстояниями между принятой реализацией $z(t)$ и соответствующими сигналами $S_i(t)$ и $S_j(t)$ в пространстве Гильберта, поэтому принимается решение о приеме того сигнала $S_i(t)$, расстояние до которого от принятой реализации $z(t)$ минимальное.

Рис. 8.1. Схема оптимального по критерию Котельникова когерентного приемника (демодулятора) на основе корреляторов



Правило (8.22) можно представить и в другом виде. После выполнения возведения в квадрат и упрощения получим

$$\int_0^T z(t)S_i(t)dt - 0,5E_i > \int_0^T z(t)S_j(t)dt - 0,5E_j; \forall j \neq i. \quad (8.23)$$

Для двоичной системы, где используются сигналы, реализации которых $S_0(t)$ и $S_1(t)$ имеют равные энергии, принимается решение о приеме символа «0», если

$$\int_0^T z(t)S_0(t)dt > \int_0^T z(t)S_1(t)dt. \quad (8.24)$$

В противном случае считается, что передан символ «1».

Из полученных выражений видно, что процедура обработки сигналов в оптимальном приемнике не содержит никаких вероятностей. Здесь имеются параметры известных реализаций и функций сигналов на его входе. Поэтому данные выражения наглядно демонстрируют все необходимые операции в процессе приема наблюдаемого колебания.

Схема демодулятора, представленная на рис. 8.1, в общем случае содержит M ветвей по числу реализаций сигналов с объединен-

ными входами, куда поступает аддитивная смесь сигнала и помехи. Каждая ветвь состоит из генератора эталонного колебания, перемножителя принимаемого и опорного (эталонного) сигналов и интегратора, с помощью которых вычисляется скалярное произведение

$$(z, S_i) = \int_0^T z(t)S_i(t)dt.$$

Устройство, реализующее данную операцию, принято называть коррелятором, или активным фильтром.

После вычисления в корреляторе скалярного произведения из него вычитается значение $C_i = E_i/2$. Далее в момент времени $t = T$, совпадающий с окончанием посылки сигнала, в решающем устройстве определяется номер ветви с максимальным сигналом, колебание на выходе которой и будет соответствовать переданному сообщению.

Рассматриваемый метод приема сигналов с полностью известными параметрами, как уже отмечалось, принято называть когерентным, а устройство его реализующее — когерентным приемником. В таком приемнике необходимо определять точное значение фазы приходящих сигналов. Это осуществляется либо по принимаемому сигналу со снятием манипуляций, либо по известному пилот-сигналу. Если фаза изменяется медленно и ее значения могут быть предсказаны по предыдущим элементам сигнала, то прием называется квазикогерентным.

8.4. Оптимальный приемник на согласованных фильтрах

Скалярное произведение наблюдаемого случайного процесса $z(t)$ и опорного сигнала $S_i(t)$ можно вычислить не только с помощью коррелятора, но и с использованием пассивного линейного фильтра с постоянными параметрами. Одним из таких фильтров является согласованный фильтр, обладающий такой передаточной функцией $K(j\omega)$, при которой в момент времени $t = T$, т. е. при снятии отсчета, отношение сигнал/шум на его выходе является максимальным.

Найдем выражение для передаточной функции $K(j\omega)$ согласованного фильтра.

Пусть $S(j\omega)$ — комплексный спектр сигнала на входе фильтра. В этом случае спектр на его выходе определяется произведением $S(j\omega)K(j\omega)$. Используя обратное преобразование Фурье, запишем выходной сигнал в момент времени $t = t_0$:

$$U(t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega)K(j\omega)e^{j\omega t_0} d\omega.$$

Пусть помехой является белый шум $n(t)$, с равномерным энергетическим спектром на всех частотах $G(\omega) = N_0/2$. Спектр шума на выходе фильтра определяется выражением

$$G_{\text{вых}}(\omega) = |K(j\omega)|^2 N_0/2.$$

Используя теорему Винера—Хинчина, запишем дисперсию помехи на выходе фильтра:

$$D[n(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\text{вых}}(\omega) d\omega = \frac{N_0}{4\pi} \int_{-\infty}^{\infty} |K(j\omega)|^2 d\omega.$$

Тогда отношение сигнал/шум в момент времени снятия отсчета $t = t_0$ будет иметь следующий вид:

$$q = \frac{|U(t_0)|^2}{D[n(t_0)]} = \frac{1}{2\pi} \frac{\left| \int_{-\infty}^{\infty} S(j\omega) K(j\omega) e^{j\omega t_0} d\omega \right|^2}{\frac{N_0}{2} \int_{-\infty}^{\infty} |K(j\omega)|^2 d\omega}.$$

Чтобы найти значение $K(j\omega)$, при котором значение q в момент $t = t_0$ является максимальным, используем известное соотношение Буняковского—Шварца:

$$\left| \int_{-\infty}^{\infty} x(f) y(f) df \right|^2 \leq \int_{-\infty}^{\infty} |x(f)|^2 df \int_{-\infty}^{\infty} |y(f)|^2 df,$$

где $x(f)$, $y(f)$ — любые комплексные функции.

При этом знак равенства имеет место только в том случае, если $x(f) = Cy^*(f)$, $C = \text{const}$, $y^*(f)$ — функция, комплексно сопряженная с $y(f)$.

Предположим теперь, что

$$y(f) = S(j\omega) e^{j\omega t_0}; \quad y^*(f) = S^*(j\omega) e^{-j\omega t_0}$$

(с учетом того, что $f = \frac{\omega}{2\pi}$).

Тогда после подстановки получим

$$q = \frac{1}{2\pi} \frac{\int_{-\infty}^{\infty} |S(j\omega) e^{j\omega t_0}|^2 d\omega \int_{-\infty}^{\infty} |K(j\omega)|^2 d\omega}{\frac{N_0}{2} \int_{-\infty}^{\infty} |K(j\omega)|^2 d\omega} \leq \frac{1}{\pi N_0} \int_{-\infty}^{\infty} |S(j\omega)|^2 d\omega = \frac{2E}{N_0},$$

где $E = \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(j\omega)|^2 d\omega$ — энергия сигнала, определяемая согласно равенству Парсеваля.

Из полученного выражения видно, что максимальное значение q на выходе фильтра не зависит от формы сигнала, а целиком определяется отношением его энергии к спектральной плотности мощности шума. Отношение q максимизируется, если передаточная функция фильтра

$$K(j\omega) = CS^*(j\omega)e^{-j\omega t_0},$$

где C — некоторая постоянная, характеризующая усиление фильтра; $S^*(j\omega)$ — функция, комплексно сопряженная со спектральной плотностью сигнала, поступающего на вход фильтра.

Спектральную плотность входного сигнала и передаточную функцию фильтра можно записать в виде

$$S(j\omega) = |S(j\omega)|e^{j\varphi_s(\omega)}; K(j\omega) = |K(j\omega)|e^{-j\varphi(\omega)}.$$

Используя выражение для передаточной функции фильтра, получим

$$|K(j\omega)| = C|S(j\omega)|; \varphi(\omega) = -(\varphi_s(\omega) + \omega t_0),$$

откуда следует, что амплитудно-частотная характеристика согласованного фильтра пропорциональна амплитудно-частотному спектру входного сигнала.

Фазочастотная характеристика равна сумме фазочастотного спектра сигнала, взятого с обратным знаком, и составляющей, обусловленной задержкой $-\omega t_0$. В результате согласованный фильтр выделяет наиболее интенсивные участки спектра и еще больше ослабляет слабые участки. При этом форма сигнала на выходе фильтра изменяется. Однако это несущественно, так как задача фильтра состоит не в точном воспроизведении вида сигнала, а в наилучшем его выделении на фоне шума. Значительная роль при этом отводится фазочастотной характеристике фильтра, которая компенсирует все фазовые сдвиги спектральных составляющих входного сигнала. Таким образом, все составляющие сигнала в момент окончания посылки $t = t_0$, складываясь, определяют пик выходного сигнала.

Найдем теперь импульсную реакцию согласованного фильтра, учитывая, что она связана с передаточной функцией посредством обратного преобразования Фурье:

$$\begin{aligned} g(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} K(j\omega)e^{j\omega t} d\omega = \frac{C}{2\pi} \int_{-\infty}^{\infty} S^*(j\omega)e^{-j\omega t_0} e^{j\omega t} d\omega = \\ &= \frac{C}{2\pi} \int_{-\infty}^{\infty} S^*(j\omega)e^{-j\omega(t_0-t)} d\omega. \end{aligned}$$

Заметим, что $S^*(j\omega) = S(-j\omega)$, тогда после замены ω на $\omega' = -\omega$ получим

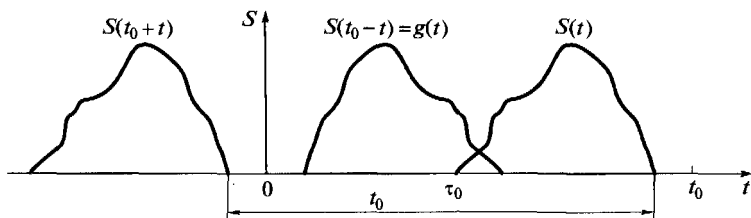


Рис. 8.2. Вид импульсной реакции согласованного фильтра

$$g(t) = \frac{C}{2\pi} \int_{-\infty}^{\infty} S(j\omega') e^{j\omega'(t_0 - t)} d\omega' = CS(t_0 - t).$$

Последнее равенство следует из того, что входной сигнал связан преобразованием Фурье с комплексной спектральной плотностью:

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

Таким образом, импульсная реакция согласованного фильтра $g(t) = CS(t_0 - t)$ целиком определяется формой сигнала. Ее графическое представление показано на рис. 8.2, где для упрощения принята постоянная $C = 1$. Здесь также показан сигнал $S(t)$, появившийся в момент времени $t = \tau_0$, а также сигнал $S(t_0 + t)$, который появится раньше на время t_0 , чем сигнал $S(t)$. График функции $S(t_0 - t)$, являющийся зеркальным отображением графика функции $S(t_0 + t)$ относительно оси ординат, соответствует графическому представлению импульсной реакции согласованного фильтра.

Отметим, что для физической реализуемости фильтра необходимо и достаточно, чтобы $g(t) = 0$ при $t < 0$. Для финитных сигналов, поступающих на вход фильтра в момент $t = 0$ и заканчивающихся в момент времени T , согласованные фильтры вполне реализуемы, если постоянная времени t_0 (момент снятия отсчета) удовлетворяет условию

$$t_0 - T, \text{ или } t_0 \geq T.$$

Следовательно, момент снятия отсчета должен совпадать с моментом окончания посылки сигнала.

Приведем основные свойства согласованных фильтров.

1. Из всех возможных линейных фильтров согласованный фильтр позволяет получить на выходе максимально возможное значение отношения энергии элемента сигнала к спектральной плотности мощности шума ($q = 2E/N_0$). Причем это значение не

зависит от формы сигнала. (Данное свойство было принято выше за определение согласованных фильтров.)

2. Согласованный фильтр инвариантен (независим) относительно момента времени поступления сигнала, т.е. времени задержки. Иными словами, для этого фильтра безразлично, когда на его вход поступит сигнал. В любом случае, если сигнал согласован по форме, на выходе фильтра будет получен максимум отношения сигнал/помеха. В отличие от согласованного фильтра коррелятор неинвариантен к задержке, и для максимизации сигнала на его выходе необходимо иметь точную тактовую синхронизацию. Заметим также, что напряжения на выходах обоих устройств совпадают только в момент окончания посылок сигнала. В остальные моменты они различны.

3. Согласованный фильтр является устройством, которое вычисляет функцию взаимной корреляции между принимаемым сигналом $\underline{U}(t)$ и опорным (ожидаемым, эталонным) сигналом $S_i(t)$ при $i=1, M$. Выходное напряжение $U(t_0)$ согласованного фильтра, как и любого линейного устройства, определяется интегралом свертки, т.е. имеет вид

$$U(t_0) = \int_{-\infty}^{\infty} U(\tau)g(t_0 - \tau)d\tau = C \int_{-\infty}^{\infty} U(\tau)S(\tau)d\tau = CB(0) = C(U(t)S(t)).$$

Это выражение с точностью до постоянного множителя C является функцией взаимной корреляции при нулевом значении аргумента или скалярным произведением принимаемой смеси и сигнала $S(t)$.

Благодаря этому свойству, оптимальный приемник можно выполнять на базе согласованных фильтров, как показано на рис. 8.3, из которого видно, что согласованный или пассивный фильтр в отличие от активного фильтра (коррелятора) заменяет сразу три элемента в приемнике: генератор опорного колебания, перемножитель и интегратор.

Казалось бы схема приемника стала проще, но на самом деле сложности, возникающие при реализации оптимальных приемников на согласованных фильтрах, оказываются весьма существенными. Сложность реализации заключается в том, что отсчет на выходе фильтра так же, как в корреляционном когерентном приемнике, должен производиться «с точностью до начальной фазы». Однако оказывается, что точность в этом случае должна быть выше, чем на выходе активного фильтра. При этом в последнем достаточно, чтобы неточность была мала по сравнению с длительностью посылки сигнала, а при согласованной фильтрации она должна быть меньше, чем период высокочастотного заполнения радиопульсов, поступающих на вход оптимального приемника.

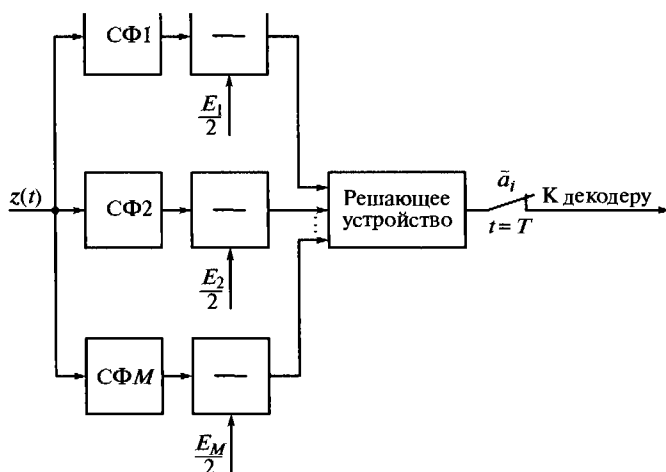


Рис. 8.3. Схема оптимального по критерию Котельникова когерентного приемника на основе согласованных фильтров

Указанные сложности реализации оптимальных приемников соизмеримы со сложностью реализации когерентного опорного колебания в активном фильтре.

8.5. Методы реализации согласованных фильтров

Согласованный фильтр финитного сигнала произвольного вида $S(t)$ можно реализовать на основе линии задержки на время T , равное длительности сигнала, с большим числом отводов. Это следует из того, что сигнал, заданный в интервале $t \in [0, T]$, легко представить в виде суммы последовательности, состоящей из $n = T/\Delta t$ прямоугольных импульсов малой длительности Δt с вы-

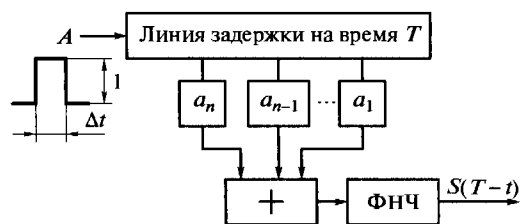


Рис. 8.4. Реализация фильтра, согласованного с сигналом произвольной формы

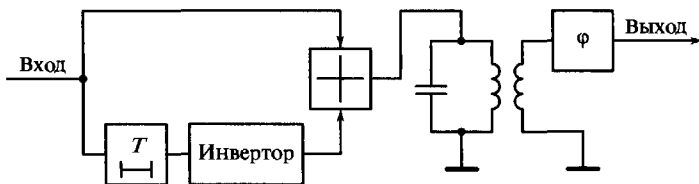


Рис. 8.5. Реализация фильтра, согласованного с прямоугольным радиоимпульсом

сотой a_k (где $k = \overline{1, n}$). Длительность импульса выбирается согласно теореме отсчетов по формуле $\Delta t = 1/(2F_c)$, где F_c — ширина спектра сигнала.

Реализация фильтра, согласованного с сигналом произвольной формы, показана на рис. 8.4.

Если на вход такого фильтра в точку A в момент времени $t = 0$ подать прямоугольный импульс с длительностью Δt и единичной амплитудой, то на отводах линии задержки появятся точно такие же импульсы, но разнесенные по времени с интервалом Δt . После прохождения через взвешивающие устройства a_k (где $k = \overline{1, n}$) эти импульсы поступят на вход фильтра низких частот (ФНЧ). Взвешивающие блоки представляют собой аттенюаторы или усилители с коэффициентом усиления a_k (с учетом знака коэффициента a_k). ФНЧ предназначен для подавления составляющих, не входящих в спектр сигнала, но существующих в схеме, поскольку на ее вход подан прямоугольный импульс.

Рассмотренная схема представляет собой линейный фильтр, называемый трансверсальным с импульсной реакцией $S(T - t)$.

В качестве линий задержки обычно применяют несколько последовательно соединенных LC -фильтров, имеющих малое затухание, большую надежность и относительно небольшие габаритные размеры.

Существуют и другие способы реализации фильтров, согласованных с сигналом заданной формы.

Для примера рассмотрим фильтр, согласованный с прямоугольным радиоимпульсом $S(t) = U_c \cos(\omega_0 t + \varphi)$, где $t \in [0, T]$, схема которого представлена на рис. 8.5.

Фильтр состоит из колебательного контура высокой добротности, практически не имеющего потерь и настроенного на частоту $\omega_0 = 2\pi f_0$, а также фазовращателя, сдвигающего фазу колебаний контура на φ_0 . Линия задержки на время T и инвертор обеспечивают гашение колебаний фильтра вне интервала $(0, T)$. Импульсная реакция фильтра с учетом сдвига фаз

$$g(t) = U_c \cos(\omega_0 t - \omega_0 T - \varphi_0) = U_c \cos(\omega_0(T - t) + \varphi_0) = U_c S(T - t).$$

8.6. Помехоустойчивость оптимального когерентного приема двоичных сигналов

Определим потенциальную помехоустойчивость оптимального приемника при передаче двоичных сигналов по каналу с постоянными параметрами в условиях аддитивного гауссовского белого шума.

Предположим, что передается сигнал $S_0(t)$, соответствующий символу «0», и на вход приемника поступает аддитивная смесь $z(t) = S_0(t) + n(t)$. Ошибочное решение в этом случае будет принято, если

$$\int_0^T z(t)S_0(t)dt - \frac{E_0}{2} < \int_0^T z(t)S_1(t)dt - \frac{E_1}{2},$$

где $E_i = \int_0^T S_i^2(t)dt$; $i = 0, 1$.

Запишем данное соотношение в виде

$$\int_0^T [S_0(t) + n(t)][S_0(t) - S_1(t)]dt < \frac{1}{2} \int_0^T [S_0^2(t) - S_1^2(t)]dt,$$

или

$$\int_0^T n(t)[S_0(t) - S_1(t)]dt < -\int_0^T S_0^2(t)dt + \int_0^T S_0(t)S_1(t)dt + \frac{1}{2} \int_0^T [S_0^2(t) - S_1^2(t)]dt,$$

откуда

$$\int_0^T n(t)[S_0(t) - S_1(t)]dt < -\frac{1}{2} \int_0^T [S_0(t) - S_1(t)]^2 dt.$$

Введем обозначение $S_\Delta(t) = [S_0(t) - S_1(t)]$, тогда эквивалентная энергия сигнала

$$E_s = \int_0^T S_\Delta^2(t)dt; \quad X = \int_0^T n(t)S_\Delta(t)dt,$$

следовательно,

$$X < -\frac{1}{2} E_s.$$

Несложно показать, что если сообщения «0» и «1» передаются равновероятно, то при передаче символа «1» получим это же со-

отношение. Иными словами, вероятности ошибки в обоях случаях равны: $P(1|0) = P(0|1) = P$, т. е. образованный в данном случае двоичный дискретный канал является симметричным.

Величина X имеет гауссовское распределение, так как она получена в результате линейных преобразований над случайным гауссовским процессом $n(t)$. Ее математическое ожидание

$$M\{X\} = M \left\{ \int_0^T n(t) S_{\Delta}(t) dt \right\} = \int_0^T M\{n(t)\} S_{\Delta}(t) dt = 0,$$

поскольку $M\{n(t)\} = 0$.

Определим дисперсию величины X :

$$\begin{aligned} D\{X\} &= M \left\{ \left[\int_0^T n(t) S_{\Delta}(t) dt \right]^2 \right\} = M \left\{ \int_0^T \int_0^T n(t) n(t') S_{\Delta}(t) S_{\Delta}(t') dt dt' \right\} = \\ &= \int_0^T \int_0^T S_{\Delta}(t) S_{\Delta}(t') M\{n(t) n(t')\} dt dt'. \end{aligned}$$

Напомним, что функция корреляции белого шума вычисляется через дельта-функцию, т. е.

$$B(\tau) = M\{n(t) n(t')\} = \frac{N_0}{2} \delta(t - t') = \frac{N_0}{2} \delta(\tau).$$

Тогда

$$D\{X\} = \int_0^T \int_0^T \frac{N_0}{2} \delta(t - t') S_{\Delta}(t) S_{\Delta}(t') dt dt' = \frac{N_0}{2} \int_0^T S_{\Delta}^2(t) dt = \frac{N_0 E_{\Delta}}{2}.$$

Теперь можно найти вероятность ошибки:

$$\begin{aligned} P &= P \left\{ x < -\frac{E_{\Delta}}{2} \right\} = \int_{-\infty}^{-E_{\Delta}/2} \omega(x) dx = \\ &= \int_{-\infty}^{-E_{\Delta}/2} \frac{1}{\sqrt{2\pi D\{x\}}} \exp \left(-\frac{x^2}{2D\{x\}} \right) dx = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{\infty} e^{-t^2/2} dt = Q(x_1). \end{aligned}$$

В данном выражении выполнена замена переменной $t = -x / \sqrt{D\{x\}}$ и введено обозначение $x_1 = -\frac{E_{\Delta}}{2\sqrt{D\{x\}}} = \sqrt{\frac{E_{\Delta}}{2N_0}}$.

Учитывая, что $Q(x_1) = 1 - F(x_1)$, где $F(x_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_1} e^{-t^2/2} dt$ — интеграл вероятностей, для которого составлены подробные таблицы (см. приложение), окончательно выражение вероятности

ошибочного приема при оптимальной процедуре обработки запишем в виде

$$P = 1 - F\left(\sqrt{\frac{E_s}{2N_0}}\right),$$

где $E_s = d^2(S_0(t), S_1(t)) = \int_0^T [S_0(t) - S_1(t)]^2 dt$ — квадрат расстояния между сигналами $S_0(t)$ и $S_1(t)$, называемый эквивалентной энергией сигналов.

Из этого выражения видно, что помехоустойчивость оптимального когерентного приемника двоичных сигналов в условиях аддитивного белого гауссовского шума определяется отношением энергии сигнала к спектральной плотности мощности шума. Следовательно, помехоустойчивость выше у системы, эквивалентная энергия сигналов которой больше независимо от формы применяемых сигналов.

При расчетах вероятности ошибки по полученному выражению следует иметь в виду, что интеграл вероятности принимает следующие значения: $F(-\infty) = 0$; $F(0) = 0,5$; $F(\infty) = 1$; $F(-x) = 1 - F(x)$.

Иногда в расчетах вместо функции $F(x)$ используют другие функции, связанные с ней.

8.7. Анализ помехоустойчивости оптимального приема двоичных сигналов

Конкретизируем полученные ранее выражения для определения вероятности ошибочного приема для двоичных сигналов с различными видами дискретной модуляции (манипуляции). Рассмотрим наиболее часто применяемые сигналы с амплитудной, частотной, фазовой и относительной фазовой модуляциями (манипуляциями) и рассчитаем их помехоустойчивость.

8.7.1. Сигналы с дискретной амплитудной модуляцией

В данном случае один из двоичных сигналов, например систему сигналов $S_1(t) = 0$, $S_0(t) \neq 0$, принято называть системой с пассивной паузой. Тогда принимаемые сигналы можно описать выражением

$$S_0(t) = U_c \cos(\omega_0 t + \varphi); t \in [0, T]; S_1(t) = 0,$$

где U_c — амплитуда сигнала; ω — круговая частота; φ — начальная фаза.

(Эти величины полагаются полностью известными.)

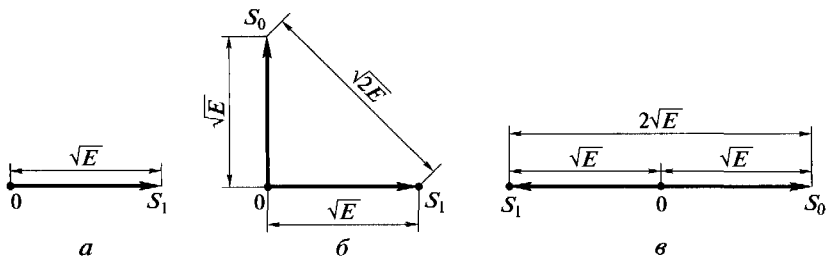


Рис. 8.6. Векторное представление для определения эквивалентной энергии сигналов с дискретными видами модуляции:

a — амплитудной; *б* — частотной; *в* — фазовой

Векторное представление сигналов с амплитудной модуляцией показано на рис. 8.6, *a*.

Квадрат расстояния между двумя сигнальными точками, или эквивалентную энергию АМ-сигналов, определим в виде

$$d^2(S_0(t), S_1(t)) = E_3 = E = \int_0^T S_0^2(t) dt.$$

Следовательно, вероятность ошибочного приема АМ-сигнала можно найти следующим образом:

$$P_{AM} = 1 - F\left(\sqrt{\frac{E_3}{2N_0}}\right) = 1 - F\left(\sqrt{\frac{h^2}{2}}\right),$$

где $h^2 = E/N_0$ — отношение энергии сигнала на входе приемника к спектральной плотности мощности белого шума.

8.7.2. Сигналы с дискретной частотной модуляцией

В отличие от АМ-сигналов ЧМ-сигналы называются сигналами с активной паузой и описываются выражениями

$$S_0(t) = U_c \cos(\omega_0 t + \varphi_0); t \in [0, T];$$

$$S_1(t) = U_c \cos(\omega_1 t + \varphi_1).$$

Кроме того, они являются ортогональными, а это означает, что их скалярное произведение равно нулю, т. е.

$$(S_0, S_1) = \frac{1}{T} \int_0^T S_0(t) S_1(t) dt = 0.$$

Пусть $\omega_0 = 2\pi k_0/T$, $\omega_1 = 2\pi k_1/T$, где k_0 и k_1 — целые числа, а φ_0 и φ_1 принимают любые значения. В этом случае можно записать

$$\begin{aligned}
& \frac{1}{T} \int_0^T S_0(t) S_1(t) dt = \frac{U_c^2}{T} \int_0^T \cos(\omega_0 t + \varphi_0) \cos(\omega_1 t + \varphi_1) dt = \\
& = \frac{U_c^2}{2T} \left\{ \int_0^T \cos[(\omega_0 + \omega_1)t + \varphi_0 + \varphi_1] + \cos[(\omega_0 - \omega_1)t + \varphi_0 - \varphi_1] dt \right\} = \\
& = \frac{U_c^2}{2T} \left\{ \int_0^T \cos \left[2\pi \frac{k_0 + k_1}{T} t + \varphi_0 + \varphi_1 \right] dt + \right. \\
& \quad \left. + \int_0^T \cos \left[2\pi \frac{k_0 - k_1}{T} t + \varphi_0 - \varphi_1 \right] dt \right\} = 0,
\end{aligned}$$

так как подынтегральные выражения содержат целое число периодов косинусоид.

Равенство обоих интегралов нулю доказывает ортогональность ЧМ-сигналов.

Квадрат расстояния между сигнальными точками, или эквивалентную энергию сигналов, определим в виде

$$\begin{aligned}
d^2(S_0(t), S_1(t)) = E_s &= \int_0^T [S_0(t) - S_1(t)]^2 dt = \int_0^T S_0^2(t) dt + \int_0^T S_1^2(t) dt - \\
& - 2 \int_0^T S_0(t) S_1(t) dt = E_0 + E_1 = 2E,
\end{aligned}$$

поскольку последний интеграл по условию ортогональности равен нулю, а энергии каждой реализации равны друг другу.

В двумерном пространстве ЧМ-сигналы можно представить в виде векторов, расположенных относительно друг друга под углом $\pi/2$. Именно в этом случае их скалярное произведение равно нулю (рис. 8.6, б).

Тогда вероятность ошибки ЧМ-сигнала

$$P_{\text{ЧМ}} = 1 - F\left(\sqrt{\frac{E}{N_0}}\right) = 1 - F(\sqrt{h^2}).$$

8.7.3. Сигналы с дискретной фазовой модуляцией

Так же как и ЧМ-сигналы, фазомодулированные сигналы являются сигналами с активной паузой, поскольку каждой их реализации соответствует двоичный информационный символ «0» или «1».

В простейшем случае, как уже отмечалось ранее, ФМ-сигнал образуется посредством скачкообразного изменения фазы несущего колебания на 180° . Получаемое в результате колебание пред-

ставляет собой последовательность информационных двухполярных посылок, умноженных на гармоническое несущее колебание. Эти сигналы можно записать в виде

$$S_0(t) = U_c \cos(\omega_c t + \varphi); t \in [0, T];$$

$$S_1(t) = U_c \cos(\omega_c t + \varphi + \pi) = -U_c \cos(\omega_c t + \varphi),$$

откуда следует, что $S_1(t) = -S_0(t)$. Поэтому такие сигналы называются противоположными, а их векторное представление показано на рис. 8.6, в.

