

Саратовский государственный университет им. Н. Г. Чернышевского

А. В. Хохлов, Т. Е. Вадивасова, А. В. Шабунин

СИГНАЛЫ

МЕТОДЫ ОПИСАНИЯ, МОДЕЛИ,
ИНФОРМАЦИОННЫЕ ВОЗМОЖНОСТИ

*Учебное пособие
для студентов физических специальностей*

Под редакцией профессора *В. С. Анищенко*

Саратов

Издательство Саратовского университета

2011

УДК 621.37:532.59:537.86:538.56
ББК 32.84
X86

Хохлов, А. В.

X86 Сигналы. Методы описания, модели, информационные возможности : учебное пособие для студентов физических специальностей / А. В. Хохлов, Т. Е. Вадивасова, А. В. Шабунин / под ред. В. С. Анищенко. – Саратов : Изд-во Саратов. ун-та, 2011. – 256 с. : ил.
ISBN 978-5-292-04061-3

В книге описаны методы представления детерминированных, случайных, цифровых и модулированных радиотехнических сигналов. Рассмотрены линейные пространства сигналов, детерминированные и случайные процессы, поля и волны, методы дискретизации и квантования аналоговых сигналов, дискретное и быстрое преобразование Фурье. Наряду с классическими методами амплитудной, частотной и фазовой модуляции анализируются случайная и цифровая модуляции сигналов.

Для студентов университетов, обучающихся по физическим специальностям, а также для аспирантов, научных работников и инженеров, специализирующихся на разработке и создании радиоэлектронных систем.

Рекомендуют к печати :

кафедра радиофизики и нелинейной динамики физического факультета
Саратовского государственного университета
доктор физико-математических наук, профессор *В. В. Астахов*
(Саратовский государственный технический университет)

Работа издана по тематическому плану 2011 года
(утвержден на Ученом совете Саратовского государственного университета,
протокол N 9 от 6 сентября 2011 года)

УДК 621.37:532.59:537.86:538.56
ББК 32.84

ISBN 978-5-292-04061-3

© Хохлов А. В., Вадивасова Т. Е,
Шабунин А. В., 2011
© Саратовский государственный
университет, 2011

Предисловие

Предлагаемое учебное пособие посвящено радиотехническим сигналам, их классификации и описанию характеристик. Эти вопросы неоднократно обсуждались в учебниках по радиофизике и радиоэлектронике. В радиоэлектронной литературе чаще всего анализируются детерминированные аналоговые и модулированные сигналы, а случайные процессы рассматриваются только как шумы и помехи. Статистическая радиофизика, напротив, исследует в основном характеристики случайных сигналов. Цифровой обработке сигналов посвящены специальные монографии. При этом описания одних и тех же сигналов используют различный подход, а иногда и разные определения.

Особенности данной книги состоят в том, что здесь объединены различные подходы к описанию радиотехнических сигналов. В ней с единых позиций рассматриваются детерминированные, случайные, цифровые и модулированные сигналы. Рассмотрены модели случайных сигналов от простых случайных процессов до марковских диффузионных процессов, задаваемых стохастическими дифференциальными уравнениями, цифровые методы описания и обработки сигналов, дискретное и быстрое преобразование Фурье, случайная модуляция и методы цифровой модуляции. Такой подход позволяет авторам сделать выводы об информационных возможностях различных сигналов и перспективах использования в современных системах связи и передачи информации.

Строгого общепринятого определения термина «сигнал», по-видимому, не существует. В научной и учебной литературе сигналами называют как процессы, передающие информационное сообщение, так и закономерно повторяющиеся периодические колебания, которые сами по себе не могут содержать информационного сообщения. Сигналами иногда называют также и случайные помехи. Авторы употребляют данный термин в самом широком смысле и называют сигналами колебательные процессы, возникающие в различных узлах радиоустройств и передаваемые системами связи.

Основу книги составили лекции, читаемые кафедрой радиофизики и нелинейной динамики на протяжении ряда лет на физическом факультете и факультете нелинейных процессов Саратовского университета, а также учебные пособия, изданные в Саратовском университете, и статьи в периодических изданиях.

Авторы книги являются опытными преподавателями и интенсивно работают в области создания различных радиоэлектронных систем, многие из которых защищены авторскими свидетельствами и патентами Российской Федерации.

Профессор *В. С. Анищенко*

Введение

Сигналы, сигналы, сигналы. Этот термин представляется всеобъемлющим. Видимо, нет области знаний, где бы не использовались сигналы.

Понятие сигнала не поддается строгому определению. По традиции сигналом (от лат. *signum* – знак) называют физический процесс, изменение характеристик которого во времени несет сообщение о каком-либо событии, состоянии объекта или о передаваемой команде управления.

Совершенно исключительное значение термин «сигнал» приобрел в физических науках: в радиофизике и теории связи, в теории информации, автоматике, биофизике и т.д. В радиофизике сигналами называют колебательные и волновые процессы естественного и искусственного происхождения, передающие информационное сообщение, или закономерно повторяющиеся периодические колебания, которые сами по себе не могут содержать информационного сообщения. Сигналами являются также и различные случайные процессы (и помехи).

Дадим краткую характеристику сигналов. В простейшем случае сигнал может быть описан некоторой заранее известной функцией. Например, гармонический сигнал можно задать в виде $x(t) = A \sin(\omega t + \varphi)$, а затухающие колебания в виде $x(t) = A \exp(-at) \sin(\omega t + \varphi)$. Такие сигналы называют детерминированными. Они заранее заданы на бесконечном интервале времени $t \in (-\infty, +\infty)$. Считается, что детерминированный сигнал создается некоторой детерминированной системой. В случае периодического сигнала порождающая его система должна быть неизменной во времени и совершать бесконечный во времени строго повторяющийся колебательный процесс в одних и тех же условиях. Ни бесконечная протяженность процесса во времени, ни его строгая периодичность в действительности не могут быть реализованы.

Любой реальный сигнал всегда содержит некоторую случайную (флуктуационную) составляющую, в той или иной степени существенную или несущественную. Если нельзя пренебречь флуктуационной составляющей сигнала, то для описания сигнала недостаточно детерминированных функций и мы должны использовать аппарат теории случайных функций и случайных процессов.

Важно отметить, что случайным может быть не только сигнал естественного происхождения (шумовая помеха), но и искусственные сигналы, несущие информационные сообщения. Действительно, если сигнал задан заранее с помощью известной функции, то какую же информацию он может содержать? Информацию мы получаем тогда, когда узнаем что-то новое о состоянии объекта, а это состояние в каждый момент времени полагается случайным, так что мы не можем знать его заранее.

В радиоэлектронике и теории связи принято разделять случайные сигналы на полезные, несущие информацию, и помехи. Такое разделение представляется весьма условным. Что представляет собой помеха в канале связи? Это случайный процесс естественного происхождения,

например собственное излучение астрономических объектов, земных покровов или грозových разрядов. Но, будучи помехами для канала связи, эти же случайные процессы служат полезными сигналами для соответствующих радиометрических исследований. К тому же описания и полезные для канала связи сигналов, и помех обычно используют одни и те же статистические модели.

Математическая теория случайных сигналов является основой статистической радиофизики и тесно связана с теорией информации в силу единства используемого математического аппарата. Разница заключается лишь в том, что статистическая радиофизика исследует случайные сигналы, т.е. физические процессы определенной (электромагнитной) природы, в то время как теория информации является математической дисциплиной и занимается не самими сигналами, а информационными сообщениями, которые содержатся в сигналах. При этом физическая природа сигналов предметом теории информации не является.

Радиофизику прежде всего интересуют физические свойства сигналов, их характеристики, методы изменения параметров несущих колебаний (модуляции), способы дискретизации и кодирования и т.п. Эти процессы удобнее всего рассматривать на примерах простых детерминированных сигналов, допускающих аналитическое представление в виде непрерывных функций гармонических колебаний или последовательностей прямоугольных импульсов. На таких моделях легко вводить модуляцию амплитуды, частоты и фазы, варьировать длительность прямоугольных импульсов, частоту повторения и задержки во времени. Поэтому детерминированные сигналы представляются идеальными моделями сигналов.

В предлагаемом учебном пособии обсуждаются различные типы сигналов, их математические модели, основные характеристики и информационные возможности. Оно содержит пять глав.

Первая глава является вводной. Здесь проводится классификация сигналов, вводятся линейные пространства сигналов и рассматриваются источники сообщений с позиций теории информации.

Вторая глава посвящена детерминированным колебаниям и волнам. Основное внимание уделяется спектральному анализу сигналов и использованию обобщенных функций Дирака.

В третьей главе проводится анализ случайных процессов, полей и волн. Рассматриваются различные виды случайных процессов и их математические модели. Основное внимание уделено марковским случайным процессам и стохастическим дифференциальным уравнениям.

Четвертая глава посвящена цифровым методам обработки сигналов. Рассматриваются методы дискретизации аналоговых сигналов и представление дискретных сигналов в цифровой форме, методы дискретного и быстрого преобразования Фурье.

Пятая глава содержит описание модулированных и узкополосных сигналов. Наряду с классическими методами модуляции здесь рассмотрены методы случайной и цифровой модуляции. В последнем разделе анализируется структура телевизионных сигналов.

К каждой главе для более детального ознакомления с данным кругом вопросов и самостоятельной работы студентов приводятся задания и примеры, а также список дополнительной литературы.

Глава 1

Элементы общей теории сигналов

Сигналы становятся объектами теоретических исследований и расчётов только при наличии строгих математических описаний – математических моделей. Ни один реальный сигнал не удастся полностью описать математически. Поэтому при создании математической модели приходится абстрагироваться от конкретной природы колебания и некоторых свойств, не существенных для рассмотрения.

Моделями электрических колебаний напряжений и токов могут служить графики и таблицы, комплексные числа и векторы, но чаще всего используются *функции времени*, векторные $\mathbf{x}(t)$ или скалярные $x(t)$. Множество функций времени образует некое *функциональное пространство*, а обобщающая концепция, позволяющая упорядочить и классифицировать это множество, должна базироваться на математическом аппарате теории функций и функционального анализа¹.

1.1. Классификация электрических сигналов

Возможны различные подходы к классификации сигналов. Вот один из них. Если функция $x(t)$ задана аналитически, то ее значения известны для любых моментов времени и она представляет не сигнал, несущий информацию, а некий колебательный процесс, который можно использовать в качестве несущего для передачи информации.

Когда несущее колебание промодулировано информационным сигналом, оно становится случайным процессом даже в отсутствии флуктуационных составляющих и описывается случайной функцией $X_{\text{мод}}(t)$.

Если одновременно с сигналом, заданным аналитически, действует флуктуационная составляющая $\xi(t)$, то сумма $x(t) + \xi(t)$ – это случайный процесс, обусловленный действием флуктуационной составляющей.

Наконец, сигнал $X_{\text{мод}}(t) + \xi(t)$ представляет сложный случайный процесс, включающий информационное сообщение и флуктуационную составляющую.

В радиоэлектронике и электротехнике сигналы классифицируют по назначению и мощности сигналов, по используемым частотам, методу модуляции и т.д., но большинство из них характеризует разделение не самих сигналов, а скорее систем приема, передачи и обработки сигналов.

¹Функциональный анализ – это раздел математики, изучающий различные пространства как множества объектов с заданными между ними формальными соотношениями, определяющими структуру, и их отображения.

Так, разделение по мощности по существу не является классификацией, так как сигналы разной мощности могут быть идентичными по структуре и различаться только амплитудой.

В настоящей книге принята следующая классификация сигналов. Прежде всего, сигналы традиционно делятся по возможности предсказать их значения в последующие моменты времени на детерминированные и случайные. Детерминированные сигналы подразделяются на периодические, квазипериодические и непериодические, а случайные сигналы – на стационарные и нестационарные, на эргодические, нормальные (гауссовские) и марковские процессы. В свою очередь марковские процессы подразделяются на марковские цепи, дискретные марковские процессы и диффузионные марковские процессы.

Детерминированные и случайные процессы могут быть скалярными или векторными, комплексными или вещественными, широкополосными или узкополосными, дифференцируемыми или недифференцируемыми.

По способу представления сигналы можно разделить на аналоговые, дискретные и цифровые. В свою очередь аналоговые сигналы могут быть непрерывными, разрывными или описываться только обобщенными функциями, а дискретные – с постоянным или с переменным шагом дискретизации, оцифрованные или неоцифрованные. Цифровые сигналы разделяются на несколько подклассов в зависимости от используемых импульсных кодов и способа кодирования.

Наконец, еще один способ классификации сигналов связан с характером передаваемой информации (радиовещание, телевидение, радиолокация, телефония, управление) и способами модуляции несущих колебаний. В зависимости от характеристик несущего колебания и вида модулирующего процесса сигналы делятся на модулированные, манипулированные или цифровые, причем модуляция может осуществляться детерминированным или случайным сигналом по амплитуде или по углу (частотная или фазовая модуляция). Способы цифровой модуляции также разделяются на подклассы: амплитудная цифровая модуляция – на импульсно-кодовую и дельта-модуляцию, фазовая – на многопозиционную, квадратурную фазовую и амплитудно-фазовую, частотная – на модуляцию с минимальным частотным сдвигом, ортогональное частотное разделение и др. Вид и структура модулированного сигнала определяют допустимую емкость информационных сообщений и информационные возможности различных модулированных сигналов.

Детерминированные и случайные сигналы. Электрические сигналы (напряжения $U(t)$ или тока $I(t)$), мгновенные значения которых в любой момент времени можно предсказать точно, т.е. с вероятностью, равной единице, называются *детерминированными* (от лат. *determino* – определяю), а сигналы, значения которых предсказуемы лишь с некоторой вероятностью, меньшей единицы, – *случайными*.

Разделение сигналов на детерминированные и случайные основывается на возможности воспроизвести процесс в ходе экспериментов. Так, детерминированные колебания произвольного числа идентичных генераторов периодического напряжения, реализуемые при одних и тех же начальных условиях, в пределах точности эксперимента будут одинаковыми и могут быть представлены математической моделью в виде из-

вестной функции $x(t)$. Детерминированные колебания используются в радиоэлектронике в качестве несущих, испытательных (тестирующих) или управляющих и очень удобны в качестве упрощенных моделей более сложных сигналов.

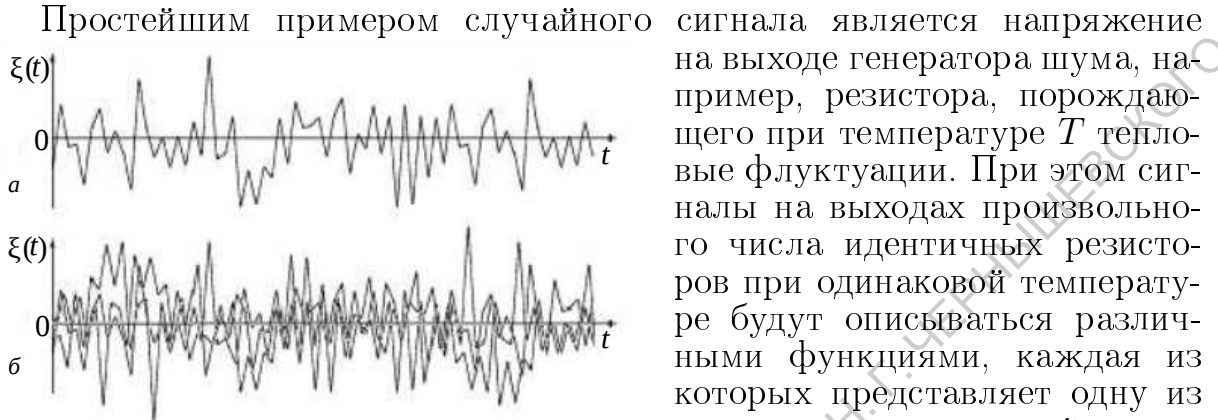


Рис. 1.1. Напряжения на выходах одного (а) и трёх (б) генераторов шума

Простейшим примером случайного сигнала является напряжение на выходе генератора шума, например, резистора, порождающего при температуре T тепловые флуктуации. При этом сигналы на выходах произвольного числа идентичных резисторов при одинаковой температуре будут описываться различными функциями, каждая из которых представляет одну из возможных *реализаций* сигнала. В качестве иллюстрации на рис. 1.1 представлены осциллограммы напряжений на выходах одного (а) и трёх (б) резисторов, полученные одновременно. Строгая математическая модель случайного сигнала должна учитывать возможность одновременного существования неограниченного множества реализаций, обладающих схожей структурой и временным поведением. На первый взгляд может показаться, что математическую модель случайных сигналов легко построить, используя распространённые методы теории вероятностей. Однако для описания поведения сигналов во времени методы классической теории вероятностей оказываются недостаточными и приходится использовать аппарат особой ветви математики – теории случайных процессов.

Между детерминированными и случайными сигналами нет резкой границы. Например, детерминированный управляющий сигнал при передаче может испытать влияние случайных воздействий и, строго говоря, перестает быть предсказуемым. Сигналы, несущие информацию, являются случайными, так как содержат непредсказуемые сообщения. Поэтому в радиоэлектронике и статистической радиофизике принято делить случайные процессы на *полезные сигналы* (несущие сообщения) и *помехи*, препятствующие правильной передаче сообщений.

Скалярные и векторные сигналы. Напряжение $U(t)$ на зажимах генератора колебаний выражает зависимость единственной переменной от времени и потому называется *скалярным* сигналом. В радиоэлектронике и статистической радиофизике используются и *векторные (многомерные)* сигналы в виде совокупности нескольких скалярных сигналов одинаковой длительности. Векторными являются напряжённости электрического $\mathbf{E}(t)$ или магнитного $\mathbf{H}(t)$ поля, обладающие тремя пространственными координатами, стереофонический сигнал, содержащий в отличие от монофонического две скалярные функции, соответствующие звуковым колебаниям правого и левого каналов. Векторными по существу являются и комплексные сигналы, если рассматривать их действительные и мнимые компоненты как скалярные функции времени. В общем случае N -мерный сигнал $\mathbf{x}(t)$ представляется выражением

$$\mathbf{x}(t) = \{x_1(t), x_2(t), x_3(t), \dots, x_N(t)\}. \quad (1.1)$$

Особенностью всех рассмотренных векторных сигналов является внутренняя связь и упорядоченность скалярных функций, так как при перестановке составляющих возникает новый векторный сигнал. Из нескольких одномерных сигналов, например входных напряжений многополюсника, тоже можно составить многомерный сигнал. Если все скалярные сигналы являются непрерывными функциями времени, то их анализ и обработка осуществляются математическими методами, разработанными для одномерных сигналов. Если же все или некоторые функции заданы только для дискретных значений времени, то обработка многомерных сигналов может значительно усложниться.

Аналоговые, дискретные и цифровые сигналы. Многие реальные физические процессы развиваются непрерывно во времени. Описывающие их колебания $x(t)$ в любые моменты времени принимают любые значения в некотором ограниченном интервале $[a, b]$ (рис. 1.2, а) или в интервале $(-\infty, \infty)$. Так выглядит сигнал на экране осциллографа.

Множество возможных значений $x(t)$ является непрерывным (образует континуум). Непрерывные по времени сигналы, с непрерывным множеством состояний называются *аналоговыми*. Если значения сигнала $x(t)$ задавать в дискретные моменты времени через некоторый интервал Δ_t , называемый *шагом дискретизации*, то получим *дискретный во времени сигнал* (рис. 1.2, б), представляющий собой последовательность отсчётов в соответствующие моменты времени. Последовательность моментов времени можно пронумеровать, т.е. перейти к дискретному времени n , принимающему значения на множестве натуральных чисел. Такую последовательность получает любой экспериментатор, периодически записывающий показания измерительного прибора.

До середины прошлого века в радиоэлектронике использовались аналоговые сигналы. Замена непрерывных сигналов дискретными позволила значительно увеличить информационные возможности сигналов, в частности, при передаче вектор-

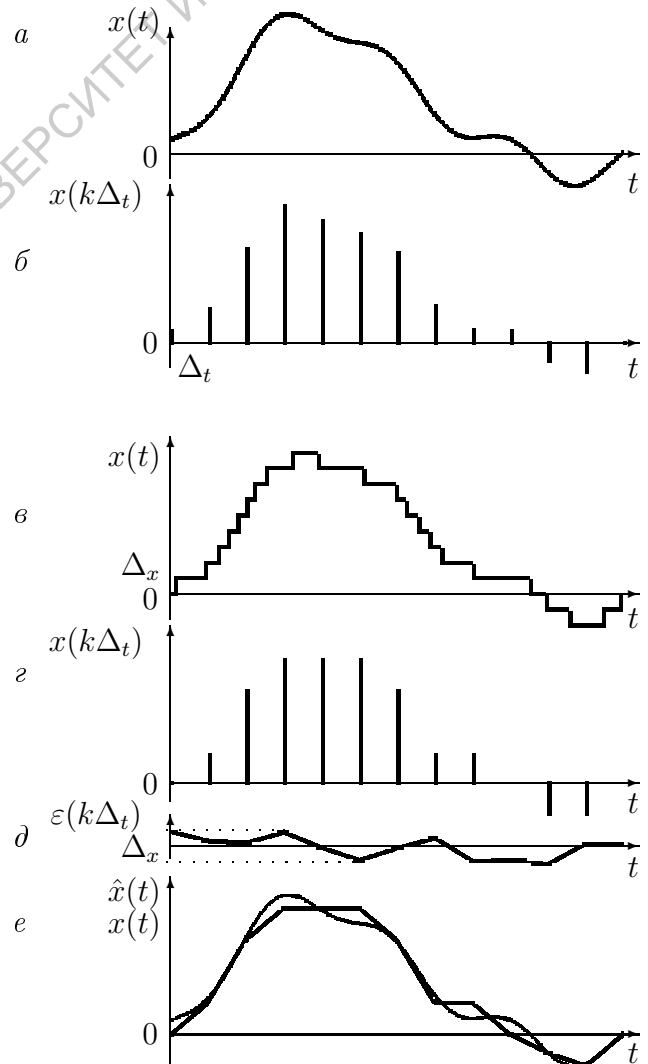


Рис. 1.2. Дискретизация аналогового сигнала

ных сигналов, и способствовала развитию методов временного разделения каналов связи.

Непрерывный или дискретизованный по времени сигнал $x(t)$ можно дискретизовать и по переменной x . Для этого диапазон возможных значений сигнала разбивается на конечное число уровней, называемых *уровнями квантования*, и возникает *квантованный сигнал* (рис. 1.2, в). Величина $\Delta_n = x_{n+1} - x_n$ называется *шагом квантования*. Пример квантованного сигнала – показания электронного вольтметра, фиксирующего два знака после запятой (шаг квантования – 10 мВ). Шаг квантования чаще всего выбирается постоянным $\Delta_n = \text{const}$, хотя квантование может проводиться и с переменным шагом, например логарифмическим.

Функция $x(t)$ может быть и дискретизована по t , и квантована по x (рис. 1.2, г). В этом случае значения $x_n(k)$ получаются заменой истинных значений $x(k\Delta_t)$, где k – номер отсчёта (рис. 1.2, б), ближайшими целыми значениями $n\Delta_x$, где n – номер уровня квантования, т.е. значения $x(t)$ заменяются дискретной последовательностью чисел $x_n(k)$. Дискретизованный по t и квантованный по x сигнал $x(t)$, уровни квантования которого представлены в цифровой форме, называется *цифровым сигналом*. Цифровые сигналы имеют конечное множество значений, определенных на конечных интервалах времени². Таким образом, происходит преобразование бесконечного числа значений аналогового сигнала в конечное число значений цифрового сигнала.

При переходе от аналоговых сигналов к квантованным всегда возникает ошибка квантования уровней $\varepsilon(k\Delta)$ (рис. 1.2, д), причём абсолютная величина ошибки не превышает $\Delta_x/2$. Восстановление непрерывного сигнала $\hat{x}(t)$ на основе квантованных отсчётов (рис. 1.2, е) сопровождается появлением фонового *шума квантования* в виде разности $x(t) - \hat{x}(t)$. Например, для звуковых сигналов шум проявляется как шипение.

1.2. Линейные пространства сигналов

Пространство как множество объектов, между которыми установлены отношения, определяющие его структуру, является одним из основных понятий современной математики. Объектами пространства могут быть числа, векторы, функции, их различные комбинации, а отношениями – гипотетические правила, связывающие два или более объектов. Примером может служить векторное (линейное) пространство как обобщающее понятие совокупности векторов трехмерного пространства.

Элементами векторного пространства (векторами в обобщённом смысле) могут быть и функции, описывающие сигналы. Основанием для этого служит тот факт, что линейная комбинация аналоговых (дискретных) сигналов тоже является аналоговым (дискретным) сигналом. Введение геометрических методов и элементов функционального анализа позволяет создать стройную теорию сигналов и установить соответствие

²Цифровые сигналы образуются при замене каждого значения $x_n(k)$ кодовой комбинацией, составленной из символов системы счисления, и содержат, как правило, одинаковое число символов. В последнее время цифровыми стали называть сигналы, принимающие только 2 значения (1 и 0 или 1 и -1), т.е. кодированные бинарными последовательностями.

между сигналами и результатами их линейных измерений.

Рассмотрим множество сигналов M и каждый сигнал будем представлять точкой. Если ввести некоторое свойство P (например гармоничности или периодичности), то сигналы, обладающие этим свойством образуют подмножество $M' \in M$. Для обозначения различия между элементами подмножества M' можно ввести отношение, называемое *расстоянием* d , т.е. каждой паре элементов $x_1(t), x_2(t) \in M'$ сопоставить *функционал*³ $d(x_1(t), x_2(t)) \geq 0$, называемый *метрикой* и удовлетворяющий следующим свойствам (аксиомам):

$$d(x_1(t), x_2(t)) = 0 \text{ тогда и только тогда, когда } x_1(t) = x_2(t);$$

$$d(x_1(t), x_2(t)) = d(x_2(t), x_1(t)); \quad d(x_1(t), x_3(t)) \leq d(x_1(t), x_2(t)) + d(x_2(t), x_3(t)),$$

где $x_1(t), x_2(t), x_3(t)$ – произвольные элементы подмножества M' .

Подмножество M' с заданной метрикой d образует *метрическое пространство сигналов*. Понятие метрики может оказаться полезным для оценки качества аппроксимации одного сигнала другим.

В радиоэлектронике в качестве метрики для аналоговых сигналов, определённых и квадратично интегрируемых в области V (на некотором ограниченном или неограниченном интервале $(a, b) \equiv V$), обычно используется квадратичная метрика в виде функционала

$$d(x_1(t), x_2(t)) = \sqrt{\int_a^b |x_1(t) - x_2(t)|^2 dt}. \quad (1.2)$$

Дискретизованный по времени сигнал $x(k\Delta_t)$ представляет последовательность вещественных чисел $x(k)$, $k = 1, 2, \dots, N$. Тогда квадратичная метрика выглядит как

$$d(x_1, x_2) = \sqrt{\sum_{k=1}^N |x_1(k) - x_2(k)|^2}. \quad (1.3)$$

Если число отсчётов N – конечно, то сигнал представляет вектор с N компонентами. Множества таких сигналов, все элементы которых являются вещественными или комплексными числами, представляют собой арифметические пространства \mathbb{R}^N или \mathbb{C}^N соответственно.

Это был первый шаг в установлении структуры пространства сигналов. Следующий шаг заключается в установлении полноты метрического пространства.

Последовательность x_i , $i = 1, 2, \dots$ точек метрического пространства будем называть *фундаментальной*, если для любого ε существует такое $N(\varepsilon)$, что $d(x_i, x_j) < \varepsilon$ для всех $i > N(\varepsilon)$, $j > N(\varepsilon)$. Если в пространстве любая фундаментальная последовательность сходится относительно заданной метрики, то это пространство называется *полным*.

Важным шагом в усовершенствовании структуры пространства сигналов является задание алгебраических операций сложения и умножения на число.

³Функционал – это оператор, преобразующий функции в числовые значения.

Множество M вещественных (комплекснозначных) функций $x(t)$, определённых и квадратично интегрируемых в области V , называется вещественным (комплексным) *линейным* или *векторным* пространством, а элементы множества – *векторами*, если оно удовлетворяет следующим условиям:

1) для любых $x_1(t), x_2(t) \in M$ определён третий элемент $x_3(t) \in M$, называемый суммой $x_1(t)$ и $x_2(t)$ и обозначаемый $x_1(t) + x_2(t)$, причём

$$x_1(t) + x_2(t) = x_2(t) + x_1(t) \quad (\text{коммутативность}),$$

$$x_1(t) + (x_2(t) + x_3(t)) = (x_1(t) + x_2(t)) + x_3(t) \quad (\text{ассоциативность});$$

2) множество M содержит нулевой вектор \emptyset , такой что $x(t) + \emptyset = x(t)$ для всех $x(t) \in M$ (существование нуля);

3) для каждого $x(t) \in M$ существует такой элемент $(-x(t))$, что $x(t) + (-x(t)) = 0$ (существование противоположного элемента);

4) для любого вещественного (комплексного) числа α и любого элемента $x(t) \in M$ определён элемент $\alpha x(t) \in M$ (*произведение* элемента $x(t)$ на число α);

5) для любых $x_1(t), x_2(t) \in M$ и любых вещественных (комплексных) чисел α, β выполняются равенства $(\alpha\beta)x_1(t) = \alpha(\beta x_1(t))$; $1 \cdot x_1(t) = x_1(t)$; $(\alpha + \beta)x_1(t) = \alpha x_1(t) + \beta x_1(t)$; $\alpha(x_1(t) + x_2(t)) = \alpha x_1(t) + \alpha x_2(t)$.

Конечное (или бесконечное) подмножество векторов e_i множества M называется *линейно независимым координатным базисом*, если из равенства $\sum_{i=1}^{\infty} \alpha_i e_i = 0$ следует, что все $\alpha_i \equiv 0$ одновременно. Если сигнал $\mathbf{x}(t)$ можно представить в виде

$$\mathbf{x}(t) = \sum_{k=1}^{\infty} c_k \mathbf{e}_k, \quad (1.4)$$

то числа c_1, c_2, \dots будут координатами сигнала $\mathbf{x}(t)$ относительно выбранного базиса.

Все сигналы из множества M можно представить в виде линейной комбинации базисных сигналов. Например, непрерывно дифференцируемый произвольное число раз сигнал $\mathbf{x}(t) = \sum_{k=0}^{\infty} c_k t^k$ можно рассматривать как разложение относительно координатного базиса $\{\mathbf{e}_0 = 1, \mathbf{e}_1 = t, \mathbf{e}_2 = t^2, \mathbf{e}_3 = t^3, \dots\}$. Число базисных векторов есть *размерность пространства*. В арифметических пространствах \mathbb{R}^N или \mathbb{C}^N число базисных векторов равно N .

Для элементов линейного пространства можно ввести числа, характеризующие их «размер». Таким образом, происходит объединение геометрических свойств метрического и алгебраических свойств линейного пространства.

Нормированные линейные пространства. Пусть задано вещественное или комплексное линейное пространство сигналов M .

Назовем *нормой* произвольного элемента линейного пространства M действительное число $\|x(t)\| \geq 0$, удовлетворяющее условиям:

1) $\|x(t)\| = 0$ тогда и только тогда, когда $x(t) \equiv 0$;

2) $\|\alpha x(t)\| = |\alpha| \|x(t)\|$ для любого числа α ;

3) $\|x_1(t) + x_2(t)\| \leq \|x_1(t)\| + \|x_2(t)\|$ для любых $x_1(t), x_2(t) \in M$.

Линейное пространство M , в котором заданы нормы элементов пространства, называется *нормированным линейным пространством*.

Отметим, что норма вектора равна расстоянию точки от начала координат. Всякое нормированное пространство становится метрическим, если в нем ввести расстояние как

$$d(x_1(t), x_2(t)) = \| x_1(t) - x_2(t) \| .$$

Нормы сигналов, как и метрики пространства, можно вводить разными способами. В радиоэлектронике и статистической радиофизике для вещественных и комплекснозначных аналоговых сигналов в качестве нормы используют *функционалы*:

$$\| x(t) \| = \sqrt{\int_a^b |x(t)|^2 dt} \quad \text{или} \quad \| x(t) \| = \sqrt{\int_{-\infty}^{\infty} x(t)x^*(t)dt}, \quad (1.5)$$

где $*$ – символ комплексного сопряжения, а для дискретных сигналов

$$\| x \| = \sqrt{\sum_{k=1}^N |x(k)|^2}, \quad (1.6)$$

где $x(k)$ – k -ый дискретный отсчёт функции $x(t)$. Если в качестве $x(t)$ рассматривать напряжение или ток, то квадрат нормы непрерывного сигнала эквивалентен энергии, а квадрат нормы дискретного сигнала – мощности, выделяемой в резисторе с единичным сопротивлением. Поскольку сигналы могут быть и детерминированными, и случайными, появляется возможность сравнивать различные сигналы по величине энергии (мощности) без учета их формы.

Метрика и норма имеют одинаковую физическую размерность, а d^2 характеризует энергию (мощность) разности двух сигналов: чем больше $d(x_1(t), x_2(t))$, тем сильнее различаются сигналы.

Евклидовы и гильбертовы пространства. В линейных пространствах сигналов кроме нормы и метрики можно ввести ещё один функционал – бинарный оператор, ставящий в соответствие каждой паре сигналов $x_1(t)$ и $x_2(t)$ скаляр $(x_1(t), x_2(t))$ – *скалярное произведение* сигналов. Скалярное произведение аналоговых сигналов, определённых для $-\infty < t < \infty$, задается формулой

$$(x_1(t), x_2(t)) = \int_{-\infty}^{\infty} x_1(t)x_2^*(t)dt, \quad (1.7)$$

а для дискретизованных по времени сигналов – формулой

$$(x_1, x_2) = \sum_{k=1}^N x_1(k)x_2^*(k). \quad (1.8)$$

Скалярные произведения сигналов удовлетворяют следующим условиям:

$$(x_1(t), x_2(t)) = (x_2(t), x_1(t))^*, \quad (x_1(t), \alpha x_2(t)) = \alpha(x_1(t), x_2(t)),$$

$$(x_1(t), x_2(t) + x_3(t)) = (x_1(t), x_2(t)) + (x_1(t), x_3(t)),$$

$(x(t), x(t)) \geq 0$, причём $(x(t), x(t)) = 0$ только при $x(t) \equiv 0$, а норму функции $x(t)$ можно определить через скалярный квадрат:

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

Скалярное произведение порождает норму, а норма – метрику, т.е. пространства со скалярным произведением являются метрическими.

Линейное пространство со скалярным произведением называется *унитарным*, а конечномерное унитарное действительное векторное пространство – *евклидовым пространством*. Евклидовым, например, является арифметическое пространство \mathbb{R}^N со стандартными операциями сложения-умножения и фиксированным в нем скалярным произведением (1.8). Можно показать, что евклидово пространство – полное, т.е. для любого ε существует такое N , что

$$|x_i(t) - x_j(t)| < \varepsilon$$

для всех $t \in V$ при $i, j > N$, и последовательность значений $\{x_i(t)\}$ функции $x(t)$ сходится.

Полное унитарное векторное пространство бесконечного числа измерений называется *гильбертовым пространством* и обозначается \mathbb{H} . Бесконечномерные гильбертовы пространства важны для приложений и представляют естественное обобщение понятия конечномерного евклидова пространства. Всякое подпространство гильбертова пространства \mathbb{H} является либо конечномерным евклидовым пространством \mathbb{R}^N , либо само представляет гильбертово пространство \mathbb{H} .

В гильбертовом пространстве справедливо фундаментальное неравенство Коши – Буняковского

$$|(x_1(t), x_2(t))| \leq \|x_1(t)\| \cdot \|x_2(t)\|. \quad (1.9)$$

Неравенство (1.9) в виде

$$\left[\sum_{k=1}^N x_1(k)x_2(k) \right]^2 \leq \sum_{k=1}^N x_1^2(k) \sum_{k=1}^N x_2^2(k)$$

было установлено Коши для N -мерного евклидова пространства, а соответствующее неравенство для пространств вещественных и комплексных функций, интегрируемых с квадратом, получено немного позднее Буняковским

$$\left| \int_{-\infty}^{\infty} x_1(t)x_2(t)dt \right|^2 \leq \int_{-\infty}^{\infty} |x_1(t)|^2 dt \int_{-\infty}^{\infty} |x_2(t)|^2 dt.$$

Рассмотрим квадрат нормы суммы двух аналоговых сигналов

$$\|x_1(t) + x_2(t)\|^2 = \int_{-\infty}^{\infty} x_1^2(t)dt + \int_{-\infty}^{\infty} x_2^2(t)dt + 2 \int_{-\infty}^{\infty} x_1(t)x_2(t)dt.$$

Здесь первые два слагаемых представляют энергии сигналов $x_1(t)$ и $x_2(t)$, а последнее – удвоенное скалярное произведение сигналов. Таким образом, сумму двух сигналов характеризуют не только энергии слагаемых,

но и энергию связи (сходства) сигналов по форме или положению в пространстве.

В евклидовом пространстве отношение скалярного произведения векторов к произведению их модулей обычно рассматривается как косинус угла φ между векторами

$$\cos \varphi = \frac{(x, y)}{\|x\| \|y\|},$$

а обращение в нуль скалярного произведения ненулевых векторов — как свидетельство об ортогональности векторов.

В комплексных пространствах угловые представления обычно не вводятся, но сигналы $x(t)$ и $y(t)$ принимаются *взаимно ортогональными* при $(x(t), y^*(t)) = 0$. Систему ортогональных функций можно сделать ортонормированной путём деления всех функций на их нормы. Ортонормальная система функций есть система попарно ортогональных сигналов, удовлетворяющая условиям

$$(x_i(t), x_j(t)) = \begin{cases} 0, & \text{если } i \neq j, \\ 1, & \text{если } i = j. \end{cases} \quad (i, j = 1, 2, \dots) \quad (1.10)$$

Избранные ортогональные базисы гильбертова пространства. Бесконечная последовательность интегрируемых с квадратом функций $\{x_k(t)\}$ является *ортонормированным (ортонормальным) базисом* в $\mathbb{H}[a, b]$ тогда и только тогда, когда выполняются следующие условия:

1) ортонормальная система функций полная (не существует дополнительных, отличных от нуля ортогональных векторов, которые можно было бы добавить, а любая $x(t) \in \mathbb{H}$, ортогональная к каждой $x_k(t)$, равна нулю почти всюду в $[a, b]$) и не содержится в какой-либо другой ортонормированной последовательности в $\mathbb{H}[a, b]$;

2) любой сигнал $x(t) \in \mathbb{H}$ может быть представлен в форме

$$x(t) = \sum_{k=1}^{\infty} c_k x_k(t), \quad \text{где } c_k = (x_k, x) \quad (k = 1, 2, \dots). \quad (1.11)$$

Формулу для c_k легко получить, если скалярно умножить $x(t)$ на $x_k(t)$ и учесть, что в силу ортонормированности из алгебраической суммы скалярных произведений в правой части полученного равенства будет отличаться от нуля только $(x_k, x_k) = 1$;

3) для любого разложения сигнала $x(t) \in \mathbb{H}[a, b]$ выполняется тождество Парсеваля

$$(x(t), x(t)) = \sum_{k=1}^{\infty} |c_k|^2. \quad (1.12)$$

Соотношение (1.11) принято называть *обобщённым рядом Фурье* сигнала $x(t)$ по ортонормальному базису $\{x_k(t)\}$, а вещественные или комплексные числа c_k — *коэффициентами Фурье* функции $x(t)$ по системе $\{x_k(t)\}$. При практическом использовании число членов ряда ограничивается определенным N и вместо $x(t)$ вводится его конечномерная аппроксимация

$$\tilde{x}(t) = \sum_{k=1}^N c_k x_k(t). \quad (1.13)$$

Возникает вопрос: сходится ли ряд (1.13), т.е. можно ли для любого ε найти такое N' и такие коэффициенты ряда, чтобы при $N > N'$ выполнялось неравенство $d = \|x(t) - \tilde{x}(t)\| < \varepsilon$?

Чтобы ответить на эти вопросы, исследуем на минимум квадрат модуля $d = \|x(t) - \tilde{x}(t)\|$ при произвольных коэффициентах c_k в (1.13)

$$\begin{aligned} d^2 &= \|x(t) - \tilde{x}(t)\|^2 = \int_{t_1}^{t_2} \left(x(t) - \sum_{k=1}^N c_k x_k(t), x(t) - \sum_{k=1}^N c_k x_k(t) \right) dt = \\ &= \int_{t_1}^{t_2} \left[\|x(t)\|^2 - 2x(t) \sum_{k=1}^N c_k x_k(t) \right] + \left[\sum_{k=1}^N c_k x_k(t), \sum_{m=1}^N c_m x_m(t) \right] dt = \\ &= \|x(t)\|^2 - 2 \sum_{k=1}^N \alpha_k c_k + \sum_{k=1}^N c_k^2 = \|x(t)\|^2 - \sum_{k=1}^N c_k^2 + \sum_{k=1}^N (\alpha_k - c_k)^2. \end{aligned}$$

Минимум d^2 достигается, когда последнее слагаемое обращается в нуль, т.е. при $\alpha_k = c_k$ ($k = 1, 2, \dots, N$).

Итак, выбирая число членов ряда (1.13), можно построить сколь угодно близкую аппроксимацию $x(t)$ частной суммой ряда $\tilde{x}(t)$ или, переходя на терминологию линейных пространств, *отобразить гильбертово пространство \mathbb{H} в евклидово пространство \mathbb{R}^N* . Правда, следует заметить, что N' будет зависеть от вида $x(t)$.

Возможность разложения непрерывных сигналов в обобщённые ряды имеет принципиальное значение, так как позволяет вместо анализа несчётного множества точек непрерывного сигнала ограничиться счётным числом коэффициентов ряда.

В математике существует огромное число ортогональных систем функций, но только некоторые из них получили распространение в радиоэлектронике и статистической радиофизике. Выбор базисных функций в пространстве сигналов определяется простотой последующего анализа.

1. Системы тригонометрических функций. Рассмотрим пространство $\mathbb{H}[-\pi, \pi]$ функций с интегрируемым квадратом. В этом пространстве функции

$$1, \cos nt, \sin nt \quad (n = 1, 2, \dots)$$

образуют полную ортонормальную систему. Если в качестве области определения функций в \mathbb{H} выбрать отрезок $[0, T]$, то ортонормальная система тригонометрических функций принимает вид

$$\Phi_0(t) = \sqrt{\frac{1}{T}}, \quad \Phi_{1C}(t) = \sqrt{\frac{2}{T}} \cos \frac{2\pi}{T}t, \quad \Phi_{1S}(t) = \sqrt{\frac{2}{T}} \sin \frac{2\pi}{T}t,$$

$$\Phi_{2C}(t) = \sqrt{\frac{2}{T}} \cos 2\frac{2\pi}{T}t, \quad \Phi_{2S}(t) = \sqrt{\frac{2}{T}} \sin 2\frac{2\pi}{T}t,$$

.....

Множитель $\sqrt{2/T}$ принято включать в коэффициенты ряда, а сам ряд (1.11) записывать в виде

$$x(t) = \frac{c_0}{2} + \sum_{k=1}^{\infty} (c_{kC} \cos k \frac{2\pi}{T} t + c_{kS} \sin k \frac{2\pi}{T} t), \quad (1.14)$$

где

$$c_0 = \frac{2}{T} \int_{-T/2}^{T/2} x(t) dt, \quad (1.15)$$

$$c_{kC} = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos k \omega_I t dt, \quad (1.16)$$

$$c_{kS} = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin k \omega_I t dt, \quad (1.17)$$

$\omega_I = 2\pi/T$ – угловая частота *первой (основной) гармоники* сигнала $x(t)$. Именно в таком виде тригонометрические ряды чаще всего используются в физике. Для радиоэлектроники такое представление не удобно, так как каждой частоте $k\omega_I$ отвечают одновременно две составляющие ряда, и необходимо устройство, способное различать синусоидальные и косинусоидальные (*квадратурные*) составляющие.

2. Система комплексных экспоненциальных функций лишена этого недостатка. Функции $\{e^{jkt}\}$ ($k = 0, \pm 1, \pm 2, \dots$), ортогональные на отрезке $[-\pi, \pi]$, нашли самое широкое распространение в качестве базисных функций в радиоэлектронике и статистической радиофизике. Разложение сигналов конечной длительности $|t| \leq T$ с интегрируемым квадратом имеет вид

$$x(t) = \sum_{k=-\infty}^{\infty} \dot{C}_k e^{jkt} \quad (1.18)$$

и называется *рядом Фурье в комплексной форме*. Из условий нормировки

$$\int_{-\pi}^{\pi} e^{jnt} e^{-jmt} dt = \begin{cases} 0, & \text{если } n \neq m, \\ 2\pi, & \text{если } n = m. \end{cases} \quad (1.19)$$

Для комплексных коэффициентов \dot{C}_k получается выражение

$$\dot{C}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} x(t) e^{-jkt} dt \quad k = 0, \pm 1, \pm 2, \dots \quad (1.20)$$

Разложение справедливо и для комплексных функций с интегрируемым квадратом модуля. Замещая в экспоненциальных функциях kt на $2\pi kt/T$ и пределы интегрирования с $[-\pi, \pi]$ на $[0, T]$, можно получить комплексный ряд Фурье в более привычном виде:

$$x(t) = \sum_{k=-\infty}^{\infty} \dot{C}_k e^{j2\pi kt/T}, \quad \dot{C}_k = \frac{1}{T} \int_0^T x(t) e^{-j2\pi kt/T} dt. \quad (1.21)$$

3. Функции Радемахера и Уолша. Интерес к функциям Радемахера и Уолша, введённым в математике ещё в 20-х годах прошлого века, связан с развитием вычислительной математики.

Зададим область определения функций интервалом $T = [0, 1]$ и построим сначала более простую систему функций Радемахера по формулам:

$$\mathbf{r}_0(t) = 1, \quad \mathbf{r}_k(t) = \text{sign}[\sin(2^k \pi t)], \quad k = 1, 2, \dots,$$

где сигнум-функция $\text{sign } x = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$ Функции Радемахера (рис. 1.3) ортонормированы на интервале $0 \leq t < 1$ и принимают значения ± 1 , но система функций неполная, так как все функции нечётны относительно середины интервала. Поэтому функции Радемахера нельзя использовать в качестве базиса для разложения сигналов в обобщённый ряд Фурье.

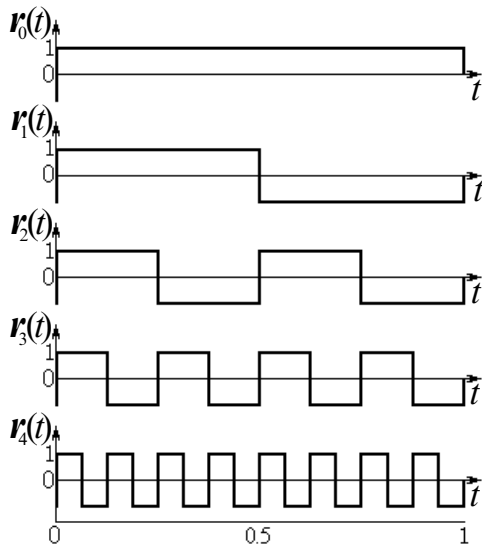


Рис. 1.3. Первые пять функций Радемахера

ния $(-1)^{(2^2 \cdot 0)} = 1$, $(-1)^{(2^2 \cdot 1/4)} = -1$, $(-1)^{(2^2 \cdot 2/4)} = 1$, $(-1)^{(2^2 \cdot 3/4)} = -1$, что хорошо согласуется с соответствующей кривой на рис. 1.3.

Развитием системы функций Радемахера явилась полная и ортонормированная последовательность функций Уолша $\{wal_k(t)\}$.

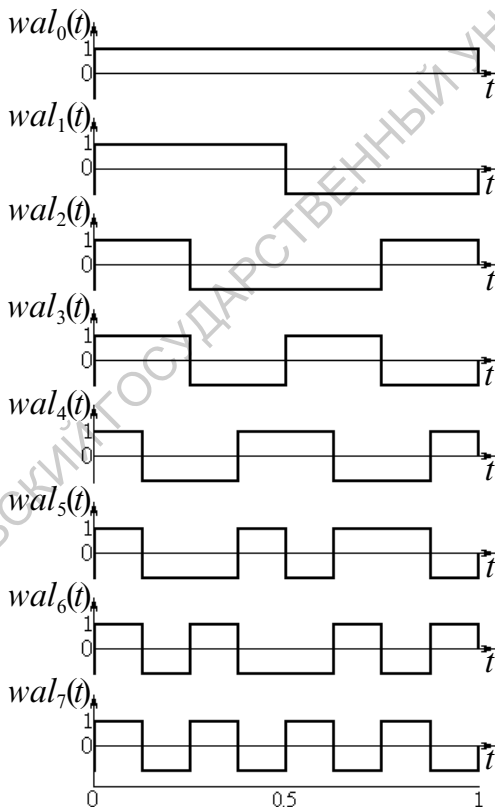


Рис. 1.4. Первые восемь функций Уолша

Последовательность функций Радемахера на том же интервале $T = [0, 1]$ можно задать иначе. Разделим область $[0, 1]$ на 2^k равных частей Δ_n и будем считать, что на интервалах Δ_n ($n = 1, 2, \dots, 2^m$) функция $r_k(t) = (-1)^{2^{k-1}n}$ и попеременно принимает постоянное значение $+1$ или -1 . Например, при $k = 2$ интервал $[0, 1]$ делится на четыре части ($2^k = 4$). Тогда функция $r_2(t)$ в четвертях интервала $[0, 1]$ последовательно принимает значения $+1, -1, +1, -1$, что хорошо согласуется с соответствующей кривой на рис. 1.3.

Достоинства несинусоидальных ортогональных функций Уолша в технике связи стали понятны уже в середине XX века, но их интенсивное применение началось лишь после того, как появилась технология изготовления интегральных микросхем и быстродействующие компьютеры.

Функции Уолша включают как чётные, так и нечётные относительно середины интервала $[0, 1]$ функции, причём все чётные функции имеют чётные номера, а нечётные функции – нечётные (рис. 1.4). Иногда даже вводятся специальные обозначения для чётных и нечётных функций Уолша $wal_{2k}(t) = cal_k(t)$, $wal_{2k-1}(t) = sal_k(t)$. Здесь первые буквы указывают на аналогию по чётности с функциями $\cos t$ и $\sin t$.

Одно из самых интересных свойств функций Уолша – мультипликативность: $wal_k(t)wal_m(t) = wal_{k \oplus m}(t)$, где $k \oplus m$ – результат поразрядного сложения по модулю двух двоичных представлений

номеров k и m . Например, результатом перемножения функций Уолша третьего (двоичный код 011) и шестого порядка (двоичный код 110) является функция пятого порядка ($0 \oplus 1 = 1, 1 \oplus 1 = 0, 1 \oplus 0 = 1$, двоичный код 101). Отсюда нетрудно заключить, что любая функция Уолша сохраняет свой порядок после умножения на функцию нулевого порядка (двоичный код не изменяется), а после умножения на себя дает функцию нулевого порядка.

Систему Уолша, состоящую из $N = 2^n$ функций всегда можно представить произведением n первых функций Радемахера. Компактный способ формирования функций различного порядка связан с процедурой превращения двоичного номера функции Уолша в код Грея и построения в соответствии с полученным кодом произведения функций Радемахера. Код Грея ($a_{k-1}^\Gamma a_{k-2}^\Gamma \dots a_0^\Gamma$) связан с двоичным кодом ($a_{k-1} a_{k-2} \dots a_0$) соотношением $a_0^\Gamma = a_0 \oplus a_1, a_1^\Gamma = a_1 \oplus a_2, \dots, a_{k-2}^\Gamma = a_{k-2} \oplus a_{k-1}, a_{k-1}^\Gamma = a_{k-1}$. Например, третья функция Уолша ($a_2 a_1 a_0 = 011$, код Грея 010) представляется второй функцией Радемахера, а пятая функция Уолша ($a_2 a_1 a_0 = 101$, код Грея 111) – произведением первых трёх функций Радемахера.

Все рассмотренные функции упорядочены по Уолшу. В ряде практических задач удобнее использовать упорядочение по Адамару $had_h(t)$ или Пэли $pal_p(t)$. Так, при упорядочении по Пэли номера функций Радемахера соответствуют двоичным номерам функций Уолша.

В общем случае по системе ортонормированных функций Уолша любую интегрируемую на интервале $[0, 1]$ функцию $x(t)$ можно представить обобщённым рядом Фурье

$$x(t) = x(0) + c_1 wal_1(t) + c_2 wal_2(t) + \dots, \quad c_k = \int_0^1 x(t) wal_k(t) dt.$$

Краткий итог раздела можно сформулировать так: введение для сигналов пространств $\mathbb{R}^N, \mathbb{C}^N$ и \mathbb{H} , нормы и метрики позволяет упростить анализ процессов сравнения, преобразования и передачи сигналов, а использование обобщённых рядов Фурье оказывается полезным при решении сложных задач радиоэлектроники и статистической радиофизики.

До сих пор рассматривались в основном детерминированные колебания, но большинство выводов этого раздела в равной степени пригодны и для случайных сигналов. Это тем более важно, что в любой радиоэлектронной системе присутствуют случайные помехи, обусловленные хаотическим движением заряженных частиц, и случайными являются все реальные сообщения, передаваемые по каналам связи, а также отдельные характеристики каналов связи.

1.3. Сигналы и информация

Важнейшей задачей различных радиоэлектронных систем является передача информации (сообщений). При этом требуется передать в течение определённого интервала времени как можно больше сведений с минимальной ошибкой. Общую схему передачи информации предложил в 1948 г. К. Шеннон. Он же впервые ввел количественные параметры информационных характеристик сигналов и каналов связи.

Сообщения источника информации отражают случайный (непредсказуемый) характер изменения его состояния. Соответственно, во времени имеет место *случайный процесс* или *случайная последовательность событий* (если рассматривать счетное множество моментов времени). Если состояние источника задается количественной характеристикой, то случайный процесс изменения состояния во времени описывается случайной функцией. Таким образом, рассмотрение информационных сигналов требует применения аппарата случайных функций. Однако во многих случаях в предположении неизменности свойств источника во времени для определения основных информационных характеристик сообщений можно ограничиться понятиями элементарной теории вероятностей, такими как случайное событие, случайная величина (дискретная и непрерывная), вероятностные характеристики случайных событий и случайных величин.

Далее, прежде, чем будут определены базовые понятия теории информации, приводятся краткие сведения из теории вероятностей. К теории случайных процессов мы обратимся в главе 3.

1.3.1. Сведения из теории вероятностей

Пусть Ω – множество (пространство) элементарных событий $\omega_i, i = 1, \dots, n$. Каждый исход некоторого испытания есть элементарное событие. С каждым элементарным событием связывается неотрицательная величина P_i , называемая *вероятностью элементарного события* ω_i . При этом выполняется условие

$$\sum_i P_i = 1.$$

С физической точки зрения вероятность характеризует частоту появления события на множестве испытаний.

В пространстве Ω можно определить систему подмножеств S , элементами которой являются события $A_i, i = 1, 2, \dots$, представляющие собой объединения элементарных событий.

Событие A называется *объединением элементарных событий* $\omega_j, \omega_k, \dots, \omega_\nu$, если оно заключается в наступлении одного из них (либо ω_j , либо ω_k, \dots , либо ω_ν). Объединение событий обозначают как

$$A = \omega_j \cup \omega_k, \dots, \cup \omega_\nu.$$

Вероятность события A равна сумме вероятностей, благоприятствующих ему элементарных событий

$$P(A) = P_j + P_k + \dots + P_\nu.$$

Для любых событий $A_1, A_2, \dots, A_m \in S$ можно определить событие, представляющее собой их *объединение* (сумму)

$$A = \bigcup_{i=1}^m A_i \in S,$$

и событие, называемое *пересечением* (произведением)

$$B = \bigcap_{i=1}^m A_i \in S.$$

Объединение A_i состоит в том, что наступает либо A_1 , либо A_2, \dots либо A_m , а пересечение B состоит в одновременном наступлении всех событий A_i .

Два события, A_1 и A_2 , называются *несовместными*, если их одновременное наступление невозможно. При этом

$$A_1 \cap A_2 = \emptyset.$$

Все элементарные события ω_i попарно несовместны. Для попарно несовместных событий A_1, A_2, \dots, A_m справедлива *теорема сложения вероятностей*:

$$P\left(\bigcup_{i=1}^m A_i\right) = \sum_{i=1}^m P(A_i).$$

Вероятность объединения двух событий, $B = A_1 \cap A_2$, в соответствии с *теоремой умножения вероятностей* определяется как

$$P(B) = P(A_1) \cdot P(A_2/A_1) = P(A_2) \cdot P(A_1/A_2),$$

где $P(A_2/A_1)$ и $P(A_1/A_2)$ – *условные вероятности событий* A_2 и A_1 при наступлении A_1 и A_2 соответственно.

События A_1 и A_2 называются *независимыми*, если

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2), \text{ т.е. } P(A_1/A_2) = P(A_1), P(A_2/A_1) = P(A_2).$$

Случайной величиной называется такая величина X , которая в результате каждого испытания может принимать то или иное заранее неизвестное значение x , т.е. принятие одного из множества значений является случайным событием.

Множество V значений x переменной X может быть дискретным (конечным или счётным) или непрерывным (возможные значения случайной величины непрерывно заполняют некоторый интервал).

Функции распределения и плотности вероятности. Полной статистической характеристикой случайной величины служит *функция распределения вероятностей* $F(x)$, определяющая вероятность P того, что случайная величина X принимает значение меньше некоторого числа x :

$$F(x) = P(X \leq x). \quad (1.22)$$

Функция распределения обладает следующими свойствами:

- 1) $F(-\infty) = 0$; $f(+\infty) = 1$; $0 \leq F(x) \leq 1$;
- 2) $F(x)$ – неубывающая функция ($F(x_1) \leq F(x_2)$ при $x_1 < x_2$);
- 3) $F(x_2) - F(x_1) = P(x_1 \leq X \leq x_2)$.

Функции распределения дискретных случайных величин являются ступенчатыми со скачками в точках x_i , а для непрерывных случайных величин – непрерывными. Если функция распределения $F(x)$ непрерывна и дифференцируема, то можно ввести *дифференциальную плотность распределения* или *плотность вероятности* случайной величины X

$$p(x) = \frac{dF(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}. \quad (1.23)$$

Плотность вероятности обладает следующими свойствами:

- 1) $p(x) \geq 0$;
- 2) $\int_{-\infty}^{+\infty} p(x) dx = 1$ (условие нормировки);

$$3) \int_{x_1}^{x_2} p(x) dx = F(x_2) - F(x_1) = P(x_1 \leq X \leq x_2).$$

Для дискретной случайной величины функция распределения $F(x)$ не дифференцируема в обычном смысле, но плотность вероятности можно получить с помощью обобщённой функции Дирака $\delta(x)$. Представим функцию распределения дискретной величины в виде

$$F(x) = \sum_i P_i \sigma(x - x_i),$$

где $P_i = P(X = x_i)$; $\sigma(x)$ – функция Хевисайда (см. раздел 2.2). Тогда

$$p(x) = \frac{dF(x)}{dx} = \sum_i P_i \delta(x - x_i). \quad (1.24)$$

Математическим ожиданием детерминированной функции $f(X)$ случайной величины X с плотностью вероятности $p(x)$ называется интеграл

$$\langle f(X) \rangle = \int_{-\infty}^{\infty} f(x) p(x) dx, \quad (1.25)$$

где символ $\langle \dots \rangle$ обозначает операцию статистического усреднения величины, указанной в угловых скобках, с соответствующей плотностью вероятности.

Вместо плотности вероятности случайной величины иногда удобно использовать *характеристическую функцию* $\Theta_X(u)$.

Характеристической функцией $\Theta_X(u)$ называется математическое ожидание случайной величины $\exp(juX)$ или преобразование Фурье от плотности распределения $p(x)$:

$$\Theta_X(u) = \langle e^{juX} \rangle = \int_{-\infty}^{\infty} e^{juX} p(x) dx \text{ для непрерывных случайных величин,}$$

$$\Theta_X(u) = \langle e^{juX} \rangle = \sum_i e^{juX_i} p_i \text{ для дискретных случайных величин.} \quad (1.26)$$

Характеристические функции удобны при определении $p(x)$ для сумм независимых случайных величин $X = \sum_k X_k$. Если известны $\Theta_k(u)$ для суммируемых случайных величин, то их плотности вероятности получают при обратном преобразовании Фурье

$$p_k(x_k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Theta_k(u) e^{-juX_k} du,$$

а характеристическая функция суммы находится простым перемножением $\Theta_k(u)$ вместо сложной операции свертки функций распределения слагаемых.

Во всех рассмотренных формулах для характеристических функций в качестве u использовалась действительная переменная, изменяющаяся от $-\infty$ до ∞ . Если в этих формулах заменить ju на лапласовскую комплексную частоту s , то формулы сохраняются, но преобразование Фурье заменяется преобразованием Лапласа.

При одновременном рассмотрении нескольких случайных величин, заданных на одном и том же вероятностном пространстве, вводится понятие их совместного распределения. В радиоэлектронике и статистической

радиофизике чаще всего используются совместные распределения двух случайных величин, поэтому ограничимся этим случаем.

Совместной функцией распределения $F(x, y)$ двух случайных величин X, Y называется вероятность одновременного выполнения двух неравенств $X < x, Y < y$

$$F(x, y) = P(X < x, Y < y). \quad (1.27)$$

Если $F(x, y)$ непрерывна и обладает непрерывной смешанной производной второго порядка, то совместная (двумерная) *плотность вероятности* $p(x, y)$ случайных величин X, Y удовлетворяет равенствам:

$$p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}; \quad F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(x, y) dx dy \quad (1.28)$$

или $F(x, y) = P(X < x, Y < y) = \sum_{x_i < x} \sum_{y_j < y} p(x_i, y_j)$ – для дискретных величин.

Функции $F(x, y)$ и $p(x, y)$ обладают следующими свойствами:

1) $F(x, y)$ – неубывающая функция своих аргументов;

2) $0 \leq F(x, y) \leq 1; \quad p(x, y) \geq 0;$

3) $F(x, -\infty) = F(-\infty, y) = 0; \quad F(\infty, \infty) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = 1;$

4) $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) dx dy;$

5) одномерные функции выражаются через двумерные по формулам: для непрерывных случайных величин

$$F(x) = F(x, \infty) = \int_{-\infty}^x \int_{-\infty}^{\infty} p(x, y) dx dy, \quad F(y) = F(\infty, y) = \int_{-\infty}^{\infty} \int_{-\infty}^y p(x, y) dx dy;$$

$$p(x) = \frac{dF(x)}{dx} = \int_{-\infty}^{\infty} p(x, y) dy, \quad p(y) = \frac{dF(y)}{dy} = \int_{-\infty}^{\infty} p(x, y) dx$$

для дискретных случайных величин

$$P(x_i) = \sum_j P(x_i, y_j), \quad P(y_j) = \sum_i P(x_i, y_j),$$

где $P(x_i, y_j)$ – совместная вероятность x_i и y_j .

Случайные величины X и Y называются *независимыми*, если

$$F(x, y) = F(x)F(y), \quad p(x, y) = p(x)p(y). \quad (1.30)$$

Когда случайные величины X и Y зависимы, их вероятностные характеристики подчиняются теореме умножения вероятностей: совместная плотность вероятности двух случайных величин $p(x, y)$ равна произведению плотности вероятности одной из них, например $p(x)$, и условной плотности вероятности $p(y|x)$ другой относительно первой:

$$p(x, y) = p(x)p(y|x) = p(y)p(x|y).$$

Из этой теоремы следует, что

$$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\int_{-\infty}^{\infty} p(x, y) dy}; \quad p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(x, y)}{\int_{-\infty}^{\infty} p(x, y) dx}. \quad (1.31)$$

Подставляя $p(x, y)$ в виде произведения плотностей вероятности в (1.31), можно получить формулы Байеса:

$$p(y|x) = \frac{p(y)p(x|y)}{\int_{-\infty}^{\infty} p(y)p(x|y)dy}; \quad p(x|y) = \frac{p(x)p(y|x)}{\int_{-\infty}^{\infty} p(x)p(y|x)dx}, \quad (1.32)$$

или для дискретных случайных величин X и Y

$$P(y_j|x_i) = \frac{P(y_j)P(x_i|y_j)}{\sum_j P(y_j)P(x_i|y_j)}; \quad P(x_i|y_j) = \frac{P(x_i)P(y_j|x_i)}{\sum_i P(x_i)P(y_j|x_i)}.$$

Формула Байеса позволяет предсказать вероятное значение y_j одной из связанных случайных величин по известным значениям x_i другой, если только известны условные вероятности $P(y_j|x_i)$. Условные плотности вероятности обладают всеми свойствами, присущими безусловным вероятностям, в частности $p(y|x) = \frac{dF(y|x)}{dy}$; $p(x|y) = \frac{dF(x|y)}{dx}$.

Числовые характеристики случайных величин. Вместо полного задания распределений вероятности случайных величин часто используется небольшое количество числовых характеристик или *моментов*: математические ожидания и дисперсии случайных величин.

Число m_k , равное математическому ожиданию $\langle X^k \rangle$, называется *начальным моментом* порядка k случайной величины X :

$$m_k(X) = \langle X^k \rangle = \begin{cases} \int_{-\infty}^{\infty} x^k p(x) dx & \text{для непрерывных случайных величин,} \\ \sum_i x_i^k P_i & \text{для дискретных случайных величин.} \end{cases}$$

Числа m_k можно трактовать как моменты инерции соответствующих порядков плоской фигуры, ограниченной кривой $y = p(x)$ и осью абсцисс.

Простейшая числовая характеристика случайной величины – начальный момент первого порядка – *среднее значение* случайной величины⁴:

$$m_1(X) = \bar{X} = \int_{-\infty}^{\infty} xp(x)dx. \quad (1.33)$$

Часто рассматривается также средний квадрат

$$m_2(X) = \overline{X^2} = \int_{-\infty}^{\infty} x^2 p(x) dx.$$

Разность $\tilde{X} = X - \bar{X}$ называется *отклонением* случайной величины от её среднего значения или *центрированной случайной величиной*, а моменты центрированной случайной величины – *центральными моментами*. Центральный момент порядка k нетрудно получить, подставляя в формулу для начального момента \tilde{X} вместо X .

Центральный момент второго порядка называется *дисперсией* случайной величины X

$$\langle (X - \bar{X})^2 \rangle = \int_{-\infty}^{\infty} (x - \bar{X})^2 p(x) dx = \sigma_X^2. \quad (1.34)$$

⁴В дальнейшем числовые характеристики будем рассматривать только для непрерывных случайных величин.

Величина $\sigma_X = +\sqrt{\sigma_X^2}$ называется *среднеквадратическим отклонением* случайной величины X . Если \tilde{X} характеризует отклонение случайной величины относительно \bar{X} , то дисперсия σ_X служит мерой разброса случайной величины относительно \bar{X} .

Центральный и начальный моменты второго порядка связаны соотношением, которое следует из (1.34):

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 p(x) dx - 2\bar{X} \int_{-\infty}^{\infty} xp(x) dx + \bar{X}^2 \int_{-\infty}^{\infty} p(x) dx = \overline{X^2} - \bar{X}^2, \quad (1.35)$$

так как $\int_{-\infty}^{\infty} xp(x) dx = \bar{X}$, а $\int_{-\infty}^{\infty} p(x) dx = 1$.

Для совокупности случайных величин X и Y с плотностью вероятности $p(x, y)$ средние значения случайных величин (начальные моменты первого порядка) согласно (1.29) имеют вид:

$$\bar{X} = \int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xp(x, y) dx dy; \quad \bar{Y} = \int_{-\infty}^{\infty} yp(y) dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yp(x, y) dx dy. \quad (1.36)$$

Аналогично определяются начальные и центральные моменты второго порядка для каждой из случайных величин

$$\begin{aligned} \overline{X^2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 p(x, y) dx dy; & \sigma_X^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})^2 p(x, y) dx dy; \\ \overline{Y^2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 p(x, y) dx dy; & \sigma_Y^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \bar{Y})^2 p(x, y) dx dy. \end{aligned} \quad (1.37)$$

Для совокупности двух случайных величин возможны ещё смешанные начальный и центральный моменты второго порядка

$$K_{XY} = \langle XY \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x, y) dx dy; \quad (1.38)$$

$$\Psi_{XY} = \langle (X - \bar{X})(Y - \bar{Y}) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})(y - \bar{Y}) p(x, y) dx dy = \overline{XY} - \bar{X}\bar{Y}. \quad (1.39)$$

При выводе последнего равенства использовались преобразования, аналогичные уже рассмотренным в (1.35).

Смешанный начальный момент K_{XY} называется *ковариационным моментом* или *ковариацией*, а смешанный центральный момент Ψ_{XY} — *корреляционным моментом* или *корреляцией*, хотя в ряде книг эти названия меняются местами. Если X и Y независимы, то согласно (1.30)

$$\Psi_{XY} = \int_{-\infty}^{\infty} (x - \bar{X})p(x) dx \cdot \int_{-\infty}^{\infty} (y - \bar{Y})p(y) dy =$$

$$= \left[\int_{-\infty}^{\infty} xp(x)dx - \bar{X} \right] \cdot \left[\int_{-\infty}^{\infty} yp(y)dy - \bar{Y} \right] = 0,$$

а Ψ_{XY} может служить мерой связи между X и Y .

В радиоэлектронике и статистической радиофизике более распространена нормированная величина – *коэффициент корреляции* (связи) случайных величин

$$R_{XY} = \frac{\Psi_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}. \quad (1.40)$$

Если случайные величины независимы, то $R_{XY}=0$, т.е. независимые случайные величины *всегда некоррелированы*.

В ряде случаев в радиоэлектронике удобно использовать условные моменты одной случайной величины относительно другой, например переменной X относительно Y . Учитывая соотношения (1.31), для условных среднего значения и дисперсии получим:

$$\langle X|y \rangle = \int_{-\infty}^{\infty} xp(x|y)dx = \frac{\int_{-\infty}^{\infty} xp(x, y)dx}{\int_{-\infty}^{\infty} p(x, y)dx}; \quad (1.41)$$

$$\sigma_{X|y}^2 = \int_{-\infty}^{\infty} [x - \langle X|y \rangle]^2 p(x|y)dx = \frac{\int_{-\infty}^{\infty} [x - \langle X|y \rangle]^2 p(x, y)dx}{\int_{-\infty}^{\infty} p(x, y)dx}. \quad (1.42)$$

В заключение определим среднее значение и дисперсию суммы случайных величин:

$$\begin{aligned} \langle X + Y \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)p(x, y)dxdy = \\ &= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} p(x, y)dy \right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} p(x, y)dx \right] dy = \\ &= \int_{-\infty}^{\infty} xp(x)dx + \int_{-\infty}^{\infty} yp(y)dy = \bar{X} + \bar{Y}, \end{aligned}$$

т.е. среднее значение суммы случайных коррелированных или некоррелированных величин равно сумме средних значений слагаемых

$$\langle (X + Y)^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + 2xy + y^2)p(x, y)dxdy = \bar{X}^2 + 2\bar{X}\bar{Y} + \bar{Y}^2.$$

Дисперсия суммы двух случайных величин имеет вид:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2K_{XY}.$$

Для некоррелированных случайных величин $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$.

1.3.2. Дискретные источники сообщений

Классификация источников сообщений. Источником информации (источником сообщений) является некий объект, который в каждый момент времени случайным образом принимает одно из возможных состояний, т.е. генерирует случайный процесс с дискретным или непрерывным временем. Различают дискретные и непрерывные источники сообщений.

В случае дискретного источника принимаемые им состояния можно пронумеровать в порядке их следования и, таким образом, рассматривать дискретное время. Принятый от такого источника сигнал представляет собой реализацию дискретной случайной последовательности.

Если состояние источника количественно задается некоторой величиной или несколькими величинами, принимающими любые значения в некотором интервале, то источник информации называется непрерывным. Сигналы, создаваемые непрерывным источником, представляют собой реализации непрерывнозначного случайного процесса.

Как отмечалось выше, информацию, заключённую в сообщении, следует отличать от сигнала, несущего сообщение. Сигнал является материальным носителем информации и имеет конкретную физическую природу: электромагнитные и акустические колебания или волны, видеоимпульсы или буквы, напечатанные на бумаге.

Теория информации представляет собой математическую дисциплину, изучающую различные количественные аспекты теории связи. К кругу вопросов теории информации относятся: определение информационных характеристик источников сообщений и каналов связи, проблемы сжатия и кодирования информации, обеспечения секретности передаваемых сообщений, борьба с помехами при передаче информации по каналам связи.

Дать строгое определение понятию «информация» едва ли возможно, так как оно является фундаментальным и не сводится к другим, более простым понятиям. Однако теория информации даёт строгое определение таких информационных параметров, как количество информации в сообщении или скорость передачи информации.

Полученное при испытании (измерении) *количество информации есть мера снятой неопределённости состояния источника*. Неопределённость состояния источника, в свою очередь, определяется *энтропией источника*. Далее рассмотрим такие информационные характеристики, как энтропия источника, количество информации в сообщении, а также совместные характеристики двух дискретных и непрерывных источников.

Информационные характеристики дискретных источников сообщений. *Дискретным источником сообщений X* называется объект, который в каждую единицу времени случайным образом принимает некоторое состояние из множества возможных состояний $X = \{x_i\}_{i=1,2,\dots,M}$. Состояния x_i называются *элементарными сообщениями* или *буквами алфавита источника*. Сообщение дискретного источника X представляет собой случайную последовательность элементарных сообщений:

$$x_{i_1}, x_{i_2}, \dots, x_{i_n}, \dots,$$

где индексы i_ν принимают значения $1, 2, \dots, M$ и соответствуют номерам букв в алфавите источника, а индекс $\nu = 1, 2, \dots$ представляет номер буквы в последовательности.

Дискретный источник X характеризуется безусловными (априорными) вероятностями элементарных сообщений $P(x_i)$, а также вероятностями различных последовательностей элементарных сообщений конечной длины $P(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ или условными вероятностями $P(x_{i_n}/x_{i_{n-1}}, \dots, x_{i_1})$. Если вероятностные характеристики не зависят от

начала нумерации последовательных элементарных сообщений, то источник называется *стационарным*. Далее будем предполагать рассматриваемые источники стационарными.

Безусловной (априорной) энтропией дискретного источника X называется величина $H(X)$, определяемая по известной формуле Шеннона:

$$H(X) = - \sum_i P(x_i) \log P(x_i), \quad (1.43)$$

где от выбора основания логарифма зависят единицы измерения энтропии и информации. Чаще всего рассматривается логарифм по основанию 2, а энтропия источника и количество информации в сообщении измеряются в *битах*. Энтропия $H(X)$, в отличие от термодинамической энтропии Больцмана, называется *информационной* или *энтропией Шеннона*. Она характеризует степень неопределённости состояния источника перед испытанием.

Убыль энтропии источника после испытания определяет среднее количество информации, содержащееся в результате испытания (т.е. в полученном элементарном сообщении):

$$I(X) = H^{pr}(X) - H^{ps}(X), \quad (1.44)$$

где $H^{pr}(X) = H(X)$ – априорная энтропия, характеризующая источник до испытания, а $H^{ps}(X)$ – апостериорная энтропия, характеризующая источник после испытания (после получения элементарного сообщения). Если в результате испытания состояние источника становится точно известным, что всегда имеет место в отсутствии помех, то апостериорная энтропия равна нулю и $I(X) = H(X)$.

Формулу (1.43) можно рассматривать как усреднение по априорному распределению величины $-\log P(x_i)$. Соответственно, в отсутствие помехи эта величина дает количество информации, содержащееся в конкретном элементарном сообщении x_i :

$$I(x_i) = -\log P(x_i). \quad (1.45)$$

Чем менее вероятно сообщение, тем большее количество информации оно содержит.

Для источника с равновероятными состояниями из (1.45) получаем формулу Хартли:

$$I(x_i) = H(X) = \log M, \quad \text{для любого } i. \quad (1.46)$$

Итак, чем больше у источника возможных состояний, тем больше его энтропия и априорное количество информации в результате испытания.

Теорема. На множестве дискретных источников с заданным числом состояний M максимальная априорная энтропия соответствует источнику с равновероятными состояниями (рекомендуется доказать самостоятельно).

Пусть имеются два дискретных источника сообщений: X с состояниями x_i , $i = 1, 2, \dots, M$ и Y с состояниями y_j , $j = 1, 2, \dots, N$. В общем случае эти источники предполагаются статистически зависимыми. Для них,

кроме безусловных энтропий $H(X)$, $H(Y)$ и средних количеств информации $I(X)$, $I(Y)$ в элементарном сообщении, определяемых по формуле Шеннона, можно ввести следующие характеристики:

1) *количество условной информации в данном элементарном сообщении при данном условии*

$$I(x_i|y_j) = -\log P(x_i|y_j) \quad (1.47)$$

– количество информации в сообщении x_i при условии, что от источника Y получено сообщение y_j ;

$$I(y_j|x_i) = -\log P(y_j|x_i) \quad (1.48)$$

– количество информации в сообщении y_j при условии, что от источника X получено сообщение x_i , где $P(x_i|y_j)$, $P(y_j|x_i)$ – условные вероятности;

2) *среднее количество условной информации, усреднённое по совместному распределению $P(x_i, y_j)$*

$$I(X|Y) = -\sum_i \sum_j P(x_i, y_j) \log P(x_i|y_j) \quad (1.49)$$

– в сообщении источника X ;

$$I(Y|X) = -\sum_i \sum_j P(x_i, y_j) \log P(y_j|x_i) \quad (1.50)$$

– в сообщении источника Y ;

3) *условные энтропии источников:*

$H(X|Y) = I(X|Y)$ – энтропия X при заданном Y ,

$H(Y|X) = I(Y|X)$ – энтропия Y при заданном X ;

4) *количество взаимной информации в данных элементарных сообщениях*

$$I(x_i; y_j) = I(x_i) - I(x_i|y_j) \quad (1.51)$$

– количество информации о сообщении x_i , содержащееся в сообщении y_j ;

$$I(y_j; x_i) = I(y_j) - I(y_j|x_i) \quad (1.52)$$

– количество информации о сообщении y_j , содержащееся в сообщении x_i . Легко видеть, что

$$I(x_i; y_j) = I(y_j; x_i) = \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}; \quad (1.53)$$

5) *среднее количество взаимной информации*

$$I(X; Y) = I(Y; X) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}; \quad (1.54)$$

6) *количество информации в данной паре сообщений от источников X и Y*

$$I(x_i y_j) = I(y_j x_i) = -\log P(x_i, y_j); \quad (1.55)$$

7) *среднее количество информации в паре сообщений от двух источников*

$$I(XY) = I(YX) = -\sum_i \sum_j P(x_i, y_j) \log P(x_i, y_j); \quad (1.56)$$

8) совместная энтропия двух источников

$$H(XY) = H(YX) = - \sum_i \sum_j P(x_i, y_j) \log P(x_i, y_j). \quad (1.57)$$

Для определённых выше информационных характеристик двух источников справедливы следующие соотношения:

$$I(X; Y) = I(Y; X) = H(X) - H(X|Y) = H(Y) - H(Y|X); \quad (1.58)$$

$$\begin{aligned} H(XY) &= H(X) + H(X|Y) = H(Y) + H(Y|X) = \\ &= H(X) + H(Y) - I(X; Y); \end{aligned} \quad (1.59)$$

$$H(XY) \leq H(X) + H(Y). \quad (1.60)$$

Количество информации в элементарном сообщении дискретного источника при наличии помех. Если имеются помехи, то полу-

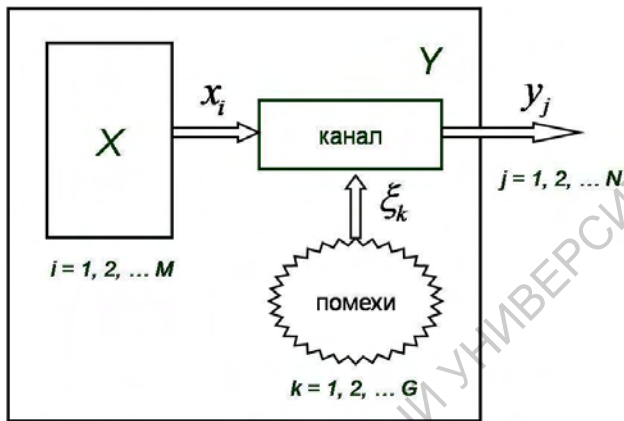


Рис. 1.5. Действие помехи на сообщения дискретного источника

Условная информация $I(x_i|y_j)$ определяет количество информации «уничтоженное» помехой ξ_k . В среднем полученное элементарное сообщение содержит количество информации $I_{\text{ш}}(X)$, равное средней взаимной информации источников X и Y :

$$I_{\text{ш}}(X) = I(X; Y) = H(X) - H(X|Y). \quad (1.61)$$

Это и есть среднее количество информации в элементарном сообщении от источника X при наличии помех:

$$I_{\text{ш}}(X) = H^{pr}(X) - H^{ps}(X),$$

где $H^{pr}(X) = H(X)$, $H^{ps}(X) = H(X|Y)$. Таким образом, если шум преобразует источник X в источник Y , то условная энтропия $H(X|Y)$ определяет среднее количество информации, уничтожаемое помехами, и является характеристикой ненадёжности канала. Статистическая независимость источников X и Y означает, что помехи уничтожили информацию полностью: $H(X|Y) = H(X)$, следовательно, $I_{\text{ш}}(X) = 0$. Если состояния источников X и Y связаны детерминированным образом, то это равнозначно отсутствию помех: $H(X|Y) = 0$, следовательно, $I_{\text{ш}}(X) = H(X) = I(X)$.

чател ь вместо множества элементарных сообщений x_i , $i = 1, 2, \dots, M$ дискретного источника X получит множество сообщений y_j , $j = 1, 2, \dots, N$, т.е. получит сообщения как бы от другого источника Y (рис. 1.5).

Сообщение y_j несет следующее количество информации о посланном сообщении x_i :

$$I(x_i; y_j) = I(x_i) - I(x_i|y_j).$$

Количество информации в последовательности элементарных сообщений. Избыточность дискретного источника с памятью. Источник X называется *источником без памяти, простым источником* или *источником Бернулли*, если его состояние x_{i_n} в любой момент времени n не зависит от предыдущих состояний источника. Источник Бернулли полностью характеризуется безусловными вероятностями элементарных сообщений $P(x_i)$. Источник X называется *источником с памятью* или *сложным источником*, если его состояние x_{i_n} в момент n зависит от предыдущих состояний x_{i_ν} в моменты $\nu = n-1, n-2, n-3, \dots$. Чтобы задать источник с памятью, нужно знать не только безусловные вероятности $P(x_i)$, но и условные вероятности или вероятности различных последовательностей элементарных сообщений.

Пусть X – дискретный источник с состояниями $x_i, i = 1, 2, \dots, M$. Рассмотрим n -буквенное сообщение (т.е. блок из n элементарных сообщений)

$$(x_{i_1}, x_{i_2}, \dots, x_{i_k}, \dots, x_{i_n}), \quad x_{i_k} = 1, \dots, M.$$

Количество информации в данном блоке можно определить как

$$I(x_{i_1}, x_{i_2}, \dots, x_{i_n}) = -\log P(x_{i_1}, x_{i_2}, \dots, x_{i_n}),$$

где $P(x_{i_1}, x_{i_2}, \dots, x_{i_n})$ – вероятность данного блока из n элементарных сообщений. Среднее количество информации в блоке длиной n есть

$$I(X_1, X_2, \dots, X_n) = -\sum_{i_1} \sum_{i_2} \dots \sum_{i_n} P(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \log P(x_{i_1}, x_{i_2}, \dots, x_{i_n}).$$

Эта же величина есть энтропия блока сообщений длиной n :

$$I(X_1, X_2, \dots, X_n) = H(X_1, X_2, \dots, X_n).$$

Среднее количество информации в расчёте на одно элементарное сообщение (энтропия на букву) с учетом статистики блоков длиной n есть

$$I_n(X) = \frac{1}{n} H(X_1, X_2, \dots, X_n).$$

Если элементарные сообщения в последовательности независимы, то $H(X_1, X_2, \dots, X_n) = nH(X)$. Таким образом, $I_n(X) = I(X)$, т.е. получаем среднее количество информации в элементарном сообщении по безусловному распределению. В общем случае, когда в информационной последовательности присутствуют корреляции (статистические взаимосвязи) между элементарными сообщениями, получаем $I_n(X) \leq I(X)$.

Можно увеличивать длину рассматриваемых блоков и в пределе при $n \rightarrow \infty$ получить истинное среднее количество информации в элементарном сообщении (с учетом всех корреляций) или же *истинную энтропию источника*:

$$\begin{aligned} I_{\text{и}}(X) &= H_{\text{и}}(X) = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n) = \\ &= -\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i_1} \dots \sum_{i_n} P(x_{i_1}, x_{i_2}, \dots, x_{i_n}) \log P(x_{i_1}, x_{i_2}, \dots, x_{i_n}). \end{aligned} \quad (1.62)$$

Очевидно, что $H_{\text{и}}(X) \leq H(X) \leq H(X_0)$, где $H(X_0) = \log M$ – энтропия источника с M равновероятными состояниями. Источник X , для которого $H_{\text{и}}(X) = H(X_0)$, называется *безызбыточным*.

Чем больше взаимосвязь элементарных сообщений в последовательности, тем меньше истинная энтропия по сравнению с $H(X_0)$ и больше избыточность источника. Для того чтобы охарактеризовать, насколько источник X отличается от безызбыточного, вводят *коэффициент избыточности* источника:

$$\rho(X) = 1 - \frac{H_{\text{и}}(X)}{H(X_0)}. \quad (1.63)$$

Очевидно, $\rho(X) \in [0; 1]$ (или $\rho(X) \in [0; 100\%]$). Значение $\rho(X) = 0$ соответствует безызбыточному источнику, для которого среднее количество информации в элементарном сообщении максимально. Если $\rho(X) = 1$, это означает, что среднее количество информации в элементарном сообщении равно нулю.

Таким образом, избыточность источника в общем случае определяется двумя причинами: во-первых, неравенством безусловных вероятностей различных элементарных сообщений, во-вторых, наличием статистической зависимости между элементарными сообщениями в последовательности. Безызбыточный источник – это простой источник (без памяти) с равновероятными состояниями. Для него количество информации в элементарном сообщении (букве) максимально на множестве всех источников с заданным размером алфавита.

1.3.3. Непрерывные источники сообщений

Информационные характеристики непрерывных источников сообщений. *Непрерывный источник сообщений* X – это объект, который в каждую единицу времени случайным образом принимает некоторое состояние из непрерывного множества возможных состояний. Обычно состояние непрерывного источника в каждый момент времени задается непрерывной случайной величиной X , принимающей значения в некотором интервале $[a, b]$ (он может быть и бесконечным). Изменение состояния во времени есть случайный процесс, описываемый случайной функцией $X(t)$.

Непрерывный источник X задан, если для любой совокупности моментов времени $t_i, i = 1, 2, \dots, n$ известна n -мерная плотность вероятности $p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n)$ ⁵. Однако для определения основных информационных характеристик достаточно знать одномерное распределение $p_1(x, t)$. Далее всюду будем считать, что источник X – стационарный, т.е. его статистические характеристики не зависят от выбора начала отсчёта времени. Соответственно, одномерная плотность вероятности не зависит от времени: $p_1(x, t) = p(x)$ (индекс «1» для краткости опустим).

Безусловная энтропия непрерывного источника X не может задаваться формулой Шеннона (1.43), введённой для дискретного источника.

⁵Многомерная плотность вероятности и другие характеристики случайного процесса будут рассматриваться позже (см. гл. 3).

Чтобы определить энтропию непрерывного источника, произведем дискретизацию (квантование) множества состояний источника X с шагом Δx (рис. 1.6). Энтропию квантованного источника $H_{\Delta x}(X)$ получим по формуле (1.43), а затем произведем предельный переход $\Delta x \rightarrow 0$:



Рис. 1.6. Квантование множества состояний непрерывного источника X

$$H(X) = \lim_{\Delta x \rightarrow 0} H_{\Delta x}(X).$$

Считая шаг квантования малым, для вероятности i -го состояния x_i имеем $P(x_i) \approx p(x_i)\Delta x$. Тогда

$$\begin{aligned} H(X) &= \lim_{\Delta x \rightarrow 0} \left[- \sum_i p(x_i)\Delta x \log[p(x_i)\Delta x] \right] = \\ &= - \int_a^b p(x) \log p(x) dx - \int_a^b p(x) dx \lim_{\Delta x \rightarrow 0} \log \Delta x. \end{aligned}$$

С учетом нормировки

$$\int_a^b p(x) dx = 1$$

получим

$$H(X) = - \int_a^b p(x) \log p(x) dx - \lim_{\Delta x \rightarrow 0} \log \Delta x. \quad (1.64)$$

Первое слагаемое в (1.64) является ограниченным и называется *безусловной (априорной) дифференциальной энтропией непрерывного источника X* . Второе слагаемое с учетом отрицательного знака дает в пределе $+\infty$. Таким образом, полная безусловная энтропия непрерывного источника бесконечна: $H(X) = +\infty$.

Назовем

$$h(X) = - \int_a^b p(x) \log p(x) dx. \quad (1.65)$$

безусловной дифференциальной энтропией. Она характеризует распределение случайной величины X и в отличие от безусловной энтропии дискретного источника может принимать как положительные, так и отрицательные значения, а также обращаться в 0. Важно отметить, что безусловная дифференциальная энтропия $h(X)$ сама по себе не определяет количества информации в мгновенном состоянии непрерывного источника.

Теорема. Безусловная дифференциальная энтропия непрерывного источника X с заданной дисперсией σ_X^2 удовлетворяет неравенству

$$h(X) \leq \log \sqrt{2\pi e \sigma_X^2},$$

где e – основание натурального логарифма. Максимальное значение энтропия принимает в случае источника X_0 с гауссовским распределением⁶:

$$h(X_0) = \log \sqrt{2\pi e \sigma_X^2}. \quad (1.66)$$

⁶Теорему рекомендуется доказать самостоятельно

Пусть X и Y – два непрерывных (в общем случае статистически зависимых) источника, мгновенные состояния которых в произвольный момент времени t задаются, соответственно, случайными величинами $X(t)$, $x \in [a, b]$ и $Y(t)$, $y \in [c, d]$. Для них, кроме безусловных энтропий $h(X)$, $h(Y)$, задаваемых формулой (1.65), можно ввести следующие характеристики:

1) *условные дифференциальные энтропии источников*

$$h(X|Y) = - \int_a^b \int_c^d p(x, y) \log p(x|y) dx dy, \quad (1.67)$$

$$h(Y|X) = - \int_a^b \int_c^d p(x, y) \log p(y|x) dx dy, \quad (1.68)$$

где $p(x, y)$ – совместная плотность вероятности случайных величин X и Y ; $p(x|y)$, $p(y|x)$ – условные плотности вероятности;

2) *совместную дифференциальную энтропию двух источников*

$$h(XY) = - \int_a^b \int_c^d p(x, y) \log p(x, y) dx dy; \quad (1.69)$$

3) *среднее количество взаимной информации на один отсчёт* $I(X; Y) = I(Y; X)$. Также как для дискретных источников среднее количество взаимной информации двух непрерывных источников, приходящееся на один отсчёт (в один момент времени), определяется как $I(X; Y) = H(X) - H(X|Y)$ или $I(X; Y) = H(Y) - H(Y|X)$. Полная безусловная энтропия $H(X)$ источника X определяется по формуле (1.64). Полную условную энтропию $H(X|Y)$ аналогичным образом можно определить как

$$H(X|Y) = \lim_{\Delta x \rightarrow 0} \lim_{\Delta y \rightarrow 0} H_{\Delta x \Delta y}(X|Y) = h(XY) - \lim_{\Delta x \rightarrow 0} \log \Delta x. \quad (1.70)$$

Тогда

$$I(X; Y) = h(X) - h(X|Y) = h(Y) - h(Y|X). \quad (1.71)$$

Среднее количество взаимной информации двух непрерывных источников определяется разностью безусловной и условной дифференциальных энтропий и является ограниченной величиной.

С учетом равенства $p(x, y) = p(x|y)p(y)$ из (1.69) следует, что

$$I(X; Y) = I(Y; X) = \int_a^b \int_c^d p(x, y) \frac{\log p(x, y)}{p(x)p(y)} dx dy \quad (1.72)$$

или

$$I(X; Y) = I(Y; X) = \int_a^b \int_c^d p(x, y) I(x; y) dx dy,$$

где

$$I(x; y) = I(y; x) = \frac{\log p(x, y)}{p(x)p(y)}$$

– количество взаимной информации в конкретных значениях x и y .

Для определённых выше информационных характеристик двух непрерывных источников справедливы следующие соотношения:

$$\begin{aligned} h(XY) &= h(X) + h(X|Y) = h(Y) + h(Y|X) = \\ &= h(X) + h(Y) - I(X; Y); \end{aligned} \quad (1.73)$$

$$h(XY) \leq h(X) + h(Y). \quad (1.74)$$

Сумма в (1.74) соответствует независимым X и Y .

Количество информации в сообщении непрерывного источника при наличие помех. Если имеются помехи, то вместо случайного сигнала $X(t)$ от источника X получатель принимает случайный сигнал $Y(t) = X(t) + \xi(t)$. Среднее количество информации на один отсчёт от непрерывного источника в присутствии помех есть

$$I_{\text{ш}}(X) = I(X; Y) = h(X) - h(X|Y) = h(Y) - h(Y|X). \quad (1.75)$$

Если источники X и Ξ статистически независимы, то $h(Y|X) = h(\xi)$ – дифференциальная энтропия помехи.

Информационные характеристики дискретных и непрерывных источников сообщений составляют базис для определения параметров (различных) каналов передачи информации, в частности для оценки их пропускной способности. Оказывается, скорость передачи информации в большей степени зависит не от свойств канала, а от количества взаимной информации, которую в среднем содержит принятый сигнал.

1.4. Контрольные вопросы и упражнения

1. Чем отличается сигнал от помехи?
2. В чем различие детерминированных и случайных сигналов?
3. Как подразделяют сигналы на аналоговые, дискретные и цифровые?
4. Что представляет собой метрическое пространство сигналов? Какими свойствами должна обладать метрика?
5. Вычислите норму сигнала $x(t) = A \sin \pi t/T$ при $0 < t < T$.
6. Какие линейные пространства называются гильбертовыми, а какие – евклидовыми?
7. Используя неравенство Коши – Буняковского (1.9), покажите, что метрика $d(x_1(t), x_2(t)) = \|x_1(t) - x_2(t)\|$, где $\|x(t)\| = (x(t), x(t))^{1/2}$, удовлетворяет аксиомам метрики.
8. Какие сигналы являются ортогональными? Что называется ортонормальным базисом гильбертова пространства?
9. Покажите, что векторы вещественного пространства $x + y$ и $x - y$ ортогональны, если $\|x\| = \|y\| = 1$.
10. Как связаны между собой функции Радемахера и функции Уолша?
11. Покажите, что расстояние между функциями $wal_0(t)$ и $wal_1(t)$ совпадает с расстоянием между $wal_1(t)$ и $wal_2(t)$.
12. Какими свойствами обладает плотность вероятности случайной величины? В чем заключается смысл теоремы Байеса?
13. Что такое среднее квадратичное отклонение случайной величины?
14. На пяти пластинах (по одной букве на каждой) написано «СЛОВО».

Пластины перемешаны. Используя условные вероятности показать, что вероятность получить слово «ВОЛ» при последовательном извлечении трех пластин составляет $1/30$.

15. Как определить понятия: «количество информации», «информационное сообщение», «сигнал»? Какова связь между ними?

16. Докажите, что безусловная энтропия дискретного источника с фиксированным числом состояний максимальна в случае равновероятных состояний.

17. Как выражается совместная энтропия двух дискретных источников X и Y через безусловные и условные энтропии?

18. Какое количество информации содержится в сообщении о том, что две подброшенные монеты упали на орла?

19. Чему равно среднее количество информации в сообщении от дискретного источника X в присутствии помех, если в результате действия помех посылаемые и получаемые сообщения оказываются статистически независимыми?

20. Какова избыточность источника X с четырьмя состояниями, принимаемыми с вероятностями $1/2$, $1/4$, $1/8$, $1/8$? Последовательные состояния считать независимыми.

21. Для какого распределения с заданной дисперсией безусловная дифференциальная энтропия $h(X)$ максимальна и чему она равна?

22. Вычислить дифференциальную энтропию $h(X)$ для случая равномерного распределения X в интервале $[0; 1]$.

Список литературы

Колмогоров, А. Н., Фомин, С. В. Элементы теории функций и функционального анализа. М. : Наука, 1976.

Френкс, Л. Теория сигналов / пер. с англ. М. : Сов. радио, 1974.

Рыжак, И. С. Избранные вопросы теории цепей и обработки сигналов. М. : Блок-информ-экспресс, 2003.

Зиновьев, А. Л., Филиппов, Л. И. Введение в теорию сигналов и цепей : учеб. пособие. М. : Высш. шк., 1968.

Гнеденко, Б. В. Курс теории вероятностей : учебник. 8-е изд., испр. и доп. М. : Эдиториал УРСС, 2005.

Шахтарин, Б. И. Случайные процессы в радиотехнике : в 2 т. / 3-е изд., перераб. Т.1: Линейные преобразования. М. : Гелиос АЗВ, 2006.

Хартмут, Х. Ф. Передача информации ортогональными функциями / пер. с англ. М. : Связь, 1975.

Сиберт, У. М. Цепи, сигналы, системы : в 2 ч. / пер. с англ. М. : Мир, 1988.

Галлогер, Р. Теория информации и надёжная связь. М. : Сов. радио, 1974.

Игнатов, В. А. Теория информации и передача сигналов. М. : Сов. радио, 1979.

Теория информации и кодирование / Б. Б. Самсонов, Е. М. Плохов, А. И. Филоненков, Т. В. Кречет. М. : Феникс, 2002.

Лидовский, В. В. Теория информации : учеб. пособие. М. : Компания Спутник, 2004.

Глава 2

Детерминированные колебания и волны. Методы описания

2.1. Виды детерминированных сигналов

Как отмечалось выше, сигналы, несущие информацию, являются случайными, а детерминированные колебания используются в радиоэлектронике в качестве несущих, испытательных или управляющих.

Детерминированные сигналы делятся на периодические, квазипериодические и непериодические. Колебания, мгновенные значения которых повторяются через равные промежутки времени, называются *периодическими* (рис. 2.1), а наименьший промежуток времени T , по истечении которого все мгновенные значения повторяются, – *периодом* колебаний. Тогда

$$x(t + T) \equiv x(t), \quad -\infty < t < \infty. \quad (2.1)$$

Простейшим видом периодических сигналов являются гармонические колебания (напряжения, тока, электрического поля \mathbf{E}).

Гармонические колебания. *Гармоническим* называется периодическое изменение физической величины (напряжения, заряда или тока) по синусоидальному (или косинусоидальному) закону (рис. 2.2):

$$x(t) = A \cdot \cos(\omega t + \varphi_0), \quad -\infty < t < \infty \quad (2.2)$$

где $x(t)$ – мгновенное значение физической величины; A – амплитуда; $\varphi(t) = (\omega t + \varphi_0)$ – мгновенная (полная) фаза колебаний в момент времени t ; ω – круговая или угловая частота, выражаемая в радианах в секунду; φ_0 – начальная фаза колебаний. Продолжительность одного полного колебания $T = 2\pi/\omega$ – период гармонического колебания выражается в секундах или долях секунды, а величина $f = 1/T$, равная числу полных колебаний в единицу времени, называется *частотой гармонического колебания* и выражается в герцах (Гц). Гармонические колебания являются идеализацией уже потому, что реальные

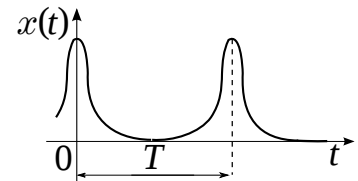


Рис. 2.1. Периодическое колебание

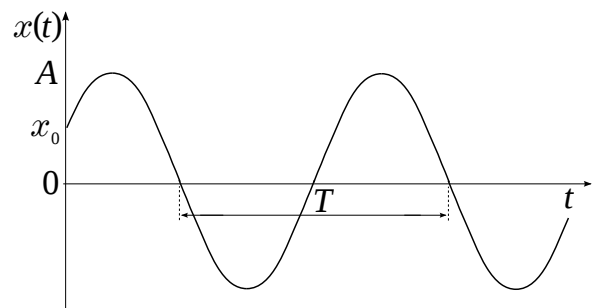


Рис. 2.2. Гармоническое колебание

колебания ограничены во времени, т.е. имеют начало и конец.

