

**Парадигма развития науки**  
**Методологическое обеспечение**

**А.Е. Кононюк**

**ИНФОРМАЦИОЛОГИЯ**  
**ОБЩАЯ ТЕОРИЯ ИНФОРМАЦИИ**

**Книга 4**

**Киев**  
**«Освіта України »**  
**2011**

**УДК 51 (075.8)**

**ББК В161.я7**

**К65.**

Рецензент: *Н.К. Печурин* - д-р техн. наук, проф. (Национальный авиационный университет).

**Кононюк А.Е.**

**К65 Информациология. Общая теория информации**

**К.: "Освіта України ", 2011. Книга 4, 488 с.**

**ISBN 978-966-7699-50-8**

В настоящей работе в последовательной систематизированной форме излагаются основы информациологии как всеобъемлющей теории о естественной и искусственной информации – информации природы Вселенной и информации, созданной человеком. Дается формализация понятий информации и информационных процессов как абстрактных понятий познания и формализация самой теории информациологии. Концепция информациологии дается в аспекте социальной и природной значимости информации для отдельного человека, коллектива и общества в целом, а также – в свете информационного единства человека и природы, единства всех форм и типов информации, всех процессов информационного взаимодействия, процессов самоинформатизации Вселенной и процессов социальной (государственной) информатизации. Изложены методологические основы информациологии, базирующиеся на фундаментальном принципе информациологического подхода, на системном сочетании интеграционного и дифференцированного подходов к исследованию. Монография может быть полезной для людей самых различных специальностей – филологов, информациологов, математиков, лингвистов, правоведов, методистов, преподавателей ВУЗов, научных работников, магистров, аспирантов, докторантов, всех, кто интересуется проблемами информации, информатики, информатизации и информациологии в целом.

**ББК В161.я7**

**ISBN 978-966-7699-50-8**

**©А.Е. Кононюк, 2011**

## **Оглавление**

17. Пространство и метрология сигналов.....	6
17.1. Основные понятия и определения.....	6
17.2. Типы сигналов.....	27
17.3. Системы преобразования сигналов.....	36
17.4. Информационная емкость сигналов.....	41
17.5. Пространство сигналов.....	51
17.6. Мощность и теория сигналов.....	67
17.7. Пространства функций.....	70
17.8. Функции корреляции сигналов.....	77
17.9. Математическое описание шумов и помех.....	82
18. Динамическая и спектральная форма представления сигналов.....	85
18.1. Разложение сигналов по единичным импульсам.....	85
18.2. Свертка (конволюция) сигналов.....	91
18.3. Введение в спектральное разложение сигналов.....	99
18.4. Разложение сигнала по гармоническим функциям.....	103
18.5. Непрерывные преобразования Фурье и Лапласа.....	114
18.6. Свойства преобразований Фурье.....	123
18.7. Спектры некоторых сигналов.....	131
18.8. Введение в энергетические спектры сигналов.....	140
18.9. Энергетические спектры сигналов.....	141
18.10. Введение в корреляцию сигналов.....	147
18.11. Автокорреляционные функции сигналов.....	148
18.12. Взаимные корреляционные функции сигналов.....	155
18.13. Спектральные плотности корреляционных функций.....	159
19. Дискретизация сигналов и преобразований.....	165
19.1. Введение в дискретизацию сигналов.....	165
19.2. Задачи дискретизации функций.....	167
19.3. Равномерная дискретизация.....	169
19.4. Дискретизация по критерию наибольшего отклонения.....	189
19.5. Адаптивная дискретизация.....	190
19.6. Квантование сигналов.....	192
19.7. Децимация и интерполяция данных.....	193
19.8. Введение в дискретные преобразования сигналов.....	194
19.9. Преобразование Фурье.....	195
19.10. Преобразование Лапласа.....	201
19.11. Z-преобразование сигналов.....	202
19.12. Дискретная свертка (конволюция).....	208
20. Случайные процессы и сигналы.....	211
20.1. Введение в случайные процессы и сигналы.....	211

20.2. Случайные процессы и функции.....	213
20.3. Функции спектральной плотности .....	225
20.4. Преобразования случайных функций.....	234
20.5. Модели случайных сигналов и помех.....	242
21. Стационарные линейные системы.....	246
21.1. Введение в стационарные линейные системы.....	247
21.2. Линейные системы.....	255
21.3. Импульсная характеристика системы.....	256
21.4. Передаточные функции цифровых систем.....	261
21.5. Частотные характеристики систем.....	263
21.6. Реакция систем на случайные сигналы.....	265
21.7. Структурные схемы систем.....	269
21.8. Нерекурсивная фильтрация сигналов.....	272
21.9. Преобразование Хартли.....	280
21.10. Свойства преобразования.....	284
21.11. Дискретное преобразование Хартли .....	291
22. Многомерные сигналы и системы.....	295
22.1. Двумерные и многомерные сигналы.....	295
22.2. Двумерные системы.....	301
22.3. Частотные характеристики сигналов и систем.....	305
22.4. Дискретизация двумерных сигналов.....	309
22.5. Многомерный спектральный анализ.....	315
23. Телекоммуникации и связь.....	318
23.1. Основные сведения.....	318
23.2. Каналы связи.....	323
23.3. Информационные характеристики источника дискретных сообщений.....	333
23.4. Информационные характеристики дискретных каналов связи.....	342
23.5. Информационные характеристики источника непрерывных сообщений.....	350
23.6. Информационные непрерывных каналов связи.....	352
23.7. Согласование физических характеристик сигнала и канала .....	356
23.8. Согласование статистических свойств источника сообщений и канала связи.....	359
23.9. Вычисление количества информации при передаче сообщений по дискретному каналу связи с шумами.....	364
23.10. Взаимная информация между произвольным числом дискретных и непрерывных ансамблей.....	373
23.11. Импульсно- модулированные сигналы.....	388
23.12. Модуляция символьных и кодов данных.....	391

23.13. Введение в модулированные сигналы.....	394
23.14. Амплитудная модуляция.....	395
23.15. Сигналы с угловой модуляцией.....	408
23.16. Внутриимпульсная частотная модуляция.....	416
23.17. Импульсно-модулированные сигналы.....	418
23.18. Модуляция символьных и кодовых данных.....	421
23.19. Аналитические сигналы.....	424
23.20. Примеры применения аналитических сигналов.....	431
23.21. Преобразование Гильберта.....	435
23.22. Свойства преобразования Гильберта.....	441
23.23. Вычисление преобразования Гильберта.....	443
24. Вейвлетные преобразования сигналов.....	445
24.1. Базисные функции вейвлет-преобразования.....	445
24.2. Свойства вейвлет-преобразования.....	452
24.3. Вейвлет-преобразование простых сигналов.....	453
24.4. Введение в вейвлетный кратномасштабный анализ.....	459
24.5. Принципы кратномасштабного анализа.....	460
24.6. Математические основы кратномасштабного анализа....	466
24.7. Быстрое вейвлет-преобразование.....	474
24.8. Фильтры дуальной декомпозиции и реконструкции сигналов.....	477
24.9. Ортогональные и биортогональные вейвлеты.....	480
Литература.....	484

## **17. Пространство и метрология сигналов**

### **17.1. Основные понятия и определения**

**Понятие сигнала.** В XVIII веке в теорию математики вошло понятие функции, как определенной зависимости какой-либо величины  $y$  от другой величины – независимой переменной  $x$ , с математической записью такой зависимости в виде  $y(x)$ . Довольно скоро математика функций стала базовой основой теории всех естественных и технических наук. Особое значение функциональная математика приобрела в технике связи, где временные функции вида  $s(t)$ ,  $v(f)$  и т.п., используемые для передачи информации, стали называть **сигналами**.

В технических отраслях знаний термин "сигнал" (signal, от латинского signum – знак) очень часто используется в широком смысловом диапазоне, без соблюдения строгой терминологии. Под ним понимают и техническое средство для передачи, обращения и использования информации - электрический, магнитный, оптический сигнал; и физический процесс, представляющий собой материальное воплощение информационного сообщения - изменение какого-либо параметра носителя информации (напряжения, частоты, мощности электромагнитных колебаний, интенсивности светового потока и т.п.) во времени, в пространстве или в зависимости от изменения значений каких-либо других аргументов (независимых переменных); и смысловое содержание определенного физического состояния или процесса, как, например, сигналы светофора, звуковые предупреждающие сигналы и т.п. Все эти понятия объединяет конечное назначение сигналов. Это **определенные сведения, сообщения, информация о каких-либо процессах, состояниях или физических величинах объектов материального мира, выраженные в форме, удобной для передачи, обработки, хранения и использования этих сведений**.

Термин “сигнал” очень часто отождествляют с понятиями “данные” (data), “информация” (information) и воздействие. Действительно, эти понятия взаимосвязаны и не существуют одно без другого, но относятся к разным категориям.

Как мы уже говорили, понятие информации имеет много определений, от наиболее широкого (информация есть формализованное отражение реального мира) до практического (сведения и данные, являющиеся объектом хранения, передачи, преобразования, восприятия и управления). В настоящее время мировая наука все больше склоняется к точке зрения, что информация,

наряду с материей и энергией, принадлежит к фундаментальным философским категориям естествознания и относится к одному из свойств объективного мира, хотя и несколько специфичному. Что касается **“данных”** (от латинского *datum* – факт), то это совокупность фактов, результатов наблюдений, измерения каких-либо физических свойств объектов, явлений или процессах материального мира, представленных в формализованном виде, количественном или качественном. Это не информация, а только атрибут информации - сырье для получения информации путем соответствующей обработки и интерпретации (истолкования).

Наука и техника интернациональны, и используют, в основном, общепринятые термины, большинство из которых англоязычны. Термин *“signal”* в мировой практике является общепринятым для характеристики формы представления данных, при которой данные рассматриваются как результат некоторых измерений объекта исследований в виде последовательности значений скалярных величин (аналоговых, числовых, графических и пр.) в зависимости от изменения каких-либо переменных значений (времени, энергии, температуры, пространственных координат, и пр.). С учетом этого, в дальнейшем под термином **“сигнал”** в узком смысле этого слова будем понимать каким-либо образом упорядоченное отображение изменения физического состояния какого-либо объекта – материального носителя сигнала. На это формализованное отображение переносятся данные о характере изменения в пространстве, во времени или по любой другой переменной определенных физических величин, физических свойств или физического состояния объекта исследований. А так как данные содержат информацию, как об основных целевых параметрах объекта исследований, так и о различных сопутствующих и мешающих факторах измерений, то в широком смысле этого слова можно считать, что **сигнал является носителем общей измерительной информации.** При этом материальная форма носителей сигналов (механическая, электрическая, магнитная, акустическая, оптическая и любая другая), равно как и форма отображения данных в каких-либо физических параметрах или процессах носителей, значения не имеет. **Информативным параметром сигнала может являться любой параметр носителя сигнала, функционально и однозначно связанный со значениями информационных данных.**

Наиболее распространенное представление сигналов - в электрической форме в виде зависимости напряжения от времени  $U(t)$  (рис. 17.1). Так, например, сигнал изменения напряженности

магнитного поля по профилю аэро съемки – это и временная последовательность изменения электрического напряжения на выходе датчика аэромагнитометра, и запись этого напряжения на ленте регистратора, и последовательные значения цифровых отсчетов при обработке лент регистратора и вводе сигнала в ЭВМ.

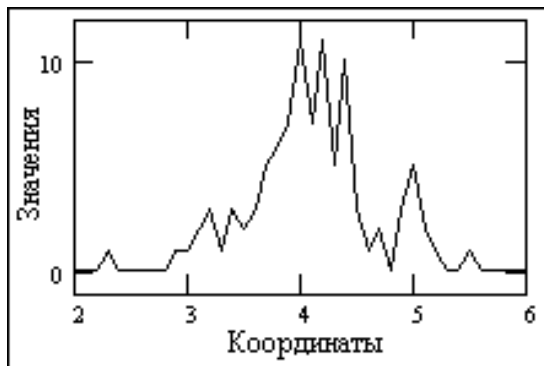


Рис. 17.1. Сигнал.

*С математической точки зрения сигнал представляет собой функцию, т.е. зависимость одной величины от другой, независимой переменной. По содержанию это информационная функция, несущая сообщение о физических свойствах, состоянии или поведении какой-либо физической системы, объекта или среды, а целью обработки сигналов можно считать извлечение определенных информационных сведений, которые отображены в этих сигналах (кратко - полезная или целевая информация) и преобразование этих сведений в форму, удобную для восприятия и дальнейшего использования.*

Под "анализом" сигналов (analysis) имеется в виду не только их чисто математические преобразования, но и получение на основе этих преобразований выводов о специфических особенностях соответствующих процессов и объектов. Целями анализа сигналов обычно являются:

- Определение или оценка числовых параметров сигналов (энергия, средняя мощность, среднее квадратическое значение и пр.).
- Разложение сигналов на элементарные составляющие для сравнения свойств различных сигналов.
- Сравнение степени близости, "похожести", "родственности" различных сигналов, в том числе с определенными количественными оценками.



Математический аппарат анализа сигналов весьма обширен, и широко применяется на практике во всех без исключения областях науки и техники.

С понятием сигнала неразрывно связан термин **регистрации сигналов**, использование которого также широко и неоднозначно, как и самого термина сигнал. В наиболее общем смысле **под этим термином можно понимать операцию выделения сигнала и его преобразования в форму, удобную для дальнейшего использования**. Так, при получении информации о физических свойствах каких-либо объектов, под регистрацией сигнала понимают процесс измерения физических свойств объекта и перенос результатов измерения на материальный носитель сигнала или непосредственное энергетическое преобразование каких-либо свойств объекта в информационные параметры материального носителя сигнала (как правило - электрического). Но так же широко термин регистрации сигналов используют и для процессов выделения уже сформированных сигналов, несущих определенную информацию, из суммы других сигналов (радиосвязь, телеметрия и пр.), и для процессов фиксации сигналов на носителях долговременной памяти, и для многих других процессов, связанных с обработкой сигналов.

Применительно к настоящей теории под термином регистрации будем понимать **регистрацию данных** (data logging) которые проходят через конкретную систему или точку системы и определенным образом фиксируются на каком-либо материальном носителе или в памяти системы.

Что касается процесса получения информации при помощи технических средств, обеспечивающих опытным путем нахождение соотношения измеряемой величины с принятой по определению образцовой единицей этой величины, и представление измеренного соотношения в какой-либо физической или числовой форме информационного сигнала, то для этого процесса будем применять, в основном, термин **детектирования**.

**Шумы и помехи** (noise). При детектировании сигналов, несущих целевую для данного вида измерений информацию, в сумме с основным сигналом одновременно регистрируются и мешающие сигналы - **шумы и помехи** самой различной природы (рис. 17.2). К помехам относят также искажения полезных сигналов при влиянии различных дестабилизирующих факторов на процессы измерений, как, например, влияние микрокаверн в стенках скважины на измерения в рентгенорадиометрических методах каротажа, грозовых разрядов на электроразведочные методы измерений и т.п. Выделение полезных составляющих из общей суммы зарегистрированных сигналов или

максимальное подавление шумов и помех в информационном сигнале при сохранении его полезных составляющих является одной из основных задач первичной обработки сигналов (результатов наблюдений).



Рис. 17.2. Сигнал с помехами.

Типы помех разделяют по источникам их возникновения, по энергетическому спектру, по характеру воздействия на сигнал, по вероятностным характеристикам и другим признакам.

Источники помех бывают внутренние и внешние.

**Внутренние шумы** могут быть присущи физической природе источников сигналов, как, например, тепловые шумы электронных потоков в электрических цепях или дробовые эффекты в электронных приборах, или возникают в измерительных устройствах и системах передачи и обработки сигналов от влияния различных дестабилизирующих факторов - температуры, повышенной влажности, нестабильности источников питания, влияния механических вибраций на гальванические соединения, и т.п.

**Внешние источники** шумов бывают искусственного и естественного происхождения. К **искусственным источникам помех** относятся промышленные помехи - двигатели, переключатели, генераторы сигналов различной формы и т.д. **Естественными источниками помех** являются молнии, флюктуации магнитных полей, всплески солнечной энергии, и т.д.

Электрические и магнитные поля различных источников помех вследствие наличия индуктивных, емкостных и резистивных связей создают на различных участках и цепях сигнальных систем паразитные разности потенциалов и токи, накладывающиеся на полезные сигналы.

Помехи подразделяются на флюктуационные, импульсные и

периодические.

**Флюктуационные или шумовые помехи** представляют хаотический и беспорядочный во времени процесс в виде нерегулярных случайных всплесков различной амплитуды. Как правило, флюктуационные помехи распределены по нормальному закону с нулевым средним и оказывают существенное влияние только на сигналы низкого уровня.

**Импульсные помехи** во многом похожи на шумовые помехи и проявляются как в виде отдельных импульсов, так и в виде последовательности импульсов, форма и параметры которых имеют случайный характер. Причинами импульсных помех являются резкие броски тока и напряжения в промышленных установках, транспортных средствах, а также природные электрические явления. Распределение импульсных помех, как правило, симметричное с произвольной плотностью распределения.

**Периодические помехи** вызываются периодическими низкочастотными или высокочастотными полями линий электропередач, силовых электроустановок и др. Если основная мощность помех сосредоточена на отдельных участках диапазона частот, например, на частоте напряжения промышленной сети или кратна этой частоте, то такие помехи называют сосредоточенными.

В зависимости от характера воздействия на сигнал помехи разделяют на аддитивные и мультипликативные.

**Аддитивные (налагающиеся) помехи** суммируются с сигналом, не зависят от его значений и формы и не изменяют информативной составляющей самого сигнала.

**Мультипликативные или деформирующие помехи** могут изменять форму информационной части сигнала, иметь зависимость от его значений и от определенных особенностей в сигнале и т.п. При известном характере мультипликативных помех возможна коррекция сигнала на их влияние.

Следует заметить, что деление сигналов на полезные и мешающие (шумовые) является достаточно условным. Источниками мешающих сигналов также могут быть определенные физические процессы, явления или объекты. При выяснении природы мешающих сигналов они могут переводиться в разряд информационных. Так, например, вариации диаметра скважин является мешающим фактором практически для всех ядерно-физических методов каротажа. Вместе с тем этот же фактор, при соответствующем методическом и аппаратном обеспечении, может дать возможность бесконтактного определения диаметра скважин в качестве дополнительного информационного параметра.

**Размерность сигналов.** Простейшими сигналами

являются одномерные сигналы, как, например, сейсмические импульсы  $s(t)$ , измерения каких-либо параметров геофизических полей (электрических, магнитных, и пр.) по профилям на поверхности земли  $s(x)$  или по стволу скважины  $s(h)$ , и т.п. Значения одномерных сигналов зависят только от одной независимой переменной, как, например, на рис. 17.1 и 17.2.

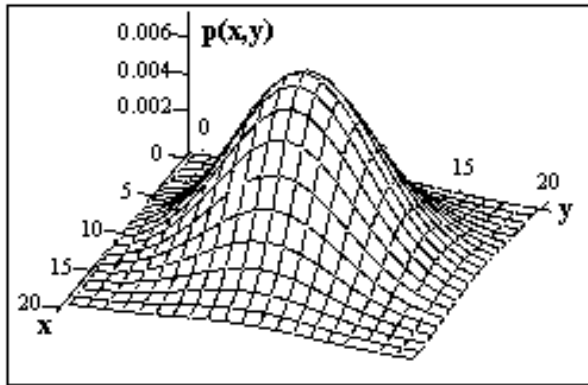


Рис. 17.3. Двумерный сигнал.

В общем случае сигналы являются многомерными функциями пространственных, временных и прочих независимых переменных - сейсмическая волна вдоль линии профиля  $s(x,t)$ , аномалия гравитационного поля на поверхности наблюдений  $s(x,y)$ , пространственно - энергетическое распределение потока ионизирующих частиц или квантов от источника излучения  $s(x,y,z,E)$  и т.п. Все большее применение находят также многомерные сигналы, образованные некоторым множеством одномерных сигналов, как, например, комплексные каротажные измерения нескольких физических параметров горных пород по стволу скважины одновременно.

Многомерные сигналы могут иметь различное представление по своим аргументам. Так, полный акустический сигнал сейсмического профиля дискретен по пространству (точкам расположения приемников) и непрерывен по времени.

**Многомерный сигнал может рассматриваться, как упорядоченная совокупность одномерных сигналов.** С учетом этого при анализе и обработке сигналов многие принципы и практические методы обработки одномерных сигналов, математический аппарат которых развит достаточно глубоко, распространяются и на

многомерные сигналы. Физическая природа сигналов для математического аппарата их обработки значения не имеет.

Вместе с тем обработка многомерных сигналов имеет свои особенности, и может существенно отличаться от одномерных сигналов в силу большего числа степеней свободы. Так, при дискретизации многомерных сигналов имеет значение не только частотный спектр сигналов, но и форма раstra дискретизации. Пример не очень полезной особенности - многомерные полиномы сигнальных функций, в отличие от одномерных, не разлагаются на простые множители. Что касается порядка размерности многомерных сигналов, то ее увеличение выше двух практически не изменяет принципы и методы анализа данных, и сказывается, в основном, только на степени громоздкости формул и чисто техническом усложнении вычислений.

Учитывая эти факторы, при рассмотрении общей теории анализа, преобразований и обработки сигналов ограничимся, в основном, одно- и двумерными сигнальными функциями, а в качестве универсальных независимых переменных (аргументов функций) будем использовать, как правило, переменную "t" для одномерных сигналов и переменные "x,t" или "x,y" для двумерных сигналов, безотносительно к их физическому содержанию (пространство, время, энергия и пр.).

***Математическое описание сигналов.*** Сигналы могут быть объектами теоретических исследований и практического анализа только в том случае, если указан способ их ***математического описания.***

Математическое описание позволяет абстрагироваться от физической природы сигнала и материальной формы его носителя, проводить классификацию сигналов, выполнять их сравнение, устанавливать степень тождества, моделировать системы обработки сигналов.

Большинство сигналов, встречающихся на практике, представлены во временной области функциями времени. При отображении сигналов на графике одной из координат (независимой) является ось времени, а другой координатой (зависимой) – ось амплитуд. Тем самым мы получаем ***амплитудно-временное представление сигнала.*** В общем случае описание сигнала задается функциональной зависимостью определенного информационного параметра сигнала от независимой переменной (аргумента) –  $s(x)$ ,  $y(t)$  и т.п. Такая форма описания и графического представления сигналов называется ***динамической*** (сигнал в реальной динамике его поведения по аргументам). Функции математического описания сигналов могут быть как вещественными, так и комплексными. Выбор математического аппарата описания определяется простотой и

удобством его использования при анализе и обработке сигналов.

Отметим двойственность применения описания сигналов функциями типа  $s(t)$  и т.п. С одной стороны  $s(t)$  – это величина, равная значению функции в момент времени  $t$ . С другой стороны мы обозначаем через  $s(t)$  и саму функцию, т.е. то правило, по которому каждому значению  $t$  ставится в соответствие определенная величина  $s$ . В большинстве аналитических выражений это не вызывает недоразумений и при однозначном соответствии значений сигналов их аналитическим выражениям принимается по умолчанию.

Сделаем также одно замечание по терминологии описания сигналов. В теоретических работах по анализу сигналов конкретные значения величины сигнала (отсчеты значений по аргументу) часто именуют **координатами сигнала**. В отраслях знаний, связанных с геологией и горным делом, и в геофизической практике в том числе, этот термин используется по своему прямому смысловому назначению – **пространственных координат результатов измерений**, и является неизменным атрибутом всех геолого-геофизических данных. С учетом последнего фактора условимся применять термин “координата” по своему традиционному смысловому назначению в качестве обобщающего термина для независимых переменных сигнальных функций. При этом под понятием **координат значений сигнала** будем понимать не только какие-либо пространственные координаты, как это непосредственно имеет место для результатов измерений при геолого-геофизических съемках, но и любые другие аргументы, на числовой оси которых отложены значения или отсчеты сигнала и рассматривается динамика его изменения (пример на рис. 17.1).

**Спектральное представление сигналов.** Кроме привычного динамического представления сигналов и функций в виде зависимости их значений от определенных аргументов (времени, линейной или пространственной координаты и т.п.) при анализе и обработке данных широко используется **математическое описание сигналов по аргументам, обратным аргументам динамического представления**. Так, например, для времени обратным аргументом является **частота**. Возможность такого описания определяется тем, что любой сколь угодно сложный по своей форме сигнал, не имеющий разрывов второго рода (бесконечных значений на интервале своего задания), можно представить в виде суммы более простых сигналов, и, в частности, в виде суммы простейших гармонических колебаний, что выполняется при помощи преобразования Фурье. Соответственно, математически разложение сигнала на гармонические составляющие описывается функциями значений амплитуд и начальных фаз колебаний по непрерывному или дискретному аргументу – частоте

изменения функций на определенных интервалах аргументов их динамического представления. Совокупность амплитуд гармонических колебаний разложения называют **амплитудным спектром** сигнала, а совокупность начальных фаз – **фазовым спектром**. Оба спектра вместе образуют полный частотный спектр сигнала, который по точности математического представления тождественен динамической форме описания сигнала.

Линейные системы преобразования сигналов описываются дифференциальными уравнениями, причем для них верен **принцип суперпозиции**, **согласно которому реакция систем на сложный сигнал, состоящий из суммы простых сигналов, равна сумме реакций от каждого составляющего сигнала в отдельности**. Это позволяет при известной реакции системы на гармоническое колебание с определенной частотой определить реакцию системы на любой сложный сигнал, разложив его в ряд гармоник по частотному спектру сигнала. Широкое использование гармонических функций при анализе сигналов объясняется тем, что они являются достаточно простыми ортогональными функциями и определены при всех значениях непрерывных переменных. Кроме того, они являются собственными функциями времени, сохраняющими свою форму при прохождении колебаний через любые линейные системы и системы обработки данных с постоянными параметрами (изменяются только амплитуда и фаза колебаний). Немаловажное значение имеет и то обстоятельство, что для гармонических функций и их комплексного анализа разработан мощный математический аппарат.

Примеры частотного представления сигналов приводятся ниже (рис. 17.5 – 17.12).

Кроме гармонического ряда Фурье применяются и другие виды разложения сигналов: по функциям Уолша, Бесселя, Хаара, полиномам Чебышева, Лаггера, Лежандра и др. Главное условие однозначности и математической идентичности отображения сигналов - **ортогональность функций разложения**. Но при качественном анализе сигналов могут применяться и неортогональные функции, выявляющие какие-либо характерные особенности сигналов, полезные для интерпретации физических данных.

**Математические модели сигналов.** Теория анализа и обработки физических данных базируется на математических моделях соответствующих **физических полей и физических процессов**, на основе которых создаются **математические модели сигналов**. Математические модели сигналов дают возможность обобщенно, абстрагируясь от физической природы, судить о свойствах сигналов, предсказывать изменения сигналов в изменяющихся условиях,

заменять физическое моделирование процессов математическим. С помощью математических моделей имеется возможность описывать свойства сигналов, которые являются главными, определяющими в изучаемых процессах, и игнорировать большое число второстепенных признаков. Знание математических моделей сигналов дает возможность классифицировать их по различным признакам, характерным для того или иного типа моделей. Так, сигналы разделяются на неслучайные и случайные в зависимости от возможности точного предсказания их значений в любые моменты времени. Сигнал является неслучайным и называется **детерминированным**, если математическая модель позволяет осуществлять такое предсказание. Детерминированный сигнал задается, как правило, математической функцией или вычислительным алгоритмом, а математическая модель сигнала может быть представлена в виде

$$s = F(t, z, \omega, \dots; A, B, C, \dots),$$

где  $s$  – информативный параметр сигнала;  $t, z, \omega, \dots$  – независимые аргументы (время, пространственная координата, частота и др.);  $A, B, C, \dots$  – параметры сигналов.

Модель должна быть, по возможности, проще, минимизирована по количеству независимых аргументов и адекватна изучаемому процессу, что во многом предопределяет результаты измерений. Рассмотрим этот вопрос на примере геофизических данных.

Под геофизическим полем понимают собственное или индуцированное определенным внешним воздействием распределение какой-либо физической величины, создаваемое геологическим объектом или геологической структурой в пространстве, во времени или по любому другому аргументу (независимой переменной). В простейшем случае геофизический сигнал - это изменение какой-либо составляющей геофизического поля, т.е. сечение поля по одному из аргументов. В пределе все геофизическое поле в целом может рассматриваться как первичный многомерный сигнал в прямом физическом отображении, с которого путем измерений могут сниматься формализованные копии определенных составляющих (сечений) сигнала на материальные носители информации.