Среди всех рассмотренных сигналов фазомодулированные сигналы имеют наибольшее расстояние между концами векторов в векторном пространстве. Поэтому естественно ожидать, что они будут наиболее различимыми, а следовательно, самыми помехоустойчивыми.

Квадрат расстояния между сигнальными точками, или эквивалентную энергию сигналов, определим в виде

$$d^2(S_0(t), S_1(t)) = E_s = \int_0^T [S_0(t) - S_1(t)]^2 dt = 4 \int_0^T S_0^2(t) dt = 4E.$$

Следовательно, вероятность ошибки ФМ-сигнала

$$P_{\text{ФМ}} = 1 - F\left(\sqrt{\frac{4E}{2N_0}}\right) = 1 - F\left(\sqrt{2h^2}\right),$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Если сравнить вероятности ошибок для ФМ-, АМ- и ЧМ-сигналов, то нетрудно увидеть, что последние занимают промежуточное место между двумя первыми. При этом энергия при переходе от ФМ-сигналов (противоположных) к ЧМ-сигналам (ортогональным) увеличивается в два раза, что эквивалентно 3 дБ.

Сравнение ЧМ- и АМ-сигналов показывает, что для обеспечения одинаковой верности приема при использовании ЧМ-сигналов потребуется энергии затратить в два раза меньше.

Несмотря на то что ФМ-сигналы являются самыми помехоустойчивыми практическая реализация их когерентного приема связана с серьезными трудностями, возникающими при формировании опорного (эталонного) колебания, синфазного с принимаемым сигналом.

Как уже отмечалось, в спектре ФМ-сигнала при изменениях фазы на 180° в отличие от сигналов с амплитудной и частотной модуляциями отсутствует колебание на несущей частоте f_c . Следовательно, из принимаемого сигнала нельзя выделить необходи-

мое эталонное напряжение для подстройки генератора опорного колебания в приемнике. Для восстановления несущей частоты можно использовать нелинейные преобразования, основанные на методах, предложенных отечественными учеными А. А. Пистолькорсом (1933), В. И. Сифоровым (1935), а также американским специалистом Д. Костасом (1956) и др.

Рассмотрим метод Пистолькорса как наиболее простой. Сначала частота принимаемого колебания удваивается, что приводит к снятию модуляции. Это следует из того факта, что при передаче сигнала $S_0(t) = U_c \cos(\omega_c t + \varphi)$ получаем колебание вида $U_c \cos 2(\omega_c t + \varphi)$, а при передаче сигнала $S_1(t) = U_c \cos(\omega_c t + \varphi + \pi)$ — колебание вида $U_c \cos 2(\omega_c t + \varphi + \pi) = U_c \cos 2(\omega_c t + \varphi)$.

После фильтрации и деления частоты на два выделяется колебание с частотой f_c и малым уровнем помех, которое можно использовать для подстройки опорного генератора. Однако здесь возникает принципиальная сложность, вызванная неоднозначностью фазы после деления частоты. С равной вероятностью фаза может принять значение 0 и 180° , что приводит к явлению обратной работы, когда символы «0» принимаются как «1», и наоборот. Причем этот эффект практически неустраним, так как перескок фазы происходит случайно под воздействием помех в канале связи. Данный недостаток присущ всем без исключения рассмотренным методам. Поэтому ФМ-сигналы в «классическом» варианте применяются при формировании оценки фазы по пилот-сигналу.

8.7.4. Сигналы с относительной фазовой модуляцией

Сигналы с относительной фазовой модуляцией позволяют полностью устранить явление обратной работы. В этом случае передаваемая информация вкладывается не в абсолютное значение фазы каждой посылки сигнала, а определяется разностью фаз последующей и предыдущей посылок.

Данный способ передачи информации, предложенный отечественным ученым Н. Г. Петровичем (1954), заключается в том, что каждый информационный символ «0» передается повторением той реализации, которая являлась предыдущим элементом, а символ «1» — изменением этой реализации на противоположную.

Систему ОФМ-сигналов можно рассматривать как сочетание кодирования и дискретной фазовой модуляции.

Иными словами, сначала производится перекодирование символов $a_k = (0, 1)$ (где $k = 1, 2, \dots$) по правилу $a'_k = a'_{k-1}$ при $a_k = 0$ и $a'_k = 1 - a'_{k-1}$ при $a_k = 1$. При этом посылка a_0 не несет информации и выбирается произвольно, а необходима она лишь для начала процесса перекодирования.

После этой операции осуществляется обычная фазовая модуляция, где в качестве управляющей (манипулирующей) применяется перекодированная последовательность символов $a'_k = (0, 1)$, где $k = 0, 1, \dots$

Приемник полностью известных ОФМ-сигналов реализуется аналогично оптимальному когерентному приемнику ФМ-сигналов. Различие заключается лишь в том, что перед принятием решения о переданном сигнале производится обратное перекодирование принимаемой последовательности

$$\tilde{A}' = (\tilde{a}'_1, \tilde{a}'_2, \dots, \tilde{a}'_k, \dots)$$

в соответствии с правилом

$$\tilde{a}'_k = \tilde{a}'_k \oplus \tilde{a}'_{k-1},$$

где \oplus — знак сложения по модулю два.

Эта операция реализуется посредством сравнения напряжений принимаемой посылки с предыдущей, задержанной на время T , которое равно длительности элемента сигнала. При совпадении полярностей сигналов принимается решение о передаче символа «0», в противном случае регистрируется символ «1». Данный вид приема принято называть методом сравнения полярностей.

Схема устройства, реализующего метод сравнения полярностей, представлена на рис. 8.7.

Опорный тракт обычно выполняется по схеме Пистолькорса с делением, фильтрацией и умножением частоты на два. При этом в отличие от фазовой модуляции в данном случае скачок фазы не опасен, так как ошибочно будет зарегистрирован лишь один элемент, а остальные принимаются верно при условии, что в канале отсутствуют помехи.

Сказанное иллюстрирует табл. 8.1, где для сравнения показан также принцип ФМ. Здесь изменение фазы обозначается стрел-

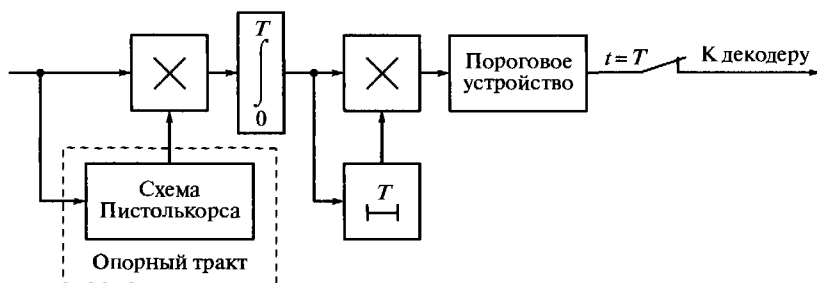


Рис. 8.7. Схема оптимального когерентного приемника сигналов с относительной фазовой модуляцией

Параметры сигналов	ОФМ	ФМ
Информация a_k	0111001010	0111001010
Перекодированные посылки a'_k	00101110011	—
Фаза сигнала	↑↑↓↑↓↓↑↑↑↓	↑↓↓↑↑↓↑↓↑
Фаза на приеме	↑↑↓↑↑↑↑↓↑↑ *	↑↓↓↑↓↓↑↓↑↓ *
Принятые посылки a_k	0110001010	0110110101

ками: стрелка, направленная вверх, соответствует фазе 0, а стрелка, направленная вниз — фазе 180° . Звездочками обозначены моменты перескока фазы, начиная с которого прием ФМ-сигнала производится с ошибками (обратная работа). В случае ОФМ только один символ ОФМ-сигнала принимается неверно, а остальные принимаются безошибочно.

Оценим потенциальную помехоустойчивость когерентного приемника ОФМ-сигналов в условиях аддитивного белого шума. Ошибка в приеме возникает при условии, что посылка k принята неверно, а $k - 1$ — правильно, и наоборот. Поскольку в рассматриваемом случае ошибки возникают независимо, вероятность каждого из сочетаний этих событий составляет $P_{\text{ФМ}}(1 - P_{\text{ФМ}})$, где $P_{\text{ФМ}}$ — вероятность ошибочного приема ФМ-сигналов.

Окончательно вероятность ошибки приема ОФМ-сигналов запишем в виде

$$P_{\text{ОФМ}} = 2P_{\text{ФМ}}(1 - P_{\text{ФМ}}) = 2 \left[1 - F(\sqrt{2h^2}) \right] F(\sqrt{2h^2}) \approx 2P_{\text{ФМ}}.$$

Полученное значение оказывается примерно вдвое больше, чем вероятность ошибки для ФМ-сигналов. Однако это и есть необходимая «плата» за возможность устранения случайных перескоков фазы в канале, приводящих к появлению обратной работы.

8.8. Некогерентный прием двоичных сигналов с дискретной частотной модуляцией

Рассмотрим теперь каналы передачи информации, в которых фаза сигналов изменяется достаточно быстро, и получить ее точную оценку не удастся, а следовательно, невозможно реализовать когерентный прием.

В таких ситуациях применяют метод обработки, при котором предполагается, что начальная фаза сигнала не известна и может принимать произвольное значение в интервале $(0, 2\pi)$. Данный метод получил название некогерентного приема.

Как и ранее, полагаем, что сообщения от источника передаются равновероятно, помехой является белый гауссовский шум $n(t)$ с нулевыми средними значениями и спектральной плотностью мощности N_0 . Тогда при неизвестной фазе сигнала на входе приемника получим колебание вида

$$z(t) = \mu U_{ci} \cos(\omega t + \theta) + n(t); \quad i = \overline{1, M}.$$

С использованием известного тригонометрического тождества данное выражение можно представить в следующем виде:

$$z(t) = \mu U_{ci} \cos \omega t \cos \theta - \mu U_{ci} \sin \omega t \sin \theta + n(t),$$

или

$$z(t) = \mu(S_i(t) \cos \theta - \widehat{S}_i(t) \sin \theta) + n(t),$$

где μ — известный постоянный коэффициент передачи канала; $\widehat{S}_i(t)$ — преобразование Гильберта от $S_i(t)$; θ — случайная начальная фаза сигнала, равномерно распределенная в интервале $(0, 2\pi)$ и имеющая плотность вероятностей $\omega(\theta) = 1/(2\pi)$.

Для синтеза некогерентного приемника используем выражение (8.21) для определения функционала отношения правдоподобия и после преобразования запишем условный функционал отношения правдоподобия (при условии, что значение θ фиксировано) в следующем виде:

$$\Lambda_i[z(t)/\theta] = \exp \left\{ -\frac{2\mu}{N_0} \int_0^T z(t) [S_i(t) \cos \theta + \widehat{S}_i(t) \sin \theta] dt - \frac{\mu^2}{N_0} \int_0^T [S_i(t) \cos \theta + \widehat{S}_i(t) \sin \theta]^2 dt \right\}.$$

Отметим также, что выражение под знаком второго интеграла представляет собой энергию E_{ni} сигнала на выходе передатчика. Поэтому отношение сигнал/помеха на входе приемника

$$h_i^2 = \mu^2 E_{ni} / N_0 = E_i / N_0.$$

Введем следующие обозначения:

$$y_i = \mu \int_0^T z(t) S_i(t) dt; \quad \widehat{y}_i = \mu \int_0^T z(t) \widehat{S}_i(t) dt.$$

Тогда искомый функционал можно записать в виде

$$\Lambda_i [z(t)/\theta] = \exp \left\{ \frac{2}{N_0} (y_i \cos \theta + \widehat{y}_i \sin \theta) \right\} e^{-h^2},$$

или

$$\Lambda_i [z(t)/\theta] = \exp(-h^2) \exp \left\{ \frac{2}{N_0} V_i \cos(\theta - \varphi_i) \right\},$$

где

$$V_i = \sqrt{y_i^2 + \widehat{y}_i^2}; \quad \varphi_i = \operatorname{arctg} \frac{\widehat{y}_i}{y_i}.$$

Определим безусловный функционал отношения правдоподобия посредством усреднения условного функционала отношения правдоподобия по фазе θ , т.е. определим математическое ожидание случайной величины $\Lambda_i[z(t)/\theta_i]$:

$$\begin{aligned} \Lambda_i [z(t)] &= \int_0^{2\pi} \omega(\theta) \Lambda_i [z(t)/\theta] d\theta = \\ &= \frac{1}{2\pi} e^{-h^2} \int_0^{2\pi} \exp \left\{ \frac{2}{N_0} V_i \cos(\theta - \varphi_i) \right\} d\theta = \\ &= e^{-h^2} I \left(\frac{2V_i}{N_0} \right), \end{aligned}$$

где $I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x \cos(\theta - \varphi_i)} d\theta$ — модифицированная функция Бесселя нулевого порядка.

После логарифмирования приходим к следующему правилу некогерентного оптимального приема сигналов со случайной фазой: посылка под номером i регистрируется, если

$$\ln I_0(2V_i/N_0) - h^2_i > \ln I_0(2V_j/N_0) - h^2_j; \quad \forall j \neq i; \quad j = \overline{1, M}.$$

Схема, реализующая данную процедуру, получила название квадратурной. В ней выполняются практически все указанные ранее операции в процессе приема сигналов. Такая схема для частного случая, т.е. для двоичной системы ЧМ-сигналов, представлена на рис. 8.8.

Из приведенной схемы видно, что приемник содержит две ветви (по числу реализаций сигнала). С помощью генераторов и фазовращателей на $\pi/2$ (преобразователей Гильберта) формируются синфазные и квадратурные составляющие $S_i(t)$, $\widehat{S}_i(t)$ (где $i = 0, 1$), поступающие на множители, на другие входы которых подается наблюдаемое колебание $z(t)$. После перемножения и инте-

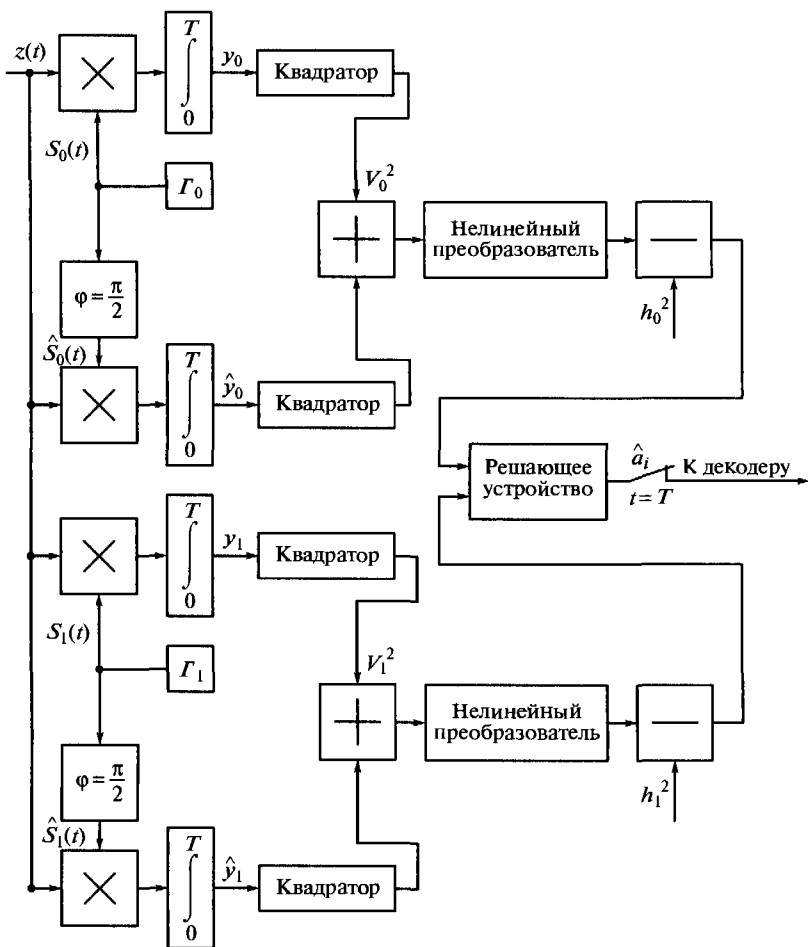


Рис. 8.8. Схема оптимального по критерию Котельникова некогерентного приемника двоичных сигналов с дискретной частотной модуляцией на основе корреляторов (фаза сигнала неизвестна)

рирования формируются переменные y_i , \hat{y}_i , а затем переменные V_i^2 , которые после прохождения нелинейных безынерционных преобразователей подаются на пороговые устройства, где из них вычитаются постоянные h_i^2 ($i = 0, 1$). Затем по наибольшему из напряжений на выходах ветвей принимается решение о переданном сигнале $S_i(t)$.

Если для передачи информации использовать сигналы с одинаковой энергией, учитывая, что при $x > 0$ функция $\ln I_0(x)$ моно-

тонно возрастает, то процедура обработки будет иметь вид (знак неравенства соответствует передаче i -го сигнала):

$$V_i > V_j; \forall j \neq i, j = \overline{1, M}.$$

В схеме приемника в этом случае не требуются блоки вычитания и нелинейные преобразователи. При использовании двоичных ЧМ-сигналов можно записать

$$V_0 > V_1.$$

Если это неравенство выполняется, то регистрируется символ «0», а если не выполняется — символ «1».

Отметим, что некогерентный приемник можно реализовать на основе согласованных фильтров. Схема такого приемника представлена на рис. 8.9. Каждая из ветвей приемника содержит согласованный фильтр и детектор, выделяющий огибающую на выходе фильтра. При снятии отсчета напряжения сравниваются, и по наибольшему из них принимается решение о переданном символе.

Проанализируем теперь помехоустойчивость рассмотренного метода приема при передаче двоичных сигналов. Например, при передаче сигнала $S_0(t)$ ошибочное решение будет принято, если шум превосходит сигнал, т.е. $V_0 < V_1$. Тогда вероятность этого события

$$P(V_0 < V_1) = \int_{V_0}^{\infty} \omega(V_1) dV_1.$$

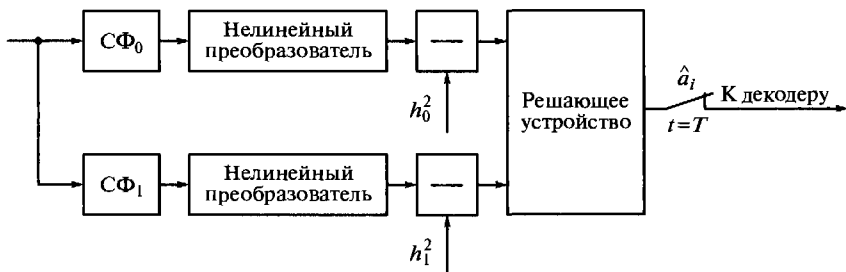


Рис. 8.9. Схема оптимального по критерию Котельникова некогерентного приемника двоичных сигналов с дискретной частотной модуляцией на основе согласованных фильтров (фаза сигнала неизвестна)

Здесь

$$V_0 = \sqrt{(\hat{\xi}_0 + E \cos \theta)^2 + (\hat{\xi}_0 + E \sin \theta)^2};$$

$$V_1 = \sqrt{\hat{\xi}_1^2 + \hat{\xi}_1^2};$$

$$\xi_i = \mu \int_0^T n(t) S_i(t) dt; \hat{\xi}_i = \mu \int_0^T n(t) \hat{S}_i(t) dt; i = 0, 1.$$

При этом величины ξ_i и $\hat{\xi}_i$ являются гауссовскими с нулевыми математическими ожиданиями и дисперсиями

$$D\{\xi_i\} = D\{\hat{\xi}_i\} = 0,5 N_0 E.$$

Рассмотрим далее случай, когда сигналы являются ортогональными в усиленном смысле:

$$\int_0^T S_i(t) S_j(t) dt = 0, \int_0^T \hat{S}_i(t) S_j(t) dt = 0; \forall j \neq i; i, j = \overline{1, M},$$

т.е. их скалярное произведение, как уже отмечалось ранее, равно нулю и в том случае, если один из них заменен сопряженным ему сигналом. К числу таких сигналов относятся и сигналы с дискретной частотной модуляцией. Заметим, что в этом случае удается сравнительно легко получить выражение для вероятности ошибки.

Отметим также, что случайная величина V_0 , являющаяся смесью сигнала и помехи, в данном случае может принимать любые значения в интервале $0 \leq V_0 < \infty$. Поэтому для нахождения вероятности ошибки усредним значение вероятности $P(V_0 < V_1)$ по всем значениям V_0 :

$$P_{\text{ош}} = \int_0^{\infty} \omega(V_0) \left[\int_{V_0}^{\infty} \omega(V_1) dV_1 \right] dV_0.$$

Нетрудно увидеть, что случайная величина V_1 распределяется по закону Рэлея:

$$\omega(V_1) = \frac{2V_1}{N_0 E} \exp\left(-\frac{V_1^2}{N_0 E}\right); V_1 \geq 0.$$

Следовательно, величина V_0 , представляющая собой совокупность сигнала и гауссовского шума, распределяется по закону Райса:

$$\omega(V_0) = \frac{2V_0}{N_0 E} \exp\left(-\frac{V_0^2 + E^2}{N_0 E}\right) I_0\left(\frac{2V_0}{N_0}\right); V_0 \geq 0.$$

Соответствующие подстановки данных величин и последующие вычисления позволяют получить окончательно следующее выражение для определения вероятности ошибки:

$$P_{\text{ош}} = \frac{1}{2} \exp\left(-\frac{E}{2N_0}\right) = \frac{1}{2} \exp\left(-\frac{h^2}{2}\right),$$

где $h^2 = E/N_0$ — отношение энергии сигнала к спектральной плотности мощности помехи.

Расчеты, выполняемые по этому выражению и соотношению для определения вероятности ошибки при когерентном приеме ЧМ-сигналов, позволяют отметить следующее. При переходе от когерентного приема к некогерентному энергетический проигрыш оказывается небольшим, причем с увеличением требований к верности приема (с уменьшением вероятности ошибки) этот проигрыш уменьшается. В связи с этим, а также ввиду значительных сложностей получения в приемнике опорных колебаний, синфазных с принимаемым сигналом, как правило, переходят к некогерентному приему даже в тех случаях, когда фаза сигнала изменяется очень медленно.

8.9. Некогерентный оптимальный прием сигналов с относительной фазовой модуляцией

Пусть имеется канал передачи информации, который описывается моделью с неопределенной фазой. Пусть изменения фазы происходят сравнительно медленно, и в интервале двух соседних посылок фазовые соотношения не нарушаются, тогда оказывается возможным применение ОФМ-сигналов. Многочисленные эксперименты показали, что практически такие условия выполняются во многих реальных каналах. Например, в каналах космической (спутниковой), тропосферной, радиорелейной, коротковолновой радиосвязи и некоторых других.

С учетом правила формирования ОФМ-сигналов можно записать, что информационному символу «0» соответствует реализация

$$S_0(t) = U_c \cos(\omega t + \varphi); \quad t \in [0, 2T],$$

а символу «1» — реализация

$$S_1(t) = \begin{cases} U_c \cos(\omega t + \varphi), & t \in [0, T]; \\ -U_c \cos(\omega t + \varphi), & t \in [0, 2T], \end{cases}$$

где φ — случайная фаза сигнала, неизвестная при приеме, которая зависит, в частности, от символа, переданного в $(k-1)$ -м тактовом интервале или ранее ($k = 1, 2, \dots$).

Можно показать, что эти реализации представляют собой систему сигналов ортогональных в усиленном смысле. Однако в отличие от ЧМ-сигналов их следует рассматривать в интервале $2T$, а не T .

Получим соотношение, определяющее процедуру обработки принимаемых ОФМ-сигналов. В соответствии с правилом оптимального некогерентного приема в демодуляторе должны быть сформированы следующие величины:

$$V_{0k} = (X_{k-1} + X_k)^2 + (Y_{k-1} + Y_k)^2;$$

$$V_{1k} = (X_{k-1} - X_k)^2 + (Y_{k-1} - Y_k)^2; k = 1, 2, \dots$$

Здесь

$$X_{k-1} = \int_{(k-1)T}^{kT} z(t) \cos \omega t dt; Y_{k-1} = \int_{(k-1)T}^{kT} z(t) \sin \omega t dt;$$

$$X_k = \int_{kT}^{(k+1)T} z(t) \cos \omega t dt; Y_k = \int_{kT}^{(k+1)T} z(t) \sin \omega t dt,$$

где $z(t)$ — смесь сигнала и помехи на входе приемника.

Выражение для V_{0k} соответствует сигналу $S_0(t)$, а выражение для V_{1k} — сигналу $S_1(t)$. Тогда при передаче символа «0» решение принимается, если

$$(X_{k-1} + X_k)^2 + (Y_{k-1} + Y_k)^2 > (X_{k-1} - X_k)^2 + (Y_{k-1} - Y_k)^2.$$

После выполнения преобразований получим

$$X_{k-1}X_k + Y_{k-1}Y_k > 0.$$

Данную процедуру можно реализовать с помощью схемы, построенной на основе активных фильтров или корреляторов. Схема такого некогерентного приемника, являющегося оптимальным в условиях помех типа белого шума и называемого взаимокорреляционным, показана на рис. 8.10. Схема содержит опорный генератор и фазовращатель, позволяющие получить напряжения $A \cos \omega t$ и $A \sin \omega t$. Свертка принимаемого сигнала с опорным колебанием осуществляется с помощью корреляторов, на выходах которых получают величины X_{k-1} , X_k , Y_{k-1} , Y_k . Далее посредством сравнения с заданным значением порога в момент времени окончания посылки с длительностью T выносится решение о переданном сигнале.

Поскольку ОФМ-сигналы, как отмечалось, являются ортогональными в усиленном смысле, то для оценки помехоустойчивости оптимального приемника в канале с неопределенной фазой можно использовать полученное ранее расчетное соотношение. Правда, при этом следует отметить, что при длительности $2T$ двух посылок, участвующих в вынесении решения, энергия имеет значение $2E$. Тогда вероятность ошибки

$$P_{\text{ош}} = \frac{1}{2} e^{-h^2}.$$

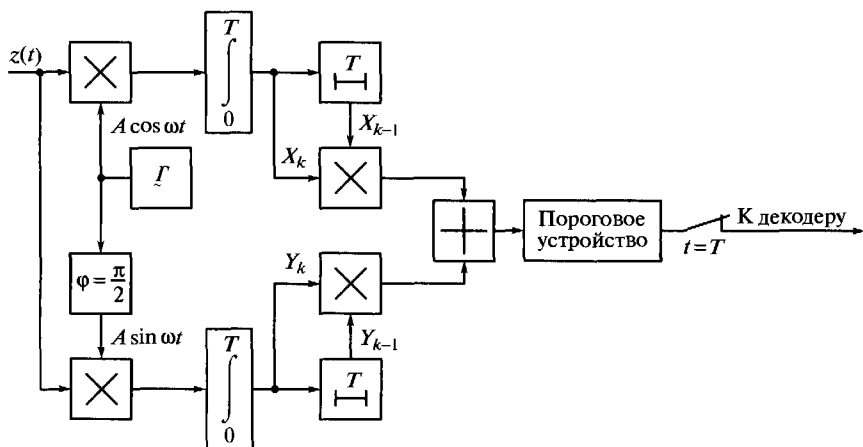


Рис. 8.10. Схема оптимального некогерентного взаимокорреляционного приемника сигналов с относительной фазовой модуляцией

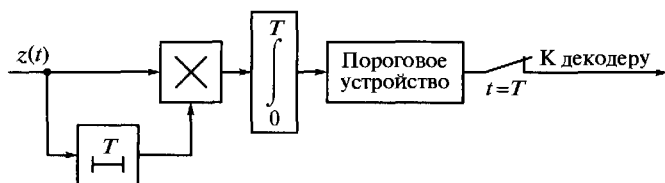


Рис. 8.11. Схема автокорреляционного некогерентного приемника сигналов с относительной фазовой модуляцией

Сравнение помехоустойчивости ОФМ- и ЧМ-сигналов при обработке в условиях неопределенной фазы в канале показывает, что энергия последних в два раза меньше, что эквивалентно 3 дБ.

Помимо рассмотренной схемы в технике телекоммуникаций применяется автокорреляционный некогерентный приемник ОФМ-сигналов, содержащий перемножитель и интегратор (рис. 8.11). В качестве опорного колебания в таком приемнике используется посылка сигнала, задержанная на время T , равное ее длительности. В корреляторе производится сравнение фазы текущей посылки с фазой предыдущей посылки (опорной). Поэтому данный метод приема сигналов называется также методом сравнения фаз. Поскольку в таком приемнике вычисляется коэффициент автокорреляции двух соседних посылок, он называется автокорреляционным.

8.10. Прием дискретных сигналов в условиях замираний

Как уже отмечалось, более общей моделью для описания процессов передачи сигналов в реальных радиоканалах является модель, где случайным образом изменяется не только фаза колебаний, но и амплитуда. Это явление принято характеризовать переменным коэффициентом передачи канала μ .