Формула (2.2) представляет *вещественную модель* таких сигналов. Первая, вторая и последующие производные по времени функции $x(t)$ также представляют гармонические функции:

$$\frac{dx(t)}{dt} = -A\omega \sin(\omega t + \varphi_0), \quad \frac{d^2x(t)}{dt^2} = -A\omega^2 \cos(\omega t + \varphi_0), \dots$$

Подставляя (2.2) в последнее соотношение, легко получить для гармонического сигнала $x(t)$ уравнение

$$\frac{d^2x(t)}{dt^2} + \omega^2 x(t) = 0, \quad (2.3)$$

называемое *дифференциальным уравнением гармонических колебаний*.

Гармонические колебания обладают одной определённой и строго постоянной частотой, т.е. являются *однотональными*¹. Среднее за период значение гармонической функции равно нулю:

$$\overline{x(t)} = \frac{1}{T} \int_0^T x(t) dt = \frac{A}{T} \int_0^T \cos(\omega t + \varphi_0) dt = \frac{A}{\omega T} (\sin(\omega t + \varphi_0)) \Big|_0^T = 0. \quad (2.4)$$

Мгновенная мощность $p(t)$, выделяемая гармоническим сигналом $x(t) = A \cos(\omega t + \varphi_0)$ в резисторе с сопротивлением 1 Ом, имеет вид:

$$p(t) = (A \cos(\omega t + \varphi_0))^2 = \frac{A^2}{2} + \frac{A^2}{2} \cos(2\omega t + 2\varphi_0)$$

и представляет сумму постоянной составляющей $A^2/2$ и косинусоидального колебания с частотой 2ω . Величину $p(t)$ нельзя считать энергетическим параметром рассматриваемого сигнала, так как она периодически изменяется от 0 до A^2 . Поэтому в качестве *квадратичных* (энергетических) параметров гармонических сигналов используются либо мощность, усреднённая за период колебаний, – *средний квадрат гармонической функции*

$$P_{\text{ср}} = \overline{x^2(t)} = \frac{1}{T} \int_0^T x^2(t) dt = \frac{A^2}{T} \int_0^T \cos^2(\omega t + \varphi_0) dt = \frac{A^2}{2}, \quad (2.5)$$

либо *среднеквадратичная величина* (СКВ), равная положительному значению квадратного корня из $\overline{x^2(t)}$, т.е. СКВ = $A/\sqrt{2}$.

Комплексное представление гармонических колебаний. Метод комплексных амплитуд. В радиоэлектронике и статистической физике часто используется символическое представление вещественных гармонических функций времени комплексными числами, получившее название *символического метода* или *метода комплексных амплитуд*².

Дополняя вещественный сигнал $x(t) = A \cos(\omega t + \varphi_0)$ мнимой компонентой $j \sin(\omega t + \varphi_0)$, получим комплексный сигнал:

$$\dot{x}(t) = A \exp[j(\omega t + \varphi_0)] = A[\cos(\omega t + \varphi_0) + j \sin(\omega t + \varphi_0)] = x(t) + y(t).$$

¹Гармонические колебания иногда называют *монохроматическими* (от греч. *μόνος* – один и *χρόμα* (*chrómatos*) – цвет), хотя этот термин введён для электромагнитного излучения.

²Метод комплексных амплитуд ввел в теорию переменных токов немецкий физик Ч. Штейнмец, а широко использовал академик В. Ф. Миткевич (Россия).

Если ввести $\dot{A} = Ae^{j\varphi_0}$ как комплексную амплитуду колебания, то действительная часть комплексной функции $\dot{x}(t)$ совпадает с исходным сигналом $x(t)$, а сигнал $y(t)$, обычно называемый сопряжённым, представляет собой $x(t)$, сдвинутый по фазе на $-\pi/2$.

Таким образом, вещественное гармоническое колебание (2.2) представлено комплексной моделью в виде числа \dot{A} . Каждая \dot{A} вводится для вполне определённого колебания, т.е. является функцией частоты.

Обратный переход от комплексной модели к вещественной осуществляется путём умножения комплексной амплитуды колебания на $\exp(j\omega t)$ и вычисления вещественной части полученного выражения:

$$x(t) = \operatorname{Re}[\dot{A}e^{j\omega t}] = \operatorname{Re}[Ae^{j(\omega t + \varphi_0)}] = A \cos(\omega t + \varphi_0). \quad (2.6)$$

Отметим основные свойства комплексных амплитуд, важные для практического использования:

1) комплексная амплитуда суммы гармонических колебаний одинаковой частоты равна сумме комплексных амплитуд слагаемых;

2) при умножении на комплексное число $Se^{j\alpha}$ модуль комплексной амплитуды увеличивается в S раз, а аргументы сомножителей суммируются;

3) произведение комплексно-сопряжённых амплитуд $\dot{A}\dot{A}^* = A^2$, где $*$ – знак комплексного сопряжения, совпадает с удвоенной средней за период мощностью:

$$P_{\text{cp}} = \frac{1}{2}\dot{A}\dot{A}^* = \frac{A^2}{2},$$

а половина произведения сдвинутых по фазе сигналов, например, напряжения $U \cos(\omega t + \alpha_1)$ и тока $I \cos(\omega t + \alpha_2)$ соответствует средней за период комплексной мощности

$$P_{\text{cp}} = \frac{1}{2}\dot{U}\dot{I}^* = \frac{1}{2}UIe^{j(\alpha_1 - \alpha_2)} = \frac{UI}{2} \cos(\alpha_1 - \alpha_2) + j\frac{UI}{2} \sin(\alpha_1 - \alpha_2)$$

и состоит из резистивной $P_{\text{cp}R} = \operatorname{Re}[P_{\text{cp}}] = \frac{1}{2}UI \cos(\alpha_1 - \alpha_2)$ и реактивной $P_{\text{cp}X} = \operatorname{Im}[P_{\text{cp}}] = \frac{1}{2}UI \sin(\alpha_1 - \alpha_2)$ составляющих, представляющих активную и реактивную мощности соответственно.

Поскольку дифференцирование (интегрирование) комплексной экспоненты по времени эквивалентно умножению (делению) на $j\omega$, комплексная амплитуда временной производной $x(t) = A \cos(\omega t + \varphi_0)$ равна $j\omega\dot{A}$ (при k -кратном дифференцировании – $(j\omega)^k\dot{A}$), а комплексная амплитуда интеграла – $\frac{1}{j\omega}\dot{A}$, т.е. линейные дифференциальные уравнения с вещественными коэффициентами легко приводятся к эквивалентным алгебраическим уравнениям относительно $j\omega$.

Суперпозиции двух или нескольких гармонических колебаний различных частот образуют более сложные колебания: периодические и квазипериодические сигналы

$$x(t) = f(u(t), v(t)), \quad \text{где } u(t + T_1) \equiv u(t), \quad v(t + T_2) \equiv v(t).$$

В общем случае и периодические, и квазипериодические сигналы удовлетворяют одному и тому же соотношению:

$$x(t) = f(\omega_1 t, \omega_2 t, \omega_3 t, \dots, \omega_N t). \quad (2.7)$$

У периодических колебаний отношение любых частот гармонических составляющих является *рациональным числом*: $\omega_i/\omega_j = m/n$, где m и n – целые числа. Только в этом случае существует отрезок времени (период), на котором периоды составляющих колебаний укладываются целое число раз. Так, детерминированный сигнал

$$x(t) = A_1 \sin \frac{2\pi}{T}t + A_2 \sin \frac{4\pi}{T}t + \sin \frac{10\pi}{T}t \quad (2.8)$$

содержит три гармонических колебания. Отношения частот $1/2$, $1/5$ и $2/5$ – рациональные числа. Частота первой гармоники 2π , а период $T=1$. У квазипериодических колебаний отношение хотя бы одной пары частот является *иррациональным числом*. Примером квазипериодического детерминированного сигнала может служить следующее выражение:

$$x(t) = A_1 \sin \frac{2\pi}{T}t + A_2 \sin \frac{2\pi\sqrt{3}}{T}t + \sin \frac{10\pi}{T}t. \quad (2.9)$$

Различие в структуре сигналов, рассчитанных по формулам (2.8) и (2.9), иллюстрирует рис. 2.3.

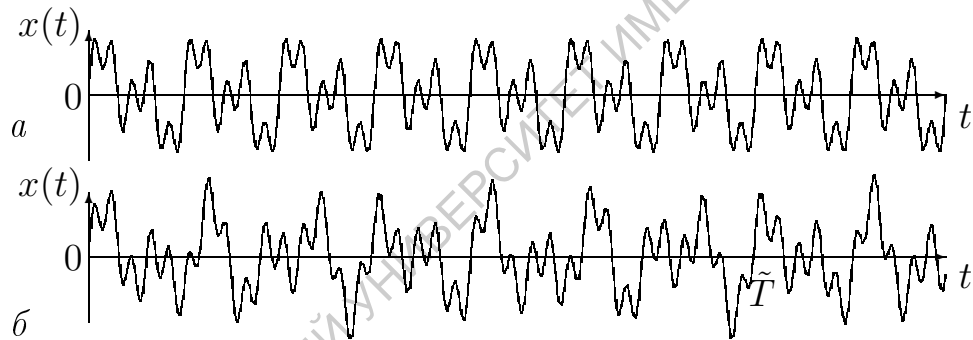


Рис. 2.3. Периодическое (а) и квазипериодическое (б) колебания

Видно, что колебание, изображённое на рис. 2.3, а, является периодическим, а колебание на рис. 2.3, б воспринимается как непериодическое.

Периодические и квазипериодические сигналы по определению существуют сколь угодно долго и поэтому характеризуют стационарные (установившиеся) состояния радиосистемы. Переходы из одного стационарного состояния в другое описываются непериодическими (переходными) сигналами. Среди непериодических функций в радиоэлектронике особая роль отводится импульсным сигналам: экспоненциальным, прямоугольным, косинусоидальным, пилообразным. В статистической радиофизике часто используются колоколообразные (гауссовы) импульсы.

Импульсные сигналы. *Электрическим импульсом* (от лат. *impulsus*

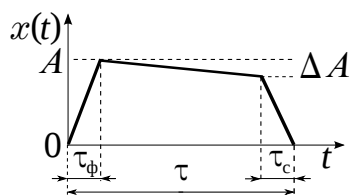


Рис. 2.4. Основные характеристики видеоимпульса

– удар, толчок) называется сигнал, существующий в течение конечного (как правило, достаточно короткого) интервала времени τ , называемого *длительностью*, и обладающий нулевой или пренебрежимо малой амплитудой за его пределами. Импульсы могут быть одиночными или образовывать последовательности (периодические или непериодические). Основными характеристиками импульса (рис. 2.4) являются его ам-

плитуда A (наибольшие значения переменной), длительность τ , длительности фронта (резкого нарастания физической величины) и среза (τ_f и τ_c), а также спад вершины ΔA .

Импульсные процессы чрезвычайно широко распространены в природе. Импульсный характер имеют многие сейсмические явления, волновые процессы в нелинейных и активных средах. Совокупностью импульсов различной амплитуды и длительности является электромагнитное излучение заряженных частиц. В радиоэлектронике различают два основных класса импульсных процессов: видеоимпульсы и радиоимпульсы.

Рассмотренные выше импульсные сигналы получили название *видеоимпульсов* (рис. 2.5, $a - d$), так как впервые они стали широко использоваться в телевидении.

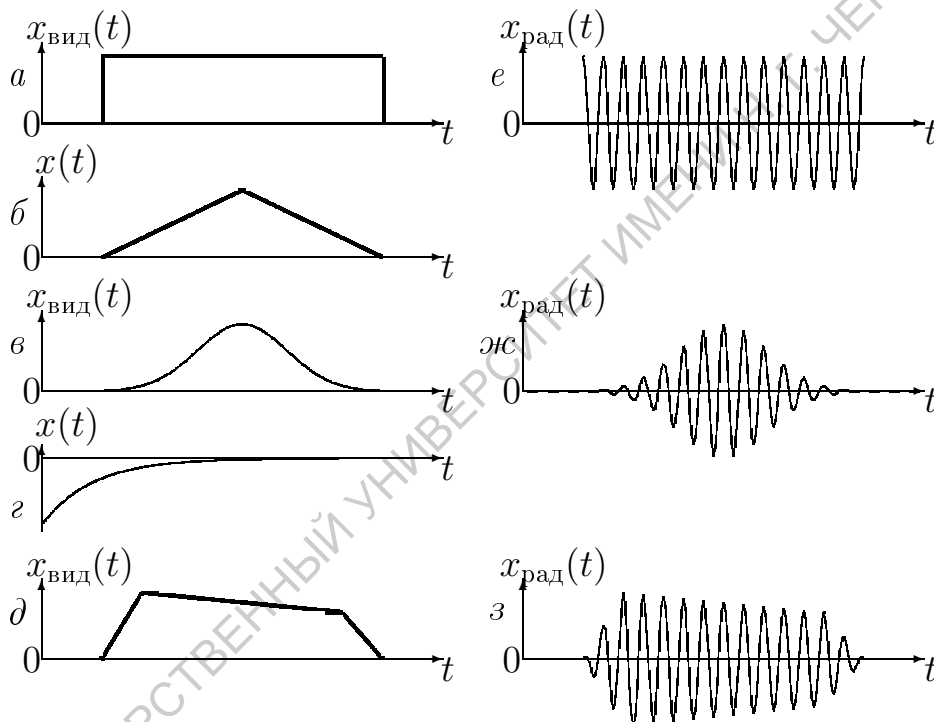


Рис. 2.5. Видеоимпульсы ($a - d$) и радиоимпульсы ($e - з$) различной формы

Радиоимпульсами называются высокочастотные или сверхвысокочастотные (СВЧ) колебания, огибающие которых изменяются по закону видеоимпульсов (рис. 2.5, $e - з$), т.е. для радиоимпульсов $x_{\text{рад}}(t)$ справедлива следующая математическая модель:

$$x_{\text{рад}}(t) = x_{\text{вид}}(t) \cdot \cos \omega_0 t, \quad (2.10)$$

где $x_{\text{вид}}(t)$ – функция, описывающая видеоимпульс; ω_0 – круговая частота несущего высокочастотного колебания. Различают видео- и радиоимпульсы прямоугольной (рис. 2.5, a, e), колоколообразной (рис. 2.5, $в, ж$), треугольной (рис. 2.5, $б$), трапецидальной (рис. 2.5, $д, з$), экспоненциальной (рис. 2.5, $з$) и других форм.

В общем случае сигналы обладают достаточно сложной структурой, и не всегда удается указать соответствующую аппроксимирующую функцию. Поэтому сложный сигнал либо представляется в виде наложения элементарных сигналов, задержанных во времени, либо в виде суперпо-

зиции ортонормальных сигналов, действующих одновременно. Первый подход называется *временным* или *динамическим представлением* сигналов, второй – *частотным* или *спектральным представлением*.

Временное (динамическое) представление отражает естественное изменение сигналов во времени, причём разность значений сигнала, разделённых некоторым интервалом времени, задается элементарным сигналом ступенчатой или импульсной формы. Если временной интервал между отсчётами устремить к нулю, то динамическое представление принимает интегральную форму и обеспечивает точное описание сигналов.

2.2. Динамическое представление сигналов

Элементарные и простые детерминированные сигналы. Важнейшая роль в теоретической радиоэлектронике отводится элементарным сигналам – функциям включения постоянного и кратковременного импульсного воздействий, математическими моделями которых являются идеализированные функции Хевисайда и Дирака.

Функция Хевисайда описывает мгновенный переход из «нулевого» в «единичное» состояние (рис. 2.6, а) и обычно задается соотношением

$$\sigma(t - t_0) = \begin{cases} 0, & t < t_0, \\ 1/2, & t = t_0, \\ 1, & t > t_0. \end{cases} \quad (2.11)$$

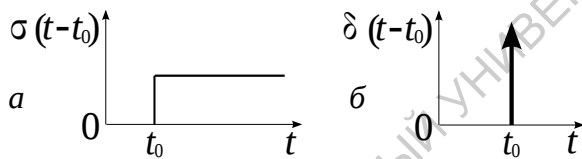


Рис. 2.6. Функция включения Хевисайда (а) и дельта-функция Дирака (б)

При этом предполагается, что $\sigma(t)$ задана для всех $-\infty < t < \infty$. На практике это условие не выполняется, и функция Хевисайда является идеализированной моделью реальных функций включения. Наряду с приведённым определением для теоретического анализа оказываются полезными различные асимптотические аппроксимации $\sigma(t)$, например

$$\sigma(t) = \lim_{\alpha \rightarrow 0} e^{-\alpha t}, \quad t \geq 0. \quad (2.12)$$

Дельта-функция, введённая П. Дираком в 1926 г., не является функцией в обычном смысле и относится к разряду так называемых «обобщенных» функций.

Дельта-функцией Дирака принято называть сингулярную импульсную функцию $\delta(t)$, имеющую единичную площадь, принимающую бесконечно большое значение в точке $t = t_0$ и равную нулю во всех остальных точках:

$$\delta(t - t_0) = \begin{cases} \infty, & t = t_0, \\ 0, & t \neq t_0, \end{cases} \quad \int_{-\infty}^{\infty} \delta(t) dt = 1. \quad (2.13)$$

Символическое изображение дельта-функции представлено на рис. 2.6, б. Функцию $\delta(t)$ легко ввести как предел сжатия прямоугольного импульса, имеющего длительность Δ и высоту $1/\Delta$. Дельта-функция обладает *фильтрующим свойством*, согласно которому

$$\int_{-\infty}^{\infty} x(t) \delta(t - t_0) dt = x(t_0). \quad (2.14)$$

Можно показать, что дельта-функция есть производная функции Хевисайда, а функция Хевисайда – интеграл от функции Дирака.

Важнейшее свойство элементарных функций заключается в том, что их нельзя представить наложением каких-либо других функций, но любой аналоговый сигнал разлагается на элементарные сигналы, возникающие в последовательные моменты времени.

Наряду с элементарными в радиоэлектронике и статистической радиофизике постоянно фигурируют простые сигналы. Некоторые из них изображены на рис. 2.7. Прямоугольный импульс (рис. 2.7, а) и периодическая последовательность прямоугольных импульсов (рис. 2.7, д) удобны при изучении инерционности элементов радиосистем, треугольные (рис. 2.7, б) и пилообразные (рис. 2.7, в) импульсы используются в узлах развёртки электронных осциллографов, а косинусоидальные импульсы (рис. 2.7, г) полезны при анализе отсечек в нелинейных системах. На рис. 2.7, е показано, как прямоугольный импульс можно представить суперпозицией двух элементарных функций. Для периодической последовательности прямоугольных импульсов аналогичное представление имеет вид:

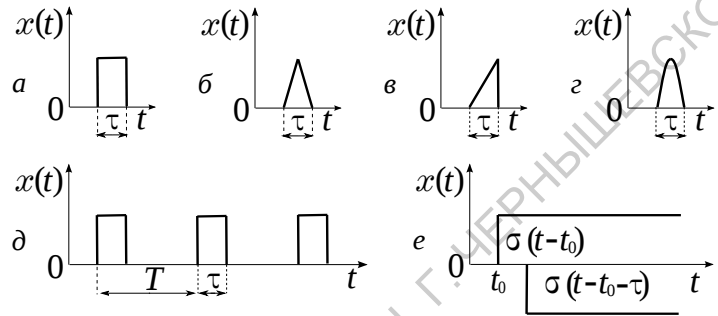


Рис. 2.7. Простые детерминированные сигналы

$$x(t) = \sum_{k=0}^{\infty} (\sigma(t - t_0 - kT) - \sigma(t - t_0 - \tau - kT)). \quad (2.15)$$

Рассмотрим разложение произвольных сигналов по элементарным функциям подробнее.

Разложение сигналов по функциям Хевисайда. Представим произвольный сигнал $x(t)$ (рис. 2.8) суммой начального скачка и множества смещённых во времени на Δ ступенчатых сигналов. Тогда

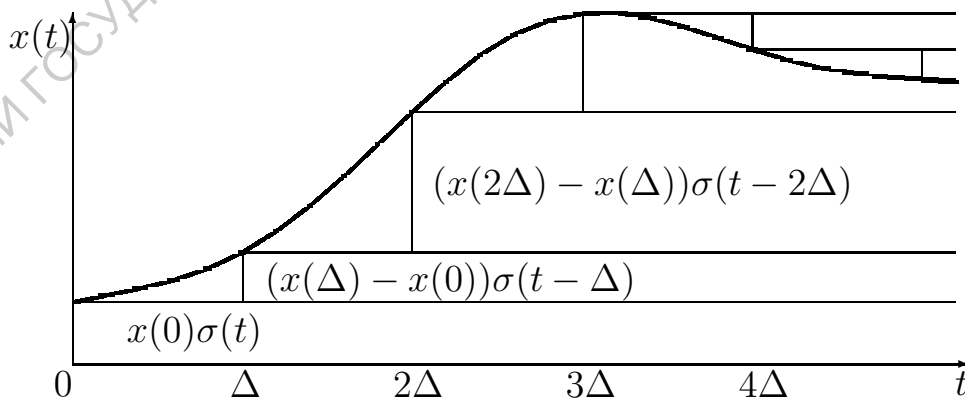


Рис. 2.8. Приближенное представление сигнала совокупностью σ -функций

$$\tilde{x}(t) = x(0)\sigma(t) + (x(\Delta) - x(0))\sigma(t - \Delta) + (x(2\Delta) - x(\Delta))\sigma(t - 2\Delta) + \dots =$$

$$\begin{aligned}
&= x(0)\sigma(t) + \sum_{k=1}^{\infty} [x(k\Delta) - x((k-1)\Delta)]\sigma(t - k\Delta) = \\
&= x(0)\sigma(t) + \sum_{k=1}^{\infty} \frac{x(k\Delta) - x((k-1)\Delta)}{\Delta} \sigma(t - k\Delta)\Delta.
\end{aligned}$$

Переходя к пределу при $\Delta \rightarrow 0$ и заменяя суммирование интегрированием по переменной τ ($d\tau \sim \Delta$), получим

$$x(t) = \lim_{\Delta \rightarrow 0} \tilde{x}(t) = x(0)\sigma(t) + \int_0^t \frac{dx}{d\tau} \sigma(t - \tau) d\tau. \quad (2.16)$$

Соотношение (2.16) выражает значение сигнала в любой момент времени через его значение в начальный момент времени и скорость изменения сигнала ($\frac{dx(t)}{dt}$).

Разложение сигналов по дельта-функциям Дирака. Представим сигнал $x(t)$ бесконечной последовательностью плотно примыкающих друг к другу прямоугольных импульсов длительностью Δ (рис. 2.9).

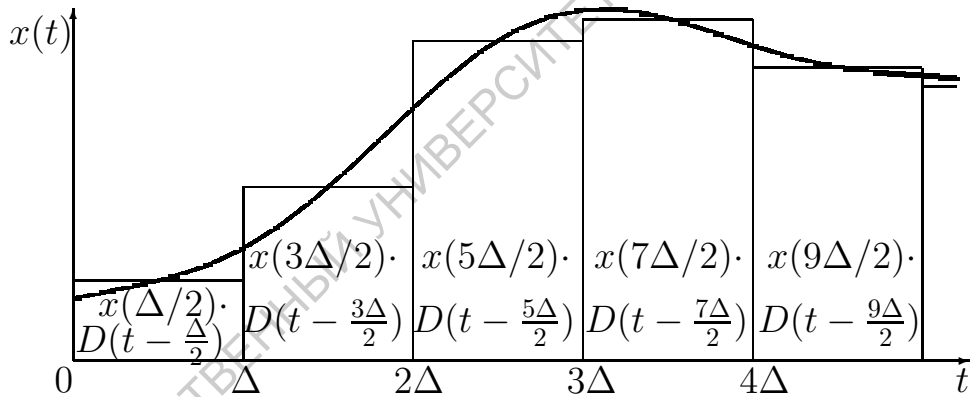


Рис. 2.9. Приближенное представление сигнала совокупностью прямоугольных импульсов

Если каждый прямоугольный импульс представить разностью двух сдвинутых во времени функций Хевисайда (см. формулу (2.15)), то сигнал $x(t)$ аппроксимируется приближенным выражением:

$$\tilde{x}(t) = \sum_{k=-\infty}^{\infty} x\left[\left(k + \frac{1}{2}\right)\Delta\right] \left(\sigma(t - k\Delta) - \sigma[t - (k + 1)\Delta]\right). \quad (2.17)$$

Переходя к пределу при $\Delta \rightarrow 0$, заменяя суммирование интегрированием по переменной $\tau = k\Delta$ ($d\tau \sim \Delta$) и учитывая, что при каждом определенном $t = k\Delta$ отлично от нуля только одно слагаемое с номером k , получим

$$x(t) = \lim_{\Delta \rightarrow 0} \tilde{x}(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau. \quad (2.18)$$

Это – известная формула (2.14), описывающая фильтрующие свойства дельта-функций: интеграл по времени от произведения непрерывного

сигнала $x(t)$ и дельта-функции дает значение непрерывной функции в точке, где сосредоточена дельта-функция. Естественно возникает потребность дать физическую трактовку построенным интегральным представлениям сигналов. К сожалению, это возможно только после рассмотрения физических свойств линейных радиоэлектронных систем.

2.3. Периодические сигналы и основы гармонического анализа

Частотные спектры периодических сигналов. В радиоэлектронных системах часто встречаются периодические сигналы различной формы. Анализ таких сигналов удастся упростить, представляя их в виде совокупности элементарных сигналов. Если в качестве элементарных использовать гармонические колебания, то всякую функцию $x(t)$, заданную во всех точках периода T , однозначную, кусочно-непрерывную и имеющую конечное число разрывов первого рода на период (условия Дирихле), можно представить тригонометрическим рядом Фурье (1.14):

$$x(t) = \frac{c_0}{2} + \sum_{k=1}^{\infty} (c_{kC} \cos k \frac{2\pi}{T} t + c_{kS} \sin k \frac{2\pi}{T} t).$$

Эту формулу нетрудно привести к виду

$$x(t) = \frac{c_0}{2} + \sum_{k=1}^{\infty} c_k \cos(k \frac{2\pi}{T} t - \varphi_k) = \frac{c_0}{2} + \sum_{k=1}^{\infty} c_k \cos(k\omega_1 t - \varphi_k), \quad (2.19)$$

где $c_0/2$ – постоянная составляющая; ω_1 – круговая частота первой гармоники сигнала; $c_k = \sqrt{c_{kC}^2 + c_{kS}^2}$; $k\omega_1 = 2\pi k/T$ и $\varphi_k = \arctg(c_{kS}/c_{kC})$ – амплитуда, угловая частота и начальная фаза k -й гармоники соответственно. Разложение (2.19) для каждой $x(t)$ единственно.

Совокупность амплитуд c_k и соответствующих частот $k\omega_1$ образуют *амплитудный спектр*, а совокупность фаз φ_k и частот $k\omega_1$ – *фазовый спектр*. Они изображаются графически в виде *спектральных диаграмм*. При этом каждой гармонике сигнала соответствует вертикальный отрезок (рис. 2.10), длина которого пропорциональна амплитуде или начальной фазе, а местоположение на горизонтальной оси соответствует частоте. Спектральные линии периодического сигнала образуют эквидистантный *дискретный* или *линейчатый* спектр. Только амплитудный или только фазовый спектр не характеризует сигнал полностью. Например, сдвиг сигнала во времени не влияет на амплитудный спектр, но изменяет его фазовый спектр.

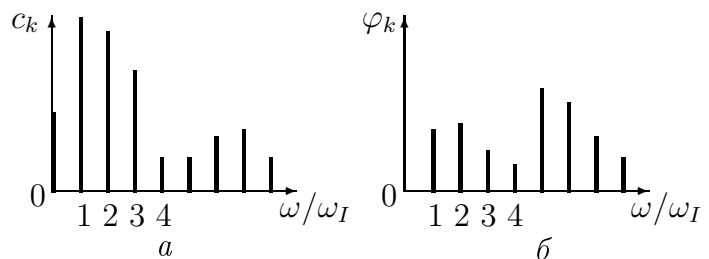


Рис. 2.10. Амплитудный (а) и фазовый (б) спектры периодического сигнала

Представление непрерывных периодических сигналов суммой постоянной составляющей и совокупности гармонических колебаний с кратными частотами называется *спектральным разложением в базисе гармонических функций* или *гармоническим анализом*.

Ряды, получаемые в результате гармонического анализа, сходятся, а необходимое число N базисных функций выбирается с учетом допустимой среднеквадратичной ошибки аппроксимации ε .

Используя теорему Парсеваля (1.12) и расписывая скалярное произведение в виде интегралов по области $[0, T]$, получим соотношение

$$\frac{1}{T} \int_0^T x^2(t) dt = \frac{c_0^2}{4} + \frac{1}{2} \sum_{k=1}^{\infty} c_k^2. \quad (2.20)$$

Если $x(t)$ представляет напряжение или ток, то левая часть совпадает со средней мощностью сигнала, выделяемой на единичном сопротивлении, а числа $c_k^2/2$ характеризуют распределение мощности по гармоникам. Графическое изображение зависимости $c_k^2/2(k\omega_1)$ представляет спектр мощности периодического сигнала (см. ниже рис. 2.13, z).

Комплексные частотные спектры. Для разложения периодических сигналов в частотный спектр в качестве ортонормального базиса можно использовать систему комплексных экспоненциальных функций. При этом комплексный ряд Фурье можно сразу записать в виде (1.18) или получить из тригонометрического ряда (2.19), представляя $\cos(k\frac{2\pi}{T}t - \varphi_k)$ по формуле Эйлера суммой экспонент. Тогда

$$x(t) = \frac{1}{2} \left(c_0 + \sum_{k=1}^{\infty} c_k e^{j(k\omega_1 t - \varphi_k)} + c_k e^{-j(k\omega_1 t - \varphi_k)} \right) = \sum_{k=-\infty}^{\infty} \dot{C}_k e^{jk\omega_1 t}, \quad (2.21)$$

$$\dot{C}_k = \frac{1}{T} \int_0^T x(t) e^{-j2\pi kt/T} dt, \quad (2.22)$$

где $\dot{C}_0 = c_0/2$, $\dot{C}_k = c_k e^{-j\varphi_k}/2$ и $\dot{C}_{-k} = \dot{C}_k^* = c_k e^{j\varphi_k}/2$.

Совокупность \dot{C}_k называется *комплексным спектром сигнала* $x(t)$. В отличие от вещественного ряда (2.19) комплексный ряд (2.21) содержит гармоники с отрицательными частотами (рис. 2.11).

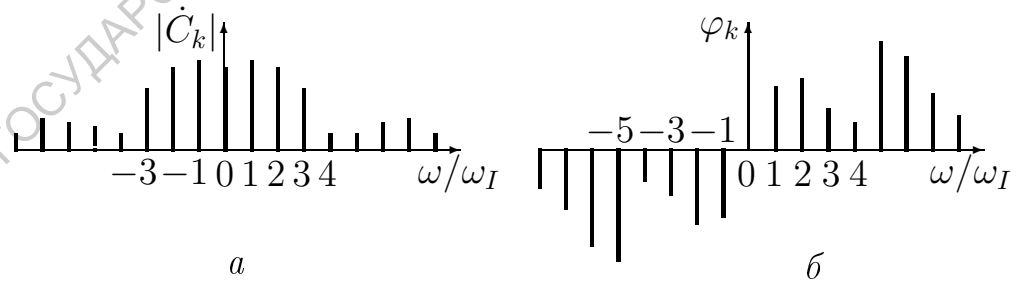


Рис. 2.11. Амплитудный (а) и фазовый (б) комплексные спектры периодического сигнала

В каждом \dot{C}_k содержится информация об амплитуде и начальной фазе. Модули комплексных амплитуд \dot{C}_k и \dot{C}_{-k} одинаковы и вдвое меньше соответствующей амплитуды вещественного спектра, а фазы \dot{C}_k совпадают с фазами составляющих вещественного спектра при $k > 0$ и противоположны им при $k < 0$. (Сравните рис. 2.11 и 2.10.) Отрицательные частоты не имеют физического смысла и порождены комплексным представлением гармонических колебаний. Любой частоте $k\omega_1$ в (2.21) соответствует пара слагаемых, сумма которых совпадает с действительной

функцией частоты $k\omega_1$. Поэтому всегда возможен переход от комплексного спектра к вещественному и наоборот.

Пример. Спектр последовательности прямоугольных видеоимпульсов. Для последовательности импульсов с амплитудой E , длительностью τ и периодом повторения T (рис. 2.12) согласно (2.21) имеем

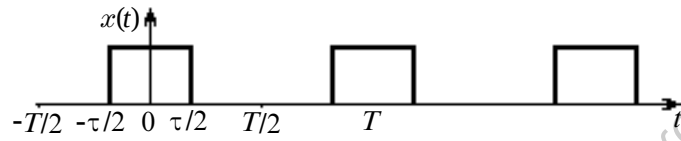


Рис. 2.12. Последовательность видеоимпульсов

$$C_0 = E \frac{\tau}{T}, \quad C_k = \frac{1}{T} \int_{-\tau/2}^{\tau/2} E e^{-jk\omega_1 t} dt = \frac{E}{T} \cdot \frac{e^{-jk\omega_1 t}}{-jk\omega_1} \Big|_{-\tau/2}^{\tau/2} =$$

$$= \frac{E}{-jk\omega_1 T} (e^{-jk\omega_1 \tau/2} - e^{jk\omega_1 \tau/2}) = E \frac{\tau}{T} \cdot \frac{\sin k\omega_1 \tau/2}{k\omega_1 \tau/2} = E \frac{\tau}{T} \cdot \frac{\sin k\pi f_1 \tau}{k\pi f_1 \tau}, \quad (2.23)$$

где $\omega_1 = 2\pi/T$ ($f_1 = 1/T$) – частоты основной (первой) гармоники $x(t)$. Так как сигнал $x(t)$ является чётной функцией, все C_k вещественны.

Амплитудный и фазовый спектры рассматриваемой последовательности импульсов, т.е. зависимости амплитуд и фаз от частот гармоник, представлены на рис. 2.13, а и 2.13, б соответственно.

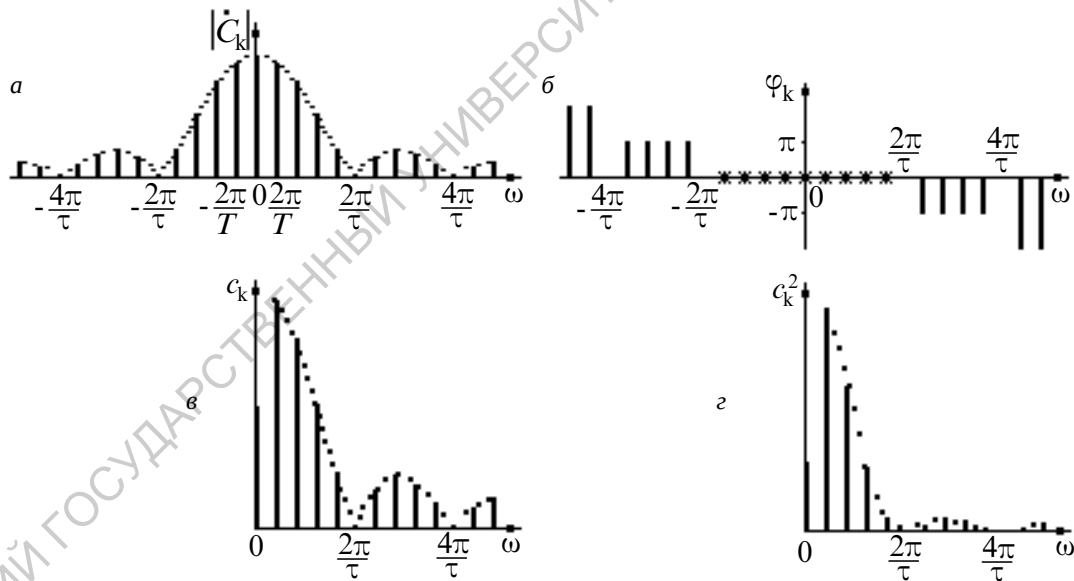


Рис. 2.13. Спектры последовательности прямоугольных импульсов: комплексный амплитудный (а), фазовый (б), вещественный амплитудный (в), спектр мощности (г)

Сигнал $x(t)$ можно представить в виде вещественного ряда Фурье:

$$x(t) = \sum_{k=-\infty}^{\infty} E \frac{\tau}{T} \cdot \frac{\sin k\omega_1 \tau/2}{k\omega_1 \tau/2} e^{jk\omega_1 t} = E \frac{\tau}{T} + \sum_{k=1}^{\infty} 2E \frac{\tau}{T} \cdot \frac{\sin k\omega_1 \tau/2}{k\omega_1 \tau/2} \cos k\omega_1 t =$$

$$= \frac{E}{q} \left(1 + 2 \sum_{k=1}^{\infty} \frac{\sin k\pi/q}{k\pi/q} \cos k\omega_1 t \right) = \frac{E}{q} \left(1 + 2 \sum_{k=1}^{\infty} \text{sinc}(k/q) \cos k\omega_1 t \right), \quad (2.24)$$

где величина $q = T/\tau$ характеризует длительность паузы между импульсами и называется *скважностью последовательности*, а функция $\text{sinc } z = \sin(\pi z)/(\pi z)$. Амплитудный вещественный спектр, построенный по формуле (2.24), представлен на рис. 2.13, в. Все спектральные диаграммы на рис. 2.13 построены для скважности 5.

Проанализируем полученные результаты.

1. Огибающая спектра, согласно (2.24), определяется соотношением

$$\frac{2\tau}{T} \cdot \left| \frac{\sin k\omega_1\tau/2}{k\omega_1\tau/2} \right| = \frac{2}{q} \cdot \left| \frac{\sin k\pi/q}{k\pi/q} \right| = \frac{2}{q} \cdot |\text{sinc}(k/q)|, \quad (2.25)$$

имеет *лепестковую структуру* и принимает нулевые значения на частотах $\omega = 2K\pi/\tau$

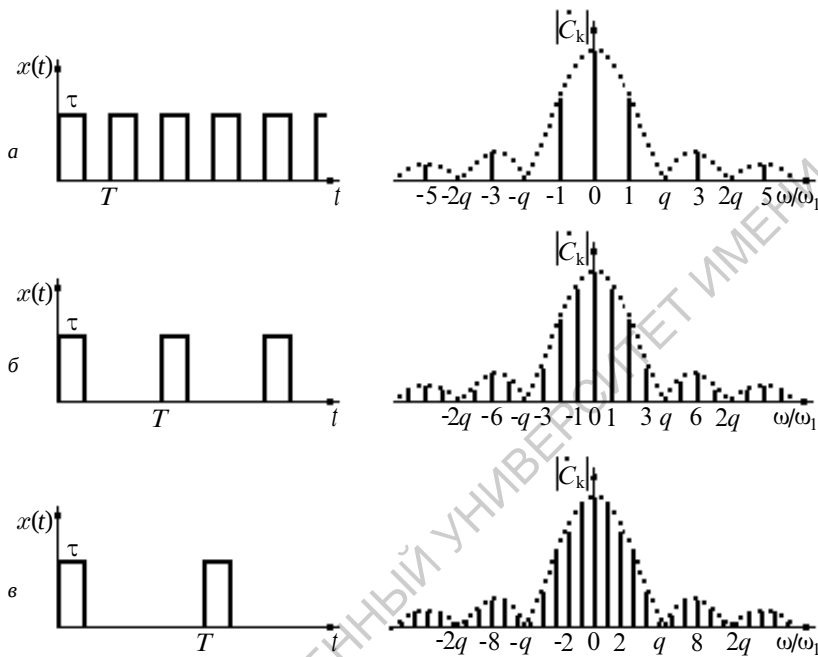


Рис. 2.14. Последовательности видеоимпульсов и их амплитудные спектры при различной скважности

Последовательность прямоугольных импульсов, длительность τ которых равна половине периода (рис. 2.14, а), называется *меандр*.

2. Уменьшение длительности импульсов τ сопровождается перемещением нулей огибающей в область все более высоких частот, т.е. *спектр расширяется*, а сами амплитуды уменьшаются.

3. Увеличение T при неизменном значении τ вызывает рост числа спектральных линий в каждом лепестке диаграммы и уменьшение расстояний между ними. Так, при удвоении скважности (рис. 2.14, б) новые гармоники располагаются между гармониками первоначального ряда (рис. 2.14, а). При $T \rightarrow \infty$, т.е. при переходе от последовательности к одиночному импульсу, спектр *из дискретного превращается в непрерывный*, состоящий из бесконечно большого числа линий малой амплитуды ($|\dot{C}_k| \rightarrow 0$), расположенных сколь угодно близко друг к другу.

4. При восстановлении гладких функций $x(t)$ конечным числом членов ряда (2.24) сходимость в среднем квадратичном обеспечивает достаточно точную аппроксимацию кривой. Если же функция содержит

или на гармониках с номерами $k = K \cdot q$. Первый нуль огибающей спектра последовательности прямоугольных импульсов имеет круговую частоту $\omega = 2\pi/\tau$, следующий $\omega = 4\pi/\tau$ и т.д. Тогда в нуль обращаются гармоники с номерами $k = q, 2q, \dots$. В частности, при $q=2$ (рис. 2.14, а) все чётные гармоники отсутствуют, а амплитуды нечётных уменьшаются обратно пропорционально номерам гармоник.

разрывы первого рода (рис. 2.15), то в окрестностях точек разрыва крутизна изменения функции уменьшается, и даже при очень большом

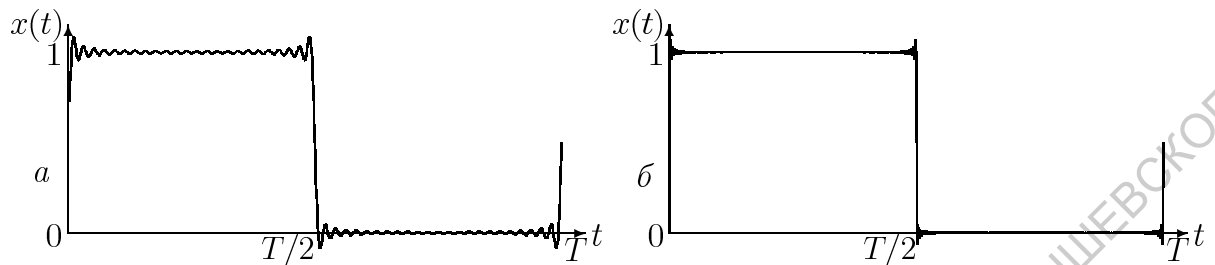


Рис. 2.15. Аппроксимация $x(t)$ укороченным рядом Фурье: $a - N=47$, $b - N=147$

числе N членов ряда возникают выбросы, достигающие 9% от величины разрыва. При этом крутизна изменения функции с ростом N возрастает, а наибольшая величина выбросов практически не изменяется. Появление неустранимых погрешностей при аппроксимации разрывных функций рядом Фурье с конечным числом членов ряда получило название явления Гиббса³.

2.4. Гармонический анализ неперiodических сигналов

Пусть сигнал $x(t)$ задан в промежутке (t_1, t_2) и равен нулю за его пределами. Используя (1.21), можно было бы разложить $x(t)$ в комплексный ряд. Однако восстановить $x(t)$ путём суммирования экспоненциальных функций не удастся: возникает совсем другой сигнал – периодическое повторение исходной функции. Более естественным представляется следующий прием: для $x(t)$ выбирается отрезок $[-T/2, T/2]$, содержащий промежуток (t_1, t_2) , и дополнительным повторением отрезка $[-T/2, T/2]$ создается периодический процесс $x_{\text{пер}}(t)$. Его можно разложить в ряд

$$x_{\text{пер}}(t) = \sum_{k=-\infty}^{\infty} \dot{C}_k e^{j2\pi kt/T}, \quad \dot{C}_k = \frac{1}{T} \int_{-T/2}^{T/2} x_{\text{пер}}(t) e^{-j2\pi kt/T} dt,$$

а затем устремить период повторения T к бесконечности.

Переход от ряда Фурье к интегралу Фурье. Спектральная функция. Рассмотрим спектр периодического сигнала $x_{\text{пер}}(t)$. Он линейчатый и в силу большого T состоит из близко расположенных друг к другу линий. Выделим небольшой интервал частот $\Delta\omega$ в окрестности частоты $\omega = 2\pi k/T$. Пусть $\Delta\omega = 2\pi n/T$, т.е. содержит n гармоник ($n \ll k$). Поскольку период T велик, а $\Delta\omega$ мало, все гармоники имеют близкие частоты, их амплитуды можно считать практически одинаковыми, а совокупность $2n$ гармоник (n с положительными и n с отрицательными частотами) заместить эквивалентным гармоническим колебанием с частотой ω и некоторой комплексной амплитудой $\Delta\dot{C} = 2n\dot{C}_k$:

³Это явление открыл в 1848 г. Г. Уиллбрейам и «переоткрыл» Дж. Гиббс в 1898 г.

$$\Delta\dot{C} = \frac{2n}{T} \int_{-T/2}^{T/2} x_{\text{пер}}(t) e^{-j2\pi kt/T} dt.$$

Используя введённые выше обозначения, окончательно получим:

$$\Delta\dot{C} = \frac{\Delta\omega}{\pi} \int_{-T/2}^{T/2} x_{\text{пер}}(t) e^{-j\omega t} dt. \quad (2.26)$$

Если рассмотреть отношение $\Delta\dot{C}$ к $\Delta\omega/\pi$ и устремить T к бесконечности, то $\Delta\omega \rightarrow 0$, вспомогательный сигнал $x_{\text{пер}}(t)$ превращается в $x(t)$, а отношение $\frac{\Delta\dot{C}}{\Delta\omega}$ – в производную $X(\omega) = \frac{d\dot{C}}{d\omega}$. Это комплексная амплитуда гармоник спектра, приходящаяся на единицу частотного интервала, т.е. спектральная плотность гармоник. Поскольку спектральные гармоники имеют размерность функции $x(t)$ (измеряются в вольтах В или в амперах А), размерность спектральной плотности – В/Гц или А/Гц.

Назовем

$$X(\omega) = \lim_{\Delta\omega \rightarrow 0} \pi \frac{\Delta\dot{C}}{\Delta\omega} = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (2.27)$$

спектральной плотностью или спектральной функцией непериодического сигнала $x(t)$ ⁴. Спектральная плотность непериодических сигналов является непрерывной комплексной функцией частоты и представляет интеграл Фурье, а формула (2.27) называется преобразованием Фурье сигнала $x(t)$. Чтобы представить $x(t)$ интегралом Фурье, она должна быть абсолютно интегрируемой.

В радиоэлектронике наряду с комплексной спектральной плотностью часто применяется вещественная спектральная плотность сигналов⁵, определённая только для положительных частот от 0 до $+\infty$ соотношением

$$X(\omega) = 2 \int_0^{\infty} x(t) \cos \omega t dt.$$

Для восстановления сигнала по известной его спектральной функции $X(\omega)$ используется обратное преобразование Фурье:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega. \quad (2.28)$$

Прямое и обратное преобразования Фурье. Формулы (2.27) и (2.28) так же, как (2.21) и (2.22), составляют пару преобразований Фурье $X(\omega) = \mathbf{F}[x(t)]$, $x(t) = \mathbf{F}^{-1}[X(\omega)]$, причём (2.27) анализирует функцию $x(t)$ по составу спектральных составляющих, а (2.28) восстанавливает или синтезирует $x(t)$ из гармонических функций. Теорема Фурье

⁴Этот термин предпочтительнее, что подтверждает государственный стандарт ГОСТ 16465-70.

⁵Вещественную спектральную функцию в математике называют косинус-интегралом Фурье.

утверждает, что процесс анализа-синтеза выполняется без потерь информации, т.е. восстановленный сигнал идентичен исходному сигналу, если функция $x(t)$ абсолютно интегрируема $\left(\int_{-\infty}^{\infty} |x(t)| dt < \infty\right)$, имеет конечное число максимумов и минимумов, а также конечное число разрывов.

Пара преобразований Фурье становится полностью симметричной, если в (2.28) перейти от круговой частоты ω к частоте $f = \omega/2\pi$. Тогда

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df. \quad (2.29)$$

Полная симметрия интегральных преобразований Фурье позволяет ввести принцип дуальности времени и частоты.

- Если $X(f)$ является преобразованием Фурье функции $x(t)$, то $x(-f)$ представляет преобразование Фурье функции $X(t)$.

Итак, периодический или непериодический физический процесс полностью определяется заданием либо функции времени либо его комплексного спектра. Оба эти представления равносильны.

Представляя энергию сигнала $x(t)$ в виде $\int_{-\infty}^{\infty} |x(t)|^2 dt$, можно записать важное соотношение, известное как равенство Парсеваля⁶:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df.$$

Функция $|X(f)|^2$ называется энергетическим спектром сигнала.

Спектральная плотность произведения двух сигналов. Произведения двух сигналов или сигнала и временной функции системы часто встречаются в радиоэлектронике и статистической радиофизике, а спектральное представление произведения функций оказывается очень эффективным при решении различных задач.

Теорема. Преобразование Фурье произведения двух функций является свёрткой их преобразований Фурье.

Доказательство. Пусть $X(\omega)$ и $Y(\omega)$ – спектральные плотности сигналов $x(t)$ и $y(t)$ соответственно. Спектральная функция $S(\omega)$ произведения этих сигналов удовлетворяет соотношению

$$S(\omega) = \int_{-\infty}^{\infty} x(t) \cdot y(t) e^{-j\omega t} dt.$$

Если функцию $x(t)$, представленную в виде интеграла Фурье

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) e^{j\Omega t} d\Omega,$$

⁶Строго говоря, Парсеваль ввел аналогичное равенство для рядов Фурье, а его обобщение на интегралы Фурье, вероятно, было сделано Релеем.

подставить в выражение для $S(\omega)$ и изменить порядок интегрирования, то получим

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) \left(\int_{-\infty}^{\infty} y(t) e^{-j(\omega-\Omega)t} dt \right) d\Omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) Y(\omega - \Omega) d\Omega.$$

Выражение

$$S(\omega) = X(\omega) * Y(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) Y(\omega - \Omega) d\Omega \quad (2.30)$$

называется *интегралом свёртки* или *свёрткой* функций $X(\omega)$ и $Y(\omega)$. Таким образом, спектральная плотность произведения двух сигналов есть свёртка их спектральных плотностей.

Интеграл свёртки $S(\omega)$ допускает геометрическое толкование в виде численного значения площади, ограниченной произведением функций $X(\omega)$ и $Y(\omega)$.

Используя принцип дуальности частоты и времени, можно записать теорему о свёртке в инверсной формулировке: спектральная плотность свёртки двух функций равна произведению их спектральных плотностей:

$$\mathbf{F} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} x(t) y(\tau - t) dt \right] = X(\omega) \cdot Y(\omega). \quad (2.31)$$

Представленная теорема справедлива и для периодических процессов. Она имеет широчайшее применение в различных областях современного естествознания, от астрономии до социологии, а в теоретической радиоэлектронике позволяет быстро и в компактной форме находить решения очень сложных задач теории фильтрации и определять характеристики взаимной связи коротких сигналов.

2.4.1. Спектральные плотности интегрируемых сигналов

Спектральная плотность экспоненциального видеоимпульса.

Пусть $x(t) = \begin{cases} Ee^{-\alpha t}, & \text{при } t \geq 0 \quad (\alpha > 0 \text{ и вещественная}) \\ 0, & \text{при } t < 0. \end{cases}$

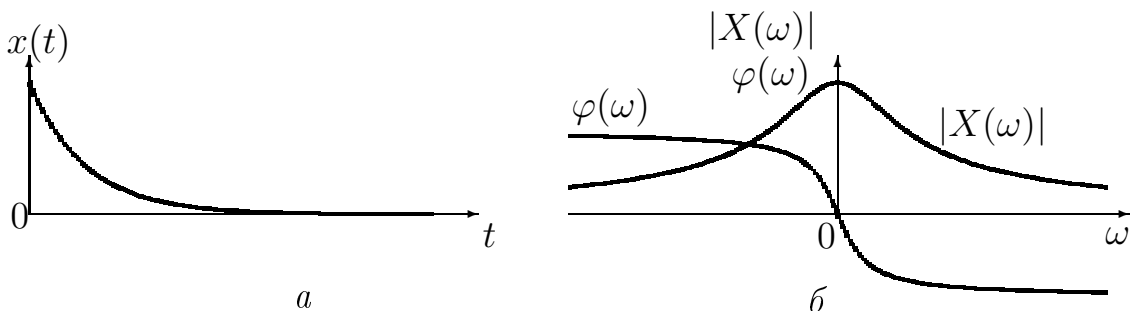


Рис. 2.16. Экспоненциальный видеоимпульс (а) и частотные зависимости модуля $|X(\omega)|$ и аргумента $\varphi(\omega)$ его спектральной функции (б)

Фурье-преобразование функции $x(t)$ (рис. 2.16, а) имеет вид:

$$X(\omega) = \int_{-\infty}^{\infty} Ee^{-\alpha t} e^{-j\omega t} dt = \frac{E}{-(\alpha + j\omega)} e^{-(\alpha + j\omega)t} \Big|_0^{\infty} = \frac{E}{\alpha + j\omega}. \quad (2.32)$$

Частотные зависимости модуля $|X(\omega)| = E/\sqrt{\alpha^2 + \omega^2}$ и аргумента $\varphi(\omega) = -\text{arctg}(\omega/\alpha)$ спектральной функции изображены на рис. 2.16, б. Эти спектральные диаграммы иллюстрируют важнейшее свойство симметрии спектральных функций, характерное для преобразования Фурье вещественных сигналов $x(t)$.

- Если $x(t)$ вещественная функция, то модуль (и реальная часть) её спектральной плотности являются *чётными функциями*, а аргумент (и мнимая часть) – *нечётными функциями*, т.е.

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

Согласно (2.28) функцию $x(t)$ можно представить следующим интегралом Фурье:

$$x(t) = Ee^{-\alpha t} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{E}{\alpha + j\omega} e^{j\omega t} d\omega = \int_{-\infty}^{\infty} \frac{E}{\alpha + j2\pi f} e^{j2\pi f t} df.$$

Спектральная плотность гауссова импульса. Пусть видеоимпульс описывается функцией Гаусса (рис. 2.17, а)

$$x(t) = Ee^{-\pi(t/\tau)^2},$$

где τ – некоторая характеристическая длительность («ширина» импульса на уровне $1/e$).

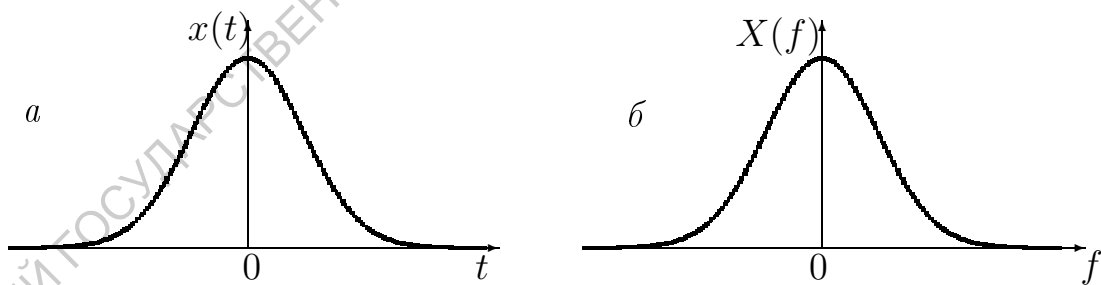


Рис. 2.17. Гауссов видеоимпульс (а) и его спектральная функция (б)

Функция Гаусса быстро убывает при $|t| \rightarrow \infty$, т.е. абсолютно интегрируема. При вычислении её спектральной плотности можно получить

$$X(f) = \int_{-\infty}^{\infty} Ee^{-\pi(t/\tau)^2} e^{-j2\pi f t} dt = E\tau e^{-\pi(f\tau)^2}.$$

Итак, спектральная плотность гауссова импульса вещественна (рис. 2.17, б) и описывается такой же функцией, как и $x(t)$:

$$x(t) = Ee^{-\pi(t/\tau)^2}, \quad X(f) = E\tau e^{-\pi(f\tau)^2}. \quad (2.33)$$

Построенная математическая модель имеет важное значение для теории сигналов, так как в этом случае симметрией обладают не только

интегральные соотношения для функции $x(t)$ и её спектральной функции $X(f)$, но и сами функции.

Спектральная плотность дельта-функции. Дельта-функция, хотя и принимает бесконечно большое значение при $t = 0$, обладает единичной площадью и поэтому абсолютно интегрируема. Согласно (2.27) для спектральной плотности функции $\delta(t)$ имеем:

$$X_{\delta}(\omega) = \int_{-\infty}^{\infty} \delta(t)e^{-j\omega t} dt = \int_{-\infty}^{\infty} \delta(t) dt = 1,$$

так как $\delta(t) \neq 0$ только в точке $t=0$, где экспонента равна единице.

Итак, дельта-функция имеет сплошной и равномерный спектр для всех $-\infty < \omega < \infty$ (рис. 2.18).

Спектральная плотность $\delta(t)$ вещественна и равна единице.

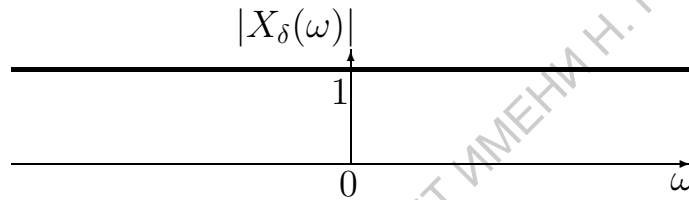


Рис. 2.18. Спектральная плотность дельта-функции

Обратное преобразование Фурье удовлетворяет соотношению

$$\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_{\delta} e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} d\omega. \quad (2.34)$$

Последняя формула имеет большое практическое значение, так как с учетом дуальности частоты ω и времени t приводится в виду:

$$\delta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega t} dt. \quad (2.35)$$

Спектральная плотность прямоугольного видеоимпульса.

Пусть прямоугольный видеоимпульс задан аналитическим выражением:

$$x(t) = \begin{cases} E, & \text{при } -\tau/2 < t < +\tau/2, \\ 0, & \text{при } t < -\tau/2 \text{ и } t > +\tau/2. \end{cases}$$

Тогда

$$X(\omega) = E \int_{-\tau/2}^{\tau/2} e^{-j\omega t} dt = \frac{E}{-j\omega} e^{-j\omega t} \Big|_{-\tau/2}^{\tau/2} = E\tau \frac{\sin \omega\tau/2}{\omega\tau/2}$$

или

$$X(f) = E\tau \frac{\sin \pi f\tau}{\pi f\tau} = E\tau \text{sinc}(f\tau). \quad (2.36)$$

Используя обратное преобразование Фурье, синтезируем $x(t)$ в виде:

$$x(t) = E\tau \int_{-\infty}^{\infty} \text{sinc}(f\tau) e^{j2\pi ft} df.$$

Согласно (2.36) спектральная плотность одиночного прямоугольного импульса вещественна и представляет известную функцию sinc ($f\tau$).

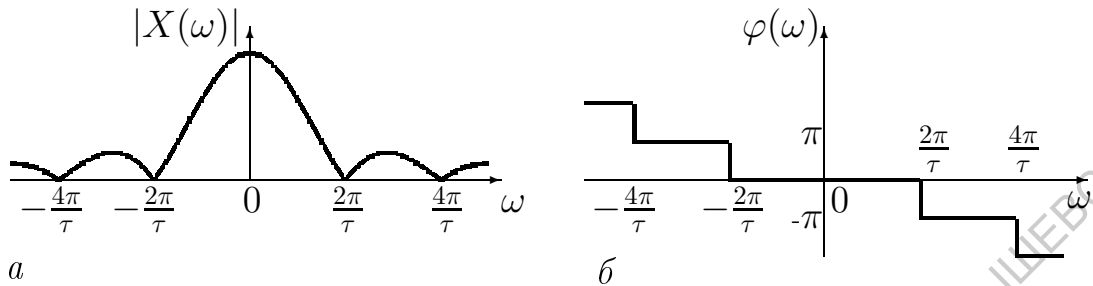


Рис. 2.19. Модуль (а) и аргумент (б) спектральной плотности прямоугольного импульса

Сопоставление спектральных диаграмм рис. 2.19, а, б и 2.13, а, б показывают, что огибающие спектров одинаковы. Нули огибающих располагаются при одинаковых значениях частоты ($\omega = 2\pi n/\tau$). Единственное отличие состоит в том, что спектр одиночного импульса сплошной, а последовательности импульсов – линейчатый.

Наконец, используя принцип дуальности частоты и времени, можно показать, что преобразование Фурье функции $x(t) = AF \operatorname{sinc}(Ft)$ имеет вид

$$X(f) = \begin{cases} A, & \text{при } -F/2 < f < F/2, \\ 0, & \text{при } f < -F/2 \text{ и } f > F/2. \end{cases}$$

Спектральная плотность радиоимпульса. Радиоимпульс представляет собой произведение импульсной огибающей $x_{\text{вид}}(t)$ и гармонического колебания с частотой ω_0 (см. раздел 2.1 и формулу (2.10)):

$$x_{\text{рад}}(t) = x_{\text{вид}}(t) \cdot \cos \omega_0 t.$$

Если спектральная плотность видеоимпульса $X_{\text{вид}}(\omega)$ известна, то спектральная плотность сигнала $x_{\text{рад}}(t)$ по формуле (2.27) принимает вид:

$$\begin{aligned} X_{\text{рад}}(\omega) &= \int_{-\infty}^{\infty} x_{\text{вид}}(t) \cos \omega_0 t e^{-j\omega t} dt = \\ &= \frac{1}{2} \int_{-\infty}^{\infty} x_{\text{вид}}(t) e^{-j(\omega - \omega_0)t} dt + \frac{1}{2} \int_{-\infty}^{\infty} x_{\text{вид}}(t) e^{-j(\omega + \omega_0)t} dt = \\ &= \frac{1}{2} X_{\text{вид}}(\omega - \omega_0) + \frac{1}{2} X_{\text{вид}}(\omega + \omega_0). \end{aligned} \quad (2.37)$$

- Спектральная плотность радиоимпульса отличается от видеоимпульса с такой же огибающей *переносом спектра в область высоких частот*.

При этом в окрестностях $\pm\omega_0$ формируются две области спектра с огибающими, подобными огибающей видеоимпульса и вдвое меньшей амплитудой, а нули спектральной функции с частотами $\pm\omega_{\text{н}} = \frac{2K\pi}{\tau}$ смещаются к частотам $\pm\omega_0 \pm \omega_{\text{н}}$.

В частности, для прямоугольных видео- и радиоимпульсов (рис. 2.20)

$$X_{\text{вид}}(\omega) = E\tau \frac{\sin \omega\tau/2}{\omega\tau/2}, \quad X_{\text{вид}}(f) = E\tau \frac{\sin \pi f\tau}{\pi f\tau} = E\tau \text{sinc} f\tau,$$

$$X_{\text{рад}}(\omega) = \frac{E\tau}{2} \left(\frac{\sin(\omega - \omega_0)\tau/2}{(\omega - \omega_0)\tau/2} + \frac{\sin(\omega + \omega_0)\tau/2}{(\omega + \omega_0)\tau/2} \right),$$

$$X_{\text{рад}}(f) = \frac{E\tau}{2} (\text{sinc}[\tau(f - f_0)] + \text{sinc}[\tau(f + f_0)]).$$

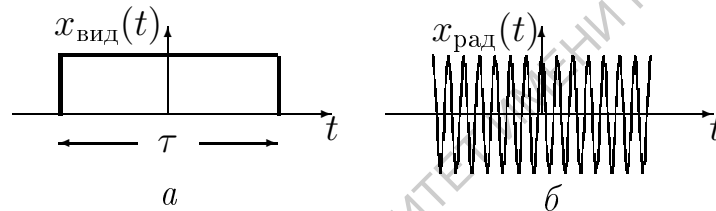


Рис. 2.20. Прямоугольные видеоимпульс (а) и радиоимпульс (б) одинаковой длительности

Итак, спектральная плотность радиоимпульса представляет наложение двух непрерывных функций вида $\text{sinc}(f\tau)$, принимающих максимальные значения на $\pm f_0$ (рис. 2.21). По мере удаления от максимумов sinc -функции быстро затухают, и при больших значениях $f_0\tau$ их «перекрытием» можно пренебречь.

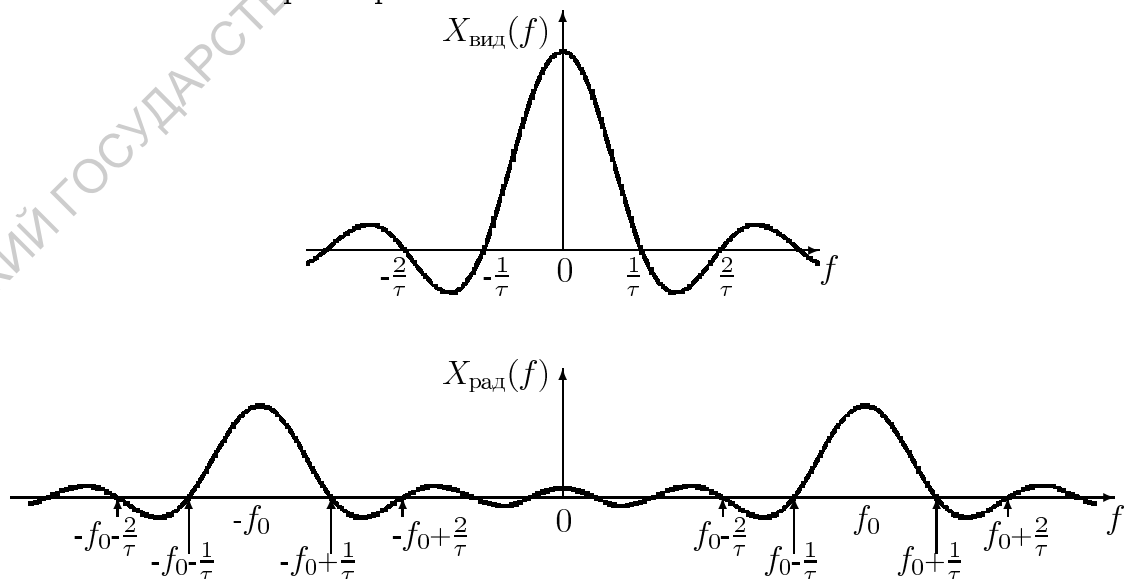


Рис. 2.21. Спектральные плотности прямоугольных видео- и радиоимпульса

2.4.2. Связь ширины спектра с длительностью и формой импульсных сигналов

Шириной спектра сигнала будем называть интервал частот, в пределах которого модуль спектральной функции превышает некоторое наперед заданное значение. Обычно ширина спектров сигналов, заданных в виде временного изменения напряжения или тока, определяется на уровне $|X(f)|/10$, так как этот интервал содержит основную часть мощности сигнала (до 90 % и более).

Как следует из рассмотренных спектральных диаграмм, ширина спектра прямоугольных импульсов произвольной длительности τ удовлетворяет соотношению $\Delta f \simeq 1.8/\tau$, треугольных импульсов – соотношению $\Delta f \simeq 1.5/\tau$, а гауссова импульса – неравенству $\Delta f \leq 1/\tau$. Таким образом,

- произведение ширины спектра импульсного сигнала и его длительности является величиной постоянной, зависящей только от формы импульса

$$\tau \Delta f \geq 1/\pi. \quad (2.38)$$

Длительность процесса и ширина его спектра не могут одновременно иметь произвольно малые значения. Соотношение (2.38) напоминает формулу Гейзенберга и по аналогии с ней иногда называется соотношением неопределённости. Им руководствуются при расчёте полосы пропускания радиоустройств, а в грубых оценках, достаточных для практики, даже полагают $\Delta f \simeq 1/\tau$.

Произведение $D = \tau \Delta f$ в радиоэлектронике называется *базой сигнала*. Если база сигнала D мала (порядка единиц), то сигнал называется *простым*. Когда D достигает значений 100 и более, то сигнал называется *сложным*. Все рассмотренные до сих пор сигналы были простыми. В качестве примера сложных сигналов в дальнейшем будет рассмотрен радиоимпульс, несущая частота которого изменяется по линейному закону (ЛЧМ-сигнал).

Реальные синусоидальные сигналы имеют конечную длительность (начало и конец) и, строго говоря, не являются гармоническими. Спектральный метод позволяет установить условия, при которых для реальных синусоидальных колебаний справедлива модель гармонического сигнала.

Синусоидальный сигнал конечной протяжённости Δt можно считать радиоимпульсом с $\tau = \Delta t$. Его спектральная плотность описывается функцией $\text{sinc}(f\Delta t)$ и расстояние между первыми нулями огибающей равно $2/\Delta t$.

Пусть однотоновый сигнал с частотой f_0 длится в течение 20 с. Ширина главного лепестка его спектра составляет $\Delta f = 2/\Delta t = 0.1$ Гц. Если сигнал имеет частоту 100 кГц, то относительная ширина его спектра ($\Delta f/f_0 = 2/f_0\Delta t$) составляет 10^{-6} или 0.0001%, и монохроматичностью можно пренебречь. Но если сигнал имеет частоту 100 Гц (рис. 2.22, а), то $\Delta f/f_0$ составляет 0.1%, и его едва ли целесообразно представлять моделью в виде гармонической функции. При увеличении длительности сигнала спектр быстро сужается (рис. 2.22, б, в), но превращение в дельта-функцию (рис. 2.22, г) происходит только для $\Delta t \rightarrow \infty$.

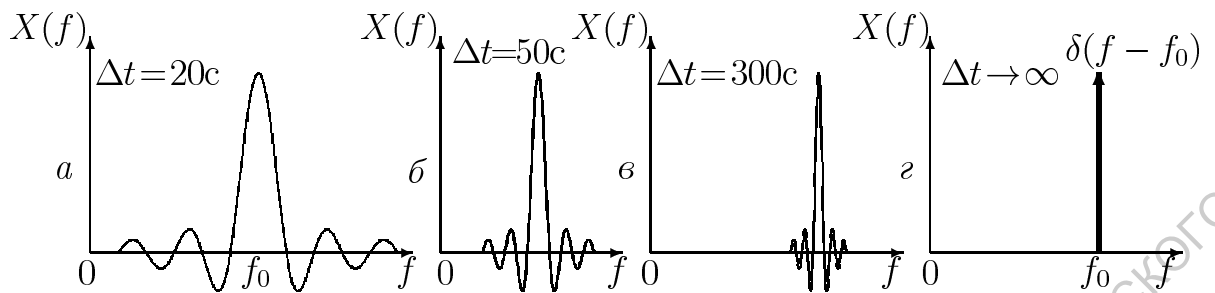


Рис. 2.22. Спектральные плотности синусоидального сигнала конечной длительности

- Синусоидальное колебание конечной длительности Δt имеет *сплошной спектр конечной ширины*, и правомерность его представления идеализированной моделью гармонического сигнала зависит от величины $\Delta f / f_0 = 2 / f_0 \Delta t$.

2.4.3. Спектральные плотности неинтегрируемых сигналов

Если функция не удовлетворяет условию абсолютной интегрируемости, то Фурье-преобразование нельзя использовать, и для построения спектральных плотностей сигналов возможны искусственные приемы с обобщёнными функциями Дирака или преобразованием Лапласа.

1. Спектральная плотность постоянного сигнала. Спектральную плотность сигнала единичной амплитуды

$$x(t) = 1, \quad -\infty < t < \infty$$

получим, используя дуальность частоты и времени для функции Дирака и её спектральной плотности. По формуле (2.27) находим

$$X_{(x(t)=1)}(\omega) = \int_{-\infty}^{\infty} 1 \cdot e^{-j\omega t} dt = \int_{-\infty}^{\infty} e^{-j\omega t} dt.$$

Так как зависимости $x(t)$ и $X_{\delta}(\omega)$ идентичны, по формуле (2.37) имеем

$$X_{(x(t)=1)} = 2\pi\delta(\omega).$$

Итак, спектральная плотность постоянного сигнала (напряжения или тока) содержит единственную вещественную составляющую в виде функции Дирака на нулевой частоте.

2. Спектральные плотности функций Хевисайда $\sigma(t)$ и $1/\pi t$. Воспользуемся асимптотической моделью функции Хевисайда в виде экспоненциального импульса (формула (2.12)). Тогда

$$\dot{X}_{\sigma}(\omega) = \lim_{\alpha \rightarrow 0} \frac{1}{\alpha + j\omega}.$$

Представим $\dot{X}_{\sigma}(\omega)$ суммой действительной и мнимой компонент:

$$\dot{X}_{\sigma}(\omega) = \lim_{\alpha \rightarrow 0} \left(\frac{\alpha}{\alpha^2 + \omega^2} + \frac{j\omega}{\alpha^2 + \omega^2} \right).$$

Предел первого слагаемого (вещественная составляющая спектральной плотности) обращается в бесконечность на нулевой частоте и в нуль на всех других частотах, т.е. представляет произведение функции Дирака и некоторого масштабного множителя. Можно показать, что величина множителя равна π . Предел второго слагаемого (мнимая составляющая спектральной плотности) описывается выражением $1/j\omega$. Обе компоненты спектральной плотности изображены на рис. 2.23, а полная спектральная плотность функции $\sigma(t)$ описывается соотношением

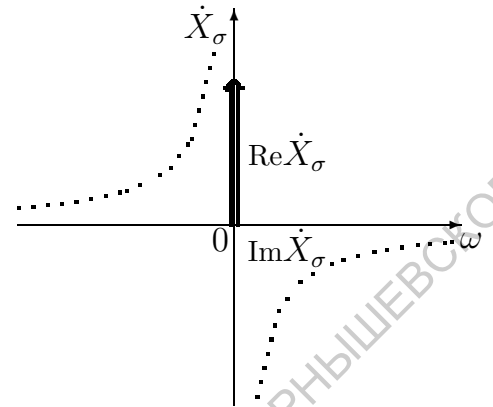


Рис. 2.23. Спектральная плотность функции Хевисайда

$$\dot{X}_\sigma = \pi\delta(\omega) + \frac{1}{j\omega}. \quad (2.39)$$

Спектральные плотности $\sigma(t)$ и постоянного сигнала различаются. Функция включения, в отличие от $x(t) = 1$, содержит скачок огибающей сигнала и, как следствие, реактивную составляющую спектральной плотности. Вещественные составляющие отличаются вдвое из-за разницы в длительностях сигналов.

Если в качестве дуальной для $X(\omega) = \frac{1}{j\omega}$ ввести временную функцию $g(t) = \frac{1}{\pi t}$, то спектральная плотность последней $G(\omega)$ будет соответствовать функции включения с мнимой единицей в качестве множителя:

$$\dot{G}(\omega) = \mathbf{F}\left[\frac{1}{\pi t}\right] = -j \operatorname{sgn}(\omega) = \begin{cases} -j & \text{при } \omega > 0, \\ j & \text{при } \omega < 0. \end{cases} \quad (2.40)$$

Справедливость (2.40) нетрудно показать, если представить $G(\omega)$ асимптотической формулой

$$\dot{G}(\omega) = -j \operatorname{sgn}(\omega) = \lim_{\alpha \rightarrow 0} [-j \operatorname{sgn}(\omega) e^{-\alpha|\omega|}]$$

и вычислить $g(t)$ в виде обратного преобразования Фурье:

$$g(t) = \frac{j}{2\pi} \lim_{\alpha \rightarrow 0} \left(\int_{-\infty}^0 e^{(\alpha+jt)\omega} d\omega - \int_0^{\infty} e^{-(\alpha+jt)\omega} d\omega \right) = \frac{1}{\pi t}.$$

3. Спектральные плотности периодических сигналов. Периодические, в частности гармонические сигналы, заданы для $-\infty < t < \infty$ и поэтому не являются абсолютно интегрируемыми, но их спектральные плотности возможно получить с использованием обобщённых функций:

1) *экспоненциальный сигнал.* Спектральная функция $X_{\text{exp}}(\omega)$ сигнала $x(t) = Ae^{j\omega_0 t}$ по формулам (2.27) и (2.37) принимает вид:

$$X_{\text{exp}}(\omega) = A \int_{-\infty}^{\infty} e^{-j(\omega-\omega_0)t} dt = 2\pi A\delta(\omega - \omega_0); \quad (2.41)$$

2) *косинусоидальный сигнал.* По формуле Эйлера функция

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

а её спектральная плотность с учетом (2.41) принимает вид:

$$X_{\cos}(\omega) = \pi A \delta(\omega - \omega_0) + \pi A \delta(\omega + \omega_0); \quad (2.42)$$

3) *синусоидальный сигнал*. По формуле Эйлера функция

$$x(t) = A \sin \omega_0 t = \frac{A}{2j} (e^{j\omega_0 t} - e^{-j\omega_0 t}),$$

а её спектральная плотность с учетом (2.41) имеет вид:

$$X_{\sin}(\omega) = -j\pi A \delta(\omega - \omega_0) + j\pi A \delta(\omega + \omega_0). \quad (2.43)$$

Проанализируем полученные результаты. При разложении функции $x(t) = A \cos \omega_0 t$ в вещественный ряд Фурье спектр сигнала содержит единственную линию с амплитудой A на частоте ω_0 , а при разложении того же сигнала в комплексный ряд Фурье – две линии с амплитудами $A/2$ на частотах $-\omega_0$ и ω_0 . Спектральная плотность косинусоидального сигнала вещественна и состоит из двух δ -функций с условной амплитудой πA на тех же частотах $-\omega_0$ и ω_0 (рис. 2.24, а). А вот спектральная плотность синусоидального сигнала оказывается мнимой, и δ -функции на частотах $-\omega_0$ и ω_0 имеют разные знаки (рис. 2.24, б). Из $X_{\cos}(\omega)$ и $X_{\sin}(\omega)$ легко составить спектральную плотность экспоненциального сигнала. Умножая $X_{\sin}(\omega)$ на j (рис. 2.24, в) и суммируя с $X_{\cos}(\omega)$, получим $X_{\exp}(\omega)$ (рис. 2.24, г), соответствующую формуле (2.41).

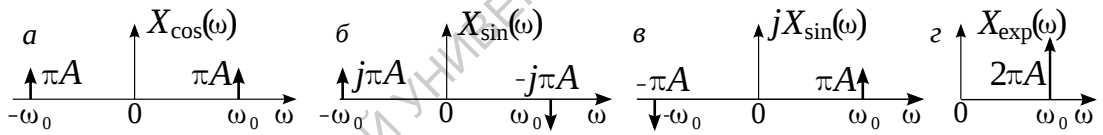


Рис. 2.24. Спектральные плотности гармонических сигналов с частотой ω_0

4) *произвольный периодический сигнал*. Используя соотношение (2.27) для спектральной плотности произвольных сигналов и полученную выше формулу (2.41), вместо разложения произвольного периодического сигнала в ряд Фурье можно получить эквивалентное представление периодического сигнала его спектральной плотностью в виде:

$$X(\omega) = \int_{-\infty}^{\infty} \left[\sum_{k=-\infty}^{\infty} \dot{C}_k e^{jk\omega_1 t} \right] e^{-j\omega t} dt = \sum_{k=-\infty}^{\infty} \dot{C}_k \int_{-\infty}^{\infty} e^{-j(\omega - k\omega_1)t} dt = 2\pi \sum_{k=-\infty}^{\infty} \dot{C}_k \delta(\omega - k\omega_1), \quad (2.44)$$

где ω_1 – первая (основная) гармоника периодического сигнала.

4. Спектральная плотность периодической последовательности δ -импульсов. Дискретные сигналы представляют последовательности равноотстоящих по времени дискретных отсчётов $x(kT) = x[k]$ (см. рис. 1.2, б), где $k = 0, \pm 1, \pm 2, \dots$ – дискретное время, а $T = \Delta_t$ – интервал дискретизации. Дискретный сигнал $x_T(t)$ удобно рассматривать как произведение непрерывной функции $x(t)$ и периодической последовательности δ -импульсов $s_T(t)$. Тогда

$$x_T(t) = x(t)s_T(t) = x(t) \sum_{k=-\infty}^{\infty} \delta(t - kT), \quad x[k] = x(t)\delta(t - kT). \quad (2.45)$$

Проблема Фурье-анализа дискретных сигналов сводится прежде всего к отысканию спектральной плотности последовательности δ -импульсов.

Будем искать спектральную плотность $S_T(\omega)$ функции $s_T(t)$ в два этапа. Сначала рассмотрим спектральную плотность $S_N(\omega)$ конечного числа $2N + 1$ одинаковых импульсов, а затем совершим предельный переход при $N \rightarrow \infty$. Спектральная плотность $S_T(\omega)$ имеет вид:

$$S_N(\omega) = \int_{-\infty}^{\infty} \left(\sum_{k=-N}^N \delta(t - kT) \right) e^{-j\omega t} dt = \sum_{k=-N}^N \int_{-\infty}^{\infty} \delta(t - kT) e^{-j\omega t} dt = \sum_{k=-N}^N e^{-j\omega kT}. \quad (2.46)$$

Если представить (2.46) в виде геометрической прогрессии

$$S_N(\omega) = e^{j\omega NT} + e^{j\omega(N-1)T} + \dots + e^{-j\omega NT}, \quad (2.47)$$

то сумма $2N + 1$ членов прогрессии описывается известной формулой

$$S_N = \frac{e^{j\omega NT} (1 - e^{-j\omega(2N+1)T})}{1 - e^{-j\omega T}} = \frac{e^{j\omega(N+\frac{1}{2})T} - e^{-j\omega(N+\frac{1}{2})T}}{e^{j\omega T/2} - e^{-j\omega T/2}}.$$

Используя формулу Эйлера, получим для $S_N(\omega)$ выражение:

$$S_N(\omega) = \frac{\sin \omega(N + \frac{1}{2})T}{\sin \omega T/2}. \quad (2.48)$$

Согласно (2.48) огибающие обращаются в нуль при $\omega(N + \frac{1}{2})T = \pm k\pi$. Тогда частоты первых нулей огибающих (ω_n) удовлетворяют равенству $\omega_n = \pm\pi/(N + \frac{1}{2})T$, т.е. с ростом N ширина лепестков уменьшается.

Амплитуды главных лепестков одинаковы. Главный лепесток нулевой структуры достигает максимума при $\omega \rightarrow 0$, и в соответствии с (2.47) его величина равна $2N + 1$. Таким образом, амплитуды главных лепестков с ростом N неограниченно возрастают.

Спектральная плотность девяти равноудалённых δ -функций ($N = 4$) изображена на рис. 2.25 и представляет последовательность девяти идентичных многолепестковых импульсных структур, отстоящих на расстоянии $2\pi/T$.

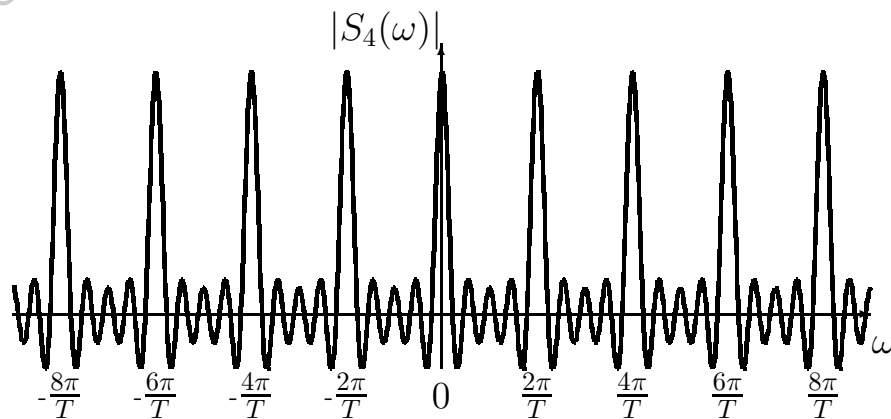


Рис. 2.25. Спектральная плотность последовательности девяти равноудалённых δ -импульсов

Теперь выполним предельный переход при $N \rightarrow \infty$. Тогда каждый импульс спектральной функции трансформируется в δ -функцию, а спектральная плотность бесконечной последовательности одинаковых δ -импульсов во временной области принимает вид бесконечной последовательности одинаковых δ -импульсов в частотной области:

$$S_T(\omega) = \mathbf{F} \left[\sum_{k=-\infty}^{\infty} \delta(t - kT) \right] = \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta(\omega - k\frac{2\pi}{T}). \quad (2.49)$$

5. Спектральная плотность суперпозиции периодического и непериодического сигналов. Спектральное описание суперпозиции перекрывающихся по частотному диапазону периодического (описываемого совокупностью коэффициентов ряда Фурье) и непериодического (описываемого спектральной плотностью) сигналов можно осуществить, если использовать эквивалентное представление периодического сигнала его спектральной плотностью (рис. 2.26). Тогда

$$X_{\Sigma}(\omega) = X_1(\omega) + X_2(\omega),$$

где $X_1(\omega)$ и $X_2(\omega)$ – спектральные плотности периодического и непериодического сигналов соответственно.

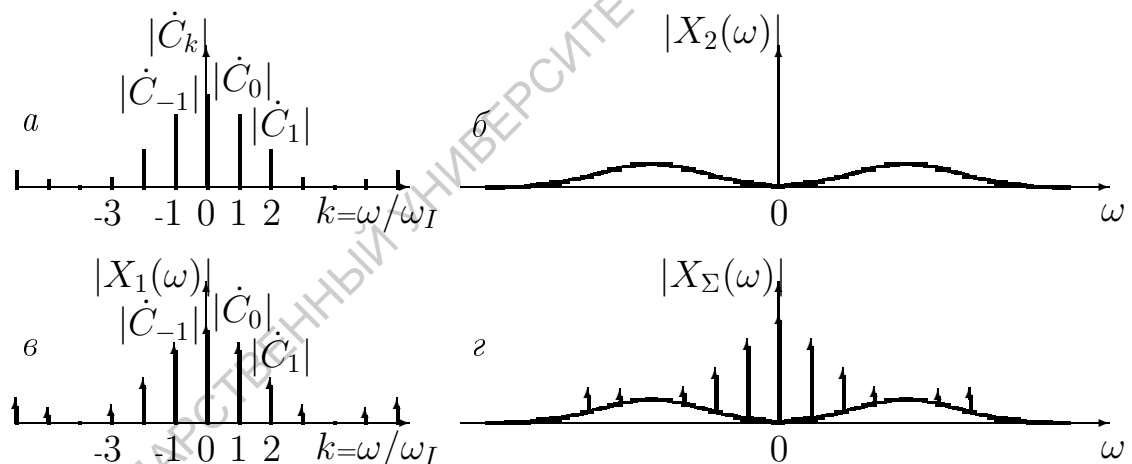


Рис. 2.26. Спектральная плотность суперпозиции периодического и непериодического сигналов: комплексный спектр периодического сигнала $x_1(t)$ (а), спектральная плотность непериодического сигнала $x_2(t)$ (б), спектральная плотность $x_1(t)$ (в), спектральная плотность суммарного сигнала (г)

2.5. Спектральные плотности комплексных функций

Преобразование Фурье должно выполняться и для произвольных комплексных функций. Покажем это.

Пусть $z(t) = x(t) + jy(t)$. Спектральные плотности действительных абсолютно интегрируемых функций $x(t)$ и $y(t)$ удовлетворяют соотношениям

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt, \quad Y(\omega) = \int_{-\infty}^{\infty} y(t)e^{-j\omega t} dt. \quad (2.50)$$

Используя обратные преобразования Фурье, можно записать

$$\begin{aligned} z(t) = x(t) + jy(t) &= \int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega + j \int_{-\infty}^{\infty} Y(\omega)e^{j\omega t}d\omega = \\ &= \int_{-\infty}^{\infty} [X(\omega) + jY(\omega)]e^{j\omega t}d\omega = \int_{-\infty}^{\infty} Z(\omega)e^{j\omega t}d\omega, \end{aligned} \quad (2.51)$$

где $Z(\omega) = \int_{-\infty}^{\infty} z(t)e^{-j\omega t}dt$ – спектральная плотность комплексного сигнала $z(t) = x(t) + jy(t)$.

Хотя представленные соотношения по форме не отличаются от аналогичных преобразований для вещественных сигналов, их спектральные плотности различаются. Пусть $X(\omega) = X_1 + jX_2$, $Y(\omega) = Y_1 + jY_2$, где X_1, Y_1 – вещественные, а X_2, Y_2 – мнимые компоненты спектральных плотностей $X(\omega)$ и $Y(\omega)$. Тогда $Z(\omega) = W_1 + jW_2 = X_1 - Y_2 + j(X_2 + Y_1)$, т.е. спектры комплексных функций теряют свойство симметрии, характерное для вещественных сигналов (см. раздел 2.4.1).

Важнейшей разновидностью комплексных сигналов являются аналитические сигналы.

Аналитический сигнал и его спектральная плотность. Комплексный сигнал $z(t) = x(t) + jy(t)$ называется *аналитическим*⁷, если функции $x(t)$ и $y(t)$ связаны преобразованиями Гильберта

$$y(t) = \mathbf{H}[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \Rightarrow \text{прямое преобразование}, \quad (2.52)$$

$$x(t) = \mathbf{H}^{-1}[y(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{y(\tau)}{\tau - t} d\tau \Rightarrow \text{обратное преобразование}. \quad (2.53)$$

Функцию $y(t)$ часто называют сопряжённой по Гильберту с $x(t)$. Поскольку подынтегральные выражения в (2.52) и (2.53) при $t = \tau$ обращаются в бесконечность, интегралы в обоих случаях понимают в смысле главного значения по Коши

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

В качестве примера найдем по формуле (2.52) сопряжённый сигнал для функции $x(t) = \cos \omega_0 t$. Для этого введем новую переменную $u = t - \tau$ и представим $\cos \omega_0(t - u)$ в виде суммы $\cos \omega_0 t \cos \omega_0 u + \sin \omega_0 t \sin \omega_0 u$. Тогда

⁷ Аналитические сигналы, используемые в радиоэлектронике, представляют весьма частный случай аналитических функций, заданных в математике, так как вместо комплексных рассматривают только вещественный аргумент – время t .

$$\begin{aligned} \mathbf{H}[\cos \omega_0 t] &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\cos \omega_0 \tau}{t - \tau} d\tau = \frac{\cos \omega_0 t}{\pi} \int_{-\infty}^{\infty} \frac{\cos \omega_0 u}{u} du + \frac{\sin \omega_0 t}{\pi} \int_{-\infty}^{\infty} \frac{\sin \omega_0 u}{u} du = \\ &= \sin \omega_0 t \quad \left(\text{так как } \int_{-\infty}^{\infty} \frac{\cos x}{x} dx = 0, \text{ а } \int_{-\infty}^{\infty} \frac{\sin x}{x} dx = \pi^8 \right). \end{aligned}$$

Таким образом, комплексная функция $z(t) = e^{j\omega_0 t} = \cos \omega_0 t + j \sin \omega_0 t$ представляет собой простейший аналитический сигнал. Аналогичные вычисления дают $\mathbf{H}[\sin \omega_0 t] = -\cos \omega_0 t$.

Рассмотрим спектральную плотность аналитических сигналов. Если спектральная плотность вещественного сигнала $x(t)$ есть $X(\omega)$, то спектральную плотность $Y(\omega)$ сопряжённого сигнала $y(t)$ можно получить по теореме о свёртке (см. раздел 2.4, формула (2.30)) как произведение спектральных плотностей сигнала $x(t)$ и функции $g(t) = \frac{1}{\pi t}$ (см. раздел 2.4.3 и формулу (2.40)). Спектральная плотность сопряжённого сигнала $y(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau = \int_{-\infty}^{\infty} x(\tau) g(t-\tau) d\tau$ с учетом (2.40) принимает вид:

$$Y(\omega) = X(\omega)G(\omega) = -j \operatorname{sgn}(\omega)X(\omega) = \begin{cases} -jX(\omega) & \text{при } \omega > 0, \\ jX(\omega) & \text{при } \omega < 0, \end{cases} \quad (2.54)$$

т.е. оказывается мнимой, а её модуль совпадает с модулем спектральной плотности вещественной составляющей. Преобразование Гильберта сохраняет модуль спектральной плотности и лишь изменяет её аргумент: при $\omega < 0$ происходит сдвиг по фазе на $+\pi/2$, при $\omega > 0$ – сдвиг по фазе на $-\pi/2$.

Спектральная плотность аналитического сигнала является суммой спектральных плотностей физического и сопряжённого сигналов, т.е.

$$\begin{aligned} Z(\omega) = \mathbf{F}[z(t)] &= X(\omega) + jY(\omega) = [1 + \operatorname{sgn}(\omega)]X(\omega) = \\ &= \begin{cases} 2X(\omega) & \text{при } \omega > 0, \\ 0 & \text{при } \omega < 0. \end{cases} \end{aligned} \quad (2.55)$$

Итак, спектральная плотность аналитического сигнала является односторонней. Она отлична от нуля только в области положительных частот и обладает удвоенным по сравнению с физическим сигналом модулем.

Свойства преобразований Гильберта. Сформулируем основные свойства преобразований Гильберта, необходимые при работе с аналитическими сигналами.

1. Линейность: $\mathbf{H}[\sum_{i=1}^n (\alpha_i x_i(t))] = \sum_{i=1}^n (\alpha_i \mathbf{H}[x_i(t)]) = \sum_{i=1}^n (\alpha_i y_i(t))$.
2. Ортогональность: $\int_{-\infty}^{\infty} x(t)y(t)dt = 0$.
3. Временной сдвиг: $\mathbf{H}[x(t - \alpha)] = y(t - \alpha)$.

⁸Si(t) = $\int_0^t \frac{\sin x}{x} dx$ представляет интегральный синус, который при $t \rightarrow \infty$ стремится к $\pi/2$.

4. Масштабирование времени: $\mathbf{H}[x(\alpha t)] = y(\alpha t)$.
5. Двукратное преобразование Гильберта: $\mathbf{H}[\mathbf{H}[x(t)]] = -x(t)$.
6. Свойство чётности и нечётности: если $x(t)$ – чётная, то $y(t)$ – нечётная и наоборот.
7. H -преобразование свёртки: $\mathbf{H}[x_1(t)*x_2(t)] = y_1(t)*x_2(t) = x_1(t)*y_2(t)$.
8. Свойство «модуляции» – важное следствие H -преобразования свёртки: если $x(t)$ – низкочастотный сигнал с ограниченным спектром, то

$$\mathbf{H}[x(t) \cos \omega_0 t] = x(t) \sin \omega_0 t,$$

$$\mathbf{H}[x(t) \sin \omega_0 t] = -x(t) \cos \omega_0 t.$$

Взаимная зависимость $x(t)$ и $y(t)$, односторонний характер спектральной плотности аналитических сигналов и свойство «модуляции» играют очень важную роль в теории узкополосных сигналов.

2.6. Корреляционный анализ сигналов и энергетические спектры

Пусть детерминированный сигнал $x(t)$ имеет конечную длительность. Функция

$$r_{xx}(\tau) = 2\pi\varphi(\tau) = \int_{-\infty}^{\infty} x(t)x(t-\tau)dt = \langle x(t)x(t-\tau) \rangle, \quad (2.56)$$

характеризующая взаимную связь значений $x(t)$, разделённых интервалом времени τ , называется *автокорреляционной* (от греч. *autos* – сам и позднелат. *correlatio* – соотношение) функцией (АКФ) сигнала $x(t)$.

АКФ, по существу, усредняет произведение $x(t)x(t-\tau)$ на бесконечном интервале времени и при $\tau = 0$ равна энергии сигнала $x(t)$. Для практической реализации АКФ необходимо осуществить регулируемую задержку, обеспечивающую получение сигнала $x(t-\tau)$, перемножение сигналов $x(t)$ и $x(t-\tau)$ и интегрирование произведения этих сигналов, обеспечивающее вычисление площади их перекрытия, пропорциональной АКФ. Построение АКФ для прямоугольного импульса длительности t_n иллюстрирует рис. 2.27. При $\tau = 0$ (рис. 2.27, а) сигналы полностью перекрываются, что соответствует $r_{xx}(0)$ (рис. 2.27, з). При $\tau \neq 0$ перекрытие сигналов и АКФ уменьшаются, причём запаздывания $\pm\tau$ (рис. 2.27, б и 2.27, в) создают одинаковые значения АКФ. Заметим, что АКФ импульсных сигналов с длительностью t_n обращаются в нуль при задержке $\tau = t_n$. Это позволяет определять длительность

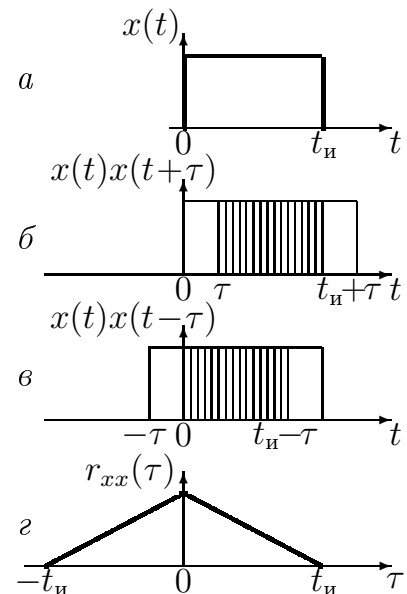


Рис. 2.27. Построение АКФ для прямоугольного импульса

ти сверхкоротких импульсов, измерение которых другими методами подчас просто невозможно.

На рис. 2.28 представлены АКФ радиоимпульса прямоугольной формы

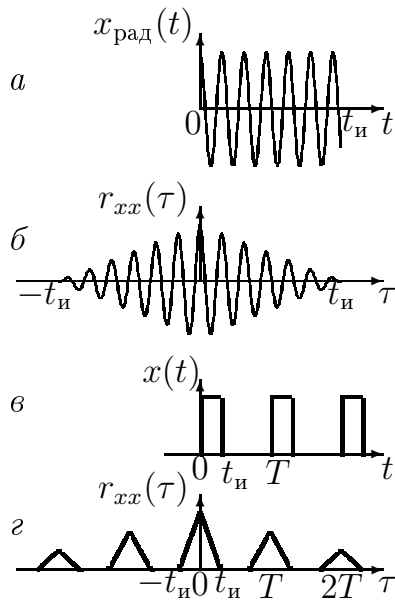


Рис. 2.28. Радиоимпульс (а), «пачка» импульсов (в) и их АКФ (б, г)

и трёх равноудалённых прямоугольных видеоимпульсов («пачки» импульсов). Как и в первом примере (рис. 2.27), при $\tau = 0$ происходит полное перекрытие сигналов, и r_{xx} принимает наибольшее значение. Для $\tau_1 < \tau_2 < \tau_3 < \dots < T$, где T – период несущего колебания, АКФ радиоимпульса сначала уменьшается до нуля (лепестки косинусоиды перекрываются все в меньшей степени), а затем изменяет знак (перекрываются положительные и отрицательные лепестки косинусоид). При дальнейшем увеличении τ число перекрывающихся лепестков сокращается, амплитуда огибающей АКФ линейно уменьшается пока не обратится в нуль. АКФ «пачки» ($n = 3$) импульсов (рис. 2.28, г) при всех τ положительна и содержит пять $(2n - 1)$ лепестков треугольной формы с линейно убывающей амплитудой. Для периодических сигналов усреднение на бесконечном интервале времени адекватно усреднению за период. Поэтому АКФ периодического сигнала можно определить в виде:

$$r_{xx}(\tau) = \frac{1}{T} \int_0^T x(t)x(t - \tau)dt.$$

АКФ периодического сигнала является периодической функцией с тем же периодом, что и $x(t)$. Размерность АКФ тоже изменяется. Теперь она равна средней мощности периодического сигнала. Важнейшее свойство АКФ – чётность. Это свойство нетрудно проиллюстрировать, если ввести замену переменных $u = t - \tau$, $t = u + \tau$. Тогда

$$r_{xx}(\tau) = \varphi(\tau) = \int_{-\infty}^{\infty} x(u + \tau)x(u)du.$$

Если сигнал $x(t)$ обладает спектральной плотностью $X(\omega)$, то $\mathbf{F}[x(-t)] = X^*(\omega)$, а спектральная плотность АКФ $r_{xx}(\omega)$ по теореме о свёртке удовлетворяет равенству

$$r_{xx}(\omega) = \dot{X}(\omega)\dot{X}^*(\omega) = |X(\omega)|^2. \quad (2.57)$$

Функция $r_{xx}(\omega) = |X(\omega)|^2$ является вещественной, имеет размерность $\text{В}^2\text{с}/\text{Гц}$ и называется *энергетическим спектром* или *спектральной плотностью энергии* сигнала $x(t)$.