Геофизическим полям в определенных условиях их регистрации соответствуют определенные математические модели сигналов, т.е. их описание на каком-либо формальном языке. Математическое описание не может быть всеобъемлющим и идеально точным и, по существу, всегда отображает не реальные объекты, а их упрощенные (гомоморфные) модели. Модели могут задаваться



таблицами, графиками, функциональными зависимостями, уравнениями состояний и переходов из одного состояния в другое и т.п. Формализованное описание может считаться математической моделью оригинала, если оно позволяет с определенной точностью прогнозировать состояние и поведение изучаемых объектов путем формальных процедур над их описанием.

Неотъемлемой частью любой математической модели сигнала является область определения сигнала, которая устанавливается интервалом задания независимой переменной. Примеры задания интервала для переменных:

$$a \leq x \leq b, \quad x \in [a, b].$$

$$a < y \leq b, \quad y \in (a, b].$$

$$a < z < b, \quad z \in (a, b).$$

Пространство значений независимой переменной обычно обозначается через индекс  $R$ . Так, например,  $R := (-\infty, +\infty)$ ,  $x \in R$ .

Кроме задания области определения сигнала могут быть также заданы виды численных значений переменных (целые, рациональные, вещественные, комплексные).

Математические модели полей и сигналов на первом этапе обработки и анализа результатов наблюдений должны позволять в какой-то мере игнорировать их физическую природу и возвращать ее в модель только на заключительном этапе интерпретации данных.

**Виды моделей сигналов.** При анализе физических данных используются два основных подхода к созданию математических моделей сигналов.

Первый подход оперирует с **детерминированными сигналами**, значения которых в любой момент времени или в произвольной точке пространства (а равно и в зависимости от любых других аргументов) являются априорно известными или могут быть достаточно точно определены (вычислены). Такой подход удобен в прямых задачах геофизики (расчеты полей для заданных моделей сред), в задачах активных воздействий на среду при заранее известных параметрах и форме сигнала воздействия (вибрационная сейсморазведка, электромагнитные методы каротажа и пр.), а также при использовании хорошо известных и достоверных геолого-геофизических данных. Для описания неслучайных сигналов используются также **квазидетерминированные модели**, в которых значения одного или нескольких параметров априорно неизвестны, и считаются случайными величинами с малой случайной компонентой, влиянием которой можно пренебречь.

Второй подход предполагает **случайный характер сигналов**, закон изменения которых во времени (или в пространстве) носит

случайный характер, и которые принимают конкретные значения с некоторой вероятностью. Модель такого сигнала представляет собой описание статистических характеристик случайного процесса путем задания законов распределения вероятностей, корреляционной функции, спектральной плотности энергии и др.

Случайность может быть обусловлена как собственной физической природой сигналов, что характерно, например, для методов ядерной геофизики, так и вероятностным характером регистрируемых сигналов как по времени или месту их появления, так и по содержанию. С этих позиций случайный сигнал может рассматриваться как отображение случайного по своей природе процесса или физических свойств объекта (процесса), которые определяются случайными параметрами или сложным строением среды, результаты измерений в которой трудно предсказуемы.

Между этими двумя видами сигналов нет резкой границы. Строго говоря, ***детерминированных процессов и отвечающих им детерминированных сигналов в природе не существует***. Даже сигналы, хорошо известные на входе в среду (при внешнем воздействии на нее), по месту их регистрации всегда осложнены случайными помехами, влиянием дестабилизирующих факторов и априорно неизвестными параметрами и строением самой среды. С другой стороны, модель случайного поля часто аппроксимируется методом суперпозиции (сложения) сигналов известной формы. Детерминированные модели могут использоваться и для изучения чисто случайных процессов, если уровень полезного сигнала в этом процессе значительно выше уровня статистических флюктуаций, что имеет место, например, при регистрации ионизирующих излучений от горных пород.

На выбор математической модели поля в том или ином методе предметной области в немалой степени влияет также сложность математического аппарата обработки сигналов и сложившиеся традиции интерпретации результатов наблюдений. Не исключается и изменение модели, как правило, с переводом из вероятностной в детерминированную, в процессе накопления информации об изучаемом явлении или объекте.

***Классификация сигналов*** осуществляется на основании существенных признаков соответствующих математических моделей сигналов. Все сигналы разделяют на две крупных группы: детерминированные и случайные. Классификация сигналов внутри групп приведена на рис. 17.4.



Рис. 17.4. Классификация сигналов.

С математических позиций группы сигналов обычно называют множествами, в которые объединяют сигналы по какому-либо общему свойству. Принадлежность сигнала  $s$  к множеству  $L_P$  записывается в виде  $L_P = \{s; P\}$ , где  $P$  – определенное свойство данного множества сигналов.

**Классификация детерминированных сигналов.**

Обычно выделяют два класса детерминированных сигналов: периодические и непериодические.

К **множеству периодических** относят гармонические и полигармонические сигналы. Для периодических сигналов выполняется общее условие  $s(t) = s(t + kT)$ , где  $k = 1, 2, 3, \dots$  - любое целое число (из множества целых чисел  $I$  от  $-\infty$  до  $\infty$ ),  $T$  - период, являющийся конечным отрезком независимой переменной. Множество периодических сигналов:

$$L_P = \{s(t); s(t+kT) = s(t), -\infty < t < \infty, k \in I\}.$$

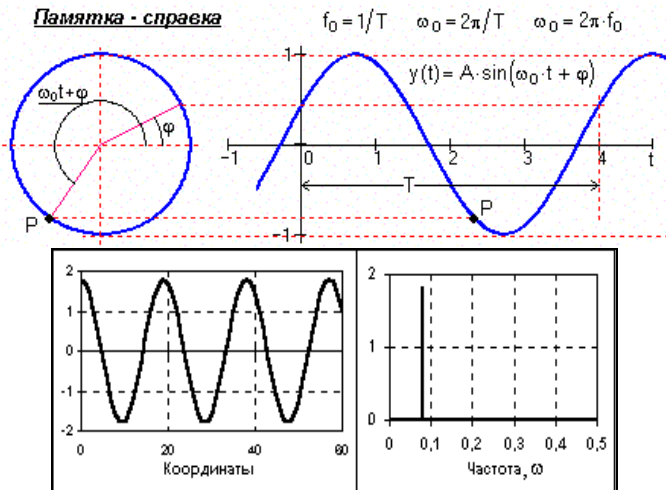


Рис. 17.5. Гармонический сигнал и спектр его амплитуд.

**Гармонические сигналы** (синусоидальные), описываются следующими формулами:

$$\begin{aligned} s(t) &= A \cdot \sin(2\pi f_0 t + \phi) = A \cdot \sin(\omega_0 t + \phi), \\ s(t) &= A \cdot \cos(\omega_0 t + \phi), \end{aligned} \quad (17.1)$$

где  $A$ ,  $f_0$ ,  $\omega_0$ ,  $\phi$  - постоянные величины, которые могут исполнять роль информационных параметров сигнала:  $A$  - амплитуда сигнала,  $f_0$  - циклическая частота в герцах,  $\omega_0 = 2\pi f_0$  - угловая частота в радианах,  $\phi$  и  $\phi$  - начальные фазовые углы в радианах. Период одного колебания  $T = 1/f_0 = 2\pi/\omega_0$ . При  $\phi = \phi - \pi/2$  синусные и косинусные функции описывают один и тот же сигнал. Частотный спектр сигнала представлен амплитудным и начальным фазовым значением частоты  $f_0$  (при  $t = 0$ ).

**Полигармонические сигналы** составляют наиболее широко распространенную группу периодических сигналов и описываются суммой гармонических колебаний:

$$s(t) = \sum_{n=0}^N A_n \sin(2\pi f_n t + \phi_n), \quad (17.2)$$

или непосредственно функцией  $s(t) = y(t \pm kT_p)$ ,  $k = 1, 2, 3, \dots$ , где  $T_p$  - период одного полного колебания сигнала  $y(t)$ , заданного на одном периоде. Значение  $f_p = 1/T_p$  называют фундаментальной частотой колебаний. Полигармонические сигналы представляют собой сумму определенной постоянной составляющей ( $f_0=0$ ) и произвольного (в пределе - бесконечного) числа гармонических составляющих с произвольными значениями амплитуд  $A_n$  и фаз  $\phi_n$ , с периодами, кратными периоду фундаментальной частоты  $f_p$ . Другими словами, на периоде фундаментальной частоты  $f_p$ , которая равна или кратно меньше минимальной частоты гармоник, укладывается кратное число периодов всех гармоник, что и создает периодичность повторения сигнала. Частотный спектр полигармонических сигналов дискретен, в связи с чем второе распространенное математическое представление сигналов - в виде спектров (рядов Фурье).

В качестве примера на рис. 17.6 приведен отрезок периодической сигнальной функции, которая получена суммированием постоянной составляющей (частота постоянной составляющей равна 0) и трех гармонических колебаний с разными значениями частоты и начальной фазы колебаний. Математическое описание сигнала задается формулой:

$$s(t) = \sum_{k=0}^3 A_k \cdot \cos(2 \cdot \pi \cdot f_k \cdot t + \phi_k),$$

где:  $A_k = \{5, 3, 4, 7\}$  - амплитуда гармоник;  $f_k = \{0, 40, 80, 120\}$  - частота в герцах;  $\phi_k = \{0, -0.4, -0.6, -0.8\}$  - начальный фазовый угол

колебаний в радианах;  $k = 0, 1, 2, 3$ . Фундаментальная частота сигнала 40 Гц.



Рис. 17.6. Модель сигнала.

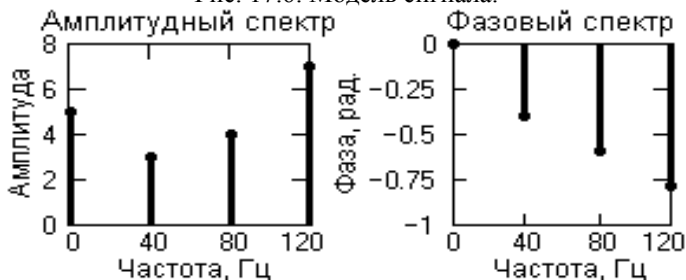


Рис. 17.7. Спектр сигнала.

Частотное представление данного сигнала (спектр сигнала) приведено на рис. 17.7. Обратим внимание, что частотное представление периодического сигнала  $s(t)$ , ограниченного по числу гармоник спектра, составляет всего восемь отсчетов и весьма компактно по сравнению с временным представлением.

Периодический сигнал любой произвольной формы может быть представлен в виде суммы гармонических колебаний с частотами, кратными фундаментальной частоте колебаний  $f_p = 1/T_p$ . Для этого достаточно разложить один период сигнала в ряд Фурье по тригонометрическим функциям синуса и косинуса с шагом по частоте, равным фундаментальной частоте колебаний  $\Delta f = f_p$ :

$$s(t) = \sum_{k=0}^K (a_k \cos 2\pi k \Delta f t + b_k \sin 2\pi k \Delta f t), \quad (17.3)$$

$$a_0 = (1/T) \int_0^T s(t) dt, \quad a_k = (2/T) \int_0^T s(t) \cos 2\pi k \Delta f t dt, \quad (17.4)$$

$$b_k = (2/T) \int_0^T s(t) \sin 2\pi k \Delta f t dt. \quad (17.5)$$

Количество членов ряда Фурье  $K = k_{\max}$  обычно ограничивается максимальными частотами  $f_{\max}$  гармонических составляющих в сигналах так, чтобы  $f_{\max} < K \cdot f_p$ . Однако для сигналов с разрывами и скачками имеет место  $f_{\max} \rightarrow \infty$ , при этом количество членов ряда ограничивается по допустимой погрешности аппроксимации функции  $s(t)$ . Одночастотные косинусные и синусные гармоники можно объединить и представить разложение в более компактной форме:

$$s(t) = \sum_{k=0}^K S_k \cos(2\pi k \Delta f t - \varphi_k), \quad (17.5')$$

$$S_k = \sqrt{a_k^2 + b_k^2}, \quad \varphi_k = \operatorname{arctg}(b_k/a_k). \quad (17.6)$$

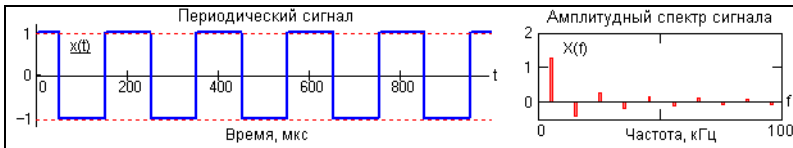


Рис. 17.8. Прямоугольный периодический сигнал (меандр).

Пример представления прямоугольного периодического сигнала (меандра) в виде амплитудного ряда Фурье в частотной области приведен на рис. 17.8. Сигнал четный относительно  $t=0$ , не имеет синусных гармоник, все значения  $\varphi_k$  для данной модели сигнала равны нулю. Информационными параметрами полигармонического сигнала могут быть как определенные особенности формы сигнала (размах от минимума до максимума, экстремальное отклонение от среднего значения, и т.п.), так и параметры определенных гармоник в этом сигнале. Так, например, для прямоугольных импульсов информационными параметрами могут быть период повторения импульсов, длительность импульсов, скважность импульсов (отношение периода к длительности). При анализе сложных периодических сигналов информационными параметрами могут также быть:

- Текущее среднее значение за определенное время, например, за время периода:

$$(1/T) \int_t^{t+T} s(t) dt.$$

- Постоянная составляющая одного периода:

$$(1/T) \int_0^T s(t) dt.$$

- Среднее выпрямленное значение:

$$(1/T) \int_0^T |s(t)| dt.$$

- Среднее квадратичное значение:

$$\sqrt{\frac{1}{T} \int_0^T x(t)^2 dt}.$$

**К непериодическим** сигналам относят почти периодические и аperiodические сигналы. Основным инструментом их анализа также является частотное представление.

**Почти периодические сигналы** близки по своей форме к полигармоническим. Они также представляют собой сумму двух и более гармонических сигналов (в пределе – до бесконечности), но не с кратными, а с произвольными частотами, отношения которых (хотя бы двух частот минимум) не относятся к рациональным числам, вследствие чего фундаментальный период суммарных колебаний бесконечно велик.

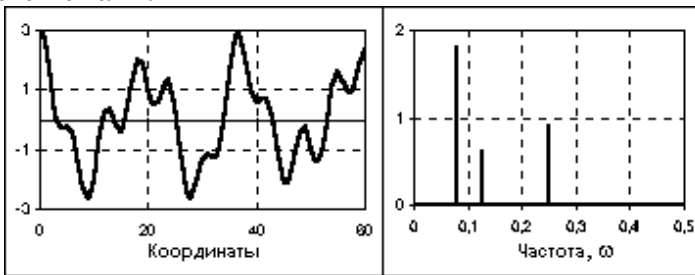


Рис. 17.9. Почти периодический сигнал и спектр его амплитуд.

Так, например, сумма двух гармоник с частотами  $2f_0$  и  $3.5f_0$  дает периодический сигнал ( $2/3.5$  – рациональное число) с фундаментальной частотой  $0.5f_0$ , на одном периоде которой будут укладываться 4 периода первой гармоники и 7 периодов второй. Но если значение частоты второй гармоники заменить близким значением  $\sqrt{12} f_0$ , то сигнал перейдет в разряд непериодических, поскольку отношение  $2/\sqrt{12}$  не относится к числу рациональных чисел. Как правило, почти периодические сигналы порождаются физическими процессами, не связанными между собой. Математическое отображение сигналов тождественно полигармоническим сигналам (сумма гармоник), а частотный спектр также дискретен.

**Аperiodические сигналы** составляют основную группу непериодических сигналов и задаются произвольными функциями времени. На рис. 17.10 показан пример аperiodического сигнала,

заданного формулой на интервале  $(0, \infty)$ :

$$s(t) = \exp(-a \cdot t) - \exp(-b \cdot t),$$

где  $a$  и  $b$  – константы, в данном случае  $a = 0.15$ ,  $b = 0.17$ .

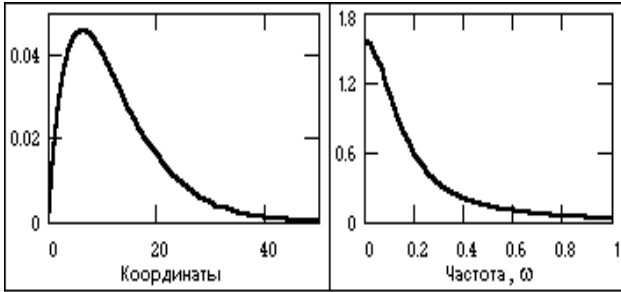


Рис. 17.10. Аperiodический сигнал и модуль спектра.

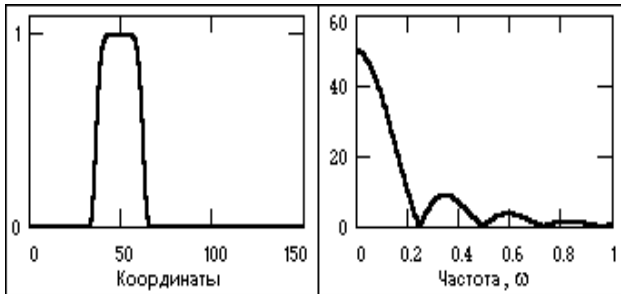


Рис. 17.11. Импульсный сигнал и модуль спектра.

К аperiodическим сигналам относятся также импульсные сигналы, которые в радиотехнике и в отраслях, широко ее использующих, часто рассматривают в виде отдельного класса сигналов. Импульсы представляют собой сигналы, как правило, определенной и достаточно простой формы, существующие в пределах конечных временных интервалов. Сигнал, приведенный на рис. 17.11, относится к числу импульсных.

Частотный спектр аperiodических сигналов непрерывен и может содержать любые гармоники в частотном интервале  $[0, \infty]$ . Для его вычисления используется интегральное преобразование Фурье, которое можно получить переходом в формулах (17.3) от суммирования к интегрированию при  $\Delta f \rightarrow 0$  и  $k\Delta f \rightarrow f$ .

$$s(t) = \int_0^{\infty} (a(f) \cos 2\pi ft + b(f) \sin 2\pi ft) df = \int_0^{\infty} S(f) \cos(2\pi ft - \varphi(f)) df. \quad (17.7)$$



$$a(f) = \int_0^T s(t) \cos 2\pi f t \, dt, \quad b(f) = \int_0^T s(t) \sin 2\pi f t \, dt, \quad (17.8)$$

$$S(f) = \sqrt{a(f)^2 + b(f)^2}, \quad \varphi(f) = \operatorname{arctg} (b(f)/a(f)). \quad (17.9)$$

Частотные функции  $a(f)$ ,  $b(f)$  и  $S(f)$  представляют собой не амплитудные значения соответствующих гармоник на определенных частотах, а распределения спектральной плотности амплитуд этих гармоник по частотной шкале. Формулы (17.8-17.9) обычно называют формулами прямого преобразования Фурье, формулы (17.7) – обратного преобразования.

Если нас не интересует поведение сигнала за пределами области его задания  $[0, T]$ , то эта область может восприниматься, как один период периодического сигнала, т.е. значение  $T$  принимается за фундаментальную частоту периодический колебаний, при этом для частотной модели сигнала может применяться разложение в ряды Фурье по области его задания (17.3-17.6).

В классе импульсных сигналов выделяют подкласс радиоимпульсов. Пример радиоимпульса приведен на рис. 17.12.

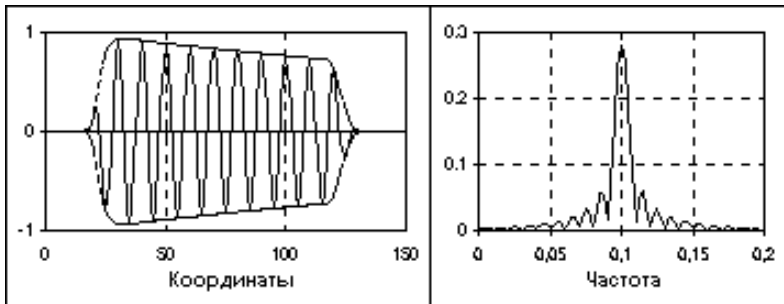


Рис. 17.12. Радиоимпульс и модуль его спектра.

Уравнение радиоимпульса имеет вид:

$$s(t) = u(t) \cos(2\pi f_0 t + \varphi_0).$$

где  $\cos(2\pi f_0 t + \varphi_0)$  – гармоническое колебание заполнения радиоимпульса,  $u(t)$  – огибающая радиоимпульса. Положение главного пика спектра радиоимпульса на частотной шкале соответствует частоте заполнения  $f_0$ , а его ширина определяется длительностью радиоимпульса. Чем больше длительность радиоимпульса, тем меньше ширина главного частотного пика.

**С энергетических позиций** сигналы разделяют на два типа: с ограниченной (конечной) энергией и с бесконечной энергией.

Для множества сигналов с ограниченной энергией должно выполняться условие:

$$L_{2E} = \{s; \int_{-\infty}^{\infty} |s(t)|^2 dt < \infty\}.$$

О сигналах  $s(t)$  данного множества принято говорить, что они *интегрируемы с квадратом*. Очевидно, что этому множеству могут соответствовать только сигналы, стремящиеся к нулю на бесконечности:  $\lim_{|t| \rightarrow \infty} s(t) \rightarrow 0$ .

Как правило, к этому типу сигналов относятся аperiodические и импульсные сигналы, не имеющие разрывов 2-го рода при ограниченном количестве разрывов 1-го рода. Любые периодические, полигармонические и почти периодические сигналы, а также сигналы с разрывами и особыми точками 2-го рода, уходящими в бесконечность, относятся к сигналам с бесконечной энергией. Для их анализа применяются специальные методы.

Для бесконечных по энергии сигналов, в том числе для периодических, ограничение по энергии может задаваться для определенного интервала (периода)  $T = t_1 - t_2$ :

$$L_{2E}(T) = \{s; \int_{t_1}^{t_2} |s(t)|^2 dt < \infty\}.$$

Иногда в отдельный класс выделяют сигналы конечной длительности, отличные от нуля только на ограниченном интервале аргументов (независимых переменных). Такие сигналы обычно называют *финитными*.

*С позиций временной динамики* сигналы подразделяются на стационарные и нестационарные. Стационарными называются сигналы, частотный спектр которых не изменяется во времени и не зависит от интервала задания сигналов. К ним относятся периодические и почти периодические сигналы. Большинство практических сигналов являются нестационарными на достаточно больших интервалах задания, но могут содержать в своем составе стационарные частотные составляющие. Так, модулированные сигналы радио и телевидения относятся к числу нестационарных, но имеют стационарные несущие частоты.

**Классификация случайных сигналов.** *Случайным сигналом называют функцию времени, значения которой заранее неизвестны, и могут быть предсказаны лишь с некоторой вероятностью.* Случайный сигнал отображает случайное физическое явление или физический процесс, причем, зарегистрированный в единичном наблюдении, сигнал не воспроизводится при повторных наблюдениях и не может быть описан явной математической зависимостью. При регистрации случайного сигнала реализуется

только один из возможных вариантов (исходов) случайного процесса, а достаточно полное и точное описание процесса в целом можно произвести только после многократного повторения наблюдений и вычисления определенных статистических характеристик ансамбля реализаций сигнала. В качестве основных статистических характеристик случайных сигналов принимают:

а) закон распределения вероятности нахождения величины сигнала в определенном интервале значений;

б) спектральное распределение мощности сигнала.

Случайные сигналы подразделяют на стационарные и нестационарные. Случайные стационарные сигналы сохраняют свои статистические характеристики в последовательных реализациях случайного процесса. Что касается случайных нестационарных сигналов, то их общепринятой классификации не существует. Как правило, из них выделяют различные группы сигналов по особенностям их нестационарности.

### 17.2. Типы сигналов

Выделяют следующие типы сигналов, которым соответствуют определенные формы их математического описания.

**Аналоговый сигнал** (analog signal) является непрерывной функцией непрерывного аргумента, т. е. определен для любого значения аргументов. Источниками аналоговых сигналов, как правило, являются физические процессы и явления, непрерывные в динамике своего развития во времени, в пространстве или по любой другой независимой переменной, при этом регистрируемый сигнал подобен (“аналогичен”) порождающему его процессу. Пример математической записи сигнала:  $y(t) = 4.8 \exp[-(t-4)^2/2.8]$ . Пример графического отображения данного сигнала приведен на рис. 17.13, при этом как сама функция, так и ее аргументы, могут принимать любые значения в пределах некоторых интервалов  $y_1 \leq y \leq y_2, t_1 \leq t \leq t_2$ .



Рис. 17.13. Аналоговый сигнал.

Если интервалы значений сигнала или его независимых переменных не ограничиваются, то по умолчанию они принимаются равными от  $-\infty$  до  $+\infty$ . Множество возможных значений сигнала образует **континуум - непрерывное пространство, в котором любая сигнальная точка может быть определена с точностью до бесконечности.**

Примеры сигналов, аналоговых по своей природе - изменение напряженности электрического, магнитного, электромагнитного поля во времени и в пространстве.

**Дискретный сигнал** (discrete signal) по своим значениям также является непрерывной функцией, но определенной только по дискретным значениям аргумента. По множеству своих значений он является конечным (счетным) и описывается дискретной последовательностью отсчетов (samples)  $y(n\Delta t)$ , где  $y_1 \leq y \leq y_2$ ,  $\Delta t$  - интервал между отсчетами (**интервал** или **шаг дискретизации**, sample time),  $n = 0, 1, 2, \dots, N$ . Величина, обратная шагу дискретизации:  $f = 1/\Delta t$ , называется **частотой дискретизации** (sampling frequency). Если дискретный сигнал получен дискретизацией (sampling) аналогового сигнала, то он представляет собой последовательность отсчетов, значения которых в точности равны значениям исходного сигнала по координатам  $n\Delta t$ .

Пример дискретизации аналогового сигнала, приведенного на рис. 17.13, представлен на рис. 17.14.



Рис. 17.4. Дискретный сигнал

При  $\Delta t = \text{const}$  (равномерная дискретизация данных) дискретный сигнал можно описывать сокращенным обозначением  $y(n)$ . В технической литературе в обозначениях дискретизированных

функций иногда оставляют прежние индексы аргументов аналоговых функций, заключая последние в квадратные скобки -  $y[t]$ . При неравномерной дискретизации сигнала обозначения дискретных последовательностей (в текстовых описаниях) обычно заключаются в фигурные скобки -  $\{s(t_i)\}$ , а значения отсчетов приводятся в виде таблиц с указанием значений координат  $t_i$ . Для числовых последовательностей (равномерных и неравномерных) применяется и следующее числовое описание:  $s(t_i) = \{a_1, a_2, \dots, a_N\}$ ,  $t = t_1, t_2, \dots, t_N$ . Примеры дискретных геофизических сигналов - результаты вертикального электрического зондирования (дискретная величина разности токовых электродов), профили геохимического опробования, и т.п.

**Цифровой сигнал** (digital signal) квантован по своим значениям и дискретен по аргументу. Он описывается квантованной решетчатой функцией

$$y_n = Q_k[y(n\Delta t)],$$

где  $Q_k$  - функция квантования с числом уровней квантования  $k$ , при этом интервалы квантования могут быть как с равномерным распределением, так и с неравномерным, например - логарифмическим. Задается цифровой сигнал, как правило, в виде дискретного ряда (discrete series) числовых данных - числового массива по последовательным значениям аргумента при  $\Delta t = \text{const}$ , но в общем случае сигнал может задаваться и в виде таблицы для произвольных значений аргумента.



Рис. 17.15. Цифровой сигнал

По существу, цифровой сигнал по своим значениям (отсчетам) является формализованной разновидностью дискретного сигнала при

округлении отсчетов последнего до определенного количества цифр, как это показано на рис 17.15. Цифровой сигнал конечен по множеству своих значений. Процесс преобразования бесконечных по значениям аналоговых отсчетов в конечное число цифровых значений называется **квантованием по уровню**, а возникающие при квантовании ошибки округления отсчетов (отбрасываемые значения) – **шумами (noise) или ошибками (error) квантования (quantization)**.