Причины и сущность замираний были кратко рассмотрены ранее, поэтому на этом вновь останавливаться не будем, а рассмотрим процедуру обработки сигналов для таких каналов. Применяв общий подход, вытекающий из критерия Котельникова, и правила максимального правдоподобия, можно найти выражения, описывающие процедуру обработки. При этом оказывается, что в случае использования системы сигналов с активной паузой (ОФМ- и ЧМ-сигналов) в канале с замираниями оптимальные приемники будут такими же, как и в канале с неопределенной фазой. Это следует из того, что процедура обработки не зависит от распределения амплитуд сигналов, в отличие от помехоустойчивости приема, которая зависит от распределения этих амплитуд.

Для выполнения оценки помехоустойчивости полученное ранее правило вынесения решения при некогерентной обработке ЧМ-сигналов можно представить в следующем виде (считая, что передается сигнал $S_0(t)$, соответствующий символу «0»):

$$X_0^2 + Y_0^2 - X_1^2 - Y_1^2 > 0.$$

Здесь

$$X_i = \int_0^T z(t) \cos \omega_i t dt; Y_i = \int_0^T z(t) \sin \omega_i t dt,$$

где $i = 0, 1$ — случайные гауссовские значения; $z(t)$ — аддитивная смесь сигнала с помехой, являющейся белым шумом.

В указанном виде правило вынесения решения соответствует устройству, которое называется K -демодулятором, поскольку в условиях белого гауссовского шума данное соотношение представляет собой квадратичную форму случайных гауссовских величин.

Теперь рассмотрим модель рэлеевских замираний, для которой коэффициент передачи μ подчиняется следующему распределению:

$$\omega(\mu) = \frac{\mu}{\sigma_\mu^2} \exp\left(-\frac{\mu^2}{2\sigma_\mu^2}\right); \mu > 0,$$

где σ_μ^2 — дисперсия коэффициента передачи μ .

При данной модели замираний квадратурные составляющие смеси сигналов и помех подчиняются гауссовскому распределению с нулевым математическим ожиданием, а плотность распределения вероятностей выходного напряжения x рассматриваемого приемника

$$\omega(x) = \begin{cases} \frac{1}{2(\alpha + \beta)} \exp\left(-\frac{x}{2\alpha}\right), & x > 0; \\ \frac{1}{2(\alpha + \beta)} \exp\left(-\frac{|x|}{2\beta}\right), & x < 0, \end{cases}$$

где α и β — соответственно дисперсии случайных величин $V_0 = X_0^2 + Y_0^2$ и $V_1 = X_1^2 + Y_1^2$.

Нетрудно увидеть, что при использовании ЧМ-сигналов, когда передается посылка $S_0(t)$, дисперсия $\alpha = \bar{E} + N_0$, а дисперсия $\beta = N_0$, где $\bar{E} = \mu^2 E_n$ — энергия сигнала на входе приемника; E_n — энергия на выходе передатчика.

Вероятность ошибки при этом можно найти с помощью выражения

$$P_{\text{ЧМ}} = \int_0^{\infty} \omega(x) dx = \frac{1}{2(\alpha + \beta)} \int_0^{\infty} \exp\left(-\frac{|x|}{2\beta}\right) dx = \frac{\beta}{\alpha + \beta} = \\ = \frac{N_0}{\bar{E} + 2N_0} = \frac{1}{\bar{h}^2 + 2}; \quad \bar{h}^2 = \frac{\mu^2 E_n}{N_0}.$$

Несложно также показать, что при использовании ОФМ-сигналов вероятность ошибки

$$P_{\text{ош}} = \frac{1}{2(1 + \bar{h}^2)}.$$

В теории связи получены соотношения для оценки помехоустойчивости и при других законах замираний. В частности, при замираниях Райса для ЧМ-сигналов

$$P_{\text{ош}} = \frac{1 + q^2}{2 + 2q^2 + \bar{h}^2} \exp\left(-\frac{\bar{h}^2 q^2}{2 + 2q^2 + \bar{h}^2}\right),$$

где q^2 — отношение мощностей регулярной и флукутирующей составляющих.

При отсутствии регулярной составляющей ($q^2 = 0$) из этой формулы следует выражение для определения вероятности ошибки при замираниях по закону Рэлея.

Известно также аналогичное выражение для определения вероятности ошибки при использовании ОФМ-сигналов:

$$P_{\text{ош}} = \frac{1 + q^2}{2(1 + q^2 + \bar{h}^2)} \exp\left(-\frac{\bar{h}^2 q^2}{1 + q^2 + \bar{h}^2}\right).$$

Из представленных соотношений для каналов с замираниями видно, что вероятности ошибок уменьшаются с увеличением отношения сигнал / шум (\bar{h}^2), но характер зависимостей для разных каналов при этом оказывается различным. Так, в канале с неопределенной фазой эта зависимость является экспоненциальной, а в канале с замираниями Рэлея — обратно пропорциональной. Поэтому для обеспечения заданной достоверности приема при замираниях приходится значительно увеличивать мощности передатчиков.

Например, вероятность ошибки не более 10^{-4} при некогерентном приеме ЧМ-сигналов в отсутствие замираний обеспечивается при $\bar{h}^2 = 18$, а при рэлеевских замираниях для этого необходимо иметь $\bar{h}^2 = 10000$, что возможно только при значительном увеличении мощностей передатчиков.

Эффективным способом повышения помехоустойчивости связи в условиях замираний является *разнесенный прием*. Сущность такого приема заключается в том, что в приемнике решение о переданной реализации сигнала принимается на основе обработки нескольких (двух и более) копий одного сигнала, пришедших разными путями, но при этом несущих одну и ту же информацию.

В настоящее время известны следующие методы разнесенного приема:

- *пространственное разнесение*, при котором один и тот же переданный сигнал принимается одновременно несколькими антеннами. При этом антенны удалены друг от друга на расстояние $(10 \dots 20)\lambda$, где λ — длина радиоволны, что в диапазоне коротких волн соответствует расстоянию порядка нескольких сотен метров и является приемлемым для декорреляции замираний на выходах разнесенных антенн. Для получения необходимого эффекта достаточно двух антенн. Прием в этом случае называется *сдвоенным*;

- *разнесение по углу прихода сигнала*, при котором для приема используется несколько антенн с очень узкими диаграммами направленности (около 1°). В диапазоне коротких волн этот метод практически не применяется, но распространен в каналах тропосферной радиосвязи;

- *поляризационное разнесение*, при котором сигнал принимается на две антенны с горизонтальной и вертикальной поляризацией. В процессе распространения электромагнитная волна поляризуется в двух плоскостях, расположенных перпендикулярно по отношению друг к другу, причем флюктуации в каждой из них происходят независимо. В связи с этим сдвоенный поляризационно разнесенный прием оказывается весьма полезным для повышения достоверности связи;

- *временное разнесение*, при котором некоторый символ сообщения передается n раз через промежутки времени, превышающие интервал корреляции между реализациями сигнала;
- *частотное разнесение*, при котором один и тот же сигнал передается на разнесенных между собой несущих частотах, что позволяет получить различные варианты этого замирающего сигнала.

Далее полученные в ветвях разнесения копии сигнала комбинируются или складываются. Существует несколько методов комбинирования. Одним из наиболее эффективных методов является автовыбор ветви с наиболее сильным сигналом, который после обработки поступает получателю.

Контрольные вопросы

1. Назовите основные статистические критерии качества приема сигналов и дайте их характеристики.
2. Как формулируется задача синтеза оптимального приемника на основе критерия идеального наблюдателя (правила максимума правдоподобия)?
3. Поясните процедуру обработки сигналов в оптимальном когерентном приемнике и что собой представляет коррелятор.
4. Дайте сравнительную оценку помехоустойчивости когерентного приема двоичных дискретных сигналов с типовыми видами модуляции и покажите ее связь с представлением сигналов в виде векторов в функциональном пространстве.
5. Что называется согласованным фильтром и каковы его основные свойства?
6. Поясните принцип формирования сигналов с относительной фазовой модуляцией, каким достоинством они обладают и какова схема их обработки в канале с постоянными параметрами и аддитивным белым шумом.
7. Как реализуются взаимокорреляционный и автокорреляционный методы приема сигналов с относительной фазовой модуляцией?
8. Как оценивается помехоустойчивость некогерентного приема в каналах с замираниями?

Глава 9

ОСНОВЫ ТЕОРИИ ПЕРЕДАЧИ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

9.1. Особенности оптимального приема непрерывных сообщений. Потенциальная помехоустойчивость приема

9.1.1. Общие сведения

Непрерывные сообщения (речь, музыка, телеметрическая информация и т. д.), как правило, передаются с использованием аналоговой модуляции. В этом случае принимаемая смесь сигнала и помехи

$$z(t) = S(t, \lambda(t)) + n(t); 0 \leq t \leq T. \quad (9.1)$$

Шум $n(t)$ будем считать белым гауссовским с односторонней спектральной плотностью N_0 . Сигнал в выражении (9.1) может зависеть от сообщения линейно (при АМ и БМ) или нелинейно (при ЧМ и ФМ), откуда и названия: линейные и нелинейные методы модуляции. Различают также прямые и интегральные методы модуляции. При прямых методах сообщение воздействует на основной параметр сигнала — амплитуду, фазу и т. д. (соответственно это АМ, БМ и ФМ). К интегральным методам относится частотная модуляция, при которой сообщение воздействует на основной параметр — фазу — через интеграл.

Под сообщением в выражении (9.1) можно понимать также набор некоторых параметров принимаемого сигнала (фазу, частоту, время задержки и т. д.), которые необходимо оценивать на приемной стороне для обеспечения синхронизации при приеме сигнала. Рассматриваемая модель принимаемого сигнала вида (9.1) самая простая. На практике сообщение может быть векторным и включать в себя различные параметры, подлежащие оцениванию. Задача оптимального приема непрерывного сообщения состоит в получении оценки, или фильтрации, этого сообщения, в наименьшей степени отличающейся от сообщения или оптимальной в смысле выбранного критерия. В качестве такого критерия часто используется среднеквадратичная ошибка воспроизведения сообщения $\epsilon^2 = \overline{[\lambda(t) - \hat{\lambda}(t)]^2} = N_{\text{вых}}$. Теория оптимального приема

непрерывных сообщений впервые была разработана В.А. Котельниковым.

Мощность передаваемого сообщения принято считать заданной, т.е. $P_{\text{вых}} = \lambda^2(t)$, а в качестве мощности ошибки фильтрации можно принять значение ее среднеквадратичной ошибки, т.е. $N_{\text{вых}} = \varepsilon^2$. Тогда отношение сигнал/шум на выходе демодулятора $\rho_{\text{вых}} = \left(\frac{P}{N}\right)_{\text{вых}}$. Отношение сигнал/шум на входе приемника обо-

значим $\rho_{\text{вх}} = \left(\frac{P}{N}\right)_{\text{вх}}$. Часто в качестве критерия оптимальности (помехоустойчивости) применяют не средний квадрат ошибки ε^2 , а отношение сигнал/шум на выходе демодулятора $\rho_{\text{вых}}$. Помехоустойчивость систем передачи непрерывных сообщений также удобно оценивать значением выигрыша

$$g = \frac{\left(\frac{P}{N}\right)_{\text{вых}}}{\left(\frac{P}{N}\right)_{\text{вх}}} = \frac{\rho_{\text{вых}}}{\rho_{\text{вх}}}, \quad (9.2)$$

которое зависит от вида модуляции и используемого способа приема. При $g > 1$ отношение сигнал/помеха при демодуляции улучшается. При $g < 1$ имеет место проигрыш.

Определим отношение сигнал/шум в числителе и в знаменателе выражения (9.2). Под пикфактором информационного сообщения принято понимать отношение максимального значения этого сообщения к действующему: $\Pi = \frac{|\lambda(t)|_{\text{max}}}{\sqrt{\lambda^2(t)}}$. Произведя нор-

мировку сообщения $|\lambda(t)|_{\text{max}} = 1$, получим $\overline{\lambda^2(t)} = \frac{1}{\Pi^2}$, тогда

$P_{\text{вых}} = \overline{\lambda^2(t)} = \frac{1}{\Pi^2}$ и $\rho_{\text{вых}} = \left(\frac{P}{N}\right)_{\text{вых}} = \frac{1}{\Pi^2 N_{\text{вых}}}$, где мощность шума на

выходе приемника (демодулятора), имеющего полосу пропускания по низкой частоте $[0 - F]$,

$$N_{\text{вых}} = \int_0^F G(f) df, \quad (9.3)$$

где $G(f)$ — спектральная плотность мощности шума на выходе оптимального демодулятора.

В теории оптимального приема непрерывных сообщений доказано, что для прямых методов модуляции

$$G(f) = \frac{N_0}{\left\{ \frac{\partial S[t, \lambda(t)]}{\partial \lambda} \right\}^2}, \quad (9.4)$$

а для интегральных методов модуляции спектральную плотность мощности шума можно вычислить по формуле

$$G(f) = \frac{N_0}{\left\{ \frac{\partial S[t, \lambda(t)]}{\partial \lambda} \right\}^2} (2\pi f)^2. \quad (9.5)$$

На входе приемника (в полосе пропускания демодулятора $F_{\text{вх}}$)

$$\rho_{\text{вх}} = \left(\frac{P}{N} \right)_{\text{вх}} = \frac{\overline{S^2(t)}}{N_0 F_{\text{вх}}}.$$

Итак, выражение (9.2) можно представить в следующем виде:

$$g = \frac{\rho_{\text{ввых}}}{\rho_{\text{вх}}} = \frac{N_0 F_{\text{вх}}}{\Pi^2 N_{\text{ввых}} \overline{S^2(t)}}.$$

Сравнивать различные системы по значению g не совсем удобно. Действительно, если $\left(\frac{P}{N} \right)_{\text{ввых}}$, $P_{\text{вх}}$ и N_0 у двух демодуляторов одинаковые, то они равноценны по качеству приема. Однако при этом мощности шумов $N_{\text{вх}} = N_0 F_{\text{вх}}$ для двух приемников (например, с АМ и ЧМ) будут различными, поэтому и выигрыши g будут отличаться. Чтобы исключить этот недостаток вводят обобщенный выигрыш, выраженный через отношение мощностей сигналов не к мощностям помех, а к их средним спектральным плотностям:

$$g' = \frac{\rho'_{\text{ввых}}}{\rho'_{\text{вх}}};$$

$$\rho'_{\text{ввых}} = \frac{P_{\text{ввых}}}{N_{0\text{ввых}}} = \frac{P_{\text{ввых}}}{\frac{N_{\text{ввых}}}{F_{\text{ввых}}}} = \left(\frac{P}{N} \right)_{\text{ввых}} F; \quad \rho'_{\text{вх}} = \frac{P_{\text{вх}}}{N_{0\text{вх}}} = \frac{P_{\text{вх}}}{\frac{N_{\text{вх}}}{F_{\text{вх}}}} = \left(\frac{P}{N} \right)_{\text{вх}} F_{\text{вх}},$$

причем $F = F_{\text{ввых}}$.

Тогда $g' = g \frac{F}{F_{\text{вх}}}$, т.е. окончательно получим

$$g' = \frac{\rho_{\text{ввых}}}{\rho_{\text{вх}}} = \frac{N_0 F}{\Pi^2 N_{\text{ввых}} \overline{S^2(t)}}. \quad (9.6)$$

Итак, обобщенный выигрыш g' характеризует потенциальную помехоустойчивость приема при различных видах модуляции и определяется выражением (9.6). Для его вычисления необходимо определить значение $N_{\text{вых}}$ по формуле (9.3), подставив в нее спектральную плотность мощности шума $G(f)$, которая находится по выражению (9.4) для прямых методов модуляции (АМ, БМ, ФМ) и по выражению (9.5) для интегральных методов модуляции (ЧМ).

9.1.2. Определение обобщенного выигрыша для различных методов модуляции

Амплитудная модуляция

При амплитудной модуляции сигнал на входе приемного устройства имеет вид

$$S(t, \lambda(t)) = A[1 + m_a \lambda(t)] \cos(\omega_0 t + \varphi),$$

где m_a — коэффициент амплитудной модуляции.

Если $\overline{\lambda(t)} = 0$ и $\overline{\lambda^2(t)} = \frac{1}{\Pi^2}$, то вследствие независимости сообщения и высокочастотного колебания среднее по времени значение квадрата сигнала на входе

$$\overline{S^2(t)} = A^2 [1 + 2m_a \overline{\lambda(t)} + m_a^2 \overline{\lambda^2(t)}] \cos^2(\omega_0 t + \varphi) = \frac{A^2}{2} \left[1 + \frac{m_a^2}{\Pi^2} \right]. \quad (9.7)$$

Для определения спектральной плотности мощности по формуле (9.4) найдем

$$\frac{\partial S(t, \lambda(t))}{\partial \lambda} = A m_a^2 \cos(\omega_0 t + \varphi),$$

откуда среднее значение квадрата производной

$$\overline{\left[\frac{\partial S(t, \lambda(t))}{\partial \lambda} \right]^2} = \overline{A^2 m_a^2 \cos^2(\omega_0 t + \varphi)} = \frac{1}{2} A^2 m_a^2.$$

Тогда

$$G(f) = \frac{N_0}{\overline{\left[\frac{\partial S(t, \lambda(t))}{\partial \lambda} \right]^2}} = \frac{2N_0}{A^2 m_a^2},$$

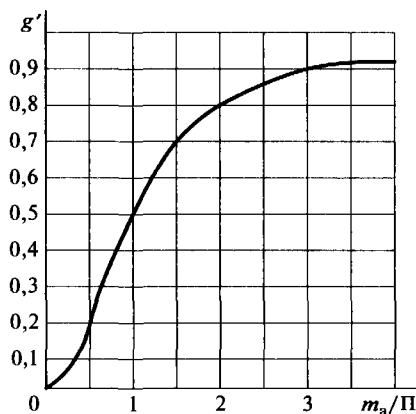


Рис. 9.1. Зависимость обобщенного выигрыша g' от отношения m_a/Π при амплитудной модуляции

а мощность шума на выходе приемника

$$N_{\text{вых}} = \int_0^F G(f) df = \frac{2N_0 F}{A^2 m_a^2}. \quad (9.8)$$

Подставив в формулу (9.6) значение мощности шума на выходе [см. формулу (9.8)] и среднее по времени значение квадрата сигнала на входе [см. формулу (9.7)], получим выражение для определения обобщенного выигрыша при амплитудной модуляции сигналов:

$$g' = \frac{m_a^2}{m_a^2 + \Pi^2} = \frac{m_a^2 / \Pi^2}{1 + m_a^2 / \Pi^2}. \quad (9.9)$$

На рис. 9.1 представлена зависимость обобщенного выигрыша g' от отношения m_a/Π . В соответствии с формулой (9.9) значение g' может изменяться в пределах от нуля до единицы, однако практически оно может быть значительно меньше единицы. Если положить $m_a = 1$, то можно записать

$$g' = \frac{1}{1 + \Pi^2}. \quad (9.10)$$

Таким образом, из расчета видно, что обобщенный выигрыш при амплитудной модуляции существенно меньше единицы и отношение сигнал/помеха на выходе приемника значительно меньше этого отношения на его входе (в полосе пропускания приемника).

Балансная и однополосная модуляция

Сигнал на выходе приемного устройства при балансной модуляции сигналов можно представить в виде

$$S(t, \lambda(t)) = A m_6 \lambda(t) \cos(\omega_0 t + \varphi),$$

где m_6 — коэффициент балансной модуляции.

По сравнению с амплитудной модуляцией в рассматриваемом случае сигнал не содержит регулярной (неинформативной) составляющей $A \cos(\omega_0 t + \varphi)$.

Среднее значение квадрата сигнала

$$\overline{S^2(t, x)} = A^2 m_6^2 \overline{x^2(t) \cos^2(\omega_0 t + \varphi)} = \frac{A^2 m_6^2}{2\pi^2}, \quad (9.11)$$

а среднее значение квадрата производной

$$\left[\frac{\partial S(t, \lambda(t))}{\partial \lambda} \right]^2 = A^2 m_6^2 \overline{\cos^2(\omega_0 t + \varphi)} = \frac{A^2 m_6^2}{2}.$$

Следовательно, спектральная плотность мощности помех на выходе приемника не зависит от частоты:

$$G(f) = \frac{N_0}{\left[\frac{\partial S(t, \lambda(t))}{\partial \lambda} \right]^2} = \frac{2N_0}{A^2 m_6^2}.$$

Тогда мощность шума на выходе приемника

$$N_{\text{вых}} = \int_0^F G(f) df = \frac{2N_0 F}{A^2 m_6^2}. \quad (9.12)$$

Подставив выражения (9.12) и (9.11) в (9.6), получим

$$g' = 1. \quad (9.13)$$

Таким образом, обобщенный выигрыш при балансной модуляции оказывается значительно больше, чем при амплитудной, и он всегда равен единице.

Можно также показать, что и при однополосной модуляции обобщенный выигрыш $g' = 1$.

Высокий выигрыш в обоих случаях достигается за счет исключения неинформативной составляющей на несущей частоте сигнала.

Фазовая модуляция

При фазовой модуляции сигнал на входе приемного устройства имеет вид

$$S(t, \lambda(t)) = A \cos[\omega_0 t + m_\phi \lambda(t) + \varphi_0],$$

где m_ϕ — индекс фазовой модуляции, характеризующий наибольшее изменение фазы высокочастотного колебания в процессе модуляции его нормированным сообщением $\lambda(t)$.

Для нахождения обобщенного выигрыша определим среднее по времени значение квадрата сигнала:

$$\overline{S^2(t, \lambda(t))} = A^2 \overline{\cos^2[\omega_0 t + m_\phi \lambda(t) + \phi_0]} = \frac{A^2}{2}. \quad (9.14)$$

Среднее значение квадрата производной

$$\left[\frac{\partial S(t, \lambda(t))}{\partial \lambda} \right]^2 = A m_\phi \overline{\{-\sin[\omega_0 t + m_\phi \lambda(t) + \phi_0]\}^2} = \frac{A^2 m_\phi^2}{2}.$$

Тогда спектральную плотность мощности шума на выходе приемника, которая постоянна по диапазону частот, запишем в виде

$$G(f) = \frac{N_0}{\left[\frac{\partial S(t, \lambda(t))}{\partial \lambda} \right]^2} = \frac{2N_0}{A^2 m_\phi^2}.$$

Мощность шума на выходе приемника

$$N_{\text{вых}} = \int_0^F G(f) df = \frac{2N_0 F}{A^2 m_\phi^2}. \quad (9.15)$$

Подставив выражения (9.14) и (9.15) в формулу (9.6), получим выражение для определения обобщенного выигрыша при фазовой модуляции сигналов:

$$g' = \frac{m_\phi^2}{\Pi^2}. \quad (9.16)$$

На рис. 9.2 представлена графическая зависимость (9.16), из которой следует, что обобщенный выигрыш при фазовой модуляции может быть существенно больше единицы. Следовательно, фазовая модуляция оказывается более помехоустойчивой по сравнению с амплитудной модуляцией. Однако при реализации фазовой модуляции возникает необходимость учета многозначности фазы

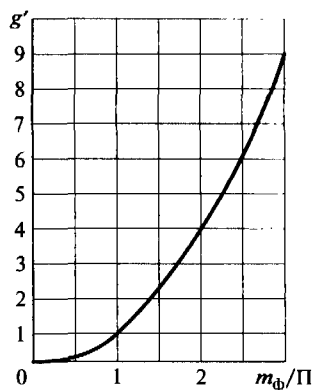


Рис. 9.2. Зависимость обобщенного выигрыша g' от отношения m_ϕ/Π при фазовой модуляции

сигнала. При увеличении индекса m_ϕ растет ширина спектра сигнала и, следовательно, ширина полосы канала связи, необходимая для передачи сообщений.

Частотная модуляция

Сигнал на входе приемного устройства при частотной модуляции имеет вид

$$S(t, x) = A \cos[\omega_0 t + \Omega_0 x(t) + \phi_0],$$

где $x(t) = \int_0^t \lambda(t) dt$.

Среднее по времени значение квадрата сигнала

$$\overline{S^2(t, x)} = A^2 \overline{\cos^2[\omega_0 t + \Omega_0 x(t) + \phi_0]} = \frac{1}{2} A^2. \quad (9.17)$$

Спектральная плотность мощности в этом случае вычисляется по формуле (9.5). Среднее по времени значение квадрата производной от сигнала по параметру $x(t)$ запишем в виде

$$\overline{\left[\frac{\partial S(t, x)}{\partial x} \right]^2} = A^2 \Omega_0^2 \overline{\sin^2[\omega_0 t + \Omega_0 x(t) + \phi_0]} = \frac{A^2 \Omega_0^2}{2}.$$

В соответствии с формулой (9.5)

$$G(f) = \frac{N_0 (2\pi f)^2}{\overline{\left[\frac{\partial S(t, x)}{\partial x} \right]^2}} = \frac{2N_0}{A^2 \Omega_0^2} (2\pi f)^2.$$

Таким образом, спектральная плотность мощности помех на выходе приемника пропорциональна квадрату частоты, а мощность помех на выходе приемника

$$N_{\text{вых}} = \int_0^F G(f) df = \frac{2N_0 (2\pi F)^2}{3A^2 \Omega_0^2} F = \frac{2N_0 F}{3A^2 m_f^2}. \quad (9.18)$$

Здесь индекс частотной модуляции равен отношению наибольшего отклонения частоты сигнала f_{max} от среднего значения f_0 к ширине полосы пропускания приемника по низкой частоте F

$$m_f = \frac{\Omega_0}{2\pi F} = \frac{F_{\text{max}}}{F}.$$

Подставив выражения (9.17) и (9.18) в формулу (9.6), получим обобщенный выигрыш при частотной модуляции

$$g' = 3 \frac{m_f^2}{\Pi^2}. \quad (9.19)$$

На рис. 9.3 представлена графическая зависимость (9.19), из которой видно, что частотная модуляция может иметь высокий обобщенный выигрыш g' , если отношение m_f/Π сравнительно большое. Увеличение выигрыша сопровождается расширением спектра сигнала, поэтому частотная модуляция применяется в основном в тех средствах связи, в которых используется диапазон частот, позволяющий выделить одному корреспонденту достаточно широкую полосу, и где желателен прием сообщений с высоким качеством. Частотная модуляция получила широкое применение в радиорелейных станциях (для вторичного уплотнения), для художественного радиовещания и звукового сопровождения телевизионных программ в диапазоне УКВ и т. д.

Однако частотная модуляция имеет существенный недостаток, характерный для сигналов с малой амплитудой: начиная со значения отношения $P/N_{\text{вх}} = 5$, называемого порогом помехоустойчивости, при уменьшении сигнала помехи подавляют его, при этом отношение сигнал/помеха на выходе приемника резко падает, вследствие чего связь прекращается.

Мощность сигнала, соответствующая порогу помехоустойчивости,

$$P_{\text{вх}} = 5N_{\text{вх}} = 5N_0 2m_f F = 10N_0 F m_f$$

Тогда

$$\frac{P_{\text{вх}}}{N_0 F} = 10m_f; \rho_{\text{вых}} = g' \frac{P_{\text{вх}}}{N_0 F}$$

Таким образом, с ростом m_f , т. е. с увеличением g' , а следовательно, и повышением качества приема увеличивается мощность сигнала на входе, соответствующая порогу помехоустойчивости. В логарифмическом масштабе последняя зависимость представляется семейством параллельных прямых вида

$$\lg \rho_{\text{вых}} = \lg g' + \lg \frac{P_{\text{вх}}}{N_0 F},$$

смещенных на постоянное значение $\lg g'$ (рис. 9.4). Для частотной модуляции при значении $P_{\text{вх}}$, меньшем порогового, наступает резкое уменьшение $\rho_{\text{вых}}$.

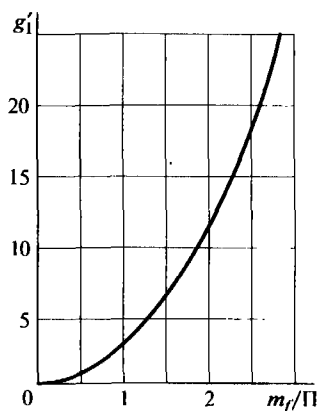


Рис. 9.3. Зависимость обобщенного выигрыша g' от отношения m_f/Π при частотной модуляции

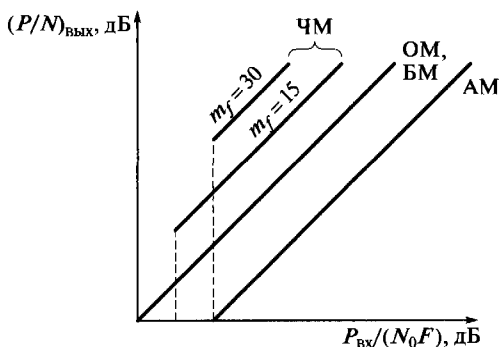


Рис. 9.4. Зависимости отношения $(P/N)_{\text{вых}}$ от отношения $P_{\text{вх}}/(N_0 F)$ для различных видов модуляции

Для снижения порога помехоустойчивости в целях обеспечения связи при малом уровне сигнала целесообразно уменьшить девиацию его частоты применением более узкополосного тракта приема. Эта задача успешно решается с помощью устройств с обратной связью по частоте (ОСЧ), позволяющих снизить девиацию частоты сигнала в тракте усиления по промежуточной частоте (УПЧ), сузить полосу пропускания приемника, снизить порог помехоустойчивости и обеспечить связь при меньшей девиации частоты сигнала.