Автокорреляционная функция детерминированного сигнала и его энергетический спектр связаны преобразованием Фурье:

$$r_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r_{xx}(\omega) e^{j\omega\tau} d\omega, \quad r_{xx}(\omega) = \int_{-\infty}^{\infty} r_{xx}(\tau) e^{-j\omega\tau} d\tau.$$

Эти соотношения часто называют формулами Винера – Хинчина.

Для оценки степени связи между различными сигналами $x(t)$ и $y(t)$ используются *взаимные корреляционные функции* (ВКФ)

$$r_{xy}(\tau) = \int_{-\infty}^{\infty} y(t)x(t - \tau)dt.$$

Если ВКФ $r_{xy}(\tau)$ удовлетворяет условию интегрируемости, то справедливы следующие формулы Винера – Хинчина:

$$r_{xy}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r_{xy}(\omega) e^{j\omega\tau} d\omega, \quad r_{xy}(\omega) = \int_{-\infty}^{\infty} r_{xy}(\tau) e^{-j\omega\tau} d\tau,$$

где $r_{xy}(\omega)$ – *взаимная спектральная плотность* функций $x(t)$ и $y(t)$. Взаимная спектральная плотность является комплексной функцией частоты.

ВКФ может быть асимметричной относительно оси ординат и не обязательно достигает максимума при $\tau = 0$. Значение ВКФ при $\tau = 0$ соответствует взаимной энергии сигналов.

Формулы Винера – Хинчина постулируют взаимно однозначную связь корреляционных и спектральных характеристик сигналов и широко используются при анализе структуры сложных, прежде всего случайных, колебаний.

2.7. Преобразование Лапласа как обобщение преобразования Фурье

Интегральное преобразование Лапласа было введено французским математиком и физиком Пьером Симоном де Лапласом в 1779 г. и стало использоваться в электротехнике в виде *одностороннего преобразования Лапласа* (функция времени задана только для $t \geq 0$) или **L**-преобразования почти через сто лет.

Прямое и обратное преобразование Лапласа. Пусть $x(t)$ существует при $t \geq 0$, но для $t \rightarrow \infty$ не становится исчезающе малой, т.е. не является абсолютно интегрируемой. Введем вспомогательную функцию $x_1(t) = x(t)e^{-\sigma t}$ и выберем параметр σ так, чтобы $x_1(t)$ была абсолютно интегрируемой⁹. Минимальное значение σ , при котором вспомогательная функция абсолютно интегрируема, называется *абсциссой сходимости*.

⁹Ввести абсолютно интегрируемую вспомогательную функцию не всегда возможно. Если $x(t)$ растет быстрее, чем любая экспонента, например $x(t) = e^{t^2}$ (гауссов импульс), то области сходимости не существует.

Итак, функция $x_1(t)$ абсолютно интегрируема и её можно разложить на спектральные составляющие с помощью преобразования Фурье. Спектральная плотность $x_1(t)$ удовлетворяет соотношению

$$\int_{-\infty}^{\infty} x_1(t)e^{-j\omega t} dt = \int_{-\infty}^{\infty} x(t)e^{-(\sigma+j\omega)t} dt = F(\sigma + j\omega). \quad (2.58)$$

Введем комплексную лапласовскую переменную $s = \sigma + j\omega$ и, учитывая, что $x(t)$ и $x_1(t)$ отличны от нуля при $t \geq 0$, получим из (2.46) следующее соотношение:

$$F(s) = \mathbf{L}[x(t)] = \int_0^{\infty} x(t)e^{-st} dt. \quad (2.59)$$

Интегральное преобразование (2.59) называется *прямым преобразованием Лапласа*, а функция $F(s)$, зависящая от комплексной переменной $s = \sigma + j\omega$, – *лапласовским* или *L-изображением* функции $x(t)$.

Интеграл (2.59) по виду напоминает прямое преобразование Фурье (2.27), хотя нижние пределы в интегралах различаются, а вместо мнимой частоты $j\omega$ экспонента в **L**-преобразовании содержит комплексную переменную $s = \sigma + j\omega$, часто называемую *комплексной частотой*.

Для практического использования **L**-преобразования необходимо правило восстановления $x(t)$ по заданной функции комплексной частоты $F(\sigma + j\omega)$ или обращения уже описанного процесса.

Поскольку спектральная плотность вспомогательной абсолютно интегрируемой функции $x_1(t)$ известна ($F(\sigma + j\omega)$), её можно восстановить, используя обратное преобразование Фурье. Тогда

$$x_1(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\sigma + j\omega)e^{j\omega t} d\omega. \quad (2.60)$$

Чтобы перейти от $x_1(t)$ к искомой функции $x(t)$, достаточно умножить обе части равенства (2.60) на $e^{\sigma t}$, внести эту экспоненту под знак интеграла и заменить интегрирование по ω интегрированием по комплексной переменной $s = \sigma + j\omega$ ($d\omega = ds/j$). Тогда получим:

$$x(t) = \mathbf{L}^{-1}[F(s)] = \frac{1}{2\pi j} \int_{\sigma-j\infty}^{\sigma+j\infty} F(s)e^{st} ds. \quad (2.61)$$

Это интегральное соотношение называется *обратным преобразованием Лапласа* и восстанавливает оригинал в виде вещественной функции времени ($x(t)$) по её лапласовскому изображению $F(s)$.

Построенная пара преобразований Лапласа

$$F(s) = \mathbf{L}[x(t)] \quad x(t) = \mathbf{L}^{-1}[F(s)]$$

аналогична паре преобразований Фурье. Но преобразование Фурье представляет сигнал бесконечным числом гармонических составляющих, а преобразование Лапласа – бесконечным числом элементарных экспоненциально затухающих синусоидальных функций различных частот с бесконечно малой амплитудой. Поэтому интегральное преобразование Фурье рассматривают как частный случай двустороннего преобразования

Лапласа при $s = j\omega$, а $F(s)$ иногда называют *спектральной функцией Лапласа*.

Рассмотрим L -изображения широко распространённых функций:

– изображение $F(s)$ функции Хевисайда $\sigma(t)$ имеет вид:

$$F(s) = \mathbf{L}[\sigma(t)] = \int_0^{\infty} e^{-st} dt = -\frac{e^{-st}}{s} \Big|_0^{\infty} = \frac{1}{s}; \quad (2.62)$$

– изображение вещественной экспоненты:

$$F(s) = \mathbf{L}[e^{-\alpha t} \sigma(t)] = \int_0^{\infty} e^{-(\alpha+s)t} dt = -\frac{e^{-(\alpha+s)t}}{s+\alpha} \Big|_0^{\infty} = \frac{1}{s+\alpha}; \quad (2.63)$$

– изображение мнимой экспоненты:

$$F(s) = \mathbf{L}[e^{\pm j\omega t} \sigma(t)] = \int_0^{\infty} e^{-(s \mp j\omega)t} dt = -\frac{e^{-(s \mp j\omega)t}}{s \mp j\omega} \Big|_0^{\infty} = \frac{1}{s \mp j\omega}; \quad (2.64)$$

– изображение функции $\cos \omega t$:

$$\begin{aligned} F(s) &= \mathbf{L}[\cos \omega t \cdot \sigma(t)] = \frac{1}{2} \int_0^{\infty} (e^{j\omega t} + e^{-j\omega t}) e^{-st} dt = \\ &= \frac{1}{2} \left(\frac{1}{s - j\omega} + \frac{1}{s + j\omega} \right) = \frac{s}{s^2 + \omega^2}; \end{aligned} \quad (2.65)$$

– изображение функции $\sin \omega t$:

$$\begin{aligned} F(s) &= \mathbf{L}[\sin \omega t \cdot \sigma(t)] = \frac{1}{2j} \int_0^{\infty} (e^{j\omega t} - e^{-j\omega t}) e^{-st} dt = \\ &= \frac{1}{2j} \left(\frac{1}{s - j\omega} - \frac{1}{s + j\omega} \right) = \frac{\omega}{s^2 + \omega^2}. \end{aligned} \quad (2.66)$$

Основные свойства преобразований Фурье и Лапласа. Наиболее важные теоремы преобразований Фурье и Лапласа совпадают при замене комплексной переменной s на $j\omega$, а лапласовских изображений – спектральными функциями.

1. *Теорема линейности.* Изображение (спектральная плотность) алгебраической суммы сигналов есть алгебраическая сумма их изображений (спектральных плотностей). Если $F_i(s) = \mathbf{L}[x_i(t)]$, а $X_i(j\omega) = \mathbf{F}[x_i(t)]$, то

$$F(s) = \mathbf{L}\left[\sum_{i=1}^N \alpha_i x_i(t)\right] = \sum_{i=1}^N \alpha_i F_i(s), \quad X(j\omega) = \mathbf{F}\left[\sum_{i=1}^N \alpha_i x_i(t)\right] = \sum_{i=1}^N \alpha_i X_i(j\omega). \quad (2.67)$$

2. *Теорема о спектральной плотности вещественных процессов.* Спектральная плотность произвольного вещественного процесса удовлетворяет условию

$$X(-j\omega) = X^*(j\omega),$$

т.е. реальная часть спектральной плотности является чётной, а мнимая – нечётной функцией. Для доказательства теоремы достаточно представить $X(\pm j\omega)$ в виде обратного преобразования Фурье и записать экспоненту в виде алгебраической суммы косинуса и синуса.

3. *Теорема об изменении времени его масштаба сигнала.* Временное растяжение сигналов $x(kt)$ $0 < k < 1$ вызывает сжатие по частоте изображения или спектральной плотности сигнала и увеличение всех спектральных составляющих в k раз (рис. 2.29):

$$\mathbf{L}[x(kt)] = \frac{1}{k} F\left(\frac{s}{k}\right); \quad \mathbf{F}[x(kt)] = \frac{1}{k} X\left(\frac{j\omega}{k}\right).$$

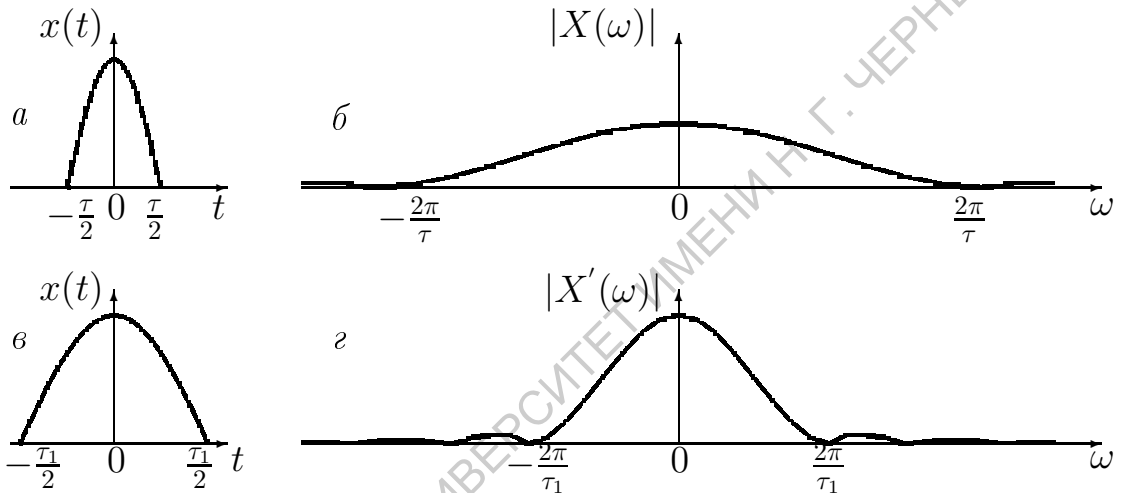


Рис. 2.29. Огибающие исходного (а) и растянутого во времени (в) видеоимпульсов и их спектральные плотности ((б) и (г) соответственно)

4. *Теорема о смещении сигнала во времени.* Если $F(s) = \mathbf{L}[x(t)]$, а $X(j\omega) = \mathbf{F}[x(t)]$, то

$$\mathbf{L}[x(t - \tau)] = e^{-s\tau} F(s), \quad \mathbf{F}[x(t - \tau)] = e^{-j\omega\tau} X(j\omega). \quad (2.68)$$

Пример. Изображение прямоугольного импульса длительности τ . Такой импульс является суммой двух сдвинутых во времени функций Хевисайда (см. рис. 2.7, е). С учетом формул (2.67) и (2.68)

$$\mathbf{L}[\sigma(t) - \sigma(t - \tau)] = \frac{1}{s} - \frac{e^{-s\tau}}{s} = \frac{1}{s}(1 - e^{-s\tau}). \quad (2.69)$$

Замещая s на $j\omega$, для спектральной плотности сигнала получим:

$$X(\omega) = \frac{1}{j\omega}(1 - e^{-j\omega\tau}) = \tau \frac{\sin \omega\tau/2}{\omega\tau/2} e^{-j\omega\tau/2} = \tau \text{sinc}(f\tau) e^{-j\pi f\tau}.$$

Полученное выражение отличается от (2.34) дополнительным фазовым множителем, так как в разделе 2.4 импульс располагался симметрично относительно начала координат, а в данном случае сдвинут во времени на $\tau/2$. Модуль спектральной функции не изменился.

5. *Теорема о смещении переменной в частотной области.* Для смещения спектра, т.е. частоты ω на ω_0 или комплексной переменной s на s_0 , достаточно умножить исходную функцию $x(t)$ на экспоненту $e^{-j\omega_0 t}$ или $e^{-s_0 t}$ соответственно

$$X(\omega \pm \omega_0) = \mathbf{F}[x(t)e^{\mp j\omega_0 t}], \quad F(s \pm s_0) = \mathbf{L}[x(t)e^{\mp s_0 t}]. \quad (2.70)$$

Итак, при умножении сигнала $x(t)$ на экспоненциальный множитель спектр сигнала смещается в частотной области, а при умножении спектра сигнала на экспоненциальный множитель сигнал смещается во времени. Из этих теорем следует важный частный случай: изображения синусоиды и косинусоиды с экспоненциально затухающими амплитудами с учетом формул (2.55) и (2.58) имеют вид:

$$\mathbf{L}[e^{-\alpha t} \cos \omega t] = \frac{s + \alpha}{(s + \alpha)^2 + \omega^2}, \quad \mathbf{L}[e^{-\alpha t} \sin \omega t] = \frac{\omega}{(s + \alpha)^2 + \omega^2}. \quad (2.71)$$

(Теоремы 1 – 5 предлагается доказать самостоятельно.)

6. *Теорема об изображении свёртки двух функций.* Для спектральных плотностей сигналов теорема была доказана в разделе 2.4. Чтобы записать её выражение в лапласовских переменных, казалось бы достаточно заменить в формуле (2.31) $j\omega$ на s . Однако интеграл свёртки в лапласовских переменных определяется иначе. Поэтому приведем доказательство

полностью. Пусть $\mathbf{L}[x(t)] = X(s)$, $\mathbf{L}[y(t)] = Y(s)$. $f(t) = \int_0^t x(\tau)y(t - \tau)d\tau$ – свёртка этих функций. Тогда

$$\begin{aligned} \mathbf{L}[f(t)] &= \int_0^{\infty} \left(\int_0^t x(\tau)y(t - \tau)d\tau \right) e^{-st} dt = \int_0^{\infty} x(\tau) e^{-s\tau} d\tau \int_0^{\infty} y(t - \tau) e^{-s(t - \tau)} dt = \\ &= \int_0^{\infty} x(\tau) e^{-s\tau} d\tau \int_0^{\infty} y(\xi) e^{-s\xi} d\xi, \end{aligned}$$

где $\xi = t - \tau$, $d\xi = dt$. Отсюда

$$\mathbf{L} \left[\int_0^t x(\tau)y(t - \tau)d\tau \right] = \int_0^{\infty} x(\tau) e^{-s\tau} d\tau \int_0^{\infty} y(\xi) e^{-s\xi} d\xi = X(s)Y(s). \quad (2.72)$$

7. *Теоремы о дифференцировании сигналов.* Спектральная плотность первой производной сигнала является произведением спектральной плотности исходного сигнала и оператора дифференцирования $j\omega$:

$$\mathbf{F} \left(\frac{dx}{dt} \right) = \int_{-\infty}^{\infty} \frac{dx}{dt} e^{-j\omega t} dt = x(t) \cdot e^{-j\omega t} \Big|_{-\infty}^{\infty} + j\omega \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt = j\omega X(\omega).$$

Это соотношение – результат интегрирования по частям и обращения в нуль первого слагаемого, так как абсолютно интегрируемая функция $x(t) \rightarrow 0$ при $t \rightarrow \pm\infty$. Оператор дифференцирования $j\omega$ уже встречался в методе комплексных амплитуд при дифференцировании гармонических сигналов. Умножение $X(\omega)$ на $j\omega$ свидетельствует об обогащении спектра высокочастотными гармониками, так как амплитуды спектральных гармоник возрастают пропорционально их частотам. Теорема о дифференцировании обобщается на случай производной n -го порядка. Тогда

$$\mathbf{F} \left(\frac{d^n x}{dt^n} \right) = (j\omega)^n X(\omega). \quad (2.73)$$

Для *изображения* первой производной сигнала теорема изменяется: из произведения изображения исходного сигнала и оператора дифференцирования s вычитается значение сигнала в начальный момент времени $t=0$. В результате интегрирования по частям получаем

$$\mathbf{L}\left(\frac{dx}{dt}\right) = \int_0^{\infty} \frac{dx}{dt} e^{-st} dt = x(t)e^{-st} \Big|_0^{\infty} + s \int_0^{\infty} x(t)e^{-st} dt = s\mathbf{L}[x(t)] - x(0). \quad (2.74)$$

Если многократно использовать формулу (2.74), то

$$\mathbf{L}\left[\frac{d^n x}{dt^n}\right] = s^n \mathbf{L}[x(t)] - s^{n-1}x(0) - s^{n-2} \frac{dx}{dt} \Big|_{t=0} - \dots - s \frac{d^{n-2}x}{dt^{n-2}} \Big|_{t=0} - \frac{d^{n-1}x}{dt^{n-1}} \Big|_{t=0}.$$

Изображение временной производной n -го порядка содержит кроме начального значения функции ещё и начальные значения её производных до $(n-1)$ -го порядка включительно. Поскольку преобразование Лапласа по определению применяется начиная с момента $t=+0$, начальные значения сигнала и его производных рассматриваются при $t=+0$. Из рассмотренных теорем как частный случай при нулевых начальных условиях и $s=j\omega$ получается теорема (2.73).

Полученные преобразования Фурье и Лапласа для обыкновенных производных сигналов позволяют свести линейные дифференциальные уравнения к алгебраическим уравнениям относительно $j\omega$ или s , при этом в лапласовских изображениях учитываются начальные условия задачи.

8. *Теорема об изображении и спектральной плотности определённого интеграла.* Изображение (спектральная плотность) определённого

интеграла с переменным верхним пределом $\Psi(t) = \int_0^t x(t) dt$ является про-

изведением спектральной характеристики исходного сигнала и оператора интегрирования $1/s$ или $1/j\omega$. Изображение функции $\Psi(t)$ получим, интегрируя по частям. Тогда

$$\mathbf{L}(\Psi(t)) = \int_0^{\infty} \Psi(t)e^{-st} dt = -\Psi(t) \frac{e^{-st}}{s} \Big|_0^{\infty} + \frac{1}{s} \int_0^{\infty} \frac{d\Psi}{dt} e^{-st} dt = \frac{\mathbf{L}[x(t)]}{s}.$$

Замещая s на $j\omega$ и переходя от \mathbf{L} - к \mathbf{F} -преобразованию, получим выражение для спектральной плотности определённого интеграла:

$$\mathbf{F}(\Psi(t)) = \frac{1}{j\omega} X(\omega). \quad (2.75)$$

И изображение, и спектральная плотность определённого интеграла не зависят от начальных условий.

9. *Теорема о производной изображения.* Производная изображения есть изображение произведения оригинала и переменной t со знаком минус, а производная порядка n – результат многократного дифференцирования:

$$\frac{dF(s)}{ds} = \mathbf{L}[-tx(t)], \quad \frac{d^n F(s)}{ds^n} = \mathbf{L}[(-t)^n x(t)]. \quad (2.76)$$

Первую формулу можно получить, дифференцируя соотношение (2.59) по комплексной переменной s , вторую – после дифференцирования предыдущих производных.

Теорема о производной изображения имеет важные следствия, обычно используемые при восстановлении оригиналов по изображениям.

9 а. Если $x(t) = \sigma(t)$, то

$$\mathbf{L}[t\sigma(t)] = \frac{1}{s^2}, \quad \mathbf{L}[t^2\sigma(t)] = \frac{2!}{s^3}, \dots, \quad \mathbf{L}[t^n\sigma(t)] = \frac{n!}{s^{n+1}}. \quad (2.77)$$

9 б. Если $x(t) = e^{-\alpha t}$, то

$$\mathbf{L}[te^{-\alpha t}] = \frac{1}{(s+\alpha)^2}, \quad \mathbf{L}[t^2e^{-\alpha t}] = \frac{2!}{(s+\alpha)^3}, \dots, \quad \mathbf{L}[t^ne^{-\alpha t}] = \frac{n!}{(s+\alpha)^{n+1}}. \quad (2.78)$$

10. Теоремы о начальном и предельном значениях функции $x(t)$. Поведение изображения производной функции $x(t)$, не содержащей особенностей при $t \rightarrow 0$, определяет при $s \rightarrow \infty$ начальное значение оригинала, а при $s \rightarrow 0$ – предельное значения функции ($t \rightarrow \infty$).

По формуле (2.74) изображение производной оригинала $x(t)$ запишется как

$$\mathbf{L}\left(\frac{dx}{dt}\right) = \int_0^{\infty} \frac{dx}{dt} e^{-st} dt = s\mathbf{L}[x(t)] - x(0).$$

При $s \rightarrow \infty$ интеграл обращается в нуль ($e^{-st} = 0$), а начальное значение интеграла принимает вид:

$$x(0) = \lim_{s \rightarrow \infty} [sF(s)],$$

т.е. начальное значение оригинала $x(t)$ определяется без нахождения обратного преобразования Лапласа.

Если в изображении производной положить $s = 0$, то соотношение принимает вид:

$$\int_0^{\infty} \frac{dx}{dt} dt = \lim_{t \rightarrow \infty} \int_0^t dx(t) = \lim_{t \rightarrow \infty} [x(t) - x(0)] = \lim_{s \rightarrow 0} [sF(s) - x(0)],$$

$$\lim_{t \rightarrow \infty} x(t) = \lim_{s \rightarrow 0} sF(s),$$

т.е. предельное значение $x(t)$ определяется без вычисления обратного преобразования.

2.8. Регулярные (детерминированные) волны

Передача различной информации: сообщений, управляющих сигналов, массивов данных на большие расстояния в радиоэлектронике и связи осуществляется через свободное околоземное или космическое пространство, а также через специализированные линии передачи с помощью волновых процессов, имеющих электромагнитную природу. Чаще всего такие электромагнитные (радио- или оптические) волны являются детерминированными, а стохастические характеристики у них возникают из-за неоднородности сред распространения или случайных отражений от границ. Волновой характер распространения электромагнитных полей следует непосредственно из уравнений Максвелла и составляет одну из фундаментальных основ радиофизики.

2.8.1. Общие сведения о волновых процессах

Волновые уравнения и их решения. *Волнами* называются изменения состояния среды или поля, распространяющиеся в пространстве с конечной скоростью. Детерминированные волны – это чаще всего периодические процессы во времени и пространстве. Различают *продольные* волны, когда колебания происходят в направлении распространения волн, и *поперечные* волны, если колебания происходят в плоскости, перпендикулярной направлению распространения. Волновые процессы описываются дифференциальным уравнением в частных производных – *волновым уравнением*:

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2 \Phi}{\partial t^2} = 0,$$

где $\Phi(x, y, z, t)$ – скалярная или векторная функция координат и времени, задающая состояние среды или поля, v – скорость распространения волны. Это трёхмерное волновое уравнение. В радиофизике часто используется одномерное волновое уравнение. Для волны, распространяющейся в направлении оси Oz , имеем:

$$\frac{\partial^2 \Phi}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2 \Phi}{\partial t^2} = 0. \quad (2.79)$$

Перепишем уравнение (2.79) в следующем виде:

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{v^2} \frac{\partial^2}{\partial t^2} \right) \Phi = \left(\frac{\partial}{\partial z} - \frac{1}{v} \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial z} + \frac{1}{v} \frac{\partial}{\partial t} \right) \Phi = 0$$

и введем новые переменные

$$\xi = z - vt, \quad \eta = z + vt.$$

Тогда

$$\frac{\partial}{\partial z} = \frac{\partial}{\partial \xi} \frac{\partial \xi}{\partial z} + \frac{\partial}{\partial \eta} \frac{\partial \eta}{\partial z} = \frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}, \quad \frac{1}{v} \frac{\partial}{\partial t} = \frac{\partial}{\partial \xi} \frac{1}{v} \frac{\partial \xi}{\partial t} + \frac{\partial}{\partial \eta} \frac{1}{v} \frac{\partial \eta}{\partial t} = \frac{\partial}{\partial \eta} - \frac{\partial}{\partial \xi},$$

$$\frac{\partial}{\partial z} - \frac{1}{v} \frac{\partial}{\partial t} = 2 \frac{\partial}{\partial \xi}, \quad \frac{\partial}{\partial z} + \frac{1}{v} \frac{\partial}{\partial t} = 2 \frac{\partial}{\partial \eta}, \quad \frac{\partial^2 \Phi}{\partial \xi \partial \eta} = 0,$$

так как $\frac{\partial \xi}{\partial z} = \frac{\partial \eta}{\partial z} = 1$ при $t = \text{const}$ и $\frac{1}{v} \frac{\partial \eta}{\partial t} = -\frac{1}{v} \frac{\partial \xi}{\partial t} = 1$ при $z = \text{const}$.

Отсюда следует, что общим решением волнового уравнения (2.79) является выражение

$$\Phi(z, t) = \Phi_1(\xi) + \Phi_2(\eta) = \Phi_1(z - vt) + \Phi_2(z + vt),$$

т.е. $\Phi(z, t)$ представляет сумму двух линейно независимых функций.

Рассмотрим значения $\Phi(z - vt)$ в двух сечениях:

$$\Phi(0, t) = \Phi(t), \quad \Phi(z_1, t) = \Phi(z_1 - vt).$$

На рис. 2.30 временные изменения $\Phi(z, t)$ в плоскости z_1 (кривая 2) повторяют временные изменения этой функции в плоскости $z=0$

(кривая 1) с запаздыванием на время $\Delta t = z_1/v$, т.е. функция $\Phi(t)$ перемещается, сохраняя свою форму, в сторону возрастающих z со скоростью v и, таким образом, представляет волновое движение. Очевидно, что функция $\Phi_2(z + vt)$ также описывает волновое движение со скоростью v , но в противоположном направлении. Таким образом, получено решение волнового уравнения в виде двух *бегущих* навстречу друг другу волн. Величина $(z \pm vt)$ называется *фазой волны*, а скорость перемещения фронта волны $(z \pm vt = \text{const})$ — *фазовой скоростью* волны

$$v_\Phi = \frac{dz}{dt} = \pm v.$$

Волновое уравнение и его решения можно записать и для сферической волны:

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Phi}{\partial r} \right) - \frac{1}{v^2} \frac{\partial^2 \Phi}{\partial t^2} = 0,$$

$$\Phi(r, t) = \frac{1}{r} \Phi_1(vt - r) + \frac{1}{r} \Phi_2(vt + r).$$

Последние соотношения используются в теории антенн.

Электромагнитные волны. Как известно, в пространстве, окружающем движущиеся заряды или вообще лишённом зарядов и токов, существует электрическое \mathbf{E} и магнитное \mathbf{H} поля, удовлетворяющие уравнениям Максвелла. Для пустого пространства они имеют вид:

$$\begin{aligned} \text{rot} \mathbf{E} &= -\mu_0 \mu \frac{\partial \mathbf{H}}{\partial t}, & \text{div} \mathbf{E} &= 0, \\ \text{rot} \mathbf{H} &= \varepsilon_0 \varepsilon \frac{\partial \mathbf{E}}{\partial t}, & \text{div} \mathbf{H} &= 0, \end{aligned} \quad (2.80)$$

где ε_0 и μ_0 — электрическая и магнитная постоянные, ε и μ — относительные диэлектрическая и магнитная проницаемости среды. Уравнения (2.80) связывают временное изменение электрического поля \mathbf{E} с пространственным изменением магнитного поля \mathbf{H} и наоборот. В пустом пространстве электрическое \mathbf{E} и магнитное \mathbf{H} поля удовлетворяют волновому уравнению. Покажем это.

Возьмём ротор от $\text{rot} \mathbf{E} = -\mu_0 \mu \frac{\partial \mathbf{H}}{\partial t}$. Используя векторное равенство $\text{rot rot} = \nabla \text{div} - \nabla^2$ и изменяя порядок операций вычисления ротора и дифференцирования по времени, получим:

$$\text{rot rot} \mathbf{E} = \nabla \text{div} \mathbf{E} - \nabla^2 \mathbf{E} = -\mu_0 \mu \frac{\partial}{\partial t} \text{rot} \mathbf{H} = -\frac{\varepsilon \mu}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2},$$

где $c = 1/\sqrt{\varepsilon_0 \mu_0} \cong 3 \cdot 10^8$ м/сек — электродинамическая постоянная, совпадающая со скоростью света в вакууме. Учитывая, что $\text{div} \mathbf{E} = 0$, $\nabla^2 = \Delta$ — оператор Лапласа, запишем волновое уравнение для вектора \mathbf{E} как

$$\Delta \mathbf{E} - \frac{1}{v^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \quad (2.81)$$

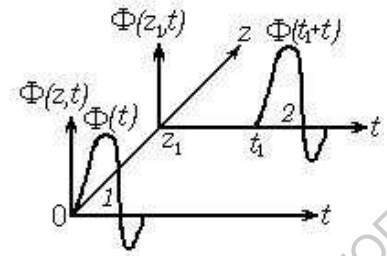


Рис. 2.30. Пространственно-временная диаграмма волны

где $v = c/\sqrt{\varepsilon\mu}$ – фазовая скорость электромагнитной волны. Аналогичное уравнение справедливо для вектора \mathbf{H} :

$$\Delta \mathbf{H} - \frac{1}{v^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0. \quad (2.82)$$

Важнейший результат уравнений Максвелла состоит в существовании электромагнитного поля как самостоятельного физического объекта. Самостоятельная реальность поля видна уже из того, что уравнения Максвелла допускают решение и в пустоте. Однако не следует думать, что электромагнитное поле берется «ниоткуда». Конечно, существует возмущение, явившееся источником волн. Но оно расположено где-то очень далеко, и установить причинно-следственные связи непросто, да и не нужно. Заметим ещё, что указать источники реликтовых микроволновых волн (тепловых и потому не детерминированных) едва ли вообще возможно.

Энергия, переносимая волной через единицу площади, перпендикулярной направлению распространения, за единицу времени определяется вектором Пойтинга $[\mathbf{E}\mathbf{H}]$, а средняя, переносимая за период волны T , энергия называется *интенсивностью электромагнитной волны*

$$I = \frac{1}{T} \left| \int_0^T [\mathbf{E}\mathbf{H}] dt \right|.$$

2.8.2. Основные свойства электромагнитных волн

Рассмотрим решение уравнений (2.81) и (2.82) в виде плоской волны, распространяющейся в направлении оси Oz . Поскольку среда распространения однородная, а волна плоская, поля в разных точках фронта волны не будут зависеть от координат фронта x и y .

Распишем $\text{rot} \mathbf{E}$ и $\text{rot} \mathbf{H}$ и будем учитывать, что координаты полей не зависят от x и y . Тогда

$$\begin{aligned} \text{rot}_x \mathbf{E} &= \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial E_y}{\partial z} = -\mu_0 \mu \frac{\partial H_x}{\partial t}, & \text{rot}_x \mathbf{H} &= -\frac{\partial H_y}{\partial z} = \varepsilon_0 \varepsilon \frac{\partial E_x}{\partial t}, \\ \text{rot}_y \mathbf{E} &= \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = \frac{\partial E_x}{\partial z} = -\mu_0 \mu \frac{\partial H_y}{\partial t}, & \text{rot}_y \mathbf{H} &= \frac{\partial E_x}{\partial z} = \varepsilon_0 \varepsilon \frac{\partial H_y}{\partial t}, \\ \text{rot}_z \mathbf{E} &= \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = 0, & \text{rot}_z \mathbf{H} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = 0. \end{aligned}$$

Из двух последних уравнений $\frac{\partial H_z}{\partial t} = 0$ и $\frac{\partial E_z}{\partial t} = 0$. Это означает, что продольные компоненты полей могут быть только постоянными. Таким образом, получен важный результат: электрическое и магнитное поля плоской волны сосредоточены в плоскости фронта волны, и волна называется *поперечной электромагнитной волной* или *T-волной* (от англ. *transversal* – поперечный). Рассматривая остальные четыре уравнения, можно сделать ещё один важный вывод: компоненты электрического поля E_x и E_y , так же как и составляющие магнитного поля H_x и H_y , не

связаны между собой, т.е. компоненты E_x и H_y можно считать координатами одной плоской волны, а E_y и H_x – другой. При этом векторы \mathbf{v} , \mathbf{E} , \mathbf{H} для той и другой волны образуют правовинтовую систему.

Монохроматические волны. Простейшим и в то же время наиболее представительным типом плоских электромагнитных волн является *монохроматическая* или *гармоническая* волна, у которой векторы \mathbf{E} и \mathbf{H} совершают гармонические колебания одинаковой частоты:

$$\mathbf{E}(\mathbf{r}, t) = \operatorname{Re}\{\mathbf{E}_m \exp[j(\omega t - \mathbf{k}\mathbf{r})]\}, \quad \mathbf{H}(\mathbf{r}, t) = \operatorname{Re}\{\mathbf{H}_m \exp[j(\omega t - \mathbf{k}\mathbf{r})]\}, \quad (2.83)$$

где $\mathbf{E}_m, \mathbf{H}_m$ – векторы комплексных амплитуд полей; ω – круговая частота волны; \mathbf{r} – радиус-вектор рассматриваемой точки; $(\omega t - \mathbf{k}\mathbf{r})$ – фаза волны.

Вектор $\mathbf{k} = \frac{\omega}{v}\mathbf{n} = \frac{2\pi}{\lambda}\mathbf{n}$ называется *волновым вектором*, а $k = |\mathbf{k}| = \omega/v$ – *волновым числом*. Здесь \mathbf{n} – единичный вектор направления распространения волны; v – фазовая скорость волны; $\lambda = vT$ – длина волны; T – период колебаний.

Вектор \mathbf{E}_m (\mathbf{H}_m) можно записать как $\mathbf{E}_m = \mathbf{E}_1 \exp(j\varphi_1) + \mathbf{E}_2 \exp(j\varphi_2)$, где \mathbf{E}_1 и \mathbf{E}_2 – взаимно ортогональные вещественные векторы, лежащие в плоскости фронта волны, например \mathbf{E}_{mx} и \mathbf{E}_{my} , а $\varphi_1 = \varphi_x$, $\varphi_2 = \varphi_y$ – их фазы. Тогда

$$E_x(z, t) = E_{mx} \cos(\omega t - kz + \varphi_x), \quad E_y(z, t) = E_{my} \cos(\omega t - kz + \varphi_y) \quad (2.84)$$

Легко показать, что для монохроматических волн волновые уравнения (2.81) и (2.82) превращаются в уравнения Гельмгольца

$$\Delta \mathbf{E} + k^2 \mathbf{E} = 0, \quad \Delta \mathbf{H} + k^2 \mathbf{H} = 0. \quad (2.85)$$

Плоские монохроматические волны, как и гармонические колебания, не пригодны для передачи сообщений. Произвольные немонахроматические волны обладают более сложной структурой и могут быть представлены совокупностью плоских монохроматических волн с различными волновыми векторами, частотами, амплитудами и начальными фазами. Такое представление аналогично спектральному представлению негармонических колебаний и поэтому часто называется *спектральным разложением* немонахроматических волн.

Немонахроматические волны. Волновые «пакеты». Пусть немонахроматическая волна является наложением монохроматических волн с частотами, заключёнными в интервале $\omega_0 - \frac{\Delta\omega}{2} \leq \omega \leq \omega_0 + \frac{\Delta\omega}{2}$, $\Delta\omega \ll \omega_0$ и все волны в этом интервале имеют одинаковую амплитуду $\mathbf{E}_m(\omega) = \mathbf{E}_{m0}$. Тогда результирующая волна имеет вид:

$$\mathbf{E}(z, t) = \mathbf{E}_{m0} \int_{\omega_0 - \frac{\Delta\omega}{2}}^{\omega_0 + \frac{\Delta\omega}{2}} \exp[-j(\omega t - k(\omega)z)] d\omega.$$

Поскольку частоты парциальных волн сосредоточены в малом интервале $\Delta\omega$, функцию $k(\omega)$ можно разложить в ряд по степеням $(\omega - \omega_0)$ и, ограничиваясь первой производной от $k(\omega)$, представить её в виде $k(\omega) = k(\omega_0) + \left(\frac{dk}{d\omega}\right)_{\omega_0} (\omega - \omega_0)$. Введем замену переменных $\xi = (\omega - \omega_0)$ и представим результирующую волну как

$$\begin{aligned}
\mathbf{E}(z, t) &= \mathbf{E}_{m0} \exp[-j(\omega_0 t - k(\omega_0)z)] \int_{-\frac{\Delta\omega}{2}}^{\frac{\Delta\omega}{2}} \exp\left\{-j\xi \left[t - \left(\frac{dk}{d\omega} \right)_{\omega_0} z \right]\right\} d\xi = \\
&= \mathbf{E}_{m0} \frac{2 \sin \left\{ \left[t - \left(\frac{dk}{d\omega} \right)_{\omega_0} z \right] \frac{\Delta\omega}{2} \right\}}{t - \left(\frac{dk}{d\omega} \right)_{\omega_0} z} \exp[-j(\omega_0 t - k(\omega_0)z)] = \\
&= \mathbf{E}_{m0} \Delta\omega \cdot \text{sinc} \left[\left(t - \left(\frac{dk}{d\omega} \right)_{\omega_0} z \right) \frac{\Delta\omega}{2} \right] \cdot \exp[-j(\omega_0 t - k(\omega_0)z)]. \quad (2.86)
\end{aligned}$$

Это плоская бегущая волна, амплитуда которой пропорциональна функции $\text{sinc}(\dots)$ и представляет собой сконцентрированное в пространстве возмущение, называемое *волновым пакетом*. Функция $\text{sinc}(\dots)$ является огибающей волнового пакета. Огибающая может иметь другую форму: все зависит от характера зависимости $\mathbf{E}_{m0}(\omega)$. Степень концентрации тем выше, чем меньше $\Delta\omega$. Волновой пакет перемещается вдоль оси Oz со скоростью

$$v_{\text{гр}} = \frac{d\omega}{dk}, \quad (2.87)$$

называемой *групповой скоростью*. Скорость $v_{\text{гр}}$ описывает скорость переноса энергии группой волн в среде без поглощения.

Дисперсия. Зависимость фазовой скорости гармонических волн от частоты называется *дисперсией волн* (от лат. *dispersio* – рассеяние), а зависимость $k(\omega)$ – *дисперсионным соотношением*.

Реальные среды в разных областях частот описываются различными дисперсионными соотношениями. Области частот, в которых $dv_{\text{ф}}/d\omega < 0$ и $dv_{\text{ф}}/d\omega > 0$, называются областями *нормальной* и *аномальной* дисперсии соответственно. На рис. 2.31 представлены графики, иллюстрирующие два вида нормальной дисперсии: в области низких частот (кривые 2) и в области высоких частот (кривые 3).

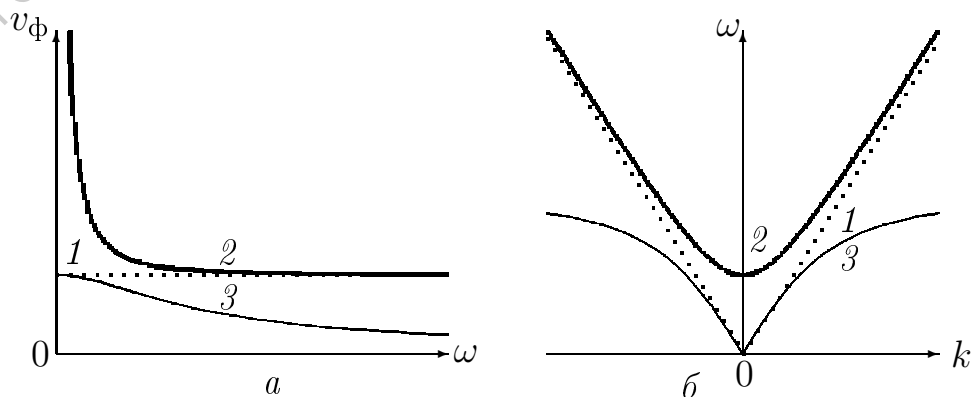


Рис. 2.31. Зависимости $v_{\text{ф}}(\omega)$ (а) и $\omega(k)$ – диаграммы Бриллюэна (б) для двух видов нормальной дисперсии. Пунктирными линиями изображены зависимости для бездисперсной среды

Из графиков видно, что при нормальной дисперсии в области низких частот $v_{\text{ф}}$ конечна, а k отлична от нуля только для частот, превышающих некую критическую частоту ($\omega > \omega_{\text{кр}}$). Критические частоты являются важной характеристикой линий передачи, в частности металлических волноводов, причём конкретной геометрии волновода и каждому типу волн соответствует собственная $\omega_{\text{кр}}$. Физические среды, в которых $v_{\text{ф}}$ не зависит от ω , а волновое число k пропорционально ω , называются *бездисперсными*.

Поляризация плоских детерминированных волн. Обратимся к равенствам (2.84) и введем ортогональные компоненты электрического поля $E_x(z, t)$ и $E_y(z, t)$ в виде вещественных детерминированных функций. В плоскости фронта волны ($z = \text{const}$) имеем

$$\begin{aligned} E_x(t) &= E_{mx} \cos(\omega t + \varphi_x), \\ E_y(t) &= E_{my} \cos(\omega t + \varphi_y). \end{aligned} \quad (2.88)$$

Исключая из (2.88) ωt , получим

$$\left(\frac{E_x}{E_{mx}}\right)^2 + \left(\frac{E_y}{E_{my}}\right)^2 - 2\frac{E_x E_y}{E_{mx} E_{my}} \cos \vartheta - \sin^2 \vartheta = 0, \quad \vartheta = \varphi_y - \varphi_x. \quad (2.89)$$

Это уравнение эллипса. Вектор электрического поля $\mathbf{E}(t)$ детерминированной плоской волны описывает эллипс в плоскости фронта волны (рис. 2.32), и волна называется *эллиптически поляризованной*.

Геометрическими параметрами поляризованной волны являются *угол ориентации эллипса* β – угол наклона большой полуоси эллипса к оси Ox и *угол эллиптичности* χ , тангенс которого равен отношению длин малой полуоси к большой. Связь параметров β и χ с амплитудами и начальными фазами проекций вектора $\mathbf{E}(t)$ выражают соотношения:

$$\operatorname{tg} 2\beta = \frac{2E_{mx} E_{my} \cos(\varphi_y - \varphi_x)}{E_{mx}^2 - E_{my}^2}, \quad \sin 2\chi = \frac{2E_{mx} E_{my} \sin(\varphi_y - \varphi_x)}{E_{mx}^2 + E_{my}^2}. \quad (2.90)$$

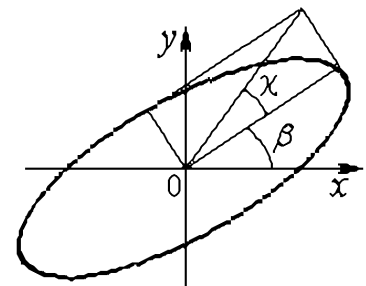


Рис. 2.32. Эллипс поляризации

Поляризацию плоской волны можно определить величинами, имеющими размерность интенсивности, что полезно для экспериментов, так как на практике приходится измерять не напряжённость электрического поля, а мощность электромагнитного излучения. Представляя компоненты вектора $\mathbf{E}(t)$ в комплексном виде $\dot{E}_x(t) = E_{mx} e^{j\varphi_x}$ и $\dot{E}_y(t) = E_{my} e^{j\varphi_y}$ и используя корреляционные функции (см. раздел 2.6), получим:

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} \dot{E}_i(t) \dot{E}_j^*(t + \tau) dt, \quad i, j = x, y.$$

Составим из них квадратную поляризационную матрицу Винера – Вольфа при $\tau = 0$

$$\hat{\mathbf{J}} = \begin{pmatrix} R_{xx}(0), & R_{xy}(0) \\ R_{yx}(0), & R_{yy}(0) \end{pmatrix}. \quad (2.91)$$

Диагональные элементы матрицы вещественны и равны средним интенсивностям x - и y -компоненты поля волны, недиагональные элементы комплексно-сопряжённые. Это позволяет скомпоновать из элементов матрицы (2.91) четыре вещественных параметра Стокса:

$$\begin{aligned} S_0 &= R_{xx}(0) + R_{yy}(0) = E_{mx}^2 + E_{my}^2 = |\mathbf{E}|^2, \\ S_1 &= R_{xx}(0) - R_{yy}(0) = E_{mx}^2 - E_{my}^2 = |\mathbf{E}|^2 \cos 2\chi \cos 2\beta, \\ S_2 &= R_{xy}(0) + R_{yx}(0) = 2E_{mx}E_{my} \cos \theta = |\mathbf{E}|^2 \cos 2\chi \sin 2\beta, \\ S_3 &= -j(R_{xy}(0) - R_{yx}(0)) = 2E_{mx}E_{my} \sin \theta = |\mathbf{E}|^2 \sin 2\chi. \end{aligned} \quad (2.92)$$

Параметры Стокса имеют следующий физический смысл. Параметр S_0 равен интенсивности эллиптически поляризованной волны. Параметры S_1 и S_2 отвечают за линейную поляризацию: S_1 – за горизонтальную ($\chi=0$, $\beta=0$, $S_1=S_0$) или вертикальную ($\chi=0$, $\beta=\pi/2$, $S_1=-S_0$), а S_2 – за линейно поляризованные волны, вектор \mathbf{E} которых ориентирован под углами $\pm\pi/4$ к осям выбранной системы координат. Наконец, последний параметр Стокса, S_3 , свидетельствует о наличии волн с правовинтовой ($S_3>0$) или левовинтовой ($S_3<0$) круговой поляризацией. Нетрудно показать, что для детерминированной волны $\det \hat{\mathbf{J}}=0$, а $S_0^2 = S_1^2 + S_2^2 + S_3^2$, т.е. волна полностью поляризована.

Поляризационная обработка сигналов широко используется в современной радиолокации, а поляризационное уплотнение каналов позволяет использовать один и тот же частотный диапазон для одновременной работы двух информационных каналов.

2.9. Контрольные вопросы и упражнения

1. Какие колебания называются периодическими и какие – квазипериодическими?
2. Каковы особенности гармонических колебаний? Почему гармонические колебания занимают особое место в радиоэлектронике и статистической радиофизике?
3. Какие характеристики используются для описания импульсных сигналов?
4. Как различаются видео- и радиоимпульсы?
5. Что такое функция включения Хевисайда и дельта-функция Дирака? Как они связаны между собой?
6. В чем смысл динамического представления сигналов? Какие виды динамического представления вам известны?
7. Что называется спектром колебания? Как получить амплитудный и фазовый спектры сигнала?
8. Изобразите спектр последовательности прямоугольных импульсов, длительность которых в 3,5 раза меньше периода.
9. Какие формы ряда Фурье используются в радиоэлектронике?

10. Изобразите спектры сигналов, представленных формулами (2.7) и (2.8) при $A_1 = 2$, $A_2 = 3$. Объясните различие в спектрах.
11. Как различаются вещественный и комплексный частотные спектры сигналов? Каков смысл понятия «отрицательная частота»?
12. Как изменится спектр периодической последовательности импульсов при увеличении периода повторения в 3 раза, в N раз?
13. Как по заданному спектру рассчитать среднюю мощность периодического сигнала?
14. В чем заключается явление Гиббса?
15. Как разложить непериодический сигнал на гармонические составляющие? Почему для этого вместо разложения в ряд Фурье необходимо использовать преобразование Фурье?
16. Когда спектр сигнала является сплошным, когда линейчатым эквидистантным и когда линейчатым неэквидистантным?
17. В чем заключается принцип дуальности частоты и времени? Как его использовать практически?
18. Как определить спектральную плотность произведения двух сигналов? В чем заключается смысл понятия «свёртка двух функций»?
19. Как различаются комплексные спектры видео- и радиоимпульсов одинаковой формы?
20. В чем заключается особенность спектра дельта-функции? Какой спектр можно назвать «белым»?
21. Как связаны между собой длительность импульса и ширина его спектра?
22. Сигнал с постоянной спектральной плотностью в интервале частот от 1 до 6 кГц продифференцировали по времени. Во сколько раз будет отличаться спектральная плотность на краях диапазона после дифференцирования?
23. Каковы особенности спектральных плотностей неинтегрируемых сигналов?
24. Покажите, что спектральная плотность группы из четырёх одинаковых импульсов, обладающих спектральной плотностью $X_1(\omega)$ и сдвинутых во времени на одинаковый интервал T , удовлетворяет равенству $X_{\Sigma}(\omega) = 4X_1(\omega)e^{-3j\omega T/2} \cos \omega T \cdot \cos \omega T/2$.
25. Какие комплексные сигналы называются аналитическими и каковы особенности спектральной плотности аналитического сигнала? Что такое сопряжённый сигнал?
26. Каков физический смысл автокорреляционной функции детерминированного сигнала? Какова связь автокорреляционной функции с энергетическим спектром сигнала?
27. Что такое комплексная частота? Каковы границы применимости преобразований Лапласа?
28. Покажите справедливость следующих \mathbf{L} -преобразований:
1) $e^{-t} - e^{-2t} \iff 1/[(s+1)(s+2)]$, 2) $1 - e^{-|\alpha|t} \iff 1/[s(s+\alpha)]$.
29. В чем общность и различие преобразований Фурье и Лапласа?
30. Как получить общее решение волнового уравнения?
31. Каков смысл фазовой скорости?
32. Что такое поперечная электромагнитная волна?
33. Как волновые уравнения превратить в уравнения Гельмгольца?

34. Что такое волновой «пакет»? Каков смысл групповой скорости?
35. Почему детерминированные волны полностью поляризованы? Как связаны параметры Стокса с параметрами поляризационного эллипса?

Список литературы

- Нефедов, В. И.* Основы радиоэлектроники и связи : учебник для вузов. 3-е изд., испр. М. : Высш. шк., 2005.
- Хохлов, А. В.* Теоретические основы радиоэлектроники : учеб. пособие. Саратов : Изд-во Сарат. ун-та, 2005.
- Баскаков, С. И.* Радиотехнические цепи и сигналы : учебник. М. : Высш. шк., 2000.
- Гоноровский, И. С.* Радиотехнические цепи и сигналы : учебник. М. : Радио и связь, 1986.
- Справочник по теоретическим основам радиоэлектроники : в 2 т. / под ред. Б. Х. Кривицкого. М. : Энергия, 1977.
- Каяцкас, А. А.* Основы радиоэлектроники. М. : Высш. шк., 1988.
- Сиберт, У. М.* Цепи, сигналы, системы : в 2 ч. / пер. с англ. М. : Мир, 1988.
- Основы радиофизики / Г. В. Белокопытов, К. С. Ржевкин, А. А. Белов, А. С. Логгинов, Ю. И. Кузнецов, И. В. Иванов / под ред. А. С. Логгинова. М. : Изд-во УРСС, 1996.
- Харкевич, А. А.* Спектры и анализ. 4-е изд. М. : Изд-во ЛКИ, 2007.
- Теоретические основы электротехники: учебник для вузов: в 3 т. 4-е изд. / К. С. Демирчян, Л. Р. Нейман, Н. В. Коровкин, В. Л. Чечерин. СПб. : Питер, 2003.
- Зиновьев, А. Л., Филиппов, Л. И.* Введение в теорию сигналов и цепей : учеб. пособие. М. : Высш. шк., 1968.
- Калинин, В. И., Герштейн, Г. М.* Введение в радиофизику : учеб. пособие / под ред. В. И. Калинина. М. : Гостехиздат, 1957.
- Астахов, В. В., Вадивасова, Т. Е., Хохлов, А. В.* Радиофизика. Задачи и упражнения : учеб. пособие. Саратов : Изд-во Сарат. ун-та, 2008.
- Горелик, Г. С.* Колебания и волны. Введение в акустику, радиофизику и оптику. 3-е изд. / под ред. С. М. Рытова. М. : Физматлит, 2008.
- Хохлов, А. В.* О поляризационном анализе плоских электромагнитных волн // Лекции по электронике СВЧ и радиофизике : 8-я зимняя школа-семинар инженеров : в 5 кн. Кн. 1. Саратов : Изд-во Сарат. ун-та, 1989.

Глава 3

Случайные процессы, поля и волны

3.1. Случайные процессы и их описание

Случайный (стохастический) процесс – это процесс изменения состояния системы во времени, протекающий по вероятностным (статистическим) законам.

Пусть множество V значений случайной переменной X есть пространство состояний системы. Множество состояний V может быть дискретным (конечным или счётным) или непрерывным. Величина X может быть векторной, скалярной, вещественной или комплексной.

Случайный процесс описывается случайной функцией $X(t)$, принимающей значения во множестве V , где аргумент t обычно имеет смысл времени. Время t может принимать непрерывное или счётное множество значений. В последнем случае говорят о *случайной последовательности* $X(n)$. Для любого фиксированного t_0 значение $X_0 = X(t_0)$ – случайная величина.

За исключением специально оговорённых случаев, будем для простоты рассматривать одну вещественную случайную переменную X и описывать случайный процесс скалярной вещественной случайной функцией $X(t)$. Случайные величины и процессы будем по возможности обозначать заглавными буквами: X , Y , $X(t)$ и т.д., а принимаемые случайными переменными значения – соответствующими строчными буквами: x , y , $x(t)$. Однако мы не задаемся целью строго придерживаться этого правила. Так, оно может нарушаться в отношении греческих букв.

Реализация случайного процесса – это детерминированная последовательность состояний системы, получаемая в результате проведения испытаний (измерений). Если случайный процесс описывается случайной функцией $X(t)$, то в результате проведённых измерений получают детерминированную функцию $x(t)$, описывающую одну из множества возможных реализаций процесса. Если бы было возможно многократно провести те же измерения в совершенно идентичных условиях, то каждый раз получалась бы другая реализация того же случайного процесса. Множество всех возможных реализаций составляет *статистический ансамбль реализаций* случайного процесса. Пример различных реализаций случайного процесса с заданным начальным состоянием, смоделированных численно, приведён на рис. 3.1.

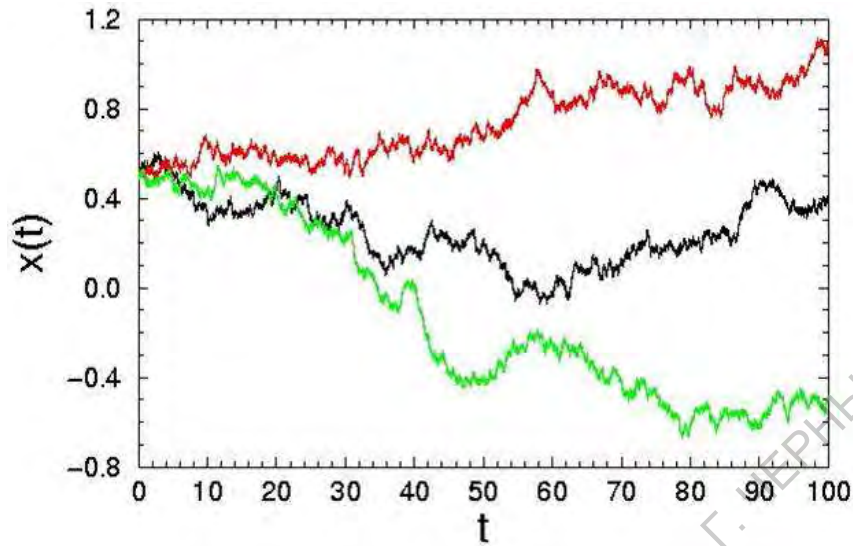


Рис. 3.1. Возможные реализации случайного процесса

Рассмотрим теперь основные характеристики непрерывнозначных случайных процессов.

Плотность вероятности, функция распределения и характеристическая функция случайного процесса. Случайный процесс $X(t)$ характеризуется n -мерной плотностью вероятности $p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n)$, которая является вещественной функцией n текущих значений случайной переменной X в n произвольных моментов времени и может быть задана следующим равенством:

$$\begin{aligned} p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) dx_1 dx_2 \dots dx_n &= \\ &= P\{X(t_1) \in [x_1; x_1 + dx_1] \wedge \dots \wedge X(t_n) \in [x_n; x_n + dx_n]\} \end{aligned} \quad (3.1)$$

или

$$\begin{aligned} p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) &= \\ &= \lim_{\Delta x_i \rightarrow 0, i=1,2,\dots,n} \frac{P\{X(t_1) \in [x_1; x_1 + \Delta x_1] \wedge \dots \wedge X(t_n) \in [x_n; x_n + \Delta x_n]\}}{\Delta x_1 \dots \Delta x_n}, \end{aligned} \quad (3.2)$$

где $P\{\dots\}$ – вероятность события, обозначенного в фигурных скобках.

Плотность вероятности должна удовлетворять следующим требованиям:

1) $p_n(\dots) \geq 0$ (неотрицательность);

2) $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) dx_1 dx_2 \dots dx_n = 1$ (нормировка);

3) для любого $k < n$

$p_k(x_1, t_1, x_2, t_2, \dots, x_k, t_k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) dx_{k+1} dx_{k+2} \dots dx_n$ (согласование);

4) для любых i и j

$p_n(x_1, t_1, \dots, x_i, t_i, \dots, x_j, t_j, \dots, x_n, t_n) = p_n(x_1, t_1, \dots, x_j, t_j, \dots, x_i, t_i, \dots, x_n, t_n)$ (инвариантность относительно перестановки пар аргументов);

5) если значения случайного процесса в некоторые моменты времени t_i^0 статистически независимы, то

$$p_n(x_1, t_1^0, x_2, t_2^0, \dots, x_n, t_n^0) = p_1(x_1, t_1^0) p_1(x_2, t_2^0) \dots p_1(x_n, t_n^0).$$

Если два случайных процесса $X(t)$ и $X'(t)$ характеризуются одной и той же n -мерной плотностью вероятности для любого n , то они называются *эквивалентными*. Все статистические характеристики таких процессов полностью совпадают.

Кроме плотности вероятности случайный процесс можно описать с помощью n -мерной функции распределения $F_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n)$ и n -мерной характеристической функции $\Theta_X(u_1, t_1, u_2, t_2, \dots, u_n, t_n)$, которые связаны с плотностью вероятности следующими соотношениями:

$$F_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p_n(x'_1, t_1, x'_2, t_2, \dots, x'_n, t_n) dx'_1 dx'_2 \dots dx'_n; \quad (3.3)$$

$$p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = \frac{\partial^n F_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n)}{\partial x_1 \partial x_2 \dots \partial x_n}; \quad (3.4)$$

$$\Theta_X(u_1, t_1, u_2, t_2, \dots, u_n, t_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_n(x_1, t_1, \dots, x_n, t_n) \exp(j \sum_{i=1}^n u_i x_i) dx_1 \dots dx_n; \quad (3.5)$$

$$p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \Theta_X(u_1, t_1, \dots, u_n, t_n) \exp(-j \sum_{i=1}^n u_i x_i) du_1 \dots du_n. \quad (3.6)$$

Случайный процесс $X(t)$ *полностью задан*, если для любого n известна одна из функций: $p_n(\dots)$; $F_n(\dots)$ или $\Theta_X(\dots)$. В этом случае можно найти любые статистические характеристики процесса.

Статистическое среднее и моментные функции. Пусть $f(x_1, x_2, \dots, x_n)$ – некоторая детерминированная функция своих аргументов. Пусть аргументы функции f есть значения случайного процесса $X(t)$ в n моментов времени: $X(t_i)$, $i = 1, 2, \dots, n$. *Статистическим средним* (математическим ожиданием) функции $f(X_1, X_2, \dots, X_n)$ на ансамбле реализаций случайного процесса $X(t)$ называется следующая детерминированная функция аргументов t_i :

$$\bar{f}(t_1, t_2, \dots, t_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) p_n(x_1, t_1, \dots, x_n, t_n) dx_1 \dots dx_n. \quad (3.7)$$

Далее будем обозначать операцию статистического усреднения с помощью угловых скобок $\langle \dots \rangle$:

$$\bar{f}(t_1, t_2, \dots, t_n) = \langle f(X_1, X_2, \dots, X_n) \rangle.$$

Моментными функциями (моментами) случайного процесса называют следующие статистические средние от степенных функций:

$$\langle X_1^{r_1} X_2^{r_2} \dots X_k^{r_k} \rangle = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{r_1} \dots x_k^{r_k} p_n(x_1, t_1, \dots, x_k, t_k) dx_1 \dots dx_k \quad (3.8)$$

либо

$$\begin{aligned} & \langle \tilde{X}_1^{r_1} \tilde{X}_2^{r_2} \dots \tilde{X}_k^{r_k} \rangle = \\ & = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 - \bar{X}(t_1))^{r_1} \dots (x_k - \bar{X}(t_k))^{r_k} p_n(x_1, t_1, \dots, x_k, t_k) dx_1 \dots dx_k, \end{aligned} \quad (3.9)$$

где $X_i = X(t_i)$ – значения случайного процесса в моменты времени t_i ; а $\tilde{X}_i = X(t_i) - \bar{X}(t_i)$ – значения *флуктуации* (отклонения от среднего); $\bar{X}(t) = \langle X(t) \rangle$ – среднее значение случайного процесса. Моменты вида (3.8) называют *начальными*, а вида (3.9) – *центральными*. Число рассматриваемых отсчётов времени k называют *размерностью момента*, а сумму показателей степеней $\sum_{i=1}^k r_i = r$ – *порядком момента*.

Моменты случайного процесса представляют собой детерминированные функции аргументов t_i или константы.

Начальные моменты случайного процесса являются коэффициентами в разложении n -мерной характеристической функции в ряд Маклорена:

$$\begin{aligned} \Theta_X(u_1, t_1, u_2, t_2, \dots, u_n, t_n) &= \\ &= \sum_{r_1=0}^{\infty} \dots \sum_{r_n=0}^{\infty} \frac{\langle X^{r_1}(t_1) X^{r_2}(t_2) \dots X^{r_n}(t_n) \rangle}{r_1! r_2! \dots r_n!} (ju_1)^{r_1} (ju_2)^{r_2} \dots (ju_n)^{r_n} \end{aligned} \quad (3.10)$$

и, соответственно, могут быть найдены по формуле

$$\langle X^{r_1}(t_1) \dots X^{r_n}(t_n) \rangle = (-j)^r \frac{\partial^r \Theta_X(u_1, t_1, \dots, u_n, t_n)}{\partial u_1^{r_1} \dots \partial u_n^{r_n}} \Big|_{u_1=u_2=\dots=u_n=0}, \quad (3.11)$$

где $r = \sum_{i=1}^n r_i$.

Аналогично (3.10) можно представить разложение натурального логарифма от характеристической функции:

$$\begin{aligned} \ln \Theta_X(u_1, t_1, u_2, t_2, \dots, u_n, t_n) &= \\ &= \sum_{r_1=0}^{\infty} \dots \sum_{r_n=0}^{\infty} \frac{\langle \langle X^{r_1}(t_1) X^{r_2}(t_2) \dots X^{r_n}(t_n) \rangle \rangle}{r_1! r_2! \dots r_n!} (ju_1)^{r_1} (ju_2)^{r_2} \dots (ju_n)^{r_n}. \end{aligned} \quad (3.12)$$

Коэффициенты в разложении, обозначенные в выражении (3.12) как $\langle \langle X^{r_1}(t_1) X^{r_2}(t_2) \dots X^{r_n}(t_n) \rangle \rangle$, называются *кумулянтными функциями* (*кумулянтами*) случайного процесса $X(t)$.

К наиболее важным моментам первого и второго порядка относятся следующие:

1) *среднее значение* случайного процесса (одномерный начальный момент порядка $r = 1$)

$$\bar{X}(t) = \langle X(t) \rangle = \int_{-\infty}^{\infty} x p_1(x, t) dx; \quad (3.13)$$

2) *средний квадрат* (одномерный начальный момент порядка $r = 2$)

$$\overline{X^2}(t) = \langle X^2(t) \rangle = \int_{-\infty}^{\infty} x^2 p_1(x, t) dx; \quad (3.14)$$

3) дисперсия (одномерный центральный момент порядка $r = 2$)

$$\sigma_X^2(t) = \langle (X(t) - \overline{X}(t))^2 \rangle = \int_{-\infty}^{\infty} (x - \overline{X}(t))^2 p_1(x, t) dx = \overline{X^2} - \overline{X}^2; \quad (3.15)$$

4) ковариационная (или автоковариационная) функция (двумерный начальный момент порядка $r = 2$)

$$K_X(t_1, t_2) = \langle X(t_1)X(t_2) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_2(x_1, t_1, x_2, t_2) dx_1 dx_2; \quad (3.16)$$

5) корреляционная¹ (или автокорреляционная) функция (двумерный центральный момент порядка $r = 2$)

$$\begin{aligned} \Psi_X(t_1, t_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \overline{X}(t_1))(x_2 - \overline{X}(t_2)) p_2(x_1, t_1, x_2, t_2) dx_1 dx_2 = \\ &= K_X(t_1, t_2) - \overline{X}(t_1)\overline{X}(t_2). \end{aligned} \quad (3.17)$$

Нормированную корреляционную функцию

$$R_X(t_1, t_2) = \frac{\Psi_X(t_1, t_2)}{\sqrt{\sigma_X^2(t_1)\sigma_X^2(t_2)}} \quad (3.18)$$

называют коэффициентом корреляции.

Корреляционная функция случайного процесса обладает следующими свойствами:

- 1) $\Psi_X^2(t_1, t_2) \leq \sigma_X^2(t_1)\sigma_X^2(t_2)$ (соответственно $|R_X(t_1, t_2)| \leq 1$);
- 2) $\Psi_X(t, t) = \sigma_X^2(t)$;
- 3) $\Psi_X(t_1, t_2) = \Psi_X(t_2, t_1)$;
- 4) если значения случайного процесса в некоторые моменты времени t'_1 и t'_2 статистически независимы, то они некоррелированы, т.е. $\Psi_X(t'_1, t'_2) = 0$. Обратное утверждение в общем случае неверно.

Характеристики совокупности случайных процессов. Для простоты рассмотрим два процесса: $X(t)$ и $Y(t)$. Их наиболее полной статистической характеристикой является совместная плотность вероятности

$$\begin{aligned} &p_n^{XY}(x_1, y_1, t_1, \dots, x_n, y_n, t_n) dx_1 dy_1 \dots dx_n dy_n = \\ &= P\{X(t_1) \in [x_1; x_1 + dx_1] \wedge Y(t_1) \in [y_1; y_1 + dy_1] \wedge \dots \\ &\dots \wedge X(t_n) \in [x_n; x_n + dx_n] \wedge Y(t_n) \in [y_n; y_n + dy_n]\}. \end{aligned} \quad (3.19)$$

Соответственно, можно ввести совместную функцию распределения

$$F_n^{XY}(x_1, y_1, t_1, x_2, y_2, t_2, \dots, x_n, y_n, t_n) =$$

¹В литературе иногда встречается другая терминология: двумерный начальный момент второго порядка называют корреляционной функцией, а двумерный центральный момент – ковариационной функцией.

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{y_n} p_n^{XY}(x'_1, y'_1, t_1, x'_2, y'_2, t_2, \dots, x'_n, y'_n, t_n) dx'_1 dx'_2 \dots dx'_n dy'_1 dy'_2 \dots dy'_n \quad (3.20)$$

и совместную характеристическую функцию

$$\Theta_{XY}(u_1, v_1, t_1, u_2, v_2, t_2, \dots, u_n, v_n, t_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_n^{XY}(x_1, y_1, t_1, \dots, x_n, y_n, t_n) \exp \left[j \sum_{i=1}^n (u_i x_i + v_i y_i) \right] dx_1 \dots dy_n. \quad (3.21)$$

Можно характеризовать совокупность процессов $X(t)$ и $Y(t)$ с помощью взаимных моментных функций, например *взаимной ковариационной функции*

$$K_{XY}(t_1, t_2) = \langle X(t_1)Y(t_2) \rangle \quad (3.22)$$

или *взаимной корреляционной функции*

$$\Psi_{XY}(t_1, t_2) = \langle X(t_1)Y(t_2) \rangle - \bar{X}(t_1)\bar{X}(t_2). \quad (3.23)$$

Следует отметить, что в общем случае $\Psi_{XY}(t_1, t_2) \neq \Psi_{XY}(t_2, t_1)$. Также $\Psi_{XY}(t_1, t_2) \neq \Psi_{YX}(t_1, t_2)$. Однако, как видно из определения (3.23), $\Psi_{XY}(t_1, t_2) \equiv \Psi_{YX}(t_2, t_1)$. Аналогичные утверждения справедливы и для ковариационной функции.

Если имеется совокупность N случайных процессов, например векторный процесс $\mathbf{X}(t)$ с компонентами $X_1(t), X_2(t), \dots, X_N(t)$, то он характеризуется совместной плотностью вероятности всех компонент. Рассматриваются также взаимные корреляционные (или ковариационные) функции различных компонент, т.е. вводится корреляционная матрица $\hat{\Psi}_{\mathbf{X}}(t_1, t_2)$ с элементами

$$\Psi_{ik}(t_1, t_2) = \langle X_i(t_1)X_k(t_2) \rangle - \bar{X}_i(t_1)\bar{X}_k(t_2).$$

Характеристики комплексного случайного процесса. Комплексный случайный процесс $Z(t)$ можно представить в виде: $Z(t) = X(t) + jY(t)$, где $X(t)$ и $Y(t)$ – вещественные случайные функции. Таким образом, вместо процесса $Z(t)$ можно рассматривать совокупность двух случайных процессов – $X(t)$ и $Y(t)$, которая полностью характеризуется плотностью вероятности $p_n^{XY}(x_1, y_1, t_1, x_2, y_2, t_2, \dots, x_n, y_n, t_n)$.

Помимо моментов вещественной и мнимой компонент $X(t)$ и $Y(t)$ и их взаимных моментов рассматриваются также моментные функции комплексного случайного процесса $Z(t)$:

среднее значение
$$\bar{Z}(t) = \bar{X}(t) + j\bar{Y};$$

дисперсия
$$\sigma_Z^2(t) = \langle |Z(t) - \bar{Z}(t)|^2 \rangle;$$

ковариационная функция

$$K_Z(t_1, t_2) = \langle Z(t_1)Z^*(t_2) \rangle.$$

Символом * обозначено комплексно-сопряжённое значение. Дисперсия комплексного процесса всегда вещественна, все остальные моментные функции в общем случае комплексные.

Корреляционную функцию комплексного процесса определяют двумя способами. Соответственно можно рассматривать две различные корреляционные функции комплексного процесса:
первая корреляционная функция –

$$\Psi_Z^I(t_1, t_2) = \langle Z(t_1)Z^*(t_2) \rangle - \bar{Z}(t_1)\bar{Z}^*(t_2);$$

вторая корреляционная функция –

$$\Psi_Z^{II}(t_1, t_2) = \langle Z(t_1)Z(t_2) \rangle - \bar{Z}(t_1)\bar{Z}(t_2).$$

Первая из них при $t_1 = t_2$ принимает вещественные значения, равные дисперсии.

3.2. Вероятностная сходимость. Непрерывность. Производная и интеграл от случайного процесса

Одним из базовых понятий теории метрических пространств, а значит, и теории сигналов является понятие сходимости. По определению последовательность элементов x_1, x_2, \dots метрического пространства M сходится к пределу $x_0 \in M$, если расстояние $d(x_n, x_0) \rightarrow 0$ при $n \rightarrow \infty$.

Неслучайная функция времени $x(t) \in M'$ сходится к $x_0 \in M'$ при $t \rightarrow t_0$, если $d(x(t), x_0) \rightarrow 0$ при $t \rightarrow t_0$.

Для случайных последовательностей и функций такие определения сходимости не подходят. Сходимость последовательностей случайных величин и случайных функций должна пониматься в вероятностном (статистическом) смысле.

Вероятностная сходимость последовательностей случайных величин. Существует несколько основных определений вероятностной сходимости.

1) Последовательность случайных величин X_n сходится к случайной величине X_0 в среднем квадратичном, если существует предел

$$\lim_{n \rightarrow \infty} \langle (X_n - X_0)^2 \rangle = 0. \quad (3.24)$$

Для среднеквадратического предела часто применяют обозначение

$$\text{l.i.m.}_{n \rightarrow \infty} X_n = X_0,$$

где «l.i.m.» является сокращением слов «limit in the mean square» (предел в среднем квадратичном). Можно также указывать характер сходимости в скобках после написания предела, например

$$\lim_{n \rightarrow \infty} X_n = X_0 \text{ (ср.кв.)},$$

или каким-то другим способом.

Замечание. Если существует предел

$$\text{l.i.m.}_{n \rightarrow \infty} X_n = a,$$

где a – заданная (не случайная) величина, то существуют обычные пределы

$$\lim_{n \rightarrow \infty} \bar{X}_n = a \quad \text{и} \quad \lim_{n \rightarrow \infty} \sigma_{X_n}^2 = 0.$$

2) Последовательность случайных величин X_n сходится к случайной величине X_0 *по вероятности*, если для любого $\varepsilon > 0$ существует предел

$$\lim_{n \rightarrow \infty} P \{ |X_n - X_0| > \varepsilon \} = 0. \quad (3.25)$$

Будем обозначать предел по вероятности как

$$\lim_{n \rightarrow \infty} X_n = X_0 \quad (\text{по вер.}).$$

3) Последовательность случайных величин X_n сходится к случайной величине X_0 *почти наверное* или *с вероятностью 1*, если

$$P \left\{ \lim_{n \rightarrow \infty} X_n = X_0 \right\} = 1, \quad (3.26)$$

где предел определяется на ансамбле реализаций случайной последовательности и рассматривается вероятность того, что он совпадает с одним из возможных значений случайной величины X_0 . Обозначим

$$\lim_{n \rightarrow \infty} X_n = X_0 \quad (\text{почти навер.}).$$

Доказаны следующие соотношения между различными типами сходимости:

1) из среднеквадратической сходимости всегда следует сходимость по вероятности;

2) из сходимости по вероятности следует среднеквадратическая сходимость только для последовательностей, удовлетворяющих условию $P\{|X_n| \leq \text{const} \neq +\infty\} = 1$;

3) из сходимости почти наверное следует сходимость по вероятности.

При рассмотрении характеристик случайных процессов чаще всего используется среднеквадратическая сходимость. Мы будем использовать её в приводимых далее определениях, хотя можно применять и другие определения вероятностной сходимости.

Сходимость и непрерывность случайной функции.

Случайная функция $X(t)$ сходится к случайной величине X_0 в среднем квадратичном, т.е.

$$\text{l.i.m.}_{t \rightarrow t_0} X(t) = X_0, \quad (3.27)$$

если для любой последовательности $\{t_n\}$: $t_n \rightarrow t_0$ при $n \rightarrow \infty$ существует предел

$$\text{l.i.m.}_{n \rightarrow \infty} X_n = X_0.$$

Случайная функция $X(t)$ непрерывна в точке t_0 в среднеквадратическом смысле, если для каждой реализации существует предел

$$\text{l.i.m.}_{\tau \rightarrow 0} X(t_0 + \tau) = X(t_0), \quad \text{т.е.} \quad \lim_{\tau \rightarrow 0} \langle (X(t_0 + \tau) - X(t_0))^2 \rangle = 0. \quad (3.28)$$

Случайная функция $X(t)$ с ограниченным средним квадратом непрерывна в точке t_0 в среднеквадратическом смысле тогда и только тогда, когда выполняются следующие условия:

- 1) $\overline{X}(t)$ – функция, непрерывная в t_0 ;
- 2) $\Psi_X(t_1, t_2)$ – функция, непрерывная в t_0 по обоим аргументам.

Если $X(t)$ непрерывна в среднем квадратичном в любой точке t , то говорят, что $X(t)$ – *непрерывная в среднем квадратичном* функция. Непрерывность функции по вероятности или почти наверное определяется аналогично, но соответствующие свойства непрерывности могут в общем случае различаться.

Случайные процессы, описываемые случайными функциями, непрерывными в среднеквадратическом смысле, по вероятности или почти наверное, называются соответственно *случайными процессами, непрерывными в среднеквадратическом смысле, по вероятности или почти наверное*. Кроме них выделяют просто *непрерывные случайные процессы*. Случайный процесс $X(t)$ называется *непрерывным*, если любая, произвольно выбранная, реализация этого процесса с вероятностью 1 является непрерывной функцией. Такая непрерывность не следует ни из среднеквадратичной непрерывности, ни из непрерывности почти наверное или непрерывности по вероятности.

Производная и интеграл от случайной функции. Случайная функция $X(t)$ дифференцируема в точке t_0 в среднеквадратическом смысле, если для каждой реализации существует предел

$$\text{l.i.m.}_{\tau \rightarrow 0} \frac{X(t_0 + \tau) - X(t_0)}{\tau} = X'(t_0). \quad (3.29)$$

Случайная величина

$$X'(t_0) = \left. \frac{dX(t)}{dt} \right|_{t_0}$$

называется *производной случайной функции* в точке t_0 .

Случайная функция $X(t)$ с ограниченным средним квадратом дифференцируема в точке t_0 в среднеквадратическом смысле тогда и только тогда, когда выполняются следующие условия:

- 1) $\overline{X}(t)$ – функция, дифференцируемая в t_0 ;
- 2) $\Psi_X(t_1, t_2)$ – функция, дифференцируемая в t_0 по обоим аргументам.

Если $X(t)$ дифференцируема в среднем квадратичном в любой точке t , то говорят, что $X(t)$ – *дифференцируемая в среднем квадратичном* функция.

Среднеквадратическим интегралом от случайной функции $X(t)$ на интервале $t \in [a, b]$ называется следующий предел:

$$\int_a^b X(t) dt = \text{l.i.m.}_{n \rightarrow \infty} \sum_{i=0}^{n-1} X(t_i)(t_{i+1} - t_i). \quad (3.30)$$

Случайная функция $X(t)$ с ограниченным средним квадратом интегрируема на интервале $t \in [a, b]$ в среднеквадратическом смысле тогда и только тогда, когда существуют римановы интегралы:

- 1) $\int_a^b \overline{X}(t) dt$;
- 2) $\int_a^b \int_a^b \Psi_X(t_1, t_2) dt_1 dt_2$.

3.3. Стационарные и эргодические случайные процессы

Стационарные случайные процессы. Стационарность случайного процесса означает, что стохастическая система находится в установившемся состоянии и её статистические характеристики не меняются во времени. Различают *стационарность в узком (строгом) смысле* и *стационарность в широком смысле*.

Случайный процесс $X(t)$ называют *стационарным в узком (строгом) смысле*, если для любого целого положительного n , любой константы T и любых моментов времени t_i , $i = 1, 2, \dots, n$ имеет место равенство

$$p_n(x_1, t_1 \pm T, x_2, t_2 \pm T, \dots, x_n, t_n \pm T) = p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n), \quad (3.31)$$

что означает инвариантность плотности вероятности (и, следовательно, всех статистических характеристик) относительно сдвига во времени. При этом одномерная плотность вероятности не зависит от времени ($p_1(x, t) \equiv p_1(x)$) и одномерные моменты являются константами. Двумерная плотность вероятности зависит только от разности моментов времени $\tau = t_2 - t_1$ ($p_2(x_1, t_1, x_2, t_2) \equiv p_2(x_1, x_2, \tau)$). Соответственно, двумерные моменты являются функциями только τ .

Случайный процесс $X(t)$ называется *стационарным в широком смысле*, если выполняются условия

$$\langle X(t) \rangle \equiv \text{const}; \quad \Psi_X(t_1, t_2) = \Psi_X(t_2 - t_1); \quad \langle X^2(t) \rangle < \infty. \quad (3.32)$$

Для процессов с конечной «мощностью» (т.е. процессов, для которых $\langle X^2(t) \rangle < \infty$) из строгой стационарности следует стационарность в широком смысле. Из стационарности в широком смысле в общем случае не следует строгая стационарность.

Корреляционная функция стационарного процесса обладает следующими свойствами:

- 1) $|\Psi_X(\tau)| \leq \sigma_X^2$, где $\tau = t_2 - t_1$;
- 2) $\Psi_X(0) = \sigma_X^2$;
- 3) $\Psi_X(-\tau) = \Psi_X(\tau)$ (чётная функция);
- 4) если корреляционная функция непрерывна при $\tau = 0$, то она непрерывна для любого значения τ .

Стационарный в строгом смысле случайный процесс, для которого выполняется условие $\lim_{\tau \rightarrow \infty} \Psi_X^2(\tau) = 0$, называется *процессом с перемешиванием*. Для него справедливо равенство, выражающее *свойство расцепления корреляций*:

$$\lim_{\tau \rightarrow \infty} p_2(x_1, x_2, \tau) = p_1(x_1)p_1(x_2). \quad (3.33)$$

Скорость перемешивания (расцепления корреляций) характеризуют *временем корреляции* $\tau_{\text{кор}}$. Чаще всего используют следующие два определения времени корреляции:

1. Время корреляции определяется как интервал, на котором огибающая корреляционной функции $\gamma_X(\tau)$ убывает в e раз (рис. 3.2), где e — основание натурального логарифма):

$$\gamma_X(\tau_{\text{кор}}) = \frac{\sigma_X^2}{e}. \quad (3.34)$$

2. Время корреляции определяется по формуле

$$\tau_{\text{кор}} = \frac{1}{\sigma_X^2} \int_0^\infty \gamma_X(\tau) d\tau. \quad (3.35)$$

Для совокупности двух случайных процессов вводится понятие стационарной связанности.

Случайные процессы $X(t)$ и $Y(t)$ называют *стационарно-связанными в узком (строгом) смысле*, если для любого целого положительного n , любой константы T и любых моментов времени $t_i, i = 1, 2, \dots, n$ совместная плотность вероятности удовлетворяет равенству:

$$\begin{aligned} p_n(x_1, y_1, t_1 \pm T, x_2, y_2, t_2 \pm T, \dots, x_n, y_n, t_n \pm T) = \\ = p_n(x_1, y_1, t_1, x_2, y_2, t_2, \dots, x_n, y_n, t_n). \end{aligned} \quad (3.36)$$

Случайные процессы $X(t)$ и $Y(t)$ называют *стационарно-связанными в широком смысле*, если

$$\Psi_{XY}(t_1, t_2) = \Psi_{XY}(t_2 - t_1). \quad (3.37)$$

Очевидно, из строгой стационарной связанности процессов $X(t)$ и $Y(t)$ следует их стационарная связанность в широком смысле, а также строгая стационарность каждого из этих процессов. Стационарность векторного или комплексного процесса предполагает стационарность (в том или ином смысле) его компонент и их стационарную связанность между собой (в соответствующем смысле).

До сих пор для определения статистических характеристик и усреднения случайных процессов использовались ансамбли реализаций, хотя в большинстве практических задач бывает известна только одна реализация случайного процесса. Естественно возникает вопрос, нельзя ли определить хотя бы некоторые характеристики, например, среднее значение или дисперсию случайного процесса посредством временной обработки этой единственной реализации. Оказывается, в общем случае этого сделать нельзя. Однако существуют случайные процессы, для которых можно определить ряд статистических характеристик, используя одну достаточно длинную реализацию. Такие процессы называют *эргодическими*.

Эргодические случайные процессы. Эргодичность случайного процесса означает, что можно получить определённые статистические характеристики процесса, заменяя статистическое усреднение (по ансамблю реализаций) усреднением по времени вдоль одной реализации.

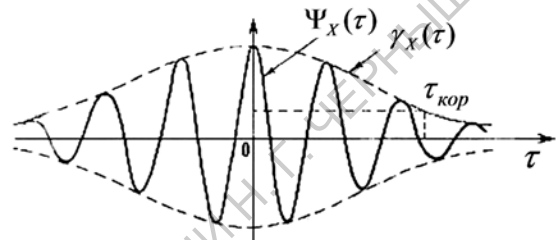


Рис. 3.2. Иллюстрация к определению времени корреляции первым способом

Можно выделить эргодичность относительно отдельных моментных функций, эргодичность 1-, 2-, ..., k-го порядка и, наконец, строгую эргодичность. Будем обозначать операцию усреднения по времени символом $\langle \dots \rangle_t$.

Пусть $X(t)$ – стационарный в строгом смысле случайный процесс². Среднее по времени значение детерминированной функции случайных аргументов $f[X(t_1), \dots, X(t_k)]$ вдоль некоторой реализации $x(t)$ случайного процесса определяется как

$$\begin{aligned} & \langle f[x(t_1), \dots, x(t_k)] \rangle_t = \\ & = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f[x(t_1), x(t_1 + \tau_1), \dots, x(t_1 + \tau_{k-1})] dt_1, \end{aligned} \quad (3.38)$$

где $\tau_i = t_{i+1} - t_i$.

Эргодичность относительно среднего значения означает выполнение следующего равенства:

$$\text{l.i.m.}_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dt = \int_{-\infty}^{\infty} p_1(x) x dx, \quad (3.39)$$

где слева стоит среднеквадратический предел случайной функции аргумента T . Если процесс $X(t)$ является эргодическим относительно среднего значения, то среднее по времени значение $\langle x(t) \rangle_t$ для почти любой реализации $x(t)$ равно среднему по ансамблю значению $\langle X(t) \rangle$:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt = \int_{-\infty}^{\infty} p_1(x) x dx.$$

Аналогично можно ввести эргодичность относительно среднего квадрата, эргодичность относительно дисперсии и т.д.

Эргодичность первого порядка означает выполнение следующего равенства:

$$\text{l.i.m.}_{T \rightarrow \infty} \frac{1}{T} \int_0^T f[X(t)] dt = \int_{-\infty}^{\infty} p_1(x) f(x) dx, \quad (3.40)$$

где $f[\dots]$ – любая детерминированная функция своего аргумента. Достаточным условием эргодичности первого порядка является свойство перемешивания.

Если процесс $X(t)$ является эргодическим первого порядка, то он также является эргодическим относительно всех одномерных моментов. В этом случае все одномерные моменты можно найти, используя усреднение по времени вдоль почти любой реализации случайного процесса.

²Для эргодичности первого порядка требование стационарности является излишним, но строгая эргодичность требует строгой стационарности.

Для процесса, обладающего свойством эргодичности первого порядка, зная какую-либо типичную реализацию, можно экспериментально определить одномерную плотность вероятности. Пусть в результате измерений нам известен отрезок реализации $x_1(t)$ на интервале времени $t \in [0; T]$. Выделим интервал значений $[x; x + \Delta x]$ (рис. 3.3) и обозначим за $T_{\Delta x}$ совокупное время тех интервалов Δt_i , в течение которых на протяжении T значение переменной принадлежало интервалу $[x; x + \Delta x]$:

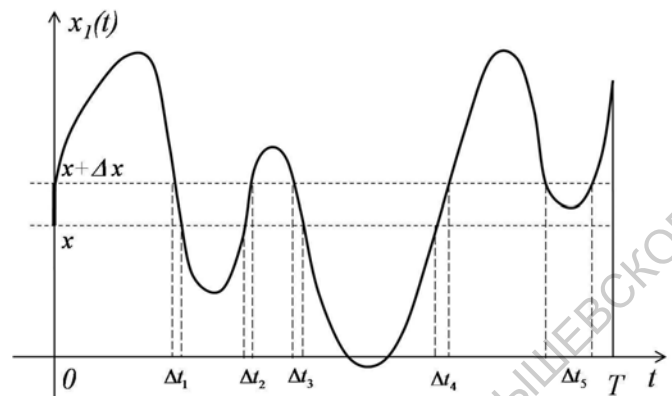


Рис. 3.3. Иллюстрация к определению одномерной плотности вероятности по реализации эргодического процесса

$$T_{\Delta x} = \sum_i \Delta t_i.$$

Используем определение (3.40), выбрав функцию $f(x_1)$ в виде

$$f(x_1) = 1, \quad x_1 \in [x; x + \Delta x] \quad \text{и} \quad f(x_1) = 0 \quad \text{в других случаях.}$$

Полагая время наблюдения T достаточно большим, а длину интервала Δx малой, на основании (3.40) приходим к приближенному равенству:

$$\frac{1}{T} \int_0^T f[x_1(t)] dt = \frac{T_{\Delta x}}{T} \approx p_1(x) \Delta x.$$

Откуда получаем

$$p_1(x) = \lim_{T \rightarrow \infty, \Delta x \rightarrow 0} \frac{T_{\Delta x}}{\Delta x T}.$$

Эргодичность второго порядка означает выполнение следующего равенства:

$$\text{l.i.m.}_{T \rightarrow \infty} \frac{1}{T} \int_0^T f[X(t), X(t + \tau)] dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) p_2(x_1, x_2, \tau) dx_1 dx_2, \quad (3.41)$$

где $f(\dots)$ – любая детерминированная функция своих аргументов.

Если процесс $X(t)$ является эргодическим второго порядка, то он обладает свойством эргодичности относительно всех двумерных моментов. В этом случае все двумерные моменты можно найти, используя усреднение по времени вдоль почти любой реализации случайного процесса. Аналогично можно определить эргодичность 3-го, 4-го, ..., k -го порядка.

Строгая эргодичность означает, что все статистические характеристики случайного процесса можно найти, используя усреднение по времени.

3.4. Спектральное представление случайных процессов

Спектральное представление сигналов в радиофизике обычно связано с разложением сигналов на сумму гармонических функций. Гармонический анализ служит важным методом исследования структуры детерминированных сигналов (см. раздел 2.3). Это связано, прежде всего, с задачами преобразования сигналов линейными системами с постоянными параметрами. В силу принципа суперпозиции каждая гармоническая компонента сигнала воздействия независимо преобразуется в гармоническую компоненту отклика. Воздействие случайного сигнала также было бы желательно разложить на гармонические компоненты.

Гармоническое разложение случайного процесса. Пусть $X(t)$ – скалярный вещественный случайный процесс с ограниченным средним квадратом ($\overline{X^2} \neq \infty$). Реализацию $x(t)$ можно представить в виде интеграла Фурье – Стилтеса

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} dC_x(\omega),$$

где $dC_x(\omega)$ – конечное или бесконечно малое (пропорциональное $d\omega$) приращение комплексной функции $C_x(\omega)$ на интервале $[\omega, \omega + d\omega)$. Разным реализациям $x(t)$ соответствуют разные функции $C_x(\omega)$. Таким образом, ансамблю реализаций случайного процесса $X(t)$ соответствует ансамбль реализаций $C_x(\omega)$, на котором определена комплексная случайная функция $C_X(\omega)$. Её можно назвать *интегральным случайным спектром* процесса $X(t)$. Случайный процесс $X(t)$ представляется в виде среднеквадратического интеграла Фурье – Стилтеса

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t} dC_X(\omega). \quad (3.42)$$

Можно представить

$$dC_X(\omega) = X(\omega)d\omega,$$

где $X(\omega)$ – *случайная спектральная плотность* или *случайный спектр* процесса $X(t)$. Тогда выражение (3.42) принимает вид интеграла Фурье³

$$X(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{j\omega t} d\omega, \quad (3.43)$$

где случайный спектр $X(\omega)$ есть прямое Фурье-преобразование от случайной функции $X(t)$:

$$X(\omega) = \int_{-\infty}^{\infty} X(t)e^{-j\omega t} dt. \quad (3.44)$$

³Коэффициент $1/2\pi$ будем включать в обратное Фурье-преобразование, а прямое преобразование записывать без коэффициента. Часто в литературе в преобразованиях Фурье можно встретить обратный способ расстановки коэффициентов.

Выражение (3.42) охватывает случаи как непрерывного, так дискретного и смешанного случайного спектра. Представление в виде (3.43) удобно для чисто непрерывного спектра, когда $X(\omega)$ является интегрируемой в среднеквадратическом смысле функцией. Необходимым и достаточным условием этого служит интегрируемость ковариационной функции $K_X(t_1, t_2)$ по обоим аргументам на интервале $(-\infty, \infty)$. Если случайный сигнал представить дискретным спектром

$$X(t) = \sum_n C_n e^{j\omega_n t},$$

то можно формально ввести случайную спектральную плотность $X(\omega)$ как

$$X(\omega) = 2\pi \sum_n C_n \delta(\omega - \omega_n).$$

Случайная спектральная плотность $X(\omega)$ или интегральный случайный спектр $C_X(\omega)$ не могут использоваться для описания спектральных свойств случайного процесса, поскольку являются случайными функциями. Для этого необходимы статистические (усредненные) спектральные характеристики.

Спектральное представление нестационарного случайного процесса. Пусть $X(t)$ – скалярный вещественный случайный процесс с ограниченным средним квадратом $\overline{X^2(t)} \neq \infty$. Если существует среднеквадратический интеграл (3.44), то можно ввести ограниченную по абсолютной величине комплексную функцию

$$G_X(\omega_1, \omega_2) = \langle X(j\omega_1) X^*(j\omega_2) \rangle, \quad (3.45)$$

которую называют *обобщенной спектральной плотностью* случайного процесса $X(t)$. Подставляя (3.44) в (3.45) легко получить

$$G_X(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_X(t_1, t_2) e^{-j(\omega_1 t_1 - \omega_2 t_2)} dt_1 dt_2, \quad (3.46)$$

где $K_X(t_1, t_2)$ – ковариационная функция процесса $X(t)$. Таким образом, обобщенная спектральная плотность является двойным преобразованием Фурье от ковариационной функции. Соответственно, ковариационная функция $K_X(t_1, t_2)$ может быть представлена в виде двойного интеграла Фурье:

$$K_X(t_1, t_2) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_X(\omega_1, \omega_2) e^{j(\omega_1 t_1 - \omega_2 t_2)} d\omega_1 d\omega_2. \quad (3.47)$$

Мы можем не требовать существования интеграла (3.44) и определять обобщенную спектральную плотность $G_X(\omega_1, \omega_2)$ формально выражением (3.46) через ковариационную функцию. Однако среднеквадратическая сходимость интеграла (3.44) равносильна интегрируемости ковариационной функции по обоим аргументам. Если ковариационная функция не является интегрируемой, то обобщенная спектральная плотность

$G_X(\omega_1, \omega_2)$ уже не будет ограниченной по абсолютной величине и значение $|G_X(\omega_1, \omega_2)|$ для некоторых ω_1 и ω_2 обращается в бесконечность.

Спектральное представление стационарных случайных процессов. Пусть $X(t)$ – стационарный в широком смысле скалярный вещественный случайный процесс с ограниченным средним квадратом $\overline{X^2}(t) \neq \infty$. Ковариационная функция стационарного процесса зависит только от $\tau = t_2 - t_1$. Легко видеть, что она не является интегрируемой по двум аргументам t_1 и t_2 . Соответственно, функция $|G_X(\omega_1, \omega_2)|$ не является ограниченной. Можно показать, что

$$G_X(\omega, \omega') = 2\pi\delta(\omega' - \omega) \int_{-\infty}^{\infty} K_X(t_1, t_2)e^{-j\omega\tau} d\tau, \quad (3.48)$$

т.е. $|G_X(\omega_1, \omega_2)|$ обращается в бесконечность на диагонали $\omega_1 = \omega_2$. По этой причине обобщённую спектральную плотность нецелесообразно использовать в качестве спектральной характеристики для стационарных процессов.

Рассмотрим случайный спектр $X(t)$ на интервале $t \in [-T/2, T/2]$:

$$X_T(\omega) = \int_{-T/2}^{T/2} X(t)e^{-j\omega t} dt. \quad (3.49)$$

Для конечных значений T среднеквадратический интеграл (3.49) существует. Статистической спектральной характеристикой стационарного случайного процесса служит *спектральная плотность мощности*⁴ (или просто *спектральная плотность*). Она может быть введена как

$$W_X(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \langle |X_T(\omega)|^2 \rangle. \quad (3.50)$$

Можно показать, что для процесса с нулевым средним значением и ковариационной функцией $K_X(\tau)$, удовлетворяющей условию

$$\int_{-\infty}^{\infty} |\tau K_X(\tau)| d\tau < \infty, \quad (3.51)$$

из определения (3.50) следует, что спектральная плотность мощности есть прямое преобразование Фурье от ковариационной функции:

$$W_X(\omega) = \int_{-\infty}^{\infty} K_X(\tau)e^{-j\omega\tau} d\tau. \quad (3.52)$$

Обратное преобразование Фурье дает ковариационную функцию:

$$K_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_X(\omega)e^{j\omega\tau} d\omega. \quad (3.53)$$

⁴Термин «плотность мощности» является достаточно условным. В тех случаях, когда рассматриваемый процесс представляет собой флуктуации тока или напряжения, величина $W_X(\omega)$ действительно пропорциональна физической мощности, приведённой к интервалу частот. Однако в общем случае процесс $X(t)$ может иметь любую физическую природу и, соответственно, $W_X(\omega)$ может иметь иную размерность.

Из (3.52) и (3.48) следует также, что

$$G_X(\omega, \omega') = 2\pi\delta(\omega' - \omega)W_X(\omega).$$

Теорема Хинчина. Можно определить спектральную плотность мощности для любого стационарного случайного процесса, воспользовавшись следующей *теоремой Хинчина*. Пусть $K_X(\tau)$ – непрерывная ковариационная функция стационарного в широком смысле случайного процесса $X(t)$. Она всегда может быть представлена в виде интеграла Фурье – Стилтеса:

$$K_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} dZ_X(\omega), \quad (3.54)$$

где $Z_X(\tau)$ – монотонно неубывающая вещественная неотрицательная ограниченная функция, называемая *спектральной функцией* или *интегральным спектром стационарного случайного процесса*. Спектральная плотность мощности, соответственно, есть

$$W_X(\omega) = \frac{dZ_X(\omega)}{d\omega}. \quad (3.55)$$

Определённая таким образом функция $W_X(\omega)$ связана с ковариационной функцией преобразованиями Фурье (3.52) – (3.53), которые часто называют *соотношениями Винера – Хинчина*. Однако при этом, если не требовать интегрируемости ковариационной функции по τ , для некоторых частот $W_X(\omega)$ может принимать бесконечные значения.

Свойства спектральной плотности мощности. Спектральная плотность мощности $W_X(\omega)$ вещественного стационарного случайного процесса $X(t)$ обладает следующими свойствами.

1. $W_X(\omega)$ – вещественная неотрицательная функция⁵.
2. $W_X(\omega) = W_X(-\omega)$ (чётная функция).
3. В силу чётности функций $K_X(\tau)$ и $W_X(\omega)$ можно переписать (3.52) – (3.53) в виде

$$W_X(\omega) = 2 \int_0^{\infty} K_X(\tau) \cos(\omega\tau) d\tau, \quad (3.56)$$

$$K_X(\tau) = \frac{1}{\pi} \int_0^{\infty} W_X(\omega) \cos(\omega\tau) d\omega. \quad (3.57)$$

4. Функция $W_X(\omega)$ при некоторых значениях аргумента может иметь особенности в виде δ -функции (т.е. спектр в общем случае является смешанным), что связано с отсутствием требования интегрируемости ковариационной функции. Например, если среднее значение процесса \bar{X} отлично от нуля, то

⁵Это утверждение справедливо и для комплексного стационарного случайного процесса.

$$W_X(\omega) = W_{\tilde{X}}(\omega) + 2\pi\delta(\omega)\overline{X^2}, \quad (3.58)$$

где

$$W_{\tilde{X}}(\omega) = \int_{-\infty}^{\infty} \Psi_X(\tau)e^{-j\omega\tau} d\tau \quad (3.59)$$

– спектр флуктуаций.