В системах цифровой обработки данных и в ЭВМ сигнал всегда представлен с точностью до определенного количества разрядов, а, следовательно, всегда является цифровым. С учетом этих факторов при описании цифровых сигналов функция квантования обычно опускается (подразумевается равномерной по умолчанию), а для описания сигналов используются правила описания дискретных сигналов. Что касается формы обращения цифровых сигналов в системах хранения, передачи и обработки, то, как правило, они представляет собой комбинации коротких одно- или двуполярных импульсов одинаковой амплитуды, которыми в двоичном коде с определенным количеством числовых разрядов кодируются числовые последовательности сигналов (массивов данных).



Рис. 17.16. Дискретно-аналоговый сигнал

В принципе, квантованными по своим значениям могут быть и аналоговые сигналы, зарегистрированные соответствующей аппаратурой (рис. 17.16), которые принято называть дискретно-аналоговыми. Но выделять эти сигналы в отдельный тип не имеет смысла - они остаются аналоговыми кусочно-непрерывными сигналами с шагом квантования, который определяется допустимой погрешностью измерений.

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. Но существуют и сигналы, которые изначально относятся к классу цифровых, как, например отсчеты количества гамма-квантов, зарегистрированных по последовательным интервалам времени.

**Преобразования типа сигналов.** Формы математического отображения сигналов, особенно на этапах их первичной регистрации (детектирования) и в прямых задачах описания физических полей и физических процессов, как правило, отражают их физическую природу. Однако последнее не является обязательным и зависит от методики измерений и технических средств детектирования, преобразования, передачи, хранения и обработки сигналов. На разных этапах процессов получения и обработки информации как материальное представление сигналов в устройствах регистрации и обработки, так и формы их математического описания при анализе данных, могут изменяться путем соответствующих операций преобразования типа сигналов.

**Операция дискретизации** (discretization) осуществляет преобразование аналоговых сигналов (функций), непрерывных по аргументу, в функции мгновенных значений сигналов по дискретному аргументу. Дискретизация обычно производится с постоянным шагом по аргументу (*равномерная дискретизация*), при этом  $s(t) \Rightarrow s(n\Delta t)$ , где значения  $s(n\Delta t)$  представляют собой отсчеты функции  $s(t)$  в моменты времени  $t = n\Delta t$ ,  $n = 0, 1, 2, \dots, N$ . Частота, с которой выполняются замеры аналогового сигнала, называется **частотой дискретизации**. В общем случае, сетка отсчетов по аргументу может быть произвольной, как, например,  $s(t) \Rightarrow s(t_k)$ ,  $k=1, 2, \dots, K$ , или задаваться по определенному закону. В результате дискретизации непрерывный (*аналоговый*) сигнал переводится в последовательность чисел.

**Операция восстановления аналогового сигнала** из его дискретного представления обратна операции дискретизации и представляет, по существу, интерполяцию данных.

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

Как известно, любая непрерывная функция может быть разложена на конечном отрезке в ряд Фурье, т.е. представлена в

спектральной форме - в виде суммы ряда синусоид с кратными (нумерованными) частотами с определенными амплитудами и фазами. У относительно гладких функций спектр быстро убывает (коэффициенты модуля спектра быстро стремятся к нулю). Для представления "изрезанных" функций, с разрывами и "изломами", нужны синусоиды с большими частотами. Говорят, что сигнал имеет *ограниченный спектр*, если после определенной частоты  $F$  все коэффициенты спектра равны нулю, т.е. сигнал представляется в виде конечной суммы ряда Фурье.

Теоремой Котельникова-Шеннона устанавливается, что если спектр сигнала ограничен частотой  $F$ , то после дискретизации сигнала с частотой не менее  $2F$  можно восстановить исходный непрерывный сигнал по полученному цифровому сигналу абсолютно точно. Для этого нужно выполнить интерполяцию цифрового сигнала "между отсчетами" специальной функцией (Котельникова-Шеннона).

На практике эта теорема имеет огромное значение. Например, известно, что диапазон звуковых сигналов, воспринимаемых человеком, не превышает 20 кГц. Следовательно, при дискретизации записанных звуковых сигналов с частотой не менее 40 кГц мы можем точно восстановить исходный аналоговый сигнал по его цифровым отсчетам, что и выполняется в проигрывателях компакт-дисков для восстановления звука. Частота дискретизации звукового сигнала при записи на компакт-диск составляет 44100 Гц.

**Операция квантования** или аналого-цифрового преобразования (АЦП; английский термин Analog-to-Digital Converter, ADC) заключается в преобразовании дискретного сигнала  $s(t_n)$  в цифровой сигнал  $s(n)=s_n \approx s(t_n)$ ,  $n=0, 1, 2, \dots, N$ , как правило, кодированный в двоичной системе счисления. Процесс преобразования отсчетов сигнала в числа называется квантованием по уровню (quantization), а возникающие при этом потери информации за счет округления – ошибками или шумами квантования (quantization error, quantization noise).

При преобразовании аналогового сигнала непосредственно в цифровой сигнал операции дискретизации и квантования совмещаются.

**Операция цифро-аналогового преобразования** (ЦАП; Digital-to-Analog Converter, DAC) обратна операции квантования, при этом на выходе регистрируется либо дискретно-аналоговый сигнал  $s(t_n)$ , который имеет ступенчатую форму (рис. 17.16), либо непосредственно аналоговый сигнал  $s(t)$ , который восстанавливается из  $s(t_n)$ , например, путем сглаживания.

Так как квантование сигналов всегда выполняется с



определенной и неустранимой погрешностью (максимум - до половины интервала квантования), то операции АЦП и ЦАП не являются взаимно обратными с абсолютной точностью.

**Алиасинг.** А что произойдет, если спектр аналогового сигнала был неограниченным или имел частоту, выше частоты дискретизации?

Предположим, что при записи акустического сигнала оркестра в помещении от какого-то устройства присутствует ультразвуковой сигнал с частотой 30 кГц. Запись выполняется с дискретизацией сигнала на выходе микрофона с типовой частотой 44.1 кГц. При прослушивании такой записи с использованием ЦАП мы услышим шумовой сигнал на частоте  $30 - 44.1/2 \approx 8$  кГц. Восстановленный сигнал будет выглядеть так, как если бы частоты, лежащие выше половины частоты дискретизации, "зеркально" от нее отразились в нижнюю часть спектра и сложились с присутствующими там гармониками. Это так называемый эффект *появления ложных (кажущихся) частот* (aliasing). Эффект аналогичен всем известному эффекту обратного вращения колес автомобиля на экранах кино и телевизоров, когда скорость их вращения начинает превышать частоту смены кадров. Природу эффекта можно наглядно видеть на рис. 17.17.

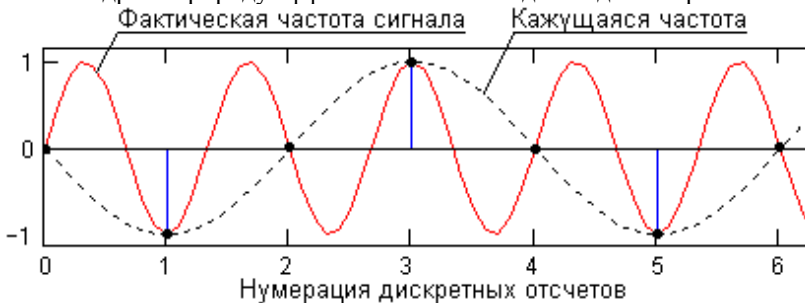


Рис. 17.17. Появление кажущейся частоты при дискретизации.

Аналогично в главный частотный диапазон дискретных сигналов "отражаются" от частоты дискретизации и все высокочастотные шумы, присутствующие в исходном аналоговом сигнале.

Для предотвращения алиасинга следует повышать частоту дискретизации или ограничить спектр сигнала перед оцифровкой *фильтрами низких частот (НЧ-фильтры, low-pass filters)*, которые пропускают без изменения все частоты, ниже заданной, и подавляют в сигнале частоты, выше заданной. Эта граничная частота называется *частотой среза (cutoff frequency)* фильтра. Частота среза анти-

алиасинговых фильтров устанавливается равной половине частоты дискретизации. В реальные АЦП почти всегда встраивается антиалиасинговый фильтр.

**Графическое отображение** сигналов общеизвестно и особых пояснений не требует. Для одномерных сигналов график – это совокупность пар значений  $\{t, s(t)\}$  в прямоугольной системе координат (рис. 17.13 – 17.16). При графическом отображении дискретных и цифровых сигналов используется либо способ непосредственных дискретных отрезков соответствующей масштабной длины над осью аргумента, либо способ огибающей (плавной или ломанной) по значениям отсчетов. В силу непрерывности физических полей и, как правило, вторичности цифровых данных, получаемых дискретизацией и квантованием аналоговых сигналов, второй способ графического отображения будем считать основным.

**Тестовые сигналы** (test signal). В качестве тестовых сигналов, которые применяются при моделировании и исследовании систем обработки данных, обычно используются сигналы простейшего типа: гармонические синус-косинусные функции, дельта-функция и функция единичного скачка.

**Дельта-функция** или функция Дирака. По определению, дельта-функция описывается следующими математическими выражениями (в совокупности):

$$\delta(t-\tau) = 0 \quad \text{при } t \neq \tau,$$
$$\int_{-\infty}^{\infty} \delta(t-\tau) dt = 1.$$

Функция  $\delta(t-\tau)$  не является дифференцируемой, и имеет размерность, обратную размерности ее аргумента, что непосредственно следует из безразмерности результата интегрирования. Значение дельта-функции равно нулю везде за исключением точки  $\tau$ , где она представляет собой бесконечно узкий импульс с бесконечно большой амплитудой, при этом площадь импульса равна 1.

Дельта-функция является полезной математической абстракцией. На практике такие функции не могут быть реализованы с абсолютной точностью, так как невозможно реализовать значение, равное бесконечности, в точке  $t = \tau$  на аналоговой временной шкале, т.е. определенной по времени также с бесконечной точностью. Но во всех случаях, когда площадь импульса равна 1, длительность импульса достаточно мала, а за время его действия на входе какой-либо системы сигнал на ее выходе практически не изменяется (реакция системы на

импульс во много раз больше длительности самого импульса), входной сигнал можно считать *единичной импульсной функцией* со свойствами дельта - функции.

При всей своей абстрактности дельта - функция имеет вполне определенный физический смысл. Представим себе импульсный сигнал прямоугольной формы  $\Pi(t-\tau)$  длительностью  $\theta$ , амплитуда которого равна  $1/\theta$ , а площадь соответственно равна 1. При уменьшении значения длительности  $\theta$  импульс, сокращаясь по длительности, сохраняет свою площадь, равную 1, и возрастает по амплитуде. Предел такой операции при  $\theta \Rightarrow 0$  и носит название дельта - импульса. Этот сигнал  $\delta(t-\tau)$  сосредоточен в одной координатной точке  $t = \tau$ , конкретное амплитудное значение сигнала не определено, но площадь (интеграл) остается равной 1. Это не мгновенное значение функции в точке  $t = \tau$ , а именно импульс (импульс силы в механике, импульс тока в электротехнике и т.п.) – математическая модель короткого действия, значение которого равно 1.

Дельта-функция обладает *фильтрующим свойством*. Суть его заключается в том, что если дельта-функция  $\delta(t-\tau)$  входит под интеграл какой-либо функции в качестве множителя, то результат интегрирования равен значению подынтегральной функции в точке  $\tau$  расположения дельта-импульса, т.е.:

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

Интегрирование в этом выражении может ограничиваться ближайшими окрестностями точки  $\tau$ .

**Функция единичного скачка** или функция Хевисайда иногда называется также функцией включения. Полное математическое выражение функции:

$$\sigma(t) = \begin{cases} 0 & \text{если } t < 0, \\ 1/2 & \text{если } t = 0, \\ 1 & \text{если } t > 0. \end{cases}$$

При моделировании сигналов и систем значение функции скачка в точке  $t=0$  очень часто принимают равным 1, если это не имеет принципиального значения.

Функция единичного скачка используется при создании математических моделей сигналов конечной длительности. При умножении любой произвольной функции, в том числе периодической, на прямоугольный импульс, сформированный из двух

последовательных функций единичного скачка

$$s(t) = \sigma(t) - \sigma(t-T),$$

из нее вырезается участок на интервале  $0-T$ , и обнуляются значения функции за пределами этого интервала.

**Функция Кронекера.** Для дискретных и цифровых систем разрешающая способность по аргументу сигнала определяется интервалом его дискретизации  $\Delta t$ . Это позволяет в качестве единичного импульса использовать дискретный интегральный аналог дельта-функции - функцию единичного отсчета  $\delta(k\Delta t - n\Delta t)$ , которая равна 1 в координатной точке  $k = n$ , и нулю во всех остальных точках. Функция  $\delta(k\Delta t - n\Delta t)$  может быть определена для любых значений  $\Delta t = \text{const}$ , но только для целых значений координат  $k$  и  $n$ , поскольку других номеров отсчетов в дискретных функциях не существует.

Математические выражения  $\delta(t-\tau)$  и  $\delta(k\Delta t - n\Delta t)$  называют также импульсами Дирака и Кронекера. Однако, применяя такую терминологию, не будем забывать, что это не просто единичные импульсы в координатных точках  $\tau$  и  $n\Delta t$ , а полномасштабные импульсные функции, определяющие как значения импульсов в определенных координатных точках, так и нулевые значения по всем остальным координатам, в пределе от  $-\infty$  до  $\infty$ .

### **17.3. Системы преобразования сигналов**

Сигналы, в любой форме материального представления, содержат определенную полезную информацию. Если при преобразованиях сигналов происходит нарушение заключенной в них информации (частичная утрата, количественное изменение соотношения информационных составляющих или параметров, и т.п.), то такие изменения называются *искажениями* сигнала. Если полезная информация остается неизменной или адекватной содержанию во входном сигнале, то такие изменения называются *преобразованиями* сигнала.

Математические преобразования сигналов осуществляются для того, чтобы получить какую-то дополнительную информацию, недоступную в исходном сигнале, или выделить из входного сигнала полезную информацию и сделать ее более доступной для дальнейшей обработки, измерений каких-либо параметров, передаче по каналам связи, и пр. **Преобразованный сигнал принято называть трансформантой исходного.**

Любые изменения сигналов сопровождаются изменением их спектра, и по характеру этого изменения разделяются на два вида: линейные и нелинейные. К *нелинейным* относят изменения, при которых в составе спектра сигналов появляются (вводятся) новые

гармонические составляющие, отсутствующие во входном сигнале. При *линейных* изменениях сигналов изменяются амплитуды и/или начальные фазы гармонических составляющих спектра (вплоть до полного подавления в сигнале определенных гармоник). И линейные, и нелинейные изменения сигналов могут происходить как с сохранением полезной информации, так и с ее искажением. Это зависит не только от характера изменения спектра сигналов, но и от спектрального состава самой полезной информации.

**Общее понятие систем.** Преобразование и обработка сигналов осуществляется в системах. Понятия сигнала и системы неразрывны, так как *любой сигнал существует в пределах какой-либо системы*. Система обработки сигналов может быть реализована как в материальной форме (специальное устройство, измерительный прибор, совокупность физических объектов с определенной структурой взаимодействия и т.п.), так и программно на ЭВМ или любом другом специализированном вычислительном устройстве. Форма реализации системы существенного значения не имеет, и определяет только ее возможности при анализе и обработке сигналов.

Безотносительно к назначению система всегда имеет *вход*, на который подается внешний входной сигнал, в общем случае многомерный, и *выход*, с которого снимается обработанный выходной сигнал (рис. 17.18).

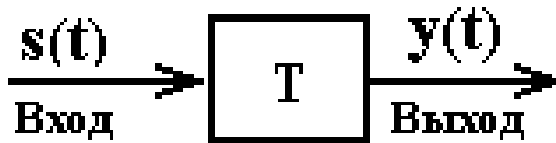


Рис. 17.18. Графическое представление системы.

Собственно *система представляет собой системный оператор (алгоритм) преобразования входного сигнала  $s(t)$  – воздействия или возбуждения, в сигнал на выходе системы  $y(t)$  – отклик или выходную реакцию системы*. Символическое обозначение операции преобразования (трансформации сигнала):

$$y(t) = T[s(t)].$$

**Системный оператор T** - это *набор правил преобразования (transformation) сигнала  $s(t)$  в сигнал  $y(t)$* . Так, например, в самом простейшем случае таким правилом может быть таблица перекодировки входных сигналов в выходные.

Для детерминированных входных сигналов соотношение между выходными и входными сигналами всегда однозначно задается

системным оператором. В случае реализации на входе системы случайного входного процесса происходит изменение статистических характеристик сигнала (математического ожидания, дисперсии, корреляционной функции и пр.), которое также определяется системным оператором.

Для полного *определения системы* необходимо задание *характера, типа и области допустимых величин входных и выходных сигналов.*

По типу обработки входных сигналов они обычно подразделяются на *системы непрерывного времени* для обработки сигналов в процессе измерений, и *цифровые системы* для обработки данных, зарегистрированных на промежуточных носителях.

*Совокупность системного оператора  $T$  и областей входных и выходных сигналов образует математическую модель системы.*

*Линейные и нелинейные системы* составляют два основных класса систем обработки сигналов.

Термин линейности (linear) означает, что система преобразования сигналов должна иметь произвольную, но в обязательном порядке линейную связь между входным сигналом (возбуждением) и выходным сигналом (откликом) с определенным изменением спектрального состава входного сигнала (усиление или подавление определенных частотных составляющих сигнала). В нелинейных (nonlinear) системах связь между входным и выходным сигналом определяется произвольным нелинейным законом с дополнением частотного состава входного сигнала частотными составляющими, отсутствующими во входном сигнале.

*Стационарные и нестационарные системы.* Система считается стационарной и имеет *постоянные параметры*, если ее свойства (математический алгоритм оператора преобразования) в пределах заданной точности не зависят от входного и выходного сигналов и не изменяются ни во времени, ни от каких-либо других внешних факторов. В противном случае система является нестационарной, и называется *параметрической* или *системой с переменными параметрами*. Среди последних большое значение имеют так называемые *адаптивные системы* обработки данных. В этих системах производится, например, оценивание определенных параметров входных и выходных сигналов, по результатам сравнения которых осуществляется подстройка параметров преобразования (переходной характеристики системы) таким образом, чтобы обеспечить оптимальные по производительности условия обработки сигналов или минимизировать погрешность обработки.

**Основные системные операции.** К базовым линейным операциям, из которых могут быть сформированы любые линейные операторы преобразования, относятся операции **скалярного умножения, сдвига и сложения сигналов:**

$$y(t) = c \times s(t), \quad y(t) = s(t-\Delta t), \quad y(t) = a(t)+b(t).$$

Для нелинейных систем выделим важный тип безинерционных операций нелинейной трансформации сигнала, результаты которой зависят только от его входных значений. К ним относятся, например, операции квадратирования и логарифмирования сигнала:

$$y(t) = [s(t)]^2, \quad y(t) = \log[s(t)].$$

**Линейные системы.** Система считается линейной, если ее реакция на входные сигналы **аддитивна** (выполняется принцип суперпозиции сигналов) и **однородна** (выполняется принцип пропорционального подобия). Другими словами, **отклик линейной системы на взвешенную сумму входных сигналов должен быть равен взвешенной сумме откликов на отдельные входные сигналы независимо от их количества и для любых весовых коэффициентов**, в том числе комплексных.

При программной реализации линейных систем на ЭВМ особых затруднений с обеспечением линейности в разумных пределах значений входных и выходных сигналов, как правило, не возникает. При физической (аппаратной) реализации систем обработки данных диапазон входных и выходных сигналов, в котором обеспечивается линейность преобразования сигналов, всегда ограничен и должен быть специально оговорен.

**Инвариантность систем к сдвигу.** Система называется инвариантной к сдвигу, если сдвиг входного сигнала по аргументам (времени, координатам пространства и т.п.) вызывает соответствующий сдвиг выходного сигнала:

$$y(x,t) = T[s(x,t)], \quad T[s(x-\Delta x,t-\Delta t)] = y(x-\Delta x,t-\Delta t).$$

Это означает, что форма выходного сигнала зависит только от входного сигнала, и не зависит от времени поступления сигнала на вход системы. Инвариантность системы к сдвигу является одним из подтверждений постоянства ее параметров.

**Линейные системы, инвариантные к сдвигу.** Линейность и инвариантность к сдвигу являются независимыми свойствами систем и не определяют друг друга. Так, например, операция квадратирования сигнала инвариантна к сдвигу, но нелинейна.

В теории анализа и обработки данных основное место занимают системы, линейные и инвариантные к сдвигу (ЛИС - системы). Они обладают достаточно широкими практическими

возможностями при относительной простоте математического аппарата. В дальнейшем, если это специально не оговаривается, будем иметь в виду именно такие системы.

Преимущество, которое отдается ЛИС - системам в методах обработки информации, базируется на возможности разложения входного сигнала любой, сколь угодно сложной формы, на составляющие простейших форм, отклик системы на которые известен и хорошо изучен, с последующим вычислением выходного сигнала в виде суммы откликов на все составляющие входного сигнала. В качестве простейших форм разложения сигналов используются, как правило, единичные импульсы и гармонические составляющие. Разложение по единичным импульсам применяется при динамическом представлении сигнала в зависимости от реальных физических аргументов (времени, координат и пр.) и использует операцию свертки. Разложение на гармонические составляющие использует спектральное (частотное) представление сигнала и преобразование Фурье.

**Соединения ЛИС - систем.** При последовательном (каскадном) соединении систем выходной сигнал одной системы служит входным сигналом для второй и т.д. в зависимости от количества составляющих систем каскада. По отношению к общей системной операции преобразования порядок соединения входящих в нее систем значения не имеет. Так, для двух последовательно соединенных систем на рис. 17.19:

$$y(t) = T_2[T_1[s(t)]] = T_1[T_2[s(t)]]$$

Последовательное соединение



Параллельное соединение

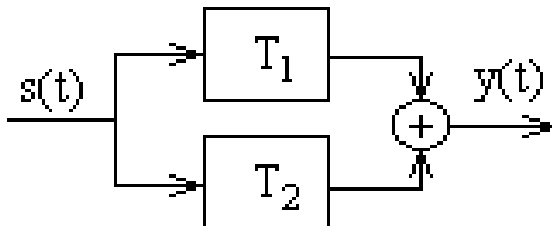


Рис. 17.19. Соединения систем.



При параллельном соединении входной сигнал поступает одновременно на входы всех составляющих систем, а выходные сигналы систем суммируются:

$$y(t) = T_1[s(t)] + T_2[s(t)] + \dots + T_N[s(t)].$$

Образуемые в результате соединений системы в целом также являются ЛИС - системами, если линейны и инвариантны к сдвигу системы, в них входящие.

**Обобщенная схема системы** цифровой обработки сигналов на рис. 17.20 приведена в качестве примера.

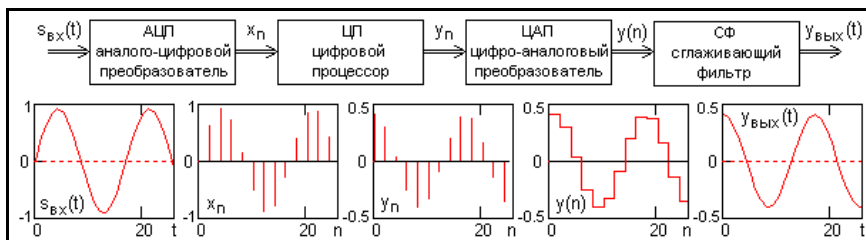


Рис. 17.20. Структурная схема системы дифференцирования сигналов.

## 17.4. Информационная емкость сигналов

Объем информации, находящейся в обращении и необходимой для функционирования и развития современного общества, нарастает примерно пропорционально квадрату развития производительных сил. В передовых по научно-техническому развитию странах мира доля рабочей силы, занятой вопросами сбора, обработки и обеспечения информацией, превышает долю рабочей силы в сфере материального производства. Применение методов и средств автоматизации на всех этапах обращения информации, эффективная организация ее хранения, обработки и обмена, приобретают все большее значение в качестве основного условия успешного функционирования экономики стран.

**Понятие информации.** Как мы уже отмечали в предыдущих разделах данной работы, в настоящее время нет общепринятого и однозначного понимания термина "Информация". Спектр бытующих понятий весьма широк, от общего философского - информация есть отражение реального мира, до узко практического - информация есть сведения, являющиеся объектом хранения, передачи и преобразования. Расхождения существуют и по вопросу места информации в материальном мире. Это свойство индивидуальных объектов или результат их взаимодействия? Присуща ли информация всем видам материи или лишь определенным образом организованной материи?

**В информатике под информацией понимается, как правило, совокупность сведений смыслового содержания, которые можно собирать, обрабатывать, передавать и т.п.** Причем именно сведений в изначальном смысле латинского слова informatio, а не данных или сигналов, которые являются носителями этих сведений. В таком понимании процессы извлечения сведений из данных и их интерпретации неразрывно связаны с разумом, а конечным результатом обработки и восприятия информации с помощью разума является раскрытие неопределенности знаний о каком-либо объекте, явлении или процессе. Но при таком подходе размывается само понятие разума.

С одной стороны, существование любого живого существа поддерживается до тех пор, пока действуют его органы чувств (датчики), преобразующие физические воздействия окружающего мира в определенным образом организованные сигналы, которые в материальной форме отображают данные об этих воздействиях. Данные собираются и интерпретируются определенной системой, которую в самой общей форме мы называем "разумом", из общей суммы данных извлекаются определенные сведения, степень неопределенности сведений об окружающей обстановке снижается, и ... лиса распутывает заячий след. ***Живое существо существует до тех пор, пока способно воспринимать и обрабатывать внешние и внутренние воздействия.*** Нет сомнений и в том, что в коллективных сообществах его члены не только способны собирать и обрабатывать информацию, но и передавать ее другим членам сообщества, как, например, в пчелиной семье точный путь до продуктивного цветочного массива. Информационный танец пчелы в этом отношении по компактности содержания ничем не уступает телеграфному сообщению. Естественно, в принятой у пчел символической форме.

С другой стороны, если информация неразрывно связана с "разумом", то в этом случае нельзя отказать в "разуме" и электронной вычислительной машине, обыгрывающей в шахматы чемпиона мира, а равно и любым устройствам технической кибернетики любого уровня, так как все они имеют определенные системы сбора, передачи, накопления, хранения и обработки информации той или иной степени сложности, и на основе этой информации способны формировать сигналы обратной связи для управления определенными процессами.

В технических отраслях знаний, где вопросы соотношения информации с разумом не стоят на первом месте, преобладает ***понимание информации в виде отображения такого всеобщего свойства материи, как многообразие, как характеристики внутренней организованности материальных систем, процессов***

**или явлений по множеству состояний, которые для них возможны.**

В такой трактовке информация существует независимо от того, воспринимается она каким-либо "разумом" или нет, и является одним из свойств материальных объектов. "Информация есть информация, а не материя и не энергия" (Норберт Винер). Это свойство в какой-то мере имеет потенциальный характер. Информация может проявлять себя только при взаимодействии объектов или процессов, может возникать (создаваться) и исчезать (уничтожаться).

Но и в такой трактовке возникает очень много вопросов, на которые трудно дать однозначные ответы. Насекомое третичного периода, неизвестное в настоящее время ученым, прилипло к капле смолы хвойного дерева. Новый слой смолы закрыл насекомое. Дерево упало, и его занесло песком. Смола превратилась в янтарь. Янтарь в потенциале содержит полную информацию о насекомом, потому как в нем десятки тысяч фрагментов ДНК - информации, достаточная для восстановления ДНК и воспроизводства насекомого, если не в настоящее время, то в ближайшем будущем. Но когда она возникла? В момент появления насекомого с его ДНК? В момент прилипания к смоле? В момент окаменения? Можно ли говорить о появлении информации, если еще не существовал субъект, способный извлечь и использовать эту информацию? Наконец, янтарь с насекомым найден и попал на глаза палеонтолога. Определен новый вид насекомого. Появилась первая частичная информация? Так может быть, информация появляется только при активном и целенаправленном воздействии на объект исследований? А если янтарь оказался непрозрачным, и его переплавили? Исчезла ли информация? И можно ли считать, что она вообще была?

Ответы на эти и подобные им вопросы тяготеют к двум полюсам, а по существу, к двум диаметрально противоположным философским позициям.