9.2. Оптимальное оценивание параметров сигналов

В настоящее время широкое применение получила основанная на одноименной теореме байесовская теория оценивания. Ее сущность состоит в следующем. Наблюдается аддитивная смесь $z(t)$ сигнала $S(t, \lambda(t))$ с помехой $n(t)$ [см. формулу (9.1)] в интервале $[0, T]$. Сигнал зависит от принимаемого сообщения $\lambda(t)$. Используя теорему отсчетов, представим принимаемую смесь $z(t)$ и сообщение в виде векторов $\mathbf{z} = (z_1, \dots, z_m)$ и $\bar{\lambda} = (\lambda_1, \dots, \lambda_m)$.

Если задана априорная плотность вероятности $w(\lambda)$, то в соответствии с формулой Байеса

$$w_{\lambda/z}(\bar{\lambda}) = \frac{w(\bar{\lambda})\omega_{z/\lambda}(\mathbf{z})}{\int_{(\lambda)} w(\lambda)\omega_{z/\lambda}(\mathbf{z})d\lambda} = k_1 w(\bar{\lambda})\omega_{z/\lambda}(\mathbf{z}). \quad (9.20)$$

Здесь $k_1 = 1 / \int_{(\lambda)} w(\bar{\lambda})\omega_{z/\lambda}(\mathbf{z})d\bar{\lambda}$, а значок (λ) под интегралом обо-

значает область интегрирования по вектору λ .

С помощью формулы (9.20) можно определить апостериорную плотность вероятностей, которая будет иметь меньшую дисперсию, чем априорная $w(\lambda)$.

Решение задачи оценивания состоит в подборе оценок $\bar{\lambda} = \{\lambda_1, \dots, \lambda_m\}$ для вектора измеряемых параметров $\bar{\lambda}$, оптимального с позиции минимума среднего риска. При этом необходимо найти качественные показатели оптимальной обработки и рассмотреть пути реализации найденных алгоритмов оценки.

Введем понятия риска и среднего риска применительно к измерению одного параметра. Совокупность возможных ситуаций при измерении соответствует совокупности различных значений параметра $\bar{\lambda}$ и его оценки $\bar{\lambda}$. При этом цена (риск) каждой ситуации определяется случайной величиной $R(\bar{\lambda}, \bar{\lambda})$, называемой функцией потерь, а средний риск — соотношением

$$\langle R \rangle = \int_{(\lambda)} d\lambda \int_{\bar{\lambda}} R(\bar{\lambda}, \bar{\lambda}) \omega(\bar{\lambda}, \bar{\lambda}) d\bar{\lambda}, \quad (9.21)$$

где $\omega(\bar{\lambda}, \bar{\lambda})$ — совместная плотность вероятностей случайных величин $\bar{\lambda}$ и $\bar{\lambda}$.

Критерий минимума среднего риска задается соотношением $\langle R \rangle = \min_{\bar{\lambda}}$, и оптимальной системой измерения по этому критерию является та система, для которой справедливо это соотношение.

Если считать функцию потерь, зависящей лишь от разности $\lambda - \bar{\lambda} = \varepsilon$, то средний риск будет определяться плотностью вероятностей $w(\varepsilon)$ и функцией потерь $R(\varepsilon)$. Для частных случаев функции $R(\varepsilon)$ можно получить частные критерии оптимального измерения. Для квадратичной функции потерь $R(\varepsilon) = \varepsilon^2$ (рис. 9.5, а) средний риск $\langle R \rangle = \int_{(\varepsilon)} \varepsilon^2 w(\varepsilon) d\varepsilon = \langle \varepsilon^2 \rangle = \min_{\varepsilon}$, что определяет мини-

мум среднего квадрата ошибки. При простой функции потерь $R(\varepsilon) = 1 - \delta(\varepsilon - \varepsilon_0)$ критерий минимума среднего риска сводится к критерию максимума апостериорной вероятности $\omega(\varepsilon_0) = \max_{\varepsilon_0}$ (рис. 9.5, б).

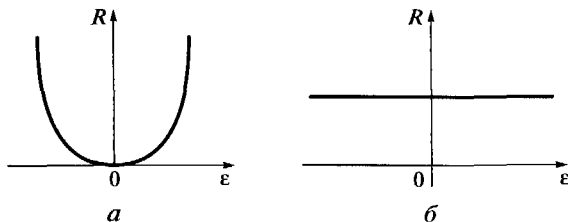


Рис. 9.5. Функции потерь квадратичная (а) и простая (б)

Формирование оценки параметра осуществляется по правилу, которое называется решающим. При фиксированных (нерандомизированных) правилах каждой реализации $z(t)$ соответствует оценка $\hat{\lambda}$, т. е. решающее правило имеет вид $\hat{\lambda}(z(t))$ или для дискретных значений реализации — $\hat{\lambda}(z)$.

Рассмотрим методику формирования решающих правил по принятым значениям реализации $\mathbf{z} = (z_1, \dots, z_m)$ при использовании в качестве критерия оптимальности критерия минимума среднего риска.

Преобразуем выражение (9.21), заменив элемент вероятности $\omega(\bar{\lambda}, \bar{\lambda})d\bar{\lambda}d\bar{\lambda}$ на $w(\bar{\lambda}, \mathbf{z})d\bar{\lambda}dz$. Тогда

$$\langle R \rangle = \int_{(z)} dz \int_{(\bar{\lambda})} R(\bar{\lambda}, \bar{\lambda}) w(\bar{\lambda}, \mathbf{z}) d\bar{\lambda}.$$

По теореме умножения $w(\lambda, \mathbf{z}) = w(\mathbf{z})w(\lambda/\mathbf{z})$, следовательно,

$$\langle R \rangle = \int_{(z)} \langle R(\bar{\lambda}/\mathbf{z}) \rangle w(\mathbf{z}) dz.$$

Здесь условный средний риск, вычисляемый при фиксированном значении \mathbf{z} ,

$$\langle R(\bar{\lambda}/\mathbf{z}) \rangle = \int_{(\bar{\lambda})} R(\bar{\lambda}, \bar{\lambda}) w(\bar{\lambda}/\mathbf{z}) d\bar{\lambda}. \quad (9.22)$$

Усреднение условного среднего риска с учетом плотности вероятностей $w(\mathbf{z})$ определяет безусловный средний риск. Итак, критерий минимума среднего риска сводится к критерию минимума условного среднего риска $\langle R(\bar{\lambda}/\mathbf{z}) \rangle > \min_{\bar{\lambda}}$, и его минимизация требует знания условной плотности вероятностей $w(\bar{\lambda}/\mathbf{z})$, которая является апостериорной плотностью вероятностей параметра $\bar{\lambda}$. Для случая скалярного параметра λ она определяется выражением, сходным с выражением (9.20) для случая векторного параметра:

$$w_{1ps}(\lambda) = w_{1/z}(\lambda) = kw(\lambda)w_{z/\lambda}(z). \quad (9.23)$$

Здесь

$$k = (w(z))^{-1} = \left[\int_{(\bar{\lambda})} w(\bar{\lambda})w_{z/\bar{\lambda}}(z) d\bar{\lambda} \right]^{-1}.$$

Условная плотность вероятностей $w(\lambda/\mathbf{z})$, рассматриваемая как функция от λ и называемая функцией правдоподобия, показывает,

на сколько одно возможное значение параметра λ более правдоподобно, чем другое при фиксированном значении реализации.

Располагая апостериорной плотностью вероятностей для скалярного параметра в виде (9.23) или для векторного параметра в виде (9.20) для конкретных функций потерь, можно найти оптимальную оценку $\hat{\lambda}$ или оценку $\hat{\lambda}$, минимизирующую средний риск. Такие оценки называют байесовскими.

Рассмотрим случай оценки скалярного параметра. При квадратичной функции потерь критерий минимума условного среднего риска (9.22) имеет вид

$$\hat{\lambda}^2 - 2\hat{\lambda} \int_{(\lambda)} \lambda w_{\lambda/z}(\lambda) d\lambda + \int_{(\lambda)} \lambda^2 w_{\lambda/z}(\lambda) d\lambda = \min_{\lambda} .$$

Оптимальная оценка находится дифференцированием левой части приведенного соотношения по $\hat{\lambda}$ и приравниванием к нулю полученного результата. Решение этого уравнения имеет вид

$$\hat{\lambda}_0 = \int_{(\lambda)} \lambda w_{\lambda/z}(\lambda) d\lambda, \quad (9.24)$$

т.е. оценка, оптимальная по критерию минимума среднего квадрата ошибки, совпадает с апостериорным средним значением. Такая оценка называется несмещенной. Условный средний риск с учетом выражения (9.24) равен условной апостериорной дисперсии:

$$\langle R(\hat{\lambda}/z) \rangle = \int_{(\lambda)} (\lambda - \hat{\lambda}_0)^2 w_{\lambda/z}(\lambda) d\lambda = \sigma_{\lambda}^2 .$$

Таким образом, при квадратичной функции потерь обеспечивается оценка параметра с минимальной дисперсией ошибки.

Возьмем теперь простую функцию потерь $R(\lambda, \hat{\lambda}) = 1 - \delta(\lambda - \hat{\lambda})$. Минимизация условного среднего риска (9.22) в этом случае определяет критерий

$$w_{\lambda/z}(\lambda) \Big|_{\lambda=\hat{\lambda}} = k w(\lambda) w_{z/\lambda}(z) \Big|_{\lambda=\hat{\lambda}} = \max_{\lambda} ,$$

т.е. оптимальная оценка обращает в максимум апостериорную плотность вероятностей оцениваемого параметра.

Если априорные сведения об оцениваемом параметре отсутствуют, то обычно предполагают, что априорная плотность вероятностей равномерна в заданном интервале неопределенности. Тогда вычисление оценки сводится к максимизации функции правдоподобия $w_{z/\lambda}(z) \Big|_{\lambda=\hat{\lambda}} = \max_{\lambda}$, а условие оптимальности оценки будет иметь вид уравнения правдоподобия $\frac{\partial}{\partial \lambda} w_{z/\lambda}(z) \Big|_{\lambda=\hat{\lambda}} = 0$.

Решение этого уравнения представляет собой оценку максимального правдоподобия $\hat{\lambda}$, которая в общем случае отличается от оценки, оптимальной по критерию минимума среднего квадрата ошибки.

В тех случаях, когда функции $w(z/\lambda)$ и $w(\lambda/z)$ являются симметричными относительно точки $\hat{\lambda}$, рассмотренные критерии определяют одну и ту же оценку $\hat{\lambda}$. Дисперсия несмещенной оценки, как доказано в теории оценок, удовлетворяет неравенству Крамера—Рао.

При равномерном априорном распределении и гауссовской аппроксимации апостериорной плотности вероятностей с математическим ожиданием λ_0 и дисперсией σ_λ^2 нижняя граница точности оценивания, определяемая неравенством Крамера—Рао, имеет вид

$$\sigma_\lambda^2 \geq \left\langle \left. \frac{\partial^2}{\partial \lambda^2} \ln w_{z/\lambda}(z) \right|_{\lambda=\hat{\lambda}} \right\rangle^{-1}, \quad (9.25)$$

где λ_0 — оценка максимального правдоподобия.

Из выражения (9.25) следует, что апостериорная дисперсия обратно пропорциональна модулю второй производной от логарифма функции правдоподобия.

Перейдем теперь от дискретных значений к непрерывным реализациям $z(t)$. Для одномерного случая апостериорная плотность вероятностей (АПВ)

$$\begin{aligned} w_{ps}(\lambda) &= w_{\lambda/z(t)}(\lambda) = \lim_{\Delta t \rightarrow 0} w_{1ps}(\lambda); \\ w_{ps}(\lambda) &= kw(\lambda)w_{z/\lambda}(z(t)), \end{aligned} \quad (9.26)$$

где $w_{z/\lambda}(z(t)) = F(\lambda)$.

Функционал отношения правдоподобия с точностью до постоянного коэффициента определяется выражением

$$F(\lambda) = \exp \left(-\frac{1}{N_0} \int_0^T (z(t) - S(t, \lambda))^2 dt \right). \quad (9.27)$$

Апостериорная плотность вероятностей (9.26) можно записать в виде

$$w_{ps}(\lambda) = kw(\lambda)F(\lambda).$$

Для нахождения оптимальных оценок можно использовать применяемое в задачах различения сигналов условное отношение правдоподобия, которое в случае векторного параметра и дискретного наблюдения представляет собой отношение плотности вероятностей $w_{z/\lambda}(z)$ к плотности вероятностей $w_{z/0}(z)$:

$$\Lambda_{1/0}(\mathbf{z}/\lambda) = \frac{w_{z/\lambda}(\mathbf{z})}{w_{z/0}(\mathbf{z})}.$$

С учетом выражения (9.20) апостериорную плотность вероятностей можно записать в виде

$$w_{z/\lambda}(\bar{\lambda}) = k_2 w(\bar{\lambda}) \Lambda_{1/0}(\mathbf{z}/\bar{\lambda}), \quad (9.28)$$

где $k_2 = k_1 w_{z/0}(\mathbf{z})$.

Перейдя к пределу в выражении (9.28) при $\Delta t \rightarrow 0$, $m \rightarrow \infty$, получим

$$w_{\lambda/z(t)}(\bar{\lambda}) = k_2 w(\lambda) \Lambda_{1/0}(z(t)/\lambda),$$

где $\Lambda_{1/0}(z(t)/\bar{\lambda})$ — функционал отношения правдоподобия.

Представим эту формулу в логарифмической форме:

$$\ln w_{\lambda/z(t)}(\bar{\lambda}) = \ln k_2 + \ln w(\bar{\lambda}) + \ln \Lambda_{1/0}(z(t)/\bar{\lambda}).$$

Если априорная плотность вероятностей $w(\lambda)$ равномерна, то максимум плотности вероятностей $w_{\lambda/z(t)}(\lambda)$ будет иметь место при максимальном значении $\Lambda_{1/0}(z(t)/\lambda)$ и, следовательно, в качестве оптимальной можно принять оценку, максимизирующую функционал отношения правдоподобия, полученную из решения уравнения

$$\frac{\partial \Lambda_{1/0}(z(t)/\bar{\lambda})}{\partial \lambda_k} = 0 \quad (9.29)$$

для всех $k = 1, 2, \dots, m$. Иными словами, в качестве оценки векторного параметра $\bar{\lambda}$ принимается решение уравнения (9.29).

Если полезный сигнал зависит не только от информационных параметров $\bar{\lambda}$, но и от вектора сопутствующих неинформационных параметров $\bar{\theta} = \{\theta_1, \dots, \theta_n\}$, т.е. $S(t) = S(t, \bar{\lambda}, \bar{\theta})$, то возможны два способа решения задачи. Первый из них связан с объединением двух векторов $\bar{\theta}$ и $\bar{\lambda}$ в один вектор $\bar{\lambda}_1$ большего размера с применением рассмотренных ранее процедур оценивания.

Второй способ заключается в следующем. Сначала для дискретного наблюдения \mathbf{z} формируется условная (в случае наблюдения \mathbf{z}) совместная апостериорная плотность вероятностей параметров $\bar{\theta}$ и $\bar{\lambda}$, которую можно записать аналогично соотношению (9.28):

$$w_{(\lambda, \theta)/z}(\bar{\lambda}, \bar{\theta}) = k_3 w(\bar{\lambda}, \bar{\theta}) \Lambda_{1/0}(\mathbf{z}/\bar{\lambda}, \bar{\theta}).$$

Представив совместную априорную плотность вероятностей параметров $w(\lambda, \theta)$ в виде произведения $w(\lambda, \theta) = w_{\lambda, \theta}(\lambda) w(\lambda)$, после подстановки этого произведения в выражение для $w_{\lambda, \theta/z}$ и

усреднения по вектору неинформационных параметров θ , получим

$$w_{\lambda/z}(\mathbf{z}) = k_3 \int_{\theta} \Lambda_{1/0}(\mathbf{z}/\bar{\lambda}, \bar{\theta}) w(\bar{\lambda}, \bar{\theta}) d\bar{\theta}.$$

Обозначив далее

$$\tilde{\Lambda}_{1/0}(\mathbf{z}/\bar{\lambda}) = \int_{(\lambda)} \Lambda_{1/0}(\mathbf{z}/\bar{\lambda}, \theta) w_{\lambda/\theta}(\bar{\lambda}) d\bar{\theta},$$

найдем

$$w_{\lambda/z}(\bar{\lambda}) = k_3 w(\bar{\lambda}) \tilde{\Lambda}_{1/0}(\mathbf{z}/\bar{\lambda}, \bar{\theta}).$$

Перейдя к пределу при $\Delta t \rightarrow 0$ в последнем выражении, получим

$$w_{\lambda/z(t)}(\lambda) = k_3 w(\lambda) \tilde{\Lambda}_{1/0}(z(t)/\lambda).$$

Здесь усредненное отношение правдоподобия имеет вид

$$\tilde{\Lambda}_{1/0}(z(t)/\lambda) = \int_{(\lambda)} \Lambda_{1/0}(z(t), \lambda, \theta) w_{\lambda/\theta}(\lambda, \theta) d\theta.$$

Прологарифмировав последнее выражение и полагая, что априорное распределение $w(\bar{\lambda})$ является равномерным, после дифференцирования по λ_k аналогично выводу выражения (9.29) получим окончательные уравнения для определения оптимальной оценки

$$\frac{\partial \tilde{\Lambda}_{1/0}(z(t)/\bar{\lambda})}{\partial \lambda_k} = 0$$

для всех $\lambda_k = \lambda_1, \dots, \lambda_m$.

Рассмотрим пример оценивания амплитуды сигнала с полностью известными параметрами.

Пусть на выходе устройства оценивания имеется сигнал в интервале $[0, T]$, все параметры которого, кроме амплитуды, известны, и шум с односторонней спектральной плотностью N_0 имеет вид.

$$z(t) = S(t, \lambda) + n(t); \quad (9.30)$$

$$S(t, \lambda) = \lambda S(t) \cos[\omega_0 t + \varphi(t) + \beta] = \lambda S_1(t);$$

$$S_1(t) = S(t) \cos[\omega_0 t + \varphi(t) + \beta].$$

Оцениванию подлежит амплитуда сигнала λ , постоянная на его длительности.

Оценивание будем производить по методу максимального правдоподобия, используя формулу (9.29), которая приобретает вид

$$\frac{\partial \ln \Lambda_{1/0}(z/\lambda)}{\partial \lambda} = 0.$$

Функционал отношения правдоподобия в случае приема смеси (9.30) имеет вид

$$\Lambda_{1/0}(z(t)/\lambda) = e^{\frac{1}{N_0} \left[2 \int_0^T z(t) \lambda S_1(t) dt - \int_0^T \lambda^2 S_1^2(t) dt \right]},$$

следовательно,

$$\frac{\partial \ln \Lambda_{1/0}(z(t)/\lambda)}{\partial \lambda} = \frac{2}{N_0} \int_0^T z(t) S_1(t) dt - \int_0^T \lambda S_1^2(t) dt = 0, \quad (9.31)$$

и оценка амплитуды сигнала определяется выражением

$$\hat{\lambda} = \frac{\int_0^T z(t) S_1(t) dt}{\int_0^T S_1^2(t) dt}. \quad (9.32)$$

Схема устройства оптимального оценивания следует непосредственно из соотношения (9.32). Определим среднее значение и дисперсию оценки $\hat{\lambda}$. Среднее значение оценки

$$M\{\hat{\lambda}\} = \{\hat{\lambda}\} = \frac{\int_0^T [\lambda S_1(t) + \langle n(t) \rangle] S_1(t) dt}{\int_0^T S_1^2(t) dt} = \lambda,$$

т.е. среднее значение оценки равно ее истинному значению. Такая оценка является состоятельной и несмещенной.

Дисперсия оценки

$$D\{\hat{\lambda}\} = M\{(\hat{\lambda} - \lambda)^2\} = \sigma_{\hat{\lambda}}^2 = M \left\{ \left[\frac{\int_0^T [\lambda S_1(t) + \langle n(t) \rangle] S_1(t) dt}{\int_0^T S_1^2(t) dt} \right]^2 \right\} =$$

$$= \frac{\int_0^T \int_0^T \langle n(t') n(t) \rangle S_1(t') S_1(t) dt' dt}{\left[\int_0^T S_1^2(t) dt \right]^2} = \frac{N_0}{2E_{(1)}}, \quad (9.33)$$

так как $\langle n(t')n(t) \rangle = R_n(t' - t) = \frac{N_0}{2} \delta(t' - t)$ и $E_{(1)} = \int_0^T S_1^2(t) dt$.

Относительная погрешность оценивания амплитуды сигнала

$$\frac{\sigma_\lambda}{\lambda} = \sqrt{\frac{N_0}{2E_{(1)}\lambda^2}} = \sqrt{\frac{N_0}{2E}}.$$

Убедимся в том, что полученная оценка является эффективной и, следовательно, достаточной. Из выражения (9.31) с учетом формулы (9.30) следует, что

$$\frac{\partial \ln \Lambda_{1/0}(z(t)/\lambda)}{\partial \lambda} = \frac{\partial \ln w(z(t)/\lambda)}{\partial \lambda} = \frac{2}{N_0} \int_0^T n(t) S_1(t) dt.$$

Тогда

$$D(\lambda) = \frac{1}{M \left\{ \left[\frac{\partial \ln w_{z/\lambda}(z)}{\partial \lambda} \right]^2 \right\}} = \frac{1}{M \left\{ \left[\frac{2}{N_0} \int_0^T n(t) S_1(t) dt \right]^2 \right\}} = \frac{N_0}{2E}.$$

Таким образом, дисперсия оценки совпадает с нижней граничной оценкой Крамера — Рао и, следовательно, она является эффективной и достаточной.

9.3. Оптимальная фильтрация непрерывных сообщений

9.3.1. Постановка и решение задачи фильтрации непрерывных сообщений

Полагаем, что наблюдаемый процесс $z(t)$ имеет вид

$$z(t) = S(t, \tilde{\lambda}(t)) + n(t). \quad (9.34)$$

Здесь $n(t)$ — гаусовский белый шум с нулевым средним значением и дельтаобразной функцией корреляции:

$$\langle n(t) \rangle = 0; \langle n(t_1)n(t_2) \rangle = \frac{N_0}{2} \delta(t_1 - t_2). \quad (9.35)$$

Полезный сигнал $S(t, \tilde{\lambda}(t))$ есть функция времени t и вектора (сообщения) $\tilde{\lambda}(t) = \{\lambda_1(t), \dots, \lambda_m(t)\}$, представляющего собой случайный векторный процесс. При этом известными считаются все

необходимые вероятностные характеристики случайного процесса $\bar{\lambda}(t)$ и помехи $n(t)$.

Задача заключается в том, чтобы на основании априорных сведений и по наблюдаемой реализации процесса $z(t)$ для каждого момента времени t сформировать апостериорную плотность вероятностей сообщения $\bar{\lambda}(t)$. В большинстве случаев требуется получить текущую оценку $\bar{\lambda}(t)$, наилучшую в смысле критерия минимума среднего риска. Различают несколько модификаций задачи построения оценок. При наблюдении процесса $z(t)$ в текущем интервале времени $[0, T]$ определяется оценка $\bar{\lambda}(t+T)$: при $T = 0$ имеет место задача текущей фильтрации; при $T > 0$ — задача фильтрации с предсказанием, или задача экстраполяции; при $T < 0$ — задача фильтрации с запаздыванием, или задача интерполяции. Критерии оптимальности, используемые в теории фильтрации, при фиксированном значении t совпадают с критериями теории оценок (см. подразд. 9.2).

Вектор $\lambda(t)$ и помеха $n(t)$ задаются либо в форме многомерных плотностей вероятностей, либо в виде дифференциальных уравнений с известными начальными условиями. В дальнейшем будем полагать, что информационное сообщение $\bar{\lambda}(t)$ является одномерным марковским процессом и определяется стохастическим дифференциальным уравнением

$$d\lambda(t)/dt = K_1(t, \bar{\lambda}) + n_\lambda(t); \quad \bar{\lambda}(0) = \lambda_0, \quad (9.36)$$

где $K_1(t, \bar{\lambda})$ — известная функция аргументов t и $\lambda(t)$.

В зависимости от вида уравнения наблюдаемого процесса (9.34) и уравнения сообщения (9.35) различают линейную фильтрацию, если уравнения (9.34) и (9.36) являются линейными относительно $\bar{\lambda}$ и нелинейную фильтрацию, если уравнение (9.34) или (9.36) содержит нелинейные функции сообщения $\bar{\lambda}$.

Считая $\lambda(t)$ случайным однокомпонентным процессом, описываемым уравнением (9.36), получим алгоритм задачи фильтрации. Случайный процесс $\lambda(t)$ является диффузионным марковским процессом, априорная плотность вероятности которого определяется уравнением Фоккера — Планка — Колмогорова:

$$\begin{aligned} \frac{\partial w(t, \lambda)}{\partial t} = & -\frac{\partial}{\partial \lambda} (K_1(t, \lambda)w(t, \lambda)) + \\ & + \frac{1}{2} \frac{\partial^2}{\partial \lambda^2} (K_2(t, \lambda) \equiv L(w(t, \lambda))). \end{aligned} \quad (9.37)$$

Здесь L — обозначения оператора Фоккера — Планка — Колмогорова. Заметим, что в рассматриваемом случае имеет место од-

нозначное соответствие между описаниями процесса $\lambda(t)$ в виде уравнений (9.36) и (9.37).

Апостериорная плотность вероятностей $w(t, \lambda)$ параметра $\lambda(t)$ удовлетворяет уравнению, которое впервые получил Р.Л. Стратонович и которое для рассматриваемого однокомпонентного марковского процесса имеет вид

$$\frac{\partial w_{ps}(t, \lambda)}{\partial t} = L\{w(t, \lambda)\} + (F(t, \lambda) - \langle F(t, \lambda) \rangle)w(t, \lambda). \quad (9.38)$$

Здесь производная по времени от логарифма функции правдоподобия

$$F(t, \lambda) = -\frac{1}{N_0} (z(t) - S(t, \lambda))^2, \quad (9.39)$$

а ее среднее значение по параметру λ

$$\langle F(t, \lambda) \rangle = \int_{-\infty}^{\infty} F(t, \lambda) w(t, \lambda) d\lambda. \quad (9.40)$$

Начальные условия для уравнения (9.37) определяются априорной плотностью вероятностей $w(0, \lambda)$ начальной координаты сообщения $\lambda(0) = \lambda_0$.

Апостериорная плотность вероятностей $w_z(t, \lambda)$ содержит всю доступную информацию о параметре $\lambda(t)$, которую можно извлечь из наблюдения $z(t)$ в интервале $[0, T]$ и из априорных сведений об этом параметре. Определив апостериорную плотность $w_z(t, \lambda)$, можно получить другие требуемые характеристики, например условное математическое ожидание $\hat{\lambda}(t)$, представляющее собой оптимальную оценку сообщения по критерию минимума среднего квадрата ошибки, или оценку, оптимальную по критерию максимума апостериорной вероятности. Для реализации оптимальных алгоритмов используются различные упрощающие предположения, которые будут рассмотрены далее.

9.3.2. Линейная фильтрация

Частным случаем рассмотренных общих выражений является линейная фильтрация сообщения. Пусть наблюдаемая смесь

$$z(t) = \lambda(t) + n(t).$$

Здесь шум имеет характеристики (9.35), а процесс $\lambda(t)$ задан линейным уравнением

$$d\lambda(t)/dt = -a\lambda(t) + n_\lambda(t), \quad (9.41)$$

где a — постоянный коэффициент, определяющий ширину спектра сообщения $\lambda(t)$.

Процесс $\lambda(t)$, заданный уравнением (9.41), можно сформировать из белого шума $n(t)$, характеристики которого известны:

$$\langle n_\lambda(t_1)n_\lambda(t_2) \rangle = \frac{N_\lambda}{2} \delta(t_1 - t_2),$$

где $\frac{N_\lambda}{2}$ — спектральная плотность шума $n_\lambda(t)$.

В качестве формирующего фильтра можно использовать фильтр нижних частот с постоянной времени $T_{RC} = 1/\alpha$.