5. Средний квадрат и дисперсию процесса $X(t)$ можно найти, интегрируя спектральную плотность мощности $W_X(\omega)$ или $W_{\tilde{X}}(\omega)$ соответственно по всем частотам:

$$\begin{aligned} \overline{X^2} &= K_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_X(\omega) d\omega, \\ \sigma_X^2 &= \Psi_X(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_{\tilde{X}}(\omega) d\omega. \end{aligned} \quad (3.60)$$

Кроме спектральной плотности мощности $W_X(\omega)$, определённой как для положительных, так и для отрицательных значений аргумента, часто используют так называемую *одностороннюю (физическую) спектральную плотность мощности* случайного процесса, которая отлична от нуля только для положительных частот и определяется следующим образом:

$$W_X^+(\omega) = 2W_X(\omega) \text{ при } \omega \geq 0 \text{ и } W_X^+(\omega) = 0 \text{ при } \omega < 0. \quad (3.61)$$

Взаимная спектральная плотность мощности двух стационарных стационарно-связанных процессов. Взаимная спектральная плотность мощности двух стационарных стационарно-связанных процессов может быть определена через следующий предел:

$$W_{XY}(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \langle X_T^*(\omega) Y_T(\omega) \rangle, \quad (3.62)$$

где $X_T(\omega)$ и $Y_T(\omega)$ – случайные спектры процессов на интервале $t \in [-T/2, T/2]$, но чаще её вводят как прямое Фурье-преобразование от взаимной ковариационной функции $K_{XY}(\tau)$:

$$W_{XY}(\omega) = \int_{-\infty}^{\infty} K_{XY}(\tau)e^{-j\omega\tau} d\tau. \quad (3.63)$$

В этом случае обратное преобразование позволяет найти взаимную ковариационную функцию:

$$K_{XY}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} W_{XY}(\omega)e^{j\omega\tau} d\omega. \quad (3.64)$$

Легко убедиться, что оба определения совпадают при выполнении условия

$$\int_{-\infty}^{\infty} |\tau K_{XY}(\tau)| d\tau < \infty. \quad (3.65)$$

Из этих определений следует, что взаимная спектральная плотность $W_{XY}(\omega)$ в общем случае будет принимать значения в области комплексных чисел. Так, если использовать определение (3.63), это следует из того обстоятельства, что взаимная ковариационная функция, в отличие от автоковариационной, не является чётной. Кроме того, $W_{XY}(\omega) \neq W_{\omega}(\tau)$.

С помощью взаимной спектральной плотности мощности определяют коэффициент когерентности двух процессов:

$$\Gamma(\omega) = \frac{|W_{XY}(\omega)|}{\sqrt{W_X(\omega)W_Y(\omega)}}, \quad (3.66)$$

который характеризует статистическую взаимосвязь их спектральных компонент.

Ширина спектра стационарного случайного процесса. Часто рассматривают такую характеристику, как *ширина спектра* стационарного случайного процесса. Если спектральная плотность мощности случайного процесса имеет один хорошо выраженный максимум, то используются следующие два определения ширины спектра:

1) *эффективная ширина спектра*

$$\Delta\omega_{\text{эфф}} = \frac{1}{W_{\tilde{X}}(\omega_0)} \int_0^{\infty} W_{\tilde{X}}(\omega) d\omega = \frac{\pi\sigma_X^2}{W_{\tilde{X}}(\omega_0)}, \quad (3.67)$$

где ω_0 – частота спектрального максимума;

2) *ширина спектра на уровне половинной мощности*

$$\Delta\omega_{1/2} = \omega_2 - \omega_1, \quad \text{где } W_{\tilde{X}}(\omega_{1,2}) = \frac{1}{2}W_{\tilde{X}}(\omega_0). \quad (3.68)$$

Ширину спектра и время корреляции стационарного случайного процесса связывает следующее соотношение (*соотношение неопределённости*): чем шире энергетический спектр случайного процесса, тем меньше время корреляции и, наоборот, чем уже спектр, тем больше время корреляции.

Примеры

1. $X(t)$ – стационарный случайный процесс с экспоненциально спадающей корреляционной функцией:

$$\Psi_X(\tau) = \sigma_X^2 e^{-\alpha|\tau|}, \quad \alpha > 0.$$

Время корреляции в этом случае есть $\tau_{\text{кор}} = 1/\alpha$. График корреляционной функции приведен на рис. 3.4, а. Используя формулу (3.59), легко найти спектр флуктуаций $W_{\tilde{X}}(\omega)$. Он имеет форму *лоренциана* (рис. 3.4, б):

$$W_{\tilde{X}}(\omega) = \frac{2\sigma_X^2\alpha}{\alpha^2 + \omega^2}.$$

Ширина спектра на уровне половинной мощности $\Delta\omega_{1/2} = 2\alpha = 2/\tau_{\text{кор}}$.

2. $X(t)$ – стационарный случайный процесс с экспоненциально спадающей корреляционной функцией:

$$\Psi_X(\tau) = \sigma_X^2 e^{-\alpha|\tau|} \cos \omega_0 \tau, \quad \alpha > 0.$$

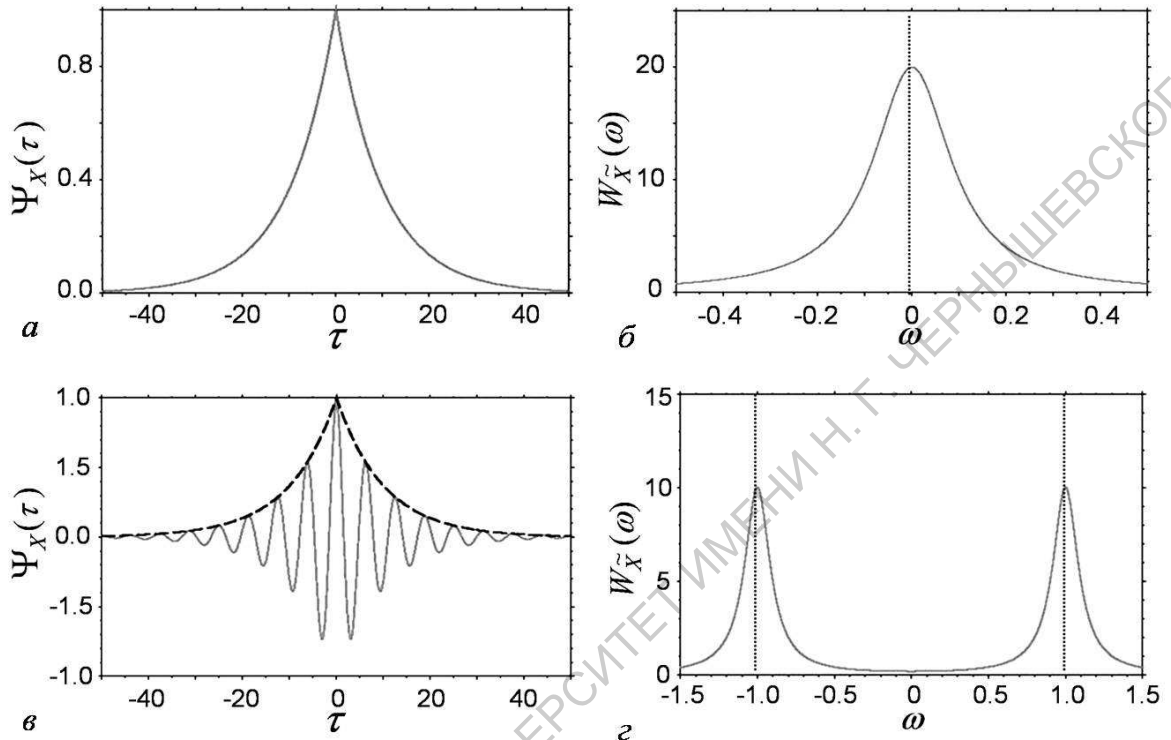


Рис. 3.4. Примеры корреляционных функций (а, в) и соответствующих им спектров флуктуаций (б, г)

График корреляционной функции приведен на рис. 3.4, в. Время корреляции определяется поведением огибающей корреляционной функции и, как и в предыдущем случае, есть $1/\alpha$. Спектр флуктуаций дается выражением

$$W_{\tilde{X}}(\omega) = \frac{\sigma_X^2 \alpha}{\alpha^2 + (\omega + \omega_0)^2} + \frac{\sigma_X^2 \alpha}{\alpha^2 + (\omega - \omega_0)^2}.$$

Он имеет форму двух лоренцианов с максимумами, смещёнными на частоты $\pm \omega_0$ (рис. 3.4, г). Если считать процесс достаточно узкополосным ($\alpha \ll 1$), то $\Delta\omega_{1/2} \approx 2\alpha$.

По спектральным свойствам различают широкополосные и узкополосные случайные процессы. Идеализированной моделью случайного процесса с бесконечно широким спектром является белый шум.

Белый шум. *Белый шум* – абсолютно случайный, стационарный в строгом смысле процесс, обладающий постоянной спектральной плотностью мощности на всех частотах $W_{\text{БШ}} \equiv W_0 \equiv \text{const}$. Легко видеть, что корреляционная функция такого процесса

$$\Psi_{\text{БШ}}(\tau) = \frac{W_0}{2\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} d\omega = W_0 \delta(\tau),$$

т.е. задается функцией Дирака, а дисперсия – бесконечна.

Таким образом, белый шум представляет последовательность некоррелированных случайных величин, а поскольку дисперсия бесконечна, то для него нельзя записать даже одномерную плотность вероятности.

Белый шум – это математическая идеализация. Строго говоря, в природе белого шума не существует, однако возможны процессы с очень малым временем корреляции, например тепловой шум резистивного сопротивления. Если во всем интересующем нас частотном диапазоне спектральная плотность процесса практически не меняется, то такой реальный процесс целесообразно заменить моделью белого шума.

Мы рассмотрели большую группу случайных процессов, стационарных и нестационарных, узкополосных и широкополосных, квазидетерминированных, эргодических и т.д. Для полного описания таких процессов использовались n -мерные плотности распределения. Они свидетельствуют о связи между значениями процесса в n произвольных моментах времени. При этом чем больше n , тем детальнее описание. Однако многие физические процессы для исчерпывающей характеристики не требуют знания большого числа p_n : достаточно ввести двумерную или даже одномерную плотность вероятности.

В двух следующих разделах главы будут проанализированы два вида таких случайных процессов: гауссовские и марковские, широко распространённые в различных областях радиоэлектроники, статистической радиофизики и связи. Сразу отметим, что двумерные плотности вероятности определяют для них распределения любого порядка.

3.5. Нормальные (гауссовские) случайные процессы

Нормальным (гауссовским) случайным процессом называется такой вещественный процесс $X(t)$, для которого любая совокупность значений $X(t_i)$, $i = 1, 2, \dots, n$ распределена по совместному нормальному (гауссовскому) закону:

$$p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = \frac{1}{\sqrt{(2\pi)^n \text{Det} \hat{\Psi}}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \tilde{\psi}_{ik}(t_i, t_k) (x_i - \bar{X}(t_i))(x_k - \bar{X}(t_k)) \right], \quad (3.69)$$

где $\hat{\Psi}$ – матрица корреляций с элементами

$$\psi_{ik} = \langle (X(t_i) - \bar{X}(t_i))(X(t_k) - \bar{X}(t_k)) \rangle,$$

а $\tilde{\psi}_{ik}$ – элементы обратной матрицы $\hat{\Psi}^{-1}$.

Можно показать, что плотности вероятности (3.69) соответствует характеристическая функция вида

$$\Theta_X(u_1, t_1, u_2, t_2, \dots, u_n, t_n) = \exp \left[j \sum_{i=1}^n u_i \bar{X}(t_i) - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \Psi_X(t_i, t_k) u_i u_k \right]. \quad (3.70)$$

Верно и обратное утверждение: если случайный процесс описывается характеристической функцией вида (3.70), то он является нормальным.

Одномерная гауссовская плотность вероятности выражается формулой

$$p_1(x, t) = \frac{1}{\sqrt{2\pi\sigma_X^2(t)}} \exp \left[-\frac{(x - \bar{X}(t))^2}{2\sigma_X^2(t)} \right],$$

где $\bar{X}(t)$ – среднее значение; $\sigma_X^2(t)$ – дисперсия случайного процесса (рис. 3.5, а). Площадь, ограниченная кривой $p_1(x)$, равна единице. Поэтому кривая нормального распределения тем сильнее сосредоточена около значения $\bar{X}(t)$, чем меньше среднееквадратическое отклонение $\sigma_X(t)$. В частности, в интервале $\bar{X}(t) - 0.67\sigma_X \div \bar{X}(t) + 0.67\sigma_X$ сосредоточено 50%, а в интервале $\bar{X}(t) - 3\sigma_X \div \bar{X}(t) + 3\sigma_X$ – 99.7% всех значений $X(t)$. Гауссовское распределение со средним $\bar{X}(t) \equiv 0$ и дисперсией $\sigma_X^2(t) \equiv 1$ называется *стандартным*.

Как следует из (3.69), двумерная гауссовская плотность вероятности может быть представлена соотношением (рис. 3.5, б):

$$p_2(x_1, t_1, x_2, t_2) = \frac{1}{2\pi\sqrt{\sigma_X^2(t_1)\sigma_X^2(t_2)(1 - R_X^2(t_1, t_2))}} \times \\ \times \exp \left[-\frac{1}{2(1 - R_X^2(t_1, t_2))} \left(\frac{(x_1 - \bar{X}(t_1))^2}{\sigma_X^2(t_1)} + \frac{(x_2 - \bar{X}(t_2))^2}{\sigma_X^2(t_2)} - \frac{2R_X(t_1, t_2)(x_1 - \bar{X}(t_1))(x_2 - \bar{X}(t_2))}{\sigma_X(t_1)\sigma_X(t_2)} \right) \right], \quad (3.71)$$

где $R_X(t_1, t_2)$ – коэффициент корреляции, $\sigma_X(t)$ – среднееквадратическое отклонение.

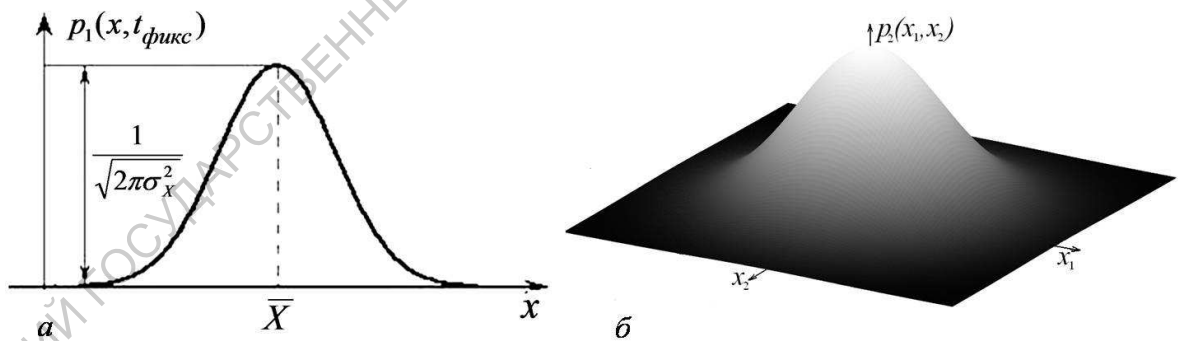


Рис. 3.5. Одномерное (а) и двумерное (б) гауссовские распределения (t, t_1, t_2 фиксированы)

Можно ввести понятие *совместно-гауссовских случайных процессов*. Процессы $X_k(t)$, $k = 1, 2, \dots, N$ называются *совместно-гауссовскими*, если любые совокупности их значений $X_k(t_i)$, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, N$ имеют совместное гауссовское распределение.

Комплексный случайный процесс $Z(t) = X(t) + jY(t)$ называется *комплексным гауссовским процессом*, если вещественные функции $X(t)$ и $Y(t)$ образуют двумерный гауссовский процесс.

Нормальные случайные процессы обладают следующими свойствами:

1) нормальный случайный процесс полностью задан, если известны функции $\bar{X}(t)$ и $\Psi_X(t_1, t_2)$;

2) некоррелированные значения нормального случайного процесса статистически независимы, т.е. при $R_X(t_i, t_j) = 0$

$$p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_X^2(t_i)}} e^{-\frac{(x_i - \bar{X}(t_i))^2}{2\sigma_X^2(t_i)}};$$

3) для нормальных случайных процессов с ограниченной дисперсией свойства стационарности в строгом смысле и в широком смысле совпадают;

4) если $X_1(t), X_2(t), \dots, X_m(t)$ – совместно нормальные случайные процессы, то их линейная комбинация $X(t) = \sum_{i=1}^m a_i(t)X_i(t)$, где $a_i(t)$ – детерминированные множители, также является нормальным случайным процессом;

5) случайный процесс, получаемый в результате линейного преобразования нормального случайного процесса, также является нормальным.

Нормальные (гауссовские) процессы занимают особое место в радиоэлектронике и статистической радиофизике, так как позволяют с хорошим приближением аппроксимировать реальные сигналы и помехи. Тепловые и дробовые шумы электронных приборов, собственные шумы приемо-передающих устройств, радиоизлучение космических объектов – вот далеко неполный перечень случайных процессов с нормальным распределением. Но многие атмосферные и промышленные помехи, а также шумы в полупроводниковых приборах имеют распределения, отличные от гауссовского. Поэтому в последние годы радиофизики, в особенности специалисты в области статистической теории связи, теории оптимального управления и распознавания образов, обратились к *рандомизированным распределениям*, представляющим вероятностные смеси стандартных распределений. Частным случаем такого распределения является полигауссовская модель, описывающая статистические свойства линейных смесей гауссовских распределений.

Случайный сигнал называют *полигауссовским*, если его плотность распределения $p_n(\dots)$ можно представить линейной комбинацией гауссовских распределений:

$$p_n(\dots) = \sum_{i=1}^n q_i p_i(\dots), \quad \sum_{i=1}^n q_i = 1,$$

где $p_i(\dots)$ – гауссовские плотности вероятности, различающиеся средними значениями, ковариациями или обеими одновременно, q_i – нормированные вероятности.

Популярности гауссовских моделей способствует широкое использование центральной предельной теоремы теории вероятностей, доказанной при достаточно общих условиях А. М. Ляпуновым.

Центральная предельная теорема теории вероятностей. Если $X_1(t), X_2(t), \dots, X_n(t)$ – независимые случайные процессы с конечными, отличными от нуля дисперсиями, то плотность вероятности их суммы $Y(t) = \sum_{i=1}^n X_i(t)$ при $n \rightarrow \infty$ асимптотически приближается к нормальной.

Реальные помехи, в частности собственные шумы радиоприёмных устройств образуются в результате наложения большого числа колебаний, и вероятностные свойства суммарного процесса в силу центральной предельной теоремы удовлетворительно согласуются со свойствами гауссовского случайного процесса.

3.6. Марковские случайные процессы

Основные определения. Для любого случайного процесса $X(t)$ n -мерная плотность вероятности может быть представлена в виде:

$$p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = p_{n-1}(x_1, t_1, x_2, t_2, \dots, x_{n-1}, t_{n-1}) \cdot p(x_n, t_n/x_{n-1}, t_{n-1}, \dots, x_1, t_1),$$

где $p(x_n, t_n/x_{n-1}, t_{n-1}, \dots, x_1, t_1)$ – условная плотность вероятности, учитывающая $n - 1$ предшествующее состояние процесса. Условная плотность вероятности неотрицательна и нормирована на единицу:

$$\int_{-\infty}^{\infty} p(x_n, t_n/x_{n-1}, t_{n-1}, \dots, x_1, t_1) dx_n \equiv 1. \quad (3.72)$$

Случайный процесс $X(t)$, для которого условная плотность вероятности зависит только от одного предшествующего состояния, т.е. для любых $t_n < t_{n-1} < t_{n-2} < \dots < t_1$ справедливо равенство

$$p(x_n, t_n/x_{n-1}, t_{n-1}, \dots, x_1, t_1) = p(x_n, t_n/x_{n-1}, t_{n-1}), \quad (3.73)$$

называется *марковским (простейшим марковским)* случайным процессом⁶. Условная плотность вероятности в этом случае называется *плотностью вероятности перехода* из предшествующего состояния в последующее.

В зависимости от дискретности или непрерывности множества состояний и множества моментов времени различают дискретные марковские последовательности (*марковские цепи*), непрерывные марковские последовательности, дискретные марковские процессы и непрерывнозначные марковские процессы, среди которых выделяют *диффузионные процессы*, которые наиболее часто встречаются в физических и радиофизических задачах.

Марковский процесс может быть векторным (многокомпонентным). Марковское свойство для векторного процесса определяется так же, как для скалярного:

$$p(\mathbf{x}_n, t_n/\mathbf{x}_{n-1}, t_{n-1}, \dots, \mathbf{x}_1, t_1) = p(\mathbf{x}_n, t_n/\mathbf{x}_{n-1}, t_{n-1}).$$

Следует отметить, что марковость векторного процесса $\mathbf{X}(t)$ в общем случае ещё не означает, что его компоненты – марковские процессы. Марковский процесс называется *однородным*, если

$$p(x_2, t_2/x_1, t_1) = p(x_2, \tau/x_1), \quad \text{где } \tau = t_2 - t_1.$$

⁶Если условная плотность вероятности зависит от k предшествующих состояний, то процесс называется *k-связанным марковским* процессом.

Марковский процесс обладает следующими свойствами.

1. Если точно известно состояние системы в момент времени t_1 , то любое состояние в момент $t_2 > t_1$ не зависит от состояний системы в моменты времени $t < t_1$ (т.е. от предыстории процесса).

2. Для марковского процесса $X(t)$ справедливо равенство

$$p_n(x_1, t_1, x_2, t_2, \dots, x_n, t_n) = p_1(x_1, t_1) \prod_{i=2}^n p(x_i, t_i / x_{i-1}, t_{i-1}) \quad (3.74)$$

и, следовательно, он полностью задан, если известны функции $p_1(x, t)$ и $p(x_2, t_2 / x_1, t_1)$.

3. Плотность вероятности перехода определяет эволюцию одномерной плотности вероятности во времени:

$$p_1(x, t) = \int_{-\infty}^{\infty} p(x, t / x_0, t_0) p_1(x_0, t_0) dx_0. \quad (3.75)$$

Уравнение Чепмена – Колмогорова. Рассмотрим состояния марковского процесса $X(t)$ в три последовательных момента времени $t_1 < t_2 < t_3$. Для плотности вероятности $p_3(x_1, t_1, x_2, t_2, x_3, t_3)$ в соответствии с (3.74) можно записать:

$$p_3(x_1, t_1, x_2, t_2, x_3, t_3) = p_1(x_1, t_1) p(x_2, t_2 / x_1, t_1) p(x_3, t_3 / x_2, t_2).$$

Интегрируя обе части этого равенства по x_2 , получаем

$$\int_{-\infty}^{\infty} p_3(x_1, t_1, x_2, t_2, x_3, t_3) dx_2 = p_1(x_1, t_1) \int_{-\infty}^{\infty} p(x_2, t_2 / x_1, t_1) p(x_3, t_3 / x_2, t_2) dx_2.$$

Учитываем условие согласованности:

$$\int_{-\infty}^{\infty} p_3(x_1, t_1, x_2, t_2, x_3, t_3) dx_2 = p_2(x_1, t_1, x_3, t_3),$$

представляем $p_2(x_1, t_1, x_3, t_3) = p_1(x_1, t_1) p(x_3, t_3 / x_1, t_1)$ и делим обе части равенства на $p_1(x_1, t_1)$. В результате получаем следующее уравнение:

$$p(x_3, t_3 / x_1, t_1) = \int_{-\infty}^{\infty} p(x_2, t_2 / x_1, t_1) p(x_3, t_3 / x_2, t_2) dx_2. \quad (3.76)$$

Уравнение (3.76) называется *уравнением Чепмена – Колмогорова*, *уравнением Смолуховского* или *обобщённым уравнением Маркова*. Оно справедливо для любых марковских процессов, в том числе и векторных.

Пример марковского процесса – броуновское движение. Броуновская частица совершает движение под действием толчков со стороны молекул среды (рис. 3.6). В силу случайного характера движения молекул толчки являются случайными и статистически независимыми. Если рассматривать изменения координаты «в крупном масштабе времени», когда за шаг по времени Δt , с которым производится измерение, частица испытывает большое число случайных толчков, то смещение ΔX можно считать не зависящим от предшествующих смещений.

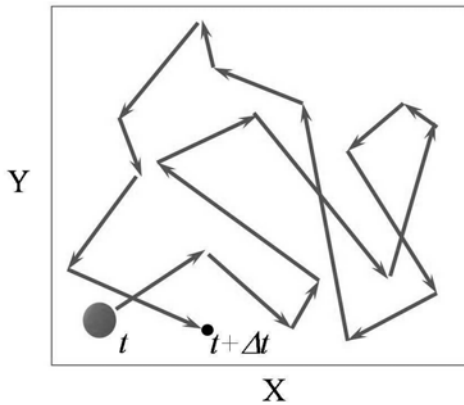


Рис. 3.6. Движение броуновской частицы как пример марковского процесса

$\Delta X_1 = X(t_2) - X(t_1)$, $\Delta X_2 = X(t_3) - X(t_2)$, ..., $\Delta X_n = X(t_{n+1}) - X(t_n)$ являются статистически независимыми случайными величинами. Тогда

$$p(x_{n+1}, t_{n+1}/x_n, t_n, \dots, x_1, t_1) = p(\Delta x_n/\Delta x_{n-1}, \dots, \Delta x_1),$$

где $p(\Delta x_n/\Delta x_{n-1}, \dots, \Delta x_1)$ – плотность вероятности приращения ΔX_n при заданных приращениях на предшествующих интервалах. В силу независимости приращений получаем

$$p(\Delta x_n/\Delta x_{n-1}, \dots, \Delta x_1) = p(\Delta x_n) = p(x_{n+1}, t_{n+1}/x_n, t_n),$$

$$p(x_{n+1}, t_{n+1}/x_n, t_n, \dots, x_1, t_1) = p(x_{n+1}, t_{n+1}/x_n, t_n),$$

что по определению означает марковское свойство.

Любую компоненту $V(t)$ скорости тяжёлой броуновской частицы (такой, что можно пренебречь силой сопротивления среды) также можно рассматривать как марковский процесс. Действительно, приращение скорости за время Δt зависит от значения V в момент t (частица испытывает больше встречных соударений, чем соударений вдогонку) и не зависит от значений скорости в предшествующие моменты времени.

3.6.1. Марковские цепи

Дискретная марковская последовательность, т.е. подкласс марковских случайных процессов с дискретным множеством состояний и с дискретным временем, называется *марковской цепью* (МЦ).

Пусть $X(i)$ – марковская цепь со значениями x_m , где i – дискретное время, m – номер состояния. Множество возможных значений x_m может быть конечным ($m = 1, 2, \dots, M < \infty$) или счётным ($m = 0, \pm 1, \pm 2, \dots, \pm \infty$). Марковскую цепь с конечным множеством состояний называют *конечной* МЦ.

Введем следующие обозначения: $P_m(i)$ – вероятность значения x_m в момент времени i ; $P(m_1, i_1, \dots, m_k, i_k)$ – совместная вероятность появления значений x_{m_1}, \dots, x_{m_k} в моменты времени i_1, \dots, i_k , соответственно; $\pi(m_n, i_n/m_{n-1}, i_{n-1}, m_{n-2}, i_{n-2}, \dots)$ – условная вероятность значения x_{m_n}

Таким образом, изменение во времени координаты броуновской частицы есть (в некотором приближении) случайный процесс с независимыми приращениями на неперекрывающихся интервалах времени. Любой случайный процесс с независимыми приращениями является простейшим марковским процессом. Обратное утверждение в общем случае неверно.

То, что из независимости приращений следует марковское свойство, можно пояснить следующим образом. Рассмотрим последовательность отсчетов времени $t_1 < t_2 < \dots < t_n < t_{n+1}$. Пусть приращения

$t_1 < t_2 < \dots < t_n < t_{n+1}$. Пусть приращения

в момент времени i_n при условии, что в предшествующие моменты времени i_{n-1}, i_{n-2}, \dots случайной переменной принимались значения $x_{m_{n-1}}, x_{m_{n-2}}, \dots$ соответственно.

Марковская цепь, для которой условная вероятность в момент i_n зависит только от состояния в непосредственно предшествующий момент времени i_{n-1} , т.е.

$$\pi(m_n, i_n/m_{n-1}, i_{n-1}, m_{n-2}, i_{n-2}, \dots) = \pi(m_n, i_n/m_{n-1}, i_{n-1}), \quad (3.77)$$

называется *простой цепью Маркова*, а условная вероятность – *вероятностью перехода* из состояния $x_{m_{n-1}}$ в момент времени i_{n-1} в состояние x_{m_n} в момент времени i_n . Будем обозначать $\pi(m_n, i_n/m_{n-1}, i_{n-1})$ как $\pi_{m_{n-1}m_n}(i_{n-1}, i_n)$. Далее будем говорить о марковских цепях, подразумевая простые МЦ.

Вероятность перехода МЦ удовлетворяет следующим условиям:

$$\pi_{m_1m_2}(i_1, i_2) \geq 0; \quad \sum_{m_2} \pi_{m_1m_2}(i_1, i_2) = 1; \quad \pi_{m_1m_2}(i_1, i_1) = \delta_{m_1m_2}, \quad (3.78)$$

где $\delta_{m_1m_2}$ – символ Кронеккера. Для МЦ равенство (3.74) превращается в следующее равенство:

$$P(m_1, i_1, \dots, m_n, i_n) = P_{m_1}(i_1) \prod_{k=2}^n \pi_{m_{k-1}m_k}(i_{k-1}, i_k). \quad (3.79)$$

Выражение (3.75) можно представить как

$$P_m(i) = \sum_{m_0} \pi_{m_0m}(i_0, i) P_{m_0}(i_0). \quad (3.80)$$

Уравнение Чепмена – Колмогорова для МЦ принимает вид

$$\pi_{m_1m_3}(i_1, i_3) = \sum_{m_2} \pi_{m_1m_2}(i_1, i_2) \pi_{m_2m_3}(i_2, i_3) \quad (3.81)$$

и называется *уравнением Маркова*.

Конечная МЦ с числом состояний M задается с помощью стохастической матрицы

$$\hat{\pi}(i_1, i_2) = \begin{bmatrix} \pi_{11}(i_1, i_2) & \pi_{12}(i_1, i_2) & \dots & \pi_{1M}(i_1, i_2) \\ \dots & \dots & \dots & \dots \\ \pi_{M1}(i_1, i_2) & \pi_{M2}(i_1, i_2) & \dots & \pi_{MM}(i_1, i_2) \end{bmatrix},$$

где π_{mm} характеризуют вероятность задержки в m -ом состоянии. При этом уравнение Маркова может быть записано в матричном виде как произведение матриц

$$\hat{\pi}(i_1, i_3) = \hat{\pi}(i_1, i_2) \hat{\pi}(i_2, i_3). \quad (3.82)$$

Обозначим через $\hat{P}(i)$ – вектор-строку, составленную из вероятностей различных значений случайной переменной в момент времени i :

$$\hat{P}(i) = (P_1(i), P_2(i), \dots, P_M(i)). \quad (3.83)$$

Тогда равенство (3.80) принимает вид

$$\widehat{P}(i) = \widehat{P}(i_0)\widehat{\pi}(i_0, i). \quad (3.84)$$

Цепь Маркова называется *однородной*, если переходные вероятности $\pi_{m_1 m_2}$ не зависят от номера шага и определяются только разностью рассматриваемых моментов времени. Поэтому конечную однородную МЦ можно описать с помощью *матрицы одношаговых переходных вероятностей* $\widehat{\pi}_1$, элементами которой являются вероятности переходов между различными состояниями за один шаг по времени:

$$\widehat{\pi}_1 = \begin{bmatrix} \pi_{11}(1) & \pi_{12}(1) & \dots & \pi_{1M}(1) \\ \dots & \dots & \dots & \dots \\ \pi_{M1}(1) & \pi_{M2}(1) & \dots & \pi_{MM}(1) \end{bmatrix}.$$

Теперь вместо (3.84) можно записать

$$\widehat{P}(i) = \widehat{P}(i_0)\widehat{\pi}_1^i, \quad (3.85)$$

где $\widehat{\pi}_1^i = \widehat{\pi}_1 \cdot \widehat{\pi}_1 \cdot \dots \cdot \widehat{\pi}_1$ (i – раз).

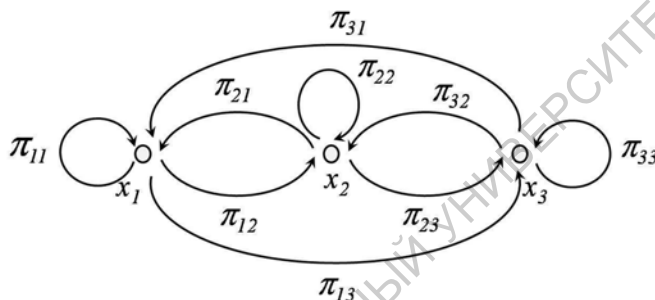


Рис. 3.7. Пример представления марковской цепи в виде графа

Однородные МЦ удобно представлять в виде графов, узлы которых соответствуют состояниям системы, рёбра с указанными направлениями показывают возможные переходы между состояниями. Над рёбрами отмечаются одношаговые вероятности соответствующих переходов. Пример изображения конечной однородной цепи с тремя состояниями приведен на рис. 3.7.

У ряда случайных процессов по истечении длительного времени устанавливается такой стационарный режим, когда состояния системы хотя и меняются случайным образом, но их вероятности $P_m(i)$ остаются постоянными $P_m^{\text{ст}} = \lim_{i \rightarrow \infty} P_m(i)$. Вероятности $P_m^{\text{ст}}$, если они существуют, называются *стационарными (финальными)*. Исследование стационарных (финальных) вероятностей представляет одну из важных задач теории марковских цепей. Для простой конечной однородной МЦ стационарное распределение вероятностей $P_m^{\text{ст}}$ ($m = 1, 2, \dots$) существует и не зависит от начального распределения $P_m(i_0)$. Вектор стационарных (финальных) вероятностей $\widehat{P}^{\text{ст}}$ является решением уравнения

$$\widehat{P}^{\text{ст}} = \widehat{P}^{\text{ст}}\widehat{\pi}_1. \quad (3.86)$$

Таким образом, для стационарных вероятностей можно записать следующую систему алгебраических уравнений:

$$P_m^{\text{ст}} = \sum_{m_0} P_{m_0}^{\text{ст}} \pi_{m_0 m}, \quad m = 1, 2, 3, \dots, M. \quad (3.87)$$

Однако система (3.87) не является линейно независимой. Для того, чтобы получить систему линейно независимых уравнений, нужно выбросить любое из уравнений системы (3.87) и дополнить систему условием нормировки: $\sum_m P_m^{\text{ст}} = 1$. Решая полученную систему M линейно независимых уравнений относительно M неизвестных, находим стационарное распределение вероятностей МЦ.

Пример марковской цепи: случайные блуждания с дискретным временем. Одномерные случайные блуждания с дискретным временем – это скалярная МЦ со счётным множеством состояний, причём возможны переходы только между «соседними» состояниями. Для любых m и i имеем: $\pi_{mm+1}(i, i+1) = p = \text{const}$ (вероятность шага вправо); $\pi_{mm-1}(i, i+1) = q = \text{const}$ (вероятность шага влево); $\pi_{mm}(i, i+1) = 1 - p - q$ (вероятность остаться на месте). Если номер состояния принимает любые целые значения ($m = 0, \pm 1, \pm 2, \dots$), то блуждания называются *неограниченными*.

Классической задачей, связанной со случайными блужданиями, является следующая: человек в каждый момент времени $i = 1, 2, \dots$ делает шаг наугад по прямой влево и вправо или остается на месте с постоянными вероятностями. Длина шага полагается постоянной. Требуется найти вероятность того, что через заданное время (число шагов) он достигнет заданной точки.

Пусть человек стартует из точки x_0 ($m = 0$). Вероятность того, что за время n он сделает n_0 «нулевых» шагов, n_1 шагов вправо и n_2 шагов влево имеет полиномиальное распределение

$$P(n_0, n_1, n_2) = \frac{n!}{n_0!n_1!n_2!} p^{n_1} q^{n_2} (1 - p - q)^{n_0},$$

где $n_0 + n_1 + n_2 = n$. Чтобы через n шагов попасть в состояние x_m , нужно, чтобы $n_1 - n_2 = m$. Тогда вероятность состояния с номером m в момент n (от начала ходьбы) есть

$$P_m(n) = \sum \frac{n!}{n_0!n_1!n_2!} p^{n_1} q^{n_2} (1 - p - q)^{n_0},$$

где сумма берется по всем n_0, n_1, n_2 , удовлетворяющих соотношениям

$$n_0 + n_1 + n_2 = n, \quad n_1 - n_2 = m.$$

В случае $p + q = 1$ («нулевые» шаги исключаются), получаем распределение Бернулли:

$$P_m(n) = \frac{n!}{n_1!n_2!} p^{n_1} q^{n_2},$$

где n_1 и n_2 однозначно определены через n и m : $n_1 = (n + m)/2$, $n_2 = (n - m)/2$. В случае $q = p = 1/2$ распределение Бернулли дает вероятность m -кратного появления одной стороны монеты при n подбрасываниях.

Случайные блуждания являются примером процесса с независимыми приращениями. Обозначим длину шага за Δ , тогда $x_m = m\Delta$, $m = 0 \pm 1, \pm 2, \dots$. Значение случайной переменной в момент времени i представляется в виде:

$$X(i) = x(0) + \sum_{k=1}^i \lambda_k,$$

где $x(0)$ – заданное начальное состояние. Случайные приращения λ_k представляют собой статистически независимые случайные величины с тремя возможными значениями: значением Δ , принимаемым с вероятностью p , значением $-\Delta$ – с вероятностью q и значением 0 – с вероятностью $1 - p - q$. Легко найти среднее значение $\bar{X}(i)$ и дисперсию $\sigma_X^2(i)$ случайных блужданий $X(i)$:

$$\bar{X}(i) = x(0) + i\bar{\lambda} = x(0) + i(p - q)\Delta, \quad \sigma_X^2(i) = i\sigma_\lambda^2 = i(p + q - (p - q)^2)\Delta^2.$$

Таким образом, дисперсия случайных блужданий растет линейно во времени.

Мы описали выше так называемые неограниченные случайные блуждания. Возможны также задачи, где блуждания рассматриваются в ограниченном интервале $[x_{\text{гр}1}, x_{\text{гр}2}]$, $x_{\text{гр}1}$ и $x_{\text{гр}2}$ – граничные значения. Иногда задается только одна граница. Имеются различные типы границ.

1. *Поглощающая граница* – в этом случае при достижении границы процесс заканчивается. Решается задача о среднем времени выхода на границу;

2. *Отражающая граница* – в этом случае, если в момент времени i достигается граница ($X(i) = x_{\text{гр}}$), то в следующий момент $i + 1$ либо с вероятностью r принимается значение $x_{\text{гр}} \pm \Delta$ (знак \pm определяется тем, где располагается граница – слева или справа), либо с вероятностью $1 - r$ значение остается равным $x_{\text{гр}}$. Случай $r = 1$ соответствует *жёсткой отражающей границе*. Если имеются две отражающие границы, то множество возможных состояний – конечно и решается задача отыскания стационарного распределения вероятностей $P_m^{\text{ст}}$.

3.6.2. Дискретные марковские процессы

Пусть теперь $X(t)$ – марковский процесс с непрерывным временем и дискретным множеством значений случайной переменной. Множество возможных значений x_m (состояний системы) может быть конечным ($m = 1, 2, \dots, M < \infty$) или счётным ($m = 0, \pm 1, \pm 2, \dots, \pm \infty$). Введем обозначения: $P_m(t)$ – вероятность состояния x_m в момент времени t ; $\pi_{m_1 m_2}(t_1, t_2)$ – вероятность перехода из состояния x_{m_1} в момент времени t_1 в состояние x_{m_2} в момент времени t_2 . Для состояний в три последовательные момента времени $t_1 < t_2 < t_3$ справедливо уравнение Чепмена – Колмогорова:

$$\pi_{m_1 m_3}(t_1, t_3) = \sum_{m_2} \pi_{m_1 m_2}(t_1, t_2) \pi_{m_2 m_3}(t_2, t_3). \quad (3.88)$$

Для дискретного марковского процесса можно ввести вероятности переходов $a_{m_1 m}(t)$ в единицу времени. При малых Δt вероятность $\pi_{m_1 m}$ перехода из состояния m_1 в состояние m и вероятность перехода в единицу времени связаны соотношением

$$\pi_{m_1 m}(t, t + \Delta t) = \delta_{m_1 m} + a_{m_1 m} \Delta t^7,$$

⁷ Это соотношение должно содержать ещё одно слагаемое $o(\Delta t)$, соответствующее членам второго и более высоких порядков малости относительно Δt , но $\lim_{\Delta t \rightarrow 0} o(\Delta t)/\Delta t = 0$, и в дальнейших выкладках это слагаемое опущено.

где $\delta_{m_1 m}$ – символ Кронеккера, т.е.

$$\begin{aligned}\pi_{m_1 m}(t, t + \Delta t) &= a_{m_1 m}(t)\Delta t, \quad m \neq m_1, \\ \pi_{mm}(t, t + \Delta t) &= 1 + a_{mm}(t)\Delta t.\end{aligned}\quad (3.89)$$

Последнее равенство выражает следующее свойство дискретных марковских процессов: при малых приращениях времени плотность вероятности π_{mm} того, что текущее состояние за время Δt не изменится, превышает вероятность изменения состояния.

Поскольку суммарная вероятность перехода в одно из возможных состояний равна единице, то

$$\sum_{m_1} \pi_{m_1 m}(t_1, t) = 1, \quad \sum_{m_1} a_{m_1 m}(t) = 0, \quad a_{m_1 m_1}(t) = - \sum_{m \neq m_1} a_{m_1 m}(t). \quad (3.90)$$

Исходя из (3.88), (3.89) и (3.90) можно получить дифференциальные уравнения для вероятностей переходов и вероятностей состояния x_m .

Рассмотрим три момента времени $t_0 < t_1 < t$, которым соответствуют состояния с номерами m_0, m_1, m . Обозначим $\Delta t = t - t_1$ и будем считать, что $\Delta t \rightarrow 0$. С учетом (3.89) и (3.90) перепишем уравнение (3.88) в виде

$$\pi_{m_0 m}(t_0, t) = \sum_{m_1 \neq m} \pi_{m_0 m_1}(t_0, t_1) a_{m_1 m}(t_1) \Delta t + \pi_{m_0 m}(t_0, t_1) (1 + a_{mm}(t_1) \Delta t).$$

Отсюда

$$\frac{\pi_{m_0 m}(t_0, t) - \pi_{m_0 m}(t_0, t_1)}{\Delta t} = \sum_{m_1} \pi_{m_0 m_1}(t_0, t_1) a_{m_1 m}(t_1).$$

При $\Delta t \rightarrow 0$, т.е. $t_1 \rightarrow t$, получаем *прямое эволюционное уравнение для вероятностей перехода*:

$$\frac{\partial \pi_{m_0 m}(t_0, t)}{\partial t} = \sum_{m_1} \pi_{m_0 m_1}(t_0, t) a_{m_1 m}(t) \quad (3.91)$$

с начальным условием $\pi_{m_0 m}(t_0, t_0) = \delta_{m_0 m}$.

Аналогично, рассматривая $\Delta t = t_1 - t_0 \rightarrow 0$, можно получить *эволюционное уравнение в обратном времени*:

$$\frac{\partial \pi_{m_0 m}(t_0, t)}{\partial t_0} = - \sum_{m_1} \pi_{m_1 m}(t_0, t) a_{m_0 m_1}(t_0) \quad (3.92)$$

с начальным условием $\pi_{m_0 m}(t, t) = \delta_{m_0 m}$. Здесь значения t и m заданы, а t_0 и m_0 считаются переменными.

Умножая прямое уравнение (3.91) на $P_{m_0}(t_0)$ и суммируя по m_0 , получаем *эволюционное уравнение для вероятностей состояний*:

$$\frac{dP_m(t)}{dt} = \sum_{m_1} P_{m_1}(t) a_{m_1 m}(t), \quad (3.93)$$

называемое *управляющим уравнением*. Управляющее уравнение может быть переписано в несколько ином виде:

$$\frac{dP_m(t)}{dt} = \sum_{m_1} [a_{m_1 m}(t)P_{m_1}(t) - a_{m m_1}(t)P_m(t)]. \quad (3.94)$$

В такой форме записи виден балансный характер уравнения, отражающего «приход–расход» вероятностей состояний. При заданных $a_{m_1 m}(t)$ система любых $M - 1$ уравнений из M уравнений (3.93) или (3.94) ($m = 1, 2, \dots, M$), дополненная условием нормировки, позволяет при заданных начальных условиях определить эволюцию распределения вероятностей состояний во времени. Если существуют стационарные (финитные) вероятности, то их можно найти при $\frac{dP_m(t)}{dt} = 0$, т.е. решая, например, следующую систему линейных алгебраических уравнений:

$$\sum_{m_1=1}^M P_{m_1}^{\text{ст}} a_{m_1 m} = 0, \quad m = 1, 2, \dots, M - 1, \quad \sum_{m=1}^M P_m^{\text{ст}} = 1. \quad (3.95)$$

Дискретные марковские процессы генерации-рекомбинации.

Дискретные марковские процессы, для которых возможны переходы только между соседними состояниями с номерами, отличающимися на единицу, называются *одношаговыми процессами, процессами генерации-рекомбинации* или *процессами «рождения – гибели»*⁸.

Введем следующие обозначения:

- 1) $a_{m+1 m}(t) = r_{m+1}(t)$ – вероятность перехода («рекомбинации») из состояния с номером $m + 1$ в состояние с номером m в единицу времени;
- 2) $a_{m-1 m}(t) = g_{m-1}(t)$ – вероятность перехода («генерации») из состояния с номером $m - 1$ в состояние с номером m в единицу времени.

Переходы в состояние с номером m из состояний с номерами $m_1 > m + 1$ или $m_1 < m - 1$ невозможны. Для коэффициентов $a_{m m}(t)$ получаем: $a_{m m}(t) = -(r_m(t) + g_m(t))$. Управляющее уравнение для одношагового процесса принимает вид:

$$\begin{aligned} \frac{dP_m(t)}{dt} = & r_{m+1}(t)P_{m+1}(t) + g_{m-1}(t)P_{m-1}(t) - \\ & - (r_m(t) + g_m(t))P_m(t), \quad m = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (3.96)$$

Если число состояний ограничено ($m = 1, 2, \dots, M$), то надо задавать граничные условия. Для отражающих границ $r_1 = g_0 = 0$, $r_{M+1} = g_M = 0$. Можно также отдельно записать вид управляющего уравнения для граничных элементов.

Для дискретных марковских процессов и особенно для одношаговых процессов с постоянными во времени вероятностями перехода r_m и g_m , так же как для марковских цепей, удобно использовать представление в виде графов. Узлы соответствуют возможным состояниям, а рёбра с

⁸Последнее название возникло в биологии, так как дискретные марковские процессы хорошо описывают статистическое поведение биологических популяций: размножения и гибели входящих в популяцию особей.

указанным направлением – возможным переходам. Над рёбрами проставляются соответствующие вероятности переходов в единицу времени.

Если вероятности переходов одинаковы для всех состояний и не зависят от времени, т.е. $r_m(t) \equiv r = \text{const}$ и $g_m(t) \equiv g = \text{const}$, то дискретный марковский процесс $X(t)$ называется *случайными блужданиями* (с непрерывным временем). В этом случае управляющее уравнение записывается как

$$\frac{dP_m(t)}{dt} = rP_{m+1}(t) + gP_{m-1}(t) - (r + g)P_m(t). \quad (3.97)$$

Рассмотрим два примера случайных блужданий с непрерывным временем: пуассоновский процесс и случайный двоичный телеграфный сигнал.

1. *Пуассоновский процесс* представляет собой неограниченные случайные блуждания, для которых $r = 0$. В начальный момент система находится в состоянии с номером $m = 1$ и далее возможны переходы только в соседние состояния с номерами $m = 2, 3, \dots$. Граф пуассоновского процесса представлен на рис. 3.8.

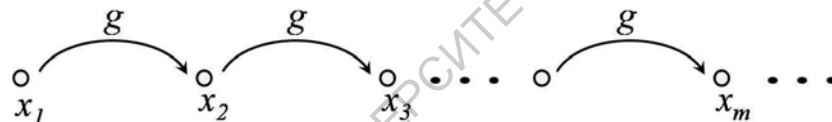


Рис. 3.8. Изображение пуассоновского процесса в виде графа

Соответствующие управляющие уравнения имеют вид:

$$\begin{aligned} \frac{dP_1(t)}{dt} &= -gP_1(t), \\ \frac{dP_m(t)}{dt} &= g(P_{m-1}(t) - P_m(t)), \quad m = 2, 3, \dots, \infty. \end{aligned} \quad (3.98)$$

Пуассоновский процесс, в частности, описывает количество событий (например числа имитируемых катодом электронов) на интервале (t_0, t) , если вероятность появления события в единицу времени (относительная частота событий) есть некоторая константа g . В этом случае пуассоновский процесс принимает целочисленные значения.

Пусть задано следующее начальное распределение: $P_1(t_0) = 1$, $P_m(t_0) = 0$, $m = 2, 3, \dots$. Найдём вероятности состояний в произвольный момент времени t . Решая однородное линейное дифференциальное уравнение первого порядка для вероятности первого состояния, получаем

$$P_1(t) = e^{-g(t-t_0)}.$$

Для вероятности второго состояния имеем неоднородное линейное уравнение первого порядка вида

$$\frac{dy}{dt} + gy = f(t),$$

решение которого с начальным условием $y(t_0) = 0$ есть

$$y(t) = e^{-g \cdot (t-t_0)} \int_{t_0}^t e^{g \cdot (\theta-t_0)} f(\theta) d\theta.$$

Аналогичный вид имеют уравнения для всех прочих состояний $m > 2$, причём в уравнении для каждого последующего состояния в правую часть $f(t)$ входит решение уравнения для предыдущего состояния. В результате последовательного решения уравнений приходим к формуле:

$$P_m(t) = \frac{g^{m-1} \cdot (t-t_0)^{m-1} e^{-g(t-t_0)}}{(m-1)!}, \quad m = 1, 2, 3, \dots$$

Очевидно, в силу переходов только в одном направлении (только «генерация») пуассоновский процесс не может быть стационарным. Вероятности состояний постоянно меняются во времени. Вероятность начального состояния P_1 экспоненциально убывает. Для любого другого состояния $m > 1$ вероятность сначала возрастает, достигает максимума в некоторый момент времени, а затем убывает.

2. *Случайный двоичный телеграфный сигнал* или *дихотомический шум* – это случайные блуждания с двумя состояниями x_1 и x_2 . Граф двоичного телеграфного сигнала изображён на рис. 3.9.

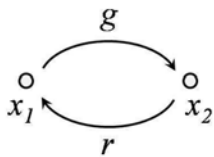


Рис. 3.9. Граф телеграфного сигнала

Соответствующая система управляющих уравнений имеет вид:

$$\begin{aligned} \frac{dP_1(t)}{dt} &= rP_2(t) - gP_1(t), \\ \frac{dP_2(t)}{dt} &= gP_1(t) - rP_2(t). \end{aligned} \quad (3.99)$$

Стационарные (финитные) вероятности можно найти, используя одно из управляющих уравнений, например первое при $\frac{dP_1(t)}{dt} = 0$. Полученное алгебраическое уравнение

$$rP_2^{\text{ст}} - gP_1^{\text{ст}} = 0 \quad \text{или} \quad gP_1^{\text{ст}} - rP_2^{\text{ст}} = 0$$

совместно с условием нормировки $P_1^{\text{ст}} + P_2^{\text{ст}} = 1$ позволяют легко получить стационарные вероятности состояний:

$$P_1^{\text{ст}} = \frac{r}{r+g}, \quad P_2^{\text{ст}} = \frac{g}{r+g}.$$

3.6.3. Диффузионные марковские процессы

Скалярный вещественный непрерывнозначный марковский процесс $X(t)$ называется *диффузионным*, если существуют следующие ограниченные пределы:

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x' - x) p(x', t + \Delta t / x, t) dx' = A(x, t); \quad (3.100)$$

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x' - x)^2 p(x', t + \Delta t/x, t) dx' = 2B(x, t) \neq 0; \quad (3.101)$$

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} (x' - x)^k p(x', t + \Delta t/x, t) dx' = C_k(x, t) \equiv 0, \quad k > 2. \quad (3.102)$$

Величина $A(x, t)$ представляет собой среднее значение локальной скорости изменения состояния и называется *коэффициентом сноса* или *дрейфа*. Величина $B(x, t)$ характеризует локальную скорость роста дисперсии приращения процесса и называется *коэффициентом диффузии*⁹.

Условия (3.100) – (3.102) означают локальный рост дисперсии приращения $\Delta X = X(t') - x(t)$ по линейному закону, т.е. для малых $\Delta t = t' - t$ справедливо

$$\sigma_{\Delta X}^2(t + \Delta t) = 2B(x, t)\Delta t. \quad (3.103)$$

Условие (3.102) исключает большие приращения процесса на малых временах. Хотя допускаются быстрые изменения значений процесса, но в противоположных направлениях. Конечные скачки появляются с нулевой вероятностью, а реализации процесса непрерывны с вероятностью единица. Примером диффузионного процесса может служить скорость броуновской частицы.

Прямое и обратное эволюционные уравнения диффузионного процесса. Плотность вероятности перехода $p(x, t/x_0, t_0)$ диффузионного процесса $X(t)$ удовлетворяет некоторым уравнениям в частных производных, которые можно получить из уравнения Чепмена – Колмогорова и условий (3.100) – (3.102). Рассмотрим три момента времени (t_0, t' и t): $t_0 < t' < t$, $t - t' = \Delta t$. Уравнение Чепмена – Колмогорова перепишем в виде:

$$p(x, t/x_0, t_0) = \int_{-\infty}^{\infty} p(x', t'/x_0, t_0) p(x, t/x', t') dx'.$$

Плотность вероятности перехода $p(x, t/x', t')$ можно представить как

$$p(x, t/x', t') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Theta(u/x', t') \exp[-ju(x - x')] du, \quad (3.104)$$

где $\Theta(u/x', t')$ – условная характеристическая функция случайного приращения $X(t) - x'$ относительно фиксированного значения x' в момент времени t' , т.е.

$$\Theta(u/x', t') = \int_{-\infty}^{\infty} p(x, t/x', t') \exp[ju(x - x')] dx.$$

Условную характеристическую функцию можно разложить в ряд Маклорена по условным моментным функциям:

$$\Theta(u/x', t') = \sum_{k=0}^{\infty} \frac{\langle (X - x')^k \rangle}{k!} (ju)^k.$$

⁹Иногда коэффициентом диффузии называют величину $2B(x, t)$.

С учетом этого разложения из (3.104) следует

$$p(x, t/x', t') = \sum_{k=0}^{\infty} \frac{\langle (X - x')^k \rangle}{k!} \frac{1}{2\pi} \int_{-\infty}^{\infty} (ju)^k \exp[-ju(x - x')] du.$$

Поскольку

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp[-ju(x - x')] du &= \delta(x - x'), \\ \frac{1}{2\pi} \int_{-\infty}^{\infty} (ju)^k \exp[-ju(x - x')] du &= (-1)^k \frac{\partial^k \delta(x - x')}{\partial x^k}, \end{aligned}$$

то получаем

$$p(x, t/x', t') = \delta(x - x') + \sum_{k=1}^{\infty} (-1)^k \frac{\langle (X - x')^k \rangle}{k!} \frac{\partial^k \delta(x - x')}{\partial x^k}. \quad (3.105)$$

Подставляем выражение (3.105) в уравнение Чепмена – Колмогорова:

$$\begin{aligned} p(x, t/x_0, t_0) &= \\ &= \int_{-\infty}^{\infty} p(x', t'/x_0, t_0) \left[\delta(x - x') + \sum_{k=1}^{\infty} (-1)^k \frac{\langle (X - x')^k \rangle}{k!} \frac{\partial^k \delta(x - x')}{\partial x^k} \right] dx'. \end{aligned}$$

Меняя местами интегрирование и суммирование в (3.106), с учетом равенств

$$\int_{-\infty}^{\infty} p(x', t'/x_0, t_0) \delta(x - x') dx' = p(x, t'/x_0, t_0)$$

и

$$\int_{-\infty}^{\infty} p(x', t'/x_0, t_0) \langle (X - x')^k \rangle \delta(x - x') dx' = p(x, t'/x_0, t_0) \langle (X - x)^k \rangle$$

получим

$$p(x, t/x_0, t_0) - p(x, t'/x_0, t_0) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \frac{\partial^k}{\partial x^k} \int_{-\infty}^{\infty} [p(x, t'/x_0, t_0) \langle (X - x)^k \rangle] dx'.$$

Согласно (3.102) при $k > 2$ интегралы обращаются в нуль. Поэтому в сумме остаются только два слагаемых. Далее делим обе части полученного равенства на Δt . Представляя $p(x, t'/x_0, t_0) = p(x, t'/x, t) \cdot (x, t/x_0, t_0)$ и переходя к пределу при $\Delta t \rightarrow 0$, получаем

$$\begin{aligned} \frac{\partial p(x, t/x_0, t_0)}{\partial t} &= \\ &= \sum_{k=1}^2 \frac{(-1)^k}{k!} \frac{\partial^k}{\partial x^k} \left[p(x, t/x_0, t_0) \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} \langle (X - x)^k \rangle p(x', t'/x, t) dx' \right]. \end{aligned}$$

С учетом условий (3.100) – (3.101) окончательно получаем следующее уравнение для $p(x, t/x_0, t_0)$:

$$\frac{\partial p(x, t/x_0, t_0)}{\partial t} = -\frac{\partial A(x, t)p(x, t/x_0, t_0)}{\partial x} + \frac{\partial^2 B(x, t)p(x, t/x_0, t_0)}{\partial x^2}. \quad (3.106)$$

Уравнение (3.106) определяет эволюцию плотности вероятности перехода в прямом времени и называется *прямым уравнением Колмогорова* или *уравнением Фоккера – Планка – Колмогорова* (ФПК). Его следует дополнить начальным условием: $p(x, t_0/x_0, t_0) = \delta(x - x_0)$.

Аналогично можно получить *обратное уравнение Колмогорова*. Для этого плотность вероятности перехода $p(x', t'/x_0, t_0)$ в (3.104) раскладывают по условным моментам и делят обе части полученного равенства на $\Delta t = t' - t_0 \rightarrow 0$. Обратное уравнение имеет вид:

$$\frac{\partial p(x, t/x_0, t_0)}{\partial t_0} = -A(x_0, t_0) \frac{\partial p(x, t/x_0, t_0)}{\partial x_0} - B(x_0, t_0) \frac{\partial^2 p(x, t/x_0, t_0)}{\partial x_0^2}, \quad (3.107)$$

а начальное условие для него есть $p(x, t/x_0, t) = \delta(x - x_0)$. Здесь величины x и t считаются заданными, а x_0 и t_0 – переменными. Обратное уравнение Колмогорова используется в задачах о времени первого достижения заданной границы.

Уравнения (3.106) и (3.107) относятся к классу параболических уравнений в частных производных и для решения требуют задания начальных и граничных условий. Начальные условия мы уже рассмотрели для каждого из уравнений, а граничные условия аналогичны тем, которые подробно рассмотрены выше для марковских цепей.

Уравнения (3.106) и (3.107) не являются независимыми. Можно показать, что операторы $L_1 : p(x, t/x_0, t_0) \rightarrow \frac{\partial v}{\partial t}$ и $L_2 : p(x, t/x_0, t_0) \rightarrow \frac{\partial v}{\partial t_0}$ являются взаимно-сопряженными.

Умножая обе части прямого уравнения (3.106) на $p_1(x_0, t_0)$ и интегрируя по x_0 , можно записать уравнение ФПК для безусловной одномерной плотности вероятности:

$$\frac{\partial p_1(x, t)}{\partial t} = -\frac{\partial A(x, t)p_1(x, t)}{\partial x} + \frac{\partial^2 B(x, t)p_1(x, t)}{\partial x^2} \quad (3.108)$$

с начальным условием $p_1(x, t_0) = p_1(x_0, t_0)$. Таким образом, диффузионный процесс $X(t)$ полностью определяется своими коэффициентами сноса и диффузии при заданных начальных и граничных условиях для уравнения ФПК.

Физический смысл уравнений ФПК нетрудно истолковать следующим образом. Введем функцию

$$J(x, t) = A(x, t)p_1(x, t) - \frac{\partial B(x, t)p_1(x, t)}{\partial x}, \quad (3.109)$$

называемую *потокм вероятности*. Поток складывается из систематической компоненты $A(x, t)p_1(x, t)$ и диффузионной компоненты $-\frac{\partial B(x, t)p_1(x, t)}{\partial x}$, а уравнение ФПК (3.108) превращается в уравнение непрерывности:

$$\frac{\partial p_1(x, t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x}. \quad (3.110)$$

Плотность вероятности $p_1(x, t)$ можно рассматривать как плотность некоторого вещества распределённого вдоль одной координаты, а поток вероятности $J(x, t)$ – как поток вещества через границы дифференциального интервала $[x, x + dx)$. Положительный поток направлен за пределы рассматриваемого интервала. В этом случае с учетом отрицательного знака в правой части равенства производная по времени от плотности $p_1(x, t)$ будет отрицательной. И наоборот, отрицательный поток направлен внутрь интервала и соответствует положительной производной от $p_1(x, t)$. Таким образом, изменения $p_1(x, t)$ (рост или убыль) связаны с наличием потока $J(x, t)$. В таком виде уравнение ФПК выражает так называемый *закон сохранения вероятности*.

Стационарный диффузионный процесс. Такие процессы используются, например, при описании равновесных флуктуаций.

Пусть $X(t)$ – стационарный (в строгом смысле) диффузионный процесс. Тогда для него $A(x, t) \equiv A(x)$, $B(x, t) \equiv B(x)$, $p_1(x, t) \equiv p_1^{\text{ст}}(x)$ и $\frac{\partial p_1(x, t)}{\partial t} = 0$. Налагая граничные условия, соответствующие неограниченному процессу

$$p_1^{\text{ст}}(x)|_{x=\pm\infty} = 0, \quad \left. \frac{dp_1^{\text{ст}}(x)}{dx} \right|_{x=\pm\infty} = 0, \quad (3.111)$$

и интегрируя дважды по x , находим решение уравнения ФПК:

$$p_1^{\text{ст}}(x) = \frac{C}{B(x)} \exp \left(\int_{x_0}^x \frac{A(x')}{B(x')} dx' \right), \quad (3.112)$$

где x_0 – произвольно выбранное значение случайной переменной; C – нормировочная константа.

Решение (3.112) можно представить в виде:

$$p_1^{\text{ст}}(x) = \exp(-\phi(x)), \quad (3.113)$$

где функция $\phi(x)$ называется *квазипотенциалом* или *термодинамическим потенциалом* и определяется как

$$\phi(x) = \ln \left(\frac{B(x)}{C} - \int_{x_0}^x \frac{A(x')}{B(x')} dx' \right). \quad (3.114)$$

Иллюстрацией полученного решения могут служить флуктуации напряжения в цепи, состоящей из резистора с сопротивлением R , находящегося при постоянной температуре T , и конденсатора с ёмкостью C . При условии, что индуктивность цепи пренебрежимо мала, флуктуации напряжения $X(t)$ на конденсаторе хорошо описываются стационарным марковским процессом и

$$p_1^{\text{ст}}(x) = \sqrt{C/2\pi kT} \exp(-Cx^2/2kT).$$

Векторный диффузионный процесс. Понятие диффузионного процесса может быть распространено и на векторный марковский процесс $\mathbf{X}(t)$, для которого текущее состояние задается вектором \mathbf{x} с

компонентами x_j , $j = 1, 2, \dots, N$. Векторный диффузионный процесс характеризуется вектором сноса $\mathbf{A}(\mathbf{x}, t)$ с компонентами

$$A_j(\mathbf{x}, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int (x'_j - x_j) p(\mathbf{x}', t + \Delta t / \mathbf{x}, t) d\mathbf{x}', \quad j = 1, 2, \dots, N \quad (3.115)$$

и матрицей диффузии $\widehat{\mathbf{B}}(\mathbf{x}, t)$ с элементами

$$B_{jk}(\mathbf{x}, t) = \frac{1}{2} \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int (x'_j - x_j)(x'_k - x_k) p(\mathbf{x}', t + \Delta t / \mathbf{x}, t) d\mathbf{x}', \quad j, k = 1, 2, \dots, N. \quad (3.116)$$

Интегралы в этих выражениях берутся по всем возможным значениям векторной переменной \mathbf{x}' . Уравнение ФПК для векторного диффузионного процесса $\mathbf{X}(t)$, записанное относительно одномерной плотности вероятности $p_1(\mathbf{x}, t)$, принимает вид:

$$\frac{\partial p_1(\mathbf{x}, t)}{\partial t} = - \sum_{j=1}^N \frac{\partial A_j(\mathbf{x}, t) p_1(\mathbf{x}, t)}{\partial x_j} + \sum_{j=1}^N \sum_{k=1}^N \frac{\partial^2 B_{jk}(\mathbf{x}, t) p_1(\mathbf{x}, t)}{\partial x_j \partial x_k}. \quad (3.117)$$

Аналогичное уравнение нетрудно записать относительно $p(\mathbf{x}, t / \mathbf{x}_0, t_0)$.

Можно ввести вектор потока вероятности $\mathbf{J}(\mathbf{x}, t)$, компоненты которого определяются как

$$J_j(\mathbf{x}, t) = A_j(\mathbf{x}, t) p_1(\mathbf{x}, t) - \sum_{k=1}^N \frac{\partial B_{jk}(\mathbf{x}, t) p_1(\mathbf{x}, t)}{\partial x_k}, \quad j = 1, 2, \dots, N. \quad (3.118)$$

Тогда уравнение (3.117) можно переписать в форме уравнения непрерывности, выражающего закон сохранения вероятности:

$$\frac{\partial p_1(\mathbf{x}, t)}{\partial t} = - \operatorname{div} \mathbf{J}(\mathbf{x}, t). \quad (3.119)$$

Поток вероятности через границу элемента объёма dV в точке \mathbf{x} в пространстве \mathbb{R}^N определяет изменение плотности вероятности в этой точке во времени.

Стационарное решение уравнения (3.117) с граничными условиями, соответствующими неограниченному процессу

$$p_1^{\text{ст}}(\mathbf{x}) \Big|_{\|\mathbf{x}\|=\infty} = 0, \quad \frac{\partial p_1^{\text{ст}}(\mathbf{x})}{\partial x_i} \Big|_{\|\mathbf{x}\|=\infty} = 0, \quad i = 1, 2, \dots, N, \quad (3.120)$$

имеет вид:

$$p_1^{\text{ст}}(\mathbf{x}) = \exp(-\phi(\mathbf{x})), \quad (3.121)$$

где квазипотенциал $\phi(\mathbf{x})$ выражается следующим образом:

$$\phi(\mathbf{x}) = \phi(\mathbf{x}_0) - \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{Z}(\mathbf{x}') d\mathbf{x}'. \quad (3.122)$$

Компоненты вектора $\mathbf{Z}(\mathbf{x})$ определяются по формуле

$$Z_j(\mathbf{x}) = \sum_{i=1}^N B_{ij}^{-1}(\mathbf{x}) \left[A_i(\mathbf{x}) - \sum_{k=1}^N \frac{\partial B_{ik}(\mathbf{x})}{\partial x_k} \right]. \quad j = 1, 2, \dots, N. \quad (3.123)$$

Здесь $B_{ij}^{-1}(\mathbf{x})$ – элементы обратной матрицы $\widehat{\mathbf{B}}^{-1}(\mathbf{x})$. Условием существования стационарного решения уравнения ФПК является существование квазипотенциала $\phi(\mathbf{x})$, что равносильно требованию потенциальности поля $\mathbf{Z}(\mathbf{x})$: $\text{rot} \mathbf{Z}(\mathbf{x}) = 0$.

3.6.4. Диффузионный процесс Винера-Леви

Пример диффузионного процесса – процесс Винера – Леви. Рассмотрим броуновскую частицу массой m , движущуюся в плоскости, трансверсальной силе тяжести в разряженной среде. Разреженность среды означает, что можно пренебречь силой сопротивления среды. Тогда уравнение движения можно записать в виде:

$$m \frac{d^2 X}{dt^2} = f(t), \quad (3.124)$$

где $X(t_0) = x_0$ – заданное начальное состояние. Сила $f(t)$ обусловлена случайными независимыми столкновениями частицы с молекулами среды. Будем для простоты считать, что начальная скорость частицы равна нулю: $p(t_0) = \left. \frac{dX}{dt} \right|_{t=t_0} = 0$. Интервал времени dt , на котором рассматриваются приращения скорости и координаты, будем считать физически малым (но не бесконечно малым), т.е. малым по сравнению с масштабами времени, характерными для эксперимента и измерительной аппаратуры. В то же время $dt \gg \tau_0$, где τ_0 – среднее время между последовательными соударениями частицы с молекулами среды. Разделив обе части уравнения (3.124) на массу частицы m и проинтегрировав один раз по t , получаем

$$\frac{dX}{dt} = \frac{1}{m} \int_{t_0}^t f(\theta) d\theta. \quad (3.125)$$

Обозначим

$$\frac{1}{m} \int_{t_0}^t f(\theta) d\theta = \xi(t),$$

где $\xi(t)$ – случайная скорость частицы, которую можно приближённо представить как

$$\xi(t) \approx \frac{1}{m} \sum_{i=1}^{i_{max}} f(\theta_i) \Delta\theta, \quad \Delta\theta = \theta_{i+1} - \theta_i.$$

Так как $\Delta\theta \gg \tau_0$, то значения слагаемых суммы можно считать статистически независимыми и, следовательно, в силу центральной предельной теоремы теории вероятности величина $\xi(t)$ имеет распределение, близкое к гауссовскому. В силу изотропности среды $\langle \xi(t) \rangle \equiv 0$. Кроме того,

из-за случайного и независимого характера и большой частоты воздействий на частицу можно считать, что время корреляции процесса $\xi(t)$ очень мало (меньше, чем физически малый интервал dt). Таким образом, мы приходим к модели гауссовского белого шума. После нормирования источника шума перепишем уравнение (3.125) в виде

$$\frac{dX}{dt} = \sqrt{2D}n(t), \quad (3.126)$$

где $n(t)$ – нормированный нормальный белый шум ($\langle n(t) \rangle \equiv 0$, $\langle n(t)n(t+\tau) \rangle = \delta(\tau)$), D – константа, характеризующая интенсивность случайных толчков (*интенсивность шума*). Величина D определяется как

$$D = \int_0^{\infty} \Psi_{\xi}(\tau) d\tau, \quad (3.127)$$

где $\Psi_{\xi}(\tau)$ – автокорреляционная функция процесса $\xi(t)$. Уравнение (3.126) описывает так называемый *винеровский процесс* или процесс *Винера – Леви*. Проинтегрировав (3.126), получаем

$$X(t) = x_0 + \sqrt{2D} \int_{t_0}^t n(\theta) d\theta, \quad t > t_0. \quad (3.128)$$

Рассмотрим основные свойства винеровского процесса.

1. Он является гауссовским процессом (в силу линейности уравнений (3.126), (3.128) и гауссовского характера воздействия $n(t)$).

2. Его среднее значение определяется начальным состоянием ($\langle X(t) \rangle = x_0$) (что следует из (3.128)).

3. Винеровский процесс является нестационарным и его дисперсия растет линейно во времени. Действительно,

$$\begin{aligned} \sigma_X^2(t) &= \langle (X(t) - x_0)^2 \rangle = \left\langle 2D \int_{t_0}^t \int_{t_0}^t n(\theta_1)n(\theta_2) d\theta_1 d\theta_2 \right\rangle = \\ &= 2D \int_{t_0}^t \int_{t_0}^t \langle n(\theta_1)n(\theta_2) \rangle d\theta_1 d\theta_2. \end{aligned}$$

С учетом того, что $\langle n(\theta_1)n(\theta_2) \rangle = \delta(\theta_2 - \theta_1)$, и свойств функции Дирака получаем

$$\sigma_X^2(t) = 2D \cdot (t - t_0). \quad (3.129)$$

Скорость линейного роста дисперсии постоянна и определяется интенсивностью случайного воздействия D . Далее будет показано, что коэффициент диффузии B винеровского процесса равен величине D .

4. Винеровский процесс – это *процесс с независимыми приращениями* на неперекрывающихся интервалах времени. Покажем справедливость этого утверждения. Обозначим

$$\Delta X_i = X(t_{i+1}) - X(t_i) = \sqrt{2D} \int_{t_i}^{t_{i+1}} n(\theta) d\theta$$

– приращение винеровского процесса за время $\Delta t_i = t_{i+1} - t_i$. Очевидно распределение величины ΔX_i является гауссовским, среднее значение $\langle \Delta X_i \rangle$ равно нулю, а дисперсия $\sigma_{\Delta X_i}^2 = 2D\Delta t_i$. Совместное распределение приращений на различных неперекрывающихся интервалах также является гауссовским. Найдем взаимную корреляцию приращений на неперекрывающихся интервалах $\Delta t_i = t_{i+1} - t_i$ и $\Delta t_j = t_{j+1} - t_j$, $j \neq i$:

$$\langle \Delta X_i \Delta X_j \rangle = 2D \int_{t_j}^{t_{j+1}} \int_{t_i}^{t_{i+1}} \delta(\theta_2 - \theta_1) d\theta_1 d\theta_2 = 0$$

(так как интервалы интегрирования не перекрываются и значение θ_2 нигде не равно θ_1). Итак, для совместного гауссовского распределения приращений из некоррелированности ΔX_i и ΔX_j следует их статистическая независимость.

5. Винеровский процесс является марковским. Свойство марковости следует из независимости приращений.

6. Винеровский процесс является диффузионным процессом с постоянным коэффициентом диффузии $B(x, t) \equiv \text{const}$ и коэффициентом сноса $A(x, t) \equiv 0$. Таким образом, это *чисто диффузионный процесс*. Покажем это, найдя пределы

$$C_k(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle (\Delta X)^k \rangle,$$

где $\Delta X = X' - x$ – случайное приращение за время $\Delta t = t' - t$ к заданному значению x в момент t . Найдем $\langle (\Delta X)^k \rangle$:

$$\langle (\Delta X)^k \rangle = \int_{-\infty}^{\infty} (\Delta x)^k p(\Delta x) d(\Delta x), \quad (3.130)$$

где $p(\Delta x)$ – гауссовское распределение с нулевым средним и дисперсией $\sigma_{\Delta X}^2 = 2D\Delta t$:

$$p(\Delta x) = \frac{1}{\sqrt{2\pi\sigma_{\Delta X}^2}} \exp \left[-\frac{(\Delta x)^2}{2\sigma_{\Delta X}^2} \right].$$

В силу нечётности подынтегральной функции в (3.130) для всех нечётных значений k получаем, что $\langle (\Delta X)^k \rangle = 0$, $k = 1, 3, 5, \dots$. Чтобы найти средние значения для чётных степеней приращений, используем замену переменных $z = \frac{\Delta x}{\sqrt{2\sigma_{\Delta X}^2}}$ и применим в (3.130) интегрирование по частям нужное число раз. В итоге получим:

$$\langle (\Delta X)^k \rangle = \frac{\sqrt{(4D\Delta t)^k}}{\sqrt{\pi}} \cdot \frac{[(k-1) \cdot (k-3) \cdot (k-5) \cdot \dots \cdot 1]}{2^{k-1}} \int_{-\infty}^{\infty} \exp(-z^2) dz, \quad k = 2, 4, \dots$$

Мы приходим к интегралу Пуассона: $\int_{-\infty}^{\infty} \exp(-z^2) dz = \sqrt{\pi}$. Для пределов $C_k(x, t)$ получаем выражения:

$$C_1(x, t) \equiv 0; \quad C_2(x, t) = 2D; \quad C_k(x, t) \equiv 0, \quad k > 2.$$

Таким образом, мы показали, что винеровский процесс является диффузионным, причём его коэффициент сноса есть $A(x, t) = C_1(x, t) \equiv 0$, а коэффициент диффузии – $B(x, t) = \frac{1}{2}C_2(x, t) \equiv D = \text{const}$.

3.6.5. Стохастические дифференциальные уравнения (СДУ)

Общие представления. Пусть поведение некоторой системы описывается дифференциальным уравнением первого порядка (его часто называют *уравнением Ланжевена*), имеющим вид

$$\frac{dX}{dt} = f(X, t) + g(X, t)n(t), \quad (3.131)$$

где f и g – детерминированные гладкие функции своих аргументов; $n(t)$ – нормированный гауссовский белый шум ($\langle n(t) \rangle \equiv 0$, $\langle n(t)n(t+\tau) \rangle = \delta(\tau)$). Множитель g характеризует интенсивность шума. Если g не зависит от состояния X , а является детерминированной функцией времени или константой, то шум, воздействующий на систему, называется *аддитивным*, если же интенсивность шумового воздействия зависит от X , то – *мультипликативным*.

Простейшее СДУ вида

$$\frac{dX}{dt} = gn(t), \quad X(t_0) = X_0, \quad g = \text{const} \quad (3.132)$$

описывает винеровский процесс $v(t)$ с коэффициентом диффузии $B_v = g^2/2$. Если винеровский процесс $v(t)$ обладает нулевым начальным значением и коэффициентом диффузии $1/2$ ($g = 1$), то $dv = n(t)dt$, т.е. производная по времени от винеровского процесса равна плотности нормированного белого гауссовского шума. Это утверждение часто используется в качестве определения белого гауссовского шума.

С учетом указанных параметров винеровского процесса СДУ (3.131) можно переписать в дифференциалах как

$$dX = f(X, t)dt + g(X, t)dv \quad (3.133)$$

или в интегральном представлении как

$$X(t) = X(t_0) + \int_{t_0}^t f(X(\theta), \theta)d\theta + \int_{t_0}^t g(X(\theta), \theta)dv(\theta). \quad (3.134)$$

В случае мультипликативного шума второй интеграл в (3.134) является так называемым *стохастическим интегралом* по винеровскому процессу. Соответствующий среднеквадратический предел от интегральных сумм

$$\begin{aligned} & \int_{t_0}^t g(X(\theta), \theta)dv(\theta) = \\ & = \text{l.i.m.}_{m \rightarrow \infty} \sum_{i=0}^{m-1} g(X(\theta'_i), \theta_i)[v(\theta_{i+1}) - v(\theta_i)] \end{aligned} \quad (3.135)$$

определяется выбором точек $\theta'_i \in [\theta_i, \theta_{i+1})$.

Стохастический интеграл (3.135) может быть определен по-разному. Пусть $\theta'_i = (1 - \nu)\theta_i + \nu\theta_{i+1}$, где $\nu \in [0, 1)$. При произвольном значении ν стохастический интеграл (3.135) называется *обобщённым стохастическим интегралом*, а уравнение (3.131) (или (3.133)) *обобщённым СДУ*. Если положить $\nu = 0$, то получаем *стохастическое дифференциальное уравнение Ито* и *стохастический интеграл Ито*. В случае $\nu = 1/2$ приходим к *СДУ Стратоновича* и *стохастическому интегралу Стратоновича*.

Свойства интеграла Ито и интеграла Стратоновича. Стохастическим интегралом в смысле Ито называется предел сходящихся в среднем квадратическом интегральных сумм (3.135) при $\theta'_i = \theta_i$. Для интеграла Ито не применимы правила обычного интегрирования. Например, можно показать, что

$$\int_{t_0}^t v(\theta)dv = \frac{1}{2} (v^2(t) - v^2(t_0) - (t - t_0)),$$

а не $\frac{1}{2} (v^2(t) - v^2(t_0))$, как в «обычном» случае.

Для вычисления интеграла Ито в СДУ (3.133) вводится замена переменных по *правилу Ито*, согласно которому для любой непрерывной функции $\Phi(x, t)$, обладающей непрерывными производными $\frac{\partial \Phi}{\partial t}$, $\frac{\partial \Phi}{\partial x}$, $\frac{\partial^2 \Phi}{\partial x^2}$, дифференциальное приращение процесса $\Phi(x(t), t)$ в фиксированной точке x в момент t будет

$$\begin{aligned} d\Phi(x, t) = & \frac{\partial \Phi(x, t)}{\partial t} dt + \frac{\partial \Phi(x, t)}{\partial x} f(x, t) dt + \frac{\partial \Phi(x, t)}{\partial x} g(x, t) dv + \\ & + \frac{1}{2} \frac{\partial^2 \Phi(x, t)}{\partial x^2} g^2(x, t) dt. \end{aligned} \quad (3.136)$$

Последнее слагаемое отсутствует, если $X(t)$ – гладкая функция.

Стохастический интеграл Ито, определённый для прямого времени, не совпадает с интегралом Ито в обратном времени, т.е.

$$\int_{t_0}^t g[X(\theta), \theta] dv(\theta) \neq - \int_t^{t_0} g[X(\theta), \theta] dv(\theta). \quad (3.137)$$

Действительно, при инверсии времени точки θ'_i будут уже не начальными точками интервалов разбиения, а конечными. Соответственно предел интегральной суммы в выражении (3.135) изменится.

Для стохастического интеграла Стратоновича, называемого также *симметризованным интегралом* (поскольку $\theta'_i = (\theta_i + \theta_{i+1})/2$ берутся симметрично – в середине интервалов разбиения), выполняются все правила обычного интегрирования. Так, например, имеем

$$\int_{t_0}^t v(\theta)dv = \frac{1}{2} (v^2(t) - v^2(t_0)).$$

Правило замены переменных также будет «обычным», а значение интеграла сохраняется при инверсии времени.

Зависимость коэффициентов сноса и диффузии случайного процесса $X(t)$ от вида СДУ. Случайный процесс $X(t)$, являющийся решением обобщённого СДУ (3.133) и удовлетворяющий детерминированному начальному условию $X(t_0) = x_0$, представляет собой диффузионный марковский процесс. Не приводя доказательства марковости, отметим только, что для него используются непрерывность функций $f(x, t)$, $g(x, t)$ и свойства приращений винеровского процесса. Для доказательства того, что процесс $X(t)$ является диффузионным, найдем пределы

$$C_k(x, t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle (\Delta X)^k \rangle,$$

где $\Delta t = t' - t$ – приращение времени; $X(t) = x$ – заданное значение переменной в момент t ; $X(t') = X'$ – случайное значение переменной в момент t' ; $\Delta X = X' - x$ – приращение процесса на интервале $\Delta t = t' - t$. Для малых интервалов времени приближённо можно считать, что

$$\Delta X \approx f(x, t)\Delta t + g(X(\theta), t)\Delta v,$$

где $\Delta v = v(t') - v(t)$ – приращение винеровского процесса; $\theta = t + \nu \cdot \Delta t$. В свою очередь, можно представить

$$X(\theta) = X(t + \nu\Delta t) \approx x + \frac{dX(t)}{dt} \nu\Delta t = x + \nu\delta X,$$

где

$$\delta X \approx \frac{dX(t)}{dt} \Delta t \approx g(x, t)\Delta v$$

(здесь мы пренебрегаем плавным изменением процесса под действием силы f и считаем, что изменение связано только со случайными скачками). Таким образом, для малого приращения ΔX получаем следующее приближенное выражение:

$$\begin{aligned} \Delta X &\approx f(x, t)\Delta t + g(x + \nu g(x, t)\Delta v, t)\Delta v \approx \\ &\approx f(x, t)\Delta t + g(x, t)\Delta v + \nu g_x(x, t)g(x, t)(\Delta v)^2, \end{aligned}$$

где $g_x(x, t) = \frac{\partial g(x, t)}{\partial x}$.

Найдем средние значения различных степеней приращения ΔX :

$$\langle (\Delta X)^k \rangle \approx \langle [f(x, t)\Delta t + g(x, t)(\Delta v + \nu g_x(x, t)(\Delta v)^2)]^k \rangle.$$

Применяя формулу бинома Ньютона, получаем

$$\langle (\Delta X)^k \rangle \approx [f(x, t)\Delta t]^k + \sum_{j=1}^k \frac{k!}{j!(k-j)!} [f(x, t)\Delta t]^{k-j} g^j(x, t) Z_j,$$

где введено обозначение

$$Z_j = \langle [\Delta v + \nu(\Delta v)^2 g_x(x, t)]^j \rangle.$$

Ещё раз применяя формулу бинома Ньютона, находим

$$Z_j = \sum_{i=0}^j \frac{j!}{i!(j-i)!} \nu^i g_x^i(x, t) \langle (\Delta v)^{i+j} \rangle.$$

В случае $j + i = 1$ имеем $\langle \Delta v \rangle = 0$. Используя диффузионный характер стандартного винеровского процесса $v(t)$, легко показать, что при малом приращении $\Delta t \rightarrow 0$ для $j + i = 2$ получаем значение $\langle (\Delta v)^2 \rangle = \Delta t$. Кроме того, для любого $r > 2$

$$\lim_{\Delta t \rightarrow 0} \frac{\langle (\Delta v)^r \rangle}{\Delta t} = 0.$$

Поэтому, если $j + i = r > 2$, то величина $\langle (\Delta v)^{i+j} \rangle$ при $\Delta t \rightarrow 0$ убывает быстрее, чем Δt . Следовательно, в сумме, входящей в правую часть выражения для Z_j , можно пренебречь всеми слагаемыми, кроме слагаемых, содержащих средний квадрат приращения винеровского процесса (для которых $i + j = 2$). С учетом этого получаем:

$$Z_1 \approx \nu g_x(x, t) \Delta t, \quad Z_2 \approx \Delta t, \quad Z_{j>2} \approx 0.$$

Наконец, находим пределы $C_k(x, t)$:

$$\begin{aligned} C_1(x, t) &= f(x, t) + \nu g_x(x, t) g(x, t); \\ C_2(x, t) &= g^2(x, t); \\ C_k(x, t) &\equiv 0, \quad k > 2. \end{aligned} \tag{3.138}$$

Таким образом, процесс $X(t)$, задаваемый обобщённым СДУ (3.133), является диффузионным процессом, причём его коэффициент сноса определяется как

$$A(x, t) = f(x, t) + \nu g_x(x, t) g(x, t), \tag{3.139}$$

а коэффициент диффузии – как

$$B(x, t) = \frac{1}{2} g^2(x, t). \tag{3.140}$$

Второе слагаемое в выражении (3.139) появляется только в случае мультипликативного шума. Оно называется *ложным сносом*. Если процесс $X(t)$ описывается СДУ Ито, то ложный снос отсутствует.

Зная вид СДУ, можно найти коэффициенты сноса и диффузии и записать уравнение ФПК. Решая его, находим вероятностное распределение процесса $X(t)$. Очевидно, что при мультипликативном шуме одно и то же по форме СДУ, понимаемое как СДУ Ито или СДУ Стратоновича, описывает различные процессы, так как разница в определении коэффициента сноса приводит к различным статистическим характеристикам процессов. Нельзя сказать, какое из определений стохастического интеграла является верным. С точки зрения математики обе концепции (Ито и Стратоновича) являются допустимыми. Однако СДУ Ито и СДУ Стратоновича являются различными моделями случайного процесса, и выбор модели определяется из физических соображений и данных эксперимента. Так, если процесс СДУ получается из разностной схемы с дискретным временем и дискретным белым шумом, путём перехода к непрерывному времени, то более подходящим оказывается СДУ Ито. Если рассматривается динамическая система, находящаяся под воздействием шума, и

реальный широкополосный шум заменяется белым, то процесс в такой системе правильнее описывать используя СДУ Стратоновича.

При численном интегрировании проще иметь дело с СДУ Ито. Поэтому, если некоторый процесс $X(t)$ описывается СДУ Стратоновича, то рассматривают эквивалентное СДУ Ито, задающее тот же самый процесс (с точки зрения его статистических характеристик). Оно имеет вид:

$$dX = [f(X, t) + \tilde{f}(X, t)]dt + g(X, t)dv, \quad (3.141)$$

где $\tilde{f}(x, t)$ – ложный снос.

Обобщение СДУ на многомерный случай. Пусть имеется система N обобщённых стохастических дифференциальных уравнений первого порядка

$$\frac{dX_j}{dt} = f_j(X_1, \dots, X_N, t) + \sum_{k=1}^N g_{jk}(X_1, \dots, X_N, t)n_k(t), \quad j=1, 2, \dots, N, \quad (3.142)$$

где $n_k(t)$, $k = 1, 2, \dots, N$ – статистически независимые нормированные источники нормального белого шума. При заданных начальных условиях $X_j(t_0) = x_j^0$, $j = 1, 2, \dots, N$ система (3.142) описывает векторный диффузионный процесс $\mathbf{X}(t) = X_1(t), X_2(t), \dots, X_N(t)$, который характеризуется вектором сноса с компонентами

$$A_j(\mathbf{x}, t) = f_j(\mathbf{x}, t) + \nu \sum_{i=1}^N \sum_{k=1}^N g_{ik}(\mathbf{x}, t) \frac{\partial g_{jk}(\mathbf{x}, t)}{\partial x_i}, \quad j = 1, 2, \dots, N \quad (3.143)$$

и матрицей диффузии с элементами

$$B_{jk}(\mathbf{x}, t) = \frac{1}{2} \sum_{i=1}^N g_{ji}(\mathbf{x}, t)g_{ki}(\mathbf{x}, t), \quad j, k = 1, 2, \dots, N. \quad (3.144)$$

Зная вектор сноса и матрицу диффузии диффузионного процесса $\mathbf{X}(t)$, можно записать уравнение Фоккера – Планка – Колмогорова (3.117) для одномерной плотности вероятности $p_1(\mathbf{x}, t)$ и аналогичное уравнение для плотности вероятности перехода $p(\mathbf{x}, t/\mathbf{x}_0, t_0)$. Решив эти уравнения и найдя функции $p_1(\mathbf{x}, t)$ и $p(\mathbf{x}, t/\mathbf{x}_0, t_0)$, мы могли бы определить все статистические характеристики процесса.

Таким образом, система, находящаяся под воздействием независимых источников гауссовского белого шума, может быть описана как стохастическими дифференциальными уравнениями, так и уравнением ФПК. Оба способа описания полностью эквивалентны. Если система исследуется теоретическими методами, то это предполагает переход к уравнению ФПК и его строгое или приближенное решение. Решить уравнение ФПК аналитически удастся только в наиболее простых случаях или с помощью каких-то существенно упрощающих задачу предположений. При проведении численных экспериментов более простым и удобным оказывается метод непосредственного интегрирования СДУ и определения тех или иных статистических характеристик процесса по результатам интегрирования. Численное решение уравнения ФПК уже для двумерных

нелинейных систем связано с рядом трудностей и требует значительно больших затрат вычислительного времени.

Примеры диффузионных процессов, задаваемых СДУ. Кроме процесса Винера – Леви, задаваемого простейшим СДУ (3.132), рассмотрим ещё два примера гауссовских диффузионных процессов, часто применяемых при численном моделировании систем с цветным шумом.

1. *Одномерный процесс Орнштейна – Уленбека.* Этот процесс описывает движение броуновской частицы с учетом силы трения (это так называемая *релеевская частица*). Он задается следующим линейным СДУ первого порядка:

$$\frac{dX}{dt} = -\gamma X + \sqrt{2D}n(t), \quad D = \text{const}, \quad X(t_0) = x_0. \quad (3.145)$$

Гауссовский белый шум $n(t)$ входит в уравнение аддитивно. Так как уравнение (3.145) задает линейное преобразование источника $n(t)$, то процесс $X(t)$ является гауссовским. В силу сформулированных ранее свойств решения СДУ процесс $X(t)$ является диффузионным, причём $A(x, t) = -\gamma x$; $B(x, t) \equiv D = \text{const}$. Стационарное решение уравнения ФПК не зависит от начального распределения и имеет вид

$$p_1^{\text{ст}}(x) = \sqrt{\frac{\gamma}{2\pi D}} \exp\left(-\frac{\gamma x^2}{2D}\right).$$

Стационарное среднее значение процесса равно нулю, а дисперсия $\sigma_X^2 = D/\gamma$. Чтобы гауссовский процесс $X(t)$ был полностью определен, нужно найти корреляционную функцию $\Psi_X(\tau)$, $\tau = t_2 - t_1$. Решение линейного неавтономного уравнения (3.145) может быть представлено в виде

$$X(dt) = x_0 e^{-\gamma(t-t_0)} + \sqrt{2D} \int_{t_0}^t e^{-\gamma(t-\theta)} n(\theta) d\theta.$$

В установившемся режиме (при $t_0 \rightarrow -\infty$) имеем

$$\begin{aligned} \Psi_X(\tau) &= K_X(\tau) = \langle X(t_1)X(t_2) \rangle = \\ &= 2De^{-\gamma(t_1+t_2)} \int_{-\infty}^{t_2} \int_{-\infty}^{t_1} e^{\gamma(\theta_1+\theta_2)} \delta(\theta_2 - \theta_1) d\theta_1 d\theta_2 \end{aligned}$$

В предположении $t_2 > t_1$ легко получить

$$\int_{-\infty}^{t_2} \int_{-\infty}^{t_1} e^{\gamma(\theta_1+\theta_2)} \delta(\theta_2 - \theta_1) d\theta_1 d\theta_2 = e^{2\gamma t_1}.$$

Если $t_1 > t_2$, аналогично получаем

$$\int_{-\infty}^{t_2} \int_{-\infty}^{t_1} e^{\gamma(\theta_1+\theta_2)} \delta(\theta_2 - \theta_1) d\theta_1 d\theta_2 = e^{2\gamma t_2}.$$

Тогда для любых t_1 и t_2 имеем

$$\Psi_X(\tau) = \frac{D}{\gamma} e^{-\gamma|\tau|}.$$

Таким образом, стационарный одномерный процесс Орнштейна – Уленбека – это нормальный экспоненциально-коррелированный диффузионный процесс с нулевым средним значением, дисперсией $\sigma_X^2 = D/\gamma$ и декрементом затухания корреляционной функции, равным γ . Соответственно спектральная плотность мощности стационарного одномерного процесса Орнштейна – Уленбека имеет форму лоренциана с максимумом на нулевой частоте и шириной 2γ на уровне половинной мощности:

$$W_X(\omega) = \frac{2D}{\gamma^2 + \omega^2}.$$

2. *Двумерный процесс Орнштейна – Уленбека.* Этот процесс описывает стохастические колебания в линейном диссипативном контуре, находящемся под воздействием аддитивного гауссовского белого шума. Процесс описывается СДУ второго порядка

$$\frac{d^2 X}{dt^2} + \gamma \frac{dX}{dt} + \omega_0^2 X = \sqrt{2D}n(t), \quad D = \text{const}, \quad X(t_0) = x_0 \quad (3.146)$$

или системой СДУ

$$\begin{aligned} \frac{dX}{dt} &= Y, \\ \frac{dY}{dt} &= -\gamma Y - \omega_0^2 X + \sqrt{2D}n(t). \end{aligned} \quad (3.147)$$

Здесь имеется только один случайный источник, входящий во второе уравнение системы. Двухкомпонентный случайный процесс $(X(t), Y(t))$ является диффузионным. Компоненты вектора сноса есть $A_1(x, y, t) = y$, $A_2(x, y, t) = -\gamma y - \omega_0^2 x$, а матрица диффузии имеет только один отличный от нуля элемент $B_{22}(x, y, t) = D$. Мы можем записать уравнение ФПК, однако его решение даже в этом простом примере оказывается достаточно сложной задачей. В то же время в силу линейности преобразования гауссовского источника $n(t)$ можно сразу сказать, что процесс $X(t)$ является гауссовским. Наличие затухания в контуре обеспечивает равенство нулю стационарного среднего значения: $\langle X(t) \rangle \rightarrow 0$ при $t_0 \rightarrow -\infty$. Находим коэффициент передачи линейного преобразования $n(t) \rightarrow X(t)$:

$$S(\omega) = \frac{\sqrt{2D}}{j\omega\gamma + \omega_0^2 - \omega^2}.$$

В силу линейности преобразования спектральные плотности мощности процессов на входе и выходе системы связаны через квадрат модуля коэффициента передачи, т.е. $W_X(\omega) = |S(\omega)|^2 W_n(\omega)$. Зная коэффициент передачи и учитывая, что $W_n(\omega) = 1$, находим

$$W_X(\omega) = \frac{2D}{\omega^2\gamma^2 + (\omega^2 - \omega_0^2)^2}.$$

При высокой добротности контура можно приближенно получить спектр процесса $X(t)$ в виде суммы лоренцианов с максимумами на частотах $\pm\omega_0$:

$$W_X(\omega) = \frac{D/2\omega_0^2}{(\gamma/2)^2 + (\omega - \omega_0)^2} + \frac{D/2\omega_0^2}{(\gamma/2)^2 + (\omega + \omega_0)^2}.$$

Ширина спектра на уровне половинной мощности равна параметру γ . Полученному «спектру» соответствует осциллирующая корреляционная функция с экспоненциально спадающей огибающей:

$$\Psi_X(\tau) = \sigma_X^2 e^{-\frac{\gamma|\tau|}{2}} \cos \omega_0 \tau, \quad \tau = t_2 - t_1.$$

Дисперсия σ_X^2 (средний квадрат) определяется как $\sigma_X^2 = \frac{D}{\gamma\omega_0^2}$.

3.7. Случайные поля и случайные волны

Целью настоящего раздела является рассмотрение статистических характеристик случайных полей и волн. Полевые и волновые радиофизические задачи во многих отношениях аналогичны задачам, решаемым в теории случайных процессов. Это позволяет использовать и обобщить многие характеристики, введенные для случайных процессов.

3.7.1. Случайные поля и их характеристики

Во многих задачах радиофизики случайная переменная \mathbf{X} зависит не только от времени t , но и от других переменных, обычно от пространственных координат. В этом случае говорят о *случайном поле* $\mathbf{X}(\mathbf{r}, t)$, где вектор $\mathbf{r} \in \mathbb{R}^M$ задает точку в арифметическом пространстве размерности M . Таким образом, случайное поле описывается случайной функцией (в общем случае вектор-функцией) от векторной переменной \mathbf{r} и скалярной переменной t .

Пространства векторов \mathbf{X} и \mathbf{r} в общем случае различны. Даже в тех случаях, когда оба пространства являются арифметическими (\mathbb{R}^3), они могут иметь неодинаковую размерность и разные системы координат, например, для одного – декартова, для другого – полярная. Так, пространство трёхмерных векторов электрического поля может быть преобразовано в шестимерное пространство амплитуд и фаз. В задачах статистической радиофизики пространство, в точках которого задается случайная переменная, представляет собой, как правило, обычное трёхмерное евклидово пространство (\mathbb{R}^3).

В каждой фиксированной точке пространства \mathbf{r}_0 случайное поле $\mathbf{X}(\mathbf{r}, t)$ превращается в случайный процесс $\mathbf{X}(\mathbf{r}_0, t)$, а в фиксированной точке \mathbf{r}_0 для каждого момента времени t_0 получаем случайную величину $\mathbf{X}(\mathbf{r}_0, t_0)$. В зависимости от характера переменной \mathbf{X} поле может быть вещественным или комплексным, скалярным или векторным.

Примерами случайных полей могут служить электромагнитное поле от нелазерного источника света, поле излучения радиопередатчика (с учетом внутренних шумов и шумов в антенне), поле температуры,

диэлектрической проницаемости в реальных турбулентных средах, тепловые поля в объёмных резонаторах и канализирующих системах.

В дальнейшем будем рассматривать электромагнитные волновые поля.

Плотности вероятностей и моментные функции полей. Рассмотрим для простоты скалярное вещественное поле $X(\mathbf{r}, t)$. Для его описания можно ввести следующие плотности вероятности:

$p_n(x_1, \mathbf{r}_1, t_1, x_2, \mathbf{r}_2, t_2, \dots, x_n, \mathbf{r}_n, t_n)$ – плотность вероятности, характеризующую статистическую взаимосвязь значений поля в n точках пространства в n моментов времени;

$p_n(\mathbf{r}, x_1, t_1, x_2, t_2, \dots, x_n, t_n)$ – плотность вероятности в одной точке пространства для n моментов времени;

$p_1(x_1, \mathbf{r}_1, x_2, \mathbf{r}_2, \dots, x_n, \mathbf{r}_n, t)$ – плотность вероятности, характеризующую значения поля в n точках пространства в один момент времени. Как и для случайных процессов, для случайных полей плотности вероятностей представляют собой неотрицательные величины и удовлетворяют условиям нормировки и самосогласованности (см. разд. 3.1).