Сторонники первой позиции считают, что информация присуща только определенным образом организованным объектам или процессам, и понимают под информацией только то, что может восприниматься, обрабатываться, осмысливаться и использоваться, т.е. является продуктом целенаправленного процесса сбора, организации, систематизации и использования сведений о материальных объектах и процессах.

Противоположная позиция, это понятие информации как свойства объектов и процессов воспринимать и перерабатывать внутреннее состояние и внешнее воздействие окружающей среды, сохранять его результаты и передавать их другим объектам. С этой позиции все материальные объекты и процессы являются источниками,

носителями и потребителями информации, на основе которой и идет развитие реального мира. По существу, это соответствует принятию материальности информации и информационной основы мироздания.

При неопределенности самого понятия информации можно достаточно обоснованно считать, что информация проявляется, хранится и передается от одного объекта к другому в материально - энергетической форме в виде сигналов. ***Сигналом, как материальным носителем информации, может быть любой физический процесс (электрический, магнитный, оптический, акустический и пр.), определенные параметры которого (амплитуда, частота, энергия, интенсивность и др.) однозначно отображают информационные данные (сообщения).***

**Количественная мера информации.** Напомним кратко метод определения количества информации, подробно рассмотренный нами в предыдущих разделах настоящей работы.

Теория любого явления начинается с появления количественных взаимоотношений между объектами исследований, т.е. при установлении принципов измеряемости каких-либо свойств объектов. Единицу количественной меры информации - БИТ (сокращение binary digit - двоичная цифра), впервые предложил Р. Хартли в 1928 году. 1 бит - это информация о двух возможных равновероятных состояниях объекта, неопределенность выбора из двух равновероятных событий. Математически это отображается состоянием 1 или 0 одного разряда двоичной системы счисления. Количество информации  $H$  (в битах), необходимое и достаточное для полного снятия неопределенности состояния объекта, который имеет  $N$  равновозможных состояний, измеряется как логарифм по основанию 2 из числа возможных состояний:

$$H = \log_2 N. \quad (17.10)$$

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

---

**Пример.** Необходимо поднять груз на определенный этаж 16-ти этажного здания (нумерация этажей 0-15,  $N = 16$ ).

Сколько бит информации полностью определяют задание?

$$H = \log_2 N = \log_2 16 = 4.$$

Следовательно, 4 бита информации необходимы и достаточны для полного снятия неопределенности выбора. В этом можно убедиться применением логики исчисления с последовательным делением пополам интервалов состояний. Например, для 9-го этажа:

---

1. Выше 7-го этажа? Да = 1. 2. Выше 11-го этажа?  
Нет = 0.  
3. Выше 9-го этажа? Нет = 0. 4. Выше 8-го этажа?  
Да = 1.

Итог: этаж номер 9 или 1001 в двоичном исчислении,  
четыре двоичных разряда.

---

Если в приведенном примере на этажах имеется по 4 квартиры с нумерацией на каждом этаже 0-3 ( $M=4$ ), то при адресации груза в квартиру потребуется еще 2 бита информации. Такой же результат получим, если вместо независимой нумерации этажей и квартир на этажах (два источника неопределенности) мы будем иметь только сквозную нумерацию квартир (один обобщенный источник):

$H = \log_2 N + \log_2 M = \log_2 16 + \log_2 4 = 6 \equiv \log_2 (N \times M) = \log_2 64 = 6$ , т.е. **количество информации** отвечает требованию **аддитивности: неопределенность объединенного источника равна сумме неопределенностей исходных источников**, что соответствует интуитивному требованию к информации: она должна быть однозначной, а ее количество должно быть одним и тем же независимо от способа задания.

Основание логарифма не имеет принципиального значения и определяет только масштаб или единицу неопределенности. Так, если за единицу неопределенности принять три равновероятных состояния, то для определения, например, одной фальшивой золотой монеты (более легкой) из 27 внешне неотличимых монет потребуется только  $H = \log_3 27 = 3$ , т.е. три взвешивания на равноплечных весах. Логику исчисления взвешиваний предлагается определить самостоятельно.

Двоичная мера информации получила общее признание в связи с простотой реализации информационной техники на элементах с двумя устойчивыми состояниями. В десятичном исчислении единицей информации является один десятичный разряд - ДИТ.

**Энтропия источника информации.** Степень неопределенности состояния объекта (или так называемого источника информации) зависит не только от числа его возможных состояний, но и от вероятности этих состояний. При неравновероятных состояниях свобода выбора для источника ограничивается. Так, если из двух возможных состояний вероятность одного из них равна 0.999, то вероятность другого состояния соответственно равна  $1-0.999 = 0.001$ , и при взаимодействии с таким источником результат практически предreshен.

В общем случае, в соответствии с теорией вероятностей, источник информации однозначно и полно характеризуется ансамблем состояний  $U = \{u_1, u_2, \dots, u_N\}$  с вероятностями состояний соответственно

$\{p(u_1), p(u_2), \dots, p(u_N)\}$  при условии, что сумма вероятностей всех состояний равна 1. Мера количества информации, как неопределенности выбора дискретным источником состояния из ансамбля  $U$ , предложена К. Шенноном в 1946 году и получила название *энтропии дискретного источника информации* или энтропии конечного ансамбля:

$$H(U) = - \sum_{n=1}^N p_n \log_2 p_n. \quad (17.11)$$

Выражение Шеннона совпадает с выражением Больцмана для энтропии физических систем при оценке степени разнообразия их состояний. Мера энтропии Шеннона является обобщением меры Хартли на случай ансамблей с неравновероятными состояниями, в чем нетрудно убедиться, если в выражении (17.11) значение  $p_n$  заменить значением  $p=1/N$  для ансамбля равновероятных состояний. *Энтропия конечного ансамбля  $H(U)$  характеризует неопределенность, приходящуюся в среднем на одно состояние ансамбля.*

Учитывая, что в дальнейшем во всех математических выражениях, касающихся энтропии, мы будем использовать только двоичное основание логарифма, индекс 2 основания логарифма в формулах будем подразумевать по умолчанию.

$u_i$	$p_i$	$u_i$	$p_i$	$u_i$	$p_i$	$u_i$	$p_i$	$u_i$	$p_i$
а	.064	з	.015	о	.096	х	.009	э	.003
б	.015	и	.064	п	.024	ц	.004	ю	.007
в	.039	й	.010	р	.041	ч	.013	я	.019
г	.014	к	.029	с	.047	ш	.006	-	.124
д	.026	л	.036	т	.056	щ	.003		
е,ё	.074	м	.026	у	.021	ъ,ь	.015		
ж	.008	н	.056	ф	.020	ы	.016		

**Пример.** Вычислить энтропию ансамбля 32 букв русского алфавита. Вероятности использования букв приведены в таблице. Сравнить энтропию с неопределенностью, которая была бы у алфавита при равновероятном их использовании.

Неопределенность на одну букву при равновероятности использования:  $H(u) = \log 32 = 5$

Энтропия алфавита по ансамблю таблицы:

$$H(u) = - 0.064 \log 0.064 - 0.015 \log 0.015 - \dots - 0.143 \log 0.143 \approx 4.42.$$

Таким образом, неравновероятность состояний снижает энтропию источника.

**Основные свойства энтропии:**

1. Энтропия является величиной вещественной и неотрицательной, т.к. значения вероятностей  $p_n$  находятся в интервале 0-1, значения  $\log p_n$  всегда отрицательны, а значения  $-p_n \log p_n$  в (1.4.2) соответственно положительны.

2. Энтропия - величина ограниченная, т.к. при  $p_n \Rightarrow 0$  значение  $-p_n \cdot \log p_n$  также стремится к нулю, а при  $0 < p_n \leq 1$  ограниченность суммы всех слагаемых очевидна.

3. Энтропия равна 0, если вероятность одного из состояний источника информации равна 1, и тем самым состояние источника полностью определено (вероятности остальных состояний источника равны нулю, т.к. сумма вероятностей должна быть равна 1).

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

$$H_{\max}(U) = - \sum_{n=1}^N (1/N) \log (1/N) = \log N.$$

5. Энтропия источника с двумя состояниями  $u_1$  и  $u_2$  при изменении соотношения их вероятностей  $p(u_1)=p$  и  $p(u_2)=1-p$  определяется выражением:

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

и изменяется от 0 до 1, достигая максимума при равенстве вероятностей. График изменения энтропии приведен на рис. 17.21.

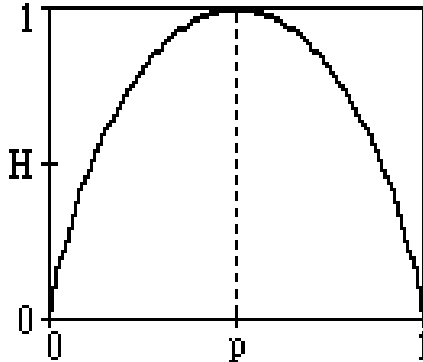


Рис. 17.21.

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

Рассмотрим это свойство на двух источниках информации  $u$  и  $v$ . При объединении источников получаем обобщенный источник информации  $(u,v)$ , который описывается вероятностями  $p(u_n v_m)$  всех

возможных комбинаций состояний  $u_n$  источника  $u$  и  $v_m$  источника  $v$ . Энтропия объединенного источника при  $N$  возможных состояниях источника  $u$  и  $M$  возможных состояниях источника  $v$ :

$$H(UV) = - \sum_{n=1}^N \sum_{m=1}^M p(u_n v_m) \log p(u_n v_m),$$

Источники статистически независимы друг от друга, если выполняется условие:

$$p(u_n v_m) = p(u_n) \cdot p(v_m).$$

С использованием этого условия соответственно имеем:

$$\begin{aligned} H(UV) &= - \sum_{n=1}^N \sum_{m=1}^M p(u_n) p(v_m) \log [p(u_n) p(v_m)] = \\ &= - \sum_{n=1}^N p(u_n) \log p(u_n) \sum_{m=1}^M p(v_m) - \sum_{m=1}^M p(v_m) \log p(v_m) \sum_{n=1}^N p(u_n). \end{aligned}$$

С учетом того, что  $\sum_{n=1}^N p(u_n) = 1$  и  $\sum_{m=1}^M p(v_m) = 1$ , получаем:

$$H(UV) = H(U) + H(V). \tag{17.12}$$

7. Энтропия характеризует среднюю неопределенность выбора одного состояния из ансамбля, полностью игнорируя содержательную сторону ансамбля. С одной стороны, это расширяет возможности использования энтропии при анализе самых различных явлений, но, с другой стороны, требует определенной дополнительной оценки возникающих ситуаций. Как это следует из рис. 17.21, энтропия состояний может быть неоднозначной, и если в каком-либо начинании действие  $u$  с вероятностью  $p_u=r$  приводит к успеху, а действие  $v$  с вероятностью  $p_v=1-r$  к неудаче, то выбор действий по оценке энтропии может оказаться и прямо противоположным, т.к. энтропия при  $p_v=r$  равна энтропии при  $p_u=r$ .

**Энтропия непрерывного источника информации**

должна быть бесконечна, т. к. неопределенность выбора из бесконечно большого числа возможных состояний бесконечно велика.

Разобьем диапазон изменения непрерывной случайной величины  $U$  на конечное число  $n$  малых интервалов  $\Delta u$ . При реализации значений  $u$  в интервале  $(u_n, u_n + \Delta u)$  будем считать, что реализовалось значение  $u_n$  дискретной случайной величины  $U'$ , вероятность реализации которой:



$$p(u_n < u < u_n + \Delta u) = \int_{u_n}^{u_n + \Delta u} p(u) du \approx p(u_n) \Delta u.$$

Энтропия дискретной величины  $U'$ :

$$H(U') = - \sum_{n=1}^N p(u_n) \Delta u \log (p(u_n) \Delta u).$$

Заменяем  $\log (p(u_n) \Delta u) = \log p(u_n) + \log \Delta u$ , принимаем во внимание, что сумма  $p(u_n)\Delta u$  по всем возможным значениям  $u_n$  равна 1, и получаем:

$$H(U') = - \sum_{n=1}^N p(u_n) \Delta u \log p(u_n) - \log \Delta u. \quad (17.12)$$

В пределе, при  $\Delta u \rightarrow 0$ , получаем выражение энтропии для непрерывного источника:

$$H(U) = - \int_{-\infty}^{\infty} p(u) \log p(u) du - \lim_{\Delta u \rightarrow 0} \log \Delta u. \quad (17.13)$$

Значение энтропии в (17.13), как и ожидалось, стремится к бесконечности за счет второго члена выражения. Для получения конечной характеристики информационных свойств непрерывных сигналов используют только первый член выражения (17.13), получивший название *дифференциальной энтропии*. Ее можно трактовать, как среднюю неопределенность выбора произвольной случайной величины по сравнению со средней неопределенностью выбора случайной величины  $U'$ , имеющей равномерное распределение в диапазоне (0-1). Действительно, для такого распределения

$$p(u_n) = 1/N, \Delta u = 1/N,$$

и при  $N \rightarrow \infty$  из (17.12) следует:

$$H(U') = - (\log N)/N - \log \Delta u \rightarrow - \lim_{\Delta u \rightarrow 0} \log \Delta u.$$

Соответственно, разность энтропий дает дифференциальную энтропию:

$$h(U) = H(U) - H(U') = - \int_{-\infty}^{\infty} p(u) \log p(u) du. \quad (17.14)$$

Дифференциальная энтропия не зависит от конкретных значений величины  $U$ :

$$h(U+a) = h(U), \quad a = \text{const},$$

но зависит от масштаба ее представления:

$$h(kU) = h(U) + \log k.$$

Практика анализа и обработки сигналов обычно имеет дело с сигналами в определенном интервале  $[a, b]$  их значений, при этом

максимальной дифференциальной энтропией обладает равномерное распределение значений сигналов:

$$h(U) = - \int_a^b p(u) \log p(u) du = \log (b-a).$$

По мере сужения плотности распределения значение  $h(U)$  уменьшается, и в пределе при  $p(u) \rightarrow \delta(u-c)$ ,  $a < c < b$  стремится к нулю.

**Информационная емкость сигналов** существенно зависит от типа сигналов и определяет требования к каналам передачи данных, равно как и технические характеристики каналов связи ограничивают информационную емкость сигналов, передаваемых по этим каналам.

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

Под **технической скоростью передачи** подразумевают число элементарных сигналов (символов), передаваемых по каналу в единицу времени. Простейший элементарный символ – однополярный электрический импульс длительностью  $\tau$  на тактовом интервале  $T$ . В дискретных каналах используют, как правило, двуполярные импульсы, положительные на первой половине интервала  $T$  и отрицательные на второй половине. Это позволяет поддерживать нулевой потенциал кабеля и выполнять тактовую синхронизацию приема-передачи сигналов. **Единицей измерения технической скорости**  $V_t = 1/T$  служит **БОД – один символ в секунду**. Полоса пропускания канала связи ограничивается предельной частотой  $F_{\text{пред}}$  по уровню затухания сигнала до уровня статистических помех, при этом значение технической скорости передачи данных не может быть выше  $F_{\text{пред}}$  без специальных устройств выделения информационных сигналов.

При известной технической скорости  $V_t$  скорость передачи информации измеряется в битах в секунду и задается соотношением:

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

где  $H(s)$  – энтропия символа. Для двоичных дискретных символов  $[0, 1]$  при постоянной амплитуде импульсов значение  $H(s)$  равно 1. При числе  $L$  возможных равновероятных уровней амплитуды импульсов (уровень помех меньше разности уровней амплитуд импульсов) значение  $H(s)$  равно  $\log L$ .

**Информационная емкость сигнала или полное количество информации** в сигнале  $S$  (сообщении, кодовой последовательности/слове) определяется полным количеством  $N = t/T$  энтропии символов в битах на интервале задания сигнала  $t$ :

$$I_t(S) = N \log L = (t/T) \log L. \quad (17.15)$$

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

Для непрерывных сигналов передача по каналам связи возможна только при условии, что максимальная информационная частота в сигнале  $F_{\max}$  не превышает предельной частоты  $F_{\text{пред}}$  передачи сигналов каналом связи. Для оценки информационной емкости непрерывного сигнала выполним его дискретизацию с интервалом  $\Delta t = 1/2F_{\max}$ . Как установлено Котельниковым и Шенноном, по мгновенным отсчетам непрерывного сигнала с таким интервалом дискретизации аналоговый сигнал может быть восстановлен без потери информации. При полной длительности сигнала  $T_s$  число отсчетов:

$$N = T_s/\Delta t = 2F_{\max} T_s.$$

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

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

Информационная емкость сигнала:

$$I(S) = 2F_{\max} T_s \log L. \quad (17.16)$$

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

В этом разделе метрология сигналов рассматривается, в основном, на уровне понятий и базовых определений, предвзяв их более подробное изучение в дальнейших разделах. Это объясняется тем, что при детальном изучении каких-либо характеристик или свойств сигналов их рассмотрение не может выполняться в отрыве от других метрологических характеристик сигналов и требует определенной ориентировки в общей метрологии сигналов, хотя бы на уровне понятий.

## **17.5. Пространство сигналов**

Важнейшее свойство аналоговых и дискретных сигналов заключается в том, что их линейные комбинации также являются аналоговыми или дискретными сигналами. Линейные комбинации цифровых сигналов, в силу их ограничения по разрядности, в принципе относятся к разряду нелинейных операций, однако последним фактором можно пренебречь, если ошибки, которые вносятся в

результаты наблюдений при квантовании отсчетов, достаточно малы по сравнению с шумами зарегистрированной информации. При дискретизации и квантовании данных непосредственно на входах в ЭВМ это условие выполняется практически всегда, поскольку ошибки определяются разрядностью ЭВМ и программными системами обработки данных, которые обычно не ниже 6-12 десятичных разрядов.

**Множества сигналов.** Сигналы обычно рассматриваются в составе определенных множеств  $L$ , объединенных каким-либо свойством  $P$ , характерным для всех и каждого из сигналов данного множества. Условное отображение множества:  $L = \{s; P\}$  – множество всех  $s$ , для которых справедливо свойство  $P$ . Определив свойство  $P$ , мы тем самым можем ограничивать сигналы, действующие в каких-либо системах, определенными типами, условиями, границами по параметрам и т.п.

---

**Пример 1. Множество гармонических сигналов.**

$$L = \{s; s(t) = A \cdot \cos(\omega t + \varphi), \quad -\infty < t < \infty\}.$$

Множество содержит гармонические сигналы с произвольными значениями амплитуд, частот и фаз.

**Пример 2. Множество периодических сигналов.**

$$L(T) = \{s; s(t) = s(t+kT), \quad -\infty < t < \infty, \quad k \in \mathbb{I}\}.$$

**Пример 3. Множество сигналов, ограниченных по амплитуде и длительности.**

$$L(K, T) = \{s; |s(t)| \leq K, \quad s(t) = 0 \text{ при } |t| > T\}.$$

---

**Терминология операций с множествами сигналов.**

Множества сигналов могут образовываться из других (ранее определенных) множеств логическими операциями объединения (индекс объединения -  $\cup$ ) и пересечения (индекс -  $\cap$ ):

$$L = S_1 \cup S_2 = \{s; s \in S_1 \text{ или } s \in S_2\},$$

$$L = S_1 \cap S_2 = \{s; s \in S_1 \text{ и } s \in S_2\}.$$

Возможно также разбиение множества сигналов на непересекающиеся подмножества, более удобные для обработки, при этом для множества  $S$ , разбитого на совокупность подмножеств  $\{S_1, S_2, S_3, \dots, S_N\}$ , должны выполняться условия:

$$S = S_1 \cup S_2 \cup S_3 \cup \dots \cup S_N,$$

$$S_n \cap S_m = \emptyset \text{ для } n \neq m.$$

Запись  $S_1 \subset S$  означает, что множество  $S_1$  входит в состав множества  $S$ , т.е. является подмножеством в составе  $S$ .

**Преобразование** элементов  $v_i$  множества  $V$  в элементы  $g_i$  множества  $G$  называется **отображением** (трансформацией, преобразованием)  $V$  в  $G$  (символьные записи:  $g = T[v]$  или  $v \rightarrow g$ ), при

этом элементы  $v$  называют **прообразом** множества  $g$ , а элементы  $g$  – **образом** множества  $v$ .

Если преобразование выполняется над числами одного множества  $R$  (например,  $x = T[y]$ ), то такое преобразование порождает **функциональную зависимость**  $x = f(y)$ .

Если преобразование выполняется над функциями одного и того же множества  $L$  (например,  $f(t) = T[g(t)]$ ), то **алгоритм преобразования**  $T[...]$  называют **оператором** преобразования  $f(t)$  в  $g(t)$ .

Преобразование  $g = T[f(t)]$  функций  $f(t)$  множества  $F$  называют **функционалом**, если результатом преобразования являются числовые значения  $g$  множества  $G$ . Примерами функционалов являются **интегралы функций в определенных пределах**.

Преобразование может выполняться **функциональными операторами** с переводом функций одной переменной, например  $t$ , в функции по другой переменной, например  $\omega$ . Типичным примером функционального оператора является преобразование Фурье. В комплексной форме:

$$S(\omega) = \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt.$$

**Пространство сигналов.** Для анализа и обработки информации, которая может быть заключена в сигналах, требуется выделять из множества сигналов сигналы с **определенными параметрами**, сравнивать сигналы друг с другом, оценивать изменение сигналов при их прохождении через системы обработки данных, и т.п. Это может выполняться только при "помещении" множества сигналов в определенные **метрические пространства** с заранее оговоренными **свойствами и единицами измерений**. Так, "квартирное пространство" любого города включает, как минимум, три структурных единицы: названия улиц, номера домов, номера квартир, что и определяет **пространство "квартирных сигналов"**. Но это пространство не является **метрическим**, так как оно не имеет **нулевой точки и единиц измерений**, по нему нельзя определить расстояние между двумя "квартирными сигналами". Положение на поверхности Земли любого объекта однозначно определяется по "координатному сигналу" в заранее сформированных метрических координатных пространствах с нулевыми точками и принятыми единицами измерений. Для практического использования определенными структурными ограничениями сформированы также различные пространства картографических проекций, жестко установленная метрология которых позволяет трансформировать информацию из одного пространства в другое, например, более удобное для обработки

определенными программами.

Главным условием превращения множество сигналов

$L\{s_1(t), s_2(t), \dots\}$ , которые имеют какие-то общие свойства, в **функциональное пространство** сигналов является выполнение условия однозначной реализации. Если **пространство значений** независимой переменной  $t$  задано выражением  $R:=(-\infty, +\infty)$ , то **пространство сигналов**  $L^p[R]$  определяет множество сигналов в этом пространстве, для которых условие однозначной реализации записывается в следующей форме:

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

Для анализа сигналов наиболее часто используется **гильбертово пространство**, сигналы в котором должны удовлетворять условию интегрирования с квадратом:

$$\int_{-\infty}^{\infty} |s(t)|^2 dt < \infty.$$

Периодические сигналы обычно рассматриваются в пространстве  $L^2[0, 2\pi]$  одного периода:

$$\int_0^{2\pi} |s(t)|^2 dt < \infty.$$

Интуитивно понятно, что метрические пространства должны иметь определенную систему координат, что позволяет рассматривать любые произвольные сигналы  $x$  и  $y$ , принадлежащие пространству, в виде векторов, соединяющих начало координат с определенными точками этого пространства, и определять расстояние между этими точками  $\rho(x,y)$ . Так как расстояние между точками должно быть числовым, а сигналы  $x$  и  $y$  представляют собой функции, то  $\rho(x,y)$  представляет собой функционал, для которого в метрическом пространстве должны быть справедливы следующие аксиомы:

- $\rho(x,y) \geq 0$ ;  $\rho(x,y) = 0$  при  $x = y$ ,
- $\rho(x,y) = \rho(y,x)$ ,
- $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$  - неравенство треугольника.

Каждый элемент векторного пространства может отображаться проекциями на координатные оси, а для обработки и преобразований сигналов могут использоваться операции векторной алгебры. Достаточно простые алгебраические взаимосвязи между сигналами характерны для линейных пространств.

**Линейное пространство сигналов.** Множество сигналов  $L$  образует линейное пространство сигналов, если для него справедливы следующие аксиомы:

1. Множество содержит такой нулевой элемент  $\emptyset$ , что для всех сигналов  $u(t) \in L$  выполняется равенство  $u(t) + \emptyset = u(t)$ .
2. Для любых сигналов  $u(t) \in L$  и  $v(t) \in L$  существует их сумма  $s(t) = u(t) + v(t)$ , которая также содержится в  $L$ . При этом операция суммирования должна быть
  - коммутативна:  $u(t) + v(t) = v(t) + u(t)$ ,
  - ассоциативна:  $u(t) + (v(t) + x(t)) = (u(t) + v(t)) + x(t)$ ,
  - однородна:  $u(t) + (-u(t)) = \emptyset$ .
3. Существует множество скалярных элементов  $\alpha$ , на которые может выполняться умножение любого сигнала  $s(t) \in L$ , при этом результат умножения является новым сигналом  $y(t) = \alpha s(t)$  в том же пространстве,  $y(t) \in L$ . Операция умножения должна быть
  - ассоциативна:
$$\alpha(\beta \cdot s(t)) = \alpha\beta \cdot s(t),$$
  - дистрибутивна:
$$\alpha(u(t) + s(t)) = \alpha u(t) + \alpha s(t), (\alpha + \beta)s(t) = \alpha s(t) + \beta s(t),$$
  - пропорциональна:
$$1 \cdot s(t) = s(t), 0 \cdot s(t) = 0.$$

---

**Пример.** Множество сигналов  $L$  состоит из импульсных сигналов произвольной формы с амплитудой не более 10 вольт. Образуют ли эти сигналы линейное пространство?

Нет, не образуют, так как не выполняется, по крайней мере, вторая аксиома линейного пространства (сумма двух сигналов с амплитудой более 5 вольт превышает 10 вольт). Требуется дополнительные структурные ограничения по параметрам сигналов.

---

Сигналы могут описываться как вещественными, так и комплексными функциями, и линейные пространства также могут быть вещественными или комплексными. Скалярные множества обычно отождествляются с множествами действительных или комплексных чисел, но на них также могут накладываться определенные ограничения. Так, например, в теории связи широко применяется бинарное скалярное множество  $\{0, 1\}$ .

Множество  $L$ , для которого выполняются приведенные выше аксиомы, при анализе сигналов и систем может рассматриваться как специальным образом сконструированное многомерное (в пределе – бесконечномерное) геометрическое пространство. Рассмотрим это на конкретном примере.

Имеем произвольный сигнал  $s(t)$ , заданный на интервале  $[a, b]$ .

Дискретизируем сигнал с равномерным шагом дискретизации и переведем в цифровую форму (представим сигнал  $N$  последовательными выборками):

$$s = (s_1, s_2, \dots, s_N).$$

В таком отображении величина  $s$  может рассматриваться в виде  $N$ -мерного вектора в  $N$ -мерном пространстве, в котором значения  $s_n$  представляют собой проекции  $s$ -вектора на координатные оси данного пространства. Двумерный вектор в двумерном пространстве – это точка с координатами  $s_1$  и  $s_2$  на рис. 17.22. Соответственно, в трехмерном пространстве сигнал  $s$  представлен точкой в трехмерном пространстве. Представить себе  $N$ -мерное пространство при  $N > 3$  можно только абстрактно, но с математических позиций такое пространство вполне реально и  $N$ -мерный сигнал  $s$  отображается вполне определенной точкой в этом пространстве с координатами  $s_n$  по осям пространства. При уменьшении интервала дискретизации сигнала до бесконечно малой величины значение  $N$  стремится к бесконечности, и пространство сигналов превращается в бесконечномерное пространство аналоговых сигналов. Следовательно, и аналоговые сигналы могут рассматриваться как предельный случай бесконечномерных векторов.

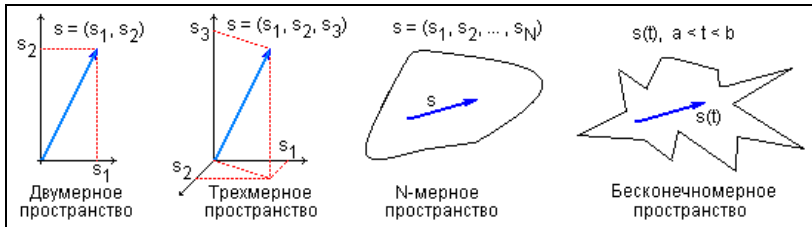


Рис. 17.22. Пространства сигналов и функций.