При линейной фильтрации гауссовских процессов, каким является рассматриваемое сообщение, апостериорная плотность вероятностей $w(t, X)$ представляется гауссовским законом. Ее параметрами служат математическое ожидание и дисперсия. Подставляя гауссовскую плотность вероятностей в уравнение (9.38), можно прийти к следующей системе уравнений:

$$\frac{\partial \hat{\lambda}(t)}{\partial t} = -\alpha \hat{\lambda}(t) + \sigma_\lambda^2(t) \frac{\partial F(t, \hat{\lambda})}{\partial \hat{\lambda}}; \quad (9.42)$$

$$\frac{\partial \sigma_\lambda^2(t)}{\partial t} = \frac{N_\lambda}{2} - 2\alpha \sigma_\lambda^2(t) + \sigma_\lambda^2(t) \frac{\partial^2 F(t, \hat{\lambda})}{\partial \hat{\lambda}^2}. \quad (9.43)$$

Уравнение (9.42) определяет алгоритм формирования оптимальной оценки, а уравнение (9.43) — эволюцию апостериорной дисперсии. Эти уравнения можно преобразовать с учетом выражений (9.39) и (9.41). В результате функция $F(t, \hat{\lambda})$ и ее производные будут иметь следующий вид:

$$\begin{aligned} F(t, \hat{\lambda}) &= -\frac{1}{N_0} (z(t) - \hat{\lambda}(t))^2; \\ \frac{\partial F(t, \hat{\lambda})}{\partial \hat{\lambda}} &= \frac{2}{N_0} (z(t) - \hat{\lambda}(t)); \\ \frac{\partial^2 F(t, \hat{\lambda})}{\partial \hat{\lambda}^2} &= -\frac{2}{N_0}. \end{aligned} \quad (9.44)$$

Подставив выражения (9.44) в уравнения (9.42) и (9.43), получим:

$$\frac{\partial \hat{\lambda}(t)}{\partial t} = -\alpha \hat{\lambda}(t) + \sigma_\lambda^2(t) \frac{2}{N_0} (z(t) - \hat{\lambda}(t)); \quad (9.45)$$

$$\frac{d\sigma_\lambda^2(t)}{dt} = \frac{N_\lambda}{2} - 2\alpha \sigma_\lambda^2(t) - \sigma_\lambda^4 \frac{2}{N_0}. \quad (9.46)$$

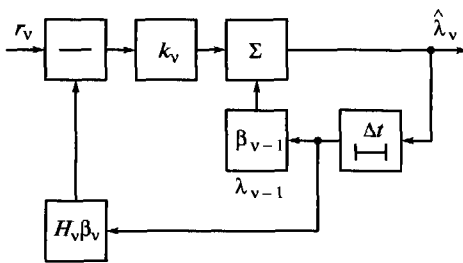


Рис. 9.6. Схема дискретного стационарного фильтра Калмана

Уравнения (9.45) и (9.46) описывают так называемый фильтр Калмана для непрерывного времени.

Рассмотрим теперь задачу фильтрации в дискретном времени. Наблюдение и сообщение в этом случае задаются скалярными разностными уравнениями:

$$z_v = H_v \lambda_v + n_v; \quad \lambda_v = \beta_{v-1} \lambda_{v-1} + n_{\lambda v},$$

где $H_v = H(t_v)$ и $\beta_v = \beta(t_v)$ — заданные функции времени; n_v и $n_{\lambda v}$ — гауссовские шумы с нулевыми средними значениями и дисперсиями D_v и $D_{\lambda v}$ соответственно; t_v — дискретное время с интервалом дискретизации $t_v - t_{v-1} = \Delta t$.

Результирующие уравнения для оценки $\hat{\lambda}_v$ и дисперсии $\sigma_{\lambda v}^2$, определяющие дискретный фильтр Калмана, носят рекуррентный характер и имеют следующий вид:

$$\hat{\lambda}_v = \beta_{v-1} \hat{\lambda}_{v-1} + H_v \left(\frac{\sigma_{\lambda v}^2}{D_v} \right) (z_v - H_v \beta_{v-1} \hat{\lambda}_{v-1}); \quad (9.47)$$

$$\frac{1}{\sigma_{\lambda v}^2} = (\beta_{v-1}^2 \sigma_{\lambda v-1}^2 + D_{\lambda v})^{-1} + \frac{H_v^2}{D_v}. \quad (9.48)$$

Схема, реализующая алгоритм (9.47), (9.48), приведена на рис. 9.6, где $K_v = H_v (\sigma_{\lambda v}^2 D_v)$. Заметим, что уравнение для дисперсии (9.48) так же, как и (9.43), не зависит от наблюдения, поэтому коэффициент $K(t_v) = K_v$ можно заранее вычислить, т. е. он представляет собой известную функцию времени.

При постоянных величинах β_v , D_v , $D_{\lambda v}$ и H_v и при условии существования установившегося значения дисперсии $\sigma_{\lambda \infty}^2 = \lim_{\sigma \rightarrow \infty} \sigma_{\lambda v}^2$

оценка может формироваться в соответствии с уравнением

$$\hat{\lambda}_v = \beta \hat{\lambda}_{v-1} + H \left(\frac{\sigma_{\lambda \infty}^2}{D_v} \right) (z_v - H \beta \hat{\lambda}_{v-1}). \quad (9.49)$$

В этом случае оценка является асимптотически оптимальной ($t \rightarrow \infty$), а фильтр, описываемый уравнением (9.49), — стационарным. Такой фильтр проще в реализации, чем нестационарный, однако в течение переходного процесса дисперсия оценки в нем больше, чем в нестационарном оптимальном фильтре.

9.3.3. Нелинейная фильтрация

Задача нелинейной фильтрации параметров (сообщения) возникает, если уравнение, описывающее сообщение, нелинейное, или, если уравнения наблюдения содержат нелинейные функции. На практике радиосигналы являются нелинейными функциями сообщений $\lambda(t)$, содержащихся, например в фазе, задержке или частоте сигнала $s(t, \lambda)$.

Рассмотрим наблюдение, заданное выражением (9.34), где сообщение $\lambda(t)$ будем считать случайным процессом, определяемым уравнением (9.36). Уравнения нелинейной фильтрации Стратоновича, характеризующие эволюцию апостериорной плотности вероятностей, определяются соотношениями (9.38)...(9.40). Интегродифференциальное уравнение (9.38) в частных производных не имеет в общем случае точного решения (за исключением линейной фильтрации), однако можно найти его приближенные решения.

Наибольшее распространение получил способ решения, основанный на аппроксимации апостериорной плотности $w(t, \lambda)$ гауссовской плотностью вероятностей, — гауссовское приближение. Это приближение тем точнее, чем больше отношение сигнал/шум и чем меньше ошибки фильтрации. При этом алгоритмы, основанные на гауссовской аппроксимации, являются оптимальными. Уравнения гауссовского приближения имеют вид (9.42), (9.43).

Получим уравнения, подобные уравнениям (9.42), (9.43) для случая нелинейной фильтрации. При наблюдении вида (9.34) первая и вторая производные от функции $F(t, \lambda)$, задаваемой выражением (9.39), в точке оценки $\hat{\lambda}$, запишем следующим образом:

$$F_{\lambda} = F'(\hat{\lambda}, t) = \left. \frac{\partial}{\partial \lambda} F(t, \lambda(t)) \right|_{\lambda=\hat{\lambda}} = \frac{2}{N_v} (z(t) - S(t, \hat{\lambda})) \frac{\partial S(t, \hat{\lambda})}{\partial \lambda}. \quad (9.50)$$

Для неэнергетических параметров возможно использование упрощенных уравнений:

$$F_{\lambda} = \frac{2}{N_v} z(t) \frac{\partial S(t, \hat{\lambda})}{\partial \lambda};$$

$$F_{\lambda\lambda} = F''(\hat{\lambda}, t) = \left. \frac{\partial^2}{\partial \lambda^2} F(t, \lambda(t)) \right|_{\lambda=\hat{\lambda}} =$$

$$= \frac{2}{N_v} \left[(z(t) - S(t, \hat{\lambda})) \frac{\partial^2 S(t, \hat{\lambda})}{\partial \lambda} \left(\frac{\partial S(t, \hat{\lambda})}{\partial \lambda} \right)^2 \right]. \quad (9.51)$$

Подставив выражения (9.50) и (9.51) в уравнения (9.42) и (9.43), получим следующие уравнения нелинейной фильтрации:

$$\dot{\hat{\lambda}} = K_1 \hat{\lambda} + \sigma_\lambda^2(t) F'(\hat{\lambda}, t); \quad (9.52)$$

$$\dot{\hat{\sigma}}_\lambda(t) = K_2(\hat{\lambda}) + 2\bar{\sigma}_\lambda^2(t) K_1(\hat{\lambda}) + \hat{\sigma}_4(t) F''(\hat{\lambda}, t). \quad (9.53)$$

Сравнив уравнения линейной фильтрации (9.45) и (9.46) с уравнениями нелинейной фильтрации (9.52) и (9.53) с учетом выражения (9.51), увидим, что в уравнение для дисперсии (9.53) в случае нелинейной фильтрации входят текущая реализация $z(t)$ и текущая оценка $\hat{\lambda}$. Следовательно, уравнение (9.53) необходимо решать совместно с уравнением (9.52) в текущем времени, что усложняет построение нелинейного фильтра.

Для упрощения нелинейного фильтра прибегают к усреднению зависящих от времени членов в уравнении (9.53), в частности к следующей замене:

$$\overline{F''(\hat{\lambda}, t)} = \overline{F_{\lambda\lambda}} = \frac{\overline{\partial^2 F(t, \lambda(t))}}{\partial \lambda} \approx -\frac{2}{N_v} \left(\frac{\partial S(t, \hat{\lambda})}{\partial \lambda} \right)^2.$$

Заметим, что черта сверху обозначает усреднение по времени. При усреднении выражения (9.53) полагаем, что существует стационарное значение апостериорной дисперсии (кумулянт):

$$\bar{\sigma}_\lambda^2(t) = k_\lambda; \quad \frac{\partial \bar{\sigma}_\lambda^2(t)}{\partial t} = 0.$$

В результате усреднения дифференциальное уравнение (9.53) преобразуется в алгебраическое, и вместо выражений (9.52) и (9.53) получаем следующие уравнения:

$$\dot{\hat{\lambda}} = K_1(\hat{\lambda}) + k_\lambda F_\lambda; \quad (9.54)$$

$$\bar{F}_{\lambda\lambda} k_\lambda^2 + 2K_1'(\hat{\lambda}) k_\lambda + K_2(\hat{\lambda}) = 0. \quad (9.55)$$

Таким образом, приведенные уравнения позволяют осуществить структурный синтез устройств нелинейной фильтрации и определить стационарные значения для дисперсий ошибок фильтрации. Получаемые в результате синтеза устройства относятся к квазиоптимальным, поскольку в переходном режиме они обладают большими ошибками, чем оптимальные устройства.

Пусть теперь процесс $\lambda(t)$ задан уравнением (9.41). Для этого процесса коэффициенты сноса и диффузии соответственно $K_1(\lambda) = -\alpha\lambda$, $K_2(\lambda) = \frac{N_\lambda}{2}$. При этом уравнения (9.54) и (9.55) приобретают следующий вид:

$$\begin{aligned} \dot{\hat{\lambda}} &= -\alpha\hat{\lambda} + k_\lambda F_\lambda; \\ \bar{F}_{\lambda\lambda} k_\lambda^2 - 2\alpha k_\lambda + \frac{N_\lambda}{2} &= 0. \end{aligned} \quad (9.56)$$

Последнее уравнение (9.56) позволяет определить среднее значение дисперсии оценки параметра $\lambda(t)$

$$k_\lambda = \frac{2\alpha \pm \sqrt{4\alpha^2 - 2\bar{F}_{\lambda\lambda} N_\lambda}}{2\bar{F}_{\lambda\lambda}}, \quad (9.57)$$

которое используется для оценки помехоустойчивости систем передачи непрерывных сообщений.

Уравнения квазиоптимальной нелинейной фильтрации можно обобщить на многомерный случай. Рассмотрим примеры применения алгоритмов нелинейной фильтрации.

9.3.4. Примеры демодуляции непрерывных сообщений на основе нелинейной фильтрации

Демодуляция ФМ-сигналов

Рассмотрим сначала демодуляцию непрерывного информационного сообщения в системе с фазовой модуляцией сигналов. Пусть по каналу связи с гауссовским шумом передается непрерывное сообщение с помощью фазомодулированных сигналов вида

$$S(t, \lambda) = A \cos[\omega_0 t + m_\phi \lambda(t) + \phi]. \quad (9.58)$$

Здесь амплитуда A , круговая частота ω_0 и начальная фаза ϕ сигнала — постоянные величины, а индекс фазовой модуляции m_ϕ характеризует крутизну фазовой характеристики.

Будем считать, что непрерывное сообщение $\lambda(t)$ аппроксимируется случайным марковским процессом, описываемым уравнением (9.41). Для определения структуры оптимального приемника и оценки его помехоустойчивости используем уравнения (9.56) и (9.57).

В соответствии с выражениями (9.39) и (9.58) сигнальная функция

$$\begin{aligned} F(\lambda, t) &= \frac{1}{N_0} [2z(t)S(t, \lambda) - S^2(t, \lambda)] = \\ &= \frac{1}{N_0} \{2z(t)A \cos[\omega_0 t + m_\phi \lambda(t) + \phi] - S^2(t, \lambda)\} = \\ &= \frac{1}{N_0} \left\{ 2z(t)A \cos[\omega_0 t + m_\phi \lambda(t) + \phi] - \right. \\ &\quad \left. - \frac{A^2}{2} [1 + \cos 2(\omega_0 t + m_\phi \lambda(t) + \phi)] \right\}, \quad (9.59) \end{aligned}$$

откуда

$$F_{\lambda} = \frac{\partial F(\lambda, t)}{\partial \lambda} = \frac{1}{N_0} \{-2z(t)Am_{\phi} \sin[\omega_0 t + m_{\phi}\lambda(t) + \phi]\}. \quad (9.60)$$

Слагаемое удвоенной частоты в формуле (9.59) опущено, так как оно отфильтровывается приемным устройством.

Заметим, что при фазовой модуляции оцениваемый параметр $\lambda(t)$ является неэнергетическим, поэтому в формуле (9.59) сигнал $S^2(t, \lambda)$ с учетом отбрасываемой второй гармоники не зависит от λ , и производные F_{λ} и $F_{\lambda\lambda}$ можно вычислить без учета этого слагаемого. Подставив F_{λ} из выражения (9.60) в формулу (9.56), получим

$$\dot{\hat{\lambda}} = -\alpha \hat{\lambda} - 2 \frac{k_{\lambda} Am_{\phi}}{N_0} z(t) \sin[\omega_0 t + m_{\phi}\lambda(t) + \phi]. \quad (9.61)$$

Полученное дифференциальное уравнение позволяет составить структурную схему оптимального приемника непрерывных сообщений, передаваемых сигналами с фазовой модуляцией (рис. 9.7, а). Здесь принимаемая смесь $z(t)$ умножается на опорное напряжение подстраиваемого гетеродина (ПГ), в результате

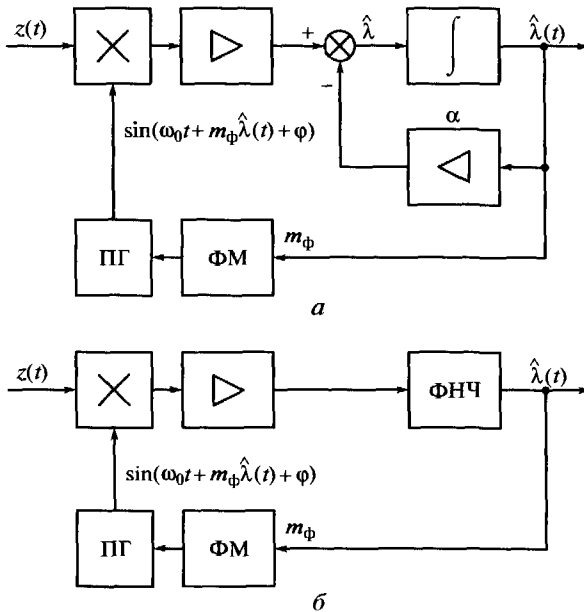


Рис. 9.7. Два варианта (а, б) структурных схем приемника непрерывного сообщения в системе с фазовой модуляцией

чего образуется низкочастотное напряжение, обусловленное рас-
согласованием фазы принимаемого сигнала и подстраиваемого
гетеродина и являющееся сообщением $\lambda(t)$, и высокочастотное
напряжение удвоенной частоты, которое в дальнейшем отфиль-
тровывается. Усилитель с коэффициентом $-2k_\lambda Am_\phi / N_0$ служит
для масштабирования, а сумматор, интегратор и элемент обрат-
ной связи являются эквивалентом безынерционного усилителя с
коэффициентом усиления $\frac{1}{\alpha}$ и инерционного звена с постоянной
времени $T = \frac{1}{\alpha}$.

Чтобы убедиться в этом, представим левую часть уравнения
(9.61) в форме $\hat{\lambda} = D\lambda$ (где D — оператор дифференцирования):

$$D\hat{\lambda} = -\alpha\hat{\lambda} - 2\frac{k_\lambda Am_\phi}{N_0} z(t) \sin[\omega_0 t + m_\phi \hat{\lambda}(t) + \phi],$$

откуда

$$\hat{\lambda} = z(t) \sin[\omega_0 t + m_\phi \hat{\lambda}(t) + \phi] \left[-2\frac{k_\lambda Am_\phi}{\alpha N_0} \right] \frac{1}{1+TD}.$$

Из этого уравнения следует второй вариант схемы оптималь-
ного приемника (рис. 9.7, б). В обоих вариантах оценка инфор-
мационного сообщения $\lambda(t)$ подается на фазовый модулятор (ФМ)
с крутизной фазовой характеристики m_ϕ для управления фазой
подстраиваемого гетеродина.

Для оценки помехоустойчивости оптимального приемника сна-
чала определим дисперсию оценки фильтрации по формуле (9.57),
а затем, используя формулу (9.60), найдем вторую производ-
ную:

$$F_{\lambda\lambda} = \frac{\partial^2 F(\lambda, t)}{\partial \lambda^2} = -\frac{2Am_\phi^2}{N_0} z(t) \cos[\omega_0 t + m_\phi \lambda(t) + \phi].$$

Для определения среднего значения $\bar{F}_{\lambda\lambda}$ необходимо учесть,
что $z(t) = S(t, \lambda) + n(t)$, причем $\langle n(t) \rangle = 0$, поэтому формула
(9.57) в рассматриваемом случае примет вид

$$k_\lambda = \frac{\sqrt{4a^2 + 2\frac{A^2 m_\phi^2}{N_0} N_\lambda - 2a}}{2\frac{A^2 m_\phi^2}{N_0}} = \frac{aN_0}{A^2 m_\phi^2} \left[\sqrt{1 + \frac{A^2 m_\phi^2}{N_0} N_\lambda} - 1 \right].$$

Обозначим отношение сигнал/помеха на входе приемника
(в полосе сообщения) $q = \frac{A^2}{2aN_0}$, тогда

$$k_\lambda = \frac{1}{2qm_\Phi^2} \left(\sqrt{1 + q \frac{m_\Phi^2 N_\lambda}{a}} - 1 \right)$$

Можно показать, что дисперсия информационного сообщения вида (9.41) в установившемся режиме $\sigma_\lambda^2 = \frac{N_\lambda}{4a}$, а дисперсия фазы сигнала $\sigma_\Phi^2 = m_\Phi^2 \sigma_\lambda^2 = \frac{m_\Phi^2 N_\lambda}{4a}$, следовательно, ошибка фильтрации

$$k_\lambda = \frac{1}{2qm_\Phi^2} (\sqrt{1 + 4q\sigma_\Phi^2} - 1).$$

Относительная дисперсия фильтрации при фазовой модуляции определяется отношением $k_\lambda / \sigma_\lambda^2$:

$$\delta_{\Phi M}^2 = \frac{k_\lambda}{\sigma_\lambda^2} = \frac{1}{2qm_\Phi^2 \sigma_\lambda^2} (\sqrt{1 + 4q\sigma_\Phi^2} - 1);$$

$$\delta_{\Phi M}^2 = \frac{1}{2q\sigma_\lambda^2} (\sqrt{1 + 4q\sigma_\Phi^2} - 1).$$

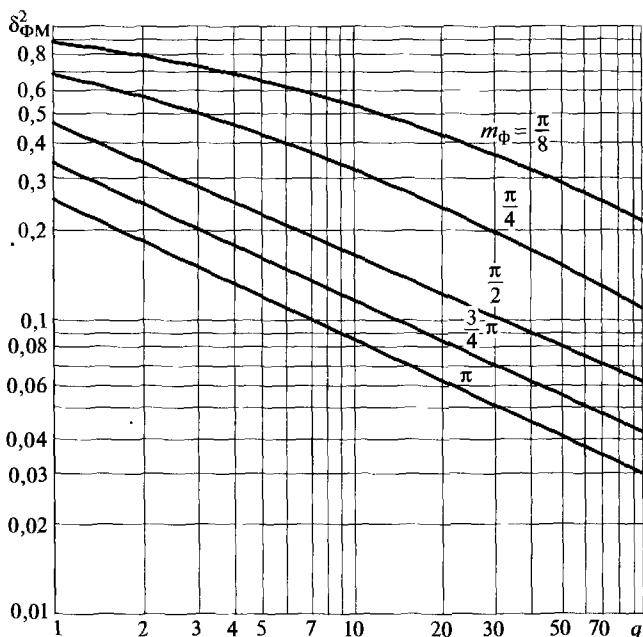


Рис. 9.8. Зависимости квадрата относительной ошибки фильтрации непрерывного сообщения в системе с фазовой модуляцией от отношения сигнал/шум при различных значениях индекса модуляции

На рис. 9.8 представлены кривые зависимостей квадрата относительной ошибки фильтрации от отношения сигнал/помеха на входе оптимального приемника при различных значениях индекса фазовой модуляции. Из графических зависимостей видно, что значение относительной ошибки фильтрации уменьшается с ростом отношения сигнал/помеха и при увеличении индекса фазовой модуляции m_{ϕ} . Однако при возрастании индекса фазовой модуляции расширяется спектр сигнала и, следовательно, — ширина полосы частот, необходимая для передачи сообщения. Возможность расширения спектра определяется используемым диапазоном радиочастот, направленностью антенных устройств и другими факторами.

Демодуляция ЧМ-сигналов

Сигналы с частотной модуляцией в месте приема можно представить следующим образом:

$$S(t, \lambda) = A \cos[\omega_0 t + \psi(t)]; \quad (9.62)$$

$$\psi(t) = m_f \int_0^t \lambda(t) dt + \varphi_0. \quad (9.63)$$

Здесь A , ω_0 , φ_0 и индекс частотной модуляции m_f — постоянные величины.

Передача непрерывных сообщений при частотной модуляции сигналов осуществляется одним из двух способов: с использованием предискажений передаваемых сообщений и без них.

При первом способе сообщение подвергается искажениям по частоте с помощью линейного устройства, частотная характеристика которого имеет коэффициент передачи, прямо пропорциональный частоте. На приемной стороне после демодуляции сообщение подвергается обратному преобразованию по частоте. Введение предискажений позволяет существенно ослабить высокочастотные шумы, возникающие на выходе наиболее часто применяемых детекторов с линейной частотной характеристикой.

Как известно, умножение спектра сообщений на частоту соответствует взятию производной от сообщения во временной области. Следовательно, введение линейных частотных предискажений, как следует из формул (9.62) и (9.63), равносильно формированию сигнала с фазовой модуляцией, для которого оптимальным является приемник, рассмотренный ранее. Таким образом, если на передающей стороне сообщения подвергаются линейным частотным предискажениям, то оптимальным для них будет приемник с фазовой демодуляцией сигналов.

Далее будем рассматривать сообщения, которые не подвергаются предсказаниям, что соответствует наличию в месте приема сигналов, представленных выражениями (9.62) и (9.63). При этом будем полагать, что сигналы принимаются на фоне гауссовых помех, а сообщение $\lambda(t)$ аппроксимируется случайным марковским процессом, описываемым уравнением (9.41).

Можно показать, что синтез оптимального приемника в рассматриваемом случае должен производиться в соответствии с дифференциальными уравнениями

$$\dot{\hat{\lambda}} = -a\dot{\lambda} + k_{\psi\lambda} F_{\psi}; \quad \dot{\hat{\psi}} = m_f \dot{\lambda} + k_{\psi\psi} F_{\psi}. \quad (9.64)$$

Поскольку частота является неэнергетическим параметром, то

$$F_{\psi} = -\frac{2}{N_0} z(t) A \sin(\omega_0 t + \psi),$$

и первое уравнение (9.64) принимает вид

$$\dot{\hat{\lambda}} = -a\dot{\lambda} + k_{\psi\lambda} \frac{2}{N_0} z(t) A \sin(\omega_0 t + \psi).$$

После введения оператора D получим

$$a\dot{\hat{\lambda}} + \hat{\lambda} = a\tilde{\lambda} \left(1 + \frac{1}{a} D \right) = -k_{\psi\lambda} \frac{2}{N_0} z(t) A \sin(\omega_0 t + \psi),$$

откуда

$$\hat{\lambda} = -\frac{k_{\lambda}}{1+TD} z(t) A \sin(\omega_0 t + \psi); \quad (9.65)$$

$$k_{\lambda} = \frac{2k_{\psi\lambda}}{aN_0}; \quad T = \frac{1}{a}.$$

Подставив формулу для оценки параметра (9.65) во второе уравнение (9.64), найдем

$$\dot{\hat{\psi}} = -\left(\frac{m_f k_{\lambda}}{1+TD} + \frac{2k_{\psi\psi}}{N_0} \right) z(t) A \sin(\omega_0 t + \psi),$$

или

$$\dot{\hat{\psi}} = -s_y \left(\frac{m_f}{s_y} \frac{k_{\lambda}}{1+TD} + k_{\psi} \right) z(t) A \sin(\omega_0 t + \psi), \quad (9.66)$$

где $k_{\psi} = \frac{2k_{\psi\psi}}{s_y N_0}$; s_y — крутизна характеристики управляющего элемента.

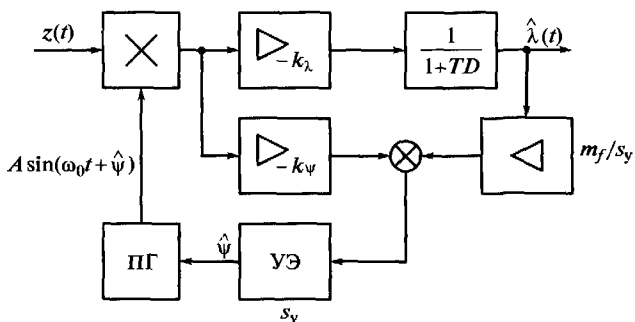


Рис. 9.9. Структурная схема приемника непрерывного сообщения в системе с частотной модуляцией

На рис. 9.9 изображена схема оптимального приемника нелинейной фильтрации для сигналов с частотной модуляцией, выполненная на основе уравнений (9.65) и (9.66). Здесь входная смесь $z(t)$ умножается на напряжение $A \sin(\omega_0 t + \hat{\psi})$ подстраиваемого гетеродина (ПГ) и поступает на безынерционный усилитель с коэффициентом передачи $-k_\lambda$ и инерционное звено с постоянной времени T , на выходе которого получаем оценку информационного сообщения $\hat{\lambda}(t)$. Как следует из формулы (9.66), оценка производной от фазы (оценка частоты) осуществляется на основе суммы двух напряжений: напряжения на выходе умножителя $z(t)A \sin(\omega_0 t + \hat{\psi})$, которое подается на сумматор с весовым коэффициентом $-k_\psi$, и напряжения на выходе инерционной цепи, подаваемого на сумматор с коэффициентом передачи m_f/s_y . С выхода сумматора напряжение поступает на управляющий элемент (УЭ) с крутизной характеристики s_y , изменяющий частоту подстраиваемого гетеродина.

Относительная ошибка фильтрации в системе с частотной модуляцией определяется по формуле

$$\delta_{\text{чМ}}^2 = \frac{1}{2q\beta_{\text{чМ}}^2} \left[1 + 2\beta_{\text{чМ}}\sqrt{q} - \sqrt{1 + 4\beta_{\text{чМ}}\sqrt{q}} \right] \sqrt{1 + 4\beta_{\text{чМ}}\sqrt{q}},$$

где $q = \frac{A^3}{2aN_0}$; $\beta_{\text{чМ}} = \sigma_\omega / a = m_f \sigma_\lambda / a$ — индекс частотной модуляции.

На рис. 9.10 представлены результаты вычислений квадрата относительной ошибки фильтрации, из которых следует, что относительная погрешность фильтрации уменьшается с ростом отношения q сигнал/помеха на входе приемника (шумы в полосе информационного сообщения) и с увеличением индекса частотной модуляции $\beta_{\text{чМ}}$. Однако увеличение $\beta_{\text{чМ}}$ сопровождается расши-

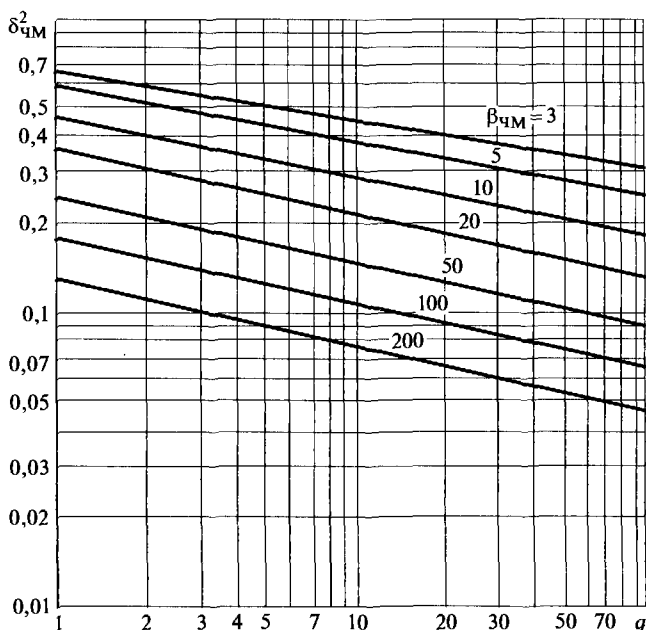


Рис. 9.10. Зависимости квадрата относительной ошибки фильтрации непрерывного сообщения в системе с частотной модуляцией от отношения сигнал/шум при различных значениях индекса модуляции

рением полосы частот, необходимой для передачи информационного сообщения.