Рассматривают также моментные функции случайного поля. Основные из них следующие:

1) *среднее значение случайного поля*

$$\bar{X}(\mathbf{r}, t) = \int_{-\infty}^{\infty} xp_1(x, \mathbf{r}, t)dx; \quad (3.148)$$

2) *средний квадрат случайного поля*

$$\bar{X}^2(\mathbf{r}, t) = \int_{-\infty}^{\infty} x^2 p_1(x, \mathbf{r}, t)dx; \quad (3.149)$$

3) *дисперсия поля*

$$\sigma_X^2(\mathbf{r}, t) = \langle (X(t) - \bar{X}(\mathbf{r}, t))^2 \rangle = \bar{X}^2(\mathbf{r}, t) - \bar{X}^2(\mathbf{r}, t); \quad (3.150)$$

4) *ковариационная функция для значений поля в двух точках пространства в два момента времени*

$$\begin{aligned} K_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) &= \langle X(\mathbf{r}_1, t_1)X(\mathbf{r}_2, t_2) \rangle = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_2(x_1, \mathbf{r}_1, t_1, x_2, \mathbf{r}_2, t_2) dx_1 dx_2; \end{aligned} \quad (3.151)$$

5) *корреляционная функция для значений поля в двух точках пространства в два момента времени*

$$\begin{aligned} \Psi_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) &= \langle (X(\mathbf{r}_1, t_1) - \bar{X}(\mathbf{r}_1, t_1)) (X(\mathbf{r}_2, t_2) - \bar{X}(\mathbf{r}_2, t_2)) \rangle = \\ &= K_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) - \bar{X}(\mathbf{r}_1, t_1)\bar{X}(\mathbf{r}_2, t_2). \end{aligned} \quad (3.152)$$

Нормированную корреляционную функцию для значений поля в двух точках пространства в два момента времени

$$R_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \frac{\Psi_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2)}{\sqrt{\sigma_X^2(\mathbf{r}_1, t_1)\sigma_X^2(\mathbf{r}_2, t_2)}} \quad (3.153)$$

называют *коэффициентом корреляции* или *коэффициентом когерентности* значений поля.

Стационарность, однородность и эргодичность полей. Как и для случайного процесса, для случайного поля можно ввести понятие *стационарности во времени*. Случайное поле называется *стационарным*, если для любой точки пространства \mathbf{r}_0 случайный процесс $\mathbf{X}(\mathbf{r}_0, t)$ является стационарным (в строгом (узком) или в широком смысле).

Кроме того, для случайного поля рассматривают свойство, аналогичное стационарности во времени, но только касающееся точек в пространстве – свойство однородности («стационарность в пространстве»). Случайное поле $X(\mathbf{r}, t)$ называется *статистически однородным в узком (строгом) смысле*, если для любого вектора \mathbf{R} и любого числа n выполняется условие:

$$\begin{aligned} p_n(x_1, \mathbf{r}_1 \pm \mathbf{R}, t_1, x_2, \mathbf{r}_2 \pm \mathbf{R}, t_2, \dots, x_n, \mathbf{r}_n \pm \mathbf{R}, t_n) = \\ = p_n(x_1, \mathbf{r}_1, t_1, x_2, \mathbf{r}_2, t_2, \dots, x_n, \mathbf{r}_n, t_n). \end{aligned} \quad (3.154)$$

Строгая однородность означает инвариантность всех статистических характеристик поля относительно выбора начала координат в пространстве. Из строгой однородности следует, что среднее значение поля одинаково во всех точках пространства: $\overline{X}(\mathbf{r}, t) = \overline{X}(t)$, а ковариационная и корреляционная функции значений поля в двух точках пространства зависят не от положения точек в отдельности, а от взаимного расположения этих точек: $K_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = K_X(\mathbf{s}, t_1, t_2)$, $\Psi_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \Psi_X(\mathbf{s}, t_1, t_2)$, где $\mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1$.

Случайное поле $X(\mathbf{r}, t)$ называется *однородным в широком смысле*, если его среднее значение не зависит от координат точек, а ковариационная и корреляционная функции являются функциями лишь разности координат $\mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1$, т.е. выполняются условия:

$$\begin{aligned} \overline{X}(\mathbf{r}, t) &= \overline{X}(t), \\ \Psi_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) &= \Psi_X(\mathbf{s}, t_1, t_2), \\ K_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) &= K_X(\mathbf{s}, t_1, t_2). \end{aligned} \quad (3.155)$$

Из однородности поля в строгом смысле всегда следует его однородность в широком смысле. Обратное в общем случае не верно.

Поле может быть однородным, но не стационарным или стационарным, но не однородным. Для многомерных пространств возможна однородность на отдельных поверхностях.

Если поле однородно и стационарно (в каком-либо смысле), то $\overline{X}(\mathbf{r}, t) = \overline{X} \equiv \text{const}$, $K_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = K_X(\mathbf{s}, \tau)$, $\Psi_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \Psi_X(\mathbf{s}, \tau)$, где $\mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1$, $\tau = t_2 - t_1$.

Однородное поле $X(\mathbf{r}, t)$ называется *изотропным*, если для любых моментов времени t_1, t_2 выполняется условие

$$\Psi_X(\mathbf{s}, t_1, t_2) = \Psi_X(s, t_1, t_2), \quad \text{где } s = \|\mathbf{s}\|, \quad (3.156)$$

т.е. корреляция значений поля в двух точках пространства зависит не от взаимного расположения этих точек, а только от расстояния между ними. Аналогично имеем для ковариационной функции $K_X(\mathbf{s}, t_1, t_2) = K_X(s, t_1, t_2)$.

Если для изотропного поля при любых t_1, t_2 имеет место следующее свойство:

$$\lim_{s \rightarrow \infty} \Psi_X(s, t_1, t_2) = 0,$$

то говорят, что имеет место *расцепление корреляций в пространстве* или *пространственное перемешивание*.

Для количественной оценки скорости затухания корреляции в пространстве при удалении двух рассматриваемых точек поля друг от друга часто вводится *радиус когерентности*. Радиус когерентности полностью аналогичен времени корреляции и также, как интервал корреляции, может быть введен двумя способами.

1. Радиус когерентности определяется интервалом, на котором огибающая корреляционной функции $\gamma_X(s, t_1, t_2)$ убывает не менее чем в e раз (e – основание натурального логарифма) при любых значениях t_1 и t_2 :

$$\gamma_X(s', t_1, t_2) = \frac{\gamma_X(0, t_1, t_2)}{e}, \quad s_{\text{ког}} = \sup_{t_1, t_2} s'(t_1, t_2). \quad (3.157)$$

2. Радиус когерентности определяется по формуле

$$s_{\text{ког}} = \sup_{t_1, t_2} \frac{1}{\gamma_X(0, t_1, t_2)} \int_0^\infty \gamma_X(s, t_1, t_2) ds. \quad (3.158)$$

Радиус когерентности имеет следующий простой смысл: если $\|\mathbf{r}_2 - \mathbf{r}_1\| < s_{\text{ког}}$ (рис. 3.10), то значения $X(\mathbf{r}_1, t_1)$ и $X(\mathbf{r}_2, t_2)$ можно считать статистически независимыми для любых t_1 и t_2 .

Для однородных случайных полей можно определить свойство *пространственной эргодичности* по аналогии с временной эргодичностью случайного процесса.

Пространственная эргодичность первого порядка означает существование следующего предела:

$$\text{l.i.m.}_{V \rightarrow \infty} \frac{1}{V} \int_D f[X(\mathbf{r}, t)] d^M r = \langle f[X(\mathbf{r}, t)] \rangle, \quad (3.159)$$

где слева стоит среднеквадратический предел случайной функции аргумента V ; V – объем области D , в которой задано случайное поле; $f(x)$ – произвольная детерминированная функция своих аргументов, угловые скобки означают усреднение по ансамблю реализаций.

Пространственная эргодичность второго порядка означает существование предела:

$$\begin{aligned} \text{l.i.m.}_{V \rightarrow \infty} \frac{1}{V} \int_D f[X(\mathbf{r}, t_1), X(\mathbf{r} + \mathbf{s}, t_2)] d^M r = \\ = \langle f[X(\mathbf{r}, t_1), X(\mathbf{r} + \mathbf{s}, t_2)] \rangle, \end{aligned} \quad (3.160)$$

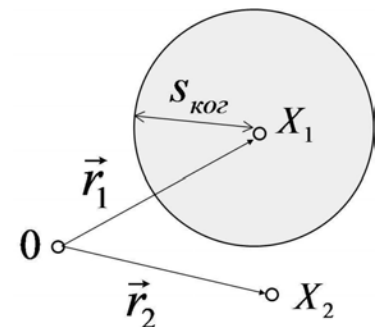


Рис. 3.10. Пояснение понятия радиуса когерентности случайного поля

где $f(x_1, x_2)$ – произвольная детерминированная функция своих аргументов.

Аналогично можно определить пространственную эргодичность 3-го, 4-го, ..., k -го порядка и *строгую пространственную эргодичность* (если поле обладает свойством пространственной эргодичности k -го порядка для любого, сколь угодно большого, k). Для стационарных случайных полей можно ввести понятие *пространственно-временной эргодичности*.

Характеристики комплексных случайных полей. Во многих задачах статистической радиофизики и оптики используются комплексные случайные поля $Z(\mathbf{r}, t) = X(\mathbf{r}, t) + jY(\mathbf{r}, t)$, где $X(\mathbf{r}, t)$ и $Y(\mathbf{r}, t)$ – скалярные вещественные, в общем случае статистически зависимые, случайные поля, представляющие собой вещественную и мнимую компоненты комплексного поля $Z(\mathbf{r}, t)$. Полное статистическое описание случайного поля $Z(\mathbf{r}, t)$ задается совместной плотностью вероятности полей $X(\mathbf{r}, t)$ и $Y(\mathbf{r}, t)$. т.е. $p_n^{XY}(x_1, y_1, \mathbf{r}_1, t_1, x_2, y_2, \mathbf{r}_2, t_2, \dots, x_n, y_n, \mathbf{r}_n, t_n)$.

Для комплексных случайных полей (так же, как и в случае комплексных случайных процессов) можно ввести моментные функции действительных и мнимых компонент, а также моментные функции комплексных случайных полей $Z(\mathbf{r}, t)$. Как и в случае комплексных случайных процессов, для комплексных полей вводятся две ковариационные и две корреляционные функции, отличающиеся своей структурой:

$$K_Z^I(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \langle Z(\mathbf{r}_1, t_1) Z^*(\mathbf{r}_2, t_2) \rangle,$$

$$\Psi_Z^I(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \langle (Z(\mathbf{r}_1, t_1) - \bar{Z}(\mathbf{r}_1, t_1)) (Z^*(\mathbf{r}_2, t_2) - \bar{Z}^*(\mathbf{r}_2, t_2)) \rangle, \quad (3.161)$$

где символ $*$ обозначает комплексно-сопряжённую величину, и

$$K_Z^{II}(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \langle Z(\mathbf{r}_1, t_1) Z(\mathbf{r}_2, t_2) \rangle,$$

$$\Psi_Z^{II}(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \langle (Z(\mathbf{r}_1, t_1) - \bar{Z}(\mathbf{r}_1, t_1)) (Z(\mathbf{r}_2, t_2) - \bar{Z}(\mathbf{r}_2, t_2)) \rangle. \quad (3.162)$$

Для однородного комплексного поля мнимая и вещественная компоненты есть однородные и однородно-связанные поля. Два поля, $X(\mathbf{r}, t)$ и $Y(\mathbf{r}, t)$, называются *однородно-связанными в строгом смысле*, если для любого вектора \mathbf{R} и любого числа n

$$p_{nn}^{XY}(x_1, y_1, \mathbf{r}_1 \pm \mathbf{R}, t_1, x_2, y_2, \mathbf{r}_2 \pm \mathbf{R}, t_2, \dots, x_n, y_n, \mathbf{r}_n \pm \mathbf{R}, t_n) =$$

$$= p_{nn}^{XY}(x_1, y_1, \mathbf{r}_1, t_1, x_2, y_2, \mathbf{r}_2, t_2, \dots, x_n, y_n, \mathbf{r}_n, t_n). \quad (3.163)$$

Два поля, $X(\mathbf{r}, t)$ и $Y(\mathbf{r}, t)$, называются *однородно-связанными в широком смысле*, если

$$\Psi_{XY}(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) = \Psi_{XY}(\mathbf{s}, t_1, t_2), \quad \mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1. \quad (3.164)$$

Спектральное представление случайных полей. Как показано выше, спектральный анализ временных функций, в том числе и случайных, значительно облегчает решение разнообразных радиофизических

задач. При рассмотрении случайных полей использование той же методики вызывает значительные трудности. Во-первых, приходится использовать менее привычное многомерное преобразование Фурье. Во-вторых, вместо временных частотных гармоник появляются пространственные гармоники поля или в разложении одновременно с пространственными формируются и временные гармоники. Поэтому в задачах со случайными полями применяют *пространственное* и *пространственно-временное спектральные разложения* полей.

1. Пространственное спектральное разложение. Пусть $X(\mathbf{r}, t)$ – скалярное вещественное случайное поле, заданное для произвольного момента времени t в пространстве \mathbb{R}^M . Воспользуемся кратным преобразованием Фурье и представим M -мерный (комплексный) пространственный спектр поля $X(\mathbf{r}, t)$ в виде:

$$X(\mathbf{k}, t) = \int_D X(\mathbf{r}, t) e^{-j\mathbf{k}\mathbf{r}} d^M r, \quad (3.165)$$

где \mathbf{k} – M -мерный волновой вектор; D – область пространства, в которой задано случайное поле; $\mathbf{k}\mathbf{r}$ – скалярное произведение векторов; $d^M r$ – M -кратный дифференциал по координатам вектора \mathbf{r} . Если существует среднеквадратический интеграл (3.165), то можно ввести ограниченную по абсолютной величине комплексную функцию

$$G_X(\mathbf{k}_1, \mathbf{k}_2, t) = \langle X(\mathbf{k}_1, t) X^*(\mathbf{k}_2, t) \rangle, \quad (3.166)$$

которую называют *обобщенной пространственной спектральной плотностью случайного поля* $X(\mathbf{r}, t)$. Подставляя (3.165) в (3.166), приходим к выражению

$$G_X(\mathbf{k}_1, \mathbf{k}_2, t) = \int_D \int_D K_X(\mathbf{r}_1, \mathbf{r}_2, t) e^{-j(\mathbf{k}_1 \mathbf{r}_1 - \mathbf{k}_2 \mathbf{r}_2)} d^M r_1 d^M r_2. \quad (3.167)$$

Можно показать, что для однородного поля из (3.167) следует

$$G_X(\mathbf{k}, \mathbf{k}', t) = W_X(\mathbf{k}, t) \delta(\mathbf{k}' - \mathbf{k}), \quad (3.168)$$

где

$$W_X(\mathbf{k}, t) = \int_D K_X(\mathbf{s}, t) e^{-j\mathbf{k}\mathbf{s}} d^M s, \quad \mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1 \quad (3.169)$$

есть *пространственная спектральная плотность мощности случайного поля* $X(\mathbf{r}, t)$. Спектральная плотность мощности $W_X(\mathbf{k}, t)$ – вещественная неотрицательная функция, ограниченная на множестве значений аргументов за исключением, возможно, отдельных точек. Используя обратное преобразование Фурье, получаем

$$K_X(\mathbf{s}, t) = \frac{1}{(2\pi)^M} \int W_X(\mathbf{k}, t) e^{j\mathbf{k}\mathbf{s}} d^M k. \quad (3.170)$$

Формулы (3.169) и (3.170) по существу представляют обобщение теоремы Винера – Хинчина на M -мерные случайные поля. Физический смысл

пространственной спектральной плотности мощности удобно рассмотреть на примере трёхмерного поля ($M=3$), чаще всего используемого в радиофизике.

2. Пространственно-временное спектральное разложение. До сих пор зависимость случайного поля от времени игнорировалась. Обратимся теперь к общему случаю и рассмотрим скалярное вещественное случайное поле, заданное в пространстве \mathbb{R}^M , но испытывающее флуктуации во времени. Используем ту же методику, что и при построении пространственного спектрального разложения, но добавим однократное преобразование Фурье по времени. Тогда

$$\begin{aligned} X(\mathbf{k}, \omega) &= \int_D \int_{-\infty}^{\infty} X(\mathbf{r}, t) e^{-j(\omega t - \mathbf{k}\mathbf{r})} d^M r dt, \\ X(\mathbf{r}, t) &= \frac{1}{(2\pi)^{M+1}} \int \int_{-\infty}^{\infty} X(\mathbf{k}, \omega) e^{j(\omega t - \mathbf{k}\mathbf{r})} d^M k d\omega, \end{aligned} \quad (3.171)$$

где $X(\mathbf{k}, \omega)$ – спектральная плотность случайного поля $X(\mathbf{r}, t)$. Теперь поле $X(\mathbf{r}, t)$ раскладывается по плоским волнам $e^{j(\omega t - \mathbf{k}\mathbf{r})}$.

Обобщённая пространственно-временная спектральная плотность поля определяется как

$$G_X(\mathbf{k}_1, \omega_1, \mathbf{k}_2, \omega_2) = \langle X(\mathbf{k}_1, \omega_1) X^*(\mathbf{k}_2, \omega_2) \rangle. \quad (3.172)$$

Очевидно, что

$$\begin{aligned} G_X(\mathbf{k}_1, \omega_1, \mathbf{k}_2, \omega_2) &= \\ &= \iiint_D \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_X(\mathbf{r}_1, t_1, \mathbf{r}_2, t_2) e^{-j(\omega_1 t_1 - \omega_2 t_2 - \mathbf{k}_1 \mathbf{r}_1 + \mathbf{k}_2 \mathbf{r}_2)} d^M r_1 d^M r_2 d\omega_1 d\omega_2. \end{aligned} \quad (3.173)$$

Для однородного стационарного поля из (3.173) следует

$$G_X(\mathbf{k}, \omega, \mathbf{k}', \omega') = W_X(\mathbf{k}, \omega) \delta(\mathbf{k}' - \mathbf{k}) \delta(\omega' - \omega), \quad (3.174)$$

где

$$\begin{aligned} W_X(\mathbf{k}, \omega) &= \int_D \int_{-\infty}^{\infty} K_X(\mathbf{s}, \tau) e^{-j(\omega \tau - \mathbf{k}\mathbf{s})} d^M s d\tau, \\ \mathbf{s} &= \mathbf{r}_2 - \mathbf{r}_1, \quad \tau = t_2 - t_1 \end{aligned} \quad (3.175)$$

есть пространственно-временная спектральная плотность мощности случайного поля $X(\mathbf{r}, t)$. Спектральная плотность $W_X(\mathbf{k}, \omega)$ – вещественная неотрицательная функция, ограниченная на множестве значений аргументов за исключением, возможно, отдельных точек. Для вещественных полей $W_X(\mathbf{k}, \omega)$ – чётная по аргументу ω . Используя обратное преобразование Фурье, получаем

$$K_X(\mathbf{s}, \tau) = \frac{1}{(2\pi)^{M+1}} \int \int_{-\infty}^{\infty} W_X(\mathbf{k}, \omega) e^{j(\omega \tau - \mathbf{k}\mathbf{s})} d^M k d\omega. \quad (3.176)$$

Для изотропных стационарных вещественных полей $W_X(\mathbf{k}, \omega) = W_X(k, \omega)$ – чётная по $\omega, k \geq 0$. Выражения (3.175) – (3.176) представляют ещё одно обобщение теоремы Винера – Хинчина на переменные во времени многомерные случайные поля и получило название *обобщенной теоремы Винера – Хинчина*.

3.7.2. Случайные волны и их характеристики

Случайные волны – это специфический вид случайных полей, или, говоря иначе, *случайные волны* представляют *случайные поля, имеющие волновую природу*. В общем случае их амплитуды, частоты и фазы являются случайными функциями. Поэтому основные статистические характеристики, полученные выше для случайных полей, будем конкретизировать для случайных волн.

Для радиофизики наибольший интерес представляет специальный класс случайных волн – квазимонохроматические волны. *Случайная квазимонохроматическая волна* является обобщением классического представления о гармонической (монохроматической) волне.

Основные определения. Рассмотрим квазиплоскую квазимонохроматическую волну:

$$\mathcal{E}(\mathbf{r}, z, t) = \dot{\mathcal{E}}_m(\mathbf{r}, z, t) \exp[j(\omega t - k_0 z)], \quad (3.177)$$

где $\mathcal{E}(\dots)$ – одна из компонент вектора электрического поля \mathbf{E} , а $\dot{\mathcal{E}}_m(\mathbf{r}, z, t)$ – её комплексная амплитуда.

Волна (3.177) распространяется вдоль оси Oz со скоростью $v_\phi = \omega/k_0$, где k_0 – абсолютная величина волнового вектора \mathbf{k}_0 , направленного по оси Oz . Радиус-вектор \mathbf{r} задает точку в плоскости фронта, перпендикулярной оси Oz . Так как комплексная амплитуда $\dot{\mathcal{E}}_m(\mathbf{r}, z, t)$ случайным образом зависит от радиус-векторов точек фронта волны и от z , то волна неоднородная. Такими свойствами обладает, например, плоская однородная волна после пропускания через безграничный тонкий слой, прозрачность и запаздывание в котором случайным образом изменяются от точки к точке, или случайные волны в волноводах. Комплексная амплитуда волны $\dot{\mathcal{E}}_m(\mathbf{r}, z, t)$ случайным образом зависит также от продольной координаты z и времени t , причём изменения амплитуды в продольном направлении (вдоль оси z) являются гораздо более «медленными», чем в поперечном направлении. Зависимость амплитуды от времени также предполагается медленной по сравнению с периодом $T = 2\pi/\omega$. Таким образом, в каждой фиксированной точке пространства \mathbf{r}_0, z_0 процесс $\mathcal{E}(\mathbf{r}_0, z_0, t)$ представляет собой гармонический шум. Вещественная волновая переменная $X(\mathbf{r}, z, t) = \text{Re}[\mathcal{E}(\mathbf{r}, z, t)]$ может быть представлена как

$$X(\mathbf{r}, z, t) = \rho(\mathbf{r}, z, t) \cos(\omega t - k_0 z + \varphi(\mathbf{r}, z, t)).$$

Функции $\rho(\mathbf{r}, z, t)$ и $\varphi(\mathbf{r}, z, t)$, представляющие собой вещественную амплитуду волны и флуктуационную компоненту фазы соответственно,

так же, как комплексная амплитуда, являются «медленными» случайными функциями z и t , что обеспечивает квазимонохроматичность волн. Среднее значение случайной переменной $X(\mathbf{r}, z, t)$ (так же, как комплексной переменной $\mathcal{E}(\mathbf{r}, z, t)$) предполагается равным нулю при любых \mathbf{r} , z и t .

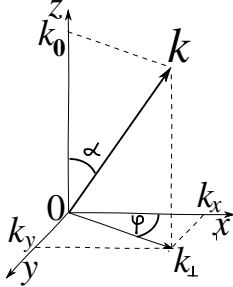


Рис. 3.11. Волновой вектор волны

Для квазиплоской квазимонохроматической волны фаза φ в трансверсальной плоскости меняется незначительно на расстояниях порядка длины волны $\lambda = \frac{2\pi}{k_0}$. В этом случае волновое поле (3.177) можно представить случайной суперпозицией *плоских волн*, волновые векторы которых располагаются вблизи оси Oz и составляют с ней малые углы (рис.3.11).

Введем угол α между направлениями векторов \mathbf{k} и \mathbf{k}_0 . Тогда $k_0 = k \cos \alpha$, $\mathbf{k} = \mathbf{k}_0 + \mathbf{k}_\perp = \mathbf{k}_0 + \mathbf{k}_x + \mathbf{k}_y$.

Для случайных волн типа (3.177) удобно рассматривать отдельно *поперечную* и *продольную корреляционные функции*. *Поперечная корреляционная функция* в заданном сечении z случайной волны (с учетом нулевого среднего значения) определяется как

$$\Psi_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z, t) = \langle \mathcal{E}(\mathbf{r}_1, z, t) \mathcal{E}^*(\mathbf{r}_2, z, t) \rangle = \langle \dot{\mathcal{E}}_m(\mathbf{r}_1, z, t) \dot{\mathcal{E}}_m^*(\mathbf{r}_2, z, t) \rangle, \quad (3.178)$$

где «*» – символ комплексного сопряжения. При $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$ поперечная корреляционная функция дает интенсивность (дисперсию) поля случайной волны

$$I(\mathbf{r}, z, t) = \Psi_{\mathcal{E}\perp}(\mathbf{r}, \mathbf{r}, z, t) = \langle |\mathcal{E}(\mathbf{r}, z, t)|^2 \rangle = \langle |\dot{\mathcal{E}}_m(\mathbf{r}, z, t)|^2 \rangle. \quad (3.179)$$

Поперечная корреляционная функция является наиболее существенной характеристикой пространственного распределения волны. Кроме неё свойства пространственного распределения случайной волны описываются *продольной корреляционной функцией*:

$$\begin{aligned} \Psi_{\mathcal{E}\parallel}(z_1, z_2, \mathbf{r}, t) &= \langle \mathcal{E}(\mathbf{r}, z_1, t) \mathcal{E}^*(\mathbf{r}, z_2, t) \rangle = \\ &= \langle \dot{\mathcal{E}}_m(\mathbf{r}, z_1, t) \dot{\mathcal{E}}_m^*(\mathbf{r}, z_2, t) \rangle \exp[jk_0(z_2 - z_1)]. \end{aligned} \quad (3.180)$$

Если предположить стационарность амплитуды $\dot{\mathcal{E}}_m(\mathbf{r}, z, t)$ во времени, то поперечная и продольная корреляционные функции, а также интенсивность поля волны окажутся не зависящими от времени. Такие случайные волны называются *стационарными*.

Пусть случайная волна (3.177) – стационарная. Между точками \mathbf{r}_1 и \mathbf{r}_2 в поперечном сечении волны можно рассматривать степень пространственной когерентности $R_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z)$, которая определяется нормированной поперечной корреляционной функцией:

$$R_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z) = \left| \frac{\Psi_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z)}{\sqrt{I(\mathbf{r}_1, z)I(\mathbf{r}_2, z)}} \right|. \quad (3.181)$$

Очевидно, что $R_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z) \in [0, 1]$. Если для любых $\mathbf{r}_1, \mathbf{r}_2, z$ выполняется равенство $R_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z) = 1$, то случайная волна называется *когерентной*. Если, напротив, $R_{\mathcal{E}\perp}(\mathbf{r}_1, \mathbf{r}_2, z) \equiv 0$, то значения поля волны в разных точках поперечного сечения статистически независимы. Такие волны называются *некогерентными*.

В наиболее простом случае однородной монохроматической волны её случайная амплитуда зависит только от точки в поперечном сечении и не меняется ни вдоль оси z , ни во времени: $\dot{\mathcal{E}}_m(\mathbf{r}, z, t) = \dot{\mathcal{E}}_m(\mathbf{r})$. В предположении однородности поля волны в поперечном сечении, поперечная корреляционная функция зависит только от расстояния \mathbf{s} между рассматриваемыми точками \mathbf{r}_1 и \mathbf{r}_2 в поперечном сечении: $\mathbf{s} = \mathbf{r}_2 - \mathbf{r}_1$, и её можно представить в виде следующего спектрального разложения:

$$\Psi_{\mathcal{E}\perp}(\mathbf{s}) = \frac{1}{(2\pi)^2} \int W_{\mathcal{E}}(\mathbf{k}_{\perp}) \exp(j\mathbf{k}_{\perp}\mathbf{s}) d^2k_{\perp},$$

где \mathbf{k}_{\perp} – поперечный волновой вектор; $W_{\mathcal{E}}(\mathbf{k}_{\perp})$ – спектральная плотность (мощности) поперечных флуктуаций. Интегрирование производится по всем точкам плоскости, перпендикулярной оси z .

Введем в поперечном сечении волны декартовы координаты ($\mathbf{k}_{\perp} \Rightarrow k_x, k_y$, $\mathbf{s} \Rightarrow s_x, s_y$). Тогда

$$\Psi_{\mathcal{E}\perp}(s_x, s_y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_{\mathcal{E}}(k_x, k_y) \exp[j(k_x s_x + k_y s_y)] dk_x dk_y, \quad (3.182)$$

$$W_{\mathcal{E}}(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi_{\mathcal{E}\perp}(s_x, s_y) \exp[-j(k_x s_x + k_y s_y)] ds_x ds_y. \quad (3.183)$$

Соотношения (3.182) и (3.183) представляют собой преобразования Фурье для переменных k_x, k_y и s_x, s_y . Переменные s_x, s_y характеризуют координаты точек в поперечном сечении волны и имеют размерность длины, а размерность переменных k_x, k_y обратно пропорциональна длине. Переменные с такой размерностью получили название *пространственных частот*. Поэтому функцию $W_{\mathcal{E}}(k_x, k_y)$ будем называть *пространственным спектром* интенсивности поля случайной волны.

В некоторых радиофизических задачах поле в поперечном сечении волны можно считать не только однородным, но и изотропным. Тогда спектральная плотность мощности поперечных флуктуаций зависит только от модуля вектора \mathbf{k}_{\perp} . Покажем это, переходя в (3.183) к полярным координатам. Как известно, переход от декартовых к полярным координатам выполняется по формулам:

$$s_x = s \cos \vartheta, \quad s_y = s \sin \vartheta, \quad ds_x ds_y = s ds d\vartheta, \quad \text{где } s = |\mathbf{s}|, \quad s \geq 0, \quad \vartheta \in [0, 2\pi].$$

Обозначим $k_x = k_{\perp} \cos \varphi$, $k_y = k_{\perp} \sin \varphi$, где $k_{\perp} = |\mathbf{k}_{\perp}|$. Тогда получим

$$k_x s_x + k_y s_y = k_{\perp} s (\cos \varphi \cos \vartheta + \sin \varphi \sin \vartheta) = k_{\perp} s \cos(\varphi - \vartheta).$$

Подставляя новые переменные в (3.183), находим

$$W_{\mathcal{E}}(k_{\perp}, \varphi) = \int_0^{\infty} s ds \int_0^{2\pi} \Psi_{\mathcal{E}\perp}(s, \vartheta) e^{jk_{\perp} s \cos(\vartheta - \varphi)} d\vartheta.$$

Используем интегральную формулу для функций Бесселя

$$J_m(z) = \frac{(-j)^m}{2\pi} \int_0^{2\pi} e^{jz \cos \vartheta} \cos m\vartheta d\vartheta$$

и, подставляя этот интеграл при $m=0$ в $W_{\mathcal{E}}(k_{\perp}, \varphi)$, приходим к

$$W_{\mathcal{E}}(k_{\perp}) = 2\pi \int_0^{\infty} s \Psi_{\mathcal{E}\perp}(s) J_0(k_{\perp} s) ds. \quad (3.184)$$

Это прямое преобразование Ханкеля. Обратное преобразование Ханкеля имеет вид:

$$\Psi_{\mathcal{E}\perp}(s) = \int_0^{\infty} k_{\perp} W_{\mathcal{E}}(k_{\perp}) J_0(k_{\perp} s) dk_{\perp}.$$

Таким образом, для квазиплоских случайных волн, однородных и изотропных в поперечном сечении, целесообразно использовать полярные координаты, а пространственные спектры представлять преобразованиями Ханкеля вместо преобразований Фурье.

Волновой спектр $W_{\mathcal{E}}(k_{\perp}) = W_{\mathcal{E}}(k \sin \alpha)$ (см. рис. 3.11) часто называют *угловым спектром*, так как он определяет среднюю интенсивность излучения под углом α к оси Oz . Поскольку волновое число k_0 определяется частотой квазимонохроматической волны ($k_0 = \frac{\omega}{v_{\text{ф}}}$, где $v_{\text{ф}}$ – фазовая скорость), а k_{\perp} от частоты не зависит, то угловые спектры оказываются различными для различных частот: чем выше частота, тем уже угловой спектр.

Поляризация случайной векторной волны. У векторных случайных полей не только комплексная амплитуда $\dot{\mathcal{E}}_m(\mathbf{r}, z, t)$, но и состояние поляризации является случайным. Поляризация электромагнитной волны (в общем случае случайной) является важной её характеристикой. Поэтому интересно рассмотреть, как влияет статистика на поляризационную структуру и характеристики волны.

Пусть задана квазиплоская квазимонохроматическая волна (3.177). Введем декартову систему координат так, чтобы ось Oz совпала с направлением распространения волны, а векторы $\mathbf{E}(\mathbf{r}, z, t)$ и $\mathbf{H}(\mathbf{r}, z, t)$ лежали в плоскости xOy . Тогда в плоскости фронта волны ($z = \text{const}$) для проекций вектора $\mathbf{E}(\mathbf{r}, z, t)$ имеем:

$$\begin{aligned} E_x(\mathbf{r}, z, t) &= \dot{E}_{mx}(\mathbf{r}, t) \exp [j(\omega t - k_0 z)] = E_{mx}(\mathbf{r}, t) \exp [j(\omega t - k_0 z + \varphi_x(\mathbf{r}, t))], \\ E_y(\mathbf{r}, z, t) &= \dot{E}_{my}(\mathbf{r}, t) \exp [j(\omega t - k_0 z)] = E_{my}(\mathbf{r}, t) \exp [j(\omega t - k_0 z + \varphi_y(\mathbf{r}, t))], \end{aligned}$$

где $E_{mx}(\mathbf{r}, t)$, $\varphi_x(\mathbf{r}, t)$ и $E_{my}(\mathbf{r}, t)$, $\varphi_y(\mathbf{r}, t)$ – случайные амплитуды и начальные фазы проекций вектора $\mathbf{E}(\mathbf{r}, z, t)$ на оси Ox и Oy соответственно.

Будем описывать поляризацию случайной волны энергетическими параметрами (см. разд. 2.7). Для этого введем набор ковариационных функций, усредняя по ансамблю реализаций обе проекции вектора $\mathbf{E}(\mathbf{r}, z, t)$

в одной точке и в один момент времени, и построим матрицу когерентности Винера – Вольфа. Тогда

$$\begin{aligned} K_{xx}(\mathbf{r}) &= \langle E_x(\mathbf{r}, z, t) E_x^*(\mathbf{r}, z, t) \rangle = \langle E_{mx}^2(\mathbf{r}) \rangle, \\ K_{xy}(\mathbf{r}) &= \langle E_x(\mathbf{r}, z, t) E_y^*(\mathbf{r}, z, t) \rangle = \langle E_{mx}(\mathbf{r}) E_{my}(\mathbf{r}) \exp[j\vartheta(\mathbf{r})] \rangle, \\ K_{yx}(\mathbf{r}) &= \langle E_y(\mathbf{r}, z, t) E_x^*(\mathbf{r}, z, t) \rangle = \langle E_{mx}(\mathbf{r}) E_{my}(\mathbf{r}) \exp[-j\vartheta(\mathbf{r})] \rangle, \\ K_{yy}(\mathbf{r}) &= \langle E_y(\mathbf{r}, z, t) E_y^*(\mathbf{r}, z, t) \rangle = \langle E_{my}^2(\mathbf{r}) \rangle, \end{aligned} \quad (3.185)$$

где $\vartheta(\mathbf{r}) = \varphi_x(\mathbf{r}) - \varphi_y(\mathbf{r})$, а поляризационная матрица Винера – Вольфа принимает вид:

$$\hat{\mathbf{J}}(\mathbf{r}) = \begin{pmatrix} K_{xx}(\mathbf{r}), & K_{xy}(\mathbf{r}) \\ K_{yx}(\mathbf{r}), & K_{yy}(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} \langle E_{mx}^2(\mathbf{r}) \rangle, & \langle E_{mx}(\mathbf{r}) E_{my}(\mathbf{r}) \exp[j\vartheta(\mathbf{r})] \rangle \\ \langle E_{mx}(\mathbf{r}) E_{my}(\mathbf{r}) \exp[-j\vartheta(\mathbf{r})] \rangle, & \langle E_{my}^2(\mathbf{r}) \rangle \end{pmatrix}. \quad (3.186)$$

Случайные электромагнитные волны поляризованы лишь частично, т.е. отношение интенсивности поляризованной части волны к полной интенсивности волны, называемое *степенью поляризации* \mathfrak{P} , меньше единицы ($0 < \mathfrak{P} < 1$). Если ввести след матрицы Винера – Вольфа

$$\text{Sp } \hat{\mathbf{J}}(\mathbf{r}) = K_{xx}(\mathbf{r}) + K_{yy}(\mathbf{r}) = \langle E_{mx}^2(\mathbf{r}) \rangle + \langle E_{my}^2(\mathbf{r}) \rangle$$

и её детерминант

$$\det \hat{\mathbf{J}}(\mathbf{r}) = K_{xx}(\mathbf{r}) K_{yy}(\mathbf{r}) - |K_{xy}(\mathbf{r}) K_{yx}(\mathbf{r})|,$$

то степень поляризации \mathfrak{P} можно представить как

$$\mathfrak{P} = \frac{\sqrt{(\text{Sp } \hat{\mathbf{J}}(\mathbf{r}))^2 - 4 \det \hat{\mathbf{J}}(\mathbf{r})}}{\text{Sp } \hat{\mathbf{J}}(\mathbf{r})}.$$

С учетом введённых обозначений параметры Стокса принимают вид:

$$\begin{aligned} S_0(\mathbf{r}) &= K_{xx}(\mathbf{r}) + K_{yy}(\mathbf{r}) = \langle E_{mx}^2 \rangle + \langle E_{my}^2 \rangle = \text{Sp } \hat{\mathbf{J}}(\mathbf{r}), \\ S_1(\mathbf{r}) &= K_{xx}(\mathbf{r}) - K_{yy}(\mathbf{r}) = \langle E_{mx}^2 \rangle - \langle E_{my}^2 \rangle = \mathfrak{P} S_0 \cos 2\tilde{\chi} \cos 2\tilde{\beta}, \\ S_2(\mathbf{r}) &= K_{xy}(\mathbf{r}) + K_{yx}(\mathbf{r}) = \langle 2E_{mx} E_{my} \cos \vartheta \rangle = \mathfrak{P} S_0 \cos 2\tilde{\chi} \sin 2\tilde{\beta}, \\ S_3(\mathbf{r}) &= -j(K_{xy}(\mathbf{r}) - K_{yx}(\mathbf{r})) = \langle 2E_{mx} E_{my} \sin \vartheta \rangle = \mathfrak{P} S_0 \sin 2\tilde{\chi}, \end{aligned} \quad (3.187)$$

где $\tilde{\chi}$ и $\tilde{\beta}$ – эквивалентные геометрические параметры поляризованной части волны. Все параметры Стокса имеют тот же физический смысл, что и в случае детерминированной волны. Единственное отличие состоит в том, что для случайной волны сумма квадратов трёх последних параметров не равна квадрату первого параметра, а равна $(\mathfrak{P} S_0)^2$.

При произвольной поляризации случайных волн коэффициент взаимной корреляции составляющих поля волны описывается формулой

$$R_{xy}(\mathbf{r}) = \sqrt{\frac{K_{xy}(\mathbf{r}) K_{yx}(\mathbf{r})}{K_{xx}(\mathbf{r}) K_{yy}(\mathbf{r})}}, \quad 0 < R_{xy}(\mathbf{r}) < 1,$$

а модуль коэффициента взаимной корреляции составляющих поля не превышает коэффициента поляризации волны: $|R_{xy}(\mathbf{r})| \leq \mathfrak{P}(\mathbf{r})$.

Круг задач, решаемых в радиофизике с использованием случайных полей и случайных волн, чрезвычайно широк. Теория случайных полей распространяется прежде всего на системы с распределёнными параметрами и тепловые флуктуации в сплошных средах и в электромагнитных полях.

Для радиоизлучения случайные неоднородности больших размеров не играют роли, поэтому обычно исходят из предположения о локальной однородности или даже локальной изотропности полей. Большой интерес представляют статистические задачи о распространении волн в средах со случайными неоднородностями, в однородных средах со случайными граничными условиями или ограниченными поверхностями со случайными свойствами (неровности граничной поверхности, случайно распределённые участки с различными коэффициентами отражения, разрывы граничных поверхностей), а также о пропускании волн через слоистые среды, у которого прозрачность и набег фазы случайным образом меняются от точки к точке и с течением времени.

Особый интерес для радиофизики представляет исследование воздействий случайных полей теплового излучения на характеристики монохроматических волн, в частности, появление целой полосы частот $|\Delta\omega|$ в окрестности ω , возникновение которой обусловлено хаотической модуляцией первичной волны меняющимися неоднородностями.

3.8. Контрольные вопросы и упражнения

1. Является ли стационарным в широком смысле случайный процесс, представляющий собой сумму стационарного в широком смысле случайного процесса и независимой от него случайной величины с ограниченным средним квадратом?

2. Записать, чему равна дисперсия случайного процесса $Y(t) = X(t) + \alpha t$, где $X(t)$ – случайный процесс с дисперсией $\sigma_X^2 = 1$, а α – не зависящая от $X(t)$ случайная величина со средним значением $\langle \alpha \rangle = \bar{\alpha}$ и средним квадратом $\langle \alpha \rangle = \bar{\alpha}^2$.

3. Известна одномерная характеристическая функция $\Theta_X(u, t)$ случайного процесса $X(t)$. Как найти средний квадрат случайного процесса, не вычисляя плотности вероятности?

4. Что означает термин «непрерывный случайный процесс»?

5. Квазидетерминированный случайный процесс $X(t)$ определяется как

$$X(t) = A \cos(\omega_0 t + \phi),$$

где A и ω_0 – заданные константы; ϕ – случайная величина, равномерно распределённая в интервале $[-\pi, \pi]$. Найти среднее значение \bar{X} , корреляционную функцию $\Psi_X(t_1, t_2)$ и коэффициент корреляции $R_X(t_1, t_2)$ процесса $X(t)$.

6. Какие случайные процессы называются процессами с перемешиванием? Как характеризуется скорость перемешивания?

7. Показать, что стационарный в широком смысле нормальный процесс

с ограниченной дисперсией является стационарным в строгом смысле.

8. Процесс $Y(t)$ представляет собой сумму $Y(t) = aX_1(t) + bX_2(t)$, где $X_1(t)$ и $X_2(t)$ являются статистически независимыми стационарными случайными процессами со спектральными плотностями мощности $W_{X_1}(\omega)$ и $W_{X_2}(\omega)$, a и b – заданные константы. Найти спектральную плотность мощности процесса $Z(t)$.

9. Спектральная плотность мощности стационарного случайного процесса $X(t)$ равна константе W_0 в ограниченной полосе частот $[-\omega_1; \omega_1]$ и обращается в ноль на всех других частотах. Найти дисперсию случайного процесса.

10. Чему равны дисперсия и время корреляции белого шума?

11. Что называется амплитудой (огибающей) стационарного случайного процесса?

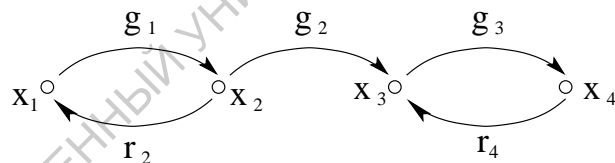
12. Какой процесс называется k -связанным марковским процессом?

13. Одношаговая матрица простой однородной марковской цепи имеет вид

$$\hat{\pi}_1 = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/3 & 1/6 \end{bmatrix}$$

Найти стационарное распределение вероятностей $P_m^{\text{ст}}$, $m = 1, 2, 3$ и соответствующее ему среднее значение \bar{X} , если цепь имеет следующие состояния: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$.

14. Найти стационарные вероятности $P_m^{\text{ст}}$, $m = 1, 2, 3, 4$ одношагового процесса, изображённого на схеме.



15. Записать управляющее уравнение для пуассоновского процесса.

16. Диффузионный процесс $X(t)$ характеризуется следующими коэффициентами сноса и диффузии: $A(x, t) \equiv 0$, $B(x, t) = B \equiv \text{const}$. Чему равна дисперсия процесса σ_X^2 ?

17. Является ли винеровский процесс гауссовским?

18. Случайный процесс $X(t)$ задается следующим СДУ:

$$\frac{dX}{dt} + X(X - 1) = Xn(t), \quad X \geq 0,$$

где $n(t)$ – нормальный белый шум со спектром $W_n = 1$. Найти стационарную плотность вероятности $p_1^{\text{ст}}(x)$ случайного процесса $X(t)$ в случае, когда СДУ понимается как: а) СДУ Ито; б) СДУ Стратоновича. Будут ли результаты качественно различными?

19. Записать уравнение Ито, эквивалентное следующему уравнению Стратоновича:

$$\frac{dX}{dt} = X + X^2\sqrt{2D}n(t),$$

где $n(t)$ – нормированный нормальный белый шум, $D \equiv \text{const}$.

20. Какое случайное поле называют однородным в строгом смысле?

21. Как определяется интервал корреляции (когерентности) однородного

изотропного случайного поля?

22. В чем заключается физический смысл понятий «пространственная гармоника» и «спектр пространственных гармоник»?

23. Как различаются пространственное и пространственно-временное спектральные разложения?

24. Какими свойствами обладает обобщённая пространственно-временная спектральная плотность стационарного поля?

25. Что понимают под поперечной и продольной корреляционными функциями случайных волн?

26. В чем состоит двойной физический смысл пространственных частот?

27. Почему при уменьшении частоты излучения угловой спектр пространственных гармоник волны сужается?

28. Какой физический смысл имеют коэффициенты матрицы Винера – Вольфа частично поляризованной волны?

29. Случайная частично поляризованная волна имеет степень поляризации 0.5, угол эллиптичности 22.5° , угол ориентации эллипса – 15° . Определить параметры Стокса волны.

30. Параметры Стокса частично поляризованной волны равны $8, \sqrt{3}, 1, 2\sqrt{3}$. Рассчитать степень поляризации волны и параметры поляризационного эллипса.

Список литературы

Стратонович, Р. Л. Случайные процессы в динамических системах. Ижевск : Изд-во ИКИ, НИЦРХД, 2009.

Рытов, С. М. Введение в статистическую радиофизику. Ч. I: Случайные процессы. М. : Наука, 1976.

Ахманов, С. А., Дьяков, Ю. Е., Чиркин, А. С. Статистическая радиофизика и оптика. Случайные колебания и волны в линейных системах. М. : Физматлит, 2010.

Тихонов, В. И., Миронов, М. А. Марковские процессы. М. : Сов. радио, 1977.

Тихонов, В. И., Харисов, В. Н. Статистический анализ и синтез радиотехнических устройств и систем. М. : Радио и связь, 2004.

Шахтарин, Б. И. Случайные процессы в радиотехнике: в 2 т. Т.1: Линейные преобразования. Т.2: Нелинейные преобразования. М. : Гелиос АРВ, 2006.

Левин, Б. Р. Теоретические основы статистической радиотехники. М. : Радио и связь, 1989.

Тихонов, В. И., Шахтарин, Б. И., Сизых, В. В. Случайные процессы. Примеры и задачи : в 4 т. Т.1 : Случайные величины и процессы. М. : Радио и связь, 2003.

Гардинер, К. В. Стохастические модели в естественных науках. М. : Мир, 1986.

Казаков, В. А. Введение в теорию марковских процессов и некоторые радиотехнические задачи. М. : Сов.радио, 1977.

Ван Кампен, Н. Г. Стохастические процессы в физике и химии. М. : Высш. шк., 1990.

Хорстхемке, В., Лефевр, Р. Индуцированные шумом переходы. М. : Мир, 1987.

Вентцель, А. Д., Фрейдлин, М. И. Флуктуации в динамических системах под действием малых случайных возмущений. М. : Наука, 1979.

Анищенко, В. С., Астахов, В. В., Вадивасова, Т. Е., Нейман, А. Б., Стрелкова, Г. И., Шиманский-Гайер, Л. Нелинейные эффекты в хаотических и стохастических системах. М. ; Ижевск : Институт компьютерных исследований, 2003.

Зверев, В. А. Радиооптика (преобразования сигналов в радио и оптике). М. : Сов.радио, 1975.

Астахов, В. В., Вадивасова, Т. Е., Хохлов, А. В. Радиоп физика. Задачи и упражнения: учеб. пособие. Саратов : Изд-во Саратов. ун-та, 2008.

Поздняк, С. И., Мелитицкий, В. А. Введение в статистическую теорию поляризации радиоволн. М. : Сов.радио, 1974.

Хохлов А. В., Перельмутер Г. И. Матричное описание поляризации трёхмерных электромагнитных волн // Лекции по электронике СВЧ и радиоп физике: 9-я зимняя шк.-семинар инженеров: в 3 кн. Кн.1. Саратов : Изд-во Саратов. ун-та, 1992.

Глава 4

Дискретные и цифровые сигналы. Методы цифровой обработки

4.1. Дискретное представление сигналов

Как показано в разд. 1.1, радиотехнические сигналы можно разделить на аналоговые, дискретные и цифровые. Значения аналоговых сигналов задаются в произвольные моменты времени и принимают любые значения в заданных пределах. Количество информации, заключенной в таких сигналах, никогда не используется полностью. Для практических целей достаточно знать значения сигналов в отдельные дискретные моменты времени.

Дискретизацией будем называть процесс представления аналогового сигнала $x(t)$, заданного на временной оси $t \in]-\infty, \infty[$, последовательностью отсчётов $\{x_i\}$ в виде бесконечно-коротких импульсов с амплитудами, равными значениям сигнала в выбранные моменты времени t_i : $x_i = x(t_i)$. Операцию дискретизации можно задать в виде интегрального оператора:

$$x_i = \int_{-\infty}^{\infty} x(t)\delta(t - t_i)dt, \quad i \in Z, \quad (4.1)$$

а сам дискретизованный сигнал записать в форме аналогового сигнала:

$$x_d(t) = \sum_{i=-\infty}^{\infty} x(t)\xi(t - t_i), \quad (4.2)$$

где $\xi(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases}$ - функция одиночного импульса.

Дискретизация может быть равномерной (регулярной) или неравномерной (адаптивной), когда на интервалах быстрого изменения значений сигналов отсчёты берутся чаще, чем на интервалах медленного изменения. В дальнейшем будем рассматривать только равномерную (регулярную) дискретизацию.

При представлении сигналов регулярными отсчётами возникает вопрос о величине шага дискретизации Δt . Понятно, что при уменьшении Δt и увеличении частоты дискретизации $\omega_d = 2\pi/\Delta t$ искажения информации уменьшаются, но одновременно растет объем обрабатываемых данных. Поэтому оптимальным ω_d следует считать такое минимальное

значение $\omega_d^{(min)}$, при котором ещё принципиально возможно восстановить аналоговый сигнал с заданной погрешностью.

Связь между спектрами аналогового и дискретного сигналов. Рассмотрим сначала, чем определяется выбор шага дискретизации по времени. Чтобы оценить те явления, которые возникают при дискретизации сигнала, удобно перейти от временного представления сигналов к спектральному. Для этого рассмотрим сначала, как связаны спектры аналогового и дискретизованного сигналов. Пусть для $-\infty < t < \infty$ задан некоторый непрерывный аналоговый сигнал $x(t)$, которому соответствует функция спектральной плотности $X(\omega)$:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \exp(-j\omega t) dt.$$

Построим другой сигнал $x_d(t)$, представляющий собой выборку исходного сигнала $x(t)$ в дискретные моменты времени $t = 0, \Delta t, 2\Delta t, 3\Delta t, \dots$ аналогично тому, как это показано на рис. 4.1.

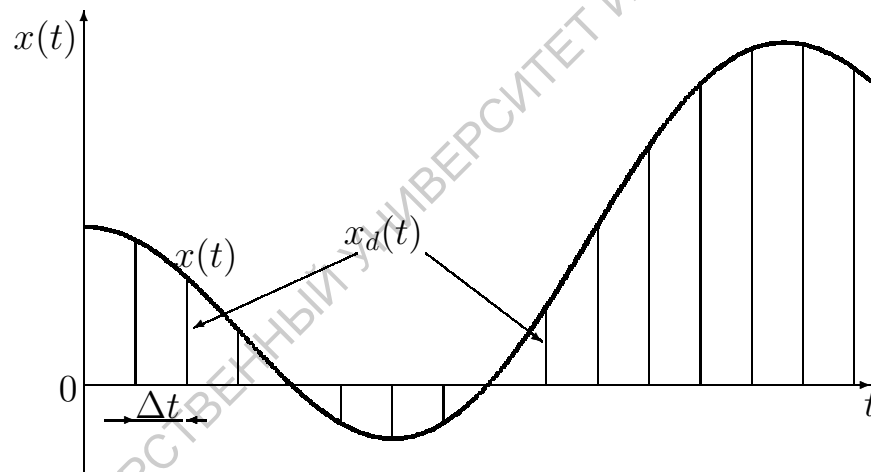


Рис. 4.1. Дискретизация по времени аналогового сигнала с равномерным шагом Δt

Для того чтобы записать $x_d(t)$ аналитически, используем формулу (4.2), где в качестве моментов времени t_i возьмём значения, кратные шагу дискретизации: $t_i = i\Delta t$. Тогда

$$x_d(t) = \sum_{i=-\infty}^{\infty} x(t) \xi(t - i\Delta t), \quad (4.3)$$

где

$$\xi(t) = \begin{cases} 1, & t = 0, \\ 0, & t \neq 0. \end{cases}$$

Если мы попробуем определить функцию спектральной плотности от (4.3) обычным образом, как

$$X_d(\omega) = \int_{-\infty}^{\infty} x_d(t) \exp(-j\omega t) dt,$$

то обнаружим, что она тождественно равна нулю. Это легко понять, если обратить внимание, что сигнал $x_d(t)$ всюду разрывен: он равен нулю почти везде, за исключением счётного числа точек. Чтобы обойти эту неприятность, введем вспомогательный сигнал

$$y(t) = \sum_{i=-\infty}^{\infty} x(t)\delta(t - i\Delta t), \quad (4.4)$$

где $\delta(t)$ – дельта функция Дирака, равная нулю везде, за исключением $t = 0$, где она обращается в бесконечность. Сигнал $y(t)$ отличается от дискретной выборки $x_d(t)$ только своей амплитудой – она у него бесконечно велика. Однако, как известно, форма спектра не зависит от амплитуды сигнала, поскольку интегральное преобразование Фурье – линейная операция. Поэтому форма спектра для сигнала $y(t)$ будет совпадать с формой спектра для сигнала $x_d(t)$, при том, что величина этого спектра, как мы увидим далее, будет ненулевой.

Построим функцию спектральной плотности для дискретизованного сигнала $y(t)$ и сравним её с $X(\omega)$:

$$Y(\omega) = \int_{-\infty}^{\infty} x(t) \left[\sum_{n=-\infty}^{\infty} \delta(t - n\Delta t) \right] \exp(-j\omega t) dt. \quad (4.5)$$

Обозначим сумму дельта-функций через $\varphi(t)$: $\varphi(t) = \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t)$. Поскольку $\varphi(t)$ является периодической функцией времени с периодом Δt , она может быть записана в виде комплексного ряда Фурье:

$$\varphi(t) = \sum_{k=-\infty}^{\infty} \dot{C}_k \exp\left(j\frac{2\pi}{\Delta t}kt\right), \quad (4.6)$$

где

$$\dot{C}_k = \frac{1}{\Delta t} \int_{-\Delta t/2}^{\Delta t/2} \delta(t) \exp(-j2\pi kt/\Delta t) dt = \frac{1}{\Delta t}.$$

Так как в интервал интегрирования $(-\Delta t/2, \Delta t/2)$ попадает только $\delta(0)$, интеграл обращается в единицу.

Подставляя \dot{C}_k и (4.6) в (4.5) и меняя порядок суммирования и интегрирования, получим:

$$Y(\omega) = \frac{1}{\Delta t} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \exp\left(-j\left[\omega - \frac{2\pi}{\Delta t}k\right]t\right) dt.$$

Выражение под интегралом представляет собой ни что иное, как спектральную плотность сигнала $x(t)$ на частоте $\omega - 2\pi k/\Delta t$, где $2\pi/\Delta t = \omega_d$. Поэтому окончательно получаем:

$$Y(\omega) = \frac{1}{\Delta t} \sum_{k=-\infty}^{\infty} X\left(\omega - \frac{2\pi}{\Delta t}k\right) = \frac{1}{\Delta t} \sum_{k=-\infty}^{\infty} X(\omega - k\omega_d). \quad (4.7)$$

Таким образом, из (4.7) следует, что спектр дискретизованного с равномерным шагом Δt сигнала представляет собой «наложение» на спектр аналогового сигнала $X(\omega)$ (рис. 4.2, а) его копий, сдвинутых на частоты, кратные частоте дискретизации $\omega_d = 2\pi/\Delta t$. Отсюда можно сделать несколько важных для практического приложения выводов.

1. Спектр дискретного сигнала есть периодическая функция с периодом, равным частоте дискретизации. Поэтому *спектры дискретных сигналов имеет смысл измерять в полосе частот от 0 до ω_d или в главном частотном диапазоне функции $X_d(\omega)$.*

2. Если спектр аналогового сигнала ограничен сверху частотой ω_h , причём $2\omega_h \leq \omega_d$, то спектр дискретизованного сигнала будет представлять собой периодическое продолжение спектра аналогового сигнала (рис. 4.2, б).

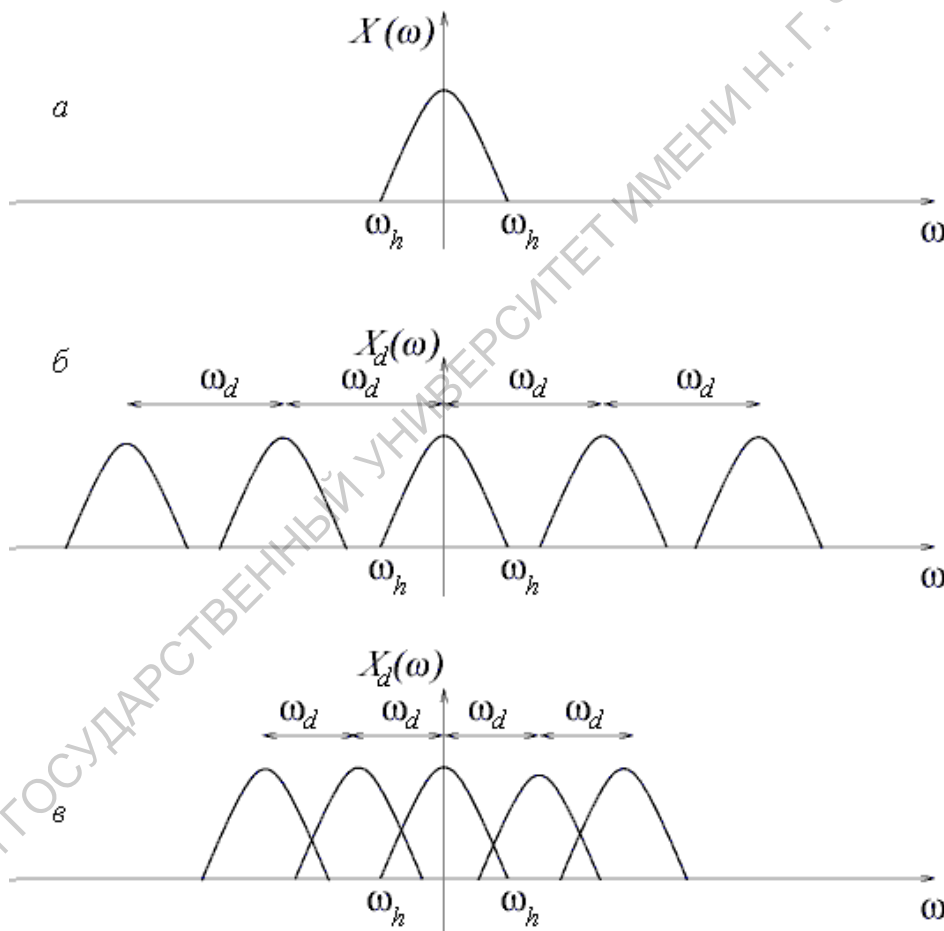


Рис. 4.2. Спектр аналогового сигнала (а) и спектры дискретизованного сигнала при $2\omega_h < \omega_d$ (б) и $2\omega_h > \omega_d$ (в)

3. Если спектр аналогового сигнала имеет сколь угодно высокие частоты или же он ограничен сверху частотой ω_h , но при этом частота дискретизации меньше удвоенной граничной частоты спектра $2\omega_h > \omega_d$, спектр периодического сигнала будет представлять собой наложение на спектр $X(\omega)$ его копий, сопровождающееся перекрытием частей спектра, как это показано на рис. 4.2, в.

Из второго и третьего пунктов можно сделать важный вывод: для того чтобы спектр дискретизованного сигнала не искажался за счёт наложения, необходимо, чтобы частота дискретизации была как минимум вдвое выше максимальной частоты в спектре аналогового сигнала. Данное правило называют критерием Котельникова (а в англоязычной литературе – критерием Найквиста).

Итак, рассмотрев свойства спектра дискретизованного сигнала, мы определили условие для шага дискретизации, которое достаточно для обработки сигналов с определённым характером спектра:

$$\omega_d^{(min)} = 2\omega_h. \quad (4.8)$$

Что произойдет, если данный критерий не будет выполнен? В этом случае, как видно из рис. 4.2, в, произойдет перекрытие спектра с собственным периодическим продолжением таким образом, что низкочастотные составляющие спектра накладываются на высокочастотные составляющие. Такое явление называют «просачиванием» высокочастотных компонент спектра в область низких частот (в англоязычной литературе – *aliasing*).

Предположим, например, что при записи акустического сигнала, спектр которого простирается до 20 кГц, в помещении от какого-то устройства присутствует ультразвуковой сигнал с частотой 30 кГц. Запись выполняется с дискретизацией сигнала на выходе микрофона со стандартной частотой 44.1 кГц, т.е. более чем в два раза превышает максимальную частоту акустического сигнала. Если бы использовалась аналоговая звукозаписывающая аппаратура, то сигнал ультразвуковой частоты после записи и последующего воспроизведения не привел бы к искажениям, поскольку он не воспринимается человеческим ухом. Иначе обстоит дело при записи дискретизованного сигнала. При прослушивании такой записи мы услышим шумовой сигнал на разностной частоте $44.1 - 30 = 14.1$ кГц, т.е. восстановленный сигнал будет выглядеть так, как если бы частоты, лежащие выше половины частоты дискретизации, зеркально от неё отразились в нижнюю часть спектра и сложились с присутствующими там гармониками. Это так называемый эффект *появления ложных (кажущихся) частот*. Эффект аналогичен всем известному эффекту обратного вращения колес автомобиля на экранах кино и телевизоров, когда скорость их вращения начинает превышать частоту смены кадров. Для предотвращения появления ложных частот следует повышать частоту дискретизации или ограничить спектр сигнала перед дискретизацией фильтрами нижних частот, которые пропускают без изменения все частоты ниже частоты среза и подавляют в сигнале частоты выше частоты среза. Частота среза антиалиасинговых фильтров устанавливается равной половине частоты дискретизации.

Итак, исследуя свойства спектра дискретных сигналов, мы обнаружили, что выбор частоты дискретизации обусловлен свойствами спектра исходного сигнала: частота дискретизации должна быть как минимум вдвое выше верхней граничной частоты спектра.

Дискретизация полосовых сигналов. Частота дискретизации – основной параметр, определяющий процесс дискретизации. Существуют

ли сигналы, для которых можно обойти правило (4.8)? Ответ на этот вопрос положительный.

Если обрабатываемый сигнал имеет полосовой спектр, т.е. кроме верхней граничной частоты ω_h он имеет и нижнюю граничную частоту ω_l , для его дискретизации можно использовать метод, известный как *полосовая дискретизация*. Данный метод позволяет существенно уменьшить частоту дискретизации и, кроме того, как будет показано ниже, осуществить перенос спектра сигнала в низкочастотную область.

Полосовая дискретизация в литературе упоминается под различными другими названиями, такими как дискретизация ПЧ, гармоническая дискретизация, субнаиквистовская дискретизация и дискретизация с пониженной частотой. Когда ширина спектра и центральная частота непрерывного входного сигнала позволяют, полосовая дискретизация не только обеспечивает снижение частоты дискретизации, но и уменьшает объем памяти, необходимый для хранения сигнала на заданном интервале.

В качестве примера рассмотрим дискретизацию сигнала с ограниченным спектром. Пусть спектр сигнала имеет полосу частот шириной $\Delta f = 5$ МГц и локализован в окрестности центральной частоты $f_c = 20$ МГц, т.е. занимает диапазон частот от $f_a = 17.5$ до $f_b = 22.5$ МГц (рис. 4.3, а). Заметим, что полный комплексный спектр сигнала, как это следует из свойств спектра вещественных сигналов, состоит из положительных частотных гармоник и зеркально симметричных им гармоник на отрицательных частотах. Поэтому на рис. 4.3, а кроме треугольника в диапазоне частот от 17.5 до 22.5 МГц показан зеркально симметричный ему треугольник с частотами от -17.5 до -22.5 МГц.

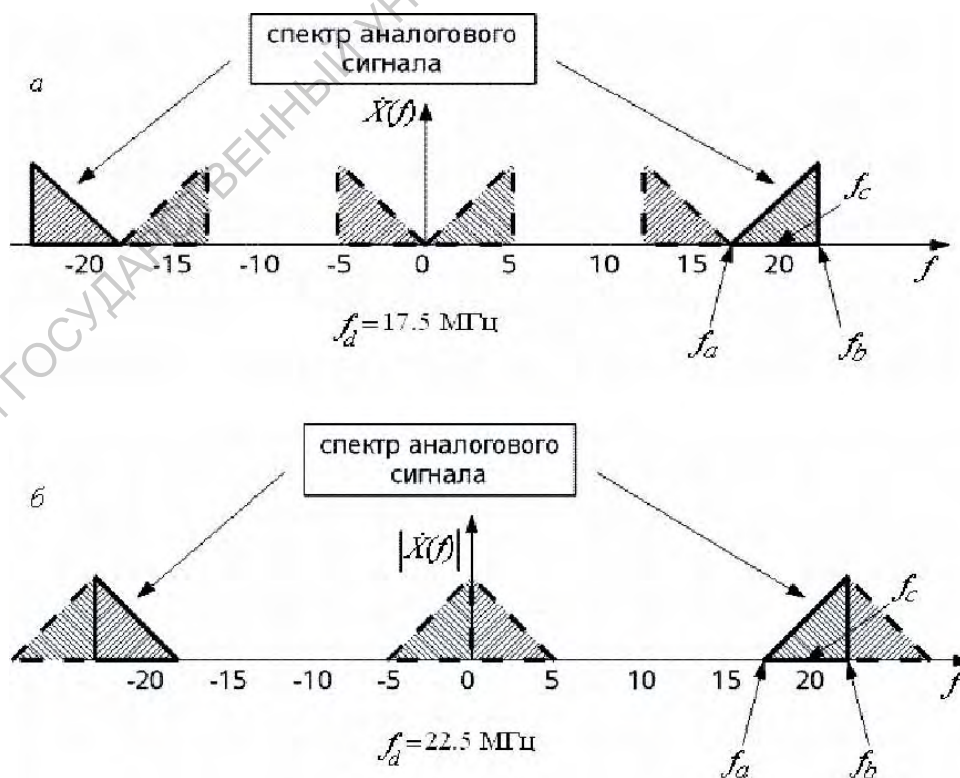


Рис. 4.3. Пример полосовой дискретизации сигнала с ограниченным спектром

Для дискретизации такого сигнала, в соответствии с критерием Найквиста (дискретизация с частотой, превышающей в два раза наивысшую частоту в спектре сигнала), нам необходимо обеспечить частоту выборки более 45 МГц (интервал дискретизации $\Delta t \cong 22.2$ нс). Оказывается, однако, что для полосовых сигналов можно использовать и меньшую частоту дискретизации.

Рассмотрим, что произойдет при частоте дискретизации $f_d = 17.5$ МГц. Как следует из формулы (4.7) при дискретизации в спектре сигнала, кроме тех компонент, которые присутствуют в спектре аналогового сигнала, будут присутствовать их «копии» на частотах, отстоящих на $\pm f_d$, $\pm 2f_d$ и т.д. В результате в спектре дискретизованного сигнала появятся компоненты, изображённые на рис. 4.3, а штриховыми линиями.

Из рисунка хорошо видно, что наложения основного спектра и копий не происходит – копии основной части спектра «промахиваются» относительно неё, попадая в те частотные области, где спектральные компоненты аналогового сигнала отсутствуют. Действительно, область основной части спектра, располагающаяся в полосе частот около частоты $-f_c$, «копируется» в области около частот $f_d - f_c = -2.5$ МГц и $2f_d - f_c = 15$ МГц. Аналогично этому область основной части спектра, располагающаяся в полосе частот около частоты f_c , «копируется» в области около частот $f_c - f_d = 2.5$ МГц и $f_c - 2f_d = -15$ МГц. Дискретизация не приводит к наложению, но переносит спектр сигнала в область низких частот.

Дискретизация с пониженной частотой может не только переносить спектр исходного сигнала в низкочастотную область, но и зеркально отражать его. Рассмотрим, например, дискретизацию того же сигнала с частотой $f_d = 22.5$ МГц. Амплитудный спектр дискретизованного сигнала показан на рис. 4.3, б. Мы видим, что в данном случае «зеркальная» часть спектра аналогового сигнала переносится в область положительных частот от 0 до 5 МГц, а составляющие на положительных частотах – в область отрицательных частот от 0 до -5 МГц. В итоге спектр в низкочастотной области оказывается зеркально симметричным спектру аналогового сигнала.

Из рассмотренных примеров ясно, что можно выбрать частоту дискретизации, не удовлетворяющую критерию Найквиста и меньшую, чем $\omega_d^{(min)} = 2\omega_h$, при которой наложения спектров не произойдет. Таковыми оказались $f_d = 22.5$ МГц и $f_d = 17.5$ МГц. Последняя почти в три раза меньше $f_d^{(min)}$. А можно ли ещё уменьшить частоту дискретизации полосового сигнала, выбрав, например, $f_d = 16$ МГц или $f_d = 14$ МГц?

Легко показать, что при $f_d = 16$ МГц области спектра от 1.5 до 6.5 МГц и от 9.5 до 14.5 МГц не перекрываются, а при $f_d = 14$ МГц наложение спектров будет иметь место (необходимые расчёты предлагается выполнить самостоятельно). Каким же образом следует выбирать частоту полосовой дискретизации? Проведем формальное рассмотрение этого вопроса.

Пусть спектр аналогового сигнала является полосовым, т.е. заключён в области частот между $f_A > 0$ и $f_B > f_A$. Обозначим центральную частоту спектра $f_c = (f_B + f_A)/2$ и его ширину $\Delta f = f_B - f_A$.

В области отрицательных частот соответственно будет находиться зеркальное отражение в диапазоне частот $-f_B \leq f \leq -f_A$. Рассмотрим «размножение» отрицательной полосы спектра при дискретизации с частотой f_d (размножение положительной полосы будет происходить симметричным образом). За счет дискретизации в спектре появятся копии отрицательной полосы спектра на частотах $-f_B + f_d \leq f \leq -f_A + f_d$, $-f_B + 2f_d \leq f \leq -f_A + 2f_d$, $-f_B + 3f_d \leq f \leq -f_A + 3f_d$ и т.д. Предположим, что внутри диапазона частот, занимаемого спектром аналогового сигнала $-f_B \leq f \leq f_B$ помещается k копий отрицательной полосы спектра. Это значит, что k -ая копия располагается в диапазоне частот $-f_B + kf_d \leq f \leq -f_A + kf_d$. Чтобы не было перекрытия с положительной полосой спектра ($f_A \leq f \leq f_B$), верхняя частота копии $-f_A + kf_d$ должна быть меньше нижней частоты основной полосы f_A :

$$-f_A + kf_d \leq f_A. \quad (4.9)$$

С другой стороны, следующая копия с номером $k+1$ должна «перескочить» через основную полосу, не пересекаясь с ней. Это означает, что нижняя граница диапазона частот этой копии $-f_B + (k+1)f_d$ должна быть выше, чем верхняя граница диапазона частот основной полосы f_B :

$$-f_B + (k+1)f_d \geq f_B. \quad (4.10)$$

Подставляя в неравенства (4.9) и (4.10) значения $f_A = f_c - \Delta f/2$ и $f_B = f_c + \Delta f/2$ и объединяя их в одно, получим критерий «правильности» выбора частоты полосовой дискретизации:

$$\frac{2f_c + \Delta f}{k+1} \leq f_d \leq \frac{2f_c - \Delta f}{k}, \quad (4.11)$$

где k – некоторое целое положительное число. Из (4.11) легко получить условие для максимального значения k :

$$\frac{2f_c + \Delta f}{k_{max} + 1} \leq \frac{2f_c - \Delta f}{k_{max}},$$

или

$$0 < k_{max} \leq \frac{f_c}{\Delta f} - \frac{1}{2}. \quad (4.12)$$

Таким образом, для рассмотренного нами случая на рис. 4.3, a максимальное значение k составляет $20/5 - 0.5 = 3$. Поэтому возможные значения k для полосовой дискретизации $k = \{1, 2, 3\}$. В первом случае ($k=1$) частота дискретизации должна лежать в диапазоне $22.5 \leq f_d \leq 35$ МГц, во втором случае ($k=2$) – $15 \leq f_d \leq 17.5$ МГц, а в третьем ($k=3$) – $11.25 \leq f_d \leq 11.66$ МГц.

Полосовая дискретизация может использоваться не только для понижения частоты выборки, но и в ряде операций, связанных с понижением частотного диапазона сигнала, например при детектировании.

4.2. Восстановление аналогового сигнала по его равномерной выборке

Восстановление непрерывного аналогового сигнала по его дискретной выборке представляет одну из важных проблем радиоэлектроники. Поскольку процедуры расчёта спектра и восстановления сигнала из спектра (прямое и обратное преобразование Фурье) взаимно однозначны, то и исходный аналоговый сигнал может быть восстановлен. Главный вопрос состоит в том, насколько точно можно восстановить аналоговый сигнал по его выборке, ведь при дискретизации необратимо теряется информация о значениях сигнала в моменты времени, промежуточные между точками отсчёта. В общем случае ошибки восстановления зависят от интервала дискретизации, структуры и спектра аналогового сигнала, выбранной системы базисных функций. Покажем, как можно осуществить процесс восстановления аналогового сигнала по дискретной выборке.

Ряд Котельникова и теорема Котельникова – Шенона.

Теорема. Аналоговый непрерывный сигнал $x(t)$, спектр которого ограничен сверху частотой ω_h , в любой момент времени t однозначно определяется его выборочными значениями, отсчитываемыми через интервалы Δt , удовлетворяющими критерию Найквиста $\omega_d = 2\pi/\Delta t \geq 2\omega_h$:

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t) \frac{\sin(\omega_d(t - k\Delta t)/2)}{\omega_d(t - k\Delta t)/2} = \sum_{k=-\infty}^{\infty} x(k\Delta t) \operatorname{sinc}(\pi t/\Delta t - \pi k). \quad (4.13)$$

Доказательство. Пусть сигнал $x(t)$ обладает конечной энергией и принадлежит гильбертовому пространству $\mathbb{H}(t)$, спектр сигнала ограничен частотами от нуля до ω_h , а $x_d(k) = x_d(k\Delta t)$ – его выборка с шагом дискретизации Δt , т.е. частота дискретизации $\omega_d = 2\pi/\Delta t$. Предположим также, что выбор шага дискретизации соответствует критерию Найквиста $\omega_d \geq 2\omega_h$. Запишем выражение сигнала через обратное преобразование Фурье функции спектральной плотности:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \exp(j\omega t) d\omega.$$

Поскольку в силу критерия Найквиста весь спектр сигнала содержится в полосе от $-0.5\omega_d$ до $0.5\omega_d$, пределы в интеграле можно заменить:

$$x(t) = \frac{1}{2\pi} \int_{-\omega_d/2}^{\omega_d/2} X(\omega) \exp(j\omega t) d\omega. \quad (4.14)$$

Теперь введем вспомогательную функцию $X_1(\omega)$, являющуюся периодическим продолжением $X(\omega)$, и подставим её в (4.14) вместо $X(\omega)$ ($X_1(\omega) = X(\omega)$, если $\omega \in [-0.5\omega_d; 0.5\omega_d]$):

$$x(t) = \frac{1}{2\pi} \int_{-\omega_d/2}^{\omega_d/2} X_1(\omega) \exp(j\omega t) d\omega. \quad (4.15)$$

Данная замена правомерна, поскольку на интервале интегрирования обе функции совпадают. С другой стороны, $X_1(\omega)$, в отличие от $X(\omega)$ – периодическая функция с периодом, равным ω_d . Следовательно, её можно разложить в ряд Фурье:

$$X_1(\omega) = \sum_{k=-\infty}^{\infty} \dot{C}_k \exp\left(j\frac{2\pi k}{\omega_d}\omega\right) = \sum_{k=-\infty}^{\infty} \dot{C}_k \exp(jk\omega\Delta t), \quad (4.16)$$

где

$$\dot{C}_k = \frac{1}{\omega_d} \int_{-\omega_d/2}^{\omega_d/2} X_1(\omega) \exp\left(-j\frac{2\pi k}{\omega_d}\omega\right) d\omega = \frac{\Delta t}{2\pi} \int_{-\omega_d/2}^{\omega_d/2} X_1(\omega) \exp(-jk\omega\Delta t) d\omega.$$

Сравнивая интегралы в этом выражении и в формуле (4.15), получим для \dot{C}_k :

$$\dot{C}_k = \Delta t \cdot x(-k\Delta t).$$

Теперь подставим \dot{C}_k в (4.16), а (4.16) в (4.15) и поменяем местами последовательность суммирования и интегрирования. Тогда

$$\begin{aligned} x(t) &= \frac{\Delta t}{2\pi} \sum_{k=-\infty}^{\infty} x(-k\Delta t) \int_{-\omega_d/2}^{\omega_d/2} \exp(j\omega(t+k\Delta t)) d\omega = \\ &= \sum_{k=-\infty}^{\infty} x(-k\Delta t) \frac{\exp(j\omega_d(t+k\Delta t)/2) - \exp(-j\omega_d(t+k\Delta t)/2)}{j\omega_d(t+k\Delta t)} = \\ &= \sum_{k=-\infty}^{\infty} x(-k\Delta t) \frac{\sin(\omega_d(t+k\Delta t)/2)}{\omega_d(t+k\Delta t)/2} = \sum_{k=-\infty}^{\infty} x(-k\Delta t) \text{sinc}(\omega_d(t+k\Delta t)/2). \end{aligned} \quad (4.17)$$

Инвертируя знаки перед $k\Delta t$, окончательно получим

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t) \text{sinc}(\omega_d(t-k\Delta t)/2) = \sum_{k=-\infty}^{\infty} x(k\Delta t) \text{sinc}(\pi t/\Delta t - \pi k). \quad (4.18)$$

Теорема доказана.

Таким образом, восстановление непрерывного аналогового сигнала при оптимальном интервале дискретизации происходит в соответствии с (4.18), т.е. при дискретизации по времени с «правильным» шагом не происходит необратимой потери информации аналогового сигнала. Правда, на сами сигналы накладывается определённое ограничение: их спектр должен быть ограничен сверху по частоте. Однако для экспериментальных сигналов это ограничение не является сдерживающим – любые сигналы с ограниченной энергией имеют ограниченный по частоте спектр. Поэтому любые радиосигналы могут быть подвергнуты дискретизации без потери их качества. Один из примеров восстановления аналогового сигнала представлен на рис. 4.4. По существу, теорема Котельникова –

Шенона иллюстрирует возможность перехода от интегральных преобразований Фурье $x(t)$ к рядам Фурье.

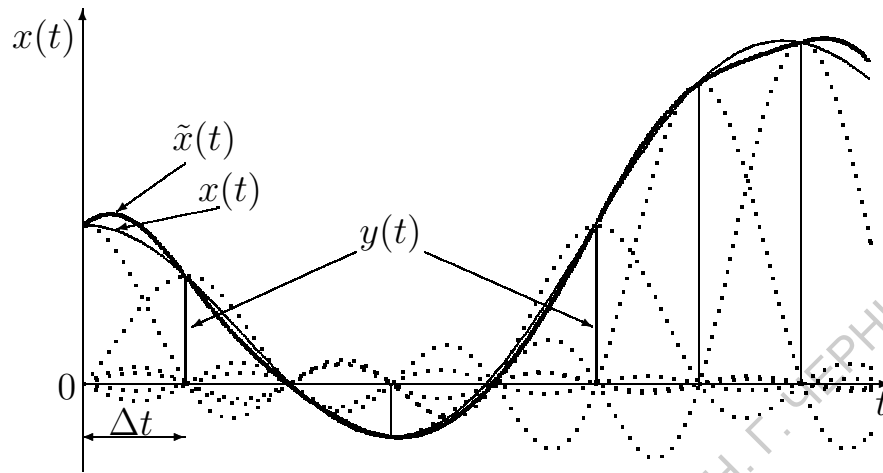


Рис. 4.4. Пример восстановления аналогового сигнала

Правая часть формулы (4.18) называется *рядом Котельникова*¹. Базисные функции

$$\varphi_k(t) = \frac{\sin(\omega_d(t - k\Delta t)/2)}{\omega_d(t - k\Delta t)/2} = \text{sinc}(\omega_d(t - k\Delta t)/2) \quad (4.19)$$

также называются *функциями Котельникова* или *функциями отсчётов*, являются ортогональными на интервале $[-\infty, \infty]$ и представляют собой последовательно сдвинутые функции $\sin t/t$.

Используя дуальность времени и частоты (см. гл. 2, разд. 2.4), можно сформулировать «дуальную теорему» Котельникова – Шенона или *теорему отсчётов в частотной области*:

комплексная спектральная плотность $X(\omega)$ ограниченного во времени интервалом $[-T/2, T/2]$ аналогового сигнала на любой частоте ω *однозначно определяется* выборочными значениями его спектра, отсчитываемые через интервал $\bar{\omega} = 2\pi/T$, называемый *частотой разрешения спектра*:

$$X(\omega) = \sum_{k=-\infty}^{\infty} X\left(k\frac{2\pi}{T}\right) \frac{\sin\left(\frac{\omega - k\frac{2\pi}{T}}{2}\right) \frac{T}{2}}{\left(\frac{\omega - k\frac{2\pi}{T}}{2}\right) \frac{T}{2}} = \sum_{k=-\infty}^{\infty} X\left(k\frac{2\pi}{T}\right) \text{sinc}\left(\frac{\omega T}{2} - \pi k\right). \quad (4.20)$$

Эту теорему предлагается доказать самостоятельно.

Заметим, что по формуле Котельникова целое число k , соответствующее номеру базисной функции и номеру отсчёта при дискретизации сигнала, должно изменяться в интервале $(-\infty, \infty)$, т.е. выборка должна содержать бесконечно большое число отсчётов, что невозможно на практике.

Восстановление сигналов с «усеченным» спектром. В общем случае длительность сигнала и ширина его спектра не могут быть одновременно ограничены конечными интервалами. Если сигнал недостаточно быстро затухает, то спектр сигнала неограничен. Конечно, его можно искусственно ограничить сверху достаточно большой частотой ω_h так,

¹Такой ряд впервые ввел в 1915 г. Е. Т. Уиттакер в теории интерполяции.

чтобы составляющие спектра за границей ω_h были пренебрежимо малы, но при этом все равно возникают погрешности восстановления.

Пусть сигнал $x(t)$ задан в интервале $[0, T]$ и равен нулю за его пределами. Предположим, что спектр «усечен» на частоте ω_h идеально, т.е. без искажений для частот $\omega \leq \omega_h$. Выбираем интервал дискретизации Δt . Чтобы избежать перекрытия спектров, задаем $\omega_d = 2\omega_h$, где $\omega_d = 2\pi/\Delta t$. Получаем число отсчетов $N = T/\Delta t$.

Рассмотрим теперь, как представить восстановленный сигнал $\tilde{x}(t)$ с усеченным спектром в виде разложения по функциям Котельникова. По условиям задачи $\Delta t = 2\pi/\omega_d = \pi/\omega_h$, число отсчетов $N = T\omega_h/\pi$ и совпадает с удвоенной базой сигнала $D = T\Delta f$, где $\Delta f = \omega_h/2\pi$ (см. разд. 2.4.2), а

$$\tilde{x}(t) = \sum_{k=1}^{2D} x(k\Delta t) \text{sinc}(\omega_d(t - k\Delta t)). \quad (4.21)$$

Функция

$$\xi(t) = x(t) - \tilde{x}(t)$$

представляет локальные погрешности восстановления (рис. 4.5). Очевидно, что функция $\xi(t)$ обращается в нуль в точках отсчета и максимальна в центрах интервалов дискретизации. Интегральную погрешность восстановления целесообразно оценивать среднеквадратическим отклонением на интервале $[0, T]$:

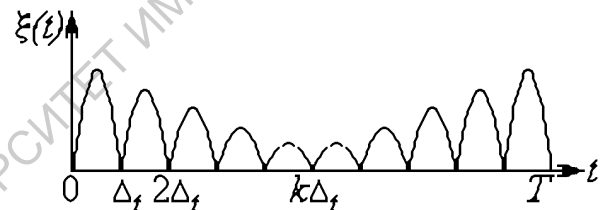


Рис. 4.5. Локальные погрешности восстановления сигнала

$$\bar{\xi} = \sqrt{\frac{1}{T} \int_0^T (x(t) - \tilde{x}(t))^2 dt}. \quad (4.22)$$

Таким образом, дискретизация усеченных сигналов возможна, но при обработке последних следует контролировать среднеквадратичные ошибки искажений. Можно показать, что для действительных или комплекснозначных функций, интегрируемых с квадратом, ряд (4.18) равномерно сходится к $x(t)$ при $2D \rightarrow \infty$.

4.3. Представление дискретных сигналов в цифровой форме

Цифровая форма представления обусловлена особенностями микропроцессорной обработки сигналов. Использование микропроцессорной техники для обработки сигналов – основная тенденция современной радиоэлектроники, поскольку:

- 1) легко реализовать чрезвычайно сложные алгоритмы обработки сигналов;
- 2) одно и то же устройство можно использовать для разных методов обработки посредством замены программы.

Большинство сигналов, с которыми приходится иметь дело при обработке данных, являются аналоговыми по своей природе, дискретизированными и квантованными в силу методических особенностей измерений или технических особенностей регистрации, т.е. преобразованными в цифровые сигналы. Процессоры могут выполнять арифметические операции над числами, заданными с конечной точностью, а точность задания числовых значений определяется в свою очередь числом двоичных разрядов регистров процессора.

Но существуют и сигналы, которые изначально относятся к классу цифровых, как, например, отсчёты количества гамма-квантов, зарегистрированных по последовательным интервалам времени, число посещений интернет-сайта в течение заданного интервала времени, сезонные изменения численности популяций животных и т.д.

Цифровые сигналы представляют собой числовые последовательности, элементы которых – вещественные числа, заданные с конечной точностью, определяемой интервалом квантования. Последнее делает их изоморфными целым числам.

Непрерывные сигналы можно преобразовать в цифровую форму, последовательно выполняя операции дискретизации по времени и квантования по уровню. Операция дискретизации по времени описана в предыдущих разделах.

Квантование сигналов. После получения дискретного сигнала $x_d(n\Delta t) = x(n)$ с ним производится операция квантования по уровню. Квантование сводится к замене бесконечного множества возможных значений непрерывного сигнала, в общем случае случайного, конечным множеством дискретных значений и выполняется следующим образом.

1. Выбирается величина *шага квантования* Δ_x .
2. Формируется множество отсчётов сигнала $x_i = i\Delta_x$, где i – целое:

$$\mathbb{Q} = \{\dots x_{-3}, x_{-2}, x_{-1}, x_0 = 0, x_1, x_2, \dots\}.$$
3. Исходный дискретный сигнал $x(n)$ замещается сигналом $x_i(n)$ так, чтобы в каждый момент времени n значение x_i из множества отсчётов \mathbb{Q} было наиболее близким к $x(n)$:

$$x(n) \rightarrow x_i, \text{ так что } |x_i - x(n)| \leq |x_j - x(n)| \text{ для всех } j \neq i.$$

Представленная операция квантования математически может быть выражена формулой:

$$x_i(n) = \Delta_x \text{INT} \left(\frac{x(n)}{\Delta_x} + \frac{1}{2} \right),$$

где $\text{INT}(\dots)$ – целая часть величины, стоящей в скобках. Число уровней квантования всегда конечно. Если интервал возможных значений простирается от x_{min} до x_{max} , то число интервалов квантования равно $(x_{max} - x_{min})/\Delta_x$, а количество дискретных уровней квантования определяется по формуле:

$$N_x = \left[\frac{x_{max} - x_{min}}{\Delta_x} \right] + 1.$$

Последовательность отсчётов $x_i(n)$, построенная на основе аналогового сигнала $x(t)$, является дискретной последовательностью вещественных чисел. По существу, сигнал $x_i(n)$ по своим значениям (отсчётам) является формализованной разновидностью дискретного сигнала $x_d(n)$ при округлении отсчётов последнего до определённого количества цифр. При квантовании сигнала, в отличие от дискретизации, происходит необратимая потеря информации. Эта потеря может быть учтена как добавление к исходному сигналу внешнего шума, называемого *шумом квантования*. Пусть $x(t)$ – исходный аналоговый сигнал, а $x_i(t)$ – сигнал, квантованный с шагом Δ_x . Тогда разница между ними

$$\xi(t) = x(t) - x_i(t)$$

и представляет шум квантования. На рис. 4.6, *а* изображено квантование гармонического сигнала единичной амплитуды с шагом $\Delta_x = 0.25$.

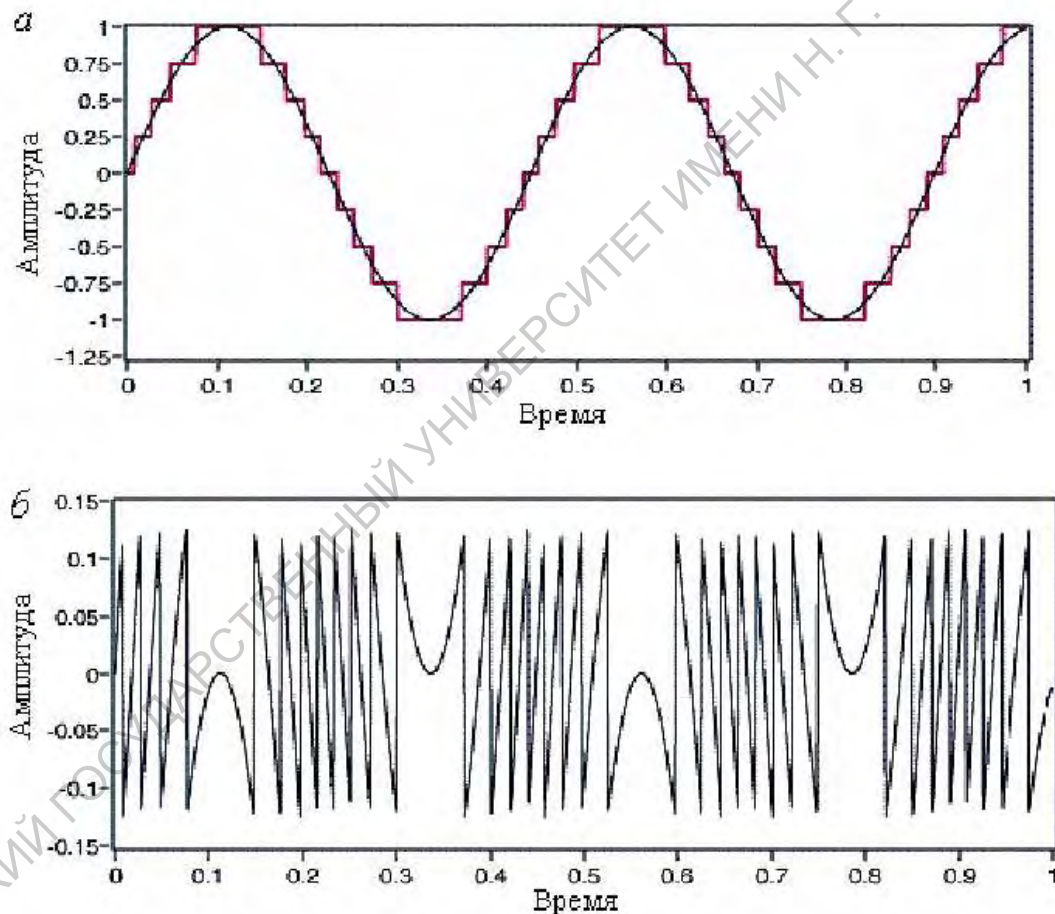


Рис. 4.6. Квантование аналогового сигнала (*а*) и шум квантования (*б*)

Из процедуры квантования ясно, что величина $\xi(t)$ представляет погрешность округления и лежит в пределах от $-0.5\Delta_x$ до $0.5\Delta_x$ (рис.4.6, *б*), причём все значения ξ равновероятны. Поэтому считается, что ошибка квантования – шумовой процесс с равномерной плотностью вероятности $p(\xi) = 1/\Delta_x$. Характер этого процесса зависит от квантуемого сигнала. Для регулярных сигналов ошибка квантования также регулярный сигнал. Однако в радиоэлектронике типичные сигналы – случайные, так как именно случайные сигналы используются для передачи информации.

Статистические свойства шума квантования будут определяться статистическими свойствами исходного сигнала. Для большинства оценок влияния этого шума предполагается, что шум квантования является некоррелированным, т.е. обладает свойствами белого шума.

Пусть ошибка квантования $\xi(t)$ – белый шум с равномерным распределением плотности вероятности ($p(\xi) = 1/\Delta_x$) и нулевым средним. Дисперсия такого процесса σ_ξ^2 определяется шагом квантования Δ_x :

$$\sigma_\xi^2 = \int_{-0.5\Delta_x}^{0.5\Delta_x} \xi^2 p(\xi) d\xi = \frac{x^3}{3\Delta_x} \Big|_{-0.5\Delta_x}^{0.5\Delta_x} = \frac{\Delta_x^2}{12}, \quad (4.23)$$

а среднеквадратическое значение этой погрешности

$$\bar{\sigma} = \Delta_x / 2\sqrt{3}.$$

Рассмотренное квантование называется *равномерным*, так как шаг Δ_x остается постоянным на всем интервале возможных значений $x_d(n)$. При квантовании сигналов в большом динамическом диапазоне шаг квантования может быть и неравномерным, например логарифмическим, т.е. пропорциональным логарифму значений входного сигнала. Неравномерное квантование, несмотря на то, что оно сложнее в реализации, чем равномерное, довольно часто используют, например, при квантовании сигналов разговорной речи.

Кодирование квантованных сигналов. Кодовые комбинации. «Кодирование сообщений» – весьма емкое понятие. В математике *кодированием* называется отображение произвольного множества A в множество последовательностей (кодовых комбинаций), составленных из символов некоторого алфавита B . Примером кодирования может служить представление натуральных чисел в m -ичной системе счисления, когда числу N соответствует кодовая комбинация в алфавите $B = \{0, 1, \dots, m-1\}$ такая, что $b_{k-1}m^{k-1} + b_{k-2}m^{k-2} + \dots + b_1m^1 + b_0m^0 = N$, где m – основание кода, k – число разрядов m -ичного (например, десятичного, восьмеричного или двоичного) числа.

При передаче дискретных сообщений квантованные отсчеты чаще всего представляются двоичными комбинациями. Кодовые комбинации с одинаковым числом элементов (с одинаковой длиной) называются *равномерными*. Хотя увеличение основания кода вызывает рост пропускной способности за счёт сокращения числа разрядов кода, большее число используемых символов усложняет линию передачи. В отсутствии помех пропускная способность должна была бы расти пропорционально основанию кода, но при наличии помех этого не происходит (ухудшается различимость отдельных символов кода).

Выбор двоичных кодов объясняется прежде всего простотой реализации цифрового представления дискретных сигналов. Наибольшее распространение получили *натуральный двоичный код, симметричный двоично-числовой код и код Грея* (см. разд. 1.2). Комбинации натурального двоичного кода представляют собой номера уровней квантования. Симметричный двоично-числовой код используется для представления биполярных квантованных отсчетов. При этом высший разряд несет информацию о знаке отсчета, а остальные разряды – об абсолютном значении в

натуральном двоичном коде. Если количество дискретных уровней квантования равно N_x , то число разрядов m -ичного кода выбирается таким, чтобы $m^k \geq N_x$, а для натурального двоичного кода – $2^k \geq N_x$.

Таким образом, разрядность k определяет точность кодирования квантованных сигналов, а значит, и точность представления дискретных сигналов в цифровой форме.

Так, в соответствии с международными соглашениями аналоговому речевому сигналу в телефонии отводится диапазон частот 4 кГц, и согласно теореме Котельникова – Шенона частота дискретизации должна составлять 8 кГц. При квантовании отсчетов амплитуд используется 256 стандартных уровней, которые кодируются 8-разрядными двоичными символами (словами). Отсюда скорость потоков данных, передающих речь, составляет 64 кбит/с. Частота 8 кГц, 8-битовая схема кодирования и скорость потока данных 64 кбит/с составляют стандарт для цифровых сетей всех видов. Последующие скорости кратны 64 кбит/с.

Код Грея $[(a_{k-1}^\Gamma a_{k-2}^\Gamma \dots a_0^\Gamma)]$ связан с натуральным двоичным кодом $[(a_{k-1} a_{k-2} \dots a_0)]$ следующими соотношениями $a_0^\Gamma = a_0 \oplus a_1$, $a_1^\Gamma = a_1 \oplus a_2$, ..., $a_{k-2}^\Gamma = a_{k-2} \oplus a_{k-1}$, $a_{k-1}^\Gamma = a_{k-1}$, где $(a_{k-1}^\Gamma a_{k-2}^\Gamma \dots a_0^\Gamma)$ – кодовая комбинация кода Грея; символ \oplus означает суммирование по модулю два. Он обладает, по сравнению с натуральным кодом, двумя особенностями: 1) любые две кодовые комбинации, соответствующие соседним уровням квантования, отличаются друг от друга только в одном разряде; 2) при переходе от одной комбинации к другой смена значений элементов в каждом разряде происходит вдвое реже, чем в натуральном двоичном коде.

Рассмотренные коды обеспечивают одинаковую погрешность восстановления из-за помех при условии, что они возникают независимо от передаваемого сигнала и соседние ошибки независимы.

Точность кодирования – не единственная определяющая характеристика цифровых кодов. Не менее важными являются скорость передачи данных, ширина полосы используемых частот, помехоустойчивость. В соответствии с этими требованиями различают несколько методов (вариантов) цифрового кодирования.

Простейший метод NRZ (Non Return to Zero без возврата к нулю) использует постоянное напряжение V как бит 1 и $V = 0$ как бит 0. При передаче нескольких единиц конец одного импульса соединяется с началом следующего единичного импульса и сигнал не возвращается к нулю.

Такие коды называются *потенциальными*. Спектр сигнала постоянно изменяется в зависимости от структуры данных. При чередовании единиц и нулей и скорости передачи информации n бит/с период первой гармоники сигнала $T = 2/n$, а ее частота $n/2$ Гц. Среднее напряжение (рис. 4.7, а) оказывается отличным от нуля, спектр сигнала смещается в область низких частот, а длинные последовательности нулей и единиц затрудняют синхронизацию.

Другой потенциальный код представляет метод NRZI с инверсией

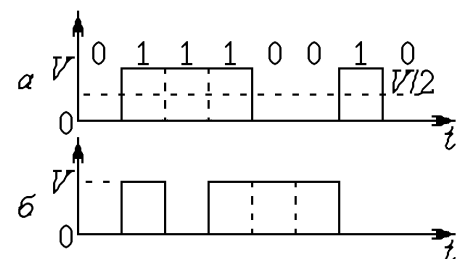


Рис. 4.7. Цифровое кодирование методом NRZ(a) и NRZI(b)

уровня при передаче единицы (Non Return to Zero with ones Inverted). Передаче нуля сохраняет ранее установленный уровень сигнала, а при передаче единицы происходит инверсия уровня (рис. 4.7, б): если предыдущее значение сигнала было равно нулю, единица устанавливается как напряжение $+V$, если же предыдущее значение было $+V$, то устанавливается $V=0$. Таким образом, при передаче длинных последовательностей единиц среднее значение V сохраняется. При чередовании единиц и нулей частота первой гармоники равна $n/4$ Гц. Оба потенциальных кода обладают хорошей помехозащищённостью.