С учетом вышеизложенного, для математического анализа систем и сигналов в линейном пространстве может использоваться математика векторов.

В линейном пространстве  $L\{u_k; k=0,1,2,\dots,K\}$  всегда можно выделить множество векторов  $\{x_k; k=0,1,2,\dots,K\}$ , для которых выполняется равенство нулю их *линейной комбинации*

$$\sum_{k=0}^K \alpha_k x_k = 0 \tag{17.17}$$

только при условии равенства нулю всех значений  $\alpha_k$ . Такое множество векторов называется *линейно независимым*. Ни один вектор этого линейно независимого множества не может быть выражен в виде какой-либо линейной комбинации других векторов этого пространства.



Такое множество векторов называется *базисом* К-мерного пространства  $L\{u_k; K\}$ . Линейные пространства сигналов имеют, как правило, не единственный базис. Выбор базиса определяется простотой и удобством его использования при обработке сигналов.

**Пример.** Имеем множество сигналов в виде числовых последовательностей, каждая из которых состоит из N чисел (N-мерные вектор-строки). Для сигналов задано скалярное пространство чисел  $R = \{\alpha, 0 \leq \alpha \leq 10\}$ . При этом пространство сигналов N-мерно и может быть определено линейной комбинацией:

$$L = \{y; y = \sum_{n=0}^N \alpha_n x_n, 0 \leq \alpha \leq 10, x_n - \text{базис пространства}\}.$$

$$x_0 = \{1, 0, 0, 0, \dots, 0\},$$

$$x_1 = \{0, 1, 0, 0, \dots, 0\},$$

$$x_2 = \{0, 0, 1, 0, \dots, 0\},$$

$$\dots$$

$$x_N = \{0, 0, 0, 0, \dots, 1\},$$

Любой сигнал в этом пространстве определен точкой с N - координатами в базисе  $x_n$ .

---

Основными метрическими параметрами линейного пространства являются норма, метрика и скалярное произведение сигналов.

**Норма сигналов** в линейном пространстве является аналогом *длины векторов*, и обозначается индексом  $\|s(t)\|$  - *норма* (norm). В математике существуют различные формы норм. При анализе сигналов обычно используются квадратичные нормы:

$$\|s(t)\| = \sqrt{\int_{-\infty}^{\infty} s^2(t) dt} . \quad (17.18)$$

Соответственно, для дискретных сигналов:

$$\|s(n)\| = \sqrt{\sum_{n=-\infty}^{\infty} s^2(n)} . \quad (17.19)$$

Для комплексных сигналов:

$$\|s(t)\| = \sqrt{\int_{-\infty}^{\infty} s(t)s^*(t) dt} , \quad (17.20)$$

где  $s^*(t)$  – величины, комплексно сопряженные с  $s(t)$ .

Линейное пространство сигналов  $L$  является нормированным, если каждому сигналу пространства  $s(t)$  однозначно сопоставлена его числовая норма  $\|s(t)\|$ , и выполняются следующие аксиомы:

1. Норма неотрицательна ( $\|s(t)\| \geq 0$ ) и равна нулю тогда и только тогда, когда сигнал равен нулю ( $\|s(t)\| = 0$ , при  $s(t) = \emptyset$ ).
2. Для любого числа  $b$  должно быть справедливо равенство:  $\|bs(t)\| = |b| \cdot \|s(t)\|$ .
3. Если  $v(t)$  и  $u(t)$  – сигналы из пространства  $L$ , то должно выполняться неравенство треугольника:  $\|v(t)+u(t)\| \leq \|v(t)\| + \|u(t)\|$ .

Пример норм для двумерных цифровых сигналов приведен на рис. 17.23.

**Метрика сигналов.** Линейное пространство сигналов  $L$  является метрическим, если каждой паре сигналов  $s(t) \in L$  и  $v(t) \in L$  однозначно сопоставляется неотрицательное число  $\rho(s(t), v(t))$  – **метрика** (metric) или **расстояние между векторами**. Пример метрики для двух векторов в двумерном пространстве приведен на рис. 17.23.

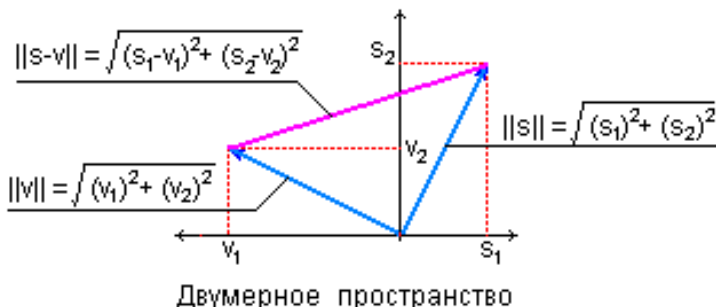


Рис. 17.23. Норма и метрика сигналов.

Для метрик сигналов в метрическом пространстве любой размерности должны выполняться аксиомы:

1.  $\rho(s(t), v(t)) = \rho(v(t), s(t))$  – рефлексивность метрики.
2.  $\rho(s(t), s(t)) = 0$  для любых  $s(t) \in L$ .
3.  $\rho(s(t), v(t)) \leq \rho(s(t), a) + \rho(a, v(t))$  для любых  $a \in L$ .

Метрика определяется нормой разности двух сигналов (см. рис. 17.23):

$$\rho(s(t), v(t)) = \|s(t) - v(t)\|. \quad (17.21)$$

В свою очередь норму можно отождествлять с расстоянием от выбранного элемента пространства до нулевого:  $\|s(t)\| = \rho(s(t), \emptyset)$ .

По метрике сигналов можно судить, например, о том, насколько точно один сигнал может быть аппроксимирован другим

сигналом, или насколько изменяется выходной сигнал относительно входного при прохождении через какое-либо устройство.

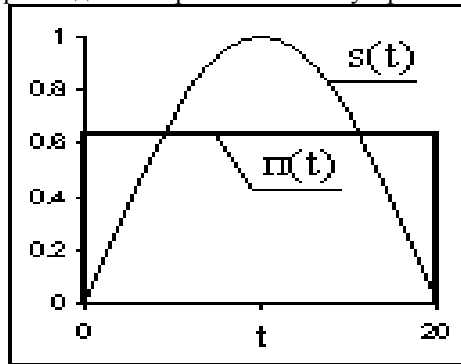


Рис. 17.24.

**Пример.** Сигнал на интервале  $(0, T)$  представляет собой половину периода синусоиды амплитудой  $A$ :  $s(t) = A \cdot \sin(\pi t/T)$ ,  $0 \leq t \leq T$ . Требуется аппроксимировать сигнал прямоугольным импульсом  $n(t)$  (см. рис. 17.24).

Если принять амплитуду импульса  $n(t)$  равной  $B$ , то квадрат расстояния между сигналами:  $\rho^2(s, n) = \int_0^T (A \sin(\pi t/T) - B)^2 dt = A^2 T/2 - 4ABT/\pi + B^2 T$ .

Для решения задачи требуется найти минимум выражения  $\rho^2(s, n)$ . Дифференцируем полученное выражение по  $B$ , приравняем нулю и, решая относительно  $B$ , находим значение экстремума:  $B = 2A/\pi \approx 0.64A$ . Это искомое значение минимума функции  $\rho^2(s, n)$  (вторая производная функции по  $B$  положительна).

При этом минимальное значение метрики:  $\rho_{\min} \approx 0.31A \sqrt{T}$ . Вычислим нормы сигналов при  $A = 1$ :

$$E_s = A^2 \int_0^T \sin^2(\pi t/T) dt = A^2 T/2 = 10.$$

Норма:  $\|s(t)\| = \sqrt{E_s} = 0.707 A \sqrt{T} \approx 3.16.$

$$E_n = B^2 \int_0^T dt = B^2 T \approx 8.1.$$

Норма:  $\|n(t)\| = \sqrt{E_n} = B \sqrt{T} \approx 2.85.$

---

**Скалярное произведение произвольных сигналов**  $u(t)$  и  $v(t)$  отражает степень их связи (сходства) по форме и положению в

пространстве сигналов, и обозначается как  $\langle u(t), v(t) \rangle$ .

$$\langle u(t), v(t) \rangle = \|u(t)\| \cdot \|v(t)\| \cos \varphi, \quad (17.22)$$

Физическую сущность скалярного произведения векторов в двумерном пространстве можно видеть достаточно наглядно (рис. 17.25). Это произведение "длины" (нормы) одного вектора на проекцию второго вектора по "направлению" первого вектора.

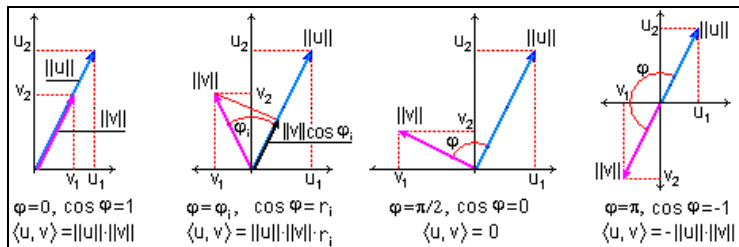


Рис. 17.25. Скалярное произведение сигналов в двумерном пространстве.

При кажущейся абстрактности скалярного произведения сигналов оно может приобретать вполне конкретный физический смысл для конкретных физических процессов, которые отображаются этими сигналами. Так, например, если  $v = F$  – сила, приложенная к телу, а  $u = s$  – перемещение тела под действием этой силы, то скалярное произведение  $W = F \cdot s$  определяет выполненную работу, при условии совпадения силы с направлением перемещения. В противном случае, при наличии угла  $\varphi$  между векторами силы и перемещения, работа будет определяться проекцией силы в направлении перемещения, т.е.  $W = s \cdot F \cdot \cos \varphi$ . Вычисление скалярного произведения обычно производится непосредственно по сигнальным функциям. Поясним это на примере двумерных сигналов с использованием рисунка 17.23. Для квадрата метрики сигналов  $s$  и  $v$  имеем:

$$\|s-v\|^2 = \|s\|^2 + \|v\|^2 - 2 \|s\| \|v\| \cos \varphi = \|s\|^2 + \|v\|^2 - 2 \langle s, v \rangle.$$

$$2 \langle s, v \rangle = \|s\|^2 + \|v\|^2 - \|s-v\|^2 = (s_1^2 + s_2^2) + (v_1^2 + v_2^2) - \{(s_1 - v_1)^2 + (s_2 - v_2)^2\} = 2 (s_1 v_1 + s_2 v_2).$$

$$\langle s, v \rangle = s_1 v_1 + s_2 v_2.$$

Обобщая полученное выражение на аналоговые сигналы:

$$\langle s(t), v(t) \rangle = \int_{-\infty}^{\infty} s(t)v(t) dt. \quad (17.23)$$

Соответственно, для дискретных сигналов в  $N$ -мерном пространстве:

$$\langle s_n, v_n \rangle = \sum_{n=1}^N s_n v_n. \quad (17.24)$$

Скалярное произведение обладает следующими свойствами:

1.  $\langle s, v \rangle \geq 0$ ;
2.  $\langle s, v \rangle = \langle v, s \rangle$ ;
3.  $\langle as, v \rangle = a \langle s, v \rangle$ , где  $a$  – вещественное число;
4.  $\langle s+v, a \rangle = \langle s, a \rangle + \langle v, a \rangle$ .

Линейное пространство *аналоговых сигналов* с таким скалярным произведением называется гильбертовым пространством  $H$  (второе распространенное обозначение -  $L^2$ ). Линейное пространство *дискретных и цифровых сигналов* - пространством Евклида (обозначение пространства -  $R^2$ ). В этих пространствах справедливо фундаментальное неравенство Коши-Буняковского (модуль косинуса в (17.22) может быть только равным или меньше 1):

$$|\langle s, v \rangle| \leq \|s\| \cdot \|v\|. \quad (17.25)$$

Для комплексного гильбертова пространства скалярное произведение вычисляется по формуле:

$$\langle s(t), v(t) \rangle = \int_{-\infty}^{\infty} s(t)v^*(t) dt. \quad (17.26)$$

При определении функций в пространстве  $L^2[a, b]$  вычисление скалярного произведения производится соответственно с пределами интегрирования от  $a$  до  $b$ .

Из выражения (17.22) следует косинус угла между сигналами:

$$\cos \varphi = \langle s(t), v(t) \rangle / (\|s\| \cdot \|v\|). \quad (17.27)$$

---

**Пример.** Имеется два смещенных во времени прямоугольных импульса с одинаковой амплитудой и длительностью:  $s_1(t) = 2$  при  $0 \leq t \leq 5$ ,  $s_1(t) = 0$  при других  $t$ ; и  $s_2(t) = 2$  при  $4 \leq t \leq 9$ ,  $s_2(t) = 0$  при других  $t$ .

Квадраты норм сигналов:

$$\|s_1\|^2 = \int_0^5 s_1^2(t) dt = 20. \quad \|s_2\|^2 = \int_4^9 s_2^2(t) dt = 20$$

$$\text{Скалярное произведение: } \langle s_1, s_2 \rangle = \int_4^5 s_1(t) s_2(t) dt = 8.$$

Отсюда имеем:

$$\cos \varphi = \langle s_1, s_2 \rangle / (\|s_1\| \cdot \|s_2\|) = 8/20 = 0.4 \quad \text{и } \varphi \approx 1.16 \text{ радиан } \approx 66^\circ$$

При полном совмещении сигналов:

$$\langle s_1, s_2 \rangle = \int_t^{t+5} s_1(t) s_2(t) dt = 20, \quad \cos \varphi = 1, \quad \varphi = 0.$$

При отсутствии перекрытия сигналов:

$$\langle s_1, s_2 \rangle = 0, \quad \cos \varphi = 0, \quad \varphi = 90^\circ.$$

---

Физическое понятие "угла" между многомерными сигналами довольно абстрактно. Однако при рассмотрении выражения (17.27) совместно с выражением для квадрата метрики сигналов

$$\rho^2(s,v) = \int_{-\infty}^{\infty} [s(t)-v(t)]^2 dt = \|s\|^2 + \|v\|^2 - 2\cdot\|s\|\cdot\|v\| \cos \varphi.$$

можно отметить следующие закономерности. При  $\varphi = 0$  ( $\cos \varphi = 1$ ) сигналы "совпадают по направлению" и расстояние между ними минимально. При  $\varphi = \pi/2$  ( $\cos \varphi = 0$ ) сигналы "перпендикулярны друг другу" (иначе говоря – ортогональны), и проекции сигналов друг на друга равны 0. При  $\varphi = \pi$  ( $\cos \varphi = -1$ ) сигналы "противоположны по направлению" и расстояние между сигналами максимально. Фактор расстояния между сигналами играет существенную роль при их селекции в многоканальных системах.

**Коэффициент корреляции сигналов.** Одновременно заметим, что значение косинуса в (17.27) изменяется от 1 до -1, и не зависит от нормы сигналов ("длины" векторов). Максимальное значение  $\cos\varphi=1$  соответствует полной тождественности относительной динамики сигналов, минимальное значение  $\cos \varphi = -1$  наблюдается при полной противоположности значений относительной динамики сигналов. По существу, коэффициент  $r = \cos \varphi$  является интегральным коэффициентом степени сходства формы сигналов по пространству их задания. С учетом этого он и получил название ***коэффициента корреляции сигналов***. На рис. 17.26 можно наглядно видеть значения коэффициента корреляции двух сигналов в зависимости от их формы и сдвига по независимой переменной.

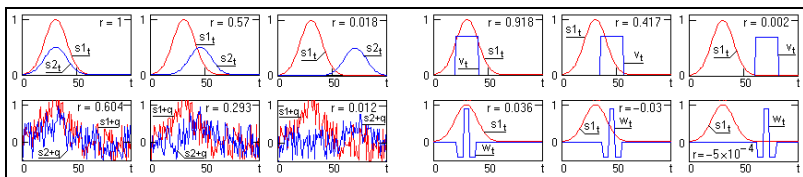


Рис. 17.26. Коэффициент корреляции сигналов.

Однако количественные значения коэффициентов корреляции существенно зависят от выбора нулевой точки сигнального пространства. Рассмотрим это более детально на конкретном примере.

На рис. 17.27 приведено изменение средней месячной температуры воздуха в трех городах земного шара в течение одного календарного года.

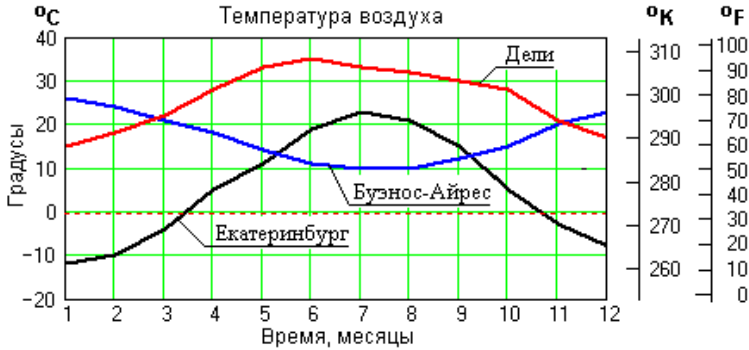


Рис. 17.27.

Характер корреляции между изменениями температур в городах достаточно хорошо виден на графиках. Вычислим (см. пример ниже) значения коэффициентов корреляции для шкалы температур по Цельсию.

**Пример.** Среднемесячная температура воздуха в городах по Цельсию:

Екатеринбург:  $E_k = \{-12, -10, -4, 5, 11, 19, 23, 21, 15, 5, -3, -8\}$ .

Дели:  $D_k = \{15, 18, 22, 28, 33, 35, 33, 32, 30, 28, 21, 17\}$ .

Буэнос-Айрес:  $B_k = \{26, 24, 21, 18, 14, 11, 10, 10, 12, 15, 20, 23\}$ .

Нумерация месяцев:  $k = 1, 2, 3, \dots, 12$ .

Норма сигналов:

$$\|E\| = \sqrt{\sum_k (E_k)^2} = 45.39, \quad \|D\| = \sqrt{\sum_k (D_k)^2} = 93.05,$$

$$\|B\| = \sqrt{\sum_k (B_k)^2} = 61.9.$$

Скалярные произведения:

$$\langle E, D \rangle = \sum_k (E_k D_k) = 2542, \quad \langle E, B \rangle = 268, \quad \langle B, D \rangle = 4876.$$

Коэффициенты корреляции:

Екатеринбург – Дели:

$$r_{ED} = \langle E, D \rangle / (\|E\| \|D\|) = 0.602.$$

Екатеринбург – Буэнос-Айрес:

$$r_{EB} = 0.095,$$

Дели – Буэнос-Айрес:

$$r_{DB} = 0.847,$$

Как следует из вычислений, полученные коэффициенты

корреляции маловыразительны. Практически не регистрируется разнонаправленная корреляция Екатеринбург - Буэнос-Айрес, и не различаются одно- (Екатеринбург – Дели) и разнонаправленные (Дели – Буэнос-Айрес) типы корреляции.

Повторим вычисления в шкале Фаренгейта ( $0^{\circ}\text{F} = -17,8^{\circ}\text{C}$ ,  $100^{\circ}\text{F} = +37,8^{\circ}\text{C}$ ), и в абсолютной шкале температур Кельвина. Дополнительно вычислим значения коэффициентов корреляции в шкале Цельсия и Фаренгейта для центрированных сигналов. Центрированный сигнал вычисляется путем определения среднего значения сигнала по интервалу его задания и вычитания этого среднего значения из исходных значений сигнала, т.е. среднее значение центрированного сигнала равно нулю. Сводные результаты вычислений приведены в таблице.

Таблица 17.1.

Коэффициенты корреляции сигналов

Пары городов	Нецентрированные сигналы			Центрированные сигналы	
	Цельсий	Фаренгейт	Кельвин	Цельсий	Фаренгейт
Екатеринбург – Дели					
Екатеринбург – Буэнос-Айрес	0.602	0.943	1	0.954	0.954
Дели – Буэнос-Айрес	0.095	0.803	0.998	-0.988	-0.988
Дели – Буэнос-Айрес	0.847	0.953	0.999	-0.960	-0.960

Как видно из таблицы, значения коэффициента корреляции нецентрированных сигналов существенно зависят от положения сигналов относительно нулевой точки пространства. При одностороннем смещении сигналов относительно нуля (шкала Фаренгейта) значение коэффициента корреляции может быть только положительным, и тем ближе к 1, чем дальше от сигналов нулевая точка (шкала Кельвина), т.к. при больших значениях сигналов-векторов значение скалярного произведения сигналов стремится к значению произведения норм сигналов.

Для получения значений коэффициентов корреляции, независимых от нуля сигнального пространства и от масштаба единиц измерений, необходимо вычислять коэффициент по центрированным сигналам, при этом в оценках коэффициента, как это видно из результатов, приведенных в таблице, появляется знаковый параметр



совпадения (или несовпадения) по "направлению" корреляции и исчезает зависимость от масштаба представления сигналов. Это позволяет вычислять коэффициенты корреляции различных сигналов вне зависимости от физической природы сигналов и их величины.

**Координатный базис пространства.** Для измерения и отображения одномерных величин достаточно одного нормированного параметра – стандарта величины или единицы ее измерения (для измерения длины – сантиметр, для измерения тока – ампер, и т.п.).

В пространстве сигналов роль такого метрологического стандарта выполняет координатный базис пространства – подмножество векторов  $\{e_1, e_2, e_3, \dots\}$  со свойствами ортогональных координатных осей с определенной единицей измерений, по которым можно разложить любой произвольный сигнал, принадлежащий этому линейному пространству.

Число базисных векторов определяет размерность векторного пространства. Так, для двумерных векторов в качестве ортогонального базиса пространства могут быть приняты векторы  $\{v_1, v_2\}$ , если выполняется условие их взаимной перпендикулярности – нулевое значение скалярного произведения  $\langle v_1, v_2 \rangle = 0$ . При  $\|v_1\| = \|v_2\| = 1$  эта пара векторов является *ортонормированным* базисом с *единичными векторами* координатных осей в качестве стандарта (единицы измерения) пространства.

---

**Пример.** Могут ли быть приняты в качестве координатного базиса двумерного пространства векторы

$$v_1 = (\sqrt{3}/2, 1/2), \quad v_2 = (-1/2, \sqrt{3}/2).$$

$$\langle v_1, v_2 \rangle = (\sqrt{3}/2) \cdot (-1/2) +$$

$$(1/2) \cdot (\sqrt{3}/2) = 0.$$

Векторы ортогональны.

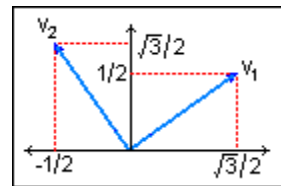
$$\|v_1\| = \sqrt{\left(\frac{\sqrt{3}}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 1.$$

$$\|v_2\| = \sqrt{\left(-\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2} = 1.$$

Векторы нормированы.

Векторы могут быть ортонормированным базисом пространства.

---



Разложение произвольного двумерного вектора - сигнала  $s$  в двумерном пространстве, по координатным осям  $v_1$  и  $v_2$  элементарно:

$$s = c_1 v_1 + c_2 v_2, \quad (17.28)$$

где коэффициенты  $c_1$  и  $c_2$  выражают значения составляющих вектора  $s$  по направлениям векторов  $v_1$  и  $v_2$ , т.е. являются проекциями вектора  $s$  на координатный базис пространства  $\{v_1, v_2\}$ . Значения проекций определяются скалярными произведениями:

$$c_1 = \langle s, v_1 \rangle, \quad c_2 = \langle s, v_2 \rangle.$$

В этом нетрудно убедиться, если вычислить скалярные произведения левой и правой части выражения (17.28) сначала с вектором  $v_1$ :

$$\langle s, v_1 \rangle = \langle (c_1 v_1 + c_2 v_2), v_1 \rangle = \langle c_1 v_1, v_1 \rangle + \langle c_2 v_2, v_1 \rangle = c_1 \langle v_1, v_1 \rangle + c_2 \langle v_2, v_1 \rangle.$$

При ортонормированности базиса  $\{v_1, v_2\}$  имеем:

$$\langle v_1, v_1 \rangle = \|v_1\|^2 = 1, \quad \langle v_2, v_1 \rangle = 0.$$

Отсюда следует:  $\langle s, v_1 \rangle = c_1$ . Аналогичным образом можно получить и выражение для значения  $c_2 = \langle s, v_2 \rangle$ .

---

**Пример.** Разложить вектор  $s = (\sqrt{3}/2, 5/2)$  по базису, представленному векторами

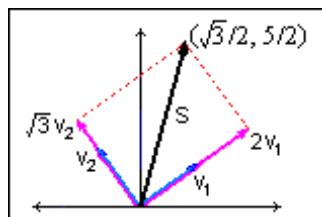
$$v_1 = (\sqrt{3}/2, 1/2) \text{ и } v_2 = (-1/2, \sqrt{3}/2)$$

из предыдущего примера.

$$s = c_1 v_1 + c_2 v_2.$$

$$c_1 = \langle s, v_1 \rangle = (\sqrt{3}/2) \cdot (\sqrt{3}/2) + (5/2) \cdot (1/2) = 2.$$

$$c_2 = \langle s, v_2 \rangle = (\sqrt{3}/2) \cdot (-1/2) + (5/2) \cdot (\sqrt{3}/2) = \sqrt{3}.$$



**Результат:** В пространстве с базисом  $\{v_1, v_2\}$  вектор  $s$  однозначно определяется двумя векторами  $v_1$  и  $v_2$ :

$$s = 2v_1 + \sqrt{3} v_2.$$


---

Множество векторов  $\{v_k, k = 1, 2, \dots, N\}$  может быть принято в качестве ортонормированного координатного базиса  $N$ -мерного пространства, если их совокупность является линейно независимой, равенство  $\sum_i a_i v_i = \emptyset$  выполняется только в случае одновременного

обращения в нуль всех числовых коэффициентов  $a_i$ , и для всех векторов этого множества при единичной норме выполняется условие взаимной ортогональности:

$$\langle v_m, v_n \rangle = \begin{cases} 0; & m \neq n \\ 1; & m = n \end{cases} \quad (17.29)$$

Выражение (17.29) обычно записывается в следующей форме:

$$\langle v_m, v_n \rangle = \delta_{mn},$$

где  $\delta_{mn}$  – импульс (символ) Кронекера.

С использованием ортонормированного базиса любой произвольный сигнал можно представить в виде линейной комбинации взвешенных базисных векторов:

$$s = c_1 v_1 + c_2 v_2 + \dots + c_N v_N = \sum_i c_i v_i,$$

где весовое значение  $c_k$  представляет собой проекцию вектора  $s$  на соответствующее координатное направление и определяется скалярным произведением:

$$c_k = \langle s, v_k \rangle.$$

Коэффициенты  $c_k$  называют коэффициентами Фурье в базисе  $\{v_k\}$ . Базисную систему  $\{v_k\}$  называют полной, если ее размерность (и размерность соответствующего пространства) равна размерности представляемых в этой системе сигналов.

Комплексное линейное пространство, векторам которого также может быть поставлено в соответствие комплексное число скалярного произведения  $\langle s, v_k \rangle$ , называют **унитарным**. Для него действительны все свойства скалярного произведения с учетом сопряжения:

1.  $\langle s, v \rangle = \langle v, s \rangle^*$ ;
2.  $\langle s, av \rangle = \langle av, s \rangle^* = a^* \langle s, v \rangle$ , где  $a$  – комплексное число.

## **17.6. Мощность и теория сигналов**

**Понятия мощности и энергии** в теории сигналов не относятся к характеристикам каких-либо физических величин сигналов, а являются их количественными характеристиками, отражающими определенные свойства сигналов и динамику изменения их значений (отсчетов) во времени, в пространстве или по любым другим аргументам.

Для произвольного, в общем случае комплексного, сигнала  $s(t) = a(t) + jb(t)$ , где  $a(t)$  и  $b(t)$  – вещественные функции, мгновенная мощность (instantaneous power) сигнала *по определению* задается выражением:

$$w(t) = s(t) s^*(t) = [a(t) + jb(t)] [a(t) - jb(t)] = a^2(t) + b^2(t) = |s(t)|^2, \quad (17.30)$$

т.е. функция распределения мгновенной мощности по аргументу сигнала равна квадрату функции его модуля, для вещественных сигналов – квадрату функции амплитуд.