Помехоустойчивость систем с частотной модуляцией оказывается значительно более высокой по сравнению с системами с амплитудной модуляцией, особенно при больших $\beta_{чм}$. При этом по помехоустойчивости частотная модуляция близка к фазовой модуляции, но имеет несколько более широкий спектр.

Контрольные вопросы

1. Поясните, что такое выигрыш и обобщенный выигрыш при приеме непрерывных сообщений и в чем состоит их различие.
2. Сравните различные методы модуляции по значению обобщенного выигрыша.
3. Что представляет собой функция потерь при оценке параметров и в чем заключается различие квадратичной и простой функций потерь?
4. В чем состоит смысл следующих критериев оптимальности при оценке параметров: критерий минимума среднего квадрата ошибки и критерий максимума апостериорной вероятности?

5. Что представляет собой решающее правило в задачах оценки параметров?

6. Чем различаются байесовская оценка и оценка максимального правдоподобия?

7. От чего зависит дисперсия ошибки оптимальной оценки параметра?

8. Поясните, в чем состоит суть текущей фильтрации, экстраполяции и интерполяции.

9. В чем заключается физический смысл уравнения Стратоновича для апостериорной плотности вероятности фильтруемого параметра?

10. Поясните принцип работы фильтра Калмана.

11. В чем состоит различие уравнений нелинейной фильтрации и линейной?

12. Поясните принцип работы демодулятора ЧМ-сигналов, полученного с использованием методов нелинейной фильтрации.

13. Поясните принцип работы демодулятора ФМ-сигналов, полученного с использованием методов нелинейной фильтрации.

Глава 10

ОСНОВЫ ТЕОРИИ ПЕРЕДАЧИ И КОДИРОВАНИЯ ИНФОРМАЦИИ

10.1. Определение количества информации

Рассмотрим источник сообщений, на выходе которого может получиться последовательность сигналов $a_1, a_2, \dots, a_k, \dots, a_n$ с вероятностями $P(a_1), P(a_2), \dots, P(a_k), \dots, P(a_n)$. При этом возникает вопрос о том, сколько информации получают при приеме любого из этих символов.

Пусть, например, $P(a_1) = 1$, тогда все остальные вероятности $P(a_k) = 0$, где $k = 1, n$. Понятно, что в этом случае заранее известно о поступлении наперед заданного сообщения. Следовательно, и количество получаемой информации равно нулю. Если теперь символы появляются с различными вероятностями, то самое маловероятное сообщение будет наиболее неожиданным. Количество получаемой информации в этой ситуации должно быть наибольшим. Иными словами, количество информации, содержащейся в сообщении, может быть функционально связано с вероятностью появления этого сообщения, т. е.

$$i(a_k) = f[P(a_k)].$$

При этом необходимо выполнение следующих требований.

1. Количество информации должно отвечать свойству аддитивности, т. е. при передаче двух или более независимых сообщений общее количество информации должно равняться сумме информации от каждого из них в отдельности.

2. Количество информации, содержащийся в заранее известном сообщении (переданном с вероятностью $P(a_k) = 1$), должно равняться нулю.

В наибольшей степени этим требованиям удовлетворяет логарифмическая функция, поэтому для определения частного количества информации, содержащегося в k -м сообщении, запишем

$$i(a_k) = -\log_x P(a_k) \geq 0.$$

Действительно, именно эта функция удовлетворяет названным условиям:

$$i(a_k) = -\log_x P(a_k) = 0; P(a_k) = 1,$$

$$i(a_i, a_k) = i(a_i) + i(a_k) = -[\log_x P(a_i) + \log_x P(a_k)]; i \neq k.$$

Заметим, что выбор основания логарифма не имеет существенного значения, но удобнее всего применять основание, равное двум ($x = 2$). Это связано с тем, что в телекоммуникационных системах и вычислительной технике чаще используются двоичные сигналы. Единица количества информации в этом случае называется бит (от *англ.* binary digit — двоичная цифра, или binary unit — двоичная единица).

В некоторых случаях, как правило в теоретических научных исследованиях, используется натуральная единица (сокращенно «нат» с основанием $x = e$), которая в $\log_2 e = 1,443$ раза больше бита. Здесь и далее будет применяться двоичная единица измерения количества информации, поэтому запись $-\log P(a_k)$ всегда означает, что основание логарифма равно двум.

В заключение отметим, что с введением понятия количества информации слово «информация» приобрело как бы два значения: абстрактное и конкретное, или качественное и количественное. С одной стороны, под информацией понимают некоторую конкретную информацию, содержащуюся в сообщении, а с другой стороны, — ее численную меру, т. е. некоторое абстрактное количество информации, вычисленное в битах, которое содержится в интересующем нас сообщении. Поэтому термин «информация» будем применять для описания конкретной информации, а «количество информации» — для определения численной меры абстрактной информации, содержащейся в передаваемых сообщениях.

10.2. Энтропия источника дискретных сообщений и ее свойства

Рассмотрим вновь источник сообщений. Будем полагать, что он является стационарным и содержит N различных дискретных сообщений. Любое из этих сообщений выдается источником случайно с вероятностью $P(a_k)$, $k = \overline{1, N}$. В общем случае вероятности могут быть различными, т. е. $P(a_1), P(a_2), \dots, P(a_N)$ — произвольные положительные числа, сумма которых равна единице. Следовательно, определенное ранее частное количество информации может быть случайным значением.

Более удобной характеристикой, которая описывает среднее количество информации, выдаваемой источником сообщений,

является энтропия. В теории связи это понятие было введено К. Шенноном. Данную величину принято определять как математическое ожидание количества информации, приходящегося на один элемент сообщения:

$$H(A) = M\{-\log P(a_k)\} = \sum_{k=1}^N P(a_k) \log \frac{1}{P(a_k)}.$$

Нетрудно заметить, что это выражение по форме идентично выражению, полученному в термодинамике для величины, также называемой энтропией, где это понятие характеризует неопределенность системы в некоторый момент времени. Величину $H(A)$ также можно рассматривать как меру неопределенности сообщения до момента его получения. Иначе говоря, это мера неожиданности, или непредсказуемости, выдачи источником сообщений.

Приведенное выражение, определяющее энтропию, является справедливым только для источников сообщений без памяти, т. е. для сообщений, которые выдаются статистически независимо. Примером такого источника может быть обезьяна, хаотично нажимающая клавиши телетайпа или клавиатуры компьютера. В то же время любой оператор компьютера, работающий с каким-либо текстом, является примером источника с памятью. В этом случае каждое последующее сообщение, логически связанное с предыдущим, содержит в себе значительно меньше неопределенности, или непредсказуемости. Следовательно, значение энтропии уменьшается.

Остановимся теперь на свойствах энтропии.

1. Энтропия любого источника сообщений всегда положительная, т. е. $H(A) \geq 0$. Это следует из условий $0 \leq P(a_k) \leq 1$, $\log P(a_k) \leq 0$, $-P(a_k) \log P(a_k) \geq 0$. Энтропия равна нулю, т. е. $H(A) = 0$, если источник выдает одно сообщение с вероятностью $P(a_k) = 1$, а остальные — с нулевой вероятностью.

2. Энтропия источника без памяти, содержащего сообщения с фиксированным объемом алфавита N , имеет максимальное значение при условии равновероятной выдачи сообщений:

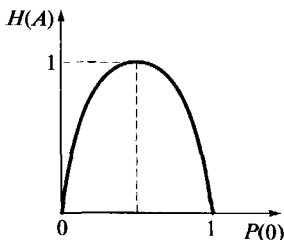
$$H_{\max}(A) = \log N \text{ при } P(a_1) = P(a_2) = \dots = P(a_N).$$

В частном случае, когда источник выдает всего два сообщения: символы $a_1 = 0$ и $a_2 = 1$, энтропия максимальная и составляет 1 бит при $P(0) = P(1) = 0,5$. Чтобы показать это, введем обозначения $P(0) = P$, $P(1) = 1 - P$. Тогда

$$\begin{aligned} H(A) &= -P(0) \log P(0) - P(1) \log P(1) = \\ &= -P \log P - (1 - P) \log(1 - P), \end{aligned}$$

откуда видно, что $H(A) = 0$ при $P(0) = 0$ и $P(1) = 1$, или при $P(0) = 1$ и $P(1) = 0$. Максимальная энтропия достигается при $P(0) = P(1) = 0,5$, т. е.

Рис. 10.1. Зависимость энтропии двоичного источника без памяти от $P(0)$



$$H_{\max}(A) = 0,5 \log 2 + 0,5 \log 2 = 1 \text{ бит.}$$

Зависимость энтропии двоичного источника без памяти от $P(0) = 1 - P(1)$ показана на рис. 10.1.

3. Энтропия аддитивна. Пусть ξ и η — сообщения от двух независимых источников. Количество информации $H(\xi, \eta)$ при совместном получении этих сообщений равно сумме значений, характеризующих уменьшение неопределенности с получением каждого из сообщений в отдельности, т.е. $H(\xi, \eta) = H(\xi) + H(\eta)$, что вытекает из свойства логарифмической функции.

10.3. Избыточность и производительность источника дискретных сообщений •

В качестве источника сообщений рассмотрим оператора, который вводит в компьютер текст на русском языке. Очевидно, что буквы в тексте появляются с разными вероятностями. Так, буква *А* передается значительно чаще, чем *Ц* или *Ю*. Кроме того, появление очередной буквы зависит от предыдущей. Ясно, что после гласных не появится *Ь*, *Ъ* или *Ы*. Весьма редким будет появление подряд трех букв *Е* (в слове «змеед»). Таким образом, на выходе источника с памятью (с зависимыми сообщениями) неопределенность оказывается меньше, чем при отсутствии памяти, когда сообщения появляются хаотично (как в примере обезьяны за компьютером). В результате уменьшается среднее количество информации, доставляемой каждой буквой.

Следовательно, если необходимо передать одинаковое количество информации от источников с памятью и без памяти, то в первом случае придется увеличить количество букв или символов. Тогда при меньшей информативности каждой буквы (символа) источника с памятью их большее число будет способствовать увеличению количества информации этого источника до значения, равного значению энтропии источника без памяти.

Таким образом, избыточность источника формально можно определить выражением

$$\rho = \frac{\log N - H(A)}{\log N} = 1 - \frac{H(A)}{\log N},$$

из которого видно, что чем больше энтропия, тем меньше избыточность источника, и наоборот. Ясно также, что избыточность принимает значения в пределах $0 \leq \rho \leq 1$.

Избыточность характеризует среднее число букв (символов) n , используемых источником сообщений для передачи заданного количества информации, относительно необходимого числа n_{\min} при равновероятном и независимом появлении букв.

Следовательно, избыточность можно определить следующим образом:

$$\rho = (n - n_{\min}) / n = 1 - n_{\min} / n.$$

Значение $\mu = H(A) / \log N = n_{\min} / n$ называется коэффициентом сжатия. Этот коэффициент показывает, до какого значения без потери информации можно сжимать передаваемые сообщения, если устранить содержащуюся в них избыточность. Например, при передаче телеграмм из текста исключают союзы и знаки препинания, которые легко восстанавливаются при чтении на основании известных правил.

Очевидно, что избыточность приводит к увеличению времени передачи сообщений и излишней загрузке каналов связи, а следовательно, к снижению эффективности их использования. Вместе с тем было бы неверным всегда рассматривать избыточность как признак несовершенства источника сообщений. В ряде случаев она бывает полезной. Наличие зависимостей (корреляции) между буквами и словами текста дает возможность восстанавливать его при искажении отдельных букв, т. е. избыточность можно использовать для повышения достоверности передачи информации в условиях воздействия помех.

Помимо избыточности важным параметром, характеризующим любой источник с фиксированной скоростью выдачи сообщений $v_{\text{и}} = 1 / T_{\text{и}}$ символов в секунду, является его производительность, которую определяют как энтропию в единицу времени (в секунду):

$$H'(A) = v_{\text{и}} H(A).$$

При максимальной энтропии, равной $\log N$, информационная скорость источника $R_{\text{и}} = \frac{\log N}{T_{\text{и}}}$, бит/с.

Физический смысл производительности — это среднее количество информации, которое выдается источником в течение одной секунды непрерывной работы.

10.4. Энтропия источника непрерывных сообщений

Рассмотрим источник непрерывных сообщений, на выходе которого в каждый момент времени t появляются сигналы $S(t)$. Эти сигналы с бесконечно малой вероятностью могут принимать бесконечное множество значений. Если бы сообщения могли передаваться по каналу абсолютно точно без искажений, то они содержали бы неограниченное количество информации. На самом деле из-за наличия помех и искажений количество информации, получаемой от такого источника, следует определять как разность энтропии до получения информации и после. Эта разность в отличие от абсолютного значения энтропии источника непрерывных сообщений конечна. Для ее определения обобщим рассмотренное ранее понятие энтропии источника дискретных сообщений на случай непрерывных сообщений, используя естественный предельный переход.

Обратимся к сигналу $S(t)$, возможные значения которого в момент времени t характеризуются плотностью распределения вероятностей $\omega(S)$, показанной на рис. 10.2.

Произведем квантование области значений S с шагом ΔS . Вероятность попадания некоторого конкретного значения сигнала S в пределы i -го шага квантования $P(S_i) \approx \omega(S_i)\Delta S$. При этом приближение тем точнее, чем меньше интервал (шаг квантования) ΔS . Энтропия такого «дискретного» сигнала

$$M \left\{ \log \frac{1}{\omega(S_i)\Delta S} \right\} = \sum_i \omega(S_i)\Delta S \log \frac{1}{\omega(S_i)\Delta S}.$$

Перейдем теперь к энтропии непрерывного сигнала, для чего устремим величину ΔS к нулю. Тогда можно записать

$$\begin{aligned} H(S) &= \lim_{\Delta S \rightarrow 0} M \left\{ \log \frac{1}{\omega(S_i)\Delta S} \right\} = \lim_{\Delta S \rightarrow 0} \sum_i \omega(S_i) \log \frac{1}{\omega(S_i)} \Delta S + \\ &+ \lim_{\Delta S \rightarrow 0} \log \frac{1}{\Delta S} \sum_i \omega(S_i)\Delta S = \int_{-\infty}^{\infty} \omega(S) \log \frac{1}{\omega(S)} dS + \\ &+ \lim_{\Delta S \rightarrow 0} \log \frac{1}{\Delta S} \int_{-\infty}^{\infty} \omega(S) dS. \end{aligned}$$

Во втором слагаемом полученного выражения $\int_{-\infty}^{\infty} \omega(S) dS = 1$ по определению. Тогда

$$H(S) = \int_{-\infty}^{\infty} \omega(S) \log \frac{1}{\omega(S)} dS + \lim_{\Delta S \rightarrow 0} \log \frac{1}{\Delta S}.$$

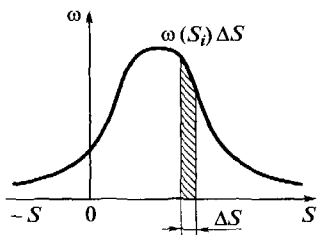


Рис. 10.2. Плотность распределения вероятностей значений сигнала источника непрерывных сообщений

Здесь первое слагаемое является конечной величиной, которая определяется плотностью распределения $\omega(S)$. Эта величина называется дифференциальной энтропией сигнала и обычно рассматривается как некоторая вспомогательная величина при проведении расчетов.

Второе слагаемое данного выражения при $\Delta S \rightarrow 0$ стремится к бесконечности независимо от значений плотности распределения вероятностей сигнала $\omega(S)$. Это означает, что при переходе от дискретной к непрерывной величине энтропия неограниченно возрастает, поскольку при $\Delta S \rightarrow 0$ вероятности попадания сигнала в такие малые пределы шага становятся бесконечно малыми значениями. В результате неожиданность, или непредсказуемость, реализаций становится неограниченно большой.

Для примера найдем дифференциальную энтропию гауссовского шума с нулевым средним значением и дисперсией σ^2 , плотность распределения вероятностей которого описывается соотношением

$$\omega(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

Подставив это выражение в формулу для дифференциальной энтропии, получим

$$h(x) = \int_{-\infty}^{\infty} \omega(x) \log \left(\sqrt{2\pi\sigma^2} e^{\frac{x^2}{2\sigma^2}} \right) dx = \log \sqrt{2\pi\sigma^2} \int_{-\infty}^{\infty} \omega(x) dx + \frac{\log e}{2\sigma^2} \int_{-\infty}^{\infty} \omega(x) x^2 dx.$$

Здесь первый интеграл равен единице, что следует из известного свойства плотности распределения вероятностей.

Второй интеграл по определению равен дисперсии σ^2 . Тогда окончательно получим

$$h(x) = \log \sqrt{2\pi\sigma^2} + \frac{\log e}{2} = \log \sqrt{2\pi e\sigma^2},$$

откуда видно, что дифференциальная энтропия гауссовского шума зависит только от значения дисперсии: с увеличением дисперсии дифференциальная энтропия монотонно возрастает.

В заключение отметим, что из всех видов возможных распределений вероятностей случайных процессов, в которых дисперсия σ^2 является фиксированной величиной, наибольшее значение дифференциальной энтропии имеет гауссовское распределение.

При любой плотности вероятностей $\omega(\xi)$ случайного процесса ξ всегда справедливо неравенство

$$h(\xi) = \int_{-\infty}^{\infty} \omega(\xi) \log \frac{1}{\omega(\xi)} d\xi \leq \log \sqrt{2\pi e \sigma^2}.$$

Знак равенства имеет место только в том случае, когда процесс является гауссовским.

10.5. Количество информации, передаваемое по дискретному каналу

Рассмотрим дискретный канал связи без памяти, на входе и выходе которого имеются ансамбли дискретных сообщений A и B соответственно. Задача заключается в том, чтобы определить количество информации, переданной по этому каналу. Иначе говоря, необходимо найти информацию, содержащуюся в принятом сообщении b_j при передаче сообщения a_i .

Пусть известны $P(a_i, b_j)$ — совместная вероятность реализаций a_i и b_j , а также $P(a_i | b_j)$ — условная вероятность реализации a_i , если на выходе канала появилось сообщение b_j . По теореме умножения вероятностей

$$P(a_i, b_j) = P(a_i)P(b_j | a_i) = P(b_j)P(a_i | b_j).$$

Введем понятие условной энтропии, которую определим как математическое ожидание, аналогичное математическому ожиданию, принятому для энтропии источника сообщений:

$$H(A | B) = \sum_i \sum_j P(a_i, b_j) \log \frac{1}{P(a_i | b_j)}.$$

Рассмотрим свойства условной энтропии.

1. $H(A | B) \geq 0$, т.е. условная энтропия всегда положительная, что следует из ее определения.

2. $H(A | B) \leq H(A)$, т.е. условная энтропия меньше или равна энтропии источника сообщений на входе канала. При этом равенство имеет место в том случае, если $P(a_i | b_j) = P(a_i)$ при всех $a \in A$ и $b \in B$, т.е. когда a и b взаимно независимы. Это означает, что с получением сообщения b_j никакой информации о сообщении a_i не поступает, следовательно, неопределенность не уменьшается. Данная ситуация может соответствовать полной потере информации при передаче из-за помех в канале связи.

Учитывая сказанное, условную энтропию обычно трактуют как количество информации, которое теряется из-за помех и не поступает получателю. Ясно, что полная потеря информации является крайним случаем и в реальных условиях практически не встречается. Величину $H(A|B)$ также называют ненадежностью.

Определим теперь количество информации, передаваемой по каналу связи $I(A, B)$ как разность между количеством информации на его входе, равным энтропии источника $H(A)$, и количеством потерянной информации, равным условной энтропии $H(A|B)$. Величину $I(A, B)$ также называют взаимной информацией и определяют выражением

$$I(A, B) = H(A) - H(A|B),$$

или

$$I(A, B) = M \left\{ \log \frac{1}{P(a_i)} \right\} - M \left\{ \log \frac{1}{P(a_i|b_j)} \right\} = M \left\{ \log \frac{P(a_i|b_j)}{P(a_i)} \right\}.$$

С учетом теоремы умножения вероятностей можно записать

$$I(A, B) = M \left\{ \log \frac{P(b_j)P(a_i|b_j)}{P(b_j)P(a_i)} \right\} = M \left\{ \log \frac{P(a_i, b_j)}{P(b_j)P(a_i)} \right\}.$$

Раскрывая последнюю формулу, получим выражение для взаимной информации в симметричном виде:

$$I(A, B) = \sum \sum P(a_i, b_j) \log \frac{P(a_i, b_j)}{P(b_j)P(a_i)}.$$

Рассмотрим основные свойства взаимной информации.

1. $I(A, B) \geq 0$, что следует из свойств энтропии, причем $I(A, B) = 0$ при обрыве канала, т. е. когда вся информация теряется из-за помех в канале.

2. $I(A, B) \leq H(A)$, причем равенство достигается при отсутствии помех, т. е. при $H(A|B) = 0$.

3. $I(A, B) = I(B, A) = H(B) - H(B|A)$, где энтропия выхода канала $H(B)$ и условная энтропия $H(B|A)$ определяются аналогично найденным ранее значениям энтропии. Данное свойство вытекает из симметрии выражения для взаимной информации.

4. $I(A, B) \leq H(B)$, что следует из предыдущего свойства, причем равенство здесь имеет место, если $H(B|A) = 0$.

5. Приняв в выражении для взаимной информации $A = B$, получим $H(A|A) = 0$, $I(A, A) = H(A)$. Таким образом, энтропию можно трактовать как меру собственной информации об ансамбле A источника сообщений.

Пусть известно время T передачи одного сообщения, тогда по аналогии с производительностью источника можно легко опреде-

лить скорость передачи информации по каналу как количество информации, переданной в единицу времени:

$$I'(A, B) = \frac{1}{T} I(A, B) = v_k I(A, B),$$

где $v_k = 1/T_n$ — скорость, т.е. число символов, поступающих в канал в течение секунды, бит/с.

10.6. Пропускная способность дискретного канала передачи информации

Ранее было показано, что количество информации, переданной по дискретному каналу, зависит от двух факторов:

- свойств источника сообщений, или от его энтропии $H(A)$;
- свойств самих каналов, определяемых их ненадежностью $H(A|B)$.

Следовательно, взаимная информация $I(A, B)$ не может в полной мере охарактеризовать канал как средство передачи сообщений. Более удобной характеристикой для этого является пропускная способность канала. Рассмотрим ее определение.

Пусть на вход дискретного канала поступают сообщения от всевозможных источников, которые описываются различными распределениями вероятностей $P(A)$. В каждом таком случае количество информации, передаваемой по каналу, будет иметь свое значение.

Пропускную способность (информационную емкость) канала определим как максимальное количество передаваемой информации, которая рассчитывается по всевозможным распределениям вероятностей на входе канала:

$$C = \max_{P(A)} I(A, B).$$

Пропускную способность можно также определить в расчете на единицу времени, бит/с:

$$C' = v_k C = \max_{P(A)} I'(A, B),$$

где v_k — скорость передачи символов по каналу.

Рассмотрим основные свойства пропускной способности.

1. $C' \geq 0$, $C' = 0$ при обрыве канала, что следует из первого свойства количества информации.

2. $C' \leq V_k \log N$, где N — объем алфавита источника сообщений. Причем $C' = V_k \log N$ при отсутствии помех в канале связи, что следует из свойств энтропии и взаимной информации.

В соответствии с определением пропускная способность характеризует предельное значение количества информации, больше которого нельзя передать по данному каналу связи. Таким образом она описывает потенциальные характеристики канала по передаче информации.

В качестве примера определим пропускную способность двоичного симметричного канала без памяти, который задают переходными вероятностями:

$$P(b_j | a_i) = \begin{cases} q = 1 - P, & j = i; \\ P, & j \neq i. \end{cases}$$

Используем для расчета пропускной способности выражение, следующее из свойств взаимной информации:

$$C = \max_{P(A)} I(B, A) = \max_{P(A)} [H(B) - H(B|A)].$$

Поскольку рассматривается канал без памяти, условную информацию определить несложно:

$$H(B|A) = M \left\{ \log \frac{1}{P(b_j | a_j)} \right\} = (1 - P) \log \frac{1}{1 - P} + P \log \frac{1}{P}.$$

Тогда

$$C = \max_{P(A)} \left[H(B) - (1 - P) \log \frac{1}{1 - P} - P \log \frac{1}{P} \right].$$

В этом выражении значение $H(B)$ зависит от распределения вероятностей, поэтому максимум количества переданной информации будет достигаться при максимальном значении этой величины. При $N = 2$ максимальное значение $H(B) = \log 2 = 1$ бит, реализуется при равновероятных и взаимонезависимых символах b_j на выходе канала, что обеспечивается только при равновероятных и взаимонезависимых символах a_i на входе канала, т.е. $P(a_1) = P(a_2) = 0,5$. Согласно формуле полной вероятности

$$\begin{aligned} P(b_j) &= \sum_{i=1}^2 P(a_i) P(b_j | a_i) = 0,5 \sum_{i=1}^2 P(b_j | a_i) = \\ &= 0,5 [(1 - P) + P] = 0,5. \end{aligned}$$

В этом случае $\max_{P(B)} \{H(B)\} = \log 2 = 1$, а следовательно, пропускная способность в расчете на один символ, бит,

$$C = 1 + P \log P + (1 - P) \log (1 - P)$$

или в расчете на единицу времени (секунду), бит/с,

$$C' = v_k [1 + P \log P + (1 - P) \log (1 - P)].$$

Анализ полученных выражений показывает, что $C = 0$ при $P = 0,5$, поскольку в этом случае выходные символы можно получить наугад, например посредством подбрасывания монеты. Ситуация, когда $P = 0,5$, получила название обрыва канала.

Если $P = 1$ или $P = 0$ (канал без помех), то пропускная способность $C = 1$. Однако это не противоречит здравому смыслу, так как при $P = 1$, чтобы правильно принимать символы, достаточно инвертировать (т.е. заменить 0 на 1, а 1 на 0) полученную последовательность.

10.7. Пропускная способность непрерывного канала передачи информации

Рассмотрим модель канала передачи с постоянными параметрами и аддитивным белым гауссовским шумом, сигнал на выходе которого описывается выражением

$$z(t) = S(t) + n(t),$$

где $S(t)$ — сигнал на выходе канала; $n(t)$ — аддитивный шум, приводящий к искажениям передаваемых сигналов.

Для определения количества информации, переданной по непрерывному каналу, вновь воспользуемся примером, который рассматривался при нахождении энтропии источника непрерывных сообщений, для чего разделим (проквантуем) области определения сигналов S и z на небольшие отрезки ΔS и Δz . В этом случае распределения вероятностей этих сигналов будут иметь вид

$$P(S_i) = P(S_i \leq S \leq S_i + \Delta S) \approx \omega(S_i)\Delta S;$$

$$P(z_j) = P(z_j \leq z \leq z_j + \Delta z) \approx \omega(z_j)\Delta z.$$

Совместная вероятность появления входных S_i и выходных z_j символов в таком дискретизированном канале

$$P(S_i, z_j) = P(S_i \leq S \leq S_i + \Delta S, z_j \leq z \leq z_j + \Delta z) \approx \omega(S_i, z_j)\Delta S\Delta z.$$

Применив предельный переход аналогичный рассмотренному ранее, определим взаимную информацию между непрерывными величинами S и Z :

$$I(S, z) = \lim_{\substack{\Delta S \rightarrow 0 \\ \Delta z \rightarrow 0}} M \left\{ \log \frac{\omega(S_i, z_j)\Delta S\Delta z}{\omega(S_i)\Delta S\omega(z_j)\Delta z} \right\} = M \left\{ \log \frac{\omega(S, z)}{\omega(S)\omega(z)} \right\}.$$

Учитывая, что $\omega(S_i, z_j) = \omega(Z)\omega(S|z)$, произведем несложные преобразования:

$$I(S, z) = \lim_{\substack{\Delta S \rightarrow 0 \\ \Delta z \rightarrow 0}} M \left\{ \log \frac{\omega(z)\omega(S|z)}{\omega(S)\omega(z)} \right\} = M \left\{ \log \frac{1}{\omega(S)} \right\} -$$

$$\begin{aligned}
 -\log \frac{1}{\omega(S|z)} \Big\} &= M \left\{ \log \frac{1}{\omega(S)} \right\} - M \left\{ \log \frac{1}{\omega(S|z)} \right\} = \\
 &= \int_{-\infty}^{\infty} \omega(S) \log \frac{1}{\omega(S)} dS - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \omega(S, z) \log \frac{1}{\omega(S|Z)} dS dz.
 \end{aligned}$$

Первое слагаемое в полученном выражении уже рассматривалось ранее и было названо дифференциальной энтропией $h(S)$. Второе слагаемое обозначим $h(S|z)$ и назовем условной дифференциальной энтропией. В результате можно записать

$$I(S, z) = h(S) - h(S|z).$$

Для взаимной информации в непрерывном канале справедливы следующие свойства.

1. $I(S, z) \geq 0$, причем $I(S, z) = 0$ в случае «обрыва» канала, когда его вход и выход независимы, т. е. когда $\omega(S|z) = \omega(S)\omega(z)$.
2. $I(S, z) = I(z, S)$, что следует из симметрии выражения для взаимной информации.
3. $I(S, z) = \infty$ при отсутствии помех в канале связи, т. е. когда $z = S$.