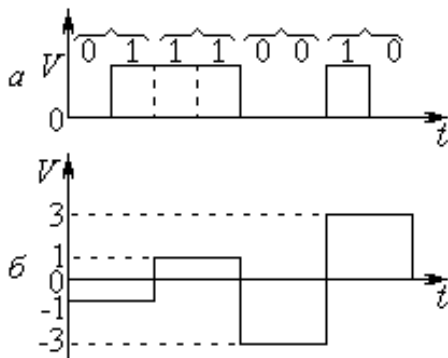


Рис. 4.8. Исходная последовательность (а) и кодирование методом 2B1Q (б)

Для повышения скорости передачи информации часто используются многоуровневые коды. В данном случае правильнее говорить не об увеличении скорости передачи, а об увеличении информационного содержания каждого символа. Примером многоуровневого потенциального кода может служить популярный код 2B1Q (2 бита за 1 такт). При этом каждой паре битов (дибиту) присваивается одно из четырех напряжений: -3 В для дибита 00, -1 В для 01, $+1$ В для 11, $+3$ В для 10 (рис. 4.8). Скорость передачи сокращается вдвое, а частота первой гармоники не превышает $n/4$ Гц.

Еще одну группу кодов образуют импульсные коды, представляемые перепадами потенциалов (рис. 4.9). При этом используется весь импульс вместе с его фронтами. Примером импульсного кода может служить метод биполярного импульсного кодирования АМІ (Alternating Mark Inversion), использующий три уровня сигнала: 0, $+1$ и -1 . Логическому нулю соответствует нулевой потенциал, а логическая единица кодируется попеременно положительным или отрицательным потенциалом (рис. 4.9, б). Для любой последовательности битов сигнал не содержит постоянной составляющей, а частота первой гармоники чередующихся импульсов

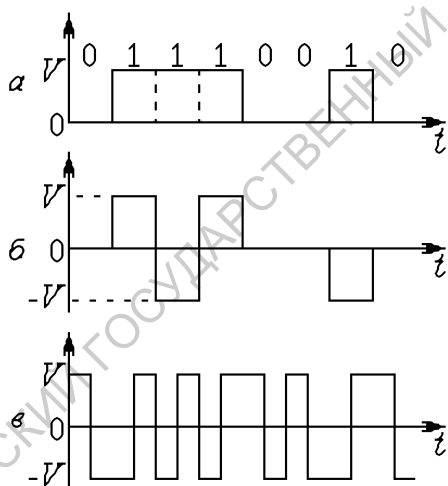


Рис. 4.9. Исходная последовательность (а) и кодирование методом АМІ (б) и манчестерским кодом (в)

равна $n/4$ Гц. Как показывают расчеты, спектр АМІ оказывается еще уже. Основной недостаток АМІ состоит в потере синхронизации, когда встречаются длинные последовательности нулей.

Примером самосинхронизирующегося импульсного кода является абсолютно биимпульсный (манчестерский) код (рис. 4.9, в), в котором единицы и нули кодируются перепадами уровней в середине тактовых интервалов: нуль – перепадом от высокого уровня ($+V$) к низкому ($-V$), а единица – от низкого уровня к высокому. Манчестерский код имеет всего

два уровня, а постоянная составляющая при любых сигналах отсутствует. Длительность каждого импульса равна половине битового интервала. Частота первой гармоники оказывается самой высокой при передаче длительных последовательностей единичных или нулевых символов.

Кроме простых двоичных кодов при кодировании используются так называемые помехоустойчивые коды, позволяющие обнаруживать и исправлять ошибки, возникающие из-за действия помех. Помехоустойчивые коды являются избыточными, так как содержат кроме информационных специальные проверочные символы, а значит, и дополнительные разряды в кодовых комбинациях. Наличие избыточных символов придает кодам корректирующие свойства.

В заключение оценим количественно влияние ошибок квантования на процесс обработки сигнала. Для этого воспользуемся часто используемой в радиоэлектронике характеристикой – *отношением сигнал/шум* (SNR – signal-to-noise ratio), которая определяется как отношение мощностей сигнала к шуму.

Пусть случайный сигнал $x(t)$ изменяется в пределах от $x_{min} = -x_{max}$ до x_{max} и этот интервал разделен на N_x уровней квантования. Поскольку и сигнал, и шум квантования представляют собой случайные процессы, вместо мощностей можно использовать их статистические эквиваленты – дисперсии σ_x^2 и σ_ξ^2 . Тогда для SNR в децибелах имеем:

$$\text{SNR} = 10 \lg \frac{\sigma_x^2}{\sigma_\xi^2} \text{ (дБ)}.$$

Используя формулу (4.23) и учитывая, что при двоичном кодировании уровней квантования $\Delta_x = 2x_{max}/N_x = 2x_{max}/2^k$, получаем дисперсию шума квантования в виде:

$$\sigma_\xi^2 = \frac{\Delta x^2}{12} = \frac{1}{12} \left(\frac{2X_{max}}{2^k} \right)^2 = \frac{X_{max}^2}{3 \times 2^{2k}}. \quad (4.24)$$

Оценить дисперсию полезного случайного сигнала трудно, так как нужно знать и форму сигнала, и его фактическую амплитуду X_m . Если же исследуемый сигнал гармонический, то отношение пикового значения сигнала к действующему составляет $\sqrt{2}$. Поэтому для гармонического сигнала $\sigma_x = \frac{X_{max}}{\sqrt{2}}$, а $\text{SNR} = 10 \lg(1.5 \cdot 2^{2k})$. Отсюда

$$\text{SNR} = 10 \lg 1.5 + 20 \frac{\log_2 2^k}{\log_2 10} = 6.02k + 1.76. \quad (4.25)$$

Для сигналов другой формы (негармонических) это соотношение будет нарушаться, однако различие будет невелико, и соотношение (4.25) вполне может использоваться для оценки SNR произвольных сигналов.

О чем же свидетельствует формула (4.25)?

Если для кодирования квантованного сигнала использован 8-разрядный двоичный код (SNR = 50 дБ), а мощность исследуемого сигнала в миллион раз превосходит мощность шума (SNR = 60 дБ), то в результате кодирования и шум, и самые слабые сигналы будут утеряны в шуме квантования, что приведет к потере качества сигнала. Не менее серьезные эффекты возникают и в других практически важных случаях.

Пусть обрабатываемый аналоговый сигнал не содержит шумовой компоненты, однако содержит в своём спектре гармонические компоненты

малой интенсивности. Тогда шум квантования может «замаскировать» эти компоненты так, что они перестанут быть различимы в спектре оцифрованного сигнала. Для оценки точности представления такого сигнала в цифровом виде можно использовать то же соотношение (4.25): SNR не должно превосходить отношение между мощностью основной гармоники и тех гармоник, которые должны быть различимы в спектре оцифрованного сигнала. Например, в рассмотренных выше примерах, 8-разрядное двоичное кодирование позволит различить в спектре выходного сигнала гармонические составляющие, мощность которых составляет -50 дБ от мощности базовой гармоники.

Таким образом, непрерывный сигнал $x(t)$ может быть преобразован в цифровой. Погрешности, вносимые при таком преобразовании, определяются формулами (4.22) и (4.23), могут быть уменьшены до требуемой величины соответствующим выбором шагов дискретизации по времени и квантования по уровню.

4.4. Цифровой спектральный анализ

Среди множества методов анализа радиосигналов спектральные методы являются наиболее распространёнными. Под спектром аналогового сигнала $x(t)$ зачастую понимают получающуюся в результате интегрального преобразование Фурье (ИПФ):

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt \quad (4.26)$$

комплекснозначную функцию спектральной плотности $X(\omega)$. Свойства спектральной плотности подробно рассмотрены в главе 2. Получены спектральные плотности для интегрируемых и не удовлетворяющих условиям абсолютной интегрируемости функций, для периодических и комплексных сигналов. Показано, что квадрат модуля функции $X(\omega)$ можно интерпретировать как плотность энергии сигнала $x(t)$, приходящуюся на частоту ω . Действительно, если принять, что $x^2(t)$ – мощность сигнала² в момент времени t , то в соответствии с равенством Парсеваля

$$\int_{-\infty}^{\infty} x^2(t) dt = \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \quad (4.27)$$

значение $\int_{-\infty}^{\infty} |X(\omega)|^2 d\omega$ представляет собой полную энергию сигнала \mathbb{E} , а $|X(\omega)|^2 = d\mathbb{E}/d\omega$.

Необходимо отметить, что формула (4.26) пригодна лишь для теоретического анализа простых и детерминированных сигналов, но никак не может служить средством для практического построения спектров по следующим причинам:

- на практике невозможно наблюдать за сигналом сколь угодно долго, любые измерения являются ограниченными во времени;

²На самом деле, если $x(t)$ – напряжение или ток в какой-либо электрической цепи, то $x^2(t)$ – величина пропорциональная мощности с коэффициентом пропорциональности в виде проводимости или сопротивления.

- при расчёте спектров по формуле (4.26) даже для сигналов конечной длительности необходимы значения сигнала в бесконечном числе точек t , т.е. $x(t)$ должна задаваться на континуальном множестве моментов времени;
- если сигнал $x(t)$ является случайным (а именно случайные сигналы представляют интерес для анализа), то и $X(\omega)$ оказывается случайной функцией частоты, и необходимы специальные методы усреднения спектров, рассмотренные в главе 3.

Итак, непосредственно использовать (4.26) для анализа реальных сигналов невозможно, однако $X(\omega)$ может служить исходной величиной для построения всех других спектральных характеристик, применяемых на практике.

Как показано в предыдущих разделах главы, при дискретизации аналоговых сигналов спектр становится периодическим, но этого недостаточно для решения практических задач. Как и в случае аналоговых сигналов, дискретные сигналы было бы желательно представить в виде разложений по дискретным базисным функциям, т.е. ввести дискретные спектральные преобразования. В настоящее время в качестве базисных функций получили широкое распространение дискретные экспоненциальные функции, а также дискретные функции Уолша и Хаара. Рассмотрим дискретные преобразования Фурье.

4.4.1. Дискретное преобразование Фурье

Начнём рассмотрение с преобразования «идеализированных сигналов», заданных на всей временной оси ($t \in (-\infty, \infty)$) и равномерно дискретизованных с интервалом Δt (так называемое дискретно-временное преобразование), а затем займёмся спектральным преобразованием ограниченных во времени сигналов.

Дискретно-временное преобразование. Пусть значения дискретного сигнала заданы в моменты времени $t_n = n\Delta t$:

$$x(n) = x_d(n\Delta t), \quad n = 0, \pm 1, \pm 2, \pm 3, \dots$$

Если интервал дискретизации достаточно мал, то интеграл в (4.26) можно заменить суммой:

$$X(\omega) \simeq \sum_{n=-\infty}^{\infty} x_d(n\Delta t) e^{-j\omega n\Delta t} \Delta t.$$

Знак приближенного равенства трансформируется в строгое равенство, если Δt устремить к нулю.

Совершим переход от «обычной» частоты ω к безразмерной частоте $\bar{\omega}$, нормировав ω на частоту дискретизации $\omega_d = 2\pi/\Delta t$, $\bar{\omega} = 2\pi\omega/\omega_d = \omega\Delta t$, и введем новую комплексную функцию $F(\bar{\omega}) = X(\bar{\omega})/\Delta t$. Тогда функцию

$$F(\bar{\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\bar{\omega}n} \quad (4.28)$$

можно назвать *комплексной спектральной плотностью дискретного сигнала* $x_d(t)$, а формулу (4.28) – «*дискретно-временным преобразованием Фурье*» (ДВПФ).

Из (4.28) легко вывести основные свойства ДВПФ.

1. Линейность:

$$F_{\sum_i \alpha_i x_i}(\bar{\omega}) = \sum_i \alpha_i F_{x_i}(\bar{\omega})$$

где α_i – любые постоянные числа.

2. Симметричность относительно нулевой частоты (для вещественных сигналов):

$$F(-\bar{\omega}) = F^*(\bar{\omega}).$$

3. При смещении сигнала на постоянный интервал времени его амплитудный спектр $|F(\bar{\omega})|$, который в основном и интересует исследователей, инвариантен к выбору начального момента времени, а фазовый – инвариантен с точностью до любой линейной функции частоты:

$$F_{x(n-N)}(\bar{\omega}) = F_{x(n)}(\bar{\omega})e^{-jN\bar{\omega}}.$$

4. Периодичность с периодом 2π : $F(\bar{\omega} + 2\pi) = F(\bar{\omega})$.

5. Симметрия относительно частоты $\bar{\omega} = \pi$ (только для вещественных сигналов): $F(\pi - \bar{\omega}) = F^*(\pi + \bar{\omega})$.

Свойства (1) – (3) выполняются для ИПФ и ДВПФ, а вот свойства (4), (5) – только для ДВПФ. Из них следует, что нет смысла измерять спектр во всем диапазоне частот, а достаточно ограничиться интервалом $\bar{\omega} \in [0 : \pi]$, поскольку для остальных значений его легко получить при помощи свойств симметрии и периодического продолжения.

Таким образом, *любые методы цифрового спектрального анализа дают значения спектральных характеристик только в диапазоне нормированных частот $\bar{\omega} \in [0 : \pi]$* . При этом нормированной частоте $\bar{\omega} = \pi$ соответствует «реальная» частота, равная половине частоты дискретизации $\omega = \omega_d/2$.

Выражение (4.28), как и интегральное преобразование Фурье, трудно использовать для построения спектров экспериментальных сигналов.

Дискретное и обратное дискретное преобразования Фурье. Обратимся к реальным дискретным сигналам, обладающим ограниченной длительностью T во времени и конечным числом отсчётов $n \in [0, N - 1]$, где $N = T/\Delta t$. По теореме Котельникова (см. разд. 4.2, формулу (4.20)) непрерывный спектр ограниченного по времени сигнала может быть точно восстановлен по значениям спектральной плотности в дискретном числе точек $\omega_k = k\omega_T$, где $\omega_T = 2\pi/T$. Поэтому на практике достаточно построить спектр в этих точках: $X_k = X(k\omega_T)$.

С этой (прикладной) точки зрения будем говорить, что спектр ограниченных во времени сигналов является дискретным. Расстояние по частоте между двумя соседними гармониками спектра равно ω_T , поэтому эту частоту будем называть *частотой разрешения* спектра.

Спектр произвольных дискретных сигналов ограничен сверху частотой Найквиста $\omega_d = 2\pi/\Delta t$. Поэтому спектр сигнала конечной длительности будет содержать ограниченное число гармоник на частотах: $0, \omega_T, 2\omega_T, \dots, (N-1)\omega_T$. Действительно, если мы разделим частотный интервал $2\pi/\Delta t$ на частоту разрешения $2\pi/T = 2\pi/N\Delta t$, то получим, что спектр содержит N гармоник.

Для нахождения этих гармоник, воспользуемся формулой (4.26) и определим значение спектральной плотности на частоте $\omega_k = k\omega_T$:

$$X(k\omega_T) = \int_0^T x(t)e^{-jk\omega_T t} dt.$$

Заменяя интеграл интегральной суммой, а переменную t на $n\Delta t$, $T = N\Delta t$, и считая интервал Δt достаточно малым, получим

$$X(k\omega_T) \simeq \Delta t \sum_{n=0}^{N-1} x(n\Delta t)e^{-jk\omega_T n\Delta t}.$$

Заменив знак приближенного равенства на строгое равенство (т.е. считая интервал $\Delta t \ll T$ и обозначая $X(k\omega_T)/\Delta t$ через X_k , получим выражение:

$$X_k = \sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N}, \quad (4.29)$$

которое используется для расчёта спектров дискретных и ограниченных во времени сигналов и называется *дискретное преобразование Фурье* (ДПФ). Иногда его также называют дискретным рядом Фурье. Величина $X(k)$ называется амплитудой k -ой гармоники спектра. Индексу k соответствует физическая частота $\omega_k = 2\pi k/N\Delta t$ или нормированная частота $\bar{\omega}_k = 2\pi k/N$. Набор величин X_k , $k=0, 1, \dots, N-1$ будем в дальнейшем называть комплексным спектром дискретного, ограниченного во времени сигнала $x(n)$. Модули этих величин $|X_k|$ образуют его амплитудный, а аргументы $\varphi_k = \text{arctg} \left(\frac{\text{Im} X_k}{\text{Re} X_k} \right)$ – фазовый спектр.

Дискретное преобразование Фурье позволяет восстанавливать дискретизированные аналоговые сигналы, имеющие ограниченную полосу частот. Для этого используем формулу (2.27) обратного преобразования Фурье для непрерывных функций

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

и полученный дискретный спектр $X(k\omega_T)$. Поскольку рассматриваемый спектр содержит конечное число гармоник, можно заменить интеграл интегральной суммой. Заменяя переменную t на $n\Delta t$ и $d\omega$ на $\omega_T = 2\pi/N\Delta t$ (интервал между гармониками спектра предполагается малым), получим

$$x(n\Delta t) \simeq \frac{1}{N\Delta t} \sum_{k=0}^{N-1} X(k\omega_T)e^{jk\omega_T n\Delta t} = \frac{1}{N\Delta t} \sum_{k=0}^{N-1} X(k\omega_T)e^{j2\pi kn/N}.$$

Вводя вместо приближенного строгое равенство и обозначая $\Delta tx(n\Delta t)$ через $x_n = x(n)$, получим *обратное дискретное преобразование Фурье* (ОДПФ) в виде:

$$x_n = x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{j2\pi kn/N}. \quad (4.30)$$

Формулы (4.29) и (4.30)³ представляют разложения сигналов $x(n)$ и $X(k)$ по системе дискретных экспоненциальных функций и выражают взаимно однозначное соответствие между последовательностями отсчетов дискретного сигнала $x(n)$ и его спектра X_k .

Функции $W_N^{kn} = e^{j2\pi kn/N}$ являются ортогональными:

$$\sum_{i=0}^{N-1} W_N^{ki} W_N^{ni} = \begin{cases} N, & (k-n) = mN, m - \text{целое число,} \\ 0, & \text{в других случаях} \end{cases}$$

и периодическими по переменной n и переменной k с периодом N :

$$W_N^{kn} = W_N^{(k+N)n} = W_N^{k(n+N)}.$$

Дискретное преобразование Фурье является линейным преобразованием вектор временных отсчетов в вектор спектральных отсчетов такой же длины. Поэтому связь этих векторов можно представить матричным уравнением

$$\mathbf{X}_k = \|\mathbf{W}\| \mathbf{x}_n, \text{ где } \|\mathbf{W}\| = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-j\frac{2\pi}{N}} & e^{-j\frac{4\pi}{N}} & \dots & e^{-j\frac{2\pi}{N}(N-1)} \\ 1 & e^{-j\frac{4\pi}{N}} & e^{-j\frac{8\pi}{N}} & \dots & e^{-j\frac{2\pi}{N}2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\frac{2\pi}{N}(N-1)} & e^{-j\frac{4\pi}{N}(N-1)} & \dots & e^{-j\frac{2\pi}{N}(N-1)^2} \end{bmatrix} \quad (4.31)$$

– *матрица преобразования*. Элемент k -й строки и n -го столбца матрицы $\mathbf{W}(k, n) = e^{-j\frac{2\pi}{N}(k-1)(n-1)}$, $1 \leq k \leq N$, $1 \leq n \leq N$.

Рассмотрим основные свойства ДПФ. Легко увидеть, что многие из них повторяют свойства ДВПФ. Из свойств дискретных экспоненциальных функций следуют:

1. **Линейность:** если последовательностям $x_i(n)$ с одинаковым периодом N соответствуют спектры $X_i(k)$, то последовательности $x(n) = \sum_i a_i x_i(n)$, где a_i – постоянные числа, отвечает спектр $X(k) = \sum_i a_i X_i(k)$.

2. **Симметричность относительно нулевой частоты** (для вещественных сигналов):

$$X(-k) = X^*(k).$$

Как видно из (4.29), постоянная составляющая спектра $X(0)$ является суммой отсчетов $x(n)$ на одном периоде

$$X(0) = \sum_{n=0}^{N-1} x(n).$$

³Множитель $1/N$ в некоторых источниках фигурирует в формуле ДПФ и отсутствует в формуле ОДПФ.

3. Преобразование ДПФ при смещении сигнала на постоянный интервал времени:

$$X_{x(n-n_0)}(k) = X_{x(n)}(k)e^{-j2\pi kn_0/N}.$$

4. Симметрия относительно гармоники с номером $N/2$ (только для вещественных сигналов и чётного N): $X(\frac{N}{2} + k) = X^*(\frac{N}{2} - k)$.

Отсюда следует, что для вещественных сигналов достаточно построить $N/2$ гармоник, а оставшиеся $N/2$ получаются из них с помощью свойства симметрии. При этом и спектр, и сигнал содержат одинаковое количество информации, так как сигнал представляет последовательность N вещественных чисел, а спектр – последовательность $N/2$ комплексных чисел, каждое из которых содержит два вещественных числа (действительную и мнимую часть). Поэтому большинство методов цифрового спектрального анализа по сигналу из N отсчётов определяет $N/2$ гармонических составляющих спектра.

5. По аналогии с обычной свёрткой двух аналоговых сигналов (см. разд. 2.4, формула (2.29)) можно ввести *линейную свёртку* дискретных сигналов $\{x(k)\}$ и $\{y(m)\}$ как вещественный дискретный сигнал $\{s(m)\}$, отсчёты которого удовлетворяют соотношению:

$$s(m) = \sum_{k=0}^{N-1} x(k)y(m-k), \quad m = 0, 1, \dots, N-1.$$

При этом дискретный спектр свёртки является произведением спектров дискретных сигналов

$$S(k) = X(k) \cdot Y(k).$$

Явление «растекания спектра». Рассмотрим теперь проблему построения спектра для ограниченного во времени сигнала с другой стороны. На практике мы обычно сталкиваемся с ситуацией, когда исследуемый сигнал $x(t)$ доступен для анализа в течение ограниченного интервала времени T , меньшего, чем длительность самого сигнала. Данный интервал принято называть *временным окном*. Как выглядит сигнал за пределами этого окна, мы не знаем. Очевидно, что введение временного окна влияет и на результаты спектрального анализа. Рассмотрим эту проблему подробнее.

Будем использовать простой гармонический сигнал $x(t) = \cos \Omega t$ (рис. 4.10, а), функция спектральной плотности которого, как известно (см. разд. 2.4.3), представляет собой сумму двух дельта-функций Дирака:

$$X(\omega) = \pi[\delta(\omega - \Omega) + \delta(\omega + \Omega)].$$

Пусть временное окно, в течение которого сигнал доступен для наблюдения, имеет длительность T . Для упрощения выкладок разместим его симметрично относительно нулевого момента времени и введем функцию прямоугольного временного окна $W(t)$, равную единице на интервале наблюдения и нулю вне его,

$$W(t) = \begin{cases} 1, & \text{если } t \in [-T/2 : T/2], \\ 0, & \text{если } t \notin [-T/2 : T/2]. \end{cases} \quad (4.32)$$

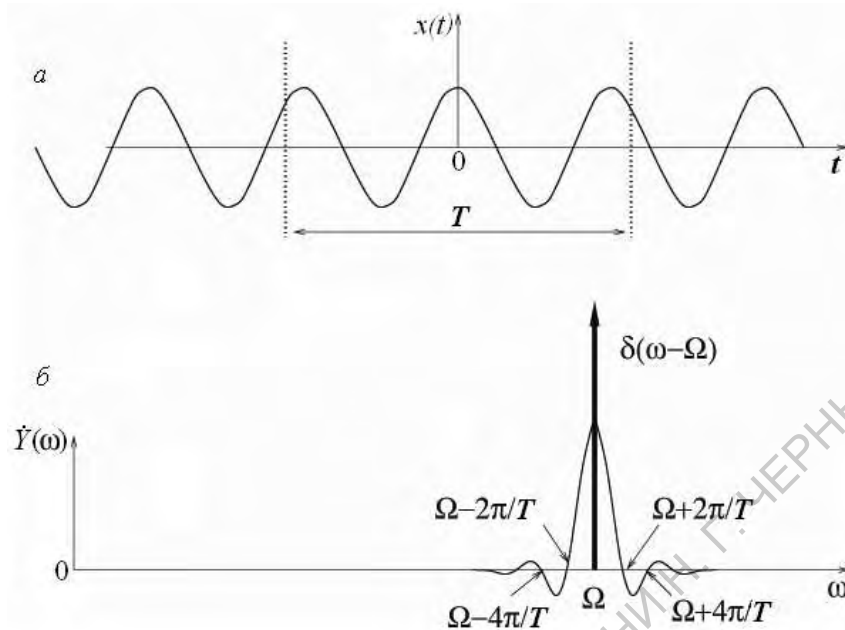


Рис. 4.10. Гармонический сигнал во временном окне (а) и его функция спектральной плотности (б)

При этом наблюдаемый сигнал $y(t)$ является произведением исходного сигнала $x(t)$ и функции окна: $y(t) = x(t)W(t)$, а его спектральная плотность $Y(\omega)$ удовлетворяет соотношению:

$$Y(\omega) = \int_{-\infty}^{\infty} y(t)e^{-j\omega t} dt = \int_{-T/2}^{T/2} x(t)e^{-j\omega t} dt = \pi \frac{T}{2} \left[\text{sinc} \left((\omega - \Omega) \frac{T}{2} \right) + \text{sinc} \left((\omega + \Omega) \frac{T}{2} \right) \right]. \quad (4.33)$$

Так как спектр симметричен относительно нулевой частоты, ограничимся областью положительных частот: $Y^+(\omega) = \pi \frac{T}{2} \text{sinc} \left((\omega - \Omega) \frac{T}{2} \right)$. Характерная форма спектра показана на рис. 4.10, б: максимальное значение спектральной плотности достигается на частоте Ω . «Горб» функции $Y(\omega)$ на этой частоте называется «основным лепестком». Основной лепесток ограничен справа и слева «нулями» функции спектральной плотности на частотах $\Omega + 2\pi/T$ и $\Omega - 2\pi/T$. Амплитуда основного лепестка прямо пропорциональна длительности временного окна T . Кроме основного лепестка спектр содержит «боковые лепестки», которые прижимаются к основному. Их ширина равна $2\pi/T$, а амплитуды быстро спадают по отношению к амплитуде основного лепестка с ростом номера: $2/3\pi$, $2/5\pi$, $2/7\pi$, ..., $2/(2m+1)\pi$, ... (m – номер лепестка).

При неограниченном увеличении длительности временного окна ($T \rightarrow \infty$) амплитуда основного лепестка стремится к бесконечности, а его ширина – к нулю. Спектральная плотность становится все больше и больше похожа на дельта-функцию: $Y^+(\omega) \rightarrow \delta(\omega - \Omega)$.

Если аналоговый сигнал является полигармоническим, например периодическим, его спектральная плотность (см. гл. 2, формула (2.43)) с

учетом функции прямоугольного временного окна $W(t)$ принимает вид:

$$Y(\omega) = \int_{-\infty}^{\infty} W(t) \left[\sum_{n=-\infty}^{\infty} \dot{C}_n \cos n\Omega t \right] e^{-j\omega t} dt = \sum_{n=-\infty}^{\infty} \dot{C}_n \int_{-T/2}^{T/2} \cos n\Omega t e^{-j\omega t} dt = \\ = \pi \frac{T}{2} \sum_{n=-\infty}^{\infty} \dot{C}_n \left[\operatorname{sinc} \left((\omega - n\Omega) \frac{T}{2} \right) + \operatorname{sinc} \left((\omega + n\Omega) \frac{T}{2} \right) \right]. \quad (4.34)$$

Полученный спектр, как и в случае простого косинусоидального сигнала, симметричен относительно нулевой частоты. Но теперь в области положительных частот, как и в области отрицательных частот, он имеет вид суперпозиции функций $\operatorname{sinc}(\dots)$ с основными лепестками, расположенными на частотах $\pm k\Omega$. В зависимости от частоты Ω и протяжённости временного окна функции $\operatorname{sinc}(\dots)$ в большей или меньшей степени пересекаются друг с другом. При неограниченном увеличении длительности временного окна $\operatorname{sinc}(\dots)$ сжимаются, амплитуды лепестков стремятся к бесконечности, а спектральная плотность приближается к (2.43).

Итак, введение временного окна, ограничивающего длительность наблюдений, приводит к существенному искажению спектров даже простых аналоговых сигналов: спектральные линии трансформируются в многолепестковые структуры типа $\operatorname{sinc}(\dots)$, но новые линии не появляются. А что же происходит со спектрами дискретных сигналов?

Будем исследовать влияние временного окна на спектры дискретных сигналов на примере все того же гармонического сигнала. При дискретном преобразовании Фурье последовательность отсчётов сигнала $x(t)$ периодически продолжалась по времени с периодом, равным длительности временного окна. Полученный спектр представлял собой набор гармоник на частотах $\omega_k = 2\pi k/T$, $k = 0, 1, 2, \dots$, амплитуды которых были пропорциональны значениям спектральной плотности $Y(\omega)$.

Теперь при исследовании влияния характеристик временного окна следует различать два характерных случая.

1. Длительность временного окна кратна периоду сигнала

$$T = m \frac{2\pi}{\Omega} = \frac{m}{F}, \quad F = \frac{\Omega}{2\pi}, \quad (4.35)$$

где m – целое число. Из (4.35) следует, что гармоника спектра с номером $k = m$ совпадает с частотой F периодического сигнала: $f_m = m/T = F$ (рис. 4.8, а, з). Все же остальные гармоники, соответствующие частотам $f_{m\pm 1} = F \pm 1/T$, $f_{m\pm 2} = F \pm 2/T$, $f_{m\pm 3} = F \pm 3/T$, ..., имеют нулевые амплитуды, так как попадают в «нули» функции спектральной плотности. Таким образом, спектр сигнала в данном случае содержит единственную гармонику на частоте F и качественно совпадает со спектром исходного сигнала $x(t)$.

2. Длительность временного окна не равна целому числу периодов сигнала $x(t)$, но соответствует целому числу периодов другой частоты $F + \Delta F$, где $\Delta F \ll F$, т.е.

$$T = \frac{m}{F + \Delta F}, \quad (4.36)$$

где m – целое число. В этом случае ни одна из гармоник f_k не совпадает с частотой сигнала F . Ближайшая к ней гармоника с номером m соответствует частоте $F + \Delta F$, а остальные – $f_{m\pm 1} = F + \Delta F \pm 1/T$, $f_{m\pm 2} = F + \Delta F \pm 2/T$, $f_{m\pm 3} = F + \Delta F \pm 3/T$, ... также «промахиваются» мимо нулей функции спектральной плотности на ту же величину ΔF . Поэтому спектр анализируемого сигнала кроме основной гармоники на частоте $F + \Delta F$ будет содержать бесконечное множество дополнительных (боковых) гармоник на частотах $F + \Delta F \pm k/T$ (рис. 4.11, б, в).

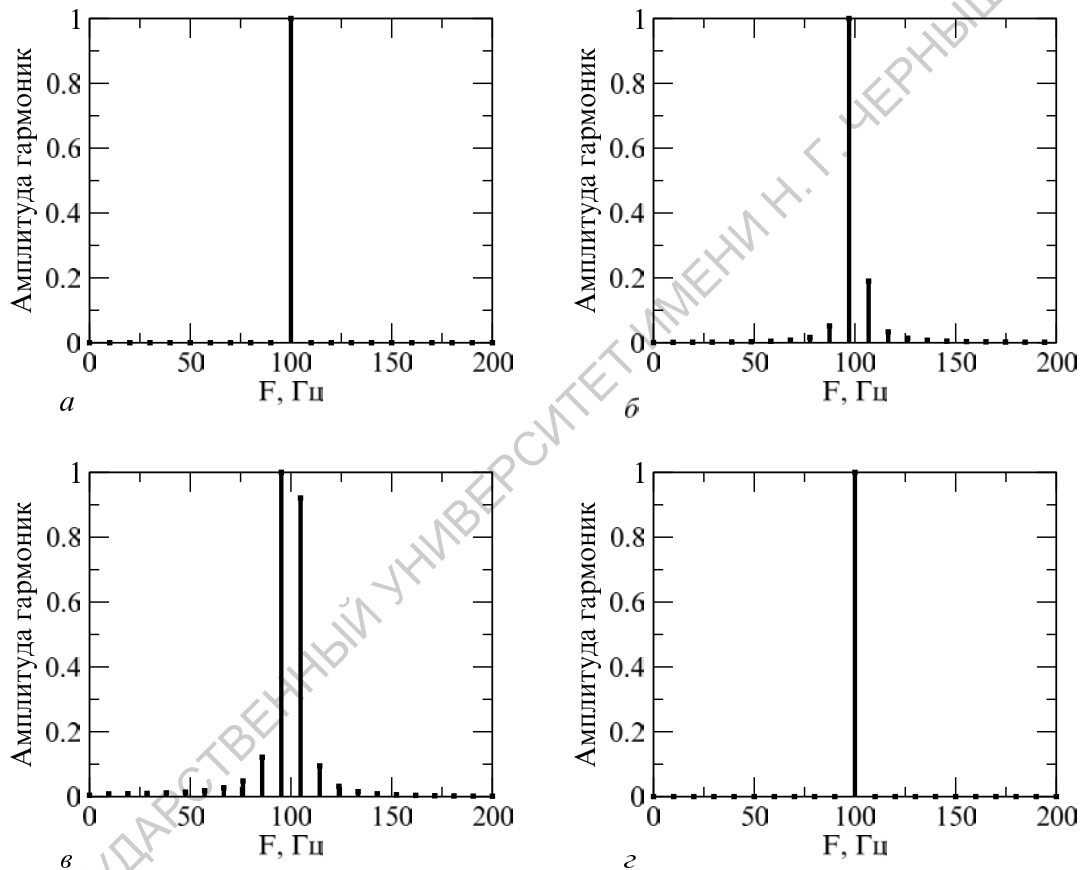


Рис. 4.11. Влияние длительности временного окна T на спектр сигнала. T кратна периоду сигнала: $T=0.1$ с, $F=100$ Гц, $m=10$ (а); T соответствует частоте 110 Гц, $m=11$ (г); T не кратна периоду сигнала: $T \cong 0.097$ с соответствует частоте 103 Гц (б), $T \cong 0.095$ с – частоте 105 Гц (в)

Итак, дискретный спектр одного и того же сигнала кардинальным образом меняется при небольшом изменении длительности временного окна. Оптимальным является такое временное окно, длительность которого составляет целое число периодов сигнала. В этом случае окно не искажает спектр (см. рис. 4.11, а). Отчего так происходит? Это легко понять, если вспомнить, что дискретный спектр строится по периодическому продолжению сигнала. Легко увидеть (см. рис. 4.11, г), что если период этого продолжения составляет целое число периодов сигнала ($m=11$), то продолженный сигнал совпадает с истинным, если же нет, то на граничных точках периодического продолжения возникают разрывы. В результате сигнал $y(t)$ более не является копией сигнала $x(t)$, а значит, и вид его спектра будет отличаться от спектра исходного сигнала:

1) частота основной гармоники спектра окажется сдвинутой относительно «истинной» частоты сигнала;

2) спектр будет содержать дополнительные гармонические составляющие, отсутствовавшие в спектре исходного сигнала (боковые лепестки).

Искажение спектра, вызванное ограниченностью временного окна, называют *растеканием спектра*. Причиной растекания спектра является совпадение исходного сигнала с его периодическим продолжением.

Явление растекания спектра мешает проведению спектральных измерений. Один и тот же сигнал может давать существенно разные спектры при разных параметрах расчёта. Как бороться с этим явлением? Ясно, что увеличение длительности временного окна уменьшает растекание, однако не всегда есть возможность наблюдать за сигналом сколь угодно долго. Если длительность окна увеличить нельзя, то можно устранить растекание, подобрав его длительность кратной периоду сигнала. Беда заключается в том, что основную массу образуют непериодические сигналы, и полностью устранить растекание не удастся. Его можно уменьшить, если у сигнала есть некоторый «характерный период» T' , т.е. интервал времени, через который он почти повторяется: $x(t + T') \simeq x(t)$. Для этого следует выбрать время наблюдения кратное T' .

Если исследуемый сигнал – шумовой без ясно выраженного периода колебаний, то растекание сигналов короткой длительности является неустранимым явлением. При этом, как уже было сказано выше, в спектре «укороченного» сигнала появляется множество дополнительных гармоник, отсутствовавших в спектре исходного. Часть этих гармоник (одна или две) располагаются в основной лепестке, часть – формируют боковые лепестки (по одной гармонике на каждый лепесток). Основной и боковые лепестки по-разному могут влиять на результат измерений. Например, широкий основной лепесток нежелателен, если нужно различить в спектре два периодических сигнала с близкими частотами и почти равными амплитудами, так как перекрытие основных лепестков помешает их разделению. Заметим, что наличие боковых лепестков в данном случае не столь важно.

Если же необходимо различить слабый сигнал на фоне сильного, частоты которых сильно различаются, то существенно «нейтрализовать» его боковые лепестки, так как спектральные составляющие слабого сигнала легко могут быть замаскированы ими.

Существует метод «перераспределения» растекания спектра между основным и боковыми лепестками. Этот метод называется «выбором формы временного окна» и заключается в специальном подборе функции $W(t)$. Временное окно простейшей формы (4.32), о котором говорилось выше, называется «прямоугольным окном». Кроме него возможны треугольное окно, окно Гаусса и другие. Каждое из этих окон характеризуется своими характеристиками: шириной основного лепестка, высотой боковых лепестков относительно основного, скоростью спадания высоты боковых лепестков с увеличением их номера. Для указанных окон эти характеристики приведены в табл. 4.1 (функция прямоугольного окна

$$w(n) = \begin{cases} 1, & n \in [0, N - 1], \\ 0, & n \notin [0, N - 1], \end{cases} \quad \alpha(n) = \left| \frac{n - (N - 1)/2}{N - 1} \right|.$$

Таблица 4.1

Характеристики временных окон

Вид окна	Прямоугольное	Треугольное	Гаусса
Вид функции $W(n)$	$w(n)$	$(1-2\alpha(n))w(n)$	$\exp[-12.5(1-2\alpha(n))^2]$
Ширина главного лепестка (на уровне половинной мощности, измеряется числом гармоник ДПФ)	0.89	1.28	1.33
Относительная высота боковых лепестков (дБ)	-13.3	-26.5	-42
Скорость спада боковых лепестков (дБ)	-6	-12	-6

Как видно из таблицы, ни одно из окон не дает возможность улучшить все расчётные характеристики, а лишь улучшает одни из них за счёт ухудшения других. Это правило является универсальным: выбор формы временного окна не позволяет уменьшить явление растекания спектра, а лишь улучшает одну из характеристик за счёт ухудшения других.

4.4.2. Быстрое преобразование Фурье

Дискретное преобразование Фурье позволяет по заданной последовательности отсчётов сигнала $x(t)$ вычислять его спектральные гармоники. Если исходная последовательность в соответствии с формулой (4.30) состоит из N комплексных отсчётов, то для вычисления ДПФ необходимо выполнить $4N^2$ умножений и $N(4N-2)$ сложений вещественных чисел или N^2 умножений и $N(N-4)$ сложений комплексных чисел. Действительно, представляя комплексный отсчёт как $x(n) = a_x + jb_x$, а комплексную функцию преобразования как $W_N^{kn} = a_W + jb_W$ и используя формулу (4.29), получим

$$X(k) = \sum_{n=0}^{N-1} [(a_x a_W - b_x b_W) + j(a_x b_W + b_x a_W)], \quad k = 0, 1, 2, \dots, N-1, \quad (4.37)$$

т.е. для вычисления каждого X_k выполняется $4N$ операций умножения. Уже при $N = 2^{10}$ (всего 1024 отсчёта) к быстродействию микропроцессора предъявляются настолько высокие требования, что проведение спектрального анализа в реальном времени становится невозможным. По-видимому, проведение спектрального анализа в реальном времени и инициировало развития многочисленных специальных методов дискретного преобразования Фурье, получивших название *быстрого преобразования Фурье* (БПФ).

Быстрым преобразованием Фурье называют совокупность алгоритмов многократного сокращения числа операций, основанных на симметрии и периодичности коэффициентов матрицы преобразований в формулах ДПФ. Для нас важно не анализировать различные алгоритмы БПФ, а понять основные принципы ускорения цифровых вычислений коэффициентов ДПФ.

Алгоритмы БПФ основаны на линейности формул ДПФ и использовании свойств симметрии и периодичности функций $W_N^{kn} = e^{\pm j2\pi kn/N}$. Когда умножение по формуле (4.31) выполняется непосредственно, вычисления происходят в полном объёме. Если же размер N матрицы преобразования $\|\mathbf{W}\|$ является составным числом, то матрица может быть факторизована, т.е. представлена произведением более простых матриц, содержащих большое количество нулей. Объём вычислений при этом заметно сокращается.

Широко известен и другой метод, основанный на представлении одномерной последовательности отсчётов сигнала двумерным массивом. Частным случаем этого метода являются алгоритмы с прореживанием по времени и по частоте. Эти методы наиболее просты и удобны для выявления сущности БПФ.

Алгоритмы прореживания для $N = 8$ иллюстрирует рис. 4.12. При прореживании по времени (рис. 4.12, а) из N -отсчётной последовательности формируются две $N/2$ -точечные подпоследовательности, состоящие из членов с чётными и нечётными номерами. При прореживании по частоте одну подпоследовательность образуют первые $N/2$ отсчётов, а другую – остальные отсчёты (рис. 4.12, б). В нижней части рис. 4.12 показано, как осуществляется дальнейшее прореживание с образованием двух $N/4$ -точечных подпоследовательностей. Процесс дробления последовательности отсчётов продолжается до тех пор, пока каждая из подпоследовательностей будет содержать только два отсчёта.

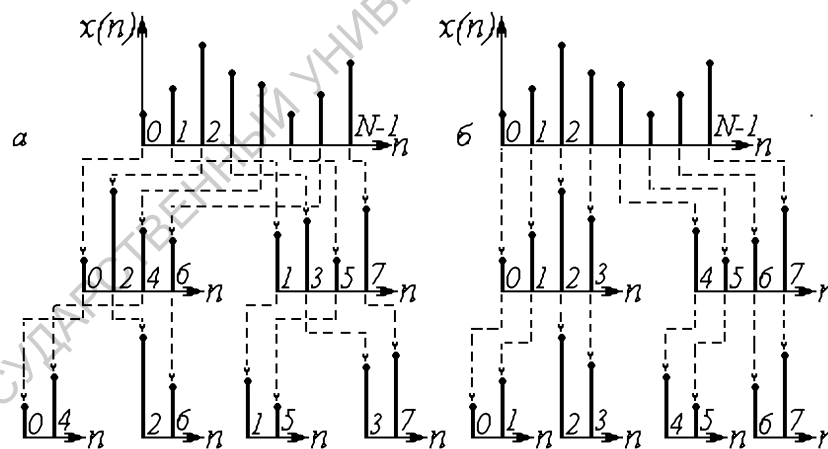


Рис. 4.12. Алгоритмы прореживания сигналов по времени (а) и по частоте (б)

БПФ с прореживанием по времени. Пусть последовательность отсчётов дискретного сигнала $\{x(n)\}$ содержит $N = 2^m$ отсчётов, где m – целое число. Разобьём её на две подпоследовательности, $\{x_1(n)\}$ и $\{x_2(n)\}$, состоящие из членов с чётными и нечётными номерами, т.е. $x_1(n) = x(2n)$, $n = 0, 1, 2, \dots, \frac{N}{2} - 1$; $x_2(n) = x(2n + 1)$, $n = 0, 1, 2, \dots, \frac{N}{2} - 1$. Подпоследовательности $x_1(n)$ и $x_2(n)$ получены из $x(n)$ с пропуском через один отсчёт, что и соответствует *прореживанию по времени*.

Используем формулу (4.29) и представим X_k в виде суммы двух слагаемых, соответствующих последовательностям отсчётов $x_1(n)$ и $x_2(n)$:

$$\begin{aligned}
X_k &= \sum_{m=0}^{N/2-1} x(2m)e^{-j2\pi 2km/N} + \sum_{m=0}^{N/2-1} x(2m+1)e^{-j2\pi k(2m+1)/N} = \\
&= \sum_{m=0}^{N/2-1} x_1(m)e^{-j2\pi km/(N/2)} + e^{-j2\pi k/N} \sum_{m=0}^{N/2-1} x_2(m)e^{-j2\pi km/(N/2)}. \quad (4.38)
\end{aligned}$$

Это сумма дискретных преобразований последовательностей отсчётов с чётными и нечётными номерами для $0 \leq k \leq N/2 - 1$, т.е.

$$X(k) = X_1(k) + X_2(k)e^{-j2\pi k/N}. \quad (4.39)$$

Коэффициенты $X_1(k)$ и $X_2(k)$ определены только для $0 \leq k \leq N/2 - 1$. Для определения $X(k)$ при $N/2 \leq k \leq N - 1$ воспользуемся следующими соображениями. ДПФ $X_1(k)$ и $X_2(k)$ при $N/2 \leq k \leq N - 1$ периодически повторяют коэффициенты, найденные для $0 \leq k \leq N/2 - 1$. Действительно, $x_1(m)$ и $x_2(m)$ от k не зависят, а $W_{N/2}^{km} = e^{-j2\pi km/(N/2)}$ – периодическая по k последовательность с периодом $N/2$. Поэтому для $0 \leq k \leq N/2 - 1$ имеем:

$$\begin{aligned}
X(k) &= X_1(k - N/2) + X_2(k - N/2)e^{-j2\pi k/N} = \\
&= X_1(k - N/2) - X_2(k - N/2)e^{-j2\pi(k-N/2)/N}, \quad (4.40)
\end{aligned}$$

так как $-e^{-j2\pi(k-N/2)/N} = -e^{-j2\pi k/N}e^{-j2\pi(-N/2)/N} = -e^{-j2\pi k/N}e^{-j2\pi} = e^{-j2\pi k/N}$.

Итак, для вычисления спектральных коэффициентов X_k получен следующий алгоритм:

$$X(k) = \begin{cases} X_1(k) + X_2(k)W_N^k, & 0 \leq k \leq N/2 - 1, \\ X_1(k - N/2) - X_2(k - N/2)W_N^{k-N/2}, & N/2 \leq k \leq N - 1. \end{cases} \quad (4.41)$$

БПФ с прореживанием по частоте. В данном случае прореживания по времени не производится, а в качестве $N/2$ -точечных подпоследовательностей выбираются первые $N/2$ отсчётов (подпоследовательность $X_1(n)$) и $N/2$ отсчётов, следующих за ними (подпоследовательность $x_2(n)$).

С учетом предложенного разбиения представим преобразование Фурье дискретной последовательности отсчётов $\{x(n)\}$ в виде:

$$\begin{aligned}
X(k) &= \sum_{m=0}^{N/2-1} x(m)e^{-j2\pi km/N} + \sum_{m=N/2}^{N-1} x(m)e^{-j2\pi km/N} = \\
&= \sum_{m=0}^{N/2-1} x_1(m)e^{-j2\pi km/N} + \sum_{m=0}^{N/2-1} x_2(m)e^{-j2\pi k(m+N/2)/N} = \\
&= \sum_{m=0}^{N/2-1} [x_1(m) + x_2(m)e^{-j2\pi k}] e^{-j2\pi km/N}. \quad (4.42)
\end{aligned}$$

Рассмотрим соотношение (4.42) отдельно для чётных ($2k$) и нечётных ($2k + 1$) номеров гармоник:

$$\begin{aligned}
 X(2k) &= \sum_{m=0}^{N/2-1} [x_1(m) + x_2(m)] e^{-j2\pi 2km/N} = \\
 &= \sum_{m=0}^{N/2-1} [x_1(m) + x_2(m)] e^{-j2\pi km/(N/2)}, \quad k = 0, 1, \dots, \frac{N}{2} - 1, \\
 X(2k + 1) &= \sum_{m=0}^{N/2-1} [(x_1(m) - x_2(m)) e^{-j2\pi m/N}] e^{-j2\pi 2km/N} = \\
 &= \sum_{m=0}^{N/2-1} [(x_1(m) - x_2(m)) e^{-j2\pi m/N}] e^{-j2\pi km/(N/2)}, \quad k = 0, 1, \dots, \frac{N}{2} - 1.
 \end{aligned}$$

Итак, для вычисления спектральных коэффициентов X_k при $k = 0, 1, \dots, \frac{N}{2} - 1$ получен следующий алгоритм:

$$X(k) \Rightarrow \begin{cases} X(2k) = \sum_{m=0}^{N/2-1} [x_1(m) + x_2(m)] e^{-j2\pi km/(N/2)}, \\ X(2k + 1) = \sum_{m=0}^{N/2-1} [(x_1(m) - x_2(m)) e^{-j2\pi m/N}] e^{-j2\pi km/(N/2)}, \end{cases} \quad (4.43)$$

т.е. имеет место прореживание по частоте, аналогичное рассмотренному выше прореживанию по времени. Полученные алгоритмы могут быть также использованы для вычисления обратного ДПФ (ОДПФ).

Формулы прямого ДПФ (4.29) и ОДПФ (4.30) имеют одинаковую структуру и отличаются знаком экспонент в функциях W_N^{kn} и коэффициентом $1/N$. В соотношении (4.30) величины $x(n)$ и $X(k)$ в общем случае комплексные. Если записать (4.30) для комплексно-сопряжённых величин, то получим:

$$\begin{aligned}
 x^*(n) &= \frac{1}{N} \sum_{k=0}^{N-1} X^*(k) [e^{j2\pi kn/N}]^* = \frac{1}{N} \sum_{k=0}^{N-1} X^*(k) e^{-j2\pi kn/N}, \quad n = 0, 1, \dots, N - 1, \\
 \text{или} \quad x(n) &= \frac{1}{N} \left[\sum_{k=0}^{N-1} X^*(k) e^{-j2\pi kn/N} \right]^*, \quad n = 0, 1, \dots, N - 1.
 \end{aligned}$$

Выражение в скобках представляет собой прямое ДПФ последовательности $X^*(k)$ и может быть вычислено как функция переменной n с помощью одного из алгоритмов БПФ.

Графы БПФ. Рассмотрим БПФ с прореживанием по времени и ограничимся 8-точечной последовательностью отсчётов. Покажем сначала, как графически в соответствии с алгоритмом (4.41) вычисляется N -точечная (8-точечная) из $N/2$ -точечных (4-точечных) подпоследовательностей $X_1(k)$ и $X_2(k)$. Как понятно из алгоритма (4.41), полная процедура вычисления БПФ является многоступенчатой (в нашем случае трёхступенчатой). На рис. 4.13 рассмотрена только одна ступень. Здесь

прямоугольниками изображены $N/2$ -точечные ДПФ для подпоследовательностей с чётными и нечётными номерами, а для $k = 0$ в соответствии с алгоритмом (4.41) на рис. 4.13, а построены $X(0)$ и $X(N/2 - k) = X(4)$.

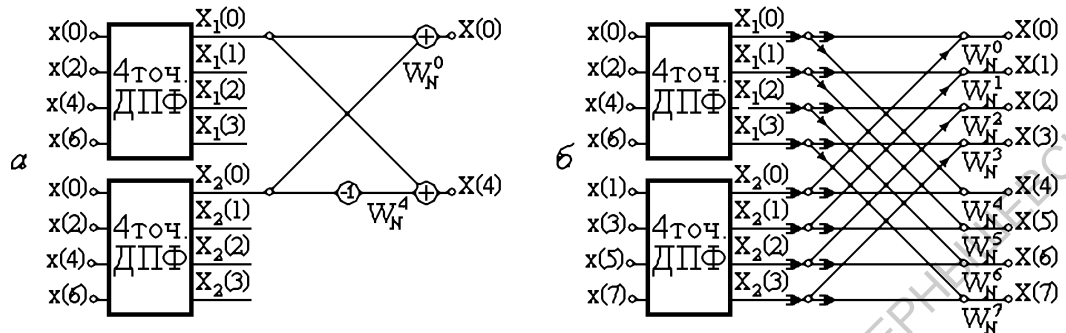


Рис. 4.13. Графическое представление процедуры получения $X(0)$ и $X(4)$ (а) и полного набора гармоник (б)

При этом графы выполняемых операции по виду напоминают повернутую на 90° «бабочку»⁴. Если внести -1 в соответствующие коэффициенты W_N , а сумматоры заменить узлами с направленными к ним ветвями, то граф ДПФ принимает вид, показанный на рис. 4.13, б. Раскрывая $N/2$ -точечные преобразования, можно изобразить трёхступенчатый граф (рис. 4.14) для полного разложения 8-точечного ДПФ. Каждая ступень графа представляет совокупность четырёх «бабочек».

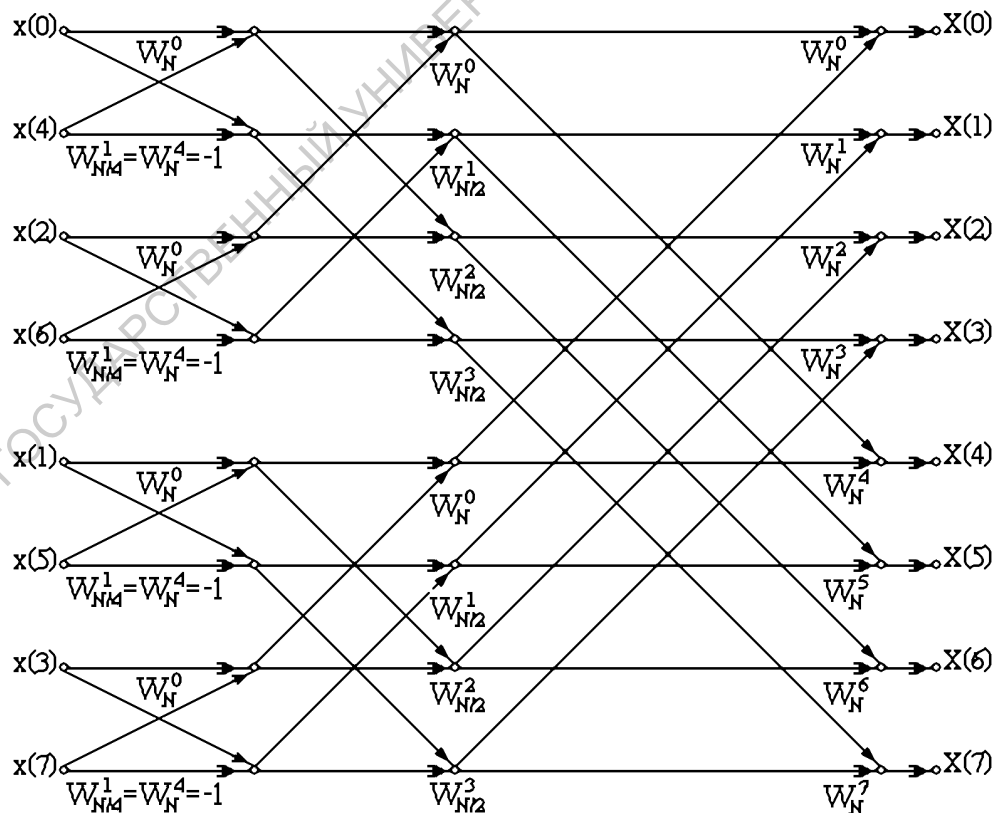


Рис. 4.14. Графическое представление полного разложения 8-точечного ДПФ с прореживанием по времени

⁴Название операция «бабочка» получило распространение в специальной литературе по цифровой обработке информации.

Начнём с первой ступени. Комплексные множители $W_{N/4}^0 = W_N^0 = 1$, $W_{N/4}^1 = W_N^4 = e^{-j2\pi 4/N} = e^{-j\pi} = -1$, т.е. при вычислении 2-точечных ДПФ операции умножения отсутствуют и выполняются только суммирование и вычитание. Теперь обратимся к операциям второй ступени. Множители $W_{N/2}^0 = W_N^0 = 1$, $W_{N/2}^2 = W_N^4 = e^{-j2\pi 2/N} = e^{-j\pi} = -1$, $W_{N/2}^1 = W_N^2 = e^{-j2\pi 2/N} = e^{-j\pi/2} = -j$, $W_{N/2}^3 = W_N^6 = e^{j\pi/2} = j$, т.е. при выполнении операций второй ступени используется единственный множитель $\pm j$, причём отрицательный знак учитывается заменой суммирования вычитанием. Наконец, на последней ступени из восьми коэффициентов остаются только три, так как $W_N^0 = 1$, $W_N^4 = -1$, а коэффициенты W_N^5 , W_N^6 , W_N^7 отличаются от W_N^1 , W_N^2 , W_N^3 только знаком.

Таким образом, при использовании БПФ общее количество комплексных умножений существенно сокращено: можно показать, что число комплексных умножений становится равным $N \log_2 N$ вместо N^2 . Кроме того, упрощается запоминание сигналов, т.е. данные, полученные на каждой ступени БПФ, засылаются в те же регистры, из которых берутся исходные данные.

Аналогичный результат получается при прореживании по частоте. Граф БПФ с прореживанием по частоте (рис. 4.15) подобен графу с прореживанием по времени. Оба они представляют последовательность ступеней, оба содержат совокупности операций типа «бабочка». Отличаются они друг от друга порядком расположения входных и выходных данных.

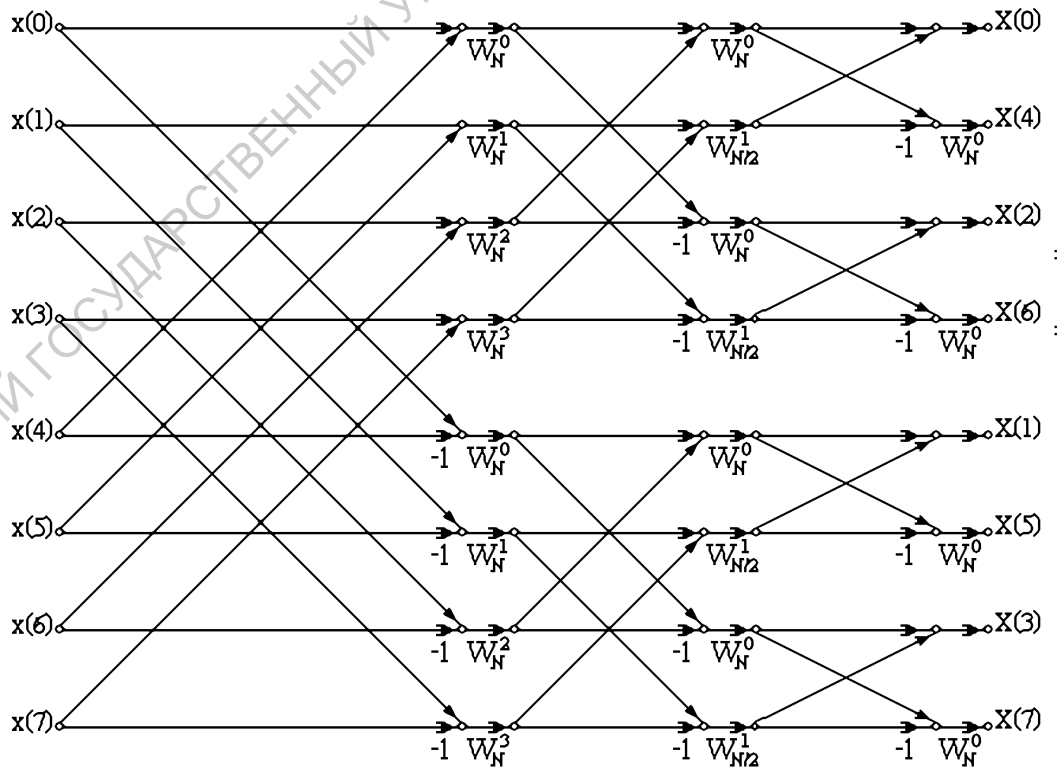


Рис. 4.15. Графическое представление полного разложения 8-точечного ДПФ с прореживанием по частоте

Если порядковые номера отсчётов входного сигнала представить в

нормальном двоичном порядке (рис. 4.15), то номера спектральных гармоник окажутся расположенными в *двоично-инверсном* порядке, когда старший разряд становится младшим и наоборот, т.е.

$$x(0) = x(000) \Rightarrow X(0) = X(000), x(1) = x(001) \Rightarrow X(4) = X(100),$$

$$x(2) = x(010) \Rightarrow X(2) = X(010), x(3) = x(011) \Rightarrow X(6) = X(110),$$

$$x(4) = x(100) \Rightarrow X(1) = X(001), x(5) = x(101) \Rightarrow X(5) = X(101),$$

$$x(6) = x(110) \Rightarrow X(3) = X(011), x(7) = x(111) \Rightarrow X(7) = X(111).$$

Если же порядковые номера отсчётов входного сигнала представлены в *двоично-инверсном* порядке, то номера спектральных гармоник окажутся расположенными в нормальном двоичном порядке (см. рис. 4.14).

В заключение отметим, что БПФ не является приближенным алгоритмом и дает такой же результат, как ДПФ. Ускорение происходит за счёт организации вычислений.

4.5. Спектр мощности дискретных случайных сигналов

Классификация и статистические характеристики случайных процессов, в том числе и дискретных случайных последовательностей, рассмотрены в предыдущей главе. В настоящем разделе предполагается подробнее рассмотреть особенности расчёта спектров мощности случайных сигналов в зависимости от объёма выборки на модели эргодических последовательностей.

Одна из трудностей, возникающих при обработке таких сигналов, состоит в том, что они не обладают конечной энергией. Неосуществим и обычно используемый подход, когда бесконечная последовательность трактуется как периодическая функция и выделяется один её период, энергия которого конечна. Примером сигналов, которые не обладают конечной энергией и не могут быть представлены как периодические, являются речевые сигналы.

Известны различные подходы к расчёту спектров мощности случайных сигналов. Рассмотрим один из них, основанный на использовании экспериментальных данных.

Пусть дано множество реализаций $x_i(n)$ дискретного случайного процесса, где i – номер реализации ($i = 1, 2, \dots, M$), n – номер отсчёта ($n = 0, 1, 2, \dots, N - 1$). Пусть все реализации имеют одинаковую длину и получены в одинаковых условиях при независимых измерениях. Для эргодических сигналов множество реализаций можно получить делением единственной длинной реализации на M более коротких.

Возьмём одну из реализаций и построим её ДПФ. Тогда для k -й гармоники спектра имеем

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n) e^{-j2\pi kn/N}.$$

Такой спектр иногда называют *периодограммой* отдельной реализации случайного процесса⁵. Так как процесс случайный, все временные реализации $x_i(n)$ различны, а значит, разными будут и соответствующие периодограммы $X_i(k)$. Функция $X_i(k)$ характеризует не спектр случайного сигнала, а лишь спектр одной его реализации и является случайной функцией частоты. При другом выборе временной реализации мы получим другую периодограмму.

Использование периодограммы $X_i(k)$ в качестве спектра случайного сигнала – характерная ошибка спектрального анализа. Для получения спектра случайного процесса необходимо провести операцию статистического усреднения по ансамблю периодограмм $\{X_i(k)\}_{i=1}^M$. Однако усреднять периодограммы нельзя, так как они не являются инвариантными к выбору начального момента времени: каждая реализация начинается после окончания предыдущей, и периодограммы в соответствии со смещением сигналов будут отличаться экспоненциальным множителем $e^{-j2\pi kn_0/N}$. Поэтому усредняют либо модули $|X_i(k)|$ (амплитудный спектр), либо, чаще всего квадраты модулей $|X_i(k)|^2$ (спектр мощности).

Рассмотрим методику построения спектра мощности $P(k)$ случайного дискретного сигнала, используя выводы статистической теории точечных оценок, согласно которой физическая величина $G(n)$ (в нашем случае $P(k)$), построенная по конечному числу выборочных точек, называется *оценкой* этой величины и обозначается как $\hat{G}(n)$. Истинная величина получается при $n \rightarrow \infty$:

$$G(n) = \lim_{n \rightarrow \infty} \hat{G}(n).$$

Оценка $\hat{G}(n)$ величины $G(n)$ называется *состоятельной*, если она сходится по вероятности $P\{\}$ к оцениваемому параметру при увеличении объёма выборки, т.е.

$$\lim_{n \rightarrow \infty} P\{|\hat{G} - G| < \varepsilon\} = 1,$$

где ε – сколь угодно малое число.

1. Пусть реализация заданной выше случайной последовательности длиной L отсчётов разделена на M более коротких подреализаций длительностью N точек, так что $L = M \times N$:

$$\left\{ \underbrace{x(0), x(1), \dots, x(N-1)}_{1\text{-я реализация}}, x(N), x(N+1), \dots, \underbrace{x((M-1)N), \dots, x(L-1)}_{M\text{-я реализация}} \right\}.$$

Получаем ансамбль реализаций $x_i(n)$, $i = 1, 2, \dots, M$, $n = 0, 1, \dots, N-1$.

2. Выбираем функцию временного окна $W(n)$ и домножаем каждую из реализаций на эту функцию⁶:

$$y_i(n) = x_i(n)W(n).$$

⁵В большинстве источников *периодограммой* принято называть оценку спектральной плотности мощности, получаемую по N отсчётам одной реализации случайного процесса.

⁶Эта операция может быть опущена, если используется прямоугольное окно. В этом случае $y_i(n) = x_i(n)$.

3. По каждой реализации посредством ДПФ строим периодограмму:

$$Y_i(k) = \sum_{n=0}^{N-1} y_i(n) e^{-j2\pi kn/N}.$$

4. Подсчитываем *выборочные* спектры мощности:

$$P_i(k) = Y_i(k) Y_i^*(k).$$

5. Подсчитываем *оценку спектра мощности*, усредняя по ансамблю выборочные спектры:

$$\hat{P}(k) = \langle P_i(k) \rangle = \frac{1}{M} \sum_{i=1}^M P_i(k). \quad (4.44)$$

При реальных измерениях длительность случайного сигнала всегда конечна, а значит, конечны длительность отдельной реализации N и число реализаций M . Ясно, что сигнал ограниченной длительности $L = M \times N$ можно разбить на подреализации множеством разных способов. Например, если общая длительность дискретного сигнала составляет 1000 отсчётов, можно «нарезать» 100 временных реализаций по 10 точек в каждой, можно – 10 реализаций по 100 точек, а можно оставить одну длинную реализацию в 1000 отсчётов. Какое из таких разбиений лучше? Чтобы ответить на этот вопрос, надо понять, за что отвечают параметры M и N . Что касается длительности под-реализации N , то ответ уже был дан в разделе 4.4.1: она определяет частоту разрешения спектра $\omega_T = 2\pi/N$. Рассмотрим теперь параметр M .

Если рассматриваемый процесс является «истинно» случайным, например белым шумом с нормальным распределением, то в теории спектров показано, что дисперсия спектра мощности, подсчитанного по одной реализации, при больших N стремится к значению квадрата спектра мощности:

$$D_{P_i}(k) = \langle P_i^2(k) - P^2(k) \rangle \xrightarrow{M \rightarrow \infty} P^2(k).$$

Иными словами, «ошибка» при расчёте спектра мощности по одной реализации сопоставима со значением самого спектра мощности. Именно поэтому спектр мощности, подсчитанный по одной периодограмме, как уже было сказано выше, не может характеризовать спектр мощности случайного процесса: оценка, полученная по одной периодограмме, является несостоятельной.

Наряду с ДПФ при цифровой обработке сигналов используются дискретное преобразование Лапласа и его модификация – z -преобразование.

4.6. Дискретные преобразование Лапласа и z -преобразование

Обратимся к формуле (4.4) для дискретизованного сигнала $x_d(n\Delta t)$ и, используя двустороннее преобразование Лапласа (2.58), найдем изо-

бражение сигнала $x_d(n\Delta t)$. Тогда

$$\begin{aligned} F(s) &= \int_{-\infty}^{\infty} x_d(n\Delta t)e^{-st} dt = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(t)\delta(t - n\Delta t)e^{-st} dt = \\ &= \sum_{n=-\infty}^{\infty} x(t) \int_{-\infty}^{\infty} \delta(t - n\Delta t)e^{-st} dt = \sum_{n=-\infty}^{\infty} x(n\Delta t)e^{-sn\Delta t}, \end{aligned} \quad (4.45)$$

где $s = \sigma + j\omega$. В последнем равенстве учтено фильтрующее свойство δ -функций.

Это *дискретное преобразование Лапласа*. Если в выражении (4.45) заменить функцию $e^{s\Delta t}$ комплексной переменной $z = e^{s\Delta t}$, то получим модифицированное преобразование Лапласа, известное как *z -преобразование*, и анализ дискретных сигналов упрощается.

Подставляя z в (4.45), получим двустороннее z -преобразование

$$X(z) = \sum_{n=-\infty}^{\infty} x(n\Delta t)z^{-n}. \quad (4.46)$$

Взаимное соответствие между z -преобразованием $X(z)$ и преобразованием Лапласа $F(s)$ легко представить в виде:

$$X(z) = F\left(\frac{1}{\Delta t} \ln z\right), \quad F(s) = X(e^{s\Delta t}).$$

По-видимому, в формулах для z -преобразования целесообразно так изменить обозначения, чтобы они соответствовали принятым при рассмотрении ДПФ, т.е. заменить $x(n\Delta t)$ на $x(n)$. Тогда

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}. \quad (4.47)$$

Рассмотрим теперь, как z -преобразование связано с ДПФ. Представим z в виде $z = \rho e^{j\omega}$ и получим

$$X(z) = \sum_{n=-\infty}^{\infty} [x(n)e^{-j\omega n}] \rho^{-n}.$$

Если в это выражение подставить $\omega = 2\pi k/N$ $0 \leq k \leq N-1$, то получим

$$X(z) = \sum_{n=-\infty}^{\infty} [x(n)e^{-j2\pi kn/N}] \rho^{-n}. \quad (4.48)$$

Выбирая равноудалённые выборки из $X(z)$ для $0 \leq n \leq N-1$ и полагая $\rho = 1$ ($|z| = 1$), видим, что (4.48) совпадает с (4.29). Таким образом, z -преобразование можно рассматривать как расширенное ДПФ.

В заключение получим формулу обратного z -преобразования. Для этого умножим обе части (4.47) на $(1/2\pi j)Z^{k-1}$ и проинтегрируем по контуру C , лежащему в области сходимости и содержащему начало координат. Согласно теории функций комплексной переменной в результате обхода контура против часовой стрелки получим

$$\frac{1}{2\pi j} \int_C X(z)z^{k-1} dz = \sum_{n=-\infty}^{\infty} x(n) \left[\frac{1}{2\pi j} \int_C z^{(k-n)-1} dz \right]. \quad (4.49)$$

Согласно теореме Коши

$$\frac{1}{2\pi j} \int_C p^{m-1} dp = \begin{cases} 1 & m = 0, \\ 0 & m \neq 0. \end{cases}$$

Итак, при $k = n$ из (4.49) получается формула обратного z -преобразования

$$x(n) = \frac{1}{2\pi j} \int_C X(z) z^{n-1} dz. \quad (4.50)$$

4.7. Контрольные вопросы и упражнения

1. Какие процессы называются дискретизацией аналоговых сигналов? Что такое равномерная дискретизация?
2. В чем состоит различие спектров аналогового и дискретизованного с равномерным шагом сигналов?
3. Спектр периодического сигнала с периодом T мкс содержит N гармоник. Какова должна быть минимальная частота выборки отсчетов (в выб/с или в Гц)?
4. Какова должна быть верхняя граничная частота фильтра (в МГц) для неискаженной выборки с частотой f_0 МГц?
5. В чем заключается метод полосовой дискретизации? Как выбирается частота полосовой дискретизации?
6. Как выбрать максимальное число k копий полосового сигнала, чтобы избежать перекрытия полос с отрицательными и положительными частотами? Как рассчитать частоту дискретизации для различных k ?
7. Какова максимальная допустимая ширина полосы спектра (в кГц), чтобы была возможна дискретизация с пониженной частотой, если значение центральной частоты спектра f_0 кГц?
8. Во сколько раз ширина полосы спектра сигнала должна быть меньше центральной частоты спектра, чтобы была возможна дискретизация с пониженной частотой?
9. Прямоугольный импульс имеет длительность 1 мс. С какой частотой необходимо дискретизировать этот сигнал, чтобы его спектр содержал первые два лепестка огибающей?
10. Что представляют собой цифровые сигналы? Как преобразовать аналоговый сигнал в цифровую форму?
11. В чем состоит кодирование сигналов? Назовите наиболее распространенные двоичные коды. Каковы особенности помехоустойчивых кодов?
12. Каково отношение сигнал-шум при 8-разрядном двоичном кодировании?
13. Подберите частоту выборки и число разрядов так, чтобы работать с сигналами, спектр которых находится в диапазоне от 0 до 10 МГц, чтобы в спектре сигнала различались гармоники, составляющие не менее -60 dB от основной гармоники сигнала.
14. Как различаются дискретно-временное и дискретное преобразования Фурье?
15. Какова нормированная частота $\bar{\omega}$, если физическая частота равна 1000 Гц, интервал дискретизации 100 мкс, а амплитуда сигнала 5 В?
16. Что понимают под частотой разрешения спектра? Как она связана с длительностью сигнала?

17. Чему равно разрешение спектра в нормированных частотах $\Delta\bar{\omega}$, если число точек сигнала, по которым рассчитывается спектр, равно 1000, а интервал дискретизации 30 мкс?
18. Каково разрешение спектра в Гц для условий, сформулированных в предыдущей задаче?
19. Сколько гармоник содержит дискретное преобразование Фурье, если число точек, по которым оно считается, равно 500, а интервал дискретизации 10 мкс?
20. В чем заключается «растекание спектра»? Как изменяются дискретные спектры сигналов при изменении длительности временного окна?
21. На каких свойствах ДПФ базируются алгоритмы быстрого преобразования Фурье (БПФ)?
22. Как формируются подпоследовательности N -отсчётной последовательности в алгоритмах прореживания по времени и по частоте?
23. Изобразите граф полного разложения 8-точечного ДПФ с прореживанием по времени.
24. Какова методика построения спектра мощности случайного дискретного процесса?
25. Как упрощается анализ дискретных сигналов при переходе от дискретного преобразования Лапласа к z -преобразованию?
26. Чему равно z -преобразование сигнала $x(n) = 3^{(n-2)}$?
27. Чему равно значение z -преобразования сигнала $x(n) = \delta(n - n_0)$ в точке $z = z_0$?

Список литературы

- Солонина, А. И., Улахович, Д. А., Арбузов, С. М., Соловова, Е. Б.* Основы цифровой обработки сигналов. СПб. : ДНУ-Питер, 2007.
- Сергиенко, А. Б.* Цифровая обработка сигналов: учебник. СПб. : Питер, 2002.
- Гольденберг, Л. М., Матюшкин, Б. Д., Поляк, М. Н.* Цифровая обработка сигналов: учеб. пособие для вузов. М. : Радио и связь, 1990.
- Лайонс, Р.* Цифровая обработка сигналов. М. : Бином-Пресс, 2006.
- Опенгейм, А., Шафер, Р.* Цифровая обработка сигналов. М. : Техносфера, 2006.
- Акимов, П. С., Сенин, А. И., Соленов, В. И.* Сигналы и их обработка в информационных системах. М. : Радио и связь, 1994.
- Залманзон, Л. А.* Преобразования Фурье, Уолша, Хаара и их применение в управлении, связи и других областях. М. : Наука, 1989.
- Марпл-мл., С. Л.* Цифровой спектральный анализ и его применения / пер. с англ. М. : Мир, 1990.
- Рабинер, Л., Гоулд, Б.* Теория и применение цифровой обработки сигналов. М. : Мир, 1978.
- Трахтман, А. М.* Введение в обобщённую спектральную теорию сигналов. М. : Сов. радио, 1972.

Глава 5

Модулированные и узкополосные сигналы

5.1. Основные задачи и методы модуляции

Сигналы, поступающие от источников сообщений, как правило, низкочастотные. Для передачи на значительные расстояния спектр сообщения необходимо перенести в область радиочастот или, как принято говорить в радиофизике, осуществить *модуляцию колебаний*¹. В таком преобразовании нуждается не одно, а сотни тысяч сообщений.

Пока существовала только радиосвязь, приходилось трансформировать сигналы с полосой частот до 10 кГц. Появилось телевидение и потребовалось переносить в область радиочастот сигналы с полосой частот до 8 МГц. Новая проблема породила новые методы модуляции. С развитием Интернета и мобильной связи комплекс проблем модуляции ещё более расширился. Хотя один канал мобильной связи занимает всего 3.5-4.0 кГц², но число каналов огромно. Поэтому возникают серьёзные проблемы эффективного использования радиочастотного диапазона, возможного временного и частотного уплотнения каналов, многократного использования одних и тех же частот, наконец, повышения помехоустойчивости. Все эти задачи в той или иной степени связаны с проблемами модуляции.

Нельзя не учитывать и проблемы демодуляции, т.е. обратного переноса информационного сигнала в область низких частот. Напомним ещё раз, что все передаваемые информационные сигналы являются случайными. Получатель должен найти нужный информационный сигнал, выделить его, учитывая статистические свойства сигнала и помех, и только после этого демодулировать. Способ демодуляции должен быть предельно простым. Как будет показано ниже, однополосная и балансная амплитудная модуляция не получили распространения именно из-за сложности демодуляции.

Модуляция и демодуляция сигналов – один из важных аспектов те-

¹В широком смысле под *модуляцией* (от лат. *modulation* – мерность, размеренность) понимают закономерное изменение или перемену состояния, например переход в другую тональность в музыке. В радиофизике и теории информации под модуляцией понимается метод преобразования сигналов для передачи сообщений по физическому каналу связи.

²Международным соглашением на стандартный канал отводится интервал частот от 300 до 3400 Гц и вводится межполосный интервал частот для снижения взаимных помех. Полная полоса частот – 4 кГц.

левизионного и спутникового вещания, а интенсивное развитие методов цифрового телевидения и телевидения высокой чёткости заставляет взглянуть на проблему телевидения по-новому. В заключительном разделе главы анализируется роль различных методов модуляции в процессах уплотнения каналов передачи сообщений.

Рассмотрение методов модуляции начнём с введения одного из основных понятий – понятия несущего колебания.

Несущим колебанием обычно называется вспомогательное высокочастотное гармоническое колебание

$$x_{\text{нес}}(t) = A \cos(\omega_0 t + \varphi_0) = A \cos \Psi(t), \quad (5.1)$$

характеристики которого изменяются в процессе модуляции.

Модуляцией называется физический процесс временного изменения одного или нескольких параметров несущего колебания в соответствии с сигналом передаваемого сообщения.

Несущее колебание (5.1) имеет три независимых параметра: *амплитуду* A , *круговую частоту* ω_0 и *текущую фазу* $\Psi = \omega_0 t + \varphi_0$.

Если при неизменных ω_0 и φ_0 для заданного сообщения $s(t)$ амплитуда колебания (5.1) изменяется по закону

$$A(t) = A + \Delta A s(t), \quad (5.2)$$

т.е. становится *информационным параметром*, то модуляция называется *амплитудной*, а колебание – *амплитудно-модулированным сигналом*.

Если при неизменных A и φ_0 информационным параметром становится частота колебания

$$\omega(t) = \omega_0 + \Delta\omega s(t), \quad (5.3)$$

то модуляция называется *частотной*, а колебание – *частотно-модулированным сигналом*. Величина $\Delta\omega$ характеризует максимальное отклонение частоты и называется *девиацией частоты*. Наконец, когда при постоянных A и ω_0 информационным параметром становится текущая фаза колебания

$$\Psi(t) = \Psi + \Delta\Psi s(t) = \omega_0 t + \varphi_0 + \Delta\Psi s(t), \quad (5.4)$$

где $\Delta\Psi$ – максимальное отклонение или *девиация* фазы, то модуляция называется *фазовой*, а результирующий сигнал – *фазо-модулированным*. (В дальнейшем используются сокращённые обозначения АМ, ЧМ и ФМ.)

Если в качестве сообщения использовать гармоническое колебание $s(t) = \cos \Omega t$, то, подставляя (5.4) в (5.1), получим для фазо-модулированного сигнала $x_{\text{ФМ}}(t)$ выражение:

$$x_{\text{ФМ}}(t) = A \cos[\omega_0 t + \varphi_0 + m \cos \Omega t]. \quad (5.5)$$

Индекс фазовой модуляции $m = \Delta\Psi$, как и в общем случае, характеризует девиацию фазы. Дифференцируя по времени текущую фазу ФМ-сигнала, получим

$$\frac{d\Psi(t)}{dt} = \omega_0 + \Delta\Psi \cdot \Omega \sin \Omega t = \omega_0 + \Delta\omega \sin \Omega t,$$

т.е. в однотоновом ФМ-сигнале девиация частоты пропорциональна девиации фазы и частоте модулирующего сигнала.

Выражение (5.3) для частоты ЧМ-сигнала нельзя непосредственно подставить в (5.1). Учитывая, что $\Psi(t) = \int_0^t \omega(\xi) d\xi + \varphi_0 = \omega_0 t + \Delta\omega \int_0^t s(\xi) d\xi$, и подставляя $\Psi(t)$ в (5.1), получим для ЧМ-сигнала выражение:

$$x_{\text{ЧМ}}(t) = A \cos[\omega_0 t + \Delta\omega \int_0^t s(\xi) d\xi + \varphi_0]. \quad (5.6)$$

Итак, соотношения (5.5) и (5.6) идентичны. Единственное различие

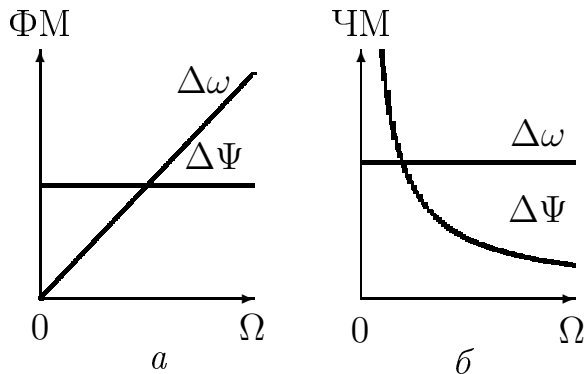


Рис. 5.1. Зависимости $\Delta\omega$ и $\Delta\Psi$ от частоты Ω модулирующего колебания для ФМ-сигнала (а) и ЧМ-сигнала (б)

однотонального модулированного по частоте сигнала девиация частоты $\Delta\omega$ не зависит от частоты модулирующего сигнала, а девиация фазы $\Delta\Psi$ обратно пропорциональна Ω (рис. 5.1, б).

Несущее колебание можно одновременно модулировать и по амплитуде, и по фазе и, таким образом, независимо передавать два низкочастотных сообщения. Тогда получаем *амплитудно-фазовую модуляцию*:

$$x(t) = A(t) \cos[\omega_0 t + \varphi(t)], \quad (5.7)$$

где $A(t)$, $\varphi(t)$ изменяются в соответствии с $s_1(t)$ и $s_2(t)$. К сожалению, $s_1(t)$ и $s_2(t)$ неравноправны. Если представить $x(t)$ в виде

$$\begin{aligned} x(t) &= A(t) \cos \varphi(t) \cos \omega_0 t - A(t) \sin \varphi(t) \sin \omega_0 t = \\ &= A_c(t) \cos \omega_0 t + A_s(t) \sin \omega_0 t, \end{aligned} \quad (5.8)$$

то $x(t)$ превращается в сумму двух АМ-колебаний с модулирующими сигналами $A_c(t)$ и $A_s(t)$, у которых несущие находятся в квадратуре (сдвинуты по фазе на 90°) и имеют одинаковую частоту ω_0 . Такая модуляция называется *квадратурной амплитудной модуляцией* (КАМ). На основе квадратурной модуляции в телевизионной системе NTSC в каждой строке передаются два цветоразностных сигнала.

Структура спектров модулированных сигналов зависит и от спектральных характеристик информационного (модулирующего) сигнала, и от вида модуляции. Обычно модулирующий сигнал *медленно изменяется во времени* по сравнению с несущим колебанием. Это означает, что *наивысшая частота в спектре передаваемого сигнала значительно ниже частоты несущего колебания* и модулированный сигнал можно считать *узкополосным* или даже *квазигармоническим*.

Модулирующий сигнал $s(t)$ может быть аналоговым или цифровым. Если модулирующий сигнал является дискретным, т.е. создает изменения информационного параметра несущего колебания в дискретные моменты времени $k\Delta_t, k = 0, 1, 2, \dots$, а сам параметр принимает только конечное число определенных значений (2, 4, 8 или более), то такой тип модуляции называется *манипуляцией*. (В дальнейшем для амплитудно-частотно- и фазо-манипулированных сигналов используются сокращенные обозначения АМн, ЧМн и ФМн.) Когда значения параметра передаются в цифровой форме, модуляция называется *цифровой*. Среди цифровых методов наиболее популярна *импульсно-кодовая модуляция*, позволяющая значительно снижать мощность сигнала.

В современной радиоэлектронике различают *одноступенчатую* и *многоступенчатую* (чаще всего *двухступенчатую*) модуляции.

При одноступенчатой модуляции передаваемый сигнал непосредственно управляет одним из параметров несущего колебания и создает АМ-, ЧМ- или ФМ-сигнал. При двухступенчатой модуляции информационный сигнал модулирует так называемую *поднесущую* в виде гармонического колебания или периодической последовательности импульсов, как правило, прямоугольных (первая ступень модуляции), а полученный сигнал используется для модуляции (обычно амплитудной) высокочастотного гармонического колебания несущей (вторая ступень модуляции). Если поднесущая – гармоническое колебание, то в результате двухступенчатой модуляции получают комбинации АМ-АМ, ЧМ-ФМ, АМ-ЧМ и другие.

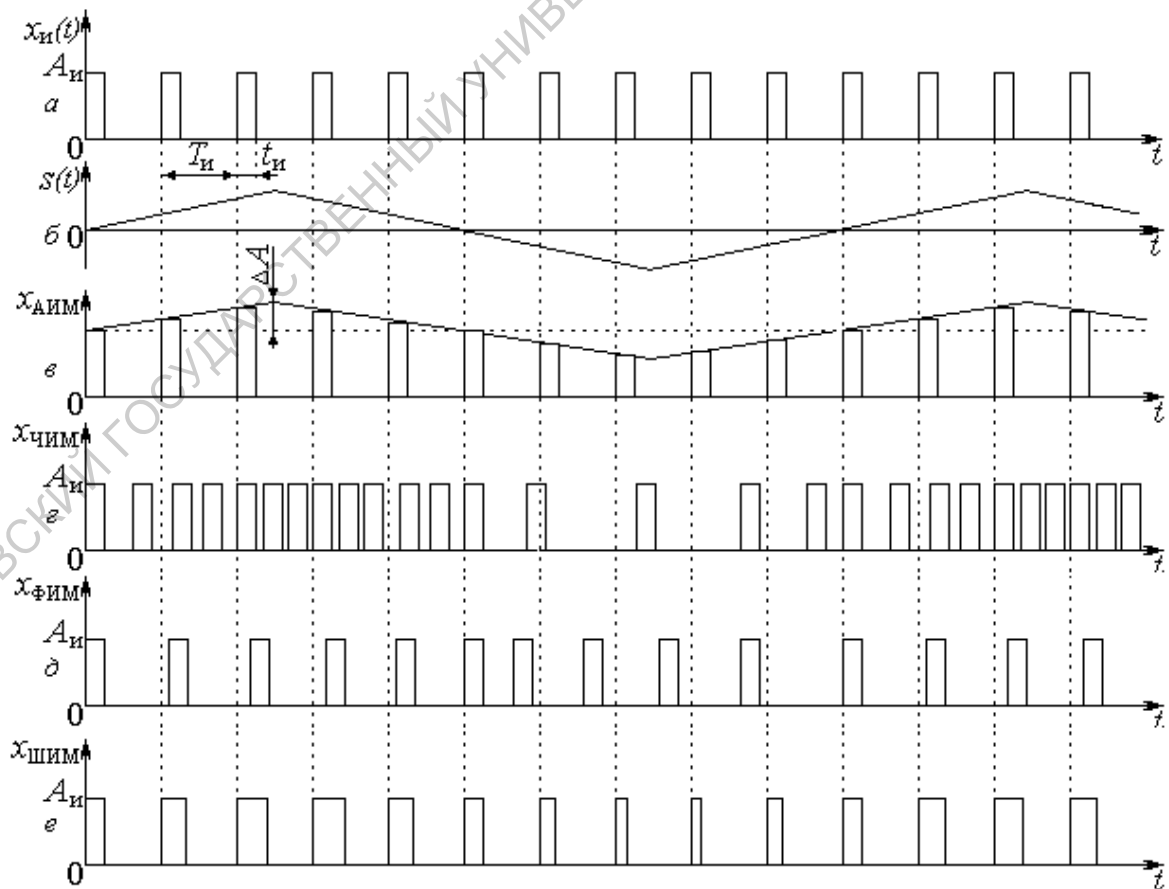


Рис. 5.2. Разные виды модуляций последовательности прямоугольных импульсов

Если же поднесущая – последовательность импульсов (рис. 5.2, а), то на первой ступени возможны четыре основных вида импульсной модуляции. Последовательность импульсов можно промодулировать сигналом $s(t)$ (рис. 5.2, б) по амплитуде (рис. 5.2, в), по частоте следования (рис. 5.2, г) или временному положению импульсов (рис. 5.2, д), а также изменяя длительности импульсов исходной последовательности (рис. 5.2, е) и получить соответственно *амплитудно-импульсную* (АИМ), *частотно-импульсную* (ЧИМ), *время-импульсную* или *фаза-импульсную* (ВИМ или ФИМ) и *широтно-импульсную* (ШИМ) модуляцию, а в результате двухступенчатой модуляции получить различные комбинации, например АИМ-АМ. Двухступенчатая модуляция АМ-ЧИМ используется для передачи стереофонических сигналов на станциях УКВ-вещания. Среди методов импульсной модуляции наибольшее распространение получила АИМ, так как при модуляции период повторения импульсов $T_{\text{и}}$ остается постоянным, и после первой последовательности импульсов можно передать вторую, задержанную на долю периода, затем третью и т.д. Система становится многоканальной, и один и тот же диапазон частот используется для передачи нескольких сообщений.

5.2. Амплитудно-модулированные сигналы

5.2.1. Модуляция гармонических колебаний

Пусть задан амплитудно-модулированный сигнал

$$x_{\text{AM}}(t) = [A + \Delta A \cdot s(t)] \cos \omega_0 t, \quad |s(t)| \leq 1, \quad (5.9)$$

где $s(t)$ – модулирующий сигнал; A – амплитуда несущего колебания; ω_0 – его круговая частота; $m = \Delta A/A$ – коэффициент модуляции.

Коэффициент модуляции является основным параметром АМ-колебания, обычно выражается

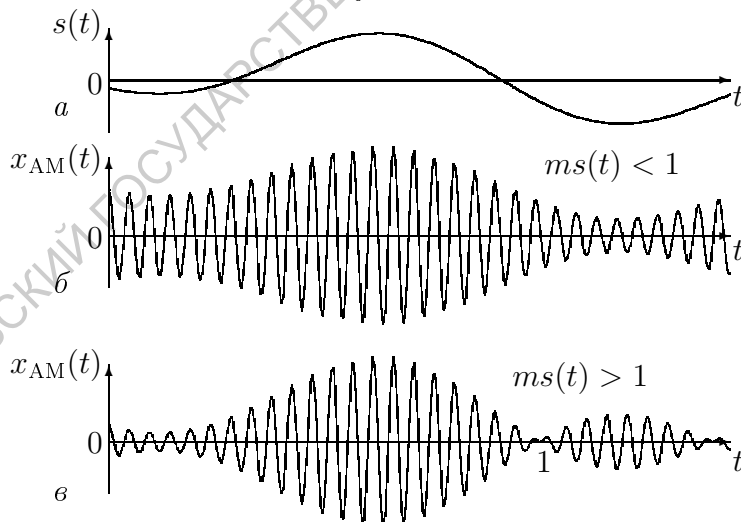


Рис. 5.3. Модулирующий (а) и АМ-сигнал при $m < 1$ и $m > 1$ (б, в), 1 – точка разрыва фазы несущего колебания

в процентах и характеризует эффективность передачи сообщения. Пока $|ms(t)| < 1$ (рис. 5.3, б), модулирующий сигнал (рис. 5.3, а) вносится в $x_{\text{AM}}(t)$ без искажений. Если же $|ms(t)| > 1$ (рис. 5.3, в), то происходит *перемодуляция* сигнала. Как видно из рис. 5.3, в, в этом случае огибающая АМ-сигнала не соответствует модулирующему сигналу, а фаза несущего колебания претерпевает разрывы (точка 1 на

рис. 5.3, в). При восстановлении передаваемого информационного сигнала из перемодулированного неизбежно возникают ошибки.

Спектральное представление амплитудно-модулированных сигналов. Пусть модулирующий сигнал $s(t)$ – произвольная интегрируемая функция. Для определения спектральной плотности $x_{AM}(t)$ запишем его как

$$x_{AM}(t) = A \cos \omega_0 t + A m s(t) \cos \omega_0 t$$

и используем прямое интегральное преобразование Фурье (см. разд. 2.4). Тогда

$$\begin{aligned} X_{AM}(\omega) &= A \int_{-\infty}^{\infty} \cos \omega_0 t e^{-j\omega t} dt + A m \int_{-\infty}^{\infty} s(t) \cos \omega_0 t e^{-j\omega t} dt = \\ &= \frac{A}{2} \left[\int_{-\infty}^{\infty} e^{-j(\omega - \omega_0)t} dt + \int_{-\infty}^{\infty} e^{-j(\omega + \omega_0)t} dt + m \int_{-\infty}^{\infty} s(t) e^{-j(\omega - \omega_0)t} dt + m \int_{-\infty}^{\infty} s(t) e^{-j(\omega + \omega_0)t} dt \right]. \end{aligned} \quad (5.10)$$

Используя формулу (2.41) для спектральной плотности экспоненциальных функций и теорему о свертке (см. разд. 2.4), получим окончательно

$$X_{AM}(\omega) = \pi A [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] + \frac{A}{2} m [X_s(\omega - \omega_0) + X_s(\omega + \omega_0)], \quad (5.11)$$

где $X_s(\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt$ – спектральная плотность модулирующего сигнала. Как отмечалось выше, функция $|X_s(\omega)|$ симметрична относительно $\omega = 0$ (рис. 5.4, а).

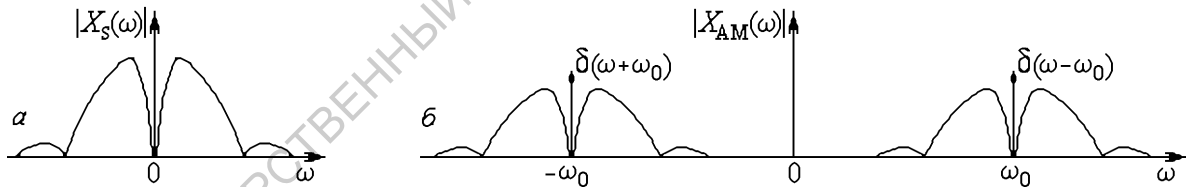


Рис. 5.4. Спектральные плотности модулирующего (а) и амплитудно-модулированного (б) сигналов

Таким образом, спектр гармонического колебания, промодулированного по амплитуде произвольным модулирующим сигналом $s(t)$, состоит из двух одинаковых частей. Каждая из них содержит дельта-функцию на частоте $\pm\omega_0$ ³ и распределение спектральной плотности модулирующего сигнала, у которого все частоты смещены к $\pm\omega_0$, а амплитуды изменены пропорционально коэффициенту модуляции m (рис. 5.4, б). Часть спектра, расположенная в области положительных частот, называется *односторонним физическим спектром* АМ-сигнала. Односторонний физический спектр симметричен относительно ω_0 , т.е. содержит верхнюю ($\omega > \omega_0$) и нижнюю ($\omega < \omega_0$) боковые полосы.

Простейшей моделью АМ-процесса может служить одностональный АМ-сигнал. Его изучение облегчает анализ основных свойств произвольных АМ-сигналов.

³Слагаемые в виде δ -функций на частотах $\pm\omega_0$ обусловлены наличием у функции $x_{AM}(t)$ неинтегрируемой гармонической компоненты $A \cos \omega_0 t$ с конечной амплитудой A .

Однотональный АМ-сигнал и его свойства. Пусть модулирующий сигнал $s(t)$ – гармоническое колебание:

$$s(t) = A_s \cos(\Omega t + \Phi_0),$$

частота которого Ω удовлетворяет условию $\Omega \ll \omega_0$. Тогда АМ-сигнал называется *однотональным*, а его мгновенное значение при $\Delta A = A_s$ описывается соотношением

$$x_{AM}(t) = A(1 + m \cos(\Omega t + \Phi_0)) \cos \omega_0 t, \quad (5.12)$$

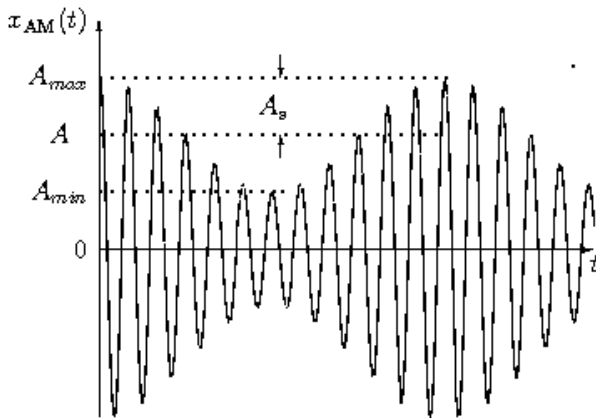


Рис. 5.5. Однотональный АМ-сигнал

где $m = A_s/A$. График (осциллограмма) однотонального АМ-колебания изображен на рис. 5.5 и показывает, что его амплитуда изменяется от $A_{min} = A(1 - m)$ до $A_{max} = A(1 + m)$, а коэффициент модуляции

$$m = \frac{A_{max} - A_{min}}{A_{max} + A_{min}}.$$

Чтобы найти спектральный состав такого сигнала, т.е. определить амплитуды c_i и частоты его гармонических составляющих, совсем не обязательно использовать Фурье-преобразование. Достаточно с помощью тригонометрических формул представить произведение функций в (5.12) в виде суммы гармонических колебаний. Тогда

$$x_{AM}(t) = A \cos \omega_0 t + \frac{mA}{2} \cos [(\omega_0 - \Omega)t - \Phi_0] + \frac{mA}{2} \cos [(\omega_0 + \Omega)t + \Phi_0]. \quad (5.13)$$

Таким образом, однотональный АМ-сигнал содержит три гармонических колебания с близкими частотами, а ширина спектра (рис.5.6) равна удвоенной частоте модулирующего сигнала. Первое колебание является несущим, и его амплитуда A не зависит от коэффициента модуляции. Два других колебания возникают в результате модуляции. Они имеют одинаковые амплитуды, пропорциональные m , начальные фазы $-\Phi_0$ и $+\Phi_0$, симметричные относительно несущего колебания, и частоты $\omega_0 - \Omega$ и $\omega_0 + \Omega$, называемые *нижней* и *верхней* боковыми частотами.

Анализ свойств модулированных сигналов в ряде случаев удобно проводить с помощью векторных диаграмм. Для АМ-сигнала векторная диаграмма представлена на рис. рис. 5.7.

Пусть вектор x_{AM} в начальный момент времени $t=0$ совпадает с осью Ox . Тогда в произвольный момент времени несущее колебание можно изобразить вектором, имеющим длину A и фазу $\omega_0 t$, а колебания боковых частот – векторами одинаковой длины $mA/2$, которые вращаются отно-

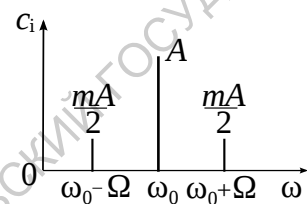


Рис. 5.6. Амплитудный спектр однотонального АМ-сигнала

сительно конца вектора A в противоположных направлениях с угловыми скоростями $-\Omega$ и Ω и имеют фазы $-\Omega t - \Phi_0$ и $\Omega t + \Phi_0$. Равнодействующая этих векторов все время остается сонаправленной с вектором несущей, а длина равнодействующей изменяется от $(1 - m)A$ до $(1 + m)A$.

Все это справедливо только для «идеального» АМ-сигнала, удовлетворяющего соотношению (5.13). Если амплитуды или модули фаз колебаний боковых частот различаются, то результирующий вектор не коллинеарен вектору несущей и будет испытывать качание относительно него, вызывая смешанную амплитудно-угловую модуляцию.

Замечания. 1. Структура спектра сохранится и для $s(t)$ с непрерывным спектром, но спектральные компоненты заполнят боковые полосы непрерывно.

2. Легко понять, что демодуляция АМ-сигналов сводится к выделению его низкочастотной огибающей. Чем проще и точнее такое выделение, тем эффективнее демодуляция.

Важнейшей характеристикой модулированных сигналов является распределение мощности между их спектральными составляющими. Вычисляя среднюю мощность однотонального АМ-сигнала за время T , значительно превышающее период модуляции ($T \gg 2\pi/\Omega$), получим

$$P_{\text{ср}} = \frac{1}{T} \int_0^T x_{\text{АМ}}^2(t) dt = \frac{A^2}{2} + \frac{m^2 A^2}{8} + \frac{m^2 A^2}{8}, \quad (5.14)$$

т.е. суммарная мощность колебаний боковых частот не превышает $1/3$ (при $m = 100\%$) от полной мощности АМ-сигнала. К тому же информация, заключённая в верхней и нижней боковых полосах, дублируется.