Аналогично для дискретных сигналов:

$$w_n = s_n s_n^* = [a_n + jb_n] [a_n - jb_n] = a_n^2 + b_n^2 = |s_n|^2, \quad (17.31)$$

Энергия сигнала (также по определению) равна интегралу от мощности по всему интервалу существования или задания сигнала. В

пределе:

$$E_s = \int_{-\infty}^{\infty} w(t) dt = \int_{-\infty}^{\infty} |s(t)|^2 dt. \quad (17.32)$$

$$E_s = \sum_{n=-\infty}^{\infty} w_n = \sum_{n=-\infty}^{\infty} |s_n|^2. \quad (17.33)$$

Мгновенная мощность  $w(t)$  является плотностью мощности сигнала, так как измерения мощности возможны только через энергию на интервалах ненулевой длины:

$$w(\tau) = (1/\Delta t) \int_{\tau - \Delta t/2}^{\tau + \Delta t/2} |s(t)|^2 dt.$$

Энергия сигналов может быть конечной или бесконечной. Конечную энергию имеют финитные сигналы и сигналы, затухающие по своим значениям в пределах конечной длительности, которые не содержат дельта-функций и особых точек (разрывов второго рода и ветвей, уходящих в бесконечность). В противном случае их энергия равна бесконечности. Бесконечна также энергия периодических сигналов.

Как правило, сигналы изучаются на определенном интервале  $T$ , для периодических сигналов - в пределах одного периода  $T$ , при этом средняя мощность (average power) сигнала:

$$W_T(\tau) = (1/T) \int_{\tau}^{\tau+T} w(t) dt = (1/T) \int_{\tau}^{\tau+T} |s(t)|^2 dt. \quad (17.34)$$

Понятие средней мощности может быть распространено и на незатухающие сигналы, энергия которых бесконечно велика. В случае неограниченного интервала  $T$  строго корректное определение средней мощности сигнала должно производиться по формуле:

$$W_s = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T w(t) dt. \quad (17.35)$$

Квадратный корень из значения средней мощности характеризует *действующее* (среднеквадратическое) значение сигнала (root mean square, RMS).

Применительно к электрофизическим системам, данным понятиям мощности и энергии соответствуют вполне конкретные физические величины. Допустим, что функцией  $s(t)$  отображается электрическое напряжение на резисторе, сопротивление которого равно  $R$  Ом. Тогда рассеиваемая в резисторе мощность, как известно, равна (в вольт-амперах):

$$w(t) = |s(t)|^2/R,$$

а полная выделенная на резисторе тепловая энергия определяется соответствующим интегрированием мгновенной мощности  $w(t)$  по интервалу задания напряжения  $s(t)$  на резисторе  $R$ . Физическая размерность мощности и энергии в этом случае определяется соответствующей физической размерностью функции напряжения  $s(t)$  и сопротивления резистора  $R$ . Для безразмерной величины  $s(t)$  при  $R=1$  это полностью соответствует выражению (17.30). В теории сигналов в общем случае сигнальные функции  $s(t)$  не имеют физической размерности, и могут быть формализованным отображением любого процесса или распределения какой-либо физической величины, при этом понятия энергии и мощности сигналов используются в более широком смысле, чем в физике. Они представляют собой специфические метрологические характеристики сигналов.

Из сравнения выражений (17.18) и (17.32) следует, что энергия и норма сигнала связаны соотношениями:

$$E_s = \|s(t)\|^2, \quad \|s(t)\| = \sqrt{E_s} \quad (17.36)$$

---

**Пример.** Цифровой сигнал задан функцией

$$s(n) = \{0, 1, 2, 3, 4, 5, 4, 3, 2, 1, 0, 0, 0, \dots\}.$$

Энергия сигнала:

$$E_s = \sum_{n=0}^N s^2(n) = 1+4+9+16+25+16+9+4+1 = 85.$$

Норма:

$$\|s(n)\| = \sqrt{85} \approx 9.22$$

---

Вычислим энергию суммы двух произвольных сигналов  $u(t)$  и  $v(t)$ :

$$E = \int_{-\infty}^{\infty} [u(t)+v(t)]^2 dt = E_u + E_v + 2 \int_{-\infty}^{\infty} u(t)v(t) dt. \quad (17.37)$$

Как следует из этого выражения, энергии сигналов (а равно и их мощности), в отличие от самих сигналов, в общем случае не обладают свойством аддитивности. Энергия суммарного сигнала  $u(t)+v(t)$ , кроме суммы энергий составляющих сигналов, содержит в себе и так называемую энергию взаимодействия сигналов или взаимную энергию:

$$E_{uv} = 2 \int_{-\infty}^{\infty} u(t)v(t) dt. \quad (17.38)$$

Нетрудно заметить, что энергия взаимодействия сигналов равна их удвоенному скалярному произведению:

$$E_{uv} = 2 \langle u(t), v(t) \rangle. \quad (17.29)$$

При обработке данных используются также понятия мощности взаимодействия двух сигналов  $x(t)$  и  $y(t)$ :

$$\begin{aligned} w_{xy}(t) &= x(t) y^*(t), \\ w_{yx}(t) &= y(t) x^*(t), \\ w_{xy}(t) &= w_{yx}^*(t). \end{aligned} \quad (17.40)$$

Для вещественных сигналов:

$$w_{xy}(t) = w_{yx}(t) = x(t) y(t). \quad (17.41)$$

С использованием выражений (17.40-17.41) интегрированием по соответствующим интервалам вычисляются значения средней мощности взаимодействия сигналов на определенных интервалах  $T$  и энергия взаимодействия сигналов.

## 17.7. Пространства функций

Пространства функций можно считать обобщением пространства  $N$ -мерных сигналов – векторов на аналоговые сигналы, как бесконечномерные векторы, с некоторыми чисто практическими уточнениями.

**Нормирование метрических параметров.** Норма функций в пространстве  $L^2[a, b]$  определяется выражением:

$$\|s(t)\| = \sqrt{\int_a^b s^2(t) dt}.$$

Нетрудно заключить, что чем больше интервал  $[a, b]$  в этой формуле, тем больше (при прочих равных условиях) будет значение нормы. При анализе и сравнении сигналов (как аналоговых, так и многомерных дискретных) такое понятие не всегда удобно, и вместо него очень часто используют понятие нормы, нормированной относительно длины интервала  $[a, b]$ . Для символического обозначения нормирования будем применять знак  $\nabla$ :

$$\|s(t)\|^\nabla = \sqrt{\frac{1}{b-a} \int_a^b s^2(t) dt}, \quad \|s_n\|^\nabla = \sqrt{\frac{1}{N} \sum_{n=1}^N S_n}.$$

Метрика сигналов (расстояние между сигналами) при

аналогичном нормировании:

$$\rho^\nabla(s(t), v(t)) = \sqrt{\frac{1}{b-a} \int_a^b [s(t) - v(t)]^2 dt}, \quad \rho^\nabla(s_n, v_n) =$$

$$= \sqrt{\frac{1}{N} \sum_{n=1}^N (s_n - v_n)^2}$$

Эти выражения применяются для вычисления среднеквадратического расхождения сигналов или среднеквадратического погрешности (стандартный индекс погрешности в абсолютных единицах измерений -  $\sigma$ ) выполнения какой-либо операции при сравнении ее результата с теоретически ожидаемым или априорно известным.

Нормированное скалярное произведение сигналов:

$$\langle s(t), v(t) \rangle^\nabla = \frac{1}{b-a} \int_a^b s(t)v(t) dt = \|s(t)\|^\nabla \|v(t)\|^\nabla \cos \varphi.$$

$$\langle s_n, v_n \rangle^\nabla = (1/N) \sum_{n=1}^N s_n v_n = \|s_n\|^\nabla \|v_n\|^\nabla \cos \varphi.$$

Косинус угла (коэффициент корреляции) между сигналами (функциями) не изменяет своих значений при вычислении как по нормированным, так и по ненормированным значениям скалярного произведения и нормы сигналов (значения нормировки в числителе и знаменателе выражения (17.27) сокращаются). Взаимная перпендикулярность функций определяется аналогично взаимной перпендикулярности векторов условием нулевого значения скалярного произведения.

Норма, метрика и скалярное произведение периодических функций обычно нормируются на длину одного периода  $T$ .

**Ортогональные сигналы.** Два сигнала называются ортогональными (orthogonal), если имеют нулевое скалярное произведение:

$$\langle u(t), v(t) \rangle = \int_a^b u(t)v(t) dt = 0.$$

Соответственно, два таких сигнала в своем функциональном пространстве являются взаимно перпендикулярными (угол между сигналами равен  $\varphi = 90^\circ$ ), полностью независимыми друг от друга (некоррелированными,  $\gamma = \cos \varphi = 0$ ), и имеют нулевую энергию взаимодействия ( $E_{uv} = 0$ ).

На рисунке 17.28 приведены примеры взаимно ортогональных

сигналов. Нулевое скалярное произведение двух левых сигналов обеспечивается их формой (равна нулю сумма положительных и отрицательных значений произведения сигналов), а двух правых - взаимным расположением (ненулевые значения сигналов не имеют общих координат).

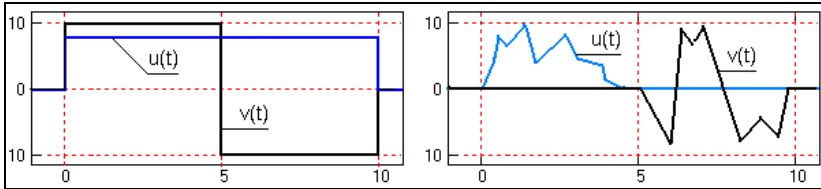


Рис. 17.28 Ортогональные сигналы.

Попутно заметим, что энергия и мощность суммы ортогональных сигналов обладают свойством аддитивности, т.к. имеют нулевое значение скалярного произведения и, соответственно, нулевую энергию взаимодействия.

**Ортонормированный базис пространства.** При распространении положений *векторного базисного пространства* на *функциональное пространство*  $L^2[a, b]$ , в качестве координатного базиса пространства мы должны использовать совокупность функций  $\{u_0(t), u_1(t), u_2(t), \dots\}$ , в пределе - бесконечную, которая должна быть *системой ортогональных функций*  $\{u_k(t), k=0, 1, 2, \dots\}$ , т.е. все функции на этом отрезке должны быть взаимно ортогональны:

$$\langle u_m(t), u_n(t) \rangle = \int_a^b u_m(t) u_n(t) dt = 0, \quad m = 1, 2, \dots; \quad n = 1, 2, \dots; \quad m \neq n.$$

Система ортогональных функций на интервале  $[a, b]$  будет *ортонормированной* (orthonormal functions), если все функции системы при  $m=n$  имеют единичную норму, т.е. выполняются условия:

$$\langle u_m(t), u_m(t) \rangle = \|u_m(t)\|^2 = \int_a^b (u_m(t))^2 dt = 1, \quad \|u_m(t)\| = 1, \quad m = 1, 2, \dots$$

Эти условия можно записать в следующей обобщенной форме:

$$\int_a^b u_m(t) \cdot u_n^*(t) dt = \delta_{m,n}.$$

Система ортогональных функций всегда может быть превращена в ортонормированную путем нормировки, т.е. *деления всех функций на их норму*.

**Разложение сигнала в ряд.** Произвольный сигнал  $s(t) \in H$  (пространство Гильберта), заданный на интервале  $[a, b]$ , может быть разложен в ряд по упорядоченной системе ортонормированных базисных функций  $u_k(t)$ :



$$s(t) = \sum_{k=0}^{\infty} c_k u_k(t). \quad (17.42)$$

Для нахождения значений коэффициентов  $c_k$  умножим обе части данного выражения на базисную функцию  $u_m(t)$  с произвольным номером  $m$  и проинтегрируем результаты по переменной  $t$ , при этом получим:

$$\int_a^b s(t)u_m(t) dt = \sum_{k=0}^{\infty} c_k \int_a^b u_m u_k dt.$$

С учетом ортонормированности функций  $u_i(t)$ , в правой части этого равенства остается только один член суммы с номером  $m = k$  при  $\int_a^b u_k u_k dt = 1$ , который, по левой части уравнения, представляет собой скалярное произведение сигнала и соответствующего  $m = k$  базисного вектора, т.е. проекцию сигнала на соответствующее базисное направление:

$$c_k = \int_a^b s(t)u_k(t) dt. \quad (17.43)$$

Таким образом, в геометрической интерпретации коэффициенты  $c_k$  представляют собой проекции вектор - сигнала  $s(t)$  на соответствующие базисные направления  $u_k(t)$ , т.е. координаты вектора  $s(t)$  по координатному базису, образованному системой ортогональных функций  $u(t)$ , в пределе - бесконечномерной. При практическом использовании количество членов ряда (17.42) ограничивается определенным значением  $N$ , при этом для любого значения  $N$  совокупность коэффициентов  $c_k$  обеспечивают наименьшее по средней квадратической погрешности приближение к заданному сигналу.

Соответственно, энергия взаимодействия двух сигналов  $x(t)$  и  $y(t)$  может вычисляться по скалярному произведению их координатных проекций, которое, с учетом взаимной ортогональности всех проекций, будет равно:

$$\begin{aligned} \langle x(t), y(t) \rangle &= \int_a^b x(t)y(t) dt = \int_a^b \left[ \sum_{n=0}^{\infty} a_n u_n(t) \right] \left[ \sum_{m=0}^{\infty} b_m u_m(t) \right] dt = \\ &= \sum_{n=0}^{\infty} a_n \cdot b_n. \end{aligned} \quad (17.44)$$

Косинус угла между векторами  $x(t)$  и  $y(t)$  с использованием выражения (17.44):

$$\cos \varphi = \sum_{n=0}^{\infty} a_n \cdot b_n / (\|x(t)\| \cdot \|y(t)\|).$$

Возможность разложения непрерывных сигналов и функций в обобщенные ряды по системам ортогональных функций имеет огромное принципиальное значение, так как позволяет вместо изучения несчетного множества точек сигнала ограничиться счетной системой коэффициентов ряда.

К системам базисных функций, которые используются при разложении сигналов, предъявляются следующие основные требования:

- для любого сигнала ряд разложения должен сходиться;
- при ограничении ряда по уровню остаточной погрешности расхождения с заданным сигналом количество членов ряда должно быть минимальным;
- базисные функции должны иметь достаточно простую аналитическую форму;
- коэффициенты разложения в ряд должны вычисляться относительно просто.

Согласно теореме Дирехле, любой сигнал  $s(t)$ , имеющий конечное число точек нарушения непрерывности первого рода, и конечный по энергии на интервале  $[a, b]$ , может быть разложен по системе ортонормальных функций, если существуют интегралы модуля сигнала и модуля его первой производной, т.е.:

$$\int_a^b |s(t)| dt < \infty, \quad \int_a^b |s'(t)| dt < \infty.$$

**Ортонормированные системы функций** хорошо известны в математике. Это полиномы Эрмита, Лежандра, Чебышева, функции Бесселя, Лагерра и целый ряд других. Выбор типа функций в качестве координатного базиса сигнального пространства, как и координатных осей для обычного трехмерного пространства (декартовы, цилиндрические, сферические и пр.), определяется удобством и простотой последующего использования при математической обработке сигналов. При спектральном анализе сигналов используются, в основном, два вида ортонормированных функций: гармонические функции и функции Уолша.

На интервале  $[-\pi, \pi]$  рассмотрим систему следующих гармонических функций:

$$\{1, \sin t, \sin 2t, \dots, \sin kt\}, \quad k = 1, 2, 3, \dots \quad (17.45)$$

Вычислим нормированные на интервал скалярные произведения системы:

$$\langle 1, \sin kt \rangle^\nabla = (1/2\pi) \int_{-\pi}^{\pi} \sin kt \, dt = (1/2k\pi) [\cos k\pi - \cos(-k\pi)] = 0,$$

$$k = 1, 2, 3, \dots$$

$$\langle \sin mt, \sin nt \rangle^\nabla = (1/2\pi) \int_{-\pi}^{\pi} \sin mt \sin nt \, dt = (1/4\pi) \int_{-\pi}^{\pi} \{ \cos (m+n)t - \cos (m-n)t \} \, dt =$$

$$= \frac{1}{4\pi(m+n)} \int_{-\pi}^{\pi} \cos (m+n)t \, dt + \frac{1}{4\pi(m-n)} \int_{-\pi}^{\pi} \cos (m-n)t \, dt = 0,$$

при  $m \neq n$ .

Следовательно, система (17.45) является системой взаимно ортогональных функций. Норма функций:

$$\|\sin kt\|^2 = (1/2\pi) \int_{-\pi}^{\pi} \sin^2 kt \, dt = (1/4\pi) \int_{-\pi}^{\pi} (1 - \cos 2kt) \, dt =$$

$$= \frac{1}{4\pi} \int_{-\pi}^{\pi} 1 \, dt - \frac{1}{8k\pi} \int_{-\pi}^{\pi} \cos 2kt \, dt = 1/2.$$

$$\|\sin kt\|^\nabla = 1/\sqrt{2}, \quad k = 1, 2, 3, \dots$$

Соответственно, для превращения системы (17.45) в ортонормированную следует разделить все функции системы на значение нормы (рис. 17.29):

$$\{1, u_k(t) = \sqrt{2} \sin kt\}, \quad k = 1, 2, 3, \dots \quad (17.46)$$

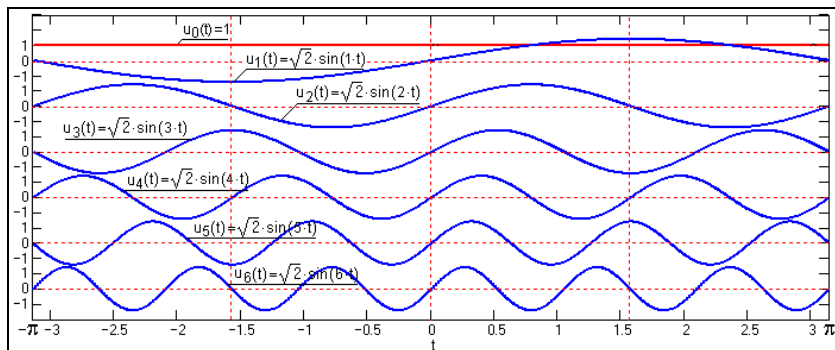


Рис. 17.29. Ортонормированный базис гармонических функций.

Аналогичным образом можно убедиться в ортонормированности косинусной системы гармонических функций:

$$\{1, u_k(t) = \sqrt{2} \cos kt\}, \quad k = 1, 2, 3, \dots, \quad (17.47)$$

и объединенной синус-косинусной системы:

$$\{1, u_k(t) = \sqrt{2} \sin kt, u_k(t) = \sqrt{2} \cos kt\}, \quad k = 1, 2, 3, \dots \quad (17.48)$$

Наибольшее распространение в качестве базисных функций частотного разложения нашли комплексные экспоненциальные функции  $\exp(pt)$  при  $p=jf$  (преобразование Фурье) и  $p=s+jf$  (преобразование Лапласа), от которых с использованием формул Эйлера

$$\begin{aligned} \exp(j\omega t) &= \cos(\omega t) + j \sin(\omega t), & \exp(-j\omega t) &= \cos(\omega t) - j \sin(\omega t), \\ \cos(\omega t) &= [\exp(j\omega t) + \exp(-j\omega t)]/2, & \sin(\omega t) &= [\exp(j\omega t) - \exp(-j\omega t)]/2j \end{aligned}$$

всегда можно перейти к вещественным синус-косинусным функциям. Термин "частотное разложение" обязан своим происхождением независимой переменной частотного представления сигналов, которая измеряется в единицах, обратных единицам времени, т.е. в единицах частоты  $f = 1/|t|$ . Однако понятие частотного преобразования не следует связывать только с временным представлением сигналов, т.к. математический аппарат преобразования не зависит от физического смысла переменных. Так, например, при переменной "x", как единице длины, значение  $f$  будет представлять собой пространственную частоту - число периодических изменений сигнала на единице длины с размерностью  $1/|x|$ .

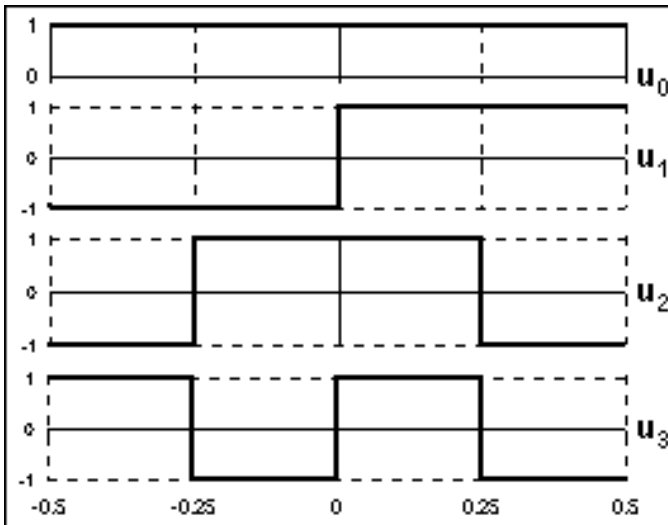


Рис. 17.30. Функции Уолша.

Ортонормированная система функций Уолша, по существу, является предельной модификацией системы периодических функций с кратными частотами, при этом функции принимают значения только

±1. Пример четырех первых функций Уолша на интервале Т от -0,5 до 0,5 приведен на рис. 17.30. Ортогональность и нормированность функций следует из принципа их построения. Стандартное математическое обозначение функций Уолша:  $wal(k,x)$ , где  $k = 0,1,2, \dots$  – порядковый номер функции,  $x = t/T$  – безразмерная координата (нормированная на интервал Т независимая переменная).

Наряду с функциями Уолша применяются также две связанные с ними системы: четные и нечетные функции  $cal(n,x) = wal(2n,x)$ , – аналогичные косинусам, и  $sal(n,x) = wal(2n-1,x)$ , – аналогичные синусам.

При разложении сигналов форма спектров Уолша практически тождественна спектрам гармонических функций.

**Разложение энергии сигнала.** Допустим, что сигнал  $s(t)$  разложен в обобщенный ряд Фурье по гармоническим функциям. Вычислим энергию сигнала непосредственной подстановкой выражения (17.43) в выражение (17.32):

$$\begin{aligned} E_s &= \int_{-\infty}^{\infty} s^2(t) dt = \int_{-\infty}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_m c_n u_m u_n dt = \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} c_m c_n \int_{-\infty}^{\infty} u_m u_n dt. \end{aligned} \quad (17.49)$$

В этом выражении, в силу ортонормированности базисной системы, отличны от нуля только члены с номерами  $m = n$ . Отсюда:

$$E_s = \int_{-\infty}^{\infty} s^2(t) dt = \sum_{n=0}^{\infty} c_n^2, \quad (17.50)$$

т.е. при корректном разложении сигнала в обобщенный ряд Фурье энергия сигнала не изменяется, и равна сумме энергии всех составляющих ряда. Это соотношение называют равенством Парсеваля.

## **17.8. Функции корреляции сигналов**

Функции корреляции сигналов применяются для интегральных количественных оценок формы сигналов и степени их сходства друг с другом.

**Автокорреляционные функции (АКФ) сигналов** (correlation function, CF). Применительно к детерминированным сигналам с конечной энергией АКФ является количественной интегральной характеристикой формы сигнала, и представляет собой интеграл от произведения двух копий сигнала  $s(t)$ , сдвинутых

относительно друг друга на время  $\tau$ :

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t) s(t+\tau) dt. \quad (17.51)$$

Как следует из этого выражения, АКФ является скалярным произведением сигнала и его копии в функциональной зависимости от переменной величины значения сдвига  $\tau$ . Соответственно, АКФ имеет физическую размерность энергии, а при  $\tau=0$  значение АКФ непосредственно равно энергии сигнала:

$$B_s(0) = \int_{-\infty}^{\infty} s(t)^2 dt = E_s.$$

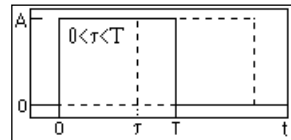
Функция АКФ является непрерывной и четной. В последнем нетрудно убедиться заменой переменной  $t = t-\tau$  в выражении (17.51):

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t-\tau) s(t) dt = \int_{-\infty}^{\infty} s(t) s(t-\tau) dt = B_s(-\tau). \quad (17.52)$$

С учетом четности, графическое представление АКФ обычно производится только для положительных значений  $\tau$ . На практике сигналы обычно задаются на интервале положительных значений аргументов от  $0-T$ . Знак  $+\tau$  в выражении (17.51) означает, что при увеличении значений  $\tau$  копия сигнала  $s(t+\tau)$  сдвигается влево по оси  $t$  и уходит за  $0$ , что требует соответствующего продления сигнала в область отрицательных значений аргумента. А так как при вычислениях интервал задания  $\tau$  обычно много меньше интервала задания сигнала, то более практичным является сдвиг копии сигнала влево по оси аргументов, т.е. применение в выражении (17.51) функции  $s(t-\tau)$  вместо  $s(t+\tau)$ .

По мере увеличения значения величины сдвига  $\tau$  для финитных сигналов временное перекрытие сигнала с его копией уменьшается и скалярное произведение в целом стремятся к нулю.

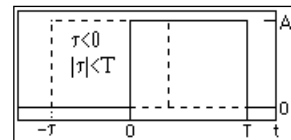
**Пример.** На интервале  $(0,T)$  задан прямоугольный импульс с амплитудным значением, равным  $A$ . Вычислить автокорреляционную функцию импульса.



При сдвиге копии импульса по оси  $t$  вправо, при  $0 \leq \tau \leq T$  сигналы перекрываются на интервале от  $\tau$  до  $T$ . Скалярное произведение:

$$B_s(\tau) = \int_{\tau}^T A^2 dt = A^2(T-\tau).$$

При сдвиге копии импульса



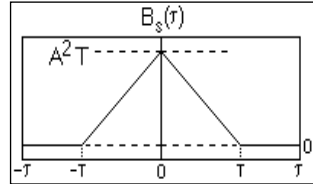
влево, при  $-T \leq \tau < 0$  сигналы перекрываются на интервале от 0 до  $T-\tau$ . Скалярное произведение:

$$B_s(\tau) = \int_0^{T+\tau} A^2 dt = A^2(T+\tau).$$

При  $|\tau| > T$  сигнал и его копия не имеют точек пересечения и скалярное произведение сигналов равно нулю (сигнал и его сдвинутая копия становятся ортогональными).

Обобщая вычисления, можем записать:

$$B_s(\tau) = \begin{cases} A^2(T - |\tau|), & |\tau| \leq T \\ 0, & |\tau| > T \end{cases}.$$



В случае периодических сигналов АКФ вычисляется по одному периоду  $T$ , с усреднением скалярного произведения и его сдвинутой копии в пределах периода:

$$B_s(\tau) = (1/T) \int_0^T s(t) s(t-\tau) dt.$$

При  $\tau=0$  значение АКФ в этом случае равно энергии, а средней мощности сигналов в пределах интервала  $T$ . АКФ периодических сигналов при этом также является периодической функцией с тем же периодом  $T$ . Для однонаправленного гармонического сигнала это очевидно. Первое максимальное значение АКФ будет соответствовать  $\tau=0$ . При сдвиге копии сигнала на четверть периода относительно оригинала подынтегральные функции становятся ортогональными друг другу ( $\cos \omega_0(t-\tau) = \cos(\omega_0 t - \pi/2) \equiv \sin \omega_0 t$ ) и дают нулевое значение АКФ. При сдвиге на  $\tau=T/2$  копия сигнала по направлению становится противоположной самому сигналу и скалярное произведение достигает минимального значения. При дальнейшем увеличении сдвига начинается обратный процесс увеличения значений скалярного произведения с пересечением нуля при  $\tau=3T/2$  и повторением максимального значения при  $\tau=T=2\pi/\omega_0$  ( $\cos \omega_0 t - 2\pi$  копии  $\equiv \cos \omega_0 t$  сигнала). Аналогичный процесс имеет место и для периодических сигналов произвольной формы (рис. 17.31).