Учитывая второе свойство взаимной информации, можно записать

$$I(S, z) = h(z) - h(z|S).$$

Для рассматриваемого канала с аддитивным шумом данное выражение принимает вид

$$I(S, z) = h(z) - h(n),$$

где $h(n)$ — дифференциальная энтропия аддитивной помехи.

Это следует из того, что при известном передаваемом сигнале $S(t)$ неопределенность принимаемого колебания определяется только случайной помехой $n(t)$.

Определим теперь пропускную способность непрерывного канала с полосой пропускания F . Это удобно сделать с использованием теоремы отсчетов, согласно которой сигналы $S(t)$ и $z(t)$ на входе и выходе канала определяются своими отсчетами, взятыми через интервалы Найквиста $1/(2F)$. Сначала найдем значение пропускной способности, приходящееся на один отсчет. Затем для определения информации, переданной по каналу за время T , просуммируем значения количеств информации, приходящихся на каждый такой отсчет.

Пропускной способностью C на один отсчет назовем максимальное количество информации $I(S, z)$, определяемое по всем возможным распределениям входных сигналов:

$$C_{\text{отсч}} = \max_{\omega(S)} I(S, z) = \max_{\omega(S)} [h(z) - h(z|S)].$$

Найдем пропускную способность при условии, что средняя мощность (дисперсия) сигнала равна P_c , а мощность (дисперсия) помехи в полосе частот F равна мощности шума $P_{\text{ш}}$. Используя найденное ранее значение дифференциальной энтропии гауссовской случайной величины $h(n) = \log \sqrt{2\pi e P_{\text{ш}}}$, запишем

$$C_{\text{отсч}} = \max_{\omega(S)} [h(z) - \log \sqrt{2\pi e P_{\text{ш}}}] .$$

Дисперсия суммарного колебания $z(t)$ на выходе канала $D\{z\} = P_c + P_{\text{ш}}$, поскольку сигнал и помеха — взаимно независимые случайные процессы. При заданном значении дифференциальной энтропии помехи $h(n)$ пропускная способность будет максимизирована при гауссовском распределении $z(t)$.

Данное условие легко выполняется, если не только помеха, но и сигнал $S(t)$ имеет гауссовское распределение. (Как известно, сумма двух гауссовских величин также имеет гауссовское распределение.)

Таким образом,

$$\max_{\omega(S)} h(z) = \log \sqrt{2\pi e (P_c + P_{\text{ш}})},$$

тогда

$$\begin{aligned} C_{\text{отсч}} &= \log \sqrt{2\pi e (P_c + P_{\text{ш}})} - \log \sqrt{2\pi e P_{\text{ш}}} = \\ &= \frac{1}{2} \log \left(1 + \frac{P_c}{P_{\text{ш}}} \right). \end{aligned}$$

Для определения окончательного выражения пропускной способности в расчете на единицу времени просуммируем в соответствии с теоремой отсчетов всю информацию, содержащуюся в $2F$ дискретных выборках:

$$C = 2FC_{\text{отсч}} = F \log \left(1 + \frac{P_c}{P_{\text{ш}}} \right).$$

Полученное выражение известно как формула Шеннона для пропускной способности непрерывного гауссовского канала с ограниченной полосой частот и ограниченной средней мощностью сигнала. Эта формула имеет большое значение в теории информации, так как указывает на возможность обмена полосы пропускания на мощность сигнала, и наоборот.

Рассмотрим подробнее, как влияет на пропускную способность изменение полосы пропускания, для чего запишем выражение

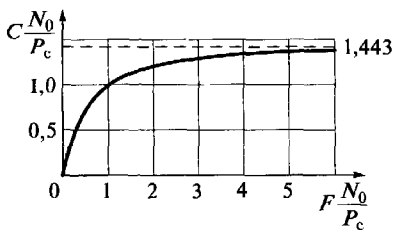


Рис. 10.3. Зависимость нормированной пропускной способности канала с постоянными параметрами и гауссовским белым шумом от полосы пропускания

мощности шума в виде известной зависимости его от односторонней спектральной плотности N_0 :

$$P_{\text{ш}} = FN_0.$$

Подставив это выражение в формулу Шеннона, получим

$$C = F \log \left(1 + \frac{P_c}{FN_0} \right) = F \log e \ln \left(1 + \frac{P_c}{FN_0} \right).$$

Анализ данного выражения показывает, что с ростом полосы частот F пропускная способность канала сначала резко возрастает, а затем ее рост замедляется по мере приближения к пределу:

$$C_{\infty} = \lim_{F \rightarrow \infty} C = \frac{P_c}{N_0} \log e \approx 1,44 \frac{P_c}{N_0} \text{ бит/с.}$$

Данное выражение следует из того, что при $|\delta| \ll 1$ значение $\ln(1 + \delta) \approx \delta$. Зависимость нормированной пропускной способности $C \frac{N_0}{P_c}$ от $F \frac{N_0}{P_c}$, представленная в виде графика на рис. 10.3,

свидетельствует о том, что пропускная способность не может увеличиваться безгранично, и всегда стремится к постоянному значению.

В заключение отметим, что в целях увеличения пропускной способности гауссовского канала нет смысла расширять полосу пропускания до бесконечности. Ее целесообразно увеличивать лишь до значения $F \approx P_c / N_0$. При этом предельное значение скорости передачи информации по каналу, определяемое отношением сигнал/шум, не зависит от полосы пропускания.

Чтобы найти это отношение сигнал/шум, будем считать, что на передачу некоторого сообщения потрачено время T . В этом случае среднее количество переданной информации

$$TI'(S, Z) < TC_{\infty} = \frac{P_c T}{N_0} \log e.$$

Следовательно, для передачи 1 бит информации необходима энергия сигнала

$$P_c T > N_0 / \log e = N_0 \ln 2 \approx 0,693 N_0, \text{ или } h^2 > 0,693.$$

10.8. Теоремы кодирования Шеннона для каналов с помехами

Пропускная способность канала, как отмечалось ранее, характеризует потенциальные свойства передачи информации с общих позиций. Более конкретно эти возможности раскрываются в фундаментальных теоремах Шеннона.

Сначала сформулируем известную теорему кодирования для каналов с помехами применительно к дискретному источнику: *если производительность источника сообщений $H'(A)$ меньше пропускной способности C' дискретного канала с помехами, т.е. $H'(A) < C'$, то при любом $\delta > 0$ существует метод кодирования (декодирования), при котором сообщения передаются получателю с вероятностью ошибки, меньшей δ (где δ — сколь угодно малая величина). Если же $H'(A) > C'$, то таких методов кодирования не существует.*

Заметим, что под кодированием здесь понимается преобразование сообщения в сигнал, а под декодированием — обратное преобразование сигнала в сообщение на выходе канала.

Из данной теоремы следует, что пропускная способность является предельным значением безошибочной скорости передачи информации в условиях помех. Однако теорема не указывает на конкретные методы построения кодирующих и декодирующих устройств. Тем не менее эта теорема чрезвычайно важна, так как она в корне изменила существовавшее до этого мнение о возможностях техники передачи информации.

Ранее считалось, что для безошибочной передачи сообщений необходимо обязательно замедлять скорость, т.е. что только $v \rightarrow 0$ можно обеспечить вероятность ошибки $P \rightarrow 0$. Таким образом, например, можно повысить верность передачи информации в каналах без памяти, если применить метод накопления. Этот метод, являющийся простейшим методом кодирования, состоит в том, что от источника символы «0» и «1» передаются по каналу с вероятностью ошибок $P < 0,5$ с помощью двух кодовых комбинаций, включающих в себя n нулей или n единиц:

$$a_1 = \underbrace{000\dots 0}_n; \quad a_2 = \underbrace{111\dots 1}_n.$$

На приеме регистрация переданного символа производится по большинству одноименных посылок в комбинации (мажоритарное кодирование). Ошибка возникает при условии, что неверно принятыми оказываются $n/2$ или более символов. Ясно, что в случае удлинения комбинации $n \rightarrow \infty$ можно обеспечить сколь угодно малую вероятность ошибки. При этом скорость передачи $v = 1/n$ будет иметь бесконечно малое значение.

Из теоремы Шеннона следует, что существуют коды, которые в отличие от метода повторения могут обеспечить безошибочную передачу информации при конечной скорости, т. е. без замедления. Правда, теорема не указывает конкретных методов построения таких кодов, а говорит лишь об их существовании. В этом смысле теорема неконструктивна, поэтому вопросы реализации конкретных кодов и методов их кодирования и декодирования являются предметом отдельного рассмотрения.

Доказательство сформулированной теоремы основано на методе так называемого случайного кодирования и является достаточно сложным. С учетом ограниченности объема данного пособия оно здесь не приводится. Отметим только, что в процессе доказательства К. Шеннон получил формулу для средней вероятности ошибочного декодирования:

$$P_{\text{ош}} \leq 2^{-T[C' - H'(A)]},$$

где T — длительность передаваемого сигнала (кодовой комбинации).

Данное соотношение показывает, что с увеличением T вероятность ошибки уменьшается и в пределе стремится к нулю. (Заметим, что по условию теоремы разность $C' - H'(A) > 0$.) Следовательно, верность связи тем выше, чем длиннее кодовая комбинация. Однако при этом затрачивается и большее время на передачу одного сообщения из-за увеличения задержек, связанных с обработкой кодовых комбинаций. Поскольку задержки времени не всегда желательны, то в этих случаях достоверность связи можно увеличить посредством менее полного использования пропускной способности канала, т. е. за счет увеличения множителя в показателе степени.

Теорема Шеннона также является справедливой при передаче дискретных сообщений по непрерывному каналу с помехами. При этом под кодированием принято понимать выбор некоторого числа реализаций сигнала $S(t)$ в интервале T и сопоставление каждой из них с символами последовательности сообщений, выдаваемыми источником за то же время T . Под декодированием понимают обратное сопоставление принятых сигналов с символами источника сообщений.

В этом случае теорема Шеннона формулируется в следующем виде: *если сообщения от любого дискретного источника с производительностью $H'(A)$ закодировать непрерывными сигналами $S(t)$, то их можно передавать по непрерывному каналу с пропускной способностью C' при вероятности ошибки меньше сколь угодно малого положительного числа лишь при выполнении условия $C' > H'(A)$. В противном случае, если $C' < H'(A)$, это невозможно.*

Заметим, что и в этом случае вероятность ошибочного декодирования сигнала с длительностью T можно оценить с помощью соотношения, аналогичного приведенному ранее.

В заключение отметим, что в отличие от непрерывного канала операции кодирования и декодирования в дискретном канале осуществляются в два этапа: сначала сообщение преобразуется в последовательность кодовых символов, а затем каждый символ кода преобразуется в соответствующий элемент сигнала. На приеме декодирование производится также в два этапа.

В непрерывном канале кодирование обеспечивает получение принципиально лучших результатов, поскольку в этом случае отсутствуют дополнительные ступени преобразования, которые связаны, как правило, с потерей информации. Однако такое кодирование является сложным в реализации. Значительно чаще и в ущерб некоторым потерям информации применяется кодирование в дискретном канале, реализация которого существенно проще.

10.9. Классификация корректирующих кодов

Целью помехоустойчивого кодирования является повышение верности передаваемых сообщений посредством целенаправленного введения избыточности в процессе преобразования информации. В результате избыточность кодовых комбинаций (кодовых слов) становится больше избыточности источника. Однако именно она способствует обнаружению и исправлению ошибок при передаче сообщений по каналам связи с помехами.

Все известные в настоящее время помехоустойчивые (корректирующие) коды классифицируются по различным признакам.

Одним из таких признаков является основание кода m (иногда применяется обозначение q) — объем кодового алфавита, т. е. число различных символов в алфавите, которое используется для построения кода. Наиболее простыми и широко распространенными являются двоичные коды, у которых $m = 2$.

Также коды подразделяются на блочные и непрерывные. Блочными называются коды, у которых каждый очередной знак сообщения от источника преобразуется в некоторый конечный блок кодовых символов (кодовое слово, кодовую комбинацию). В непрерывных кодах разделения на отдельные блоки или слова не производится. Кодовые символы здесь определяются всей последовательностью знаков сообщения.

Для блочных кодов особое значение имеет понятие длины кодового слова. Для двоичных кодов длина определяется числом символов «0» и «1», входящих в кодовый блок. Если длина всех

комбинаций одинаковая, то, как уже отмечалось, код называется равномерным, в противном случае код является неравномерным. Простейшими примерами таких кодов, являются равномерный телеграфный код МТК-2 и неравномерные коды Морзе, Хаффмана, Шеннона—Фано и др.

Пусть длина кодового слова равномерного кода равна n , тогда число различных блоков M удовлетворяет неравенству $M \leq m^n$.

Если $M = m^n$, то код называется простым, или примитивным. Такой код не содержит избыточности и не является помехоустойчивым. Любая ошибка в кодовом слове преобразует его в другое кодовое слово, что не обнаруживается. В данном коде все возможные комбинации используются для передачи сообщений.

При $M < m^n$ код называется избыточным, или помехоустойчивым. Это означает, что не все возможные кодовые слова применяются для передачи информации: одна их часть относится к разрешенным комбинациям, а другая — к запрещенным. Именно на этом свойстве основан принцип помехоустойчивого кодирования. При этом на приеме в декодирующем устройстве всегда должно быть известно, какие из комбинаций являются разрешенными, а какие относятся к запрещенным. Если при передаче в результате воздействия помех комбинация преобразуется в запрещенную, это обнаруживается. При определенных условиях ошибки могут быть исправлены. Разумеется, что ошибки не обнаруживаются, если одно разрешенное кодовое слово преобразуется в другое разрешенное кодовое слово.

Приведенная классификация кодов является достаточно общей и всегда может быть дополнена и конкретизирована.

В дальнейшем будут рассматриваться, главным образом, двоичные коды с объемом кодового алфавита $m = 2$.

10.10. Основные характеристики блочных корректирующих кодов

Блочные корректирующие коды принято обозначать как коды (n, k) , где n — общее число символов в кодовом блоке (длина кодового слова); k — число информационных символов в блоке. Например, широко известный код Хэмминга с длиной блока, равной семи, и четыремя информационными символами обозначается как код $(7, 4)$.

Любой блочный корректирующий код в каждой своей комбинации имеет ровно $r = n - k$ проверочных (избыточных) символов. Так как в кодовом слове k символов являются информационными, определяющими остальные r проверочных символов, то из общего числа m^n кодовых слов на долю разрешенных прихо-

дится $M = m^k$ комбинаций, т.е. при $m = 2$ число разрешенных комбинаций $M = 2^k$. Данная величина называется объемом кода.

Скоростью кода будем называть величину $R = \frac{\log_2 M}{n \log_2 m}$. При $m = 2$ скорость кода $R = \frac{k}{n}$, бит/символ.

Когда все кодовые слова передаются равновероятно и независимо друг от друга, $\log_2 M$ представляет собой энтропию (собственную информацию), которая приходится на каждое кодовое слово. В этом случае R — это собственная информация в расчете на один кодовый символ.

Важным параметром является вес кодового слова $w(V_i)$ — число символов в слове V_i (где $i = \overline{1, M}$), отличающихся от нуля.

Расстоянием Хэмминга $\rho(V_i, V_j)$ между двумя кодовыми словами принято называть числом позиций (разрядов, элементов), в которых они отличаются друг от друга. Очевидно, что для любых $V_i \neq V_j$ можно записать

$$1 \leq \rho(V_i, V_j) \leq n.$$

Рассматривая в корректирующих кодах множества разрешенных кодовых комбинаций, нетрудно увидеть, что в общем случае расстояние Хэмминга между произвольно выбранными парами кодовых слов является различным. Поэтому минимальным кодовым расстоянием d для заданного кода принято называть минимальное расстояние по Хэммингу между всеми парами его несопадающих кодовых слов:

$$d = \min_{V_i \neq V_j} \rho(V_i, V_j).$$

Очевидно, что для примитивного кода $d = 1$. Вместе с тем существуют корректирующие коды, в которых расстояние Хэмминга между любыми парами разрешенных кодовых слов одинаковое. Такие коды называются эквидистантными.

По аналогии с рассмотренным ранее помехоустойчивым приемом сигналов в теории кодирования применяется метод декодирования по максимуму правдоподобия. Пусть V_i — переданное кодовое слово, а Z — принятый блок. Тогда правило декодирования можно записать в виде

$$P(Z | V_i) > P(Z | V_j); \forall j = \overline{1, M}; j \neq i,$$

или

$$\max_i P(Z | V_i).$$

При равновероятной передаче всех кодовых слов данная процедура декодирования обеспечивает максимально возможную среднюю вероятность правильного их приема.

В симметричном канале без памяти декодирование по максимуму правдоподобия соответствует декодированию по минимуму расстояния Хэмминга, которое формально можно представить соотношением

$$V_i = \min_j \rho(Z, V_j).$$

Данное правило означает, что переданным является то кодовое слово, которое отличается от принятого слова в наименьшем числе символов (позиций, разрядов). Заметим, что если в канале имеется память или он обладает несимметрией, то декодирование по минимуму расстояния Хэмминга не будет являться оптимальным.

В теории кодирования весьма важным является понятие кратности ошибок. Обычно говорят, что произошла ошибка с кратностью t , если в кодовой блоке искажено t символов. По существу кратность ошибки — это расстояние Хэмминга между переданным и принятым кодовыми словами.

С минимальным кодовым расстоянием связаны понятия обнаруживающей и исправляющей способностей кода. Так, если корректирующий код с $d > 1$ используется в режиме обнаружения ошибок, то все ошибки с кратностью $t \leq d - 1$ будут обнаружены гарантированно. Действительно, при кратности ошибок t , не превышающей значение d в принятом кодовом слове, это кодовое слово нельзя отнести к разрешенным, поскольку оно отличается от любого другого числом элементов, меньшим d . Это противоречит определению d . Следовательно, такое кодовое слово будет запрещенным, т. е. ошибки обнаруживаются. Если же $t > d$, то кодовый блок может оказаться разрешенным с необнаруженными ошибками. Разумеется, в такой ситуации некоторые ошибки могут быть все-таки обнаружены, но без полной гарантии. Очевидно, что минимальное кодовое расстояние, при котором могут обнаруживаться любые одиночные ошибки, $d = 2$.

В теории кодирования также доказано, что если корректирующий код используется в режиме исправления ошибок, то он способен исправлять с гарантией любые ошибки с кратностью до $t \leq \frac{d}{2} - 1$ (при четном d) или $t \leq (d - 1)/2$ (при нечетном d). Только в этом случае любая искаженная помехами комбинация, попавшая в разряд запрещенных, будет декодироваться в разрешенное кодовое слово, которое по расстоянию Хэмминга наиболее близко к передававшейся комбинации.

Наконец, при использовании корректирующего кода для одно-временного обнаружения и исправления ошибок и стираний его кодовое расстояние должно удовлетворять условию $d > 2t_o + t_c$, где t_o — число гарантированно исправляемых ошибок; t_c — число исправляемых стираний.

10.11. Линейные двоичные коды для корректирования ошибок

Среди известных в настоящее время кодов наиболее изученными являются линейные блочные корректирующие коды. Построение этих кодов базируется на тех разделах высшей алгебры, предметом изучения которых являются множества дискретных элементов и операции с ними. Кроме того, эти коды не сложно технически реализовать на основе цифровых интегральных схем. Именно поэтому данные коды широко распространены.

Линейным двоичным блочным кодом называется множество двоичных последовательностей длиной n , которое содержит чисто нулевую последовательность (000...0) и обладает следующим свойством: сумма по модулю два любых пар последовательностей из этого множества также является элементом этого множества.

Данные коды в некоторых случаях называются групповыми, поскольку они представляют собой подгруппу для группы всех двоичных последовательностей длиной n .

В классе линейных кодов наибольший интерес представляют так называемые систематические (n, k) -коды. В этих кодах первые k символов кодового слова являются информационными. Последующие $r = n - k$ элементов являются проверочными (избыточными), формирующимися на основе линейных операций (сложение по модулю два) с информационными символами.

Поясним подробнее принцип построения линейных кодов, используя понятие линейной независимости. Известно, что совокупность кодовых слов V_1, V_2, \dots, V_k называется линейно независимой, если справедливо соотношение

$$\alpha_1 V_1 + \alpha_2 V_2 + \dots + \alpha_k V_k \neq 0,$$

при всех значениях α_i за исключением

$$\alpha_1 = \alpha_2 = \dots = \alpha_k = 0, \text{ где } \alpha_i \in \{0, 1\}.$$

В теории кодирования доказывается, что из общего числа $M = 2^k$ разрешенных кодовых слов линейного кода можно выбрать совокупность, содержащую k произвольных ненулевых слов, которые отвечают свойству линейной независимости. Эта совокупность, как уже отмечалось, называется линейным базисом. Далее

все элементы базиса, т.е. выбранные кодовые слова, могут быть сложены поразрядно по модулю два друг с другом в различных сочетаниях. Общее число таких комбинаций равно 2^k , что как раз и соответствует числу разрешенных кодовых блоков. Следовательно, линейный блочный код можно определить с помощью элементов базиса, содержащих k линейно независимых кодовых комбинаций.

Такие комбинации принято записывать в виде прямоугольной таблицы, называемой порождающей матрицей $G_{n,k}$. Данная матрица, содержащая k строк и n столбцов, записывается обычно в так называемой канонической форме

$$G_{n,k} = [I_k B_{k \times (n-k)}]$$

или в развернутом виде:

$$G_{n,k} = \begin{bmatrix} 100 \dots 0 & b_{1,k+1} \dots b_{1,n} \\ 010 \dots 0 & b_{2,k+1} \dots b_{2,n} \\ \vdots & \vdots \\ \underbrace{000 \dots 1}_k & \underbrace{b_{k,k+1} \dots b_{k,n}}_{r=n-k} \end{bmatrix}$$

Здесь I_k — единичная матрица размерном $k \times k$ с единицами на главной диагонали и нулями в других местах. Строки этой матрицы представляют собой информационные символы, вырабатываемые источником сообщений. Матрица $B_{k \times (n-k)}$ в своих строках содержит проверочные символы корректирующего кода.

Порождающая матрица кода, содержащая k линейно независимых кодовых слов, необязательно записывается в канонической форме. Однако посредством элементарных преобразований ее строк, т.е. посредством перестановок или сложения строк по модулю два друг с другом, порождающую матрицу всегда можно представить в канонической форме.

Отметим, что если аналогичные процедуры произвести со столбцами, то будет получен новый корректирующий код, характеристики которого отличаются от характеристик исходного кода. В случае когда столбцы только переставляются, вес кодовых слов не изменяется. В результате получают новый линейный код, эквивалентный исходному.

Указанные операции со строками порождающей матрицы (кодowymi словами) могут привести к заданию нулевой комбинации, которая по определению будет принадлежать к данному коду. Если теперь выбрать любую пару ненулевых кодовых слов V_i и V_j , то расстояние Хэмминга между ними $\rho(V_i, V_j)$ будет равно весу $w(V_k)$ некоторой третьей комбинации V_k , которая, в свою очередь, так-

же принадлежит этому коду. Путем последовательного перебора несложно найти некоторое кодовое слово, которое будет иметь минимальный вес и, следовательно, минимальное расстояние Хэмминга относительно нулевого кодового слова. Из этого следует важное свойство линейного кода: минимальное кодовое расстояние линейного корректирующего кода равно минимальному весу его ненулевого кодового слова, т. е.

$$d = \min_{V_i \in V} w(V_i); \quad \forall V_i \neq 0.$$

Таким образом, если для линейного кода необходимо определить значение минимального кодового расстояния, с которым связаны его корректирующие свойства, то сделать это значительно легче на основе изучения списка весов, а не хэмминговских расстояний.

Отметим также, что использование линейных корректирующих кодов, которые задаются своими порождающими матрицами, решает проблему сложности процессов кодирования.

Действительно, вместо хранения в памяти кодирующего устройства всего объема кода, т. е. $M = 2^k$ кодовых слов длиной n или $n2^k$ бит информации, в общем случае достаточно запомнить только порождающую матрицу, содержащую $nk = n \log_2 M$ бит информации.

В качестве примера рассмотрим метод задания линейного корректирующего кода (7, 4), который имеет $M = 2^4 = 16$ разрешенных кодовых слов:

1. 0001110	5. 1001001	9. 1010010	13. 1011100
2. 0010101	6. 1100100	10. 0101101	14. 1101010
3. 0100011	7. 0011011	11. 1110001	15. 1111111
4. 1000111	8. 0110110	12. 0111000	16. 0000000

Здесь минимальный вес ненулевых кодовых слов $w(V_i) = 3$, поэтому $d = 3$. Следовательно, данный код способен с гарантией исправить одну ошибку.

Для задания рассматриваемого кода с помощью порождающей матрицы, имеющей каноническую форму, выберем первые четыре кодовых слова.

Тогда

$$G_{7,4} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Кроме рассмотренного метода задания линейных кодов в теории кодирования широко применяется и другой метод, основанный на использовании так называемых проверочных матриц:

$$H_{n,k} = [A_{(n-k) \times k} I_{n-k}],$$

где $A_{(n-k) \times k} = B_{k \times (n-k)}^T$ — матрица, содержащая $n - k$ строк и k столбцов; T — знак транспонирования матрицы B (т. е. взаимной замены строк и столбцов); I_{n-k} — единичная матрица с $n - k$ строками и таким же числом столбцов.

В развернутом виде проверочная матрица имеет вид

$$H_{n,k} = \begin{bmatrix} a_{1,1} \dots a_{1,k} & 100 \dots 0 \\ a_{2,1} \dots a_{2,k} & 010 \dots 0 \\ \vdots & \vdots \\ a_{(n-k),1} \dots a_{(n-k),k} & 000 \dots 1 \end{bmatrix}.$$

Проверочную матрицу кода можно построить следующим образом. Сначала записывается единичная матрица I_{n-k} . Затем слева к ней приписывается матрица $A_{(n-k) \times k}$, каждая строка которой содержит символы, взятые из столбцов матрицы $B_{n \times (n-k)}$. Данные символы являются проверочными (избыточными) в кодовых словах, поэтому матрица получила название проверочной. Для рассматриваемого кода она будет иметь вид

$$H_{7,4} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Сущность проверочной матрицы заключается в следующем. Символы, отличные от нуля, расположенные на позициях, которые соответствуют информационным элементам кодового слова, указывают, с помощью каких информационных элементов формируются проверочные символы. Элементы, не равные нулю, стоящие на позициях проверочных элементов, показывают, какой именно проверочный символ образован с помощью этих информационных символов.

Учитывая сказанное и с использованием приведенной ранее проверочной матрицы линейного кода (7, 4), можно записать правила формирования проверочных символов в любом из слов данного кода в виде $V_i = (v_1, v_2, \dots, v_7)$, где $i = \overline{1, 16}$:

$$v_5 = v_1 + v_3 + v_4;$$

$$v_6 = v_1 + v_2 + v_4;$$

$$v_7 = v_1 + v_2 + v_3.$$

Заметим, что сложение здесь производится по модулю два. Используем эти соотношения для кодирования последовательности, например 0101. Информационные элементы кодового слова в этом случае следующие: $v_1 = 0, v_2 = 1, v_3 = 0, v_4 = 1$, а проверочные символы имеют значения $v_5 = 0 + 0 + 1 = 1, v_6 = 0 + 1 + 1 = 0, v_7 = 0 + 1 + 0 = 1$. Таким образом получим кодовое слово 0101101, которое совпадает с кодовым блоком под номером десять.

Аналогично кодируются и остальные кодовые слова.

Если вновь рассмотреть матрицы G и H , то можно увидеть, что каждая из них содержит множество линейно независимых комбинаций или векторов. В связи с этим каждую из этих матриц считают базисом некоторого линейного пространства, состоящего из векторов длиной n . Нетрудно заметить, что эти матрицы ортогональны друг другу, т. е. скалярное произведение каждой строки матрицы G и каждой строки матрицы H равно нулю:

$$HG^T = 0.$$

Таким образом, эти матрицы можно поменять местами, т. е. матрицу H использовать в качестве порождающей, а G — как проверочную матрицу некоторого кода. Полученный корректирующий код называется дуальным (двойственным) исходному коду. Векторное пространство, соответствующее матрице H , принято называть нулевым пространством по отношению к векторному пространству, соответствующему матрице G .

10.12. Декодирование линейного кода

Введенная ранее проверочная матрица линейного кода связана с процедурой декодирования. Это легко заметить, если соотношение для проверочных символов в кодовом слове переписать в общем виде:

$$V_i H^T = 0; \quad i = \overline{1, M},$$

где V_i — произвольное слово кода; H^T — транспонированная матрица проверок.

Смысл этого соотношения заключается в том, что некоторая комбинация V_i принадлежит блочному коду (n, k) только в том случае, если она ортогональна каждой строке проверочной матрицы H . По существу это выражение составляет процедуру декодирования линейного кода.

Всякий принимаемый вектор сигнала V_i перемножается в декодирующем устройстве на хранящуюся в его памяти транспонированную матрицу проверок. В результате этого получают комбинацию, называемую *синдром*. По виду синдрома можно судить

о наличии или отсутствии ошибок: нулевой синдром указывает на отсутствие ошибок, ненулевой — на их наличие.

(Синдром может быть нулевым, если в принятом кодовом слове имеются необнаруженные ошибки, что приводит к преобразованию одного разрешенного кодового слова в другое также разрешенное.)