Поскольку все радиовещательные станции (кроме УКВ-диапазона) используют принципы амплитудной модуляции, проблема повышения эффективности использования мощности передатчиков долгое время занимала радиофизиков. В результате исследований были разработаны более экономичные методы амплитудной модуляции: модуляция с подавленной несущей, получившая название *балансной АМ*, и модуляция с подавленной боковой полосой – *однополосная АМ*.

При балансной модуляции в эфир поступают только сигналы верхней и нижней боковых полос. Чтобы понять, как при этом изменяется структура сигнала, обратимся к однотональной модуляции. В спектре АМ-сигнала теперь остаются две составляющие (два последних слагаемых в (5.13), сложение которых приводит к биениям (произведение гармонических функций в (5.12)). В результате огибающая не соответствует передаваемому сигналу и периодически обращается в нуль. Хотя в спектре отсутствует несущее колебание, высокочастотное заполнение содержит частоту ω_0 . При демодуляции АМ-сигнала несущая должна восстанавливаться в приёмнике, что значительно его усложняет.

Ещё большая эффективность использования мощности передатчика

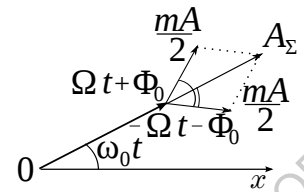


Рис. 5.7. Векторная диаграмма однотонального АМ-сигнала

достигается при однополосной амплитудной модуляции, когда в эфир передается несущая и одна боковая полоса частот или только боковая полоса частот. Строго говоря, это уже не амплитудная, а амплитудно-угловая модуляция. Обратимся к векторной диаграмме рис. 5.7. Если удалить один из векторов $mA/2$, то оставшийся вектор будет равномерно вращаться вокруг конца вектора A , а результирующий вектор изменится по амплитуде и колеблется по фазе.

Разработка различных систем, повышающих эффективность АМ, в том числе квадратурных модуляторов Хартли, не дала необходимого результата и проблема остается.

Амплитудно-манипулированные сигналы. Высокочастотное колебание, промодулированное по амплитуде последовательностью прямоугольных импульсов, называется *амплитудно-манипулированным сигналом* (АМн) (рис. 5.8).

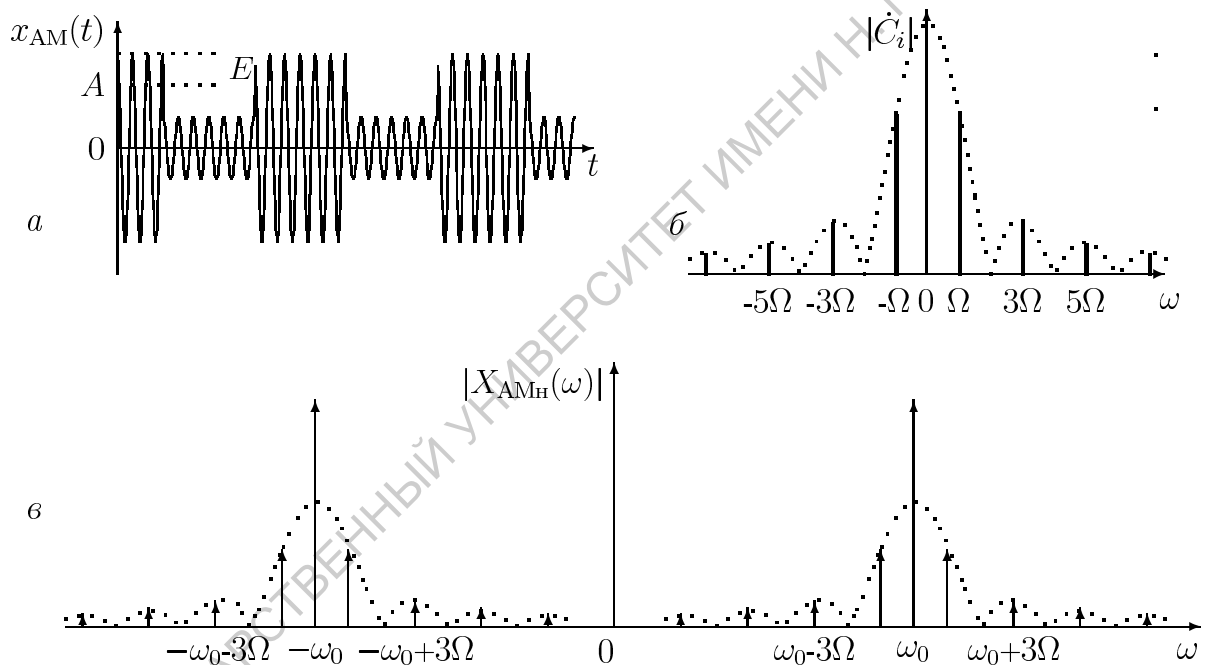


Рис. 5.8. Временная реализация амплитудно-манипулированного сигнала (а), спектр прямоугольной огибающей (б) и спектр АМн-сигнала (в)

Пусть скважность последовательности $q = 2$. Последовательность импульсов (см. разд. 2.3) представляет прямоугольное напряжение $s(t)$, а амплитуда импульсов – размах (удвоенную амплитуду) модулирующего напряжения. Используя для $s(t)$ соотношение (2.22), можно записать как

$$s(t) = \sum_{i=-\infty}^{\infty} \dot{C}_i e^{ji\Omega t}, \quad \dot{C}_i = \frac{E}{T} \int_{-T/2}^{T/2} e^{-ji\Omega t} dt = E \left(\frac{\sin(i\pi/2)}{i\pi/2} - \frac{\sin i\pi}{i\pi} \right),$$

где $\Omega = 2\pi/T$ – частота первой гармоники $s(t)$. Коэффициенты разложения \dot{C}_i с четными номерами (в том числе и \dot{C}_0) обращаются в нуль, и для мгновенных значений АМн-сигнала имеем:

$$x_{\text{АМн}}(t) = \left(A + \sum_{i=-\infty}^{\infty} \dot{C}_i e^{ji\Omega t} \right) \cos \omega_0 t = A \left(1 + m \sum_{\substack{i=-\infty \\ i \neq 0}}^{\infty} \frac{\sin(i\pi/2)}{i\pi/2} e^{ji\Omega t} \right) \cos \omega_0 t,$$

где $m = E/A$ – коэффициент модуляции. Наиболее полно АМн-сигнал описывает частотное распределение спектральной плотности:

$$\begin{aligned} X_{\text{АМн}}(\omega) &= \int_{-\infty}^{\infty} x_{\text{АМн}}(t) e^{-j\omega t} dt = A \int_{-\infty}^{\infty} \left[1 + m \sum_{\substack{i=-\infty \\ i \neq 0}}^{\infty} \frac{\sin i\pi/2}{i\pi/2} e^{ji\Omega t} \cos \omega_0 t \right] e^{-j\omega t} dt = \\ &= \pi A \left\{ \delta(\omega - \omega_0) + \delta(\omega + \omega_0) + m \sum_{\substack{i=-\infty \\ i \neq 0}}^{\infty} \frac{\sin i\pi/2}{i\pi/2} [\delta(\omega - (\omega_0 - i\Omega)) + \delta(\omega + (\omega_0 - i\Omega))] \right\}. \end{aligned} \quad (5.15)$$

Спектр АМн-сигнала (рис. 5.8, в) дискретный, так как модулирующий сигнал периодический. Как и в общем случае, частотное распределение спектральной плотности симметрично относительно $\omega = 0$. Односторонний (физический) спектр содержит δ -функцию на частоте несущего колебания и спектральную плотность модулирующего сигнала, все частоты которого смещены на ω_0 . Ширина спектра зависит от частоты переключений модулирующего напряжения, т.е. в конечном счете от частоты дискретизации сигнала сообщения. Когда модулирующий сигнал принимает несколько дискретных значений, спектр АМн-сигнала усложняется. Этот вопрос более подробно обсуждается при рассмотрении цифровой амплитудной манипуляции.

5.2.2. Амплитудная модуляция импульсных последовательностей

Пусть в качестве поднесущего колебания задана периодическая последовательность прямоугольных видеоимпульсов и каждый из импульсов описывается функцией $\eta(t) = A_{\text{и}}(\sigma(t) - \sigma(t - t_{\text{и}}))$ (см. разд. 2.7, формула (2.68)). Тогда

$$x_{\text{нec}}(t) = A_{\text{и}} \sum_{i=-\infty}^{\infty} \eta(t - iT_{\text{и}}),$$

где $A_{\text{и}}$, $t_{\text{и}}$ и $T_{\text{и}}$ – амплитуда, длительность и период повторения импульсов, а амплитудно-модулированный сигнал принимает вид:

$$x_{\text{АИМ}}(t) = A_{\text{и}} [1 + ms(t)] \sum_{i=-\infty}^{\infty} \eta(t - iT_{\text{и}}), \quad (5.16)$$

где $s(t)$ – модулирующий (информационный) сигнал; m – коэффициент модуляции. Временная диаграмма $x_{\text{АИМ}}(t)$ представлена на рис. 5.2, в. Для определения частотного распределения спектральной плотности сигнала проанализируем (5.16).

Несущее колебание $x_{\text{нec}}(t)$ как периодическую функцию можно представить разложением в ряд Фурье:

$$x_{\text{нec}}(t) = \sum_{k=-\infty}^{\infty} \dot{C}_k e^{j2\pi kt/T_n} = \sum_{k=-\infty}^{\infty} \dot{C}_k e^{jk\omega_I t},$$

где $\dot{C}_k = A_n \frac{t_n}{T_n} \frac{\sin(k\omega_I t_n/2)}{k\omega_I t_n/2} = A_n \frac{t_n}{T_n} \text{sinc}(k\omega_I t_n/2)$ – комплексные амплитуды гармоник (см. разд. 2.3); $\omega_I = 2\pi/T_n$ – круговая частота первой гармоники $x_{\text{нec}}(t)$. Если спектр модулирующего сигнала $s(t)$ представить спектральной плотностью $X_s(\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt$, то с учетом введенных обозначений спектральная плотность АИМ-сигнала принимает вид:

$$\begin{aligned} X_{\text{АИМ}}(\omega) &= \int_{-\infty}^{\infty} x_{\text{АИМ}}(t) e^{-j\omega t} dt = \int_{-\infty}^{\infty} [1 + ms(t)] \sum_{k=-\infty}^{\infty} \dot{C}_k e^{-j(\omega - k\omega_I)t} dt = \\ &= \sum_{k=-\infty}^{\infty} \dot{C}_k \int_{-\infty}^{\infty} e^{-j(\omega - k\omega_I)t} dt + m \sum_{k=-\infty}^{\infty} \dot{C}_k \int_{-\infty}^{\infty} s(t) e^{-j(\omega - k\omega_I)t} dt = \\ &= 2\pi \sum_{k=-\infty}^{\infty} \dot{C}_k \delta(\omega - k\omega_I) + m \sum_{k=-\infty}^{\infty} \dot{C}_k X_s(\omega - k\omega_I). \end{aligned} \quad (5.17)$$

Спектральная плотность АИМ-сигнала согласно (5.17) содержит спектральную плотность немодулированной последовательности прямоугольных импульсов $x_{\text{нec}}(t)$ и смещенные к частотам $k\omega_I$ спектральные плотности модулирующего сигнала $s(t)$. Таким образом, около каждой из спектральных гармоник последовательности импульсов формируется двухполосный спектр АМ-сигнала. Если наибольшая частота ω_s в спектре $s(t)$ меньше $\omega_I/2$, то спектры АМ-сигналов не перекрываются и результирующий спектр имеет вид, представленный на рис. 5.9. Любой из АМ-сигналов можно использовать для демодуляции информационного сигнала. Однако из формулы (5.17) следует и более привлекательное решение. Во второй сумме один из интегралов при $k = 0$ представляет *низкочастотный спектр* модулирующего сигнала (заштрихованная область на рис. 5.9), и для демодуляции достаточно выделить эту низкочастотную часть спектра.

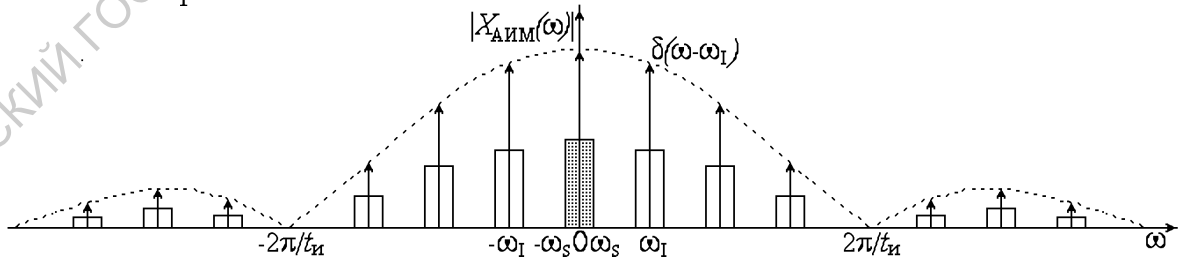


Рис. 5.9. Частотное распределение модуля спектральной плотности АИМ-сигнала

Как отмечалось выше, если период повторения импульсов таков, что между ними можно передавать ещё несколько задержанных на соответствующее время последовательностей, модулированных другими информационными сигналами, то имеет место многоканальная связь. Для этих сигналов будет использоваться один и тот же частотный диапазон, и ширина спектра, пропорциональная длительности импульсов, не изменится.

5.3. Сигналы с угловой модуляцией

У произвольного гармонического колебания

$$x(t) = A \cos(\omega_0 t + \varphi_0) = A \cos \Psi(t)$$

набег текущей фазы $\Psi(t)$ является линейной функцией времени. Гармоническое колебание $A \cos(\omega_0 t + \varphi_0)$ – проекция равномерного вращательного движения точки по окружности единичного радиуса. Фаза $\Psi(t)$ задает текущий угол точки относительно положительного направления оси координат, ее значение измеряется в градусах или радианах. Изменение фазы в единицу времени представляет угловую скорость вращательного движения:

$$\omega(t) = \frac{d\Psi(t)}{dt}, \quad (5.18)$$

называемую также круговой частотой вращения. Единицы измерения круговой частоты вращения ω и частоты колебаний $f = \omega/2\pi$ рассмотрены в разд. 2.1.

Текущее значение фазы можно вычислить, если известны круговая частота ω и начальная фаза φ_0 :

$$\Psi(t) = \int_0^t \omega(\xi) d\xi + \varphi_0. \quad (5.19)$$

Для немодулированного колебания Ψ – линейная функция времени, а круговая частота ω – константа: $\omega(t) = \omega_0$. При угловой модуляции значение фазы перестает быть линейной функцией времени и, соответственно, частота перестает быть постоянной. Воспользуемся этими соотношениями для анализа структуры сигналов с угловой модуляцией.

5.3.1. Фазо-модулированные сигналы и их спектры

Фазовой модуляцией будем называть вид угловой модуляции, при которой фаза сигнала изменяется в соответствии с законом:

$$\Psi(t) = \omega_0 t + \varphi_0 + ms(t), \quad (5.20)$$

где $-1 \leq s(t) \leq 1$ – нормированный информационный сигнал с единичной амплитудой и нулевым средним; постоянный коэффициент m – индекс (коэффициент) фазовой модуляции. Учитывая (5.20), можно записать фазо-модулированный сигнал в виде:

$$x_{\text{ФМ}}(t) = A \cos[\omega_0 t + \varphi_0 + ms(t)]. \quad (5.21)$$

Конкретное значение начальной фазы φ_0 для рассмотрения фазовой модуляции является несущественным и в дальнейшем принимается равным нулю.

Спектральное представление фазо-модулированных сигналов. Преобразуем выражение (5.21) к виду

$$x_{\text{ФМ}}(t) = A \cos \omega_0 t \cos(ms(t)) - A \sin \omega_0 t \sin(ms(t)). \quad (5.22)$$

Обозначим $x_1(t) = A \cos \omega_0 t \cos(ms(t))$, $x_2(t) = A \sin \omega_0 t \sin(ms(t))$, найдем спектральные плотности сигналов $x_1(t)$ и $x_2(t)$ по отдельности и

спектральную плотность полного сигнала как $X_{\text{ФМ}}(\omega) = X_1(\omega) - X_2(\omega)$. Тогда

$$\begin{aligned} X_1(\omega) &= A \int_{-\infty}^{\infty} \cos \omega_0 t \cos(ms(t)) \exp(-j\omega t) dt = \\ &= \frac{A}{2} \int_{-\infty}^{\infty} \cos(ms(t)) [\exp(-j(\omega - \omega_0)t) + \exp(-j(\omega + \omega_0)t)] dt = \\ &= \frac{A}{2} [X_c(\omega - \omega_0) + X_c(\omega + \omega_0)]. \end{aligned} \quad (5.23)$$

Здесь использовано обозначение $X_c(\omega) = \int_{-\infty}^{\infty} \cos(ms(t)) \exp(-j\omega t) dt$ для спектральной плотности сигнала $\cos(ms(t))$. Аналогично найдем

$$X_2(\omega) = \frac{A}{2j} [X_s(\omega - \omega_0) - X_s(\omega + \omega_0)], \quad (5.24)$$

где $X_s(\omega) = \int_{-\infty}^{\infty} \sin(ms(t)) \exp(-j\omega t) dt$ – спектральная плотность сигнала $\sin(ms(t))$. Учитывая (5.23) и (5.24), получаем для полной спектральной плотности фазо-модулированного сигнала выражение:

$$X_{\text{ФМ}}(\omega) = \frac{A}{2} [X_c(\omega - \omega_0) + jX_s(\omega - \omega_0)] + \frac{A}{2} [X_c(\omega + \omega_0) - jX_s(\omega + \omega_0)]. \quad (5.25)$$

Выражение (5.25) можно записать в более компактной форме, если ввести комплексный сигнал $z(t) = \cos(ms(t)) + j \sin(ms(t)) = \exp(jms(t))$. Тогда

$$X_{\text{ФМ}}(\omega) = \frac{A}{2} Z(\omega - \omega_0) + \frac{A}{2} \tilde{Z}(\omega + \omega_0), \quad (5.26)$$

где $Z(\omega)$ – спектральная плотность $z(t)$; $\tilde{Z}(\omega)$ – спектральная плотность сигнала $z^*(t)$.

Проанализируем соотношение (5.26). Мы видим, что спектральная плотность ФМ-сигнала состоит из двух слагаемых: $X_{\text{ФМ}}^+ = \frac{A}{2} Z(\omega - \omega_0)$ и $X_{\text{ФМ}}^- = \frac{A}{2} \tilde{Z}(\omega + \omega_0)$. Первое из них соответствует сигналу $\exp(jms(t))$, смещенному на ω_0 в сторону положительных частот, а второе – сопряженному сигналу $\exp(-jms(t))$, смещенному на $-\omega_0$ в сторону отрицательных частот. Таким образом, при фазовой модуляции, как и при амплитудной модуляции, происходит перенос спектра на частоты $\pm\omega_0$. Но в отличие от АМ-сигналов, у которых спектральная плотность информационного (модулирующего) сигнала транслируется без изменений, при фазовой модуляции спектральная плотность модулирующего сигнала подвергается воздействию нелинейного преобразования. Важно понять, каким образом это преобразование влияет на ширину спектра? Для того чтобы ответить на этот вопрос, обратимся сначала к ФМ-сигналам с малыми индексами модуляции, а затем перейдем к ФМ-сигналам с большими значениями индекса модуляции.

Спектральное представление ФМ-сигналов с малым индексом модуляции. Как известно, экспоненциальную функцию можно представить в виде бесконечного степенного ряда:

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + \dots$$

С учетом этого представления введенный нами вспомогательный комплексный сигнал $z(t)$ запишем как

$$z(t) = 1 + jms(t) - \frac{m^2}{2}s^2(t) - j\frac{m^3}{6}s^3(t) + \dots + j^n\frac{m^n}{n!}s^n(t) + \dots \quad (5.27)$$

Будем считать, что индекс модуляции $m \ll 1$. Тогда в формуле (5.27) можно пренебречь всеми, кроме первых двух, слагаемыми, т.е.

$$z(t) \cong 1 + jms(t).$$

Спектральную плотность такого сигнала нетрудно представить как

$$Z(\omega) = 2\pi\delta(\omega) + jmS(\omega),$$

где $S(\omega)$ – спектральная плотность модулирующего сигнала. Подставляя полученное выражение $Z(\omega)$ в $X_{\Phi M}^+(\omega)$, получим

$$X_{\Phi M}^+ = \frac{A}{2}[2\pi\delta(\omega - \omega_0) + jmS(\omega - \omega_0)]. \quad (5.28)$$

Сопоставим выражение (5.28) с полученной ранее спектральной плотностью АМ-сигнала. И в том, и в другом случае спектр модулированного сигнала содержит дельта-функцию на частоте несущей и произведение коэффициента модуляции m и смещенной на эту частоту спектральной плотности модулирующего сигнала. Если модулирующий сигнал $s(t)$ обладает ограниченным спектром $S(\omega) = 0$ для $\omega \notin [-\omega_h : \omega_h]$ (ω_h – верхняя граничная частота спектра), то ширина спектра модулированного сигнала равна удвоенной граничной частоте⁴:

$$\Delta\omega = 2\omega_h.$$

Отличие спектров АМ- и ФМ-сигналов (при малых m) заключается только в наличии множителя j , который соответствует сдвигу фазы спектра $S(\omega)$ относительно фазы несущей на угол в 90° . Таким образом, амплитудные спектры АМ- и ФМ-сигналов при малых индексах модуляции совпадают, а фазовые спектры различаются: в случае амплитудной модуляции фаза спектральной плотности огибающего сигнала не меняется, а в случае фазовой модуляции поворачивается на $\pi/2$.

В частном случае *однотональной* модуляции $s(t) = \cos \Omega t$. Спектральная плотность такого сигнала

$$S(\omega) = \pi[\delta(\omega - \Omega) + \delta(\omega + \Omega)].$$

Следовательно, часть спектра, соответствующая положительным частотам, будет иметь вид:

$$X_{\Phi M}^+ = \pi A[\delta(\omega - \omega_0) + j\frac{m}{2}\delta(\omega - \omega_0 - \Omega) + j\frac{m}{2}\delta(\omega - \omega_0 + \Omega)].$$

⁴В литературе иногда указывают, что ширина спектра модулированных колебаний вдвое превосходит ширину спектра модулирующего сигнала. Это неверно, так как модулирующий сигнал может иметь полосовой спектр (чаще всего так и бывает) и его ширина меньше, чем ω_h .

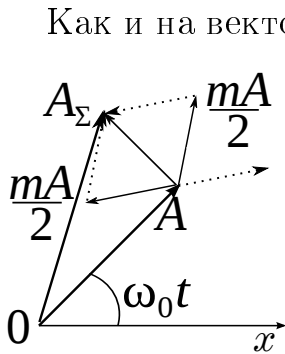


Рис. 5.10. Векторная диаграмма ФМ-сигнала с малым индексом модуляции

5.10) его длина изменяется, так как для большей наглядности рисунка длины векторов боковых колебаний выбраны чрезмерно большими и условие $m \ll 1$ не выполняется.

Изменение спектра при увеличении индекса модуляции. При больших индексах модуляции в разложении $z(t)$ будем учитывать слагаемое, квадратичное по m :

$$z(t) \cong 1 + jms(t) - \frac{m^2}{2}s^2(t).$$

В этом случае к спектральной плотности вида (5.28) необходимо добавить слагаемое $-0.5m^2S_2(\omega)$, где $S_2(\omega)$ – функция спектральной плотности для сигнала $s^2(t)$. Тогда

$$X_{\text{ФМ}}^+(\omega) = \frac{A}{2}[\delta(\omega - \omega_0) + jms(\omega - \omega_0) - \frac{m^2}{2}S_2(\omega - \omega_0)].$$

Как известно (см. разд. 2.4), спектр произведения сигналов равен интегральной свертке спектров сомножителей. В данном случае

$$S_2(\omega) = \int_{-\infty}^{\infty} S(\xi)S(\omega - \xi)d\xi.$$

Пусть спектр сигнала $s(t)$ ограничен сверху частотой ω_h . Тогда

$$S_2(\omega) = \int_{-\omega_h}^{\omega_h} S(\xi)S(\omega - \xi)d\xi.$$

Величина $S(\omega - \xi)$ отлична от нуля только в диапазоне значений аргумента $-\omega_h \leq \omega - \xi \leq \omega_h$, т.е. $-\omega_h + \xi \leq \omega \leq \omega_h + \xi$. Так как переменная интегрирования ξ меняется в тех же пределах, спектр сигнала $s^2(t)$ будет отличным от нуля в диапазоне частот $\omega \in [-2\omega_h : 2\omega_h]$.

Таким образом, верхняя граничная частота спектра сигнала $s^2(t)$ вдвое больше верхней граничной частоты самого сигнала, т.е. при учете в разложении (5.27) слагаемого, квадратичного по m , получаем спектр ФМ-сигнала, отличный от нуля в диапазоне $\omega \in [\omega_0 - 2\omega_h : \omega_0 + 2\omega_h]$. Ширина спектра становится равной $4\omega_h$. Легко увидеть, что учет кубического слагаемого в (5.27) даст нам ширину спектра в $6\omega_h$, а учет слагаемого n -ой степени по индексу модуляции – ширину спектра в $2n\omega_h$.

Из этих рассуждений можно сделать следующие выводы.

1. Ширина спектра ФМ-сигнала при конечном значении индекса модуляции является неограниченно большой, даже если спектр модулирующего сигнала ограничен.
2. Если ограничить точность расчета спектра, введя его эффективную ширину, не учитывающую те компоненты, амплитуда которых меньше некоторого заданного значения, то эффективная ширина спектра будет оставаться конечной при конечном индексе модуляции.
3. Эффективная ширина спектра растет с ростом индекса модуляции.

5.3.2. Частотно-модулированные сигналы и их спектры

Частотно-модулированными сигналами будем называть сигналы, у которых круговая частота меняется во времени по закону передаваемого информационного сигнала, т.е. $\omega(t) = \omega_0 + ms(t)$. С учетом соотношения между частотой и фазой сигнала ЧМ-сигнал имеет вид:

$$x_{\text{ЧМ}}(t) = A \cos \left(\omega_0 t + m \int_0^t s(\xi) d\xi + \varphi_0 \right)$$

или

$$x_{\text{ЧМ}}(t) = A \cos (\omega_0 t + m\zeta(t) + \varphi_0), \quad (5.29)$$

где $\zeta(t) = \int_0^t s(\xi) d\xi$ – первообразная от информационного сигнала $s(t)$. Из сопоставления (5.20) и (5.29) видно, что формы ФМ- и ЧМ-сигналов ничем не отличаются. Отличие заключается лишь в том, что, если в случае ФМ-сигналов информационный сигнал добавляется к фазе сигнала аддитивно, то в случае ЧМ-сигнала к фазе добавляется его первообразная. Поэтому все свойства спектров, которые были определены для ФМ-сигналов, остаются справедливыми и для ЧМ-сигналов. Однако в выражениях для спектров спектральную плотность информационного сигнала $S(\omega)$ следует заменить спектральной плотностью $S_\zeta(\omega)$ первообразной от $s(t)$. В общем случае спектр частотно-модулированного сигнала совпадает по форме со спектром ФМ-сигнала, т.е.

$$X_{\text{ЧМ}}(\omega) = \frac{A}{2} Z(\omega - \omega_0) + \frac{A}{2} \tilde{Z}(\omega + \omega_0),$$

но вспомогательный сигнал $z(t)$ представляет собой теперь нелинейное преобразование от сигнала $\zeta(t)$:

$$z(t) = \exp(jm\zeta(t)).$$

В случае малых индексов модуляции $m \ll 1$ спектр сигнала для положительных частот будет иметь вид:

$$X^+(\omega) = \frac{A}{2} \left[2\pi\delta(\omega - \omega_0) + \frac{m}{\omega - \omega_0} S(\omega - \omega_0) \right].$$

Спектральное представление ЧМ-сигналов с произвольным индексом модуляции. Пусть передаваемое сообщение (информационный сигнал) имеет вид: $s(t) = \cos \Omega t$. Тогда частота модулированного сигнала удовлетворяет соотношению:

$$\omega(t) = \omega_0 + m \int_0^t \cos \Omega \xi d\xi = \omega_0 + \frac{m}{\Omega} \sin \Omega t.$$

Обозначим m/Ω через β и назовем *индексом частотной модуляции* для однотонального сигнала с частотой Ω . Тогда ЧМ-сигнал (5.29) принимает вид:

$$x_{\text{ЧМ}}(t) = A \cos(\omega_0 t + \beta \sin \Omega t + \varphi_0). \quad (5.30)$$

Чтобы определить спектр этого ЧМ-сигнала, его нужно разложить на гармонические составляющие. Разложение проще осуществить для комплексного представления $x_{\text{ЧМ}}(t) = \text{Re}[A \exp j(\omega_0 t + \beta \sin \Omega t + \varphi_0)]$.

Рассмотрим экспоненциальную функцию

$$z(t) \exp j(\omega_0 t + \varphi_0) = \exp j(\beta \sin \Omega t) \exp j(\omega_0 t + \varphi_0).$$

Функцию $z(t)$, обладающую такой структурой, можно представить рядом Фурье:

$$z(t) = e^{j\beta \sin \Omega t} = \sum_{k=-\infty}^{\infty} J_k(\beta) e^{jk\Omega t}. \quad (5.31)$$

Подставляя (5.31) в $x_{\text{ЧМ}}(t)$ и вычисляя $\text{Re}(\cdot)$, окончательно получим:

$$x_{\text{ЧМ}}(t) = A \sum_{k=-\infty}^{\infty} J_k(\beta) \cos[(\omega_0 + k\Omega)t + \varphi_0]. \quad (5.32)$$

Итак, спектр однотонального модулированного по частоте (или в общем случае по углу) колебания эквидистантный и содержит бесконечное число боковых гармоник, попарно симметричных относительно частоты несущего колебания. Амплитуды всех составляющих, в том числе и несущей, пропорциональны значениям функций Бесселя. Функции Бесселя представляют собой медленно затухающие квазипериодические функции (рис. 5.11).

Значения β , при которых функции Бесселя нулевого, первого и второго порядка обращаются в нуль, представлены в табл. 5.1, а значения $J_0(\beta)$, $J_1(\beta)$ – $J_6(\beta)$ при фиксированных значениях β – в табл. 5.2.

Таблица 5.1

Значения первых пяти корней функций Бесселя $J_0(\beta)$, $J_1(\beta)$ и $J_2(\beta)$

β	β_1	β_2	β_3	β_4	β_5
$J_0(\beta)$	2.405	5.52	8.65	11.79	14.93
$J_1(\beta)$	3.83	7.01	10.17	13.31	16.47
$J_2(\beta)$	5.13	8.41	11.62	14.79	17.96

Таблица 5.2
Значения функций Бесселя $J_0(\beta), J_1(\beta) - J_6(\beta)$ при фиксированных значениях β

β	$J_0(\beta)$	$J_1(\beta)$	$J_2(\beta)$	$J_3(\beta)$	$J_4(\beta)$	$J_5(\beta)$	$J_6(\beta)$
$\beta = 0.5$	0.938	0.242	0.03	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$
$\beta = 1.0$	0.765	0.44	0.114	0.02	$\rightarrow 0$	$\rightarrow 0$	$\rightarrow 0$
$\beta = 1.5$	0.511	0.558	0.232	0.06	0.008	$\rightarrow 0$	$\rightarrow 0$
$\beta = 2.0$	0.224	0.576	0.352	0.13	0.034	0.007	$\rightarrow 0$
$\beta = 2.5$	-0.04	0.497	0.446	0.241	0.097	0.29	0.007

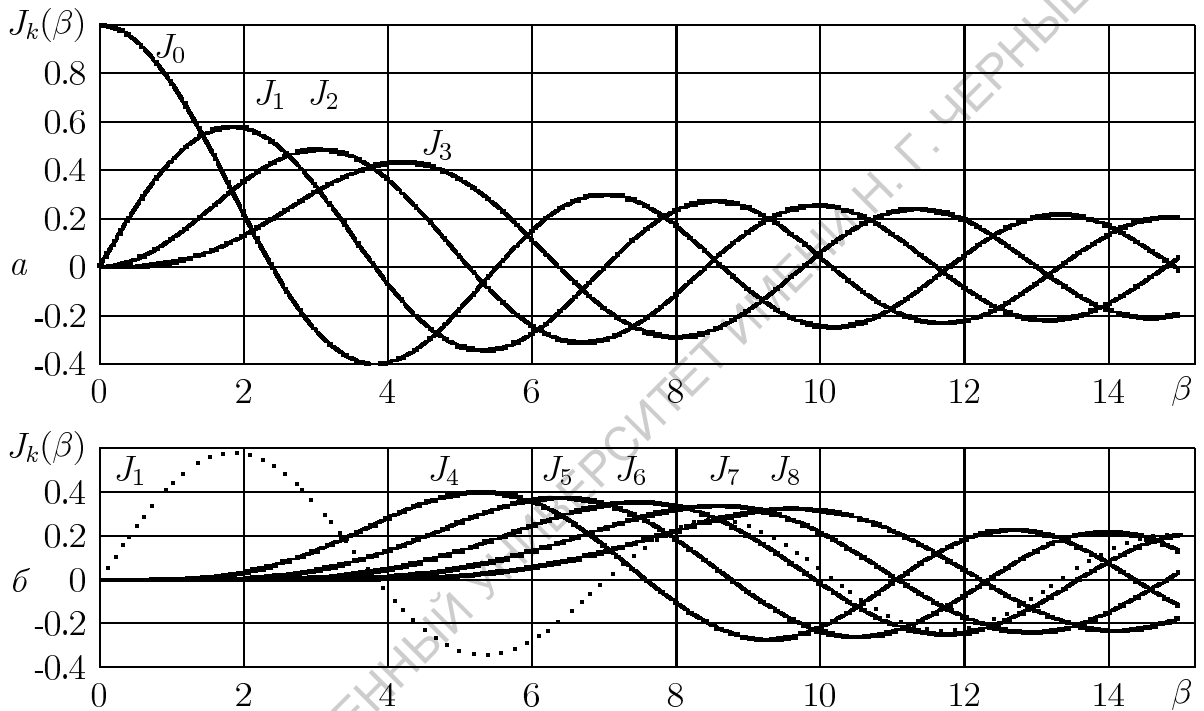


Рис. 5.11. Функции Бесселя нулевого – третьего порядка (а) и четвертого – восьмого порядка (б)

Интересно, что при больших значениях β существенную для расчетов величину имеют только те функции Бесселя, порядок которых k не превышает $\beta + 1$, т.е. реальная ширина спектра ЧМ-сигнала составляет $2(\beta + 1)\Omega = 2(\Delta\omega + \Omega) \simeq 2\Delta\omega = 2\beta\Omega$ и в 2β раз превышает ширину спектра аналогичного однотонового АМ-сигнала. Спектры двух ЧМ-сигналов при $\beta = 1$ и $\beta = 2.5$ представлены на рис. 5.12.

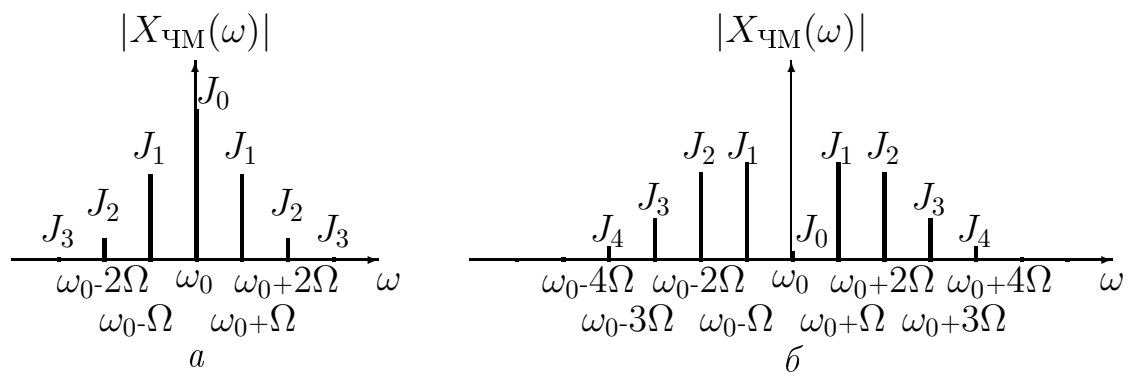


Рис. 5.12. Спектры ЧМ-сигналов с индексами модуляции $\beta = 1$ (а) и $\beta = 2.5$ (б)

Отметим, что амплитуда несущей может быть не только много меньше амплитуд боковых колебаний, но и при $\beta = 2.405, 5.52, \dots$ (см. табл. 5.1) вообще отсутствует в спектре однотонального ЧМ-сигнала.

Замечания. 1. Ширина спектра однотонального ЧМ-сигнала при больших индексах модуляции близка к удвоенной девиации частоты и не зависит от частоты модулирующего сигнала.

2. Одновременное присутствие многих составляющих, содержащих информацию о модулирующем сигнале, обеспечивает высокую достоверность при передаче сообщений и помехоустойчивость приема.

3. Необходимая широкополосность ЧМ-сигналов обеспечивается только в области метровых или более коротких волн. Поэтому ЧМ-сигналы (их проще создавать и демодулировать) используются при трансляции стереопередач в УКВ-диапазоне и звукового сопровождения телевизионных передач.

4. При демодуляции ЧМ-сигнал сначала преобразуется в АМ-сигнал с частотно-модулированной несущей и огибающей, пропорциональной $\Delta\omega s(t)$, а затем выделяется его низкочастотная огибающая.

В заключение рассмотрим еще один подкласс ЧМ-сигналов – радиоимпульсы с частотно-модулированным заполнением. Использование таких сигналов в радиолокационных системах способствует значительному повышению их чувствительности.

5.3.3. Сигналы с внутриимпульсной линейной ЧМ

Рассмотрим прямоугольный радиоимпульс с глубокой внутриимпульсной частотной модуляцией по линейному закону – ЛЧМ-импульс (рис. 5.13). Его текущая частота задается формулой

$$\omega(t) = \omega_0 + \frac{\Delta\omega}{\tau}t \quad |t| \leq \tau/2, \quad (5.33)$$

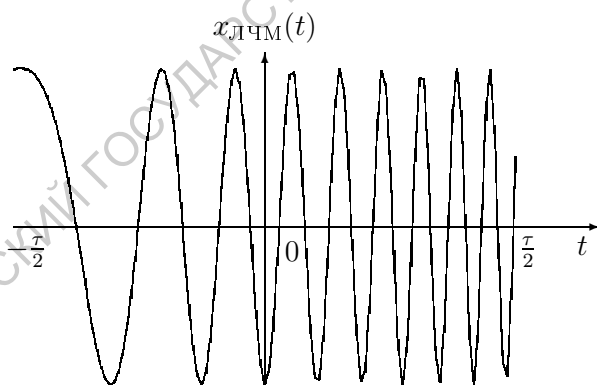


Рис. 5.13. Временная развертка ЛЧМ-импульса

где ω_0 – несущая частота; $\Delta\omega$ – девиация частоты; τ – длительность импульса. Наибольший интерес представляют ЛЧМ-импульсы с большой базой (см. разд. 2.4.2) $D = \Delta f\tau = \Delta\omega\tau/2\pi \gg 1$ ($D = 100$ и больше). Это пример сложных сигналов.

Мгновенные значения ЛЧМ-импульсов описываются формулой:

$$x_{\text{ЛЧМ}}(t) = A \cos \left(\int_0^t \omega(t) dt \right).$$

Полагая $\varphi_0 = 0$, окончательно получим:

$$x_{\text{ЛЧМ}}(t) = \begin{cases} 0, & t < -\tau/2, \\ A \cos \left(\omega_0 t + \frac{\Delta\omega}{2\tau} t^2 \right), & -\tau/2 \leq t \leq \tau/2, \\ 0, & t > \tau/2. \end{cases} \quad (5.34)$$

ЛЧМ-импульсы обладают замечательными свойствами. Во-первых, с помощью дисперсионной линии задержки, у которой $t_{\text{зад}}$ обратно пропорционально частоте поступающего сигнала, можно существенно «сжать» ЛЧМ-импульс во времени. Действительно, если начало импульса с частотой $\omega_0 - \frac{\Delta\omega}{2}$ задержать на 0.8τ и, постепенно уменьшая задержку, обеспечить $t_{\text{зад}} = 0.3\tau$ для середины импульса с частотой ω_0 , то можно получить результирующий импульс с длительностью порядка 0.2τ и тем самым увеличить амплитуду выходного сигнала и чувствительность приемной системы.

Вторая особенность ЛЧМ-сигналов состоит в структуре их спектров.

Спектральная плотность ЛЧМ-импульсов получается при вычислении интеграла Фурье от функции $x_{\text{ЛЧМ}}(t)$, заданной формулой (5.34):

$$\begin{aligned} X_{\text{ЛЧМ}}(\omega) &= A \int_{-\infty}^{\infty} \cos\left(\omega_0 t + \frac{\Delta\omega}{2\tau} t^2\right) e^{-j\omega t} dt = \\ &= \frac{A}{2} \int_{-\tau/2}^{\tau/2} e^{j[(\omega_0 - \omega)t + \frac{\Delta\omega}{2\tau} t^2]} dt + \frac{A}{2} \int_{-\tau/2}^{\tau/2} e^{-j[(\omega_0 + \omega)t + \frac{\Delta\omega}{2\tau} t^2]} dt. \end{aligned}$$

Первый интеграл представляет спектральную плотность с максимумом в области положительных частот, быстро осциллирующую в области отрицательных частот, а второй, напротив, имеет спектральный максимум в области отрицательных частот и быстро затухает в области положительных частот. Пренебрегая перекрытием этих спектров и ограничивая рассмотрение областью положительных частот, получим:

$$X_{\text{ЛЧМ}}(\omega) \cong \frac{A}{2} \int_{-\tau/2}^{\tau/2} e^{j[(\omega_0 - \omega)t + \frac{\Delta\omega}{2\tau} t^2]} dt. \quad (5.35)$$

Показатель этой экспоненты можно дополнить до полного квадрата и представить в виде $j\frac{\pi}{2}x^2 - \mu$, где μ – постоянная:

$$j[(\omega_0 - \omega)t + \frac{\Delta\omega}{2\tau} t^2] = j\frac{\Delta\omega}{2\tau} \left[t^2 + \frac{2\tau}{\Delta\omega} (\omega_0 - \omega)t + \frac{(\omega_0 - \omega)^2}{\Delta\omega^2} \tau^2 \right] - \frac{(\omega_0 - \omega)^2}{2\Delta\omega} \tau.$$

Обозначим $x = \sqrt{\frac{2\Delta f}{\tau}} \left(t + \frac{\omega_0 - \omega}{\Delta\omega} \tau \right)$. Тогда

$$X_{\text{ЛЧМ}}(\omega) \cong \frac{A}{2} \sqrt{\frac{\tau}{2\Delta f}} e^{-j\frac{(\omega_0 - \omega)^2}{2\Delta\omega} \tau} \int_{-x_1}^{x_2} e^{j\frac{\pi}{2}x^2} dx,$$

где $-x_1$ и x_2 соответствуют $-\tau/2$ и $\tau/2$, т.е.

$$x_1 = \sqrt{\frac{2\Delta f}{\tau}} \left(-\frac{\tau}{2} + \frac{\omega_0 - \omega}{\Delta\omega} \tau \right) = -\sqrt{\frac{D}{2}} \left(t + 2\frac{\omega - \omega_0}{\Delta\omega} \right), \quad (5.36)$$

$$x_2 = \sqrt{\frac{2\Delta f}{\tau}} \left(\frac{\tau}{2} + \frac{\omega_0 - \omega}{\Delta\omega} \tau \right) = \sqrt{\frac{D}{2}} \left(t - 2\frac{\omega - \omega_0}{\Delta\omega} \right), \quad (5.37)$$

$D = \Delta f \tau$ – база сигнала. Как известно,

$$\int_0^y e^{j\frac{\pi}{2}x^2} dx = \int_0^y \cos \frac{\pi}{2}x^2 dx + j \int_0^y \sin \frac{\pi}{2}x^2 dx = C(y) + jS(y),$$

где $C(y)$ – косинус-интеграл, а $S(y)$ – синус-интеграл Френеля.

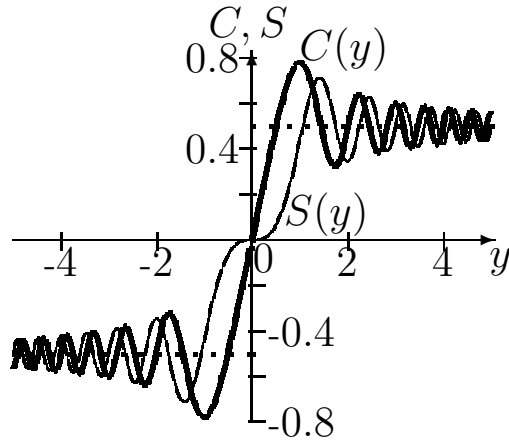


Рис. 5.14. Графики косинус- и синус-интегралов Френеля

Графики интегралов Френеля (рис. 5.14) представляют нечетные функции, т.е.

$$C(-y) = -C(y), \quad S(-y) = -S(y),$$

и с увеличением аргумента сначала быстро возрастают, а затем асимптотически стремятся к ± 0.5 (пунктирные прямые на рис. 5.14). Аргументы этих функций в соответствии с (5.36) и (5.37) являются линейными функциями частоты и в диапазоне $(\omega_0 - \Delta\omega/2 \div \omega_0 + \Delta\omega/2)$ изменяются линейно от 0 до $\sqrt{2D}$ (x_1) и от $\sqrt{2D}$ до 0 (x_2). Поэтому для больших значений D можно считать, что

$$C(y) = S(y) = -0.5 \quad \text{при } y < 0 \quad \text{и} \quad C(y) = S(y) = 0.5 \quad \text{при } y > 0.$$

Используя интегралы Френеля, получим для спектральной плотности ЛЧМ-импульса окончательное выражение

$$X_{\text{ЛЧМ}}(\omega) = \frac{A}{2} \sqrt{\frac{\tau}{2\Delta f}} e^{-j\frac{(\omega_0 - \omega)^2}{2\Delta\omega}\tau} [C(x_1) + C(x_2) + jS(x_1) + jS(x_2)]. \quad (5.38)$$

Если представить $X_{\text{ЛЧМ}}(\omega)$ в экспоненциальной форме

$$X_{\text{ЛЧМ}}(\omega) = |X_{\text{ЛЧМ}}(\omega)| e^{j\varphi(\omega)},$$

то для амплитудного и фазового спектров ЛЧМ-сигнала имеем:

$$|X_{\text{ЛЧМ}}(\omega)| = A \sqrt{\frac{\tau}{8\Delta f}} \sqrt{[C(x_1) + C(x_2)]^2 + [S(x_1) + jS(x_2)]^2}, \quad (5.39)$$

$$\varphi_{\text{ЛЧМ}}(\omega) = -\frac{(\omega_0 - \omega)^2}{2\Delta\omega}\tau + \text{arctg} \frac{S(x_1) + S(x_2)}{C(x_1) + C(x_2)}. \quad (5.40)$$

Графики зависимостей $|X_{\text{ЛЧМ}}(\omega)|$ и $\varphi_{\text{ЛЧМ}}(\omega)$ для двух значений D , рассчитанные по формулам (5.39) и (5.40), представлены на рис. 5.15 и показывают, что с увеличением базы D колебания амплитудного спектра ЛЧМ-сигнала уменьшаются, а его огибающая приближается к прямоугольной с шириной, равной девиации частоты $\Delta\omega$. Поскольку формула (5.40) состоит из двух слагаемых и первое слагаемое имеет вид квадратичной параболы, на графиках показано только второе слагаемое, которое при больших D постоянно в пределах $\Delta\omega$ и стремится к $\pi/4$.

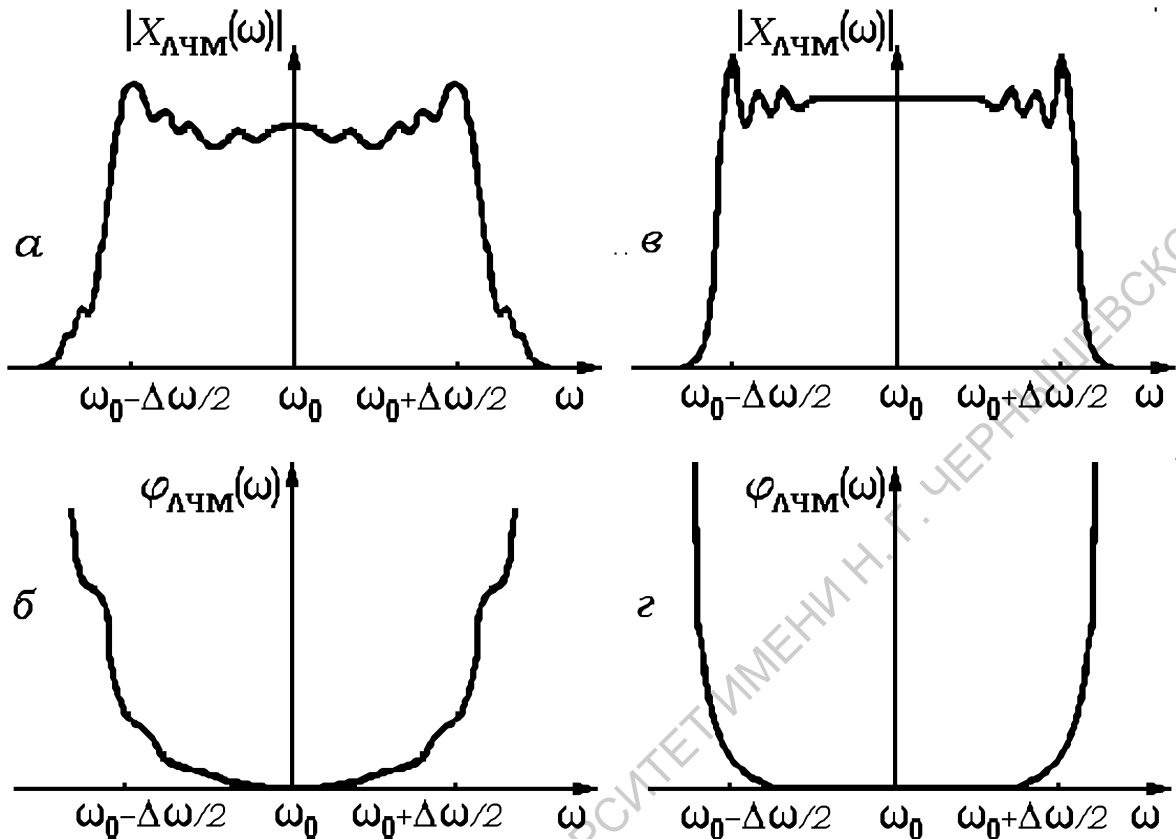


Рис. 5.15. Спектральные плотности. Амплитудные спектры ЛЧМ-сигналов при $D = 20$ (а), $D = 200$ (в) и осциллирующие составляющие фазовых спектров при $D = 20$ (б) и $D = 200$ (г)

Таким образом, ЛЧМ-сигналы при больших D обладают важной особенностью: огибающие спектров и временных реализаций идентичны. Как отмечалось выше, этим свойством обладают также гауссовы импульсы и периодические последовательности дельта-импульсов.

5.3.4. Частотно- и фазо-манипулированные сигналы

Частотно-манипулированный сигнал представляет гармоническое колебание, частота которого скачкообразно изменяется во времени в соответствии с заданным информационным сигналом. Коммутируемых частот может быть несколько, но чаще всего только две. Для упрощения анализа таких процессов используем в качестве модулирующего сигнала последовательность прямоугольных импульсов с периодом повторения T и длительностью импульсов $T/2$, ($q=2$) (рис. 5.16, а). Тогда

$$x(t) = \begin{cases} A \cos \omega_{\text{нес}1} t, & nT < t < (n + \frac{1}{2})T \quad (n - \text{целое число}), \\ A \cos \omega_{\text{нес}2} t, & (n + \frac{1}{2})T < t < (n + 1)T, \end{cases}$$

т.е. в один полупериод частота несущего колебания будет $\omega_{\text{нес}1}$, а в следующий — $\omega_{\text{нес}2}$ и т.д. (рис. 5.16, б). Этот сигнал можно представить суммой двух последовательностей прямоугольных радиоимпульсов (рис. 5.16, в, г) с различным высокочастотным заполнением.

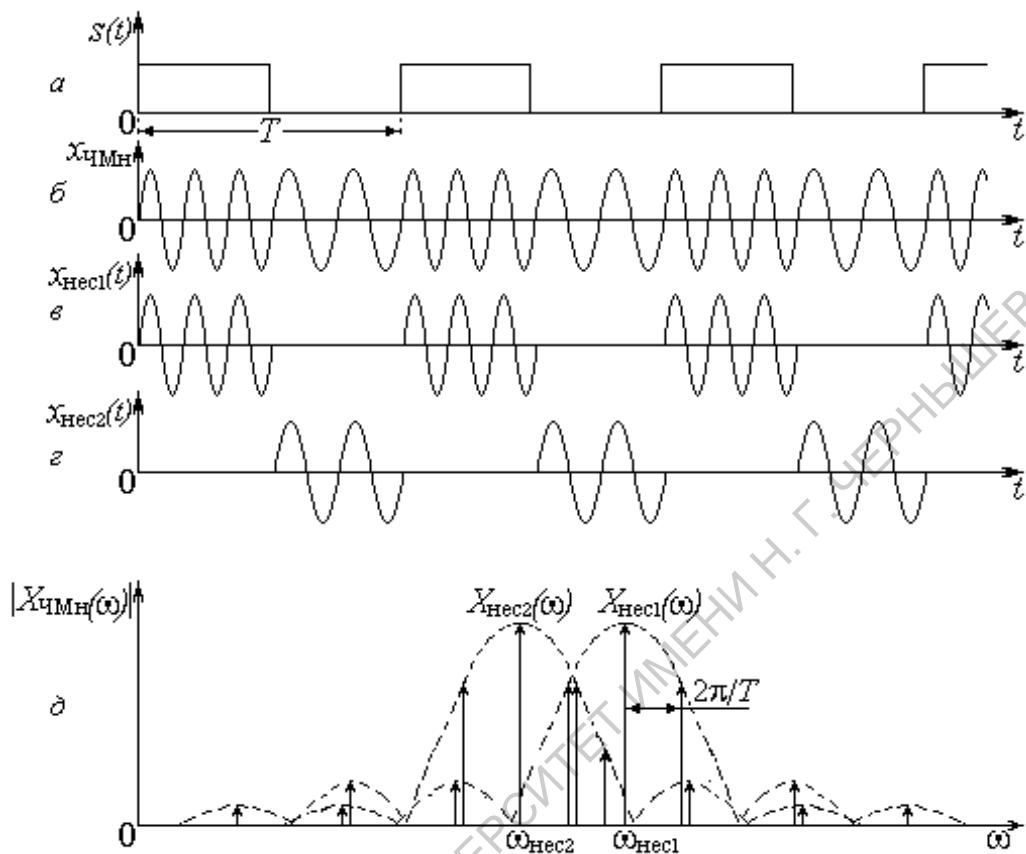


Рис. 5.16. Частотно-манипулированный сигнал, его составляющие и частотный спектр

Огибающие спектров будут идентичными (рис. 5.16, г), но сдвинутыми по частоте на $\Delta\omega = |\omega_{\text{нес1}} - \omega_{\text{нес2}}|$. Когда $\Delta\omega$ мала, спектры модулированного сигнала и последовательности радиоимпульсов практически совпадают, а ширина спектра пропорциональна частоте коммутации $2\pi/T$. Если же $\Delta\omega$ велика, то ширина спектра близка к $2\Delta\omega$. Таким образом, разность частот несущих в частотной манипуляции имеет такой же физический смысл, как девиация частоты в однотоновой частотной модуляции.

Представленный спектр является очень грубой моделью ЧМн-сигнала (хотя в таком виде часто встречается в литературе по радиоэлектронике). В общем случае длительности импульсов в модулирующих последовательностях изменяются случайным образом, и спектр ЧМн-сигнала (как и любого случайного процесса) становится сплошным. При переходе с одной частоты на другую фаза несущего колебания может претерпевать разрывы, вызывающие всплески интенсивности на частотах, кратных $2\pi/T$. Если же изменение фазы удастся сделать непрерывным⁵, то спектральная плотность представляет непрерывную функцию, а ее огибающая напоминает несколько расширенную огибающую немодулированной последовательности импульсов.

Теперь обратимся к фазовой манипуляции и в качестве модулирующего сигнала будем использовать ту же последовательность прямоуголь-

⁵Модуляцию с линейным нарастанием фазы иногда называют ЧМн с непрерывной фазой (continuous phase frequency shift keying – CPFSK).

ных импульсов с периодом повторения T и длительностью импульсов $T/2$ ($q=2$) (рис. 5.17, а). Такой ФМн-сигнал (рис. 5.17, б) удобно представить в комплексной экспоненциальной форме

$$z_{\text{ФМн}}(t) = Ae^{j[\omega t + \varphi(t)]} = \dot{A}(t)e^{j\omega t},$$

где $\dot{A}(t) = Ae^{j\pi(\sigma(t+nT) - \sigma(t+(2n-1)\frac{T}{2}))}$ – комплексная амплитуда (n – целое число); A , ω – амплитуда и частота несущего колебания; $\sigma(t)$ – функция включения Хевисайда (см. разд. 2.7, пример изображения прямоугольного импульса). Комплексную амплитуду ФМн-сигнала можно представить на векторной диаграмме: в первые полупериоды T вектор \dot{A} направлен, а во вторые – сонаправлен оси Ox .

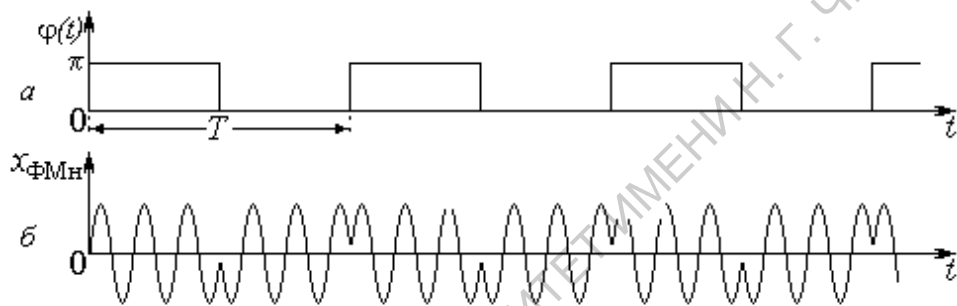


Рис. 5.17. Фазо-манипулированный сигнал

Длительности импульсов в модулирующих последовательностях, как и в случае частотной манипуляции, изменяются случайным образом. Спектр ФМн-сигнала становится сплошным и содержит всплески интенсивности, связанные со скачками фазы. Как показывают расчеты, спектральная плотность мощности ФМн-сигналов отлична от нуля в anomalously широкой полосе частот. При ограничении частот с помощью фильтрации возникает так называемая межсимвольная интерференция.

Фаза несущего колебания может изменяться и с другими интервалами ($\frac{3\pi}{2}$, $\frac{\pi}{2}$, $\frac{3\pi}{4}$, $\frac{\pi}{4}$, ...). Такая дискретизация используется при цифровой квадратурной модуляции и описана ниже.

5.4. Случайная модуляция гармонических сигналов

Колебания, параметры которых (амплитуда, частота, фаза) промодулированы случайным сигналом, типичны для задач радиофизики как в силу неизбежного присутствия случайных помех в системах передачи информации, так и по той причине, что модулирующие сигналы несут информационные сообщения и, таким образом, не могут быть детерминированными (предсказуемыми). Будем предполагать, что несущий сигнал является гармоническим, статистические характеристики случайного модулирующего сигнала известны, и рассмотрим статистические характеристики процесса, получающегося в результате случайной модуляции того или иного параметра гармонического сигнала.

5.4.1. Случайная амплитудная модуляция

Рассмотрим случайный процесс $X(t)$, представимый как

$$X(t) = A_0(1 + \xi(t)) \cos(\omega_0 t + \varphi_0), \quad (5.41)$$

где A_0, ω_0 – детерминированные константы; φ_0 – случайная величина, равномерно распределённая в интервале $[-\pi, \pi]$; $\xi(t)$ – стационарный случайный процесс, статистически не зависящий от φ_0 . Положим также, что $\bar{\xi}(t) \equiv 0$. Стационарность процесса $\xi(t)$ и равномерное распределение начальной фазы φ_0 обеспечивают стационарность промодулированных колебаний $X(t)$.

В том, что удовлетворяются требования стационарности в широком смысле, легко убедиться, получив выражения для среднего значения $\bar{X}(t)$ и ковариационной функции $K_X(t_1, t_2)$. С учетом статистической независимости $\xi(t)$ и φ_0 имеем:

$$\begin{aligned} \bar{X}(t) &= A_0 \langle (1 + \xi(t)) \rangle \langle \cos(\omega_0 t + \varphi_0) \rangle = \\ &= \frac{A_0}{2\pi} \int_{-\pi}^{\pi} \cos(\omega_0 t + \varphi_0) d\varphi_0 \equiv 0; \\ K_X(t_1, t_2) &= \langle X(t_1)X(t_2) \rangle = \frac{A_0^2}{2} (1 + \langle \xi(t_1)\xi(t_2) \rangle) \times \\ &\times \{ \langle \cos[\omega_0(t_1 + t_2) + 2\varphi_0] \rangle + \cos \omega_0(t_2 - t_1) \}. \end{aligned}$$

Так как

$$\langle \cos[\omega_0(t_1 + t_2) + 2\varphi_0] \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos[\omega_0(t_1 + t_2) + 2\varphi_0] d\varphi_0 \equiv 0,$$

то

$$K_X(t_1, t_2) = K_X(\tau) = \frac{A_0^2}{2} (1 + K_\xi(\tau)) \cos \omega_0 \tau, \quad \tau = t_2 - t_1, \quad (5.42)$$

где $K_\xi(\tau)$ – ковариационная функция модулирующего сигнала.

Спектральная плотность мощности процесса $X(t)$ есть

$$\begin{aligned} W_X(\omega) &= \int_{-\infty}^{\infty} K_X(\tau) e^{-j\omega\tau} d\tau = \\ &= \frac{A_0^2}{2} \left[\int_{-\infty}^{\infty} e^{-j\omega\tau} \cos \omega_0 \tau d\tau + \int_{-\infty}^{\infty} K_\xi(\tau) e^{-j\omega\tau} \cos \omega_0 \tau d\tau \right]. \end{aligned}$$

С учетом формулы Эйлера $\cos \omega_0 \tau = \frac{1}{2} (e^{-j\omega_0 \tau} + e^{j\omega_0 \tau})$ получим

$$\begin{aligned} W_X(\omega) &= \frac{A_0^2}{4} \left[\int_{-\infty}^{\infty} e^{-j(\omega+\omega_0)\tau} d\tau + \int_{-\infty}^{\infty} e^{-j(\omega-\omega_0)\tau} d\tau + \right. \\ &\left. + \int_{-\infty}^{\infty} K_\xi(\tau) e^{-j(\omega+\omega_0)\tau} d\tau + \int_{-\infty}^{\infty} K_\xi(\tau) e^{-j(\omega-\omega_0)\tau} d\tau \right]. \end{aligned}$$

Используя представление δ -функции $\delta(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\pm j\omega\tau} d\tau$ и соотношение

$$W_{\xi}(\omega) = \int_{-\infty}^{\infty} K_{\xi}(\tau) e^{-j\omega\tau} d\tau,$$

приходим к окончательному выражению

$$W_X(\omega) = \frac{A_0^2}{4} [2\pi\delta(\omega + \omega_0) + 2\pi\delta(\omega - \omega_0) + W_{\xi}(\omega + \omega_0) + W_{\xi}(\omega - \omega_0)], \quad (5.43)$$

где $W_{\xi}(\omega)$ – спектр модулирующего сигнала. Односторонний (физический) спектр процесса $X(t)$ есть

$$W_X^+(\omega) = \frac{A_0^2}{2} [2\pi\delta(\omega - \omega_0) + W_{\xi}(\omega - \omega_0)]. \quad (5.44)$$

Он симметричен относительно несущей частоты ω_0 и имеет две компоненты: дискретную $A_0^2\pi\delta(\omega - \omega_0)$ (δ -пик на несущей частоте) и непрерывную $-\frac{A_0^2}{2}W_{\xi}(\omega - \omega_0)$, которая повторяет спектр модулирующего сигнала.

Обозначим $A_0(1 + \xi(t))$ за $\rho(t)$, а полную фазу $\omega_0 t + \varphi_0$ – за $\Phi(t)$. Тогда $X(t) = \rho(t) \cos \Phi(t)$. Введем процесс $Y(t) = \rho(t) \sin \Phi(t)$. Пусть переменные ρ , Φ и X , Y связаны детерминированными преобразованиями

$$X = \rho \cos \Phi, \quad Y = \rho \sin \Phi. \quad (5.45)$$

Для совместной плотности вероятности имеем

$$p_1^{XY}(x, y) = \frac{p_1^{\rho\Phi}(\rho, \Phi)}{\rho},$$

где $1/\rho$ – якобиан преобразования (5.45). Учтем статистическую независимость ρ и Φ и равномерное распределение φ_0 . Тогда

$$p_1^{XY}(x, y) = \frac{p_1^{\rho}(\rho)}{2\pi\rho}.$$

Для плотности вероятности $p_1(x)$ переменной $X(t)$ получаем

$$\begin{aligned} p_1^X(x) &= \int_{-\infty}^{\infty} p_1^{XY}(x, y) dy = \frac{1}{\pi} \int_0^{\infty} \frac{p_1^{\rho}(\rho) d\rho}{\rho} \Big|_{\rho=\sqrt{x^2+y^2}} = \\ &= \frac{1}{\pi} \int_{|x|}^{\infty} \frac{p_1^{\rho}(\rho) d\rho}{\sqrt{\rho^2 - x^2}}. \end{aligned} \quad (5.46)$$

Таким образом, распределение переменной X определяется распределением случайной амплитуды $\rho(t) = A_0(1 + \xi(t))$. Можно показать⁶, что в случае релеевского распределения ρ

$$p_1^{\rho}(\rho) = \frac{\rho}{\lambda} e^{-\frac{\rho^2}{2\lambda}}, \quad \lambda = \text{const} > 0,$$

колебания $X(t)$ имеют гауссовское распределение:

⁶Рекомендуется проделать самостоятельно.

$$p_1^X(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{x^2}{2\lambda}}, \quad \lambda = \text{const} > 0.$$

В качестве примера рассмотрим случайный процесс, возникающий при передаче речевых сигналов.

Частотное распределение мощности речевого сигнала для русской и английской речи изображено на рис. 5.18 и внешне напоминает распределение Релея⁷. Уровень мощности представлен в dB SPL (Sound Power Level – уровень звуковой мощности) в дБ относительно уровня различимости. Нормальная громкость разговора соответствует 50 ÷ 60 dB SPL. Как показали исследования, низкие частоты до 300 Гц отвечают за разборчивость голоса и несут всего 3–5% информации. Максимум энергии приходится на диапазон от 300 до 2000 Гц, где содержится около 40% передаваемой информации. Частоты выше 1500 Гц характеризуют четкость и разборчивость речи и содержат около 60% всей передаваемой информации.

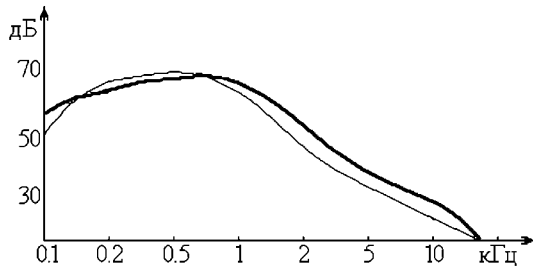


Рис. 5.18. Огибающая спектра мощности русской и английской речи (жирная кривая)

Таким образом, принятое МЭК ограничение полосы частот телефонных разговоров частотой 3400 Гц вполне согласуется с результатами таких исследований.

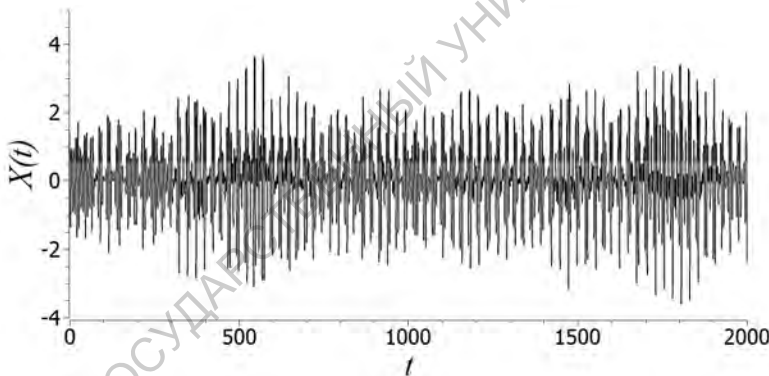


Рис. 5.19. Вид реализации случайной амплитудной модуляции гармонического сигнала гауссовским процессом со спектром (5.47)

На рис. 5.19 приведена численно смоделированная реализация процесса $X(t)$ в виде двумерного процесса Орнштейна – Уленбека (см. разд. 3.6.5), соответствующего амплитудной модуляции гармонического колебания стационарным гауссовским сигналом $\xi(t)$ со спектром

$$W_\xi(\omega) = \alpha\sigma_\xi^2 \left[\frac{1}{\alpha^2 + (\omega + \omega_1)^2} + \frac{1}{\alpha^2 + (\omega - \omega_1)^2} \right], \quad (5.47)$$

где α – параметр модулирующего сигнала, задающий ширину его спектра. Спектры и одномерные плотности вероятности сигналов $\xi(t)$ и $X(t)$ изображены на рис. 5.20.

⁷Первые исследования, по-видимому, были отражены в акустическом стандарте США (USASI-United States Acoustic Standart Interference).

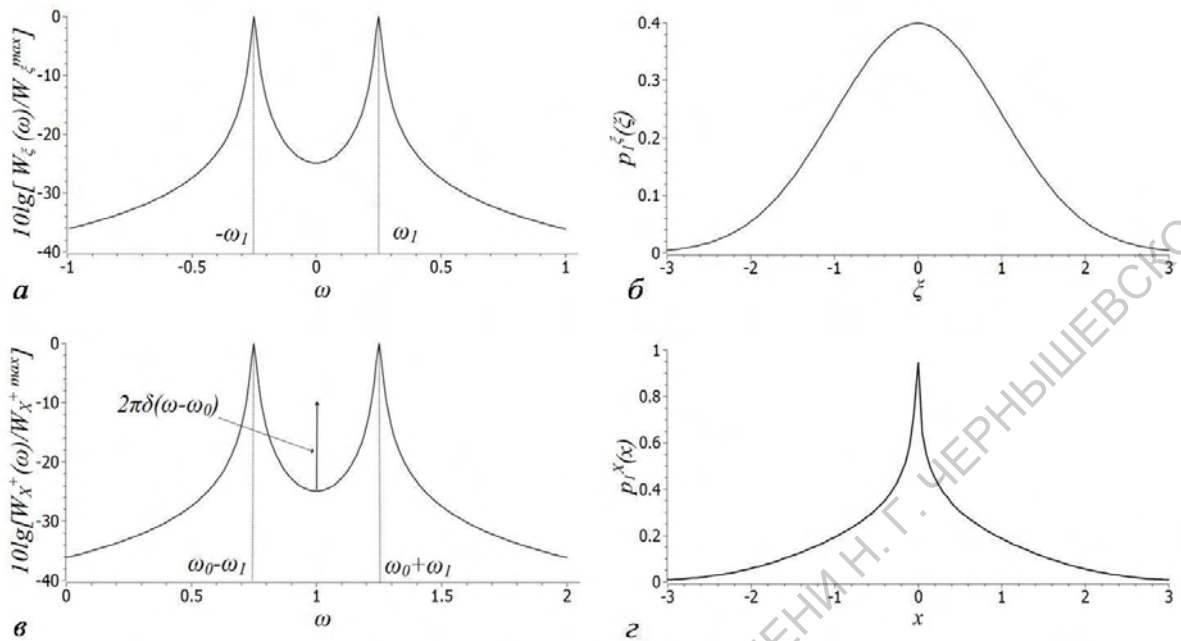


Рис. 5.20. Характеристики амплитудной модуляции гауссовским процессом со спектром (5.47): нормированный спектр (а) и однородная плотность вероятности (б) модулирующего сигнала $\xi(t)$; нормированный односторонний спектр (в) и одномерная плотность вероятности (г) процесса $X(t)$ (получена численно). Параметры сигналов: $A_0 = 1$, $\omega_0 = 1$, $\sigma_\xi^2 = 1$, $\alpha = 0.01$, $\omega_1 = 0.25$

5.4.2. Случайная угловая модуляция

Угловая модуляция, как говорилось ранее, применительно к детерминированному случаю может быть фазовой и частотной.

Случайная модуляция фазы. Для случайной фазовой модуляции имеем

$$X(t) = A_0 \cos(\omega_0 t + \varphi(t)), \quad (5.48)$$

где A_0 , ω_0 – детерминированные константы, задающие амплитуду и частоту сигнала соответственно. Слагаемое $\varphi(t)$ представляет собой случайную компоненту фазы и может рассматриваться как модулирующий сигнал при случайной фазовой модуляции.

Нас интересует прежде всего такая случайная модуляция фазы, которая соответствует стационарному процессу $X(t)$. Для среднего значения $\bar{X}(t)$ процесса (5.48) легко получить

$$\bar{X}(t) = A_0 \langle \cos \varphi(t) \rangle \cos \omega_0 t - A_0 \langle \sin \varphi(t) \rangle \sin \omega_0 t.$$

Очевидно, что промодулированный по фазе сигнал может быть стационарным процессом только в том случае, если $\langle \cos \varphi(t) \rangle \equiv 0$ и $\langle \sin \varphi(t) \rangle \equiv 0$. Но и равенство нулю этих средних значений еще не обеспечивает стационарности $X(t)$. Так, для ковариационной функции промодулированного сигнала имеем

$$\begin{aligned}
K_X(t, t + \tau) &= \frac{A_0^2}{2} [\cos 2\omega_0 t < \cos \varphi(2\varphi(t) + \Delta\varphi) > - \\
&\quad - \sin 2\omega_0 t < \sin \varphi(2\varphi(t) + \Delta\varphi) > + \\
&\quad + \cos \omega_0 \tau < \cos \Delta\varphi > - \sin \omega_0 \tau < \sin \Delta\varphi >],
\end{aligned}$$

где $\Delta\varphi = \varphi(t + \tau) - \varphi(t)$. В стационарном случае она должна зависеть только от переменной τ , что накладывает дополнительные требования на характеристики процесса $\varphi(t)$.

Можно показать, что промодулированный сигнал становится стационарным, если в качестве модулирующего сигнала $\varphi(t)$ выбрать винеровский процесс. В этом случае $\varphi(t)$ имеет гауссовское распределение со средним значением $\bar{\varphi} = \varphi(0) = \varphi_0$ и дисперсией $\sigma_\varphi^2(t) = 2B_\varphi \cdot t$ ($t \geq 0$), где B_φ – коэффициент диффузии фазы. Начальную фазу φ_0 считаем детерминированной константой. Из свойств винеровского процесса следует, что распределения величин $\Delta\varphi$ и $\theta = 2\varphi_0 + \Delta\varphi$ являются гауссовскими со следующими средними значениями и дисперсиями: $\langle \Delta\varphi \rangle = 0$, $\sigma_{\Delta\varphi}^2 = 2B_\varphi|\tau|$, $\langle \theta \rangle = 2\varphi_0$, $\sigma_\theta^2 = 8B_\varphi \cdot t_m + 2B_\varphi|\tau|$. Здесь переменная t_m совпадает с t при $\tau > 0$ и равна $t + \tau$, если $\tau < 0$.

С учетом распределений величин $\varphi(t)$, $\Delta\varphi$ и θ получаем:

$$\langle \cos \varphi(t) \rangle = \frac{1}{\sqrt{2\pi\sigma_\varphi^2}} \int_{-\infty}^{\infty} \cos \varphi \exp \left[-\frac{(\varphi - \varphi_0)^2}{2\sigma_\varphi^2} \right] d\varphi = \cos \varphi_0 \exp [-B_\varphi \cdot t],$$

$$\langle \sin \varphi(t) \rangle = \frac{1}{\sqrt{2\pi\sigma_\varphi^2}} \int_{-\infty}^{\infty} \sin \varphi \exp \left[-\frac{(\varphi - \varphi_0)^2}{2\sigma_\varphi^2} \right] d\varphi = \sin \varphi_0 \exp [-B_\varphi \cdot t],$$

$$\begin{aligned}
\langle \cos (2\varphi + \Delta\varphi) \rangle &= \langle \cos \theta \rangle = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \int_{-\infty}^{\infty} \cos \theta \exp \left[-\frac{(\theta - 2\varphi_0)^2}{2\sigma_\theta^2} \right] d\theta = \\
&= \cos 2\varphi_0 \exp [-4B_\varphi \cdot t_m] \exp [-B_\varphi|\tau|],
\end{aligned}$$

$$\begin{aligned}
\langle \sin (2\varphi + \Delta\varphi) \rangle &= \langle \sin \theta \rangle = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \int_{-\infty}^{\infty} \sin \theta \exp \left[-\frac{(\theta - 2\varphi_0)^2}{2\sigma_\theta^2} \right] d\theta = \\
&= \sin 2\varphi_0 \exp [-4B_\varphi \cdot t_m] \exp [-B_\varphi|\tau|],
\end{aligned}$$

$$\langle \cos \Delta\varphi \rangle = \frac{1}{\sqrt{2\pi\sigma_{\Delta\varphi}^2}} \int_{-\infty}^{\infty} \cos \Delta\varphi \exp \left[-\frac{(\Delta\varphi)^2}{2\sigma_{\Delta\varphi}^2} \right] d(\Delta\varphi) = \exp [-B_\varphi|\tau|],$$

$$\langle \sin \Delta\varphi \rangle = \frac{1}{\sqrt{2\pi\sigma_{\Delta\varphi}^2}} \int_{-\infty}^{\infty} \sin \Delta\varphi \exp \left[-\frac{(\Delta\varphi)^2}{2\sigma_{\Delta\varphi}^2} \right] d(\Delta\varphi) \equiv 0.$$

Установившийся режим соответствует пределу $t, t_m \rightarrow \infty$. При этом получаем, что $\langle \cos \varphi(t) \rangle \equiv 0$, $\langle \sin \varphi(t) \rangle \equiv 0$, $\langle \cos (2\varphi + \Delta\varphi) \rangle \equiv 0$ и

$\langle \sin(2\varphi + \Delta\varphi) \rangle \equiv 0$. Таким образом, $\overline{X}(t) \equiv 0$, а для ковариационной (корреляционной) функции справедливо выражение

$$K_X(\tau) = \frac{A_0^2}{2} e^{-B_\varphi|\tau|} \cos \omega_0 \tau. \quad (5.49)$$

Среднее значение промодулированного сигнала в установившемся режиме равно нулю, корреляционная функция зависит только от интервала τ , а дисперсия $\sigma_X^2 = \overline{X^2} = \frac{A_0^2}{2}$ конечна, т.к. $A_0 \neq \infty$. Следовательно, процесс $X(t)$ является стационарным, по крайней мере в широком смысле.

Ковариационной функции с экспоненциально убывающей огибающей соответствует односторонний спектр в виде «сдвинутых» лоренцианов с максимумами на частотах $\pm\omega_0$:

$$W_X(\omega) = \frac{A_0^2 B_\varphi}{2} \left[\frac{1}{B_\varphi^2 + (\omega - \omega_0)^2} + \frac{1}{B_\varphi^2 + (\omega + \omega_0)^2} \right]. \quad (5.50)$$

Время корреляции и ширина спектра процесса $X(t)$ определяются коэффициентом диффузии фазы B_φ . Если значение B_φ мало ($B_\varphi \ll \omega_0$), то процесс (5.48) можно считать гармоническим шумом с временем корреляции $\tau = \frac{1}{B_\varphi}$ и шириной спектра на уровне половинной мощности $\Delta\omega_{0.5} = 2B_\varphi$. В этом случае односторонний спектр имеет вид:

$$W_X^+(\omega) = \frac{A_0^2 B_\varphi}{B_\varphi^2 + (\omega - \omega_0)^2}. \quad (5.51)$$

Реализация сигнала (5.48), промодулированного по фазе винеровским процессом $\varphi(t)$, и его односторонний спектр (5.51) приведены на рис. 5.21.

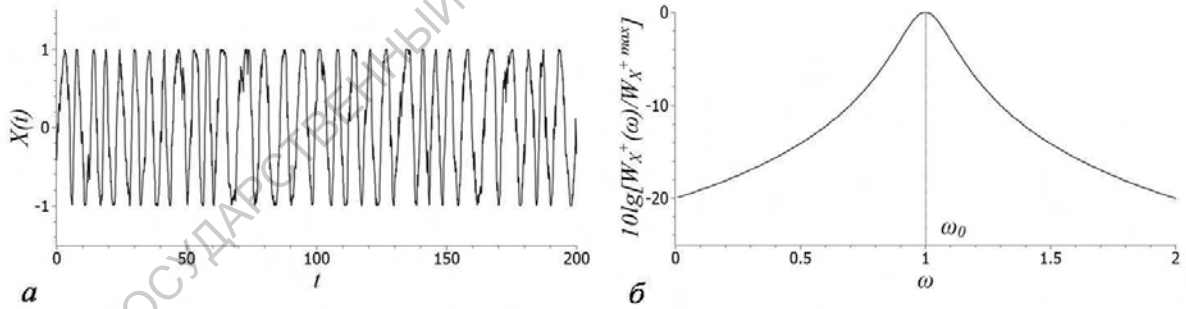


Рис. 5.21. Фазовая модуляция гармонического сигнала винеровским процессом: *a* – реализация промодулированного сигнала (численное моделирование); *б* – нормированный односторонний спектр промодулированного сигнала, построенный по формуле (5.51). Параметры сигналов: $A_0 = 1$, $\omega_0 = 1$, $B_\varphi = 0.1$

При представлении фазо-модулированных колебаний (5.48) переменную $\varphi(t) \in (-\infty; \infty)$ можно заменить на случайную фазу $\phi(t)$, принимающую значения в ограниченном интервале $[-\pi; \pi]$ или $[0; 2\pi]$:

$$\phi(t) = \varphi(t) + 2\pi\nu(t), \quad (5.52)$$

где ν – некоторая целочисленная случайная функция. В силу 2π -периодичности косинуса процесс $X(t)$ от этого не изменится, однако использовать ограниченную переменную $\phi(t)$ в некоторых случаях оказывается удобнее. Если дисперсия фазы $\sigma_\varphi^2(t)$ неограниченно растет во времени (как в случае винеровского процесса), то приведённая фаза ϕ имеет в пределе $t \rightarrow \infty$ стационарное равномерное распределение

$\lim_{t \rightarrow \infty} p_1^\phi(\phi, t) = p_1^\phi(\phi) = 1/2\pi$. В силу функциональной зависимости текущих значений ϕ и $x = f(\phi)$ имеем:

$$p_1^X(x)|dx| = p_1^\phi(\phi_1)|d\phi_1| + p_1^\phi(\phi_2)|d\phi_2|,$$

где $\phi_{1,2} = \pm \arccos\left(\frac{x}{a_0}\right) - \omega_0 t$ – две ветви обратной функции $\phi = g(x)$ в интервале значений $\phi \in [-\pi; \pi]$. Отсюда легко получить одномерную плотность вероятности процесса $X(t)$:

$$p_1^X(x) = \frac{1}{\pi \sqrt{A_0^2 + x^2}}. \quad (5.53)$$

Случайная модуляция частоты. Рассмотрим гармоническую несущую со случайной модуляцией частоты:

$$X(t) = A_0 \cos \left[\int_0^t (\omega_0 + \eta(\theta)) d\theta \right], \quad (5.54)$$

где амплитуда сигнала A_0 – детерминированная константа, а мгновенная частота $\omega(t)$ состоит из двух компонент: постоянной частоты ω_0 и переменной частоты, представляющей собой сигнал случайной модуляции $\eta(t)$. Полная мгновенная фаза $\Phi(t)$ также состоит из двух слагаемых: детерминированной фазы $\omega_0 t$ и случайной фазы $\varphi(t) = \int_0^t \eta(\theta) d\theta$. Таким образом, мы приходим к ранее рассмотренному случаю фазовой модуляции, но теперь случайный сигнал с заданными характеристиками входит под знак интеграла.

Пусть $\eta(t)$ – стационарный случайный процесс с дисперсией $\sigma_\eta^2 = \overline{\eta^2}$ (полагаем $\overline{\eta} \equiv 0$). Для дисперсии случайной фазы имеем

$$\begin{aligned} \sigma_\varphi^2(t) &= \overline{\varphi^2}(t) = \left\langle \int_0^t \int_0^t \eta(\theta_1) \eta(\theta_2) d\theta_1 d\theta_2 \right\rangle = \\ &= \int_0^t \int_0^t K_\eta(\theta_2 - \theta_1) d\theta_1 d\theta_2 = 2t \int_0^t K_\eta(\tau) d\tau - 2 \int_0^t \tau K_\eta(\tau) d\tau, \end{aligned}$$

где $\tau = \theta_2 - \theta_1$, а $K_\eta(\tau)$ – ковариационная функция процесса $\varphi(t)$. (Она совпадает с корреляционной функцией, поскольку $\overline{\varphi} \equiv 0$.) Дисперсия случайной фазы неограниченно растет с ростом t , и кривая вероятностного распределения $p_1^\varphi(\varphi)$ для выбранного t становится всё более широкой. Это означает, что для ограниченной фазы $\phi(t) \in [-\pi, \pi]$ в итоге устанавливается равномерное стационарное распределение.

Пусть функция $K_\eta(\tau)$ достаточно быстро убывает с ростом τ , так, что

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \tau K_\eta(\tau) d\tau = 0.$$

Тогда в пределе $t \rightarrow \infty$ получаем диффузионный закон роста дисперсии фазы:

$$\sigma_\varphi^2(t) \approx 2B_\varphi \cdot t,$$

где

$$B_\varphi \cong \int_0^\infty K_\eta(\tau) d\tau = \pi W_\eta(0) = \text{const}$$

есть коэффициент диффузии фазы, а $W_\eta(0)$ – спектральная плотность мощности сигнала $\eta(t)$ на нулевой частоте. Таким образом, на больших временах широкополосный сигнал $\eta(t)$ (с малым временем корреляции) можно заменить на соответствующий белый шум. В этом случае, при гауссовском распределении $\eta(t)$, для $\varphi(t)$ приходим к модели винеровского процесса. Характеристики сигнала со случайной фазой, задаваемой винеровским процессом, уже были нами рассмотрены.

5.5. Узкополосные процессы и их спектры

Узкополосными называются квазигармонические сигналы, частотные спектры которых сконцентрированы в узкой полосе частот вокруг некоторой частоты ω_0 , удалённой от начала координат:

$$x(t) = A(t) \cos(\omega_0 t + \varphi(t)) = A(t) \cos \Phi(t). \quad (5.55)$$

С узкополосными сигналами работают почти все реальные радиоэлектронные системы. Предложены различные идеализированные модели, но проблема достоверного описания узкополосных сигналов остается актуальной. Дело в том, что исследователю более или менее достоверно известны лишь границы спектра, а распределение амплитуд или мощности спектральных компонент является случайной функцией времени и подлежит определению статистическими методами. В этих условиях необходимо установить опорную частоту сигнала и найти статистические характеристики $x(t)$ или $A(t)$ и $\Phi(t)$. Сведения о структуре функций $A(t)$ и $\Phi(t)$ представляют особую ценность при получении так называемых низкочастотных представлений.

Для идеализированного АМ-радиосигнала отыскание низкочастотного представления, т.е. демодуляция сообщения, сводится к измерению огибающей сигнала. Когда АМ-радиосигнал претерпевает искажения в среде распространения, например из-за дисперсии, то восстановить низкочастотную огибающую простыми способами не удастся. Следует отметить, что выявление низкочастотных огибающих узкополосных случайных сигналов весьма полезны при решении многих задач радиоэлектроники и статистической радиофизики: описание колебаний в автогенераторах и лазерах, статистическое обоснование ширины спектральных линий и т.п.

Рассмотрение моделей узкополосных процессов и методов получения низкочастотных представлений целесообразно начать с регулярных (детерминированных) сигналов, спектральные характеристики которых заданы точно. Это позволяет на простых примерах описать различные свойства узкополосных процессов и ввести математический аппарат, необходимый для анализа характеристик других узкополосных процессов.

5.5.1. Вещественные модели детерминированных узкополосных сигналов

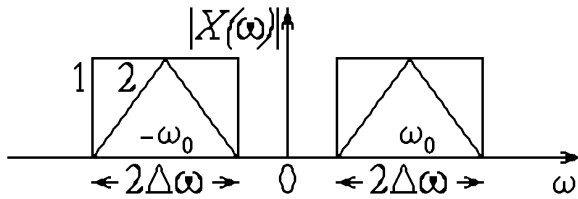


Рис. 5.22. Два распределения $X(\omega)$: 1 – $X(\omega) = \text{const}$, 2 – $X(\omega)$ – треугольное

Пусть задан детерминированный сигнал $x(t)$, спектральные компоненты которого сосредоточены в полосе частот от ω_1 до ω_2 причём $2\Delta\omega = \omega_2 - \omega_1 \ll \omega_0 = (\omega_1 + \omega_2)/2$, т.е. спектр концентрируется вокруг ω_0 , удалённой от начала координат. Распределения спектральной плотности узкополосных сигналов представлены на рис. 5.22.

Мгновенные значения соответствующей функции $x(t)$ нетрудно найти, используя обратное преобразование Фурье:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega. \quad (5.56)$$

При этом на распределение спектральной плотности $X(\omega)$ не накладывается никаких ограничений.

Пусть комплексная спектральная плотность сигнала постоянна в узкой полосе частот (см. рис. 5.22) и удовлетворяет соотношению:

$$X(\omega) = \begin{cases} 0, & \omega < -\omega_0 - \Delta\omega, \\ X_0, & -\omega_0 - \Delta\omega < \omega < -\omega_0 + \Delta\omega, \\ 0, & -\omega_0 + \Delta\omega < \omega < \omega_0 - \Delta\omega, \\ X_0, & \omega_0 - \Delta\omega < \omega < \omega_0 + \Delta\omega, \\ 0, & \omega > \omega_0 + \Delta\omega. \end{cases} \quad (5.57)$$

Математическая модель сигнала, обладающего спектральной плотностью (5.57), принимает вид:

$$x(t) = \frac{1}{2\pi} \int_{-\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} X(\omega) e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{-\omega_0 - \Delta\omega}^{-\omega_0 + \Delta\omega} X(\omega) e^{j\omega t} d\omega + \frac{1}{2\pi} \int_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} X(\omega) e^{j\omega t} d\omega. \quad (5.58)$$

Переходя к вещественной плотности и объединяя экспоненциальные функции по формуле Эйлера, получим:

$$\begin{aligned} x(t) &= \frac{X_0}{\pi} \int_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} \cos \omega t d\omega = \frac{X_0}{\pi} \frac{\sin \omega t}{t} \Big|_{\omega_0 - \Delta\omega}^{\omega_0 + \Delta\omega} = \frac{X_0}{\pi t} [\sin(\omega_0 + \Delta\omega)t - \sin(\omega_0 - \Delta\omega)t] = \\ &= \frac{2X_0\Delta\omega}{\pi} \frac{\sin \Delta\omega t}{\Delta\omega t} \cos \omega_0 t = \frac{2X_0\Delta\omega}{\pi} \text{sinc}(\Delta\omega t) \cos \omega_0 t. \end{aligned} \quad (5.59)$$

Колебание $x(t)$ представляет временную развертку детерминированного узкополосного сигнала

$$x(t) = A(t) \cos \omega_0 t, \quad A(t) = \frac{2X_0\Delta\omega}{\pi} \text{sinc}(\Delta\omega t) = A_0 \text{sinc}(\Delta\omega t).$$

Выполняя аналогичные вычисления для треугольного распределения спектральной плотности, нетрудно получить

$$x_1(t) = \frac{X_0 \Delta \omega}{\pi} \operatorname{sinc}^2(\Delta \omega t / 2) \cos \omega_0 t. \quad (5.60)$$

Эти колебания определены для $-\infty < t < \infty$ и имеют сложную структуру. На рис. 5.23 представлены их временные развертки.

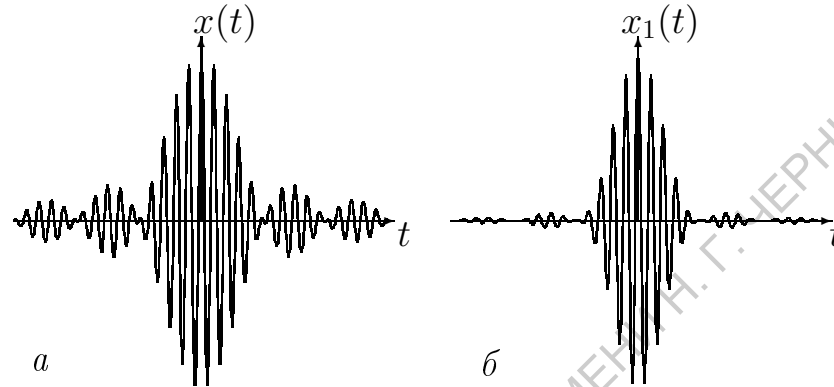


Рис. 5.23. Временные развертки узкополосных сигналов $x(t)$ (а) и $x_1(t)$ (б)

Таким образом, можно сформулировать два вывода.

1. Временные развертки узкополосных сигналов с одинаковыми $\Delta \omega$ и средней частотой ω_0 различаются из-за особенностей $X(\omega)$. Изображения разверток на рис. 5.23 имеют упрощенную форму и приведены только для того, чтобы продемонстрировать эту зависимость. Если рассмотреть всего $100 \div 200$ периодов несущего колебания, то при $\Delta \omega / \omega_0 \cong 10^{-4}$ вариации амплитуды колебаний $x(t)$ и $x_1(t)$ окажутся незаметными, что свидетельствует о медленности ее изменения.

2. Амплитуды узкополосных процессов $A(t) = \frac{2X_0 \Delta \omega}{\pi} \operatorname{sinc}(\Delta \omega t)$ и $A_1(t) = \frac{X_0 \Delta \omega}{\pi} \operatorname{sinc}^2(\Delta \omega t / 2)$ есть низкочастотные представления узкополосных сигналов.

Теперь рассмотрим обратную задачу. Пусть $A(t)$ – узкополосный низкочастотный сигнал. Его спектр сосредоточен в окрестности $\omega = 0$. При умножении $A(t)$ на $\cos \omega_0 t$ или на $\sin \omega_0 t$, где ω_0 – высокая по сравнению со всеми гармониками $A(t)$ частота, низкочастотный спектр $A(t)$ в соответствии с теоремой о смещении перемещается в область высоких частот, а колебания

$$x_1(t) = A(t) \cos \omega_0 t \quad \text{и} \quad x_2(t) = A(t) \sin \omega_0 t$$

в равной степени приобретают признаки узкополосного сигнала, отличаясь лишь начальной фазой высокочастотного колебания. Наиболее общее выражение можно получить, используя линейную комбинацию последних. Тогда

$$x(t) = A_c(t) \cos \omega_0 t - A_s(t) \sin \omega_0 t. \quad (5.61)$$

Колебания $A_c(t) \cos \omega_0 t$ и $A_s(t) \sin \omega_0 t$ называются *квадратурными компонентами* узкополосного сигнала, а функции $A_c(t)$ и $A_s(t)$ – его *синфазной* и *квадратурной* амплитудами. Сопоставляя (5.61) с (5.55), получим

$$A_c(t) = A(t) \cos \varphi(t), \quad A_s(t) = A(t) \sin \varphi(t).$$

Для реального узкополосного сигнала функции $A_c(t)$ и $A_s(t)$ можно определить экспериментально, используя вспомогательные сигналы $\cos \omega_0 t$, $\sin \omega_0 t$, перемножение сигналов и фильтрацию нижних частот (ФНЧ). Тогда

$$\begin{aligned} x(t) \cos \omega_0 t &= A_c(t) \cos^2 \omega_0 t - A_s(t) \sin \omega_0 t \cos \omega_0 t = \\ &= \frac{A_c(t)}{2} + \frac{A_c(t)}{2} \cos 2\omega_0 t - \frac{A_s(t)}{2} \sin 2\omega_0 t \Rightarrow \text{ФНЧ} \Rightarrow \frac{A_c(t)}{2}, \\ x(t) \sin \omega_0 t &= A_c(t) \cos \omega_0 t \sin \omega_0 t - A_s(t) \sin^2 \omega_0 t = \frac{A_c(t)}{2} \sin 2\omega_0 t - \\ &- \frac{A_s(t)}{2} + \frac{A_s(t)}{2} \cos 2\omega_0 t \Rightarrow \text{ФНЧ} \Rightarrow -\frac{A_s(t)}{2}. \end{aligned}$$

После фильтрации возникают низкочастотные колебания, пропорциональные синфазной или квадратурной амплитуде, т.е. формируется *низкочастотное представление* узкополосного сигнала.

Заметим, что опорная частота выбирается произвольно и поэтому экспериментально найденные значения $A_c(t)$ и $A_s(t)$, а следовательно, и низкочастотное представление $A(t)$ оказываются неоднозначными: даже незначительное изменение опорной частоты приведет к появлению новых значений $A_c(t)$ и $A_s(t)$. Неоднозначность удастся устранить при введении комплексного описания узкополосных сигналов и хорошо разработанного аппарата аналитических функций (см. разд. 2.5). Но сначала рассмотрим комплексное представление узкополосных сигналов.

5.5.2. Комплексное представление детерминированных узкополосных сигналов

Комплексное представление узкополосных сигналов является естественным развитием символического метода: вместо комплексных амплитуд используются медленно изменяющиеся комплексные функции времени $\dot{A}(t)$. Запишем в комплексной форме узкополосный сигнал (5.55) и получим:

$$x(t) = \text{Re}[\dot{A}(t)e^{j\omega_0 t}], \quad (5.62)$$

где $\dot{A}(t) = A(t)e^{j\varphi(t)}$ – комплексная амплитуда. Функция $\dot{A}(t)$ называется еще *комплексной огибающей*, а вещественная неотрицательная функция $A(t)$ – *физической огибающей* узкополосного сигнала $x(t)$. Комплексная огибающая содержит полную информацию о сигнале. Действительная и мнимая компоненты комплексной огибающей имеют вид:

$$\begin{aligned} \text{Re}[\dot{A}(t)] &= A(t) \cos \varphi(t) = A_c(t), \\ \text{Im}[\dot{A}(t)] &= A(t) \sin \varphi(t) = A_s(t), \end{aligned} \quad (5.63)$$

т.е. совпадают с синфазной и квадратурной амплитудами, а фаза сигнала $\varphi \in [-\pi, \pi]$ и удовлетворяет равенству:

$$\varphi(t) = \text{arctg} \left[\frac{\text{Im}[\dot{A}(t)]}{\text{Re}[\dot{A}(t)]} \right] = \text{arctg} \left[\frac{A_s(t)}{A_c(t)} \right] \pm \frac{\pi}{2}. \quad (5.64)$$

Физическая огибающая узкополосного сигнала и ее свойства. Физическая огибающая связана с его синфазной и квадратурной амплитудами равенством

$$A(t) = \sqrt{A_c^2(t) + A_s^2(t)}. \quad (5.65)$$

Рассмотрим основные свойства физической огибающей.

1. Величины ее мгновенных значений, в отличие от комплексной огибающей, не зависят от выбора опорной частоты.

Покажем это. Пусть узкополосный сигнал задан формулой (5.62). Если в качестве опорной частоты вместо ω_0 выбрать частоту $\omega'_0 = \omega_0 + \delta\omega$, то (5.62) принимает вид:

$$x(t) = \text{Re}[\dot{A}(t)e^{j\omega'_0 t}] = \text{Re}[\dot{A}(t)e^{j\delta\omega t}e^{j\omega_0 t}].$$

Новая функция комплексной огибающей $\dot{A}'(t) = \dot{A}_s(t)e^{j\delta\omega t}$, а ее модуль (физическая огибающая) остается неизменной, так как $|e^{j\delta\omega t}| = 1$.