**MATNCAD Автокорреляционная функция полигармонического сигнала**

$T := 1$     $\Delta\tau := 0.005$     $\tau := -2T, -2T + \Delta\tau \dots 2T$    <= Параметры расчетов

$$s(t) := - \sum_{n=0}^4 \sin(2\pi \cdot n \cdot t) \quad B(\tau) := \int_{-T}^T s(t) \cdot s(t + \tau) dt \quad * \quad <= \text{Уравнения сигнала и его автокорреляционной функции}$$

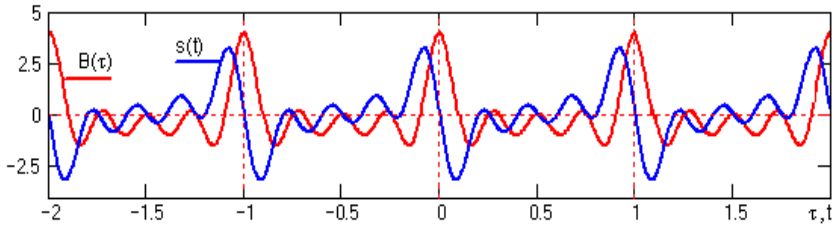


Рис. 17.31.

Отметим, что полученный результат не зависит от начальной фазы гармонического сигнала, что характерно для любых периодических сигналов и является одним из свойств АКФ.

Для сигналов, заданных на определенном интервале  $[a, b]$ , вычисление АКФ производится с нормировкой на длину интервала  $[a, b]$ :

$$B_s(\tau) = \frac{1}{b-a} \int_a^b s(t) s(t+\tau) dt. \quad (17.53)$$

Автокорреляция сигнала может оцениваться и коэффициентом автокорреляции, вычисление которого производится по формуле (по центрированным сигналам):

$$r_s(\tau) = \cos \varphi(\tau) = \langle s(t), s(t+\tau) \rangle / \|s(t)\|^2.$$

**Взаимная корреляционная функция** (ВКФ) сигналов (cross-correlation function, CCF) показывает как степень сходства формы двух сигналов, так и их взаимное расположение друг относительно друга по координате (независимой переменной), для чего используется та же формула (17.51), что и для АКФ, но под интегралом стоит произведение двух разных сигналов, один из которых сдвинут на время  $\tau$ :

$$B_{12}(\tau) = \int_{-\infty}^{\infty} s_1(t) s_2(t+\tau) dt. \quad (17.54)$$

При замене переменной  $t = t - \tau$  в формуле (17.54), получаем:

$$B_{12}(\tau) = \int_{-\infty}^{\infty} s_1(t-\tau) s_2(t) dt = \int_{-\infty}^{\infty} s_2(t) s_1(t-\tau) dt = B_{21}(-\tau)$$



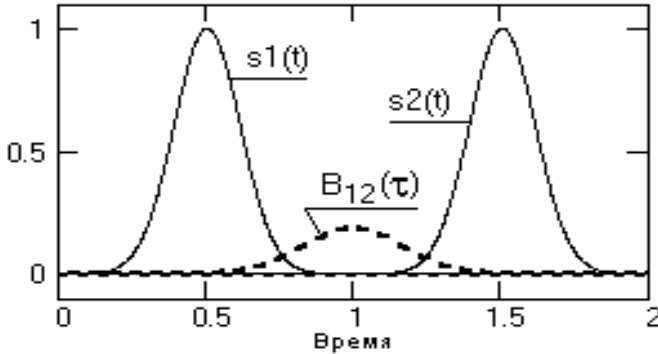


Рис. 17.32. Сигналы и ВКФ.

Отсюда следует, что для ВКФ не выполняется условие четности, а значения ВКФ не обязаны иметь максимум при  $\tau = 0$ . Это можно наглядно видеть на рис. 17.32, где заданы два одинаковых сигнала с центрами на точках 0.5 и 1.5. Вычисление по формуле (17.54) с постепенным увеличением значений  $\tau$  означает последовательные сдвиги сигнала  $s_2(t)$  влево по оси времени (для каждого значения  $s_1(t)$  для подынтегрального умножения берутся значения  $s_2(t+\tau)$ ).

При  $\tau=0$  сигналы ортогональны и значение  $B_{12}(\tau)=0$ . Максимум  $B_{12}(\tau)$  будет наблюдаться при сдвиге сигнала  $s_2(t)$  влево на значение  $\tau=1$ , при котором происходит полное совмещение сигналов  $s_1(t)$  и  $s_2(t+\tau)$ . При вычислении значений  $B_{21}(-\tau)$  аналогичный процесс выполняется последовательным сдвигом сигнала  $s_1(t)$  вправо по временной оси с постепенным увеличением отрицательных значений  $\tau$ , а соответственно значения  $B_{21}(-\tau)$  являются зеркальным (относительно оси  $t=0$ ) отображением значений  $B_{12}(\tau)$ , и наоборот. На рис. 17.33 это можно видеть наглядно.

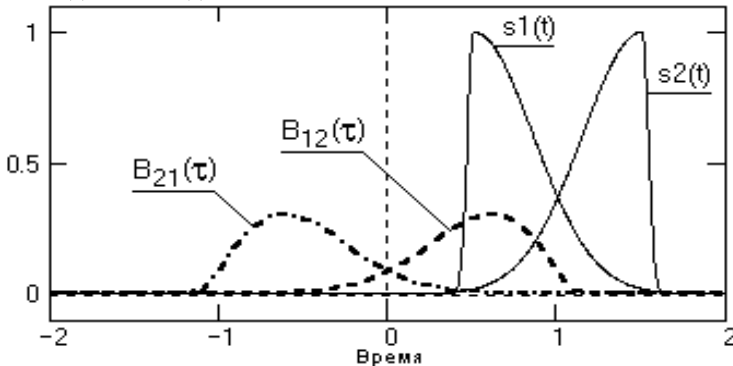


Рис. 17.33. Сигналы и ВКФ.

Таким образом, для вычисления полной формы ВКФ числовая ось  $\tau$  должна включать отрицательные значения, а изменение знака  $\tau$  в формуле (17.54) равносильно перестановке сигналов.

Для периодических сигналов понятие ВКФ обычно не применяется, за исключением сигналов с одинаковым периодом, например, сигналов входа и выхода систем при изучении характеристик систем.

Коэффициент взаимной корреляции двух сигналов вычисляется по формуле (по центрированным сигналам):

$$r_{sv}(\tau) = \cos \varphi(\tau) = \langle s(t), v(t+\tau) \rangle / \|s(t)\| \|v(t)\|. \quad (17.55)$$

Значение коэффициента взаимной корреляции может изменяться от -1 до 1.

## 17.9. Математическое описание шумов и помех

**Шумы и помехи** (noise). При детектировании сигналов в сумме с основным информационным сигналом одновременно регистрируются и мешающие сигналы - шумы и помехи самой различной природы. К помехам относят также искажения информационных сигналов при влиянии различных дестабилизирующих факторов на процессы измерений, как, например, влияние микрокаверн в стенках скважины на измерения в рентгенорадиометрических методах каротажа, грозových разрядов на электроразведочные методы измерений и т.п. Выделение информационных составляющих из зарегистрированных сигналов или максимальное подавление шумов и помех в информационном сигнале при сохранении его полезных составляющих является одной из основных задач первичной обработки сигналов (результатов наблюдений).

Если помехи известны и регулярны, как например, фон переменного тока, то борьба с ними особых затруднений не представляет. Наибольшие трудности представляет борьба со случайными (непредсказуемыми) помехами. В общей форме влияние помех на регистрируемый сигнал записывается в следующем виде:

$$y(t) = V(s(t), q(t)), \quad (17.56)$$

где  $s(t)$  – информационная (полезная) часть сигнала,  $q(t)$  – помеха.

Помеха называется **аддитивной**, и обычно именуется шумом, если выражение (17.56) представляет собой простую сумму сигнала и помехи:

$$y(t) = s(t) + q(t). \quad (17.57)$$

Если случайный процесс  $v(t)$ , оказывающий влияние на сигнал,

является неотрицательным, а его влияние выражается в форме:

$$y(t) = v(t) \cdot s(t), \quad (17.58)$$

то помеху  $v(t)$  называют *мультипликативной*.

В общем случае в сигнале могут присутствовать оба вида помех:

$$y(t) = v(t) s(t) + q(t). \quad (17.59)$$

**Природа помех.** Как правило, случайные шумовые помехи (аддитивные) порождаются различного рода физическими флуктуациями – случайными отклонениями тех или иных физических величин от своих средних значений. Природа флуктуаций обычно определяется статистической природой физических процессов. Многие физические величины представляют собой результаты усреднения определенных параметров физических процессов, дискретных и случайных по своей природе. Так, например, тепловой шум регистрируемого напряжения на резисторах электрических цепей обуславливается флуктуациями теплового движения носителей зарядов - случайностью процесса дрейфа отдельных электронов по резистору, по суммарной интенсивности движения которых и формируется падение напряжения на резисторе. Дискретной является природа электромагнитных видов излучения – дискретный квант энергии излучения (фотон) определен значением  $h\nu$ , где  $h$  – постоянная Планка,  $\nu$  - частота. Флуктуации физических величин, дискретных и случайных по своей природе, принципиально неустранимы, и речь может идти только о том, чтобы уменьшать их относительную величину имеющимися в нашем распоряжении средствами. Природа мультипликативных помех обычно связана с изменениями условий измерений, параметров каналов передачи данных и систем их обработки, т.е. когда случайные помехи накладываются не на сам сигнал непосредственно, а на системы, в которых этот сигнал формируется и обращается, вызывая опосредствованные искажения сигнала, как линейные, так и нелинейные.

**Характеристики помех.** В математическом описании помехи представляются случайными функциями времени. Случайную функцию непрерывного времени обычно называют *случайным процессом*, ее дискретный аналог – *случайной последовательностью*. Как правило, помехи относятся к классу стационарных случайных процессов, и характеризуются своими распределениями и моментами распределений, как их числовыми параметрами. Основное распределение большинства шумовых сигналов – нормальное (гауссов процесс). Это объясняется тем, что распределение сумм независимых случайных величин, из которых складываются случайные помехи, сходится к нормальному, вне зависимости от характера распределения

слагаемых (теорема Ляпунова).

Момент первого порядка выражает среднее значение (постоянную составляющую) случайного процесса:

$$M\{q\} = \bar{q} = \int_{-\infty}^{\infty} q \cdot p(q) dq. \quad (17.60)$$

где  $p(q)$  – плотность вероятностей значений  $q$ .

Центральный момент второго порядка определяет *дисперсию* процесса:

$$D\{q\} = \sigma^2 = \int_{-\infty}^{\infty} (q - \bar{q})^2 \cdot p(q) dq = \overline{q^2} - \bar{q}^2. \quad (17.61)$$

Дисперсия выражает мощность переменной составляющей процесса. Корень квадратный из значения дисперсии, т.е. значение  $\sigma$ , является средним квадратическим значением разброса случайных значений  $q$  относительно среднего значения  $\bar{q}$ .

Смешанный момент второго порядка называется функцией автокорреляции случайного процесса  $q(t)$ :

$$M\{q(t)q(t+\tau)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \cdot p(x_1, x_2) dx_1 dx_2 = B(\tau). \quad (17.62)$$

Величина  $B(\tau)$  при  $\tau=0$  равна общей мощности случайного процесса  $q(t)$ .

На практике большинство случайных процессов обладают свойством эргодичности. Оно заключается в том, что средние значения по множеству реализаций (математические ожидания, вычисляемые по плотностям распределений (17.60-17.62)) совпадают со средними значениями по времени  $T$  одной реализации процесса при  $T \Rightarrow \infty$ . Это позволяет производить оценку числовых значений параметров помех непосредственно по произвольным интервалам  $[a, b]$  задания сигналов:

$$\bar{q} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T q(t) dt \approx \frac{1}{b-a} \int_a^b q(t) dt. \quad (17.63)$$

$$\sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (q(t) - \bar{q})^2 dt \approx \frac{1}{b-a} \int_a^b (q(t) - \bar{q})^2 dt. \quad (17.64)$$

$$B(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T q(t)q(t+\tau) dt \approx \frac{1}{b-a} \int_a^b q(t)q(t+\tau) dt. \quad (17.65)$$

Спектральная плотность мощности случайного процесса (распределение мощности помех и шумов по частоте) связана с функцией автокорреляции преобразованием Фурье. В одностороннем (физическом) представлении спектра:

$$B(f) = 4 \int_0^{\infty} B(\tau) \cos 2\pi f \tau dt. \quad (17.66)$$

$$B(\tau) = \int_0^{\infty} B(f) \cos 2\pi f\tau \, df. \quad (17.67)$$

Аддитивную помеху с равномерным спектром  $B(f) = B_0 = \text{const}$  называют *белым шумом*. Мощность белого шума в полосе частот 0-F пропорциональна ширине полосы:

$$W_F = \int_0^F B(f) \, df = B_0 F.$$

При белом шуме полоса частот всегда полагается конечной, т.к. в противном случае мы получим бесконечную мощность шумов.

Сигнал с аддитивной помехой обычно характеризуют не абсолютной мощностью помехи, а отношением средних мощностей сигнала и помехи, которое кратко называют *отношением сигнал/помеха*:

$$\rho = W_c/W_q.$$

Значения случайных процессов являются некоррелированными только при неограниченной полосе частот. Любое ограничение частотной полосы вносит определенную корреляцию в процесс и независимыми друг от друга можно считать только значения процесса, отстоящие друг от друга как минимум на интервал корреляции  $\tau_0$ :

$$\tau_0 = (2/W_F) \int_0^{\infty} B(t) \, dt = 1/2F.$$

## **18. Динамическая и спектральная форма представления сигналов**

### **18.1. Разложение сигналов по единичным импульсам**

Динамическая форма представления сигналов соответствует естественному и привычному для нас математическому описанию в виде функций независимых переменных (аргументов) в реальном (текущем) масштабе времени. Динамические модели сигналов позволяют определять текущие значения сигналов в любых системах по заданным априори математическим функциям описания физических процессов в реальных физических системах или системных операций в программных системах. Достоинством динамических моделей является их универсальность, основные математические инструменты реализации - дифференциальные уравнения и интеграл Дюамеля, для цифровых сигналов - разностные уравнения и операция свертки.

Основной задачей динамической модели является математическое описание реакции системы (выходного сигнала

системы) на определенное входное воздействие (входной сигнал). Моделирование и анализ линейных стационарных систем обработки сигналов произвольной формы в динамическом представлении базируется на разложении сигналов по единичным импульсам простейшей формы.

**Единичные импульсы.** В качестве математической модели единичного импульса при анализе аналоговых сигналов используют дельта-функцию.

**Дельта-функция** или функция Дирака. По определению, дельта-функция описывается следующими математическими выражениями (в совокупности):

$$\delta(t-\tau) = 0 \text{ при } t \neq \tau, \quad \int_{-\infty}^{\infty} \delta(t-\tau) dt = 1.$$

Функция  $\delta(t-\tau)$  не является дифференцируемой, и имеет размерность, обратную размерности ее аргумента, что непосредственно следует из безразмерности результата интегрирования. Значение дельта-функции равно нулю везде за исключением точки  $\tau$ , где она представляет собой бесконечно узкий импульс с бесконечно большой амплитудой, при этом площадь импульса равна 1.

Дельта-функция является полезной математической абстракцией. На практике такие функции не могут быть реализованы с абсолютной точностью, так как невозможно реализовать значение, равное бесконечности, в точке  $t = \tau$  на аналоговой временной шкале, т.е. определенной по времени также с бесконечной точностью. Но во всех случаях, когда площадь импульса равна 1, длительность импульса достаточно мала, а за время его действия на входе какой-либо системы сигнал на ее выходе практически не изменяется (реакция системы на импульс во много раз больше длительности самого импульса), входной сигнал можно считать *единичной импульсной функцией* со свойствами дельта-функции.

**Функция Кронекера.** Для дискретных и цифровых систем в качестве единичного импульса используется дискретный интегральный аналог дельта-функции - функция единичного отсчета  $\delta(k\Delta t - n\Delta t)$ , которая равна 1 в координатной точке  $k = n$  и нулю во всех остальных точках, при этом функция  $\delta(k\Delta t - n\Delta t)$  определена только для целых значений координат  $k$  и  $n$ . Математические выражения  $\delta(t-\tau)$  и  $\delta(k\Delta t - n\Delta t)$  называют также импульсами Дирака и Кронекера. Однако, применяя такую терминологию, не следует забывать, что это не просто единичные импульсы в координатных точках  $\tau$  и  $n\Delta t$ , а импульсные функции, определяющие как значения импульсов в определенных

координатных точках, так и нулевые значения по всем остальным координатам, в пределе от  $-\infty$  до  $\infty$ .

**Разложение сигнала по единичным импульсам.**

Импульсы Дирака и Кронекера используются для разложения, соответственно, произвольных аналоговых сигналов  $s(t)$  и дискретных сигналов  $s(k\Delta t)$  в непрерывную последовательность непрерывающихся (ортогональных) импульсов:

$$s(t) = \int_{-\infty}^{\infty} s(\tau)\delta(t-\tau) d\tau. \quad (18.1)$$

$$s(k\Delta t) = \sum_{n=-\infty}^{\infty} s(n\Delta t)\delta(k\Delta t-n\Delta t). \quad (18.2)$$

Для аналоговых сигналов разложение (18.1) в физическом представлении эквивалентно сканированию значений сигнала  $s(t)$  в моменты времени  $t = \tau$  бесконечно узкой щелью, бегущей вдоль оси  $t$ . Для цифровых сигналов эта щель равна одному отсчету. Пример разложения дискретного сигнала приведен на рис. 18.1.



Рис. 18.1.

Единичные импульсные функции  $\delta(t-\tau)$ ,  $-\infty < \tau < \infty$ , и  $\delta(k\Delta t-n\Delta t)$ ,  $-\infty < n < \infty$ , образуют в бесконечномерных пространствах системы

координатных базисов  $\{\delta(t-\tau)\}$  и  $\{\delta(k\Delta t-n\Delta t)\}$ , т.к. они не перекрываются и, соответственно, взаимно ортогональны. По этим координатным системам и производится разложение сигналов  $s(t)$  и  $s(k\Delta t)$ . Совокупности проекций сигналов на координатные базисы представляют собой векторные описания сигналов.

**Импульсный отклик линейной системы.** Если на вход линейной системы в момент времени  $t = 0$  подать единичный импульс (Дирака или Кронекера, в зависимости от типа системы), то на выходе мы получим реакцию системы на единичный входной сигнал. Эта реакция называется функцией *импульсного отклика* системы или импульсной характеристикой. Она однозначно определяется оператором преобразования  $h(\cdot)$ :

$$y(t) = T[\delta(t-0)] = h(t). \quad (18.3)$$

$$y(k\Delta t) = T[\delta(k\Delta t-0)] = h(k\Delta t). \quad (18.4)$$

Импульсный отклик аналоговой системы на входную дельта-функцию также в определенной степени представляет собой математическую абстракцию идеального преобразования. С практической точки зрения под импульсным откликом можно понимать отображение реакции системы на импульсный входной сигнал произвольной формы с единичной площадью, если длительность этого сигнала пренебрежимо мала по сравнению с временной (координатной) разрешающей способностью системы. Для цифровых систем импульсный отклик однозначно определяется реакцией системы на импульс Кронекера. Функцию импульсного отклика называют также *весовой функцией системы*.

Очевидно, что в линейных и инвариантных к сдвигу системах форма импульсного отклика не зависит от времени прихода входного сигнала и определяет только его положение на временной оси. Так, если входной импульс задержан (относительно 0) на время  $t_0$ , то соответствующий выходной сигнал будет определяться выражением:

$$y(t) = T[\delta(t-t_0)] = h(t-t_0).$$

В любой системе, работающей в реальном масштабе времени, сигнала на выходе системы не может быть, если нет сигнала на ее входе. Отсюда следует односторонность импульсного отклика физических систем:

$$h(t-\tau) = 0 \text{ при } t < \tau.$$

Для программных систем, работающих с зарегистрированными массивами цифровых данных, импульсный отклик может быть и двусторонним, так как при обработке сигналов в любой текущей точке  $k\Delta t$  системе доступны как "прошлые" отсчеты  $k\Delta t-n\Delta t$ , так и "будущие" отсчеты  $k\Delta t+n\Delta t$ . Это резко расширяет



возможности программной обработки сигналов по сравнению с физическими системами.

На рисунке 18.2 приведен пример импульсного отклика  $h(t)$  элементарной физической системы преобразования электрических сигналов – динамической интегрирующей RC-цепи.

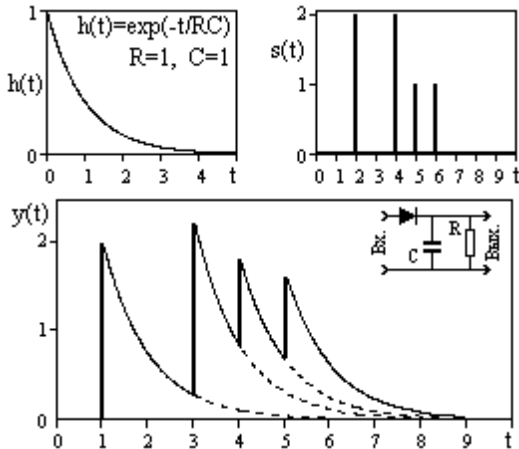


Рис. 18.2.

Подобные схемы очень часто применяются в полевых геофизических приборах (например, в радиометрах) в качестве интенсиметров - измерителей средней скорости счета импульсных потоков сигналов.

При подаче на вход RC-цепи единичного и очень короткого ( $\Delta t \ll RC$ ) импульса заряда  $\Delta q$  емкость  $C$  заряжается до напряжения  $V_0 = \Delta q/C$ , и начинает разряжаться через сопротивление  $R$ , при этом напряжение на емкости изменяется по закону

$$v(t) = V_0 \exp(-t/RC) = (\Delta q/C) \exp(-t/RC).$$

Отсюда, импульсный отклик RC-цепи на единичный входной сигнал с единичным значением заряда  $\Delta q = 1$  равен:

$$h(t) = (1/C) \exp(-t/RC),$$

где форма отклика определяется функцией экспоненты, а множитель  $(1/C)$  является масштабным преобразователем сигнала (заряда в напряжение). По существу, импульсным откликом системы определяется доля входного сигнала, которая действует на выходе системы по истечении времени  $t$  после поступления сигнала на вход (запаздывающая реакция системы).

Если функция импульсного отклика системы известна, то, с учетом принципа суперпозиции сигналов в линейной системе, можно

выполнить расчет реакции системы в любой произвольный момент времени на любое количество входных сигналов в любые моменты времени их прихода путем суммирования запаздывающих реакций системы на эти входные сигналы. На рис. 18.2 приведен пример входного сигнала  $s(t)$  для RC-цепи в виде последовательности импульсов и реакция системы  $y(t)$  на такой входной сигнал, образованная суммированием реакций системы на каждый импульс.

Допустим, что на вход RC-цепи в моменты времени  $t_1=1$  и  $t_2=2$  поступили очень короткие (по сравнению со значением RC) импульсы заряда величиной  $A$  и  $B$ . Математически это можно отобразить сигналом  $s(t) = q_1(t)+q_2(t)$ , где  $q_1(t) = A \cdot \delta(t-t_1)$  и  $q_2 = B \cdot \delta(t-t_2)$ . Выходной сигнал системы при известном импульсном отклике  $h(t)$  отобразится формулой:

$$y(t) = T[q_1(t)+q_2(t)] = T[A\delta(t-t_1)]+T[B\delta(t-t_2)] = A \cdot T[\delta(t-t_1)]+B \cdot T[\delta(t-t_2)] = A \cdot h(t-t_1)+B \cdot h(t-t_2).$$

При расчете значений выходного сигнала в произвольный момент времени  $t$  после прихода на вход системы сигналов  $q_1$  и  $q_2$ , например, для  $t = 5$ , для каждого из сигналов вычисляются значения их запаздывающих реакций:  $y_1 = A \cdot h(5-1) = A \cdot h(4)$  и  $y_2 = B \cdot h(5-2) = B \cdot h(3)$ , после чего значения запаздывающих реакций суммируются  $y = y_1+y_2$ . Пример этой операции можно видеть на рис. 18.3, где для удобства графического представления приняты значения  $A=1$  и  $B=1$ .

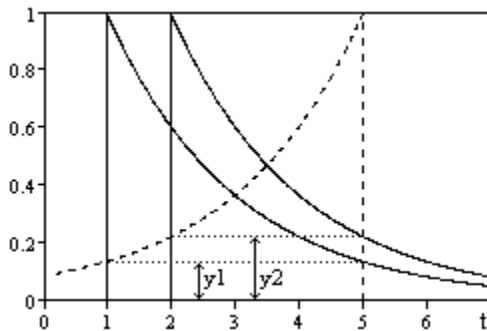


Рис. 18.3.

Сущность операции не изменяется при любых значениях  $A$  и  $B$ , а в общем случае и для любого количества импульсов.

Однако эту же операцию можно рассматривать и с другой позиции. Развернем импульсный отклик  $h(t)$  системы на  $180^\circ$  и поместим его начало  $h(0)$  непосредственно в точку, для которой нужно выполнить расчет выходного сигнала, т.е. в точку  $t=5$  для нашего

примера. Если теперь отсчет координат для функции  $h(t)$  повести назад от точки расчета по аргументу  $\tau$ , т.е. перейти на вычисление  $h(\tau)$ , где значение  $\tau$  изменяется от 0 и далее (в пределе до  $\infty$ ), то нетрудно убедиться (на рисунке это наглядно видно), что функция  $h(\tau)$  пересечет входные импульсы на тех же значениях  $y_1$  и  $y_2$ . Для этих точек пересечения первого и второго импульсов соответственно имеет место  $\tau_1 = t-t_1$  и  $\tau_2 = t-t_2$ , как и при прямом методе расчета запаздывающих реакций при расчете значений  $h(t-t_1)$  и  $h(t-t_2)$ . После умножения полученных значений  $h(\tau_1)$  и  $h(\tau_2)$  на значения входного сигнала  $A$  и  $B$ , получаем полную аналогию:

$$y_1 = A \cdot h(\tau_1) = A \cdot h(t-t_1) \text{ и } y_2 = B \cdot h(\tau_2) = B \cdot h(t-t_2),$$

и соответственно суммарный сигнал  $y = y_1 + y_2$ .

Такое, чисто математическое представление расчета более удобно для составления математических алгоритмов вычислений. Условно этот процесс для коротких входных импульсных сигналов может быть представлен в следующем виде. Для любой точки расчета  $t_i$  выходного сигнала инвертированная по координатному направлению функция импульсного отклика  $h(\tau)$  помещается в эту точку  $t_i$  и просматривается по своей координате  $\tau$  с одновременным синхронным просмотром входного сигнала  $s(t)$  назад от точки расчета (прошлые значения входного сигнала) по координатам  $t_i - \tau$ . Значения всех встреченных при просмотре импульсов  $s(t_i - \tau)$  перемножаются со значениями  $h(\tau)$  и суммируются. Тем самым, для каждой текущей точки расчета  $t_i$  в аналоговой системе выполняется операция:

$$y(t_i) = \int_0^{\infty} h(\tau) \cdot s(t_i - \tau) dt. \quad (18.5)$$

Соответственно в цифровых системах для произвольной точки  $k$ :

$$y(k\Delta t) = \sum_{n=0}^{\infty} h(n\Delta t) \cdot s(k\Delta t - n\Delta t). \quad (18.6)$$

Полученная сумма значений и будет представлять собой запаздывающую реакцию системы на все импульсы, поступившие на вход системы до текущей точки расчета выходного сигнала.

Таким образом, для линейных и стационарных систем легко определить их реакцию на любой входной сигнал, если известен импульсный отклик систем на единичный входной сигнал.

## **18.2. Свертка (конволюция) сигналов**

**Интеграл Дюамеля** позволяет определять реакцию системы на воздействие  $s(t)$  в текущем времени по ее переходной функции  $g(t)$  на единичный скачок входного воздействия:

$$y(t) = y(0) + \int_0^t \frac{\partial}{\partial \tau} g(\tau) s(t-\tau) d\tau, \quad (18.7)$$

где  $y(0)$  – начальное значение выходного сигнала системы.

Пример расчета выходного сигнала системы на ступенчатый входной сигнал приведен на рис. 18.4.

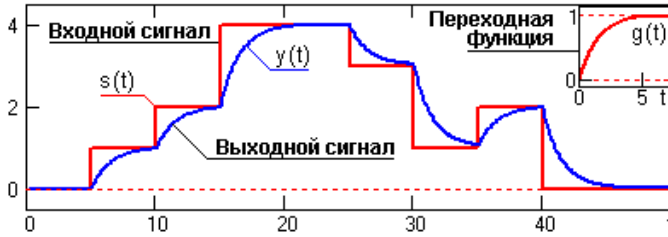


Рис. 18.4.