Проиллюстрируем сказанное на примере линейного кода (7, 4).

Пусть принятый вектор сигнала имеет вид $\tilde{V}_i = 0101101$. Тогда

$$S = \tilde{V}_i H^T = 0101101 \times \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 000,$$

т. е. синдром нулевой, $S = 000$, что свидетельствует об отсутствии ошибок в декодированном кодовом блоке.

Предположим теперь, что на вход декодера поступает вектор $\tilde{V}_i = 0100101$. В этом случае

$$S = \tilde{V}_i H^T = 0100101 \times \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 110,$$

т. е. синдром отличен от нуля, что указывает на ошибки в кодовом блоке.

Для некоторых линейных кодов можно заметить интересную особенность. Пусть в кодовой комбинации проверочные разряды расположены на номерах позиций, которые выражаются степенью числа 2 (т. е. $2^0, 2^1, 2^2, \dots, 2^{r-1}$, где r — число проверочных символов). Тогда вычисленный ненулевой синдром оказывается двоичным представлением номера разряда кодового слова, где произошла ошибка.

Первым эту особенность заметил Р. Хэмминг в 1950 г., поэтому рассмотренный линейный код получил название кода Хэм-

минга. В общем случае коды Хэмминга имеют следующие параметры: длина кодовых слов $n = 2^r - 1$; число проверочных символов составляет r ; число информационных элементов $k = 2^r - 1 - r$. Минимальное кодовое расстояние у кодов Хэмминга $d = 3$, поэтому они способны гарантированно исправлять одиночные ошибки. Помимо представленного здесь кода (7, 4) к числу кодов Хэмминга принадлежат коды (15, 11), (31, 26) и др.

Коды Хэмминга удобнее всего задавать, если проверочную матрицу H построить таким образом, чтобы i -й столбец непосредственно соответствовал двоичному представлению номера i . Тогда ненулевой синдром будет в точности являться двоичной записью номера того разряда, в котором произошла ошибка. Так, для кода (7, 4) можно выбрать следующую матрицу:

$$H_{7,4} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Коды Хэмминга являются лишь частным случаем линейных блочных кодов. В общем случае процедура декодирования оказывается значительно сложнее, поскольку по виду синдрома любого линейного кода, кроме кода Хэмминга, нельзя определить разряд в кодовом слове, где произошла ошибка. Отличие синдрома от нуля — это по существу констатация факта появления ошибки. Если код используется в режиме обнаружения, то это отличие не имеет существенного значения, поскольку такой режим применяется в системах с обратной связью, т.е. ошибочную комбинацию всегда можно исправить, переспросив ее по обратному каналу.

Допустим, на вход декодера поступает вектор \tilde{V}_i , соответствующий некоторой комбинации V_i (где $i = \overline{1, M}$), искаженной помехами, т.е. $\tilde{V}_i = V_i + E$, где E — вектор ошибок. При этом синдром несложно вычислить:

$$S = \tilde{V}_i H^T = (V_i + E) H^T = V_i H^T + E H^T = E H^T.$$

Так как первое слагаемое полученного выражения $V_i H^T = 0$ по определению, то синдром зависит только от вектора или образца ошибок, искаживших переданное кодовое слово.

Таким образом, в общем случае процедура декодирования линейного блочного кода может быть следующей. В запоминающее устройство декодера записывается таблица, состоящая из различных ненулевых синдромов и соответствующих им конфигураций (векторов) ошибок. В процессе декодирования по принимаемой комбинации вычисляется синдром, а по нему находится вектор

ошибок. Далее вектор ошибок складывается с принятой комбинацией. В результате этого происходит исправление ошибок, и получателю выдается кодовое слово, которое не должно содержать искаженных помехами символов. При таком методе декодер должен хранить 2^{n-k} векторов ошибок и столько же синдромов, включая нулевой. В этом случае таблица декодирования будет сравнительно небольшой при использовании относительно коротких кодов.

Вместе с тем из теоремы кодирования Шеннона следует, что для обеспечения высокой достоверности необходимо применять достаточно длинные коды. В результате объем таблицы декодирования будет существенно возрастать.

Например, для кода (63, 45) необходимо хранить в памяти декодера $2^{18} = 262\,144$ векторов ошибок.

Рассмотренный метод декодирования может применяться только для относительно коротких кодов и исправления ошибок небольшой кратности, чаще всего одиночных. В последнее время разработаны достаточно эффективные коды и методы их декодирования с использованием современной элементной базы, однако их изучение выходит за рамки курса теории электрической связи.

Контрольные вопросы

1. Какая единица и почему применяется для измерения количества информации?
2. Что понимается под энтропией источника дискретных сообщений и каковы свойства энтропии?
3. Каким образом определяется и как вычисляется дифференциальная энтропия?
4. Как определяется взаимная информация в дискретном и непрерывном каналах?
5. Поясните понятие пропускной способности дискретного канала.
6. Поясните формулу Шеннона и определите, для каких каналов связи она применима.
7. Сформулируйте и поясните теорему кодирования Шеннона для канала с помехами.
8. Назовите и поясните смысл основных параметров блочных корректирующих кодов.
9. Что представляют собой порождающая и проверочная матрицы и как они записываются в канонической форме?
10. Как в процессе декодирования группового кода возникает синдром?
11. В чем заключается особенность кода Хэмминга?

ПРИЛОЖЕНИЕ

Таблица интеграла вероятностей

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7703	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,90147

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97173	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99007	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995604	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605

3,0	0,9 ² 8650	0,9 ² 8694	0,9 ² 8736	0,9 ² 8777	0,9 ² 8817	0,9 ² 8856	0,9 ² 8893	0,9 ² 8930	0,9 ² 8965	0,9 ² 8999
3,1	0,9 ³ 0324	0,9 ³ 0646	0,9 ³ 0957	0,9 ³ 1260	0,9 ³ 1553	0,9 ³ 1836	0,9 ³ 2112	0,9 ³ 2378	0,9 ³ 2636	0,9 ³ 2886
3,2	0,9 ³ 3129	0,9 ³ 3363	0,9 ³ 3590	0,9 ³ 3810	0,9 ³ 4024	0,9 ³ 4230	0,9 ³ 4429	0,9 ³ 4623	0,9 ³ 4810	0,9 ³ 4991
3,3	0,9 ³ 5166	0,9 ³ 5335	0,9 ³ 5499	0,9 ³ 5659	0,9 ³ 5811	0,9 ³ 5959	0,9 ³ 6103	0,9 ³ 6242	0,9 ³ 6376	0,9 ³ 6505
3,4	0,9 ³ 6631	0,9 ³ 6752	0,9 ³ 6869	0,9 ³ 6982	0,9 ³ 7091	0,9 ³ 7197	0,9 ³ 7299	0,9 ³ 7389	0,9 ³ 7493	0,9 ³ 7785
3,5	0,9 ³ 7674	0,9 ³ 7759	0,9 ³ 7842	0,9 ³ 7922	0,9 ³ 7999	0,9 ³ 8074	0,9 ³ 8146	0,9 ³ 8215	0,9 ³ 8282	0,9 ³ 8347
3,6	0,9 ³ 8409	0,9 ³ 8469	0,9 ³ 8527	0,9 ³ 8583	0,9 ³ 8637	0,9 ³ 8689	0,9 ³ 8739	0,9 ³ 8787	0,9 ³ 8834	0,9 ³ 8879
3,7	0,9 ³ 8922	0,9 ³ 8964	0,9 ⁴ 0039	0,9 ⁴ 0426	0,9 ⁴ 0799	0,9 ⁴ 1158	0,9 ⁴ 1504	0,9 ⁴ 1838	0,9 ⁴ 2159	0,9 ⁴ 2469
3,8	0,9 ⁴ 2765	0,9 ⁴ 3052	0,9 ⁴ 3327	0,9 ⁴ 3593	0,9 ⁴ 3848	0,9 ⁴ 4094	0,9 ⁴ 4331	0,9 ⁴ 4558	0,9 ⁴ 4777	0,9 ⁴ 4999
3,9	0,9 ⁴ 5190	0,9 ⁴ 5385	0,9 ⁴ 5573	0,9 ⁴ 5753	0,9 ⁴ 5926	0,9 ⁴ 6092	0,9 ⁴ 6253	0,9 ⁴ 6406	0,9 ⁴ 6554	0,9 ⁴ 6696
4,0	0,9 ⁴ 6833	0,9 ⁴ 6964	0,9 ⁴ 7090	0,9 ⁴ 7211	0,9 ⁴ 7327	0,9 ⁴ 7439	0,9 ⁴ 7546	0,9 ⁴ 7649	0,9 ⁴ 7748	0,9 ⁴ 7843
4,1	0,9 ⁴ 7936	0,9 ⁴ 8022	0,9 ⁴ 8106	0,9 ⁴ 8186	0,9 ⁴ 8263	0,9 ⁴ 8339	0,9 ⁴ 8409	0,9 ⁴ 8477	0,9 ⁴ 8542	0,9 ⁴ 8605
4,2	0,9 ⁴ 8665	0,9 ⁴ 8723	0,9 ⁴ 8778	0,9 ⁴ 8832	0,9 ⁴ 8882	0,9 ⁴ 8931	0,9 ⁴ 8978	0,9 ⁵ 0226	0,9 ⁵ 0655	0,9 ⁵ 1066
4,3	0,9 ⁵ 1460	0,9 ⁵ 1837	0,9 ⁵ 2199	0,9 ⁵ 2545	0,9 ⁵ 2876	0,9 ⁵ 3193	0,9 ⁵ 3497	0,9 ⁵ 3788	0,9 ⁵ 4066	0,9 ⁵ 4332
4,4	0,9 ⁵ 4587	0,9 ⁵ 4831	0,9 ⁵ 5065	0,9 ⁵ 5288	0,9 ⁵ 5502	0,9 ⁵ 5706	0,9 ⁵ 5902	0,9 ⁵ 6089	0,9 ⁵ 6268	0,9 ⁵ 6439
4,5	0,9 ⁵ 6602	0,9 ⁵ 6759	0,9 ⁵ 6908	0,9 ⁵ 7051	0,9 ⁵ 7187	0,9 ⁵ 7318	0,9 ⁵ 7442	0,9 ⁵ 7561	0,9 ⁵ 7675	0,9 ⁵ 7784
4,6	0,9 ⁵ 7888	0,9 ⁵ 7987	0,9 ⁵ 8081	0,9 ⁵ 8172	0,9 ⁵ 8258	0,9 ⁵ 8340	0,9 ⁵ 8419	0,9 ⁵ 8494	0,9 ⁵ 8566	0,9 ⁵ 8634
4,7	0,9 ⁵ 8699	0,9 ⁵ 8761	0,9 ⁵ 8821	0,9 ⁵ 8877	0,9 ⁵ 8931	0,9 ⁵ 8983	0,9 ⁶ 0320	0,9 ⁶ 0789	0,9 ⁶ 1235	0,9 ⁶ 1661
4,8	0,9 ⁶ 2067	0,9 ⁶ 2453	0,9 ⁶ 2822	0,9 ⁶ 3173	0,9 ⁶ 3508	0,9 ⁶ 3827	0,9 ⁶ 4131	0,9 ⁶ 4420	0,9 ⁶ 4696	0,9 ⁶ 4958

Окончание таблицы

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
4,9	0,9 ⁶ 5208	0,9 ⁶ 5446	0,9 ⁶ 5673	0,9 ⁶ 5889	0,9 ⁶ 6094	0,9 ⁶ 6289	0,9 ⁶ 6475	0,9 ⁶ 6652	0,9 ⁶ 6821	0,9 ⁶ 6981
5,0	0,9 ⁶ 7134	0,9 ⁶ 7278	0,9 ⁶ 7416	0,9 ⁶ 7547	0,9 ⁶ 7672	0,9 ⁶ 7791	0,9 ⁶ 7904	0,9 ⁶ 8011	0,9 ⁶ 8113	0,9 ⁶ 8210
5,1	0,9 ⁶ 8302	0,9 ⁶ 8390	0,9 ⁶ 8473	0,9 ⁶ 8552	0,9 ⁶ 8627	0,9 ⁶ 8698	0,9 ⁶ 8766	0,9 ⁶ 8830	0,9 ⁶ 8891	0,9 ⁶ 8949
5,2	0,9 ⁷ 0036	0,9 ⁷ 0558	0,9 ⁷ 1154	0,9 ⁷ 1525	0,9 ⁷ 1972	0,9 ⁷ 2396	0,9 ⁷ 2797	0,9 ⁷ 3179	0,9 ⁷ 3508	0,9 ⁷ 3884
5,3	0,9 ⁷ 4209	0,9 ⁷ 4519	0,9 ⁷ 4812	0,9 ⁷ 5089	0,9 ⁷ 5353	0,9 ⁷ 5602	0,9 ⁷ 5839	0,9 ⁷ 6063	0,9 ⁷ 6276	0,9 ⁷ 6477
5,4	0,9 ⁷ 6668	0,9 ⁷ 6849	0,9 ⁷ 7020	0,9 ⁷ 7182	0,9 ⁷ 7336	0,9 ⁷ 7482	0,9 ⁷ 7619	0,9 ⁷ 7749	0,9 ⁷ 7873	0,9 ⁷ 7990
5,5	0,9 ⁷ 8101	0,9 ⁷ 8206	0,9 ⁷ 8305	0,9 ⁷ 8399	0,9 ⁷ 8488	0,9 ⁷ 8572	0,9 ⁷ 8651	0,9 ⁷ 8726	0,9 ⁷ 8797	0,9 ⁷ 8865
5,6	0,9 ⁷ 8928	0,9 ⁷ 8988	0,9 ⁸ 0452	0,9 ⁸ 0989	0,9 ⁸ 1497	0,9 ⁸ 1978	0,9 ⁸ 2431	0,9 ⁸ 2860	0,9 ⁸ 3265	0,9 ⁸ 3648
5,7	0,9 ⁸ 4009	0,9 ⁸ 4351	0,9 ⁸ 4674	0,9 ⁸ 4978	0,9 ⁸ 5266	0,9 ⁸ 5538	0,9 ⁸ 5794	0,9 ⁸ 6036	0,9 ⁸ 6265	0,9 ⁸ 6481
5,8	0,9 ⁸ 6684	0,9 ⁸ 6876	0,9 ⁸ 7058	0,9 ⁸ 7229	0,9 ⁸ 7390	0,9 ⁸ 7542	0,9 ⁸ 7686	0,9 ⁸ 7821	0,9 ⁸ 7949	0,9 ⁸ 8069
5,9	0,9 ⁸ 8182	0,9 ⁸ 8289	0,9 ⁸ 8390	0,9 ⁸ 8485	0,9 ⁸ 8575	0,9 ⁸ 8659	0,9 ⁸ 8739	0,9 ⁸ 8814	0,9 ⁸ 8884	0,9 ⁸ 8951
6,0	0,9 ⁹ 0134									

Примечания: 1. Показатель степени у цифры 9 означает число девяток после запятой, например:

$$F(2,43) = 0,9^{22451} = 0,992451.$$

2. При отрицательных значениях x значения интегральных вероятностей определяются с помощью соотношения $F(-x) = 1 - F(x)$.

СПИСОК ЛИТЕРАТУРЫ

1. *Андреев В. С.* Теория нелинейных электрических цепей : учеб. пособие для вузов / В. С. Андреев. — М. : Радио и связь, 1982. — 280 с.
2. *Баскаков С. И.* Радиотехнические цепи и сигналы: учебник для вузов / С. И. Баскаков. — 3-е изд., перераб. и доп. — М. : Высш. шк., 2000. — 462 с.
3. *Биккенин Р. Р.* Теория электрической связи. Случайные процессы. Помехоустойчивая передача дискретной информации : учеб. пособие / Р. Р. Биккенин, М. Н. Чесноков. — СПб. : СПбГУТ, 2001. — 157 с.
4. *Гоноровский И. С.* Радиотехнические цепи и сигналы : учебник для вузов / И. С. Гоноровский. — М. : Радио и связь, 1986. — 512 с.
5. *Григоровский Л. Ф.* Теория электрической связи : в 3 т. Ч. 1. Модели сигналов и методы их преобразования в системах связи: учеб. пособие / Л. Ф. Григоровский, В. И. Коржик, В. Г. Красов, В. Ф. Кушнир. — Л. : ЛЭИС, 1990. — 88 с.
6. *Григоровский Л. Ф.* Теория электрической связи : в 3 т. Ч. 2. Прохождение детерминированных сигналов через каналы связи: учеб. пособие / Л. Ф. Григоровский, В. И. Коржик, В. Г. Красов, В. Ф. Кушнир — Л. : ЛЭИС, 1991. — 96 с.
7. *Григоровский Л. Ф.* Теория электрической связи : в 3 т. Ч. 3. Модели случайных сигналов и методы анализа элементов каналов связи : учеб. пособие / Л. Ф. Григоровский, В. И. Коржик, В. Г. Красов, В. Ф. Кушнир. — СПб. : ЛЭИС, 1992. — 108 с.
8. *Кларк Дж.* мл. Кодирование с исправлением ошибок в системах цифровой связи / Дж. Кларк, Дж. Кейн : пер. с англ. [под ред. Б. С. Цыбакова]. — М. : Радио и связь, 1987. — 392 с.
9. *Клюев Н. И.* Основы теории связи / Н. И. Клюев. — Л. : ВАС, 1985. — 262 с.
10. *Котельников В. А.* Теория потенциальной помехоустойчивости / В. А. Котельников. — М. : Госэнергоиздат, 1956. — 152 с.
11. *Кушнир В. Ф.* Теория нелинейных электрических цепей : учебник для вузов / В. Ф. Кушнир, Б. А. Ферсман. — М. : Связь, 1974. — 384 с.
12. *Петрович Н. Т.* Передача дискретной информации в каналах с фазовой манипуляцией / Н. Т. Петрович. — М. : Сов. радио, 1965. — 263 с.
13. *Прокис Дж.* Цифровая связь / Дж. Прокис : пер. с англ. [под ред. Д. Д. Кловского]. — М. : Радио и связь, 2000. — 800 с.
14. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение / Б. Скляр. 2-е изд. испр. : пер. с англ. [под ред. А. В. Назаренко]. — М. : Изд. дом «Вильямс», 2007. — 1104 с.

15. *Стейн С.* Принципы современной теории связи и их применение к передаче дискретных сообщений / С. Стейн, Дж. Джонс : пер. с англ. [под ред. Л. М. Финка]. — М. : Связь, 1971. — 376 с.

16. Теория передачи информации. Терминология. Вып. 101. — М. : Наука, 1984. — 32 с.

17. Теория электрической связи : учебник для вузов / [А. Г. Зюко, Д. Д. Кловский, В. И. Коржик, М. В. Назаров и др.] / под ред. Д. Д. Кловского. — М. : Радио и связь, 1998. — 432 с.

18. *Тихонов В. И.* Статистическая радиотехника : учеб. пособие для вузов / В. И. Тихонов. — М. : Сов. радио, 1966. — 678 с.

19. *Финк Л. М.* Теория передачи дискретных сообщений / Л. М. Финк. — М. : Сов. радио, 1970. — 728 с.

20. *Фрейман И. Г.* Курс радиотехники / И. Г. Фрейман. — 2-е изд. — Л. : ГИЗ, 1928.

21. *Чердынцев В. А.* Радиотехнические системы / В. А. Чердынцев. — Высшейш. шк., 1988. — 369 с.

22. *Шеннон К. Э.* Математическая теория связи : Работы по теории информации и кибернетике / К. Э. Шеннон : пер. с англ. [под ред. Р. Л. Добрушина, О. В. Лупанова]. — М. : Иностран. лит., 1963. — 830 с.

23. *Shannon C. E.* A mathematical theory of communication. *Bell System Technical Journal*, vol. 27, pp. 379—423 and 623—656, July and October, 1948.

ОГЛАВЛЕНИЕ

Предисловие.....	3
Глава 1. Общие сведения о системах электросвязи.....	6
1.1. Понятия информации, сообщения, сигнала	6
1.2. Модель системы передачи информации	10
1.3. Помехи и искажения в каналах связи	20
1.4. Основные характеристики систем передачи информации.....	22
Глава 2. Модели сигналов.....	25
2.1. Классификация сигналов.....	25
2.2. Сигналы как элементы функциональных пространств	28
2.3. Обобщенный ряд Фурье.....	37
2.4. Спектральное представление колебаний	40
2.4.1. Спектральное представление периодических колебаний ..	40
2.4.2. Спектральное представление непериодических колебаний.....	44
2.4.3. Распределение мощности и энергии в спектре колебания. Ширина спектра колебания.....	48
2.5. Связь между временными и спектральными характеристиками колебаний	50
2.6. Теоремы о спектрах	51
2.7. Понятие функции корреляции	53
2.8. Ряд Котельникова. Теорема отсчетов.....	55
2.9. Дискретное преобразование Фурье	60
2.10. Аналитический сигнал	63
2.10.1. Общие сведения.....	63
2.10.2. Огибающая, мгновенная фаза и мгновенная частота сигнала.....	68
2.10.3. Узкополосные сигналы.....	68
Глава 3. Нелинейные преобразования в каналах электросвязи	72
3.1. Общие сведения о преобразованиях сигналов.....	72
3.2. Методы аппроксимации характеристик нелинейных преобразователей сигналов.....	74
3.2.1. Общие сведения	74
3.2.2. Полиномиальная аппроксимация	74
3.2.3. Кусочно-линейная аппроксимация	76

3.2.4. Аппроксимация трансцендентными функциями	77
3.3. Спектральный анализ колебаний в нелинейных преобразователях	79
3.3.1. Общие сведения	79
3.3.2. Метод кратных дуг	79
3.3.3. Метод трех и пяти ординат	82
3.3.4. Метод функций Бесселя	83
3.3.5. Метод угла отсечки	84
3.4. Нелинейное усиление и умножение частоты	87
3.5. Преобразование частоты сигнала	91
3.6. Перемножение сигналов	93
Глава 4. Модуляция и демодуляция сигналов	96
4.1. Общие сведения о модулированных сигналах	96
4.2. Амплитудная модуляция	98
4.3. Балансная модуляция	101
4.4. Однополосная модуляция	103
4.5. Формирование сигналов с амплитудной модуляцией	105
4.6. Методы демодуляции	106
4.6.1. Амплитудная модуляция	106
4.6.2. Детектирование сигналов с балансной и однополосной модуляцией	110
4.7. Угловая модуляция	113
4.8. Формирование и детектирование сигналов с угловой модуляцией	119
Глава 5. Основные виды дискретной модуляции сигналов	125
5.1. Дискретная амплитудная модуляция	125
5.2. Дискретная фазовая модуляция	128
5.3. Квадратурная амплитудная модуляция	136
5.4. Амплитудно-фазовая модуляция	139
5.5. Дискретная частотная модуляция	141
5.6. Дискретная частотная модуляция с непрерывной фазой	144
5.7. Дискретная частотная модуляция с минимальным сдвигом	146
5.8. Обработка ММС-сигналов на основе алгоритма Витерби	150
5.9. Амплитудно-импульсная модуляция	153
5.10. Широко-импульсная модуляция	155
5.11. Временная (фазовая) импульсная модуляция	158
5.12. Импульсно-кодовая модуляция	159
5.13. Дифференциальная импульсно-кодовая модуляция	162
5.14. Дельта-модуляция	163
Глава 6. Математические модели случайных сигналов и помех	166
6.1. Методы задания случайных процессов	166
6.2. Стационарность и эргодичность случайных процессов	171
6.3. Спектральные характеристики случайных процессов	174

6.4. Случайные процессы, определяемые одномерной и двухмерной плотностями вероятностей.....	180
6.5. Узкополосные случайные процессы.....	184
6.6. Огибающая и фаза квазигармонического случайного процесса.....	188
6.7. Некоторые модели источников сообщений.....	193
Глава 7. Модели каналов передачи информации.....	196
7.1. Классификация каналов.....	196
7.2. Модели непрерывных каналов.....	198
7.2.1. Общие сведения.....	198
7.2.2. Канал с аддитивным гауссовским шумом.....	199
7.2.3. Канал с неопределенной фазой сигнала и аддитивным шумом.....	199
7.2.4. Однолучевой гауссовский канал с флуктуациями амплитуд и фаз сигнала (с общими замираниями).....	200
7.2.5. Многолучевой гауссовский канал с частотно-селективными замираниями.....	200
7.2.6. Канал со сложной аддитивной помехой.....	201
7.2.7. Канал с межсимвольной интерференцией.....	202
7.3. Модели дискретных каналов связи.....	203
7.3.1. Общие сведения.....	203
7.3.2. Дискретный симметричный канал без памяти.....	205
7.3.3. Двоичный симметричный канал со стираниями.....	207
7.3.4. Дискретный несимметричный канал без памяти.....	208
7.3.5. Дискретный канал с памятью.....	209
7.4. Линейные преобразования случайных сигналов в каналах связи.....	210
7.5. Нелинейные преобразования случайных сигналов в каналах связи.....	213
7.6. Случайные преобразования сигналов в каналах связи.....	214
Глава 8. Элементы теории помехоустойчивого приема дискретных сообщений.....	218
8.1. Постановка задачи синтеза оптимального приемника.....	218
8.2. Статистические критерии оптимального приема сигналов.....	223
8.3. Синтез оптимального когерентного приемника в условиях аддитивного шума.....	225
8.4. Оптимальный приемник на согласованных фильтрах.....	229
8.5. Методы реализации согласованных фильтров.....	234
8.6. Помехоустойчивость оптимального когерентного приема двоичных сигналов.....	236
8.7. Анализ помехоустойчивости оптимального приема двоичных сигналов.....	238
8.7.1. Сигналы с дискретной амплитудной модуляцией.....	238
8.7.2. Сигналы с дискретной частотной модуляцией.....	239
8.7.3. Сигналы с дискретной фазовой модуляцией.....	240

8.7.4. Сигналы с относительной фазовой модуляцией.....	242
8.8. Некогерентный прием двоичных сигналов с дискретной частотной модуляцией.....	244
8.9. Некогерентный оптимальный прием сигналов с относительной фазовой модуляцией.....	250
8.10. Прием дискретных сигналов в условиях замираний	253
Глава 9. Основы теории передачи непрерывных сообщений	257
9.1. Особенности оптимального приема непрерывных сообщений. Потенциальная помехоустойчивость приема	257
9.1.1. Общие сведения.....	257
9.1.2. Определение обобщенного выигрыша для различных методов модуляции	260
9.2. Оптимальное оценивание параметров сигналов.....	266
9.3. Оптимальная фильтрация непрерывных сообщений.....	274
9.3.1. Постановка и решение задачи фильтрации непрерывных сообщений.....	274
9.3.2. Линейная фильтрация	276
9.3.3. Нелинейная фильтрация	279
9.3.4. Примеры демодуляции непрерывных сообщений на основе нелинейной фильтрации.....	281
Глава 10. Основы теории передачи и кодирования информации	290
10.1. Определение количества информации.....	290
10.2. Энтропия источника дискретных сообщений и ее свойства	291
10.3. Избыточность и производительность источника дискретных сообщений.....	293
10.4. Энтропия источника непрерывных сообщений.....	295
10.5. Количество информации, передаваемое по дискретному каналу	297
10.6. Пропускная способность дискретного канала передачи информации.....	299
10.7. Пропускная способность непрерывного канала передачи информации.....	301
10.8. Теоремы кодирования Шеннона для каналов с помехами.....	305
10.9. Классификация корректирующих кодов.....	307
10.10. Основные характеристики блочных корректирующих кодов.....	308
10.11. Линейные двоичные коды для корректирования ошибок.....	311
10.12. Декодирование линейного кода.....	315
Приложение. Таблица интеграла вероятностей.....	319
Список литературы.....	323

Учебное издание

**Биккенин Рафаэль Рифгатович,
Чесноков Михаил Николаевич**

Теория электрической связи

Учебное пособие

Редактор *В. Н. Махова*

Технический редактор *О. Н. Крайнова*

Компьютерная верстка *А. В. Бобылёва*

Корректоры *О. И. Лыкова, А. П. Сизова*

Изд. № 101108525. Подписано в печать 30.12.2009. Формат 60 × 90/16.
Гарнитура «Таймс». Печать офсетная. Бумага офс. № 1. Усл. печ. л. 21,0.
Тираж 2 000 экз. Заказ № 4595

Издательский центр «Академия», www.academia-moscow.ru

125252, Москва, ул. Зорге, д. 15, корп. 1, пом. 26б.

Адрес для корреспонденции: 129085, Москва, пр-т Мира, 101В, стр. 1, а/я 48.

Тел./факс: (495) 648-0507, 616-00-29.

Санитарно-эпидемиологическое заключение № 77.99.60.953.Д.007831.07.09 от 06.07.2009.

Отпечатано с электронных носителей издательства.

ОАО «Тверской полиграфический комбинат», 170024, г. Тверь, пр-т Ленина, 5.

Телефон: (4822) 44-52-03, 44-50-34. Телефон/факс: (4822) 44-42-15.

Home page - www.tverpk.ru Электронная почта (E-mail) - sales@tverpk.ru.

Высшее профессиональное образование

Р. Р. Биккенин
М. Н. Чесноков

ТЕОРИЯ ЭЛЕКТРИЧЕСКОЙ СВЯЗИ

Учебное пособие



Радиоэлектроника
и телекоммуникации