2. Физическая огибающая узкополосного сигнала определяет его мгновенную амплитуду. В этом легко убедиться, если ввести производную от квадрата физической огибающей:

$$A(t) \frac{dA(t)}{dt} = A_c(t) \frac{dA_c(t)}{dt} + A_s(t) \frac{dA_s(t)}{dt}.$$

В те моменты времени, когда $A_s(t) = 0$, синфазная составляющая $A_c(t)$, т.е. исходный сигнал, совпадает с физической огибающей и наступает равенство их временных производных. К тому же по определению в любой момент времени $|x(t)| \leq A(t)$, а равенство достигается в моменты времени, когда $\cos(\omega_0 t + \varphi(t)) = \pm 1$.

Таким образом, физическая огибающая и исходный сигнал никогда не пересекаются и имеют общие касательные в точках, где $A_s(t) = 0$. Физическая огибающая действительно «огibtает» узкополосный сигнал и служит его мгновенной амплитудой.

Взаимная связь спектральных плотностей узкополосного сигнала и его комплексной огибающей. Пусть $X(\omega)$ – спектральная плотность узкополосного сигнала $x(t)$, а $G(\omega)$ – спектральная плотность его комплексной огибающей $\dot{A}(t)$. Представим $X(\omega)$ в виде преобразования Фурье от $x(t) = \text{Re}[\dot{A}(t)e^{j\omega_0 t}]$ и воспользуемся известным соотношением $\text{Re}[\dot{A}] = (\dot{A} + \dot{A}^*)/2$. Тогда

$$\begin{aligned} X(\omega) &= \int_{-\infty}^{\infty} \text{Re}[\dot{A}(t)e^{j\omega_0 t}]e^{-j\omega t} dt = \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \dot{A}(t)e^{-j(\omega-\omega_0)t} dt + \frac{1}{2} \int_{-\infty}^{\infty} \dot{A}^*(t)e^{-j(\omega+\omega_0)t} dt. \end{aligned}$$

Последние интегралы представляют преобразования Фурье одной и той же функции $\dot{A}(t)$, смещенные в окрестности частот $\pm\omega_0$, т.е.

$$X(\omega) = \frac{1}{2}G(\omega - \omega_0) + \frac{1}{2}G(\omega + \omega_0). \quad (5.66)$$

Таким образом, спектральная плотность узкополосного сигнала образуется в результате смещения низкочастотного спектра комплексной огибающей в окрестности $\pm\omega_0$.

Аналитические сигналы и низкочастотное представление узкополосных колебаний. Если при введении комплексной модели узкополосного сигнала к вещественному физическому сигналу добавить в качестве мнимой части сопряженный по Гильберту сигнал, то комплексный сигнал становится аналитическим (см. разд. 2.5):

$$z(t) = x(t) + jy(t) = A(t)e^{j\Phi(t)}, \quad Z(\omega) = 2X(\omega), \quad \omega \geq 0, \quad (5.67)$$

где $A(t) = \sqrt{x^2(t) + y^2(t)}$ – физическая огибающая узкополосного сигнала; $\Phi(t) = \text{Arctg}[y(t)/x(t)]$ – мгновенная фаза сигнала.

Функция $\text{Arctg}(\dots)$ принимает любые значения от $-\infty$ до ∞ (в отличие от главного значения $\text{arctg}(\dots)$, существующего в интервале $[-\pi/2, \pi/2]$), и задает полный угол вращения траектории на плоскости xOy за время t (рис. 5.24).

Можно ввести мгновенную частоту узкополосного сигнала как

$$\omega(t) = \frac{d\Phi(t)}{dt} = \frac{y'(t)x(t) - x'(t)y(t)}{A^2(t)}.$$

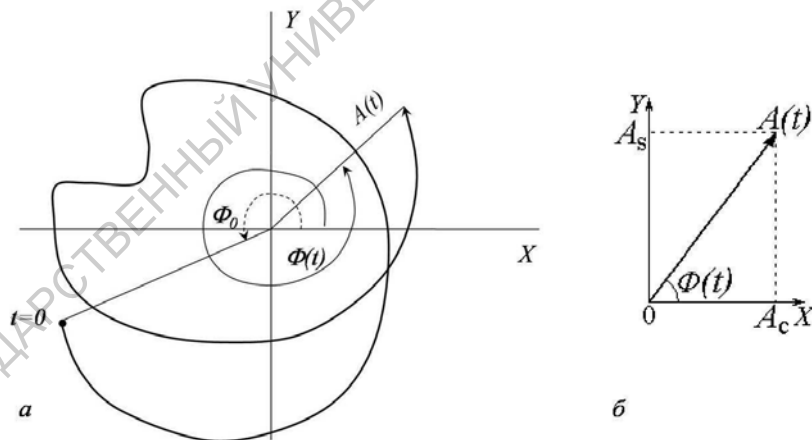


Рис. 5.24. Иллюстрация к определению огибающей $A(t)$, полной фазы $\Phi(t)$ (а), фазы $\varphi(t)$ и квадратурных компонентов $A_c(t)$, $A_s(t)$ (б) узкополосного процесса

Постоянную компоненту частоты $\omega(t)$ называют опорной частотой $\omega_{\text{оп}}$ узкополосного сигнала. Существуют и другие способы определить опорную частоту $\omega_{\text{оп}}$. Например, можно минимизировать полосу частот низкочастотного представления, если взять в качестве $\omega_{\text{оп}}$ «центр тяжести» квадрата спектральной плотности. Полагая равной нулю производную выражения

$$\int_0^{\infty} (\omega - \omega_{\text{оп}})^2 |Z(\omega)|^2 d\omega,$$

получим

$$\omega_{\text{оп}} = \int_0^{\infty} \omega |Z(\omega)|^2 d\omega \left[\int_0^{\infty} |Z(\omega)|^2 d\omega \right]^{-1}.$$

Выведенные соотношения нетрудно проиллюстрировать на простом примере. В начале раздела получена вещественная модель узкополосного сигнала с постоянной спектральной плотностью X_0 . Ему соответствует сигнал $x(t) = A_0 \text{sinc}(\Delta\omega t) \cos \omega_0 t$. Сопряженный сигнал по свойству модуляции $y(t) = A_0 \text{sinc}(\Delta\omega t) \sin \omega_0 t$. Отсюда физическая огибающая $A(t) = A_0 \text{sinc}(\Delta\omega t)$, мгновенная фаза $\Phi(t) = \omega_0 t$, опорная частота $\omega_{\text{оп}} = \omega_0$, а низкочастотное представление есть физическая огибающая и обладает постоянной спектральной плотностью в диапазоне от 0 до $\Delta\omega$.

Вводя тем или иным способом опорную частоту $\omega_{\text{оп}}$, можно определить квадратурные составляющие узкополосного сигнала:

$$A_c(t) = A(t) \cos \varphi(t), \quad A_s(t) = A(t) \sin \varphi(t),$$

где $\varphi(t) = \Phi(t) - \omega_{\text{оп}} t$ – медленная компонента фазы.

Фаза $\varphi(t)$ так же, как $\Phi(t)$, может принимать любые значения. Часто вместо нее вводят приведенную фазу $\varphi_0(t) \in [-\pi, \pi]$ (или $\varphi_0(t) \in [0, 2\pi]$): $\varphi_0(t) = \varphi(t) + 2\pi k(t)$, где k – случайная целочисленная функция, принимающая значения $0, \pm 1, \pm 2, \dots$. Очевидно, при переходе от $\varphi(t)$ к $\varphi_0(t)$ значения косинуса и синуса не изменяются. Соответственно, $A_c(t) = A(t) \cos \varphi_0(t)$ и $A_s(t) = A(t) \sin \varphi_0(t)$. Геометрическая интерпретация квадратурных компонент и угла φ_0 приведены на рис. 5.24, б. Понятно, что

$$\begin{aligned} x(t) &= \text{Re}[z(t)] = A_c \cos \omega_{\text{оп}} t - A_s(t) \sin \omega_{\text{оп}} t, \\ y(t) &= \text{Im}[z(t)] = A_s(t) \cos \omega_{\text{оп}} t + A_c(t) \sin \omega_{\text{оп}} t. \end{aligned} \quad (5.68)$$

Квадратурные составляющие узкополосного процесса так же, как огибающая $A(t)$ и фаза $\varphi(t)$, являются «медленными» функциями по сравнению с периодом $T = 2\pi/\omega_{\text{оп}}$.

Итак, преобразование Гильберта узкополосного процесса $x(t)$ позволяет ввести аналитический сигнал, огибающая которого является его низкочастотным представлением. Это наиболее универсальный способ определения огибающей, мгновенной фазы, мгновенной частоты и квадратурных составляющих узкополосного сигнала, позволяющий устранить описанные выше неоднозначности.

5.5.3. Огибающая, фаза и квадратурные составляющие случайного процесса

Среди случайных процессов также можно выделить узкополосные процессы. Стационарный случайный процесс $X(t)$ называется узкополосным, если выполняется соотношение $\Delta\omega_X \ll \omega_0$, где $\Delta\omega_X$ – ширина спектра на уровне половинной мощности $\Delta\omega_{1/2}$ (или эффективная $\Delta\omega_{\text{эфф}}$

полоса), а ω_0 – частота спектрального максимума. Реализация узкополосного случайного процесса напоминает случайно модулированный гармонический сигнал (рис. 5.25). Поэтому такой процесс часто называют *квазигармоническим* (или *гармоническим*) *шумом*.

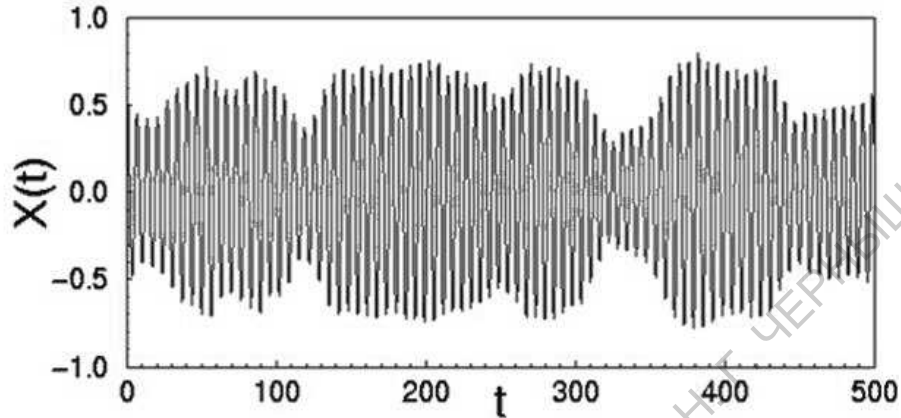


Рис. 5.25. Реализация узкополосного случайного процесса

Так как спектральная плотность мощности $W_X(\omega)$ гармонического шума существенно отлична от нуля только в окрестности ω_0 , то корреляционная функция может быть представлена в виде

$$\Psi_X(\tau) = \gamma_X(\tau) \cos \omega_0 \tau, \quad (5.69)$$

где огибающая корреляционной функции $\gamma_X(\tau)$ является медленно убывающей функцией по сравнению с периодом осцилляций $T_0 = 2\pi/\omega_0$.

Пусть $X(t)$ – стационарный узкополосный случайный процесс (обычно полагают также $\overline{X(t)} \equiv 0$). Если существует сопряженный по Гильберту процесс

$$Y(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{X(\theta)}{t - \theta} d\theta, \quad (5.70)$$

где берется главное значение Коши среднеквадратического интеграла, то можно ввести *случайный аналитический сигнал*

$$Z(t) = X(t) + jY(t). \quad (5.71)$$

Его огибающая $A(t)$, полная фаза $\Phi(t) = \omega_0 t + \varphi(t)$, «медленная» фаза $\varphi(t)$ и квадратурные составляющие определяются аналогично соответствующим характеристикам детерминированного узкополосного сигнала. В качестве опорной частоты обычно выбирают частоту спектрального максимума ($\omega_{\text{оп}} = \omega_0$) или среднюю частоту ($\omega_{\text{оп}} = \omega_{\text{ср}} = \langle \frac{d\Phi(t)}{dt} \rangle$)⁸.

Очевидно, что для случайного сигнала огибающая, фаза $\Phi(t)$ и квадратурные составляющие представляют собой случайные процессы. В принципе, если существует сопряженный по Гильберту процесс $Y(t)$, можно ввести соответствующие характеристики для любого случайного процесса $X(t)$, выбрав опорную частоту произвольным образом. Данное представление имеет смысл только для узкополосных процессов, так как

⁸Во многих случаях частота спектрального максимума совпадает со средней частотой, что означает равенство $\langle \varphi(t) \rangle \equiv 0$.

функции $A(t)$, $\varphi(t)$, $A_c(t)$ и $A_s(t)$ оказываются медленно меняющимися по сравнению с характерным временем $T_0 = 2\pi/\omega_0$. Случайный процесс представляется в виде

$$X(t) = A(t) \cos(\omega_0 t + \varphi(t))$$

и его можно рассматривать как гармонический сигнал, промодулированный по амплитуде и фазе «медленными» случайными сигналами.

Можно показать, что для нормального (вещественного) узкополосного шума $X(t)$ с $\overline{X(t)} \equiv 0$ квадратурные составляющие $A_c(t)$ и $A_s(t)$ являются совместно гауссовскими «медленными» процессами, причём $\overline{A_c(t)} \equiv \overline{A_s(t)} \equiv 0$ и $\langle A_c(t)A_s(t) \rangle \equiv 0$, т.е. одновременные значения $A_c(t)$ и $A_s(t)$ статистически независимы. Представление случайных процессов в виде гармонических сигналов с «медленной» случайной модуляцией параметров оказывается удобным при рассмотрении ряда задач.

Замечание. Огибающую, фазу и квадратурные компоненты можно ввести для любого стационарного случайного процесса $X(t)$, если существует и известен сопряженный процесс $Y(t)$. Однако когда $X(t)$ не является узкополосным, эти функции не будут «медленными» и их использование нецелесообразно.

5.6. Принципы цифровой модуляции сигналов

Современные информационные технологии включают целый комплекс цифровых методов обработки, использования и передачи информации. Как показано в разд. 4.3, для преобразования в цифровую форму непрерывный сигнал дискретизируется по времени, квантуется по уровню и полученные отсчеты кодируются в m -ичной системе счисления, например в k -разрядном двоичном коде. При этом каждому отсчету соответствует последовательность символов системы счисления, называемая *кодовой последовательностью*.

При цифровой модуляции используются главным образом равномерные потенциальные коды (см. разд. 4.3). Закодированное сообщение в виде периодической последовательности кодовых слов поступает в канал связи и в проводных и некоторых коаксиальных линиях (например в компьютерных сетях) может передаваться без модуляции. Но в подавляющем числе радиосистем связи, использующих коаксиальные и оптоволоконные линии, атмосферу Земли, околоземное и космическое пространство, модуляция высокочастотных несущих колебаний принципиально необходима.

Простейшие виды цифровой модуляции возникают уже при дискретизации и кодировании амплитудно-частотно- и фазо-манипулированных сигналов. Представление модулирующих сигналов в цифровой форме значительно расширяет возможности традиционных методов модуляции и позволяет создавать и постоянно совершенствовать системы спутниковой и мобильной связи, кабельного телевидения и телефонии.

Разнообразие цифровых систем и методов обработки сигналов породило множество видов цифровой модуляции (ЦМ). Поэтому мы не пы-

тались даже коротко описать все методы и предлагаем обсудить только основные принципы функционирования цифровой модуляции сигналов.

5.6.1. Цифровая амплитудная модуляция сигналов

Цифровая амплитудная и импульсно-кодовая модуляция. Цифровой амплитудной модуляцией будем называть дискретное изменение амплитуды несущего колебания в соответствии с изменением цифрового модулирующего сигнала.

Пусть задан непрерывный сигнал $x(t)$ (рис. 5.26, а) с ограниченным спектром $0 \leq \omega \leq \omega_h$ (см. разд. 4.2). Этим сигналом можно модулировать по амплитуде высокочастотное несущее колебание и получить АМ-сигнал. Учитывая ограниченность спектра, дискретизируем $x(t)$ с интервалом Δt ($\omega_d = 2\pi/\Delta t \geq 2\omega_h$) и получим дискретные отсчеты $x(n\Delta t) = x(n)$, где n - номер отсчета (в нашем случае $n = 0 \div 12$). Согласно теореме Котельникова по $x(n)$ исходный аналоговый сигнал восстанавливается точно.

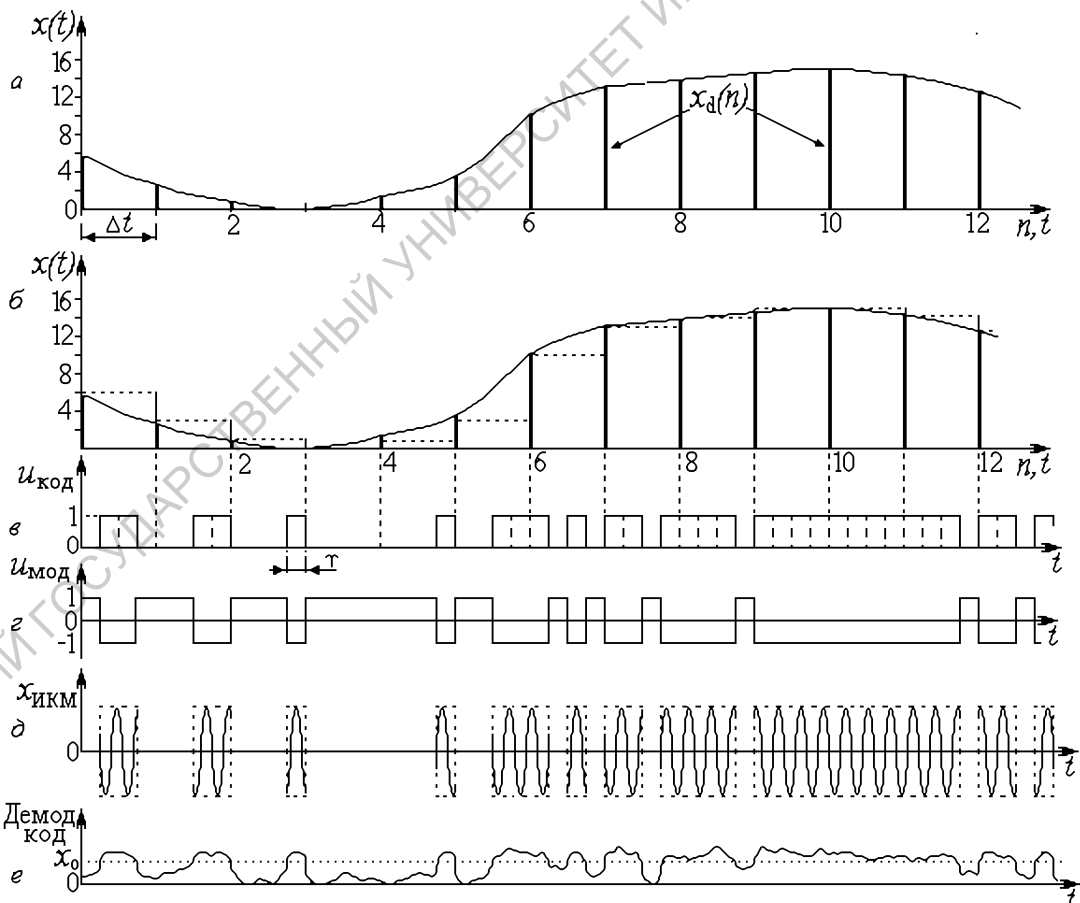


Рис. 5.26. Формирование ИКМ-сигнала и демодуляция ИКМ-сигнала, искаженного помехой

Если произвести квантование полученных отсчетных значений в двоичной системе счисления, то получим импульсный цифровой сигнал $x_i(n)$, где i - номер уровня квантования. Пролонгируем каждое значение $x(n)$ до следующего отсчета и получим потенциальный цифровой сигнал -

ступенчатую аппроксимацию $x(t)$ (рис. 5.26, б). И $x_i(n)$, и ступенчатая аппроксимация $x(t)$ представляют один и тот же цифровой модулирующий сигнал, но в разной форме (импульсной или потенциальной). От формы представления, естественно, зависит ширина спектра модулирующего сигнала. Ступенчатым цифровым сигналом (его спектр уже) тоже возможно промодулировать по амплитуде несущее колебание и получить цифровой амплитудно-манипулированный сигнал (ЦАМн-сигнал). Ширина спектра АМн-сигнала даже между первыми нулями огибающей оказывается вдвое шире, чем у соответствующего АМ-сигнала, но и помехозащищенность возрастает, так как при наличии помех значение $x(n)$ восстанавливается на интервале Δt .

Цифровая амплитудная манипуляция не получила распространения и очень быстро была вытеснена *импульсно-кодовой модуляцией* (ИКМ). При ИКМ отсчетные значения сигнала преобразуются в потенциальный NRZ код (см. рис. 4.7) – последовательность единичных импульсов $u_{\text{код}}(t)$ (рис. 5.26, в) – битов⁹. Длительности битов (такты или битовый интервал последовательности Υ) выбирается с условием, что импульсы $u_{\text{код}}(t)$ уместятся до наступления следующего отсчета $x(k)$ ($\Upsilon \leq \frac{\Delta t}{k}$). Такие сигналы можно передавать по проводным линиям без модуляции.

Как отмечалось выше, качество декодированных сигналов зависит от числа уровней квантования, а значит, и от числа разрядов в импульсно-кодовых последовательностях. Кодирование отсчетных значений сигнала выполняется как минимум 8-разрядными кодами. При 12-, 16-, а в отдельных случаях и 24-разрядном двоичном кодировании длина кодовых последовательностей увеличивается, и приходится уменьшать тактовый интервал Υ . Кроме того, в приведённом виде ИКМ-последовательность содержит ненужную при модуляции постоянную составляющую. Для ее устранения ИКМ-сигнал преобразуется в модулирующий сигнал $u_{\text{мод}}(t)$ (рис. 5.26, г): логическая единица превращается в напряжение отрицательной полярности, а логический нуль – в напряжение той же амплитуды, но положительной полярности.

Так как длительность кодовых импульсов определяется тактовым интервалом Υ , спектр сигнала, манипулированного этим напряжением (рис. 5.26, д), значительно расширяется по сравнению с АМн-сигналом. Демодуляция ИКМ-сигнала выполняется в обратной последовательности.

Демодулированный ИКМ-сигнал, искаженный воздействием помех, может заметно отличаться от модулирующего сигнала (рис. 5.26, е), но практически всегда удается задать такой уровень x_0 , чтобы пересечения его соответствовали переходу от нуля к единице и наоборот.

Таким образом, ИКМ обладает следующими недостатками:

- 1) полоса частот модулированного сигнала превышает полосу частот соответствующих АМ- и АМн-сигналов;
- 2) модулирующие и демодулированные сигналы нуждаются в точной синхронизации во времени;
- 3) при многоразрядном кодировании тактовый интервал Υ уменьшается, а спектр сигнала расширяется. Одновременно возникают труд-

⁹Строго говоря, бит является двоичной единицей количества информации, но в радиоэлектронике и в цифровой технике битами называют и единичные импульсы, передающие 0 или 1.

ности безошибочной передачи длинных кодовых последовательностей.

Понятие о дифференциальной ИКМ и дельта-модуляции. Соседние отсчетные значения реальных сигналов, например сигнала, показанного на рис. 5.26, а, из-за ограниченности спектра не могут резко отличаться по величине. Поэтому вместо кодирования отсчетных значений после квантования можно кодировать разности значений соседних отсчетов. Амплитуды разностей оказываются много меньше самих отсчетов, и число разрядов в кодовых комбинациях уменьшается в 2–3 раза. Это основная идея разностной или *дифференциальной ИКМ*. Если ещё и интервал между отсчетами сделать *адаптивным*, т.е. «подстраивать» к свойствам сигнала: при медленном изменении отсчетных значений увеличивать его величину, а при резком изменении – уменьшать, то полоса частот *адаптивного дифференциального ИКМ* сокращается почти вдвое.

Еще более эффективным способом преобразования аналоговых сигналов в цифровую форму является *дельта-модуляция*. Пусть шаг дискретизации непрерывного аналогового сигнала Δ_δ (рис. 5.27, а) значительно меньше того, который необходим по критерию Котельникова – Найквиста (интервал Δt отмечен короткими жирными линиями) и выбран так, чтобы разность соседних отсчетов не превышала шага квантования. Тогда приращение каждого нового отсчета Δx будет единичным импульсом положительной или отрицательной полярности (рис. 5.27, б) в зависимости от знака приращения. Это импульсное кодирование. Такая «модуляция» отличается простотой реализации и используется в проводных линиях связи. Восстановление аналогового сигнала осуществляется с помощью интегрирования.

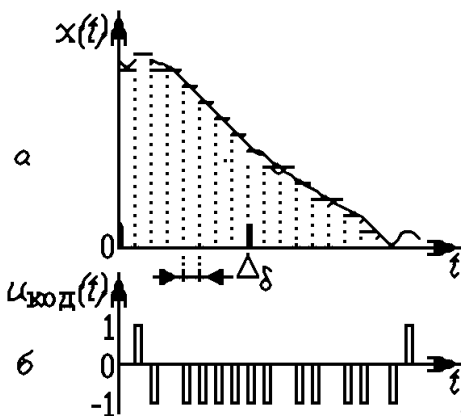


Рис. 5.27. Дельта-модуляция

5.6.2. Цифровая фазовая манипуляция сигналов

Многопозиционная фазовая манипуляция. Цифровой фазовой манипуляцией будем называть дискретное изменение фазы несущего колебания в соответствии с временным изменением цифрового модулирующего сигнала. Абсолютное значение начальной фазы несущего колебания измерять не удастся, и практически используется фазо-разностная или *относительная фазовая манипуляция* (ОФМн), когда измеряется фазовый сдвиг двух соседних интервалов.

Простейший пример цифровой фазовой манипуляции был рассмотрен в разд. 5.3, где несущее колебание модулировалось по фазе последовательностью одноразрядных двоичных чисел. Описанный принцип можно распространить на модуляцию фазы произвольным цифровым сигналом, например описанной выше импульсно-кодовой последовательностью, инвертируя фазу поразрядно. Представленный способ неэффективен из-за малой длительности единичных импульсов (тактовых интервалов) и, как следствие, большой полосы частот. Поэтому обычно применяется многопозиционная фазовая манипуляция.

Пусть отсчетные значения модулирующего сигнала представлены 3-разрядными двоичными числами (всего $N = 2^3 = 8$ различных чисел), и каждой кодовой комбинации приписано свое значение фазового сдвига, пропорциональное $2\pi/8 = \pi/4$. Интервалы между изменениями фазы увеличиваются, а полоса частот ФМн-сигнала сокращается в 3 раза. При фазовой манипуляции амплитуда несущего колебания A постоянна, а комплексные амплитуды $\dot{A}(t) = Ae^{j\Delta\varphi \cdot m_N(t)} = Ae^{j\frac{\pi}{4}m_N}$ (m_N – условный номер кодовой комбинации) для различных $\Delta\varphi$ можно представить в полярной системе координат совокупностью векторов, получившей название *сигнального созвездия* (рис. 5.28). Коды соседних векторов выбираются в соответствии с кодом Грея (см. разд. 1.2) и различаются значением одного двоичного разряда. Благодаря этому при демодуляции зашумленного сигнала удастся исправлять большинство ошибочных разрядов.

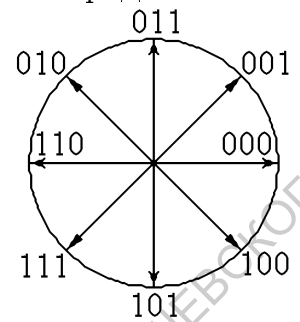


Рис. 5.28. Сигнальное созвездие

Из векторной диаграммы видно, что представленная многопозиционная ФМн содержит «скачки» фазы на 180° , которые могут вызывать значительные искажения огибающей радиосигнала. Влияние последних удастся уменьшить при использовании квадратурной ФМн.

Квадратурная фазовая манипуляция. Пусть цифровой сигнал задан последовательностью единичных импульсов $u_{\text{код}}(t)$ (рис. 5.29, а), полученной при рассмотрении ИКМ. Любые пары следующих друг за другом битов составляет одну из четырех комбинаций: (0, 0), (0, 1), (1, 0), (1, 1), и каждому такому двухбитовому числу можно сопоставить свое

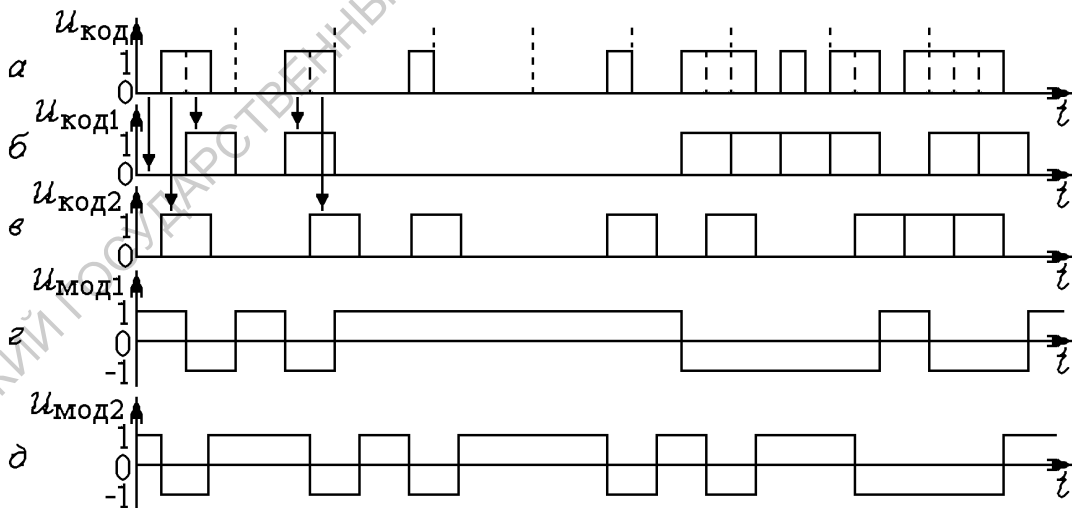


Рис. 5.29. Формирование квадратурного фазо-манипулированного сигнала

значение фазового сдвига, аналогично тому, как это делалось с 3-разрядными числами. Но при квадратурной фазовой манипуляции результирующий комплексный вектор формируется иначе. Сначала все нечетные биты выделяются в одну, а четные – в другую подпоследовательности $u_{\text{код1}}(t)$, $u_{\text{код2}}(t)$ (рис. 5.29, б, в). При этом длительности тактовых интервалов T удваиваются. Мы получаем потенциальное кодирование типа

2B1Q. Затем обе подпоследовательности преобразуются в модулирующие сигналы $u_{\text{мод1}}(t)$, $u_{\text{мод2}}(t)$ и модулируют по амплитуде квадратурные компоненты несущего колебания. Результирующий комплексный вектор $z_{\text{ФМн}}$ удовлетворяет равенству:

$$z_{\text{ФМн}}(t) = \frac{A}{\sqrt{2}}u_{\text{мод1}}(t) \cos \omega_0 t + j \frac{A}{\sqrt{2}}u_{\text{мод2}}(t) \sin \omega_0 t,$$

где A – амплитуда несущего колебания.

Это и есть квадратурная фазовая манипуляция (QPSK – Quadrature Phase Shift Keying). Ее сигнальное созвездие представлено на рис. 5.30, *а*. Все комплексные векторы имеют одинаковую амплитуду и образуют с осями координат углы $\pm 45^\circ$. Разность фаз двух соседних векторов составляют $\pm 90^\circ$. Легко проверить, что кодовая комбинация, указанная у конца каждого комплексного вектора, при любых t соответствует значениям следующих друг за другом битов. Поскольку подпоследовательность четных битов сдвинута относительно подпоследовательности нечетных на T , такую фазовую манипуляцию часто называется *квадратурной ФМн со сдвигом*.

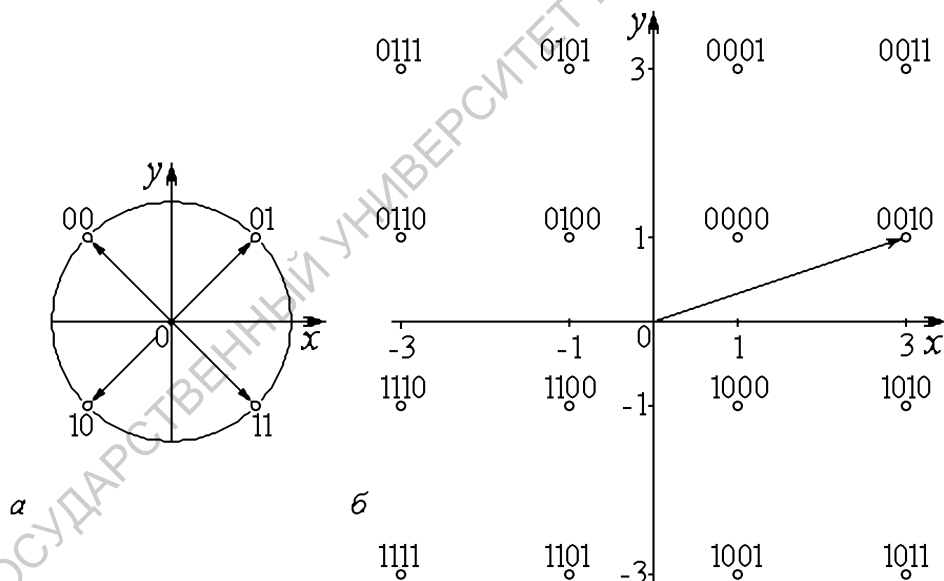


Рис. 5.30. Сигнальные созвездия квадратурной фазовой манипуляции (*а*) и четырехуровневой QAM (*б*)

Квадратурная амплитудно-фазовая манипуляция. Квадратурную амплитудно-фазовую манипуляцию (QAM – Quadrature Amplitude Modulation или QASK – Quadrature Amplitude Shift Keying) будем рассматривать как обобщение квадратурной фазовой манипуляции:

$$z_{\text{QAM}}(t) = \frac{A}{\sqrt{2}}q_1 u_{\text{мод1}}(t) \cos \omega_0 t + j \frac{A}{\sqrt{2}}q_2 u_{\text{мод2}}(t) \sin \omega_0 t, \quad (5.72)$$

где A – амплитуда несущего колебания; q_1, q_2 – некоторые масштабные множители.

При цифровой квадратурной манипуляции кодируются одновременно амплитуда и фаза несущего колебания, причем масштабные множители

квадратурных компонентов независимы и образуют последовательность $q_1, q_2 = \pm 1, \pm 3, \pm 5, \dots$. Значения масштабных множителей подобраны так, чтобы окончания комплексных векторов образовали на плоскости сигнальных созвездий равномерную квадратную сетку с шагом 2.

Пусть отсчетные значения модулирующего сигнала представлены 4-разрядными двоичными числами $B(b_0, b_1, b_2, b_3)$. Такая QAM называется *четырёхуровневой*. Как и в ФМн, из этого сигнала формируются двумерные последовательности из нечетных и четных битов. Два старших бита задают знаки реальной и мнимой компонент вектора z_{QAM} и тем самым указывают квадрант на плоскости сигнальных созвездий, в котором располагается z_{QAM} . Два младших разряда задают амплитуды реальной и мнимой квадратурных компонент. Например, комбинации битов 01 соответствуют амплитуды $A/\sqrt{2}, 3A/\sqrt{2}$. Таким образом формируются 16 различных значений комплексного вектора, отличающихся по амплитуде и фазе (отсюда и обозначение 16-QAM). Каждый из векторов отражает соответствующее сочетание двоичных символов в параметрах несущего колебания длительностью T . Когда $q_1 = q_2 = \pm 1$, получаем двухуровневую QAM, которая совпадает с квадратурной ФМн. Сигнальное созвездие 16-QAM и одно из возможных значений комплексного вектора представлены на рис. 5.30, б.

При создании QAM-сигналов не существуют ограничения на число возможных уровней. Практически могут быть реализованы 128-QAM, 256-QAM и т.д., но наибольшее распространение получили 16-QAM и 64-QAM. По существу QAM является разновидностью гармонической амплитудно-фазовой модуляции. Спектр QAM-сигнала сосредоточен в окрестности ω_0 и не превышает по ширине спектр соответствующего АМ-колебания. Системы QAM обладают повышенной помехоустойчивостью. Шумы и помехи размывают сигнальные точки созвездия, но приемник анализирует полученное значение комплексного вектора и отождествляет его с ближайшим стандартным из базы векторов QAM.

В настоящее время цифровая квадратурная модуляция применяется не только в радиоканалах, но и в высокоскоростных системах асимметричной цифровой абонентской связи по двухпроводным линиям (ADSL).

5.6.3. Цифровая частотная манипуляция сигналов

Цифровой частотной манипуляцией будем называть дискретное изменение частоты несущего колебания в соответствии с временным изменением цифрового модулирующего сигнала. Простейший пример цифровой частотной манипуляции был рассмотрен в разд. 5.3, где модулирующий сигнал представлял последовательность одноразрядных двоичных чисел. Многоуровневая частотная манипуляция, как и другие виды цифровой модуляции, использует разбиение двоичной кодовой последовательности на m -битовые символы, превращение их в двухполярные модулирующие напряжения и введение для каждого символа определенной частоты несущего колебания. Так получается 4- и 8-уровневая частотная манипуляция, используемая в современных системах радиосвязи.

Большой интерес представляет так называемая *частотная манипуляция с минимальным сдвигом* (ЧММС). При ЧММС фаза манипули-

рованного радиосигнала не испытывает скачков при переходе с одной частоты на другую, т.е. на границах радиоимпульсов. Это возможно, если разность частот $\omega_{\text{раз}} = \omega_1 - \omega_2$ подобрана так, что коэффициент взаимной корреляции обращается в нуль при $\omega_{\text{раз}} \Upsilon = \pi$. При такой $\omega_{\text{раз}}$ фаза радиосигнала за время Υ изменяется на $\pi/2$, а частоты $\omega_{1,2}$ удовлетворяют равенству $\omega_{1,2} = \omega_0 \pm \pi/2\Upsilon$. При демодуляции ЧММС используется синхронное детектирование.

Принципы многочастотной модуляции. Многочастотный способ модуляции – ортогональное частотное разделение каналов (OFDM – Orthogonal Frequency Division Multiplexing) представляет модуляцию с несколькими несущими. При этом полоса частот, отведенная для передачи данных, разбивается на узкополосные поддиапазоны (обычно 256 полос со своими несущими и шириной в 4,3 кГц). Для каждого поддиапазона, исходя из уровня помех, выбирается метод фазовой или квадратурной манипуляции. Если несущие частоты кратны друг другу и частота первой равна ω , то последняя будет иметь частоту $n\omega$, в 256 раз большую.

Пусть последовательность информационных символов разбивается на n подблоков (символов) и каждый символ модулирует свою несущую методом QAM, изначально обладающим свойством ортогональности, т.е. фактически происходит переход от последовательной к параллельной передаче данных.

Сигнал несущей с частотой $k\omega$ описывается комплексным колебанием

$$z_{QAM}(t) = \dot{C}_k \exp jk\omega t,$$

где $\omega = 2\pi/(m\Upsilon)$; $m\Upsilon$ – длительность символа (m битов). Здесь сигналы, модулирующие квадратурные компоненты несущего колебания, содержатся в \dot{C}_k . Если теперь просуммировать все несущие колебания, промодулированные своими символами, то получим

$$z_{QAM}(m\Upsilon) = \sum_{k=1}^n \dot{C}_k \exp jk\omega m\Upsilon = \sum_{k=1}^n \dot{C}_k \exp j2\pi km/N, \quad (5.73)$$

где N – количество отсчетов. Временная последовательность из таких сигналов и образует OFDM-сигнал. Заметим, что этот сигнал еще не пригоден для передачи по радиоканалу, так как это всего лишь многочастотная поднесущая. Используя ее для модуляции высокочастотного колебания, получим нужный радиосигнал.

Принятый радиосигнал прежде всего демодулируется и превращается в $z_{QAM}(t)$. Теперь его нужно развернуть в исходную последовательность цифровых символов, и сделать это желательно в реальном времени.

Сравнивая (5.73) с формулой (4.30) для обратного преобразования Фурье (см. разд. 4.4.1), легко увидеть аналогию между этими разложениями. Таким образом для получения исходной цифровой последовательности необходимо выполнить обратное дискретное преобразование Фурье. Как и в случае преобразования Фурье, формула (5.73) позволяет определить значения сигнала при суммировании ортогональных гармонических составляющих с известными комплексными амплитудами. Поэтому разработанная еще 50-х годах прошлого века OFDM-модуляция не

получила распространения, пока не появились алгоритмы БПФ и микросхемы, практически реализующие эти алгоритмы.

Система OFDM обладает многими достоинствами. Достаточно указать на возможность точной адаптации приемо-передающих устройств к условиям распространения сигнала и высокую скорость передачи данных. Главный недостаток – сложность технической реализации. Несмотря на это метод OFDM признан основным в высокоскоростных системах асимметричной цифровой абонентской связи ADSL.

5.6.4. Методы цифровой модуляции и перспективы цифрового радиовещания

В настоящее время известны три разновидности цифрового радиовещания: самая близкая к нам по методике и технологиям европейская система «Эврика-147» со стандартом DAB (Digital Audio Broadcasting – цифровое звуковое радиовещание), американская IBOC и система Всемирного цифрового радио (DRM – Digital Radio Mondiale) с частотами несущих менее 30 МГц. Все эти системы используют цифровое кодирование и сжатие сообщений, квадратурную амплитудную модуляцию (QAM), ортогональное частотное разделение каналов (OFDM), методы быстрого Фурье-преобразования (БПФ). Использование этих методов позволяет сформировать и одновременно передавать на одной несущей частоте 6 различных радиовещательных программ.

В соответствии с рассмотренными выше цифровыми методами модуляции к закодированному и сжатому MPEG-кодеком¹⁰ информационному аудиосигналу добавляются служебная информация и дополнительные данные (например, название исполняемого произведения, прогноз погоды и т.п.). Кодированные таким образом каналы объединяются в один общий канал и после квадратурной модуляции передаются для OFDM-модуляции радиочастотного колебания.

Если современная УКВ-станция требует для передачи одного канала полосу частот более 240 кГц, то цифровая радиостанция может использовать один радиоканал с полосой частот всего 10 кГц и передавать шесть программ с качеством, близким к УКВ-вещанию.

Цифровые методы радиовещания открывают и другие возможности. В частности, при создании разнесенных в пространстве радиостанций, работающих в синхронном режиме на одной и той же частоте, удается ликвидировать «эффект замирания» и зоны молчания на метровых волнах. Однако рассмотрение этих вопросов выходит за пределы нашего учебного пособия. Заметим только, что в Европе уже работают десятки цифровых радиовещательных станций, а число радиоприемников составляет несколько десятков тысяч, хотя стоимость приемников пока еще остается высокой. Следует заметить, что серьезную конкуренцию наземному цифровому радиовещанию составляет спутниковое вещание

¹⁰Наиболее распространены два вида кодеров-декодеров MPEG-кодексов (Moving Picture Experts Group) – MPEG-2 и MPEG-4. MPEG-2 уже длительное время используется для записи DVD-дисков и имеет максимальную разрядность 16 бит. MPEG-4 является более совершенным. Он содержит адаптивные механизмы компрессии, а при записи DVD-информации ищет и сохраняет ключевые кадры изображения, а затем прогнозирует и сохраняет лишь информацию об изменениях.

радиопрограмм в формате аналогового FM. Спасает только то, что для приема спутникового вещания необходима специальная антенна и соответствующий приемник, а сигналы наземного цифрового вещания принимаются на обычную штыревую антенну.

В заключение отметим, что в России в 2001–2003 гг. проводилось опытное вещание в Москве и Санкт-Петербурге. При этом транслировались с высоким качеством 6 общероссийских программ, в том числе «Радио России», «Маяк» и «Орфей».

5.7. Телевизионные сигналы и их спектры

Системы телевизионного вещания на протяжении десятилетий развиваются по различным направлениям и используют различные методы модуляции сигналов. Разработанные стандарты различаются, в основном, числом строк в кадре изображения, частотой смены кадров и способом введения цветовой информации. Сейчас на смену старым системам приходит телевидение высокой четкости (ТВЧ), использующее цифровую обработку сигналов. Изменяются не только стандарты, но и физические характеристики телевизионных сигналов, методы их формирования, передачи и приема. По существу ТВЧ является развитием традиционного (классического) телевидения и использует уже апробированные решения. Поэтому нам представлялось естественным сначала рассмотреть основные структуры эфирного аналогового телевидения и только после этого анализировать методы цифрового телевидения, обращая особое внимание на эволюцию методов модуляции. Во Франции, России и Китае принят стандарт SECAM (формат кадра $K = L_x/L_y = 4/3$, число строк $N = 625$, частота кадров в секунду – 25, полукадров – 50). На него, в основном, и будем ориентироваться.

5.7.1. Сигналы аналогового эфирного телевидения

В реальных радиосигналах всегда присутствуют случайные компоненты, содержащие передаваемые сообщения, и детерминированные элементы, позволяющие распознать передаваемые сигналы.

Полный телевизионный сигнал тоже содержит случайную и детерминированную составляющие. Случайная компонента включает три вида информации: распределение яркости в различных точках экрана, сведения о цветовой окраске элементов изображения и сигналы звукового сопровождения. Детерминированными являются несущее колебание с частотой телевизионного канала и управляющие построочной и покадровой синхронизацией импульсы, входящие в состав модулирующего сигнала. Таким образом, модулирующий сигнал изображения содержит случайные компоненты, характеризующие распределение яркости и цвета, и детерминированные последовательности импульсов для синхронизации строк и кадров телевизионного изображения.

Пространственные и временные спектры сигналов телевизионных изображений. Пусть передаваемое двумерное изображение с распределением яркости $E'(x, y)$ неподвижно, а его размеры ограни-

чены по ширине ($0 \leq x \leq L_x$) и высоте ($0 \leq y \leq L_y$). Многократно повторяя отрезки L_x и L_y по x и y , можно «сконструировать» периодическую функцию $E(x, y)$, существующую для $-\infty < x < \infty$, $-\infty < y < \infty$ и совпадающую с $E'(x, y)$ для $0 \leq x \leq L_x$, $0 \leq y \leq L_y$, и представить $E(x, y)$ в виде разложения в двумерный ряд Фурье.

Формирование электрических сигналов, адекватных изображениям, осуществляется в различных фотоэлектрических преобразователях, например в видекодах. В современных системах телевидения принято последовательное построчное и покадровое разложение изображений (рис. 5.31). Когда электронный луч в видеконе переходит на более освещенный элемент мишени, ток в сопротивлении нагрузки и напряжение $U(t)$ на нем возрастают. Таким образом, при попадании электронного луча в точку с координатами x, y на нагрузке возникает напряжение $U(t)$, пропорциональное яркости $E'(x(t), y(t))$.

При построчной (*прогрессивной*) развертке (см. рис. 5.31, а) электронный луч, начиная с левого верхнего угла, перемещается по горизонтали от начала до конца строки и обратно (период строчной развертки T_c) и одновременно, но более медленно, по вертикали (период кадровой развертки T_k).

При использовании чересстрочной развертки (рис. 5.31, б) изображение формируется из двух полей. Первое поле образуют нечетные строки, второе – четные. Прочертив половину последней нечетной строки, луч перемещается в середину верхнего края кадра и прочерчивает ее вторую половину, а затем последовательно все четные строки. Дойдя до нижнего правого угла раstra, электронный луч возвращается в верхний левый угол и процесс развертки повторяется.

Если электронный луч равномерно перемещается вдоль оси Ox со скоростью v_x , а вдоль оси Oy – со скоростью v_y , то $T_c = L_x/v_x$ и $T_k = L_y/v_y$ превращаются во временные периоды, а $f_c = 1/T_c$, $f_k = 1/T_k$ – в частоты строчной и кадровой развертки. Это позволяет представить напряжение $U(t) = \Phi[E'(x(t), y(t))]$ как

$$U(t) = \Phi[E'(x(t), y(t))] = \frac{1}{2} \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \dot{U}_{ik} e^{2\pi j(i f_c + k f_k)t},$$

или в вещественном виде:

$$U(t) = \Phi[E'(x(t), y(t))] = \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} U_{ik} \cos [2\pi(i f_c + k f_k)t + \varphi_{ik}], \quad (5.74)$$

где U_{ik} и φ_{ik} – амплитуды и фазы спектральных составляющих.

Постоянная U_{00} и низкочастотные составляющие U_{0k} ($i=0$) с частотами, кратными 50 Гц, быстро затухают и характеризуют среднюю яркость

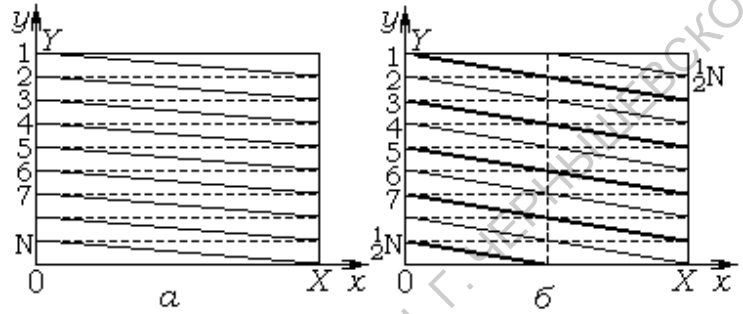


Рис. 5.31. Прогрессивная (а) и чересстрочная (б) развертка изображений. Обратный ход развертки изображен пунктирными линиями

кадра, а составляющие с частотами $if_c \pm kf_k$ образуют узкие полосы вокруг $f_c, 2f_c, \dots$ (рис. 5.32) и характеризуют яркость построчно.

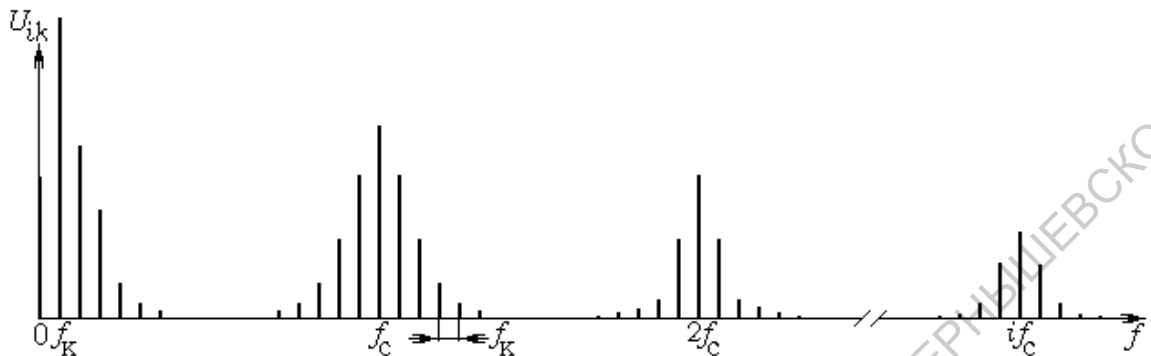


Рис. 5.32. Амплитудный спектр напряжения $U(t)$ неподвижного изображения

Разложение (5.74) является неполным, так как не содержит дискретные спектральные составляющие периодических последовательностей гасящих и синхронизирующих импульсов, которые создаются в конце каждой строки и каждого кадра для гашения обратного хода развертки.

Структура гасящих строчных импульсов для двух строк, одна из которых соответствует плавному изменению яркости, а другая – чередованию белых и черных полос, представлена на рис. 5.33, а гасящие кадровые импульсы – на рис. 5.34. Максимальной яркости (уровню белого) обычно соответствует минимальная величина видеосигнала, а амплитуды гасящих импульсов превышают уровень черного. Во время сигналов гашения дополнительно передаются импульсы, синхронизирующие развертки¹¹.

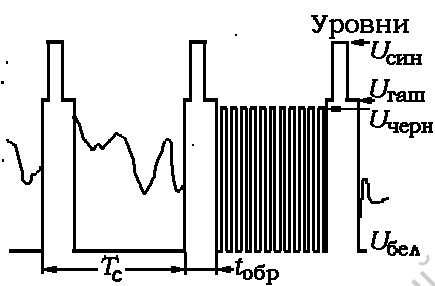


Рис. 5.33. Структура гасящих строчных импульсов

Уровни $U_{\text{син}}$, $U_{\text{гаш}}$, $U_{\text{черн}}$, $U_{\text{бел}}$

Во время сигналов гашения дополнительно передаются импульсы, синхронизирующие развертки¹¹.

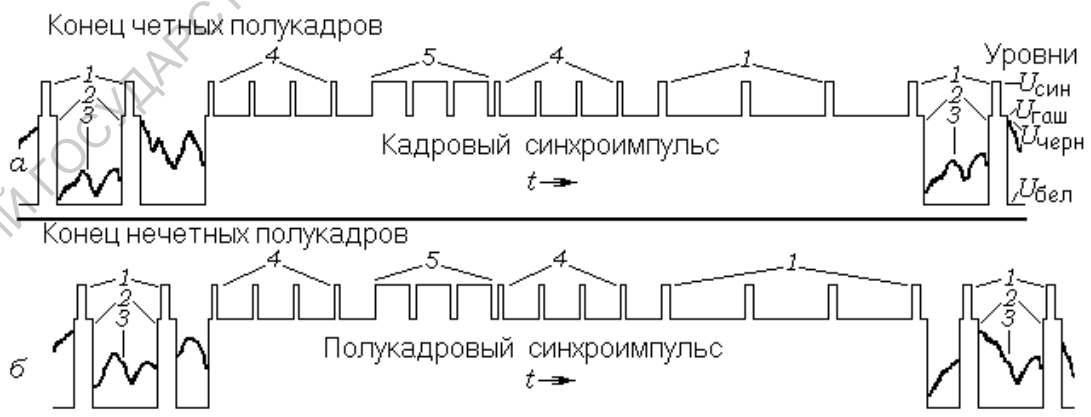


Рис. 5.34. Структура кадрового (а) и полукадрового (б) импульсов: 1 – строчные синхроимпульсы; 2 – гасящие строчные импульсы; 3 – видеосигнал; 4 – уравнивающие импульсы; 5 – кадровые синхроимпульсы

Все кадровые гасящие импульсы прогрессивной развертки одинаковы

¹¹Это так называемая негативная полярность сигналов, принятая еще в СССР. Существует и другая (позитивная) полярность, когда уровню белого соответствует максимальная величина видеосигнала, а гасящие импульсы формируются на уровне черного.

(рис. 5.34, а), а кадровые и полукадровые гасящие импульсы чересстрочной развертки представлены на рис. 5.34, а, б. Они содержат врезки с периодом строчной развертки, так называемые уравнивающие импульсы, следующие с интервалом, равным половине длительности строки, и синхронизирующие развертку в полукадрах. Таким образом, полный видеосигнал $U(t)$ состоит из переменных напряжений, мгновенные значения которых пропорциональны яркости изображения $E'(x(t), y(t))$, разделенных гасящими строчными и кадровыми импульсами. При этом спектр неподвижного изображения (см. рис. 5.32) остается линейчатым, но, как показывают расчеты, из-за синхронизирующих импульсов амплитуды гармоник (особенно кратных f_c) заметно увеличиваются.

При передаче подвижных изображений спектральные линии смещаются и расширяются. Дискретный характер спектра практически не изменяется, но каждая спектральная линия занимает некоторую полосу частот. Средняя яркость изображений непрерывно изменяется с очень низкой частотой, и в спектре телевизионного сигнала появляются так называемые «нулевые» частоты в интервале от 0 до $2 \div 3$ Гц. Передача таких низких частот по радиоканалам невозможна, и изменения средней яркости изображений передаются путем медленной амплитудной модуляции строчных гасящих импульсов.

Для оценки усредненных с учетом разнообразных сюжетов спектральных характеристик случайных сигналов используются понятия энергетического спектра $W_U(f)$ и ковариационной функции $K_U(\tau)$:

$$K_U(\tau) = \lim_{T \rightarrow \infty} \left(\frac{1}{T} \int_0^T U(t)U(t + \tau) dt \right) = \int_{-\infty}^{\infty} W_U(f) e^{j2\pi f\tau} df,$$

$$W_U(f) = \int_{-\infty}^{\infty} K_U(\tau) e^{-j2\pi f\tau} d\tau.$$

Как показывают расчеты, для реальных параметров телевизионных изображений коэффициенты корреляции между соседними элементами одной строки составляют 0.98, а между ближайшими элементами соседних строк и кадров – 0.92. Энергетическая плотность сигналов быстро убывает с ростом частоты (рис. 5.35), а основная энергия спектра сосредоточена в области низких частот: составляющие спектра в $\cong 10\%$ -ной полосе частот сигнала содержат более 90% его полной мощности, а в диапазоне частот от 50 Гц до 1.25 МГц – до 95% полной мощности.

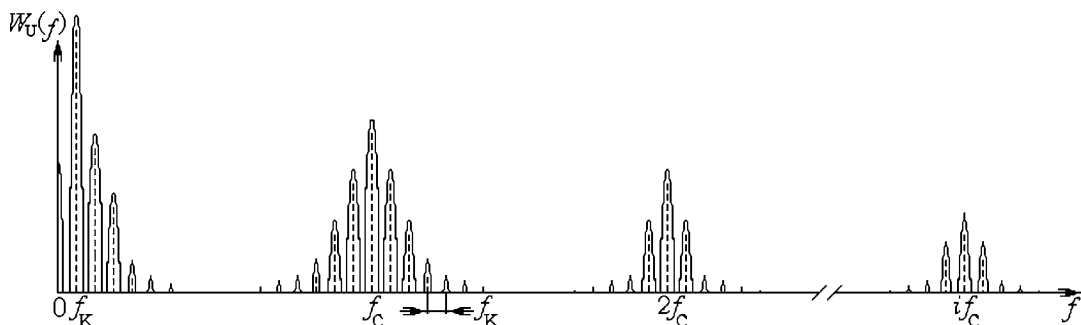


Рис. 5.35. Спектральная плотность мощности электрического сигнала

Столь высокая степень корреляции и быстрое убывание энергетической плотности сигнала представляют очень важные выводы и широко используются при построении телевизионных систем высокой четкости для предсказания значений сигналов и сокращения объема передаваемой информации.

Обратимся теперь к формированию видеосигналов изображения в системе SECAM и рассчитаем полосу частот, необходимую для передачи сигналов изображения.

Формирования видеосигналов изображения. Сигналы цветности. Пусть минимальный размер деталей изображения составляет 1 пиксел (pixel) и равен ширине строки телевизионного изображения. Тогда по вертикали укладывается 625, по горизонтали $625 \cdot 4/3 \cong 830$ пикселов, а общее число элементов в кадре $\cong 520000$. Наивысшая частота сигнала получается, когда белые и черные пикселы чередуются, т.е. распределение яркости вдоль строки представляет последовательность прямоугольных импульсов со скважностью 2. Для передачи столь мелких деталей оказывается достаточно использовать первую гармонику.

Составим из чередующихся белых и черных пикселов изображение, напоминающее шахматную доску. Чтобы такое изображение повторялось 50 раз в секунду (прогрессивная развертка), необходим синусоидальный сигнал с частотой

$$f_{max} = 625 \cdot 415 \cdot 50 \cong 13 \text{ МГц.}$$

Эта частота и определяет полосу частот телевизионных каналов. Как следует из формулы, полосу частот можно сократить, уменьшая число строк или число кадров в секунду. Но уменьшение числа строк вызывает ухудшение качества изображения, а уменьшение числа кадров – мерцание экрана¹². Противоречие между необходимостью сокращать полосу частот и качеством изображения устраняется при использовании чересстрочной развертки.

При послесвечении экрана $\cong 20$ мс каждое новое поле из нечетных строк совмещается с предыдущим полем из четных строк и наоборот, т.е. создается впечатление построчной развертки, и $f_{max} = 6.5$ МГц. Реально в развертке изображения участвуют 575–589 строк, а остальные образуют полукадровый и кадровый гасящие импульсы и $f_{max} \cong 6$ МГц.

Не вдаваясь в детали цветового зрения, отметим, что красный (Red), зеленый (Green) и синий (Blue) цвета являются независимыми, и любой цвет можно получить в виде комбинации этих *основных* цветов. Яркостный сигнал белого цвета (E_Y) является следующей комбинацией основных:

$$E_Y = 0.299E_R + 0.587E_G + 0.114E_B,$$

и для передачи любых цветов вводятся два дополнительных цветоразностных сигнала: $E_{R-Y} = E_R - E_Y = 0.701E_R - 0.587E_G - 0.114E_B$ и $E_{B-Y} = E_B - E_Y = -0.299E_R - 0.587E_G + 0.886E_B$. Тогда, например, сигнал зеленого цвета определяется равенством $E_G = E_Y - 0.51E_{R-Y} - 0.19E_{B-Y}$.

Трехцветность характеризует только крупные детали изображений, передаваемые относительно низкими видеочастотами. Мелкие детали с

¹² Для яркости экрана $30 \div 100 \text{ кд/м}^2$ и послесвечения 0.02 с критическая частота мерцаний составляет 46 Гц.

видеочастотами от 1.5 МГц различаются только по яркости и воспринимаются как черно-белые. Для получения цветного изображения крупных деталей необходимо в каждой строке кроме яркостного сигнала E_Y , имеющего полосу по крайней мере 6 МГц, передать в полосе частот от 50 Гц до 1.5 МГц два цветоразностных сигнала, почти не отличающихся по структуре от E_Y . Во всех стандартных системах цветного телевидения принято это отношение полос частот сигналов как 4:1.

Итак, для цветоразностных сигналов E_{R-Y} и E_{B-Y} необходима область частот, уже занятая яркостным сигналом. Поэтому сигналы E_{R-Y} и E_{B-Y} переносятся в область более высоких частот, где гармоники E_Y имеют пренебрежимо малые амплитуды, и в системе SECAM поочередно (E_{R-Y} – в нечетных, E_{B-Y} – в четных строках) модулируют по частоте разные гармонические колебания (поднесущие): $f_{0R} = 4.40625$ МГц, девиация частоты $\Delta f_R = \pm 0.28$ МГц, а $f_{0B} = 4.250$ МГц, $\Delta f_B = \pm 0.23$ МГц. Индексы ЧМ оказываются меньше 0.5, и спектры ЧМ-сигналов практически не отличаются от спектров при АМ (см. разд. 5.3.1).

Для создания полноцветных изображений нужны одновременно два цветоразностных сигнала, а формируется только один. Недостающая информация восполняется при задержке другого сигнала, формируемого в предыдущей строке, на время, равное периоду строчной развертки T_c . Отсюда и название системы *Sequentiel couleurs a mémoire* – «Последовательные цвета с запоминанием». Применение линии задержки ведет к тому, что сигналы строк попарно повторяются, а цветовая синхронизация осуществляется с помощью специальных сигналов опознавания.

В американской системе NTSC цветоразностные сигналы одновременно модулируют по амплитуде квадратурные компоненты гармонического колебания с частотой 3.58 МГц (поднесущую) с помощью квадратурной амплитудной модуляции (КАМ). Подчеркнем, что это лишь поднесущие колебания, которые вместе с другими компонентами видеосигнала в дальнейшем используются для модуляции несущего колебания.

Передача телевизионных сигналов по радиоканалам. Видеосигнал изображения вместе с сигналом звукового сопровождения передаются в аналоговом виде по эфирному радиоканалу. Формирование телевизионного радиосигнала осуществляется с использованием амплитудной и частотной модуляции: видеосигнал передается с помощью амплитудной модуляции, а цветоразностные сигналы и аудиосигнал – путем частотной модуляции.

Как известно, при модуляции частота несущего колебания по крайней мере в 6–8 раз должна превышать наибольшую частоту модулирующего сигнала. Поэтому для несущих $f_{из}$ сигналов изображения выбран диапазон частот выше 50 МГц (несущая частота Первого канала российского телевидения 49.75 МГц). Несущие $f_{зв}$ звукового сопровождения устанавливаются на 6.5 МГц выше несущих изображения.

Полное описание телевизионного видеосигнала (огibaющей радиосигнала) имеет вид:

$$U(t) = U_Y \{ \cos(2\pi f_{из} t) + u_R \cos[2\pi(f_{0R} t + \Delta f_R U_{R-Y}(t))] + \\ + u_B \cos[2\pi(f_{0B} t + \Delta f_B U_{B-Y}(t + T_c))] \},$$

где u_R и u_B – масштабные множители.

Обычный АМ-сигнал содержит две боковые полосы (см. разд. 5.2.1) и в нашем случае занимает спектр шириной $12 \div 13$ МГц. Для сокращения спектра нижняя боковая полоса частично подавляется.

Подавление одной боковой полосы вызывает паразитную фазовую модуляцию и нелинейные искажения, которые проявляются тем слабее, чем выше частоты гармоник модулирующего сигнала. Поэтому частоты видеосигнала $U(t)$ от 50 Гц до 1.25 МГц, содержащие до 95% его мощности, передаются двумя боковыми полосами (рис. 5.36). Искажения более высоких частот из-за отсутствия второй боковой полосы остаются практически незаметными. ЧМ-сигнал звукового сопровождения обычно занимает полосу 100 кГц (девиация частоты ± 50 кГц). Если $f_{зв} - f_{из} = 6.5$ МГц, то номинальная ширина полосы частот радиоканала составляет 8 МГц. Полный спектр телевизионного радиосигнала представлен на рис. 5.36.

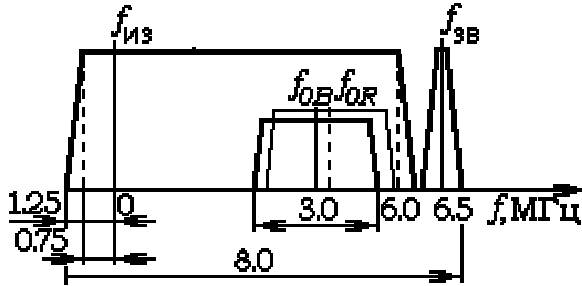


Рис. 5.36. Полосы частот телевизионного сигнала

ки незаметными. ЧМ-сигнал звукового сопровождения обычно занимает полосу 100 кГц (девиация частоты ± 50 кГц). Если $f_{зв} - f_{из} = 6.5$ МГц, то номинальная ширина полосы частот радиоканала составляет 8 МГц. Полный спектр телевизионного радиосигнала представлен на рис. 5.36.

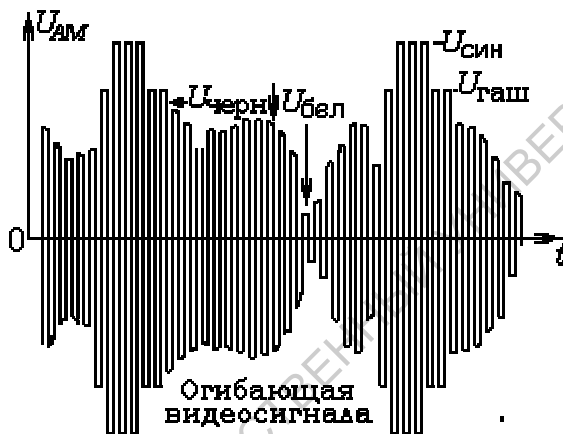


Рис. 5.37. Огибающая АМ-радиосигнала изображения

Огибающая АМ-сигнала для одной строки изображения представлена на рис. 5.37. Сигнал цветовой синхронизации на рисунке отсутствует. Если размах синхронизирующих импульсов принять за 100% ($U_{син}$), то амплитуда несущей, соответствующая видеосигналу, изменяется от $10 \div 15\%$ ($U_{бел}$) до $75 \div 80\%$ ($U_{черн}$). Напомним, что амплитуда синхронизирующих импульсов превышает уровень гасящих обратный ход ($U_{гаш}$) и зависит от среднего уровня яркости сюжета.

5.7.2. Сигналы цифрового телевидения

Развитие современного телевидения связано с несколькими проблемами. Прежде всего, это нехватка незанятых частот для трансляции новых телевизионных программ и невысокое качество изображения и звука. Несколько нам известно, в России только несколько каналов имеют стереофоническое сопровождение. Если к этому добавить большую мощность передатчиков и невысокую помехоустойчивость, то создается неприглядная картина. Между тем в ряде стран такие проблемы успешно решаются путем использования цифровых технологий. Так, по имеющимся сведениям, в Великобритании все телевидение является цифровым. Это позволяет вместо одной аналоговой программы в том же частотном диапазоне транслировать по крайней мере 4 программы с более высоким

качеством, а также передавать дополнительную и вспомогательную информацию.

Цифровым телевидением (Digital Television) принято называть радиоэлектронную систему, выполняющую преобразования оптического изображения в цифровой сигнал, сжатие, передачу и обратное преобразование цифрового сигнала в оптическое изображение.

Целью настоящего раздела является знакомство с принципами цифрового телевидения, методами обработки и модуляции телевизионных сигналов.

Дискретизация, кодирование и многопозиционная модуляция телевизионных сигналов. По принятому соглашению дискретизация и кодирование яркостного E_Y и цветоразностных сигналов E_{R-Y} и E_{B-Y} в системах цифрового телевидения осуществляются отдельно. Если наибольшую частоту аналогового яркостного радиосигнала принять равной 6.5 МГц, то согласно теореме Котельникова – Шенона минимальная частота дискретизации составит 13 МГц (см. разд. 4.2). В соответствии с рекомендацией Международного союза электросвязи (МСЭ) для дискретизации E_Y установлена частота 13.5 МГц – минимальная частота, кратная частотам 625 и 525 строк при разложении изображения и позволяющая получать целое число отсчетов в строке. Для дискретизации цветоразностных сигналов выбрана вдвое меньшая частота, равная 6.75 МГц (отношение протяженностей полос 4:2:2). Более высокое качество изображения можно получить, если частоты дискретизации трех основных цветов изображения будут одинаковыми и равными 13.5 МГц (отношение протяженностей полос 4:4:4).

Квантование и последующее кодирование составляют следующий этап в создании цифровых сигналов E_Y , E_{R-Y} и E_{B-Y} . Необходимое число уровней квантования определяется экспериментально. Приемлемым по уровню шумов квантования оказалось 8-разрядное (8-битовое) линейное кодирование. Для исключения различных нестабильностей предложено использовать только 220 из 256 уровней квантования. При этом для позитивной полярности сигналов черному и белому присваиваются 16 и 235 уровни, что обеспечивает запас для непредвиденных изменений видеосигнала. Самые последние достижения в области цифрового кодирования связаны с использованием в студийной аппаратуре высокого качества 10-разрядных преобразований. Важно отметить, что увеличение разрядности кодирования происходит отнюдь не «бесплатно», так как требует соответствующего увеличения скорости цифрового потока. При 8-разрядном кодировании яркостного сигнала E_Y с частотой дискретизации 13.5 МГц скорость цифрового потока должна быть $8 \cdot 13.5 \cdot 10^6 = 108$ Мбит/с, а полного цветного потока – 216 Мбит/с. При 10-разрядном кодировании трех основных цветов с повышенным качеством потребуется скорость цифрового потока уже 405 Мбит/с.

Если результаты кодирования передавать в канал связи с помощью двоичных импульсно-кодовых последовательностей (ИКМ), то скорость передачи данных не может превышать так называемый *барьер Найквиста*, равный $\frac{2 \text{бит/с}}{\Gamma_{\text{ц}}}$. Естественный выход состоит в использовании многопозиционной модуляции, при которой каждый передаваемый сим-

вол содержит более 1 бита информации. Как было показано выше, к таким видам модуляции относятся квадратурная амплитудная манипуляция (QAM), квадратурная фазовая манипуляция (QPSK), частотная модуляция с ортогональными несущими (OFDM).

Безусловным прорывом в цифровом кодировании явилось создание системы сжатия звуковых и видеосигналов MPEG-2, позволяющей сократить занимаемую полосу частот. Система MPEG-2 включает несколько процедур, действующих после предварительной пространственной и временной нелинейной фильтрации и устраняющих «избыточность» телевизионного сигнала. Различают три вида избыточности:

1) поскольку большую часть изменений в двух последовательных кадрах можно интерпретировать как смещение его малых фрагментов, изображение разбивают на блоки по 16×16 элементов и для каждого блока определяют набор параметров, показывающий направление и величину смещения – вектор смещения, т.е. вместо дискретизации нового кадра формируются небольшие разности между текущими и предсказуемыми блоками (пространственная избыточность). Вектор трансформируется в частотную область и концентрируется в узкой полосе низких частот;

2) описанный блок изображения из 256 элементов преобразуется в равное число коэффициентов. Квантование производится над набором коэффициентов в целом (адаптивное векторное квантование);

3) так как не все значения векторов движения и коэффициентов равновероятны, применяется статистическое кодирование с переменной длиной кодового слова, и событиям с наибольшей вероятностью присваиваются короткие кодовые слова (устраняется амплитудная избыточность). Стандарт MPEG-2 предусматривает объединение отдельных потоков изображения, звука, синхронизации одной или нескольких программ в единый цифровой поток.

На сегодняшний день в Европе, США и Японии разработаны различные стандарты цифрового телевидения. Они решают одинаковые задачи – улучшение качества вещания при сокращении частотного диапазона и все используют систему сжатия информации MPEG-2. Европейский стандарт DVB (Digital Video Broadcasting – цифровое видеовещание) разделяет цифровое телевидение на три вида: эфирное (наземное), спутниковое и кабельное, в зависимости от способа передачи цифровых телесигналов. Американский ATSC (Advanced Television Systems Committee) и японский ISDB (Integrated Services Digital Broadcasting) являются только эфирными (наземными).

Цифровое эфирное (наземное) телевидение DVB-T. Основные достоинства цифровой обработки сигналов проявляются и в цифровом телевидении, а недостатки и погрешности передачи цифровых сигналов зависят от вида телевизионного цифрового вещания. Для наземного вещания (особенно в городах) погрешности приема прежде всего связаны с многократными отражениями радиосигналов.

Когда на вход декодера одновременно поступает несколько цифровых последовательностей, сдвинутых во времени в результате отражений, резко возрастают цифровые ошибки. Для устранения описанных помех был специально разработан метод кодированной OFDM-модуляции – COFDM. Как отмечалось выше, при OFDM последовательный цифро-

вой поток преобразуется в большое число (обычно 1705 или 6817) параллельных потоков (субпотоков), каждый из которых передается на отдельной несущей. Длительность параллельного цифрового потока оказывается достаточно большой, и оценка значений принятых символов задерживается в декодере на часть интервала субблока (защитный интервал) для компенсации отражений. Оценка значений производится на основной (рабочей) части интервала. Выбор числа несущих, величины их частотного разноса, длительностей защитного и рабочего интервалов информационного символа определяют характеристики телевизионного канала. В частности, если в рабочий интервал укладывается целое число разностной частоты Δf двух соседних несущих, то они ортогональны.

В стандарте DVB-T предусмотрены два режима OFDM с одинаковой шириной спектра группового сигнала 7.61 МГц, отличающиеся длительностью рабочих интервалов информационных символов и получивших названия **8К** (рабочий интервал 896 мкс, $\Delta f = 1/896$ мкс = 1116 Гц, 6817 несущих) и **2К** (рабочий интервал 224 мкс, $\Delta f = 4464$ Гц, 1705 несущих). При такой ширине спектра группового сигнала OFDM хорошо согласуется с шириной спектра аналоговых каналов (8 МГц). Технически режим **2К** более прост и не требует микропроцессоров большой ёмкости для проведения прямого и обратного БПФ, но режим **8К** позволяет увеличить расстояние между ТВ-передатчиками и осуществлять цифровое эфирное вещание на больших территориях.

Как обычно (см. разд. 5.6.3), несущие OFDM-сигнала модулируются по фазе или по амплитуде и фазе. Стандартом DVB-T предусмотрены квадратурная фазовая или квадратурная амплитудно-фазовая манипуляции (QAM) в зависимости от требуемого качества изображения, необходимой скорости передачи данных и уровня помех. Модуляция 16-QAM обеспечивает скорости передачи более 10 Мбит/с, а 64-QAM – до 30 Мбит/с.

Для повышения помехоустойчивости и согласования форматов OFDM и MPEG-2 используется канальное кодирование (отсюда и название CODFM). Канальный кодер и декодер разделены на две части: внешнюю и внутреннюю. Это позволяет унифицировать ряд узлов для систем эфирного, спутникового и кабельного цифрового телевидения и сосредоточить их во внешней системе кодирования, а дополнительную, специфическую обработку данных проводить во внутренней системе. Такое разделение предусмотрено системой DVB-T и позволяет использовать программное обеспечение и новые интегральные схемы, разработанные для одной системы вещания, в других.

Цифровое спутниковое телевидение DVB-S. На первом этапе использовались ИСЗ, которые обращались по эллиптическим орбитам. В настоящее время применяются более простые и эффективные геостационарные спутники, движущиеся в экваториальной плоскости на высоте 35785 км. Спутник неподвижен относительно Земли и охватывает вещанием около трети ее поверхности. По радиоканалу с Земли спутник получает определенную информацию. Сигнал усиливается, преобразуется по частоте и излучается обратно на Землю.

На цифровое спутниковое телевидение изначально возлагались большие надежды. Однако для передачи оцифрованного цветного изображе-

ния (яркостного и двух цветоразностных сигналов) с хорошим качеством потребовалась скорость цифрового потока более 200 Мбит/с. При полосе пропускания спутникового ретранслятора $25 \div 35$ МГц такая скорость потока оказалась недостижимой, и первые спутниковые телевизионные системы были комбинированными цифро-аналоговыми. Только разработка алгоритмов цифрового сжатия (MPEG-1) позволила передавать изображение высокого качества со скоростью 7-9 Мбит/с. Дальнейшее развитие цифровых технологий и разработка MPEG-2 привели к созданию европейского стандарта телевизионного вещания DVB-S. Для спутникового видеовещания были выделены два участка сантиметрового диапазона волн с частотами $10.7 \div 11.75$ ГГц (*Ku*-диапазон) и $3 \div 4,5$ ГГц (*C*-диапазон).

Система спутникового вещания решает еще одну задачу: улучшение качества звукового сопровождения телепередач и создание нескольких каналов, например для одновременной передачи двуязычного вещания.

Энергетические возможности спутниковых передатчиков ограничены. Поэтому используются помехоустойчивое кодирование и эффективная модуляция с невысокими кратностями (квадратурная фазовая манипуляция QPSK вместо 16-QAM). Многопрограммное вещание осуществляется чаще всего за счет мультиплексирования N сжатых цифровых сигналов и передачи на одной несущей, но возможна также передача каждого из N сжатых сигналов на своей несущей. Количество программ, передаваемых одним спутником, зависит от необходимой скорости передачи информации и формата кодирования. Но реально, например в *Ku*-диапазоне, принимается несколько сотен программ. Важнейшее свойство системы DVB-S – возможность приема программ спутникового телевидения через небольшие параболические антенны на домашние интегральные приемники-декодеры.

Уже сегодня появилась усовершенствованная система спутникового вещания DVB-S2, позволившая вернуться к высокоскоростным способам модуляции 16-QAM и 32-QAM и использовать более эффективные форматы канального кодирования. Толчком для создания DVB-S2 стали планы запуска телевидения высокой четкости ТВЧ (HDTV – High Definition Television). Оказалось, что при переходе на ТВЧ недостаток частотного ресурса не смогут компенсировать даже более совершенные системы сжатия цифровых сигналов. К тому же качество приема, особенно в *Ku*-диапазоне, сильно зависело от погодных условий. Поэтому внедрение DVB-S2 с высокоскоростной модуляцией в настоящее время представляется перспективным для улучшения качества приема. Это подтверждают и первые экспериментальные результаты.

Звуковые сигналы отдельных каналов, как и сигналы изображения, преобразуются в цифровую форму и объединяются в общий цифровой поток, содержащий ИКМ последовательности и некоторые вспомогательные импульсы. Исходный звуковой сигнал дискретизируется с тактовой частотой 48 кГц и преобразуется в цифровую форму с точностью $10 \div 16$ бит/отсчет. Последняя соответствует качеству звучания CD-компакт-диска. Этот цифровой поток модулирует по фазе поднесущую звуковую частоту, расположенную, как и в случае аналогового сигнала, выше спектра видеосигнала. При скорости потока 2 Мбит/с образуются

либо два канала с полосой 20 кГц, либо 4 канала с полосой 14 кГц.

Особенности цифрового кабельного телевидения DVB-C. Кабельное телевидение является привлекательным способом получения доступа к большому количеству программ эфирного и спутникового вещания. Во многих случаях (в горных и удалённых районах) кабельное телевидение представляет единственную возможность получать высококачественное цветное изображение телевизионных программ. В настоящее время многие кабельные системы работают в режиме аналогового сигнала. При этом трудно рассчитывать на высокое качество приема. Переход на цифровое вещание затруднён тем, что телецентры не обеспечивают вещание в цифровом виде.

По пространственно-топологической структуре кабельная сеть может быть радиальной, кольцевой или древовидной. Радиальные сети аналогичны телефонным и могут быть объединены с ними. Большинство действующих кабельных сетей имеют древовидную структуру.

В последние годы во многих странах создаются современные кабельные сети на основе волоконно-оптических линий связи (ВОЛС). В таких линиях связи допускается частотное уплотнение. При этом полный многоканальный сигнал формируется на относительно низких частотах, а затем уже сформированный сигнал переносится на оптическую несущую. Возможно и другое решение, когда по каждому оптоволокну передается своя телепрограмма.

ВОЛС особенно эффективны при передаче информации в цифровом виде. Головная станция сети принимает аналоговые или цифровые эфирные сигналы и конвертирует их в цифровую форму. Одна или несколько спутниковых антенн принимают сигналы в формате DVB-S и преобразуют их в формат DVB-C. Помехоустойчивость ВОЛС очень высокая, и может использоваться квадратурная амплитудная модуляция от 16-QAM до 256-QAM, повышающая эффективность вещания в десятки раз.

5.7.3. На пути к телевидению высокой четкости

Телевидение высокой четкости имеет самое прямое отношение к тематике нашей книги, так как все основные идеи и достижения ТВЧ связаны с сигналами, их спектрами и методами цифрового представления.

Первый формат телевизионного экрана был заимствован у кинематографа, использовавшего киноленту с размером кадра 4×3 . Довольно долго кино- и телеформаты соответствовали друг другу. Ситуация изменилась с появлением широкоэкранных, а затем широкоформатных кинофильмов.

С 1964 года японская вещательная компания NHK, по ее выражению, «начала прокладывать путь телевидению следующего поколения» и в середине 80-х годов предложила стандарт HDTV с форматом изображения 16:9, соответствующим широкоформатным фильмам. Вслед за Японией свои стандарты предложили США и объединенная Европа. На сегодняшний день существует более 10 систем ТВЧ и постоянно разрабатываются новые стандарты ультравысокой четкости (Ultra HDTV, SHV-Super Hi-Vision и другие). При разработке стандартов высокой четкости в первую очередь достигается соответствие параметров видеоизоб-

ражения психофизиологическим особенностям зрительного восприятия и совместимость ТВЧ-системы с существующими системами цветного телевидения SECAM, PAL и NTSC.

Здесь желательны учитывать основные критерии комфортного просмотра телевизионных передач. Для изображения такими критериями являются эффект присутствия зрителя в действии, происходящем на экране телевизора, и высокое качество воспроизведения мелких деталей (максимальное сглаживание точечной структуры изображения), а для звукового сопровождения – создание равномерного объемного звучания, позволяющего точно определить позицию источника на экране. Рассмотрим подробнее, как эти критерии можно реализовать в телевидении высокой четкости.

1. На основе психофизиологических исследований и экспертных оценок установлено, что бинокулярное поле зрения человека простирается на 30° вправо и влево и несколько меньше по вертикали, а угол «ясного» зрения при просмотре телепрограмм составляет всего 30° по горизонтали и около 12° по вертикали. Было показано, что при восприятии изображения формата 16:9 с расстояния, в 3 раза превышающего высоту изображения на экране, участвуют периферические области сетчатки, уменьшающие влияние границ изображения. Чем крупнее экран, тем легче создать эффект максимального присутствия зрителя.

2. Степень воспроизведения мелких деталей в вертикальном направлении характеризует частота строк, а в горизонтальном – число пикселей в строке. Если угловую разрешающую способность человека принять равной одной угловой минуте, то слияние строк и элементов строки в изображении для зрителя будет происходить, когда число строк не менее 720 ($12^\circ \times 60'/1'$), а соответствующее число пикселей в строке не менее 1280. При таких параметрах кадра, чересстрочной развертке и последовательной передаче яркостного и цветоразностных сигналов необходимая полоса частот составила бы 46 МГц, а скорость передачи оцифрованного 8-разрядным кодом цифрового потока – 368 Мбит/с¹³. Очевидно, что эти требования невыполнимы, и только использование новейших методов сжатия телевизионных сигналов позволяет реализовать ТВЧ уже сегодня.

Система HDTV предполагает формат изображения 16:9 и поддерживает три стандарта (символы p и i означают прогрессивную (progressive scanning) и чересстрочную (interlaced scanning) развертки:

формат 720p содержит 720 строк, общее число пикселей 921000 (1280×720) и прогрессивную развертку с частотой кадров 24, 25, 30, 50 или 60 кадров/с;

формат 1080i содержит 1080 строк, общее число пикселей около двух миллионов (1920×1080) и чересстрочную развертку с частотой кадров 50 или 60 кадров/с;

формат 1080p содержит 1080 строк, такое же число пикселей (1920×1080) и прогрессивную развертку с частотой кадров 24, 25 или 30 кадров/с.

3. Звуковое сопровождение HDTV сосредоточено в одном стереофайле

¹³Полученные результаты предлагается проверить самостоятельно, используя методику, описанную при расчете полосы частот аналогового телесигнала.

и соответствует стандарту Dolby Digital 5.1, рассчитанному на 6 каналов. В телевизорах звук разводится на 6 колонок аналогично тому, как это делается в домашних кинотеатрах. Из них 5 колонок – основные широкополосные, а одна – узкополосная для низкочастотных эффектов. Цифровой звук создает «обтекающий слушателя эффект» и позволяют точно определить позицию источника на экране. Описанная система ТВЧ уже функционирует в ряде стран и начинает формироваться в России.

В последние годы все чаще обсуждаются возможности дальнейшего совершенствования ТВЧ. Лидерами здесь опять выступают японские ученые и инженеры. В системе UHDVT (Ultra HDVT) планируется создать разрешение 7680×4320 пикселей (с общим объемом 33 Мпикселя), улучшить цветопередачу и яркость изображения. При частоте 60 кадров/с развертка будет прогрессивной. Для звукового сопровождения предложен стандарт 22.2, согласно которому 10 широкополосных колонок располагаются на уровне ушей, 9 – выше и 3 – ниже этого уровня. Еще 2 узкополосных колонки предназначены для низкочастотных эффектов. Предполагается, что полная полоса частот составит 600 МГц, а скорость передачи цифровых потоков будет достигать 6.6 Гбит/с.

По данным периодической печати, Япония уже дважды демонстрировала возможности системы UHDTV (на выставке в 2005 г. и на Международном конгрессе в Голландии в сентябре 2006 г.). Изображение проецировалось на экран с диагональю до 11,4 м. Однако пока это только заманчивый проект, демонстрирующий неограниченные возможности современной радиофизики.

5.8. Контрольные вопросы и упражнения

1. Какой физический процесс называется модуляцией колебаний? Каковы виды модуляции несущего гармонического колебания и периодической последовательности импульсов?
2. Каково соотношение частот несущего и модулирующего сигналов?
3. Как различаются амплитудно-фазовая и квадратурная амплитудная модуляция?
4. Что такое двухступенчатая модуляция?
5. Почему возникают искажения при передаче перемодулированных сигналов?
6. Какие составляющие АМ-сигнала переносят информационный сигнал?
7. Представьте однотональный и многотональный АМ-сигналы в аналитическом виде. Как различаются их спектры?
8. Спектр сигнала содержит пять гармоник, амплитуды частоты и фазы которых соответственно равны: 1) 3 В, 98 кГц, 0.6 рад; 2) 4 В, 99 кГц, 0.4 рад; 3) 20 В, 100 кГц, 1.2 рад; 4) 4 В, 101 кГц, 2 рад; 5) 3 В, 102 кГц, 1.8 рад. Записать аналитическое выражение сигнала, определить его тип и параметры, изобразить спектр.
9. Изобразите спектр АИМ-сигнала. В чем особенности его спектра?
10. Какие соотношения связывают девиацию частоты и девиацию фазы в частотно- и фазо-модулированных сигналах?

11. Что такое индекс частотной модуляции? Когда его можно использовать и когда не следует?
12. Мгновенная частота УМ-колебания описывается соотношением $f(t) = 10^8 \cdot (1 + 10^{-4} \sin(2\pi \cdot 2500t))$. Определить девиацию фазы $\Delta\Psi$, индекс частотной модуляции β однотонового сигнала и записать аналитическое выражение для сигнала.
13. Как выглядит спектральное представление ЧМ-сигнала при малом и произвольном индексах модуляции?
14. В спектре однотонового ЧМ-сигнала (несущая частота 0.3 МГц, частота модулирующего колебания 10 кГц) отсутствуют гармоники с частотами 290 и 310 кГц. Используя таблицу функций Бесселя, определить девиацию частоты и рассчитать частоты и относительные амплитуды гармоник сигнала.
15. Каковы особенности спектров ЛЧМ-сигналов? Как зависит структура спектров от базы сигналов?
16. Пусть сигнал $X(t)$ со случайной модуляцией амплитуды задан выражением (5.41). Считая, что случайная амплитуда $\rho(t) = A_0(1 + \xi(t))$ распределена по закону Релея (5.47) и $\bar{\xi} = 0$, определить величину $\bar{\xi}^2$ и связь параметра релеевского распределения λ с постоянной амплитудой A_0 .
17. Пусть гармонический сигнал промодулирован по амплитуде и фазе статистически независимыми случайными сигналами и его можно представить в виде $X(t) = \rho(t) \cos(\omega_0 t + \varphi(t))$, где $\omega_0 = \text{const}$, $\rho(t)$ – стационарный случайный процесс со средним значением $\bar{\rho} = a_0$ и корреляционной функцией $\Psi_\rho(\tau) = \sigma_\rho^2 e^{-\alpha|\tau|}$, $\tau = t_2 - t_1$, а $\varphi(t)$ – винеровский процесс с коэффициентом диффузии B_φ и нулевым начальным значением. Используя независимость амплитуды и фазы, записать выражения для ковариационной функции и спектральной плотности мощности процесса $X(t)$ в установившемся режиме.
18. Что означает термин «низкочастотное представление» узкополосного сигнала?
19. Какова роль аналитических сигналов в теории узкополосных процессов?
20. Как выглядят квадратурные компоненты нормального узкополосного шума?
21. Назовите основные виды цифровой амплитудной манипуляции. В чем особенности ИКМ- и дельта-модуляции?
22. Назовите основные виды фазовой манипуляции. Что такое квадратурная фазовая манипуляция? Чем квадратурная амплитудно-фазовая манипуляция отличается от квадратурной ФМн?
23. Изобразите сигнальные созвездия для 16-QAM и 64-QAM. Что означает число 16 в обозначении 16-QAM?
24. В чем заключаются принципы многочастотной цифровой модуляции? Как используются цифровые методы в радиовещании?
25. Что такое чересстрочная развертка? Как формируется кадр при чересстрочной развертке?
26. Что представляют собой цветоразностные сигналы? Как отличается формирование цветоразностных сигналов в системах SEKAM и NTSC?

27. Как рассчитать максимальную частоту спектра телевизионного сигнала?

28. Что такое телевидение высокой четкости? Какие стандарты ТВЧ существуют сегодня и каковы перспективы развития ТВЧ?

Список литературы

Баскаков, С. И. Радиотехнические цепи и сигналы : учебник. М. : Высш. шк., 2000.

Нефедов, В. И. Основы радиоэлектроники и связи : учебник для вузов. 3-е изд., исправ. М. : Высш. шк., 2005.

Хохлов, А. В. Теоретические основы радиоэлектроники : учеб. пособие. Саратов: Изд-во Саратов. ун-та, 2005.

Гоноровский, И. С. Радиотехнические цепи и сигналы : учебник. М. : Радио и связь, 1986.

Информационные технологии в радиотехнических системах : учеб. пособие. 2-е изд., перераб. и доп./В. А. Васин, И. Б. Власов, Ю. М. Егоров и др.; под ред. И. Б. Федорова. М. : Изд-во МГТУ им. Н. Э. Баумана, 2004.

Рытов, С. М. Введение в статистическую радиофизику: в 2 ч. Ч. 1: Случайные процессы. М. : Наука, 1976.

Ахманов, С. А., Дьяков, Ю. Е., Чиркин, А. С. Статистическая радиофизика и оптика. Случайные волны в линейных системах. М. : Физматлит, 2010.

Тихонов, В. И., Харисов, В. Н. Статистический анализ и синтез радиотехнических устройств и систем. М. : Радио и связь, 2004.

Сергиенко, А. Б. Цифровая обработка сигналов: учебник. СПб. : Питер, 2002.

Вайнштейн, Л. А., Ширман, Д. Е. Разделение частот в теории колебаний и волн. М. : Наука, 1983.

Цифровая обработка сигналов и изображений в радиофизических приложениях/М. А. Бесараб, В. К. Волосюк, О. В. Горячкин и др.; под ред. В. Ф. Кравченко. М. : Физматлит, 2007.

Шахнович, И. Цифровое радиовещание – последний штрих цифрового портрета. Электроника // Наука, технология, бизнес. 2002. Вып. 3, 2002.

Быков, Р. Е. Теоретические основы телевидения. СПб. : Лань, 1998.

Крыжановский, В. Д., Костыков, Ю. В. Телевидение цветное и черно-белое. М. : Связь, 1980.

Птачек, М. Цифровое телевидение. Теория и практика / пер. с чеш. под ред. Л. С. Виленчика. М. : Радио и связь, 1990.

Заключение

Эта книга представляет первую часть большого учебника, задуманного авторами под общим названием «Радиоэлектроника и статистическая радиофизика».

Современная радиоэлектроника и статистическая радиофизика для описания свойств и возможных применений детерминированных, случайных и цифровых сигналов использует разнообразные математические и физические методы. В настоящем издании показано, как строятся модели различных сигналов, какую роль играет статистический подход к анализу частотно-временных и информационных свойств сигналов.

Авторы старались по-возможности подробно описывать различные стороны рассматриваемых явлений, их взаимную связь и пути практического использования. Нас особенно увлекало богатство развитых статистических моделей и методов модуляции сигналов, необъятные перспективы использования цифровых сигналов и цифровых технологий в радиоэлектронике. При этом многие вопросы, касающиеся радиолокации, телефонии, мобильной связи и Интернета, оказались нерассмотренными. Нас частично оправдывает то, что названные вопросы относятся не столько к сигналам, сколько к системам, а используемые в этих системах сигналы, как нам представляется, описаны достаточно подробно.

Авторы надеются, что представленная книга будет полезна студентам и аспирантам вузов и уже при первом чтении не вызовет желания отложить ее изучение до лучших времен.

Оглавление

Предисловие.	3
Введение	4
Глава 1. Элементы общей теории сигналов	6
1.1. Классификация электрических сигналов	6
1.2. Линейные пространства сигналов	10
1.3. Сигналы и информация	19
1.3.1. Сведения из теории вероятностей	20
1.3.2. Дискретные источники сообщений	26
1.3.3. Непрерывные источники сообщений	32
1.4. Контрольные вопросы и упражнения	35
Список литературы.	36
Глава 2. Детерминированные колебания и волны. Методы описания	37
2.1. Виды детерминированных сигналов.	37
2.2. Динамическое представление сигналов	42
2.3. Периодические сигналы и основы гармонического анализа.	45
2.4. Гармонический анализ непериодических сигналов	49
2.4.1. Спектральные плотности интегрируемых сигналов	52
2.4.2. Связь ширины спектра с длительностью и формой импульсных сигналов	57
2.4.3. Спектральные плотности неинтегрируемых сигналов	58
2.5. Спектральные плотности комплексных функций	62
2.6. Корреляционный анализ сигналов и энергетические спектры	65
2.7. Преобразование Лапласа как обобщение преобразования Фурье	67
2.8. Регулярные (детерминированные) волны	73
2.8.1. Общие сведения о волновых процессах	74
2.8.2. Основные свойства электромагнитных волн	76
2.9. Контрольные вопросы и упражнения	80
Список литературы.	82
Глава 3. Случайные процессы, поля и волны.	83
3.1. Случайные процессы и их описание	83
3.2. Вероятностная сходимость. Непрерывность. Производная и интеграл от случайного процесса	89
3.3. Стационарные и эргодические случайные процессы	92
3.4. Спектральное представление случайных процессов	96
3.5. Нормальные (гауссовские) случайные процессы	103
3.6. Марковские случайные процессы	106
3.6.1. Марковские цепи.	108
3.6.2. Дискретные марковские процессы	112
3.6.3. Диффузионные марковские процессы	116
3.6.4. Диффузионный процесс Винера – Леви	122
3.6.5. Стохастические дифференциальные уравнения (СДУ).	125
3.7. Случайные поля и случайные волны	132
3.7.1. Случайные поля и их характеристики	132
3.7.2. Случайные волны и их характеристики	139
3.8. Контрольные вопросы и упражнения	144
Список литературы	146

Глава 4. Дискретные и цифровые сигналы. Методы цифровой обработки.	148
4.1. Дискретное представление сигналов	148
4.2. Восстановление аналогового сигнала по его равномерной выборке.	156
4.3. Представление дискретных сигналов в цифровой форме	159
4.4. Цифровой спектральный анализ.	166
4.4.1. Дискретное преобразование Фурье	167
4.4.2. Быстрое преобразование Фурье	176
4.5. Спектр мощности дискретных случайных сигналов.	182
4.6. Дискретные преобразование Лапласа и z-преобразование	184
4.7. Контрольные вопросы и упражнения	186
Список литературы	187
Глава 5. Модулированные и узкополосные сигналы.	188
5.1. Основные задачи и методы модуляции	188
5.2. Амплитудно-модулированные сигналы.	192
5.2.1. Модуляция гармонических колебаний	192
5.2.2. Амплитудная модуляция импульсных последовательностей.	197
5.3. Сигналы с угловой модуляцией	199
5.3.1. Фазо-модулированные сигналы и их спектры	199
5.3.2. Частотно-модулированные сигналы и их спектры	203
5.3.3. Сигналы с внутриимпульсной линейной ЧМ.	206
5.3.4. Частотно- и фазоманипулированные сигналы.	209
5.4. Случайная модуляция гармонических сигналов	211
5.4.1. Случайная амплитудная модуляция	212
5.4.2. Случайная угловая модуляция.	215
5.5. Узкополосные процессы и их спектры	219
5.5.1. Вещественные модели детерминированных узкополосных сигналов.	220
5.5.2. Комплексное представление детерминированных узкополосных сигналов	222
5.5.3. Огибающая, фаза и квадратурные составляющие случайного процесса	225
5.6. Принципы цифровой модуляции сигналов	227
5.6.1. Цифровая амплитудная модуляция сигналов	228
5.6.2. Цифровая фазовая манипуляция сигналов.	230
5.6.3. Цифровая частотная манипуляция сигналов.	233
5.6.4. Методы цифровой модуляции и перспективы цифрового радиовещания.	235
5.7. Телевизионные сигналы и их спектры	236
5.7.1. Сигналы аналогового эфирного телевидения	236
5.7.2. Сигналы цифрового телевидения	242
5.7.3. На пути к телевидению высокой четкости	247
5.8. Контрольные вопросы и упражнения	249
Список литературы	251
Заключение	252

Учебное издание

*Хохлов Артур Вениаминович,
Вадивасова Татьяна Евгеньевна,
Шабунин Алексей Владимирович*

**СИГНАЛЫ
МЕТОДЫ ОПИСАНИЯ, МОДЕЛИ,
ИНФОРМАЦИОННЫЕ ВОЗМОЖНОСТИ**

*Учебное пособие
для студентов физических специальностей*

Под редакцией профессора *В. С. Анищенко*

Редактор Т. А. Трубникова
Технический редактор Л. В. Агальцова
Корректор С. Ю. Малова
Оригинал-макет подготовил А. В. Хохлов

Подписано в печать 21.11.11. Формат 60x84 1/16.
Усл. печ. л. 14,88(16). Тираж 100. Заказ 283 -Т.

Издательство Саратовского университета.
410012, Саратов, Астраханская, 83.
Типография Саратовского университета.
410012, Саратов, Б, Казачья, 121 А.