Ступенчатая форма сигнала принята для более наглядного представления процесса формирования выходного сигнала. В общем случае, форма входного воздействия может быть произвольной.

**Интеграл свертки** – это вариант интеграла Дюамеля.

Произвольный сигнал на входе системы с использованием выражений разложения сигнала может быть представлен в виде последовательной линейной комбинации взвешенных единичных импульсов:

$$y(t) = T[s(t)] = T\left[ \int_{-\infty}^{\infty} s(\tau) \delta(t-\tau) d\tau \right].$$

На основании принципа суперпозиции линейный оператор  $T$  может быть внесен под знак интеграла, т.к. последний представляет собой предельное значение суммы. При этом операция преобразования действует только по переменной  $t$ . Отсюда следует:

$$y(t) = \int_{-\infty}^{\infty} s(\tau) T[\delta(t-\tau)] d\tau = \int_{-\infty}^{\infty} s(\tau) h(t-\tau) d\tau. \quad (18.8)$$

Это выражение и представляет собой интеграл свертки (конволюции) входного сигнала с импульсной характеристикой системы. Заменой переменных  $t-\tau = \tau$  можно убедиться в том, что свертка коммутативна:

$$\int_{-\infty}^{\infty} s(\tau) h(t-\tau) d\tau \equiv \int_{-\infty}^{\infty} h(\tau) s(t-\tau) d\tau. \quad (18.9)$$

Функция  $h(\tau)$  называется *ядром свертки (kernel)* или импульсной характеристикой линейной системы.

Смысл интеграла свертки состоит в том, что входной сигнал представляется сомкнутой последовательностью следующих друг за

другом коротких импульсов, площади которых равны значению сигнала в моменты их следования при длительности импульсов, стремящейся к нулевой. Такая последовательность импульсов условно может рассматриваться в виде последовательности дельта-функций с площадями, равными площадям соответствующих импульсов. Реакция системы (18.8) находится как сумма реакций на каждый импульс, составляющий входное воздействие.

Аналогично, для дискретных сигналов, где значение  $\Delta t$ , как правило, принимается равным 1, а индексы  $k$  и  $n$  выполняют роль номеров отсчетов числовых рядов:

$$y(k) = \sum_{n=-\infty}^{\infty} h(n) s(k-n). \quad (18.10)$$

В цифровых методах обработки сигналов функцию  $h(n)$  обычно называют *оператором свертки*, а его размер по числу отсчетов – *окном оператора свертки*.

Выражения (18.8) имеют специальную форму упрощенной математической записи в символическом виде:

$$y(t) = s(t-\tau) * h(\tau) \equiv s(t) * h(t).$$

Сравнением выражений (18.9 и 18.10) с выражениями (18.5) нетрудно убедиться в их полной идентичности, за исключением нижнего предела интегрирования (суммирования). Это и понятно, так как выражения (18.5) были получены при рассмотрении реальной физической системы, работающей в реальном масштабе времени, импульсный отклик которых является односторонним (равен нулю при  $\tau < 0$ ). Для таких систем интегрирование (и суммирование) от  $-\infty$  до 0 не имеет смысла. Кроме того, в реальных физических системах импульсный отклик, как правило, отличен от нуля только на определенном интервале, и, соответственно, пределы интегрирования (суммирования) в выражениях (18.9 и 18.10) ограничиваются значениями, на которых функции  $h(\tau)$  и  $h(n)$  существует или имеет значимые значения.

Сигналы, обрабатываемые на компьютере, имеют конечную продолжительность. Допустим, сигнал  $s(k)$  отличен от нуля только на отрезке от 0 до  $K$  включительно ("имеет длину  $K+1$ "). Пусть окно оператора свертки  $h(n)$  отлично от нуля на отрезке от  $-N$  до  $N$  ( $2N+1$  отсчет). При подстановке этих сигналов в уравнение свертки, мы получим сигнал  $y(k)$ , который отличен от нуля на отрезке от  $-N$  до  $K+N$  включительно. Таким образом, длина выходного сигнала равна  $2N+K+1$ , т.е. сумме длин исходного сигнала и ядра свертки минус один.

**Техника свертки.** Для вычисления свертки по выражению

(18.9) функция импульсного отклика реверсируется по своей координате, т.е. строится в режиме обратного времени, и движется относительно функции входного сигнала в сторону возрастания значений  $t$ . В каждый текущий момент времени значения обеих функций перемножаются, а произведение интегрируется в пределах окна импульсного отклика. Полученный результат относится к той координатной точке, против которой находится значение импульсного отклика  $h(0)$ .

На рис. 18.5. приведен пример выполнения свертки прямоугольного импульса с импульсным откликом RC-цепи, площадь которого нормирована к 1.

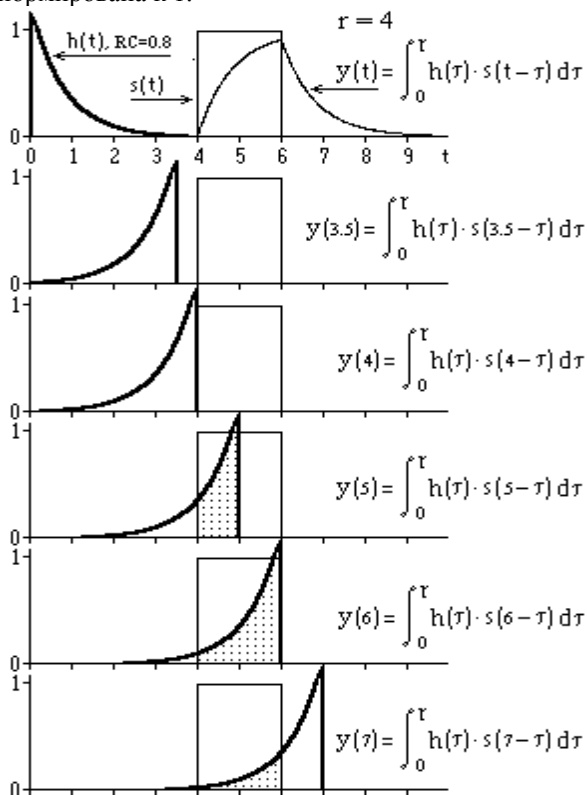


Рис. 18.5.

Если площадь импульсного отклика  $h(t)$  равна 1, то площадь выходного сигнала свертки всегда должна быть равна площади входного сигнала, что можно видеть на верхнем графике рисунка, при

этом одномасштабное сравнение входного и выходного сигналов наглядно демонстрирует характер преобразования сигнала в данной системе. На последующих графиках рисунка демонстрируется вычисление результатов свертки в ряде последовательных точек  $t_i = \{3.5, 4, 5, 6, 7\}$  временной оси. В силу отрицательного знака  $\tau$  в аргументах функции  $s(t-\tau)$  интегрирование произведения  $h(t)s(t-\tau)$  выполняется назад по времени и может ограничиваться только определенной длиной значимых значений импульсного отклика (которая в данном случае установлена равной  $r = 4$ ), а результат относится к начальной точке  $h(0)$  импульсного отклика. Так как входной сигнал, рассмотренный на рисунке, представляет собой прямоугольный импульс с амплитудой 1, то интеграл свертки в каждой текущей точке расчета равен площади импульсного отклика в пределах границ входного прямоугольного импульса (заполнено точками).

Еще более наглядна техника выполнения цифровой свертки, приведенная на рис. 18.6.

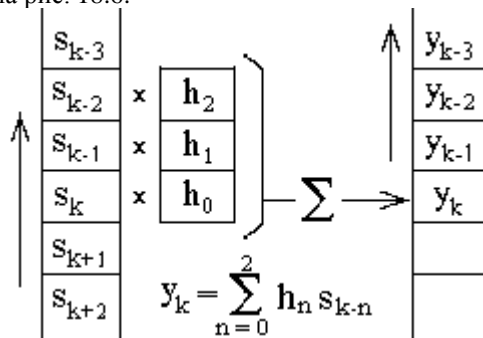


Рис. 18.6.

Для вычисления свертки массив одной из функций ( $s_k$  - входного или свертываемого сигнала) располагается по ходу возрастания номеров. Массив второй функции ( $h_n$  - более короткой), строится параллельно первому массиву в обратном порядке (по ходу уменьшения номеров первого массива или в режиме обратного времени). Для вычисления  $y_k$  значение  $h_0$  располагается против  $s_k$ , все значения  $s_{k-n}$  перемножаются с расположенными против них значениями  $h_n$  и суммируются. Результаты суммирования являются выходным значением функции  $y_k$ , после чего оператор  $h_n$  сдвигается на один номер  $k$  вперед (или функция  $s_k$  сдвигается ему навстречу) и вычисление повторяется для номера  $k+1$  и т.д.

**Свойства свертки.** Для свертки характерны следующие свойства:

- 1. Дистрибутивность:  $h(t) * [a(t)+b(t)] = h(t) * a(t)+h(t) * b(t)$ .
- 2. Коммутативность:  $h(t) * a(t) * b(t) = a(t) * b(t) * h(t)$ .
- 3. Ассоциативность:  $[a(t) * b(t)] * h(t) = h(t) * a(t) * b(t)$ .

Преобразование свертки однозначно определяет выходной сигнал  $y(t)$  для установленного значения входного сигнала  $s(t)$  при известном значении функции импульсного отклика системы  $h(t)$ .

**Обратная задача деконволюции** - *определение функции  $s(t)$  по функциям  $y(t)$  и  $h(t)$* , относится к разряду некорректных, и имеет решение только при вполне определенных условиях. Это объясняется тем, что свертка может существенно изменить частотный спектр сигнала  $y(t)$  относительно  $s(t)$  и восстановление функции  $s(t)$  становится невозможным, если определенные частоты ее спектра в сигнале  $y(t)$  полностью утрачены.

Любая практическая система должна быть **устойчивой**, т.е. *для сигналов, конечных по энергии или средней мощности, выходные сигналы также должны быть конечными по этим параметрам*. Устойчивость обеспечивается при выполнении условия абсолютной интегрируемости импульсного отклика системы:

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

Для систем с  $m$  входами и  $n$  выходами аналогично определяются парциальные импульсные отклики  $h_{ij}(t)$ ,  $i = \{1,2, \dots, n\}$ ,  $j = \{1,2, \dots, m\}$ , каждым из которых отображается сигнал на  $i$ -м выходе при поступлении сигнала  $\delta(t)$  на  $j$ -й вход. Полная совокупность импульсных откликов образует матрицу:

$$\bar{\mathbf{h}}(\tau) = \begin{vmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,m} \\ h_{2,1} & h_{2,2} & \dots & h_{2,m} \\ \cdot & \cdot & \cdot & \cdot \\ h_{n,1} & h_{n,2} & \dots & h_{n,m} \end{vmatrix},$$

а выражение свертки приобретает вид:

$$\bar{\mathbf{Y}}(t) = \int_{-\infty}^{\infty} \bar{\mathbf{h}}(\tau) \bar{\mathbf{S}}(t-\tau) d\tau.$$

Здесь (и в дальнейшем тексте) жирным шрифтом с "крышкой" выделяются векторные величины.

**Системы свертки**. Свертка выполняется системой (физическим или программным устройством). Физические системы,



работающие в реальном времени, вычисляют текущее значение выходного сигнала по всем прошлым значениям входного сигнала, и не могут иметь в своем распоряжении будущих значений входного сигнала. Операторы таких систем являются односторонними (каузальными). Вышеприведенная, нормированная к 1 по площади, функция RC-цепи  $h(t)=(1/RC)\cdot\exp(-t/RC)$ , принятая в качестве системного оператора на рис. 18.5, является именно таким односторонним каузальным оператором. При сравнении выходного сигнала такой системы с входным нетрудно заметить, что выходной сигнал сдвигается относительно входного сигнала. Для каузальных систем такой "сдвиг по фазе" существует всегда и не может быть исключен (сигнал на выходе системы не может быть раньше сигнала на ее входе).

Входным сигналом программных систем является сигнал в целом, записанный в память вычислительного устройства. При обработке таких данных в распоряжении системы при вычислении любой текущей точки выходного сигнала имеются как "прошлые" для данной точки, так и "будущие" значения входного сигнала. Это позволяет создавать системы без сдвига фазы выходного сигнала относительно входного. Для создания таких систем может использоваться два способа:

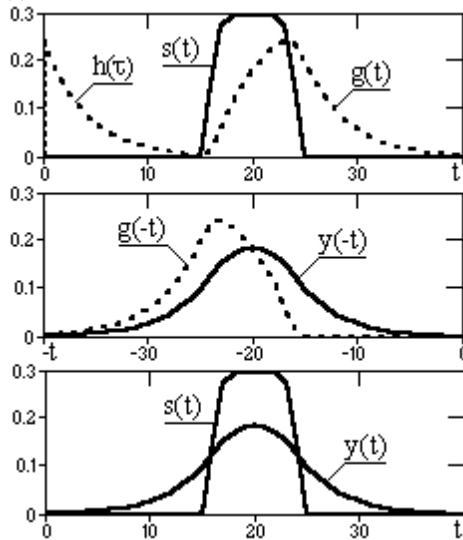


Рис. 18.7.

1. Первый способ иллюстрирует рис. 18.7. Задается система с односторонним каузальным оператором  $h(\tau)$ . Входной сигнал  $s(t)$

пропускается через систему в обычном порядке, и выполняется свертка  $g(t) = h(\tau) * s(t)$ . Затем выходной сигнал  $g(t)$  реверсируется ( $g(t) \Rightarrow g(-t)$ , конец сигнала становится его началом в порядке возрастания  $t$ ) и повторно пропускается через систему, т.е. выполняется свертка  $y(-t) = h(\tau) * g(-t)$ . Полученный сигнал снова реверсируется  $y(-t) \Rightarrow y(t)$ , и результат является окончательным выходным сигналом  $y(t)$  системы.

Три последние операции (реверс  $g(t) \Rightarrow$  свертка с  $h(\tau) \Rightarrow$  реверс выходного сигнала) эквивалентны свертке сигнала  $g(t)$  с реверсированным откликом системы  $h(-\tau)$ , и сдвиг по фазе при свертке реверсированного сигнала компенсирует сдвиг по фазе сигнала, полученный при первой свертке. Общий результат операции  $y(t) = h(\tau) * h(-\tau) * s(t)$  не имеет сдвига по фазе выходного сигнала относительно входного. Такую операцию приходится выполнять для исключения сдвига фазы при применении рекурсивных фильтров, которые всегда являются односторонними.

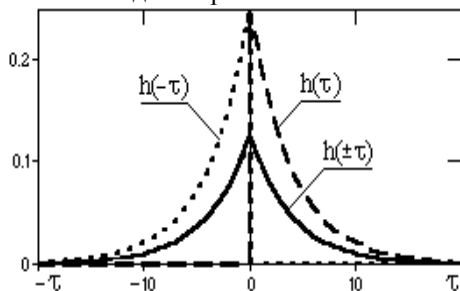


Рис. 18.8.

2. Выходной результат  $y(t) = h(\tau) * h(-\tau) * s(t)$  предыдущей операции позволяет, используя свойство коммутативности свертки, сначала выполнить свертку  $h(\tau) * h(-\tau) = h(\pm\tau)$  и получить один системный оператор  $h(\pm\tau)$  (см. рис. 18.8), обеспечивающий свертку без сдвига фазы. Этот системный оператор является двусторонним и симметричным относительно  $\tau = 0$ . Но использование его возможно только для предварительно записанных сигналов, т.к. при выполнении свертки  $y(t) = h(\pm\tau) * s(t - \tau)$  для отрицательных значений  $\tau$  требуются "будущие" значения входного сигнала  $s(t + \tau)$ . Результат свертки с симметричным оператором полностью аналогичен первой операции (сигнал  $y(t)$  на рис. 18.7).

Приведенное выше формирование двустороннего симметричного оператора свертки имеет чисто познавательный характер. На практике вполне естественным является расчет непосредственно симметричных двусторонних операторов под требуемые задачи обработки числовых данных (сигналов,

зарегистрированных в дискретной числовой форме).

**Начальные условия свертки.** В начальный момент свертки, при вычислении значений  $y(t_i)$  для значений  $t_i < \tau_{\max}$  оператора  $h(\tau)$ , функция оператора, построенная в режиме обратного времени, при  $\tau > t_i$  "зависает" для значений  $t_i - \tau$  против отсутствующих значений входной функции.

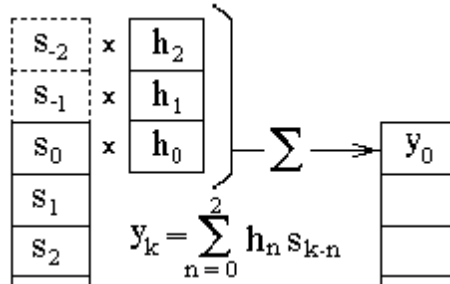


Рис. 18.9.

Пример такого зависания оператора дискретной свертки против несуществующих отсчетов  $s_{-1}$  и  $s_{-2}$  входного массива данных при вычислении отсчета  $y_0$  приведен на рис. 18.9. Зависание исключают либо заданием начальных условий - дополнительных отсчетов, чаще всего нулевых или равных первому отсчету входной функции, либо началом свертки с отсчета входной функции  $k_i = n_{\max}$  с соответствующим сокращением интервала выходной функции на интервал задания системного оператора. Для симметричных операторов со значениями  $-n$  (вперед по времени) такой же момент наступает и в конце входного массива, и требует задания конечных условий или сокращения размера выходного сигнала.

### 18.3. Введение в спектральное разложение сигналов

Спектральная (частотная) форма представления сигналов использует разложение сигнальных функций на периодические составляющие.

Периодичность гармонических колебаний исследовал еще в VI веке до нашей эры Пифагор и даже распространил его на описание гармонического движения небесных тел. Термин "spectrum" ("спектр") впервые применил И. Ньютон в 1571 году при описании разложения солнечного света, пропущенного через стеклянную призму, на многоцветную полосу. Он же дал и первую математическую трактовку

периодичности волновых движений. В 18-м веке решениями волновых уравнений (в приложении к струнам) занимались Даниил Бернулли и Леонард Эйлер. В дальнейших своих работах по математике и физике Д. Бернулли, Л. Эйлер и Ж. Лагранж показали, что ***произвольные периодические функции представляют собой суммы простейших гармонических функций – синусов и косинусов кратных частот***. Эти суммы получили название *рядов Фурье*, после того как в 1807 году французский инженер Жан Батист Фурье обосновал метод вычисления коэффициентов тригонометрического ряда, которым можно отображать с абсолютной точностью (при бесконечном числе членов ряда) или аппроксимировать с заданной точностью (при ограничении числа членов ряда) любую периодическую функцию, определенную на интервале одного периода  $T = b-a$ , и удовлетворяющую условиям Дирихле (ограниченная, кусочно-непрерывная, с конечным числом разрывов 1-го рода). Ряды Фурье в вещественной форме имеют следующий вид:

$$y(x) = (a_0/2) + \sum_{k=1}^{\infty} (a_k \cos(2\pi k f_1 x) + b_k \sin(2\pi k f_1 x)), \quad f_1 = 1/T.$$

$$a_k = (2/T) \int_a^b y(x) \cos(2\pi k f_1 x) dx, \quad b_k = (2/T) \int_a^b y(x) \sin(2\pi k f_1 x) dx.$$

Разложение сигнала на гармонические функции получило название *прямого преобразования Фурье*. Обратный процесс – синтез сигнала по синусоидам – называется *обратным преобразованием Фурье (inverse Fourier transform)*.

На первых этапах своего развития данное направление разложения функций, получившее название *гармонического анализа*, имело больше теоретический характер и использовалось, в основном, в естественных науках для выявления и изучения периодичности и состава периодических составляющих (в том числе скрытых) в различных явлениях и процессах (активность солнца, девиация магнитного поля Земли, метеорологические наблюдения, и т.п.). Теория гармонического анализа была развита в работах Дирихле, Гаусса, Римана, Чебышева, Винера и других с распространением на произвольные функции с бесконечным периодом (интегралы Фурье).

Положение резко изменилось с появлением электротехнических и радиотехнических отраслей науки и техники и развитием радиосвязи, где гармонический состав сигналов приобрел конкретный физический смысл, а математический аппарат спектрального преобразования функций стал основным инструментом анализа и синтеза сигналов и систем. В настоящее время спектральный анализ является одним из основных методов обработки

экспериментальных данных во многих отраслях науки и техники.

Спектральное преобразование функций, по существу, представляет собой представление функций в новой системе координат, т.е. перевод исходных функций на новый координатный базис. Выбор наиболее рациональной ортогональной системы координатного базиса функций, как правило, зависит от цели исследований и определяется стремлением максимального упрощения математического аппарата анализа, преобразований и обработки данных. В качестве базисных функций в настоящее время используются полиномы Чебышева, Эрмита, Лагерра, Лежандра и другие. Наибольшее распространение получило преобразование сигналов в базисах гармонических функций: комплексных экспоненциальных  $\exp(j2\pi ft)$  и вещественных тригонометрических синус-косинусных функций, связанных друг с другом формулой Эйлера. Это объясняется тем, что гармонические колебания является функциями времени, сохраняющими свою форму при прохождении через любую линейную цепь, изменяются только амплитуда и начальная фаза колебаний, что очень удобно для анализа систем преобразования сигналов.

Ряды Фурье произвольных аналоговых периодических сигналов могут содержать бесконечно большое количество членов. Однако одним из важных достоинств преобразования Фурье является то, что при ограничении (усечении) ряда Фурье до любого конечного числа его членов обеспечивается наилучшее по средней квадратической погрешности приближение к исходной функции (для данного количества членов).

Спектральный анализ часто называют частотным анализом. Термин "частотный" обязан происхождением обратной переменной  $f=1/|t|$  временного представления сигналов и функций. Понятие частотного преобразования не следует связывать только с временными аргументами, т.к. математический аппарат преобразования не зависит от физического смысла независимых переменных. Так, например, при переменной "x", как единице длины, значение  $f$  будет представлять собой пространственную частоту с размерностью  $1/|x|$  - число периодических изменений сигнала на единице длины.

В математическом аппарате частотного анализа удобно использовать угловую частоту  $\omega = 2\pi f$ . Для процессов по другим независимым переменным в технической литературе вместо индекса частоты  $f$  часто используется индекс  $\nu$ , а для угловой частоты индекс  $k = 2\pi\nu$ , который называют *волновым числом*.

## 18.4. Разложение сигнала по гармоническим функциям

**Понятие собственных функций.** Удобство использования частотного представления сигналов заключается в том, что гармонические функции являются собственными функциями операций переноса, интегрирования, дифференцирования и других линейных операций, инвариантных по координатам. Они проходят через линейные системы, не изменяя формы, а изменяют лишь фазу и амплитуду.

Допустим, что исходная функция является линейной комбинацией функций синуса и косинуса:

$$s(x) = A \sin(x) + B \cos(x).$$

Осуществим произвольный сдвиг функции по аргументу на величину  $h$ . При этом получаем:

$$s(x+h) = C \sin(x) + D \cos(x),$$

$$C = A \cos(h) - B \sin(h),$$

$$D = A \sin(h) + B \cos(h),$$

где коэффициенты  $C$  и  $D$ , как и в исходном выражении коэффициенты  $A$  и  $B$ , не зависят от аргумента, при этом  $C^2 + D^2 = A^2 + B^2$ . Таким образом, при произвольном переносе функции по аргументу (а равно и при интегрировании, дифференцировании и других линейных преобразованиях) любую линейную комбинацию синуса и косинуса можно представить линейной комбинацией этих же функций.

**Экспоненциальная комплексная запись гармонических функций** делает это свойство еще нагляднее. Для произвольной гармонической функции имеем:

$$\cos(\omega t - \varphi) = A \cos(\omega t) + B \sin(\omega t),$$

где  $A = \cos(\varphi)$ ,  $B = \sin(\varphi)$ ,  $\varphi$  - начальный фазовый угол колебания при  $t=0$ . Переходя к комплексной записи данной функции с использованием тождеств Эйлера

$\cos(\omega t) = [\exp(j\omega t) + \exp(-j\omega t)]/2$ ,  $\sin(\omega t) = [\exp(j\omega t) - \exp(-j\omega t)]/2j$ , получаем:

$$\cos(\omega t - \varphi) = C \exp(j\omega t) + C^* \exp(-j\omega t),$$

где:  $C = 0,5 \exp(-j\varphi)$ ,  $C^* = 0,5 \exp(j\varphi)$  - величина, комплексно сопряженная с  $C$ . Применяя в качестве гармонической составляющей разложения сигнала функцию  $\exp(j\omega t)$ , можно рассматривать вторую функцию  $\exp(-j\omega t)$ , комплексно сопряженную с первой, как такую же составляющую, но с отрицательной частотой. Естественно, что отрицательная частота является чисто математической абстракцией, но нужно помнить, что пара таких комплексно сопряженных

составляющих в сумме всегда дает вещественную функцию, т.е. является отображением (образом) вещественной функции в новом математическом пространстве (базис - комплексные экспоненциальные функции).

Экспоненциальные функции также являются собственными функциями линейных операций. Для операции переноса с использованием экспоненциальных функций:

$$\exp[j\omega(t+h)] = \exp(j\omega h) \cdot \exp(j\omega t) = H(\omega) \exp(j\omega t),$$

где  $H(\omega) = \exp(j\omega h)$  - собственное значение операции переноса, независимое от переменной.

Для операции дифференцирования:

$$d[\exp(j\omega t)]/dt = j\omega \exp(j\omega t), \quad H(\omega) = j\omega.$$

Для операции интегрирования:

$$\int \exp(j\omega t) dt = (1/j\omega) \exp(j\omega t), \quad H(\omega) = 1/j\omega.$$

В общей форме, для любых линейных операций преобразования:

$$T[\exp(j\omega t)] = H(\omega) \exp(j\omega t),$$

где  $T[.]$  - произвольный линейный оператор,  $H(\omega)$  - собственное значение операции, независимое от аргумента.

У специалистов - практиков существует предубеждение против использования комплексных функций с их мнимыми частотами. Поэтому в дальнейшем будем использовать и вещественные функции, и их комплексные аналоги, по крайней мере, до тех пор, пока простота и удобство использования последних не станет очевидным.

**Ряды Фурье.** Разложению в ряды Фурье подвергаются периодические сигналы. Периодическую функцию любой формы, заданную на интервале одного периода  $T = b-a$  и удовлетворяющую на этом интервале условиям Дирехле (ограниченная, кусочно-непрерывная, с конечным числом разрывов 1-го рода), можно представить в виде ряда Фурье:

$$s(t) = \sum_{n=-\infty}^{\infty} S_n \exp(jn\Delta\omega t), \quad S_n = S(n\Delta\omega), \quad \Delta\omega = 2\pi/T, \quad (18.11)$$

где весовые коэффициенты  $S_n$  ряда определяются по формуле:

$$S_n = (1/T) \int_a^b s(t) \exp(-jn\Delta\omega t) dt. \quad (18.12)$$

Ряд Фурье представляет собой ансамбль комплексных экспонент  $\exp(jn\Delta\omega t)$  с частотами, образующими арифметическую прогрессию. Функцию весовых коэффициентов  $S(n\Delta\omega)$  принято называть комплексным спектром периодического сигнала или фурье-

образом функции  $s(t)$ . Спектр периодического сигнала является дискретной функцией, т.к. он определен только для целых значений  $n$  с шагом по частоте, обратным периоду:  $\Delta\omega = 2\pi/T$  (или  $\Delta f = 1/T$ ). Первую частотную составляющую спектра при  $n=1$ , равную  $\omega_1 = 1 \cdot \Delta\omega = 2\pi/T$  (или  $f_1 = 1/T$ ), называют *основной* частотой сигнала (*первой гармоникой*), остальные частоты дискретного спектра  $n\omega_1$  при  $n>1$  называют *гармониками* сигнала. Значения  $S(n\Delta\omega)$  по положительным и отрицательным значениям  $n$  являются комплексно сопряженными. Шаг по частоте  $\Delta\omega$  между двумя соседними синусоидами из разложения Фурье называется *частотным разрешением* спектра.

С чисто математических позиций множество функций  $\exp(jn\Delta\omega t)$ ,  $-\infty < n < \infty$  образует бесконечномерный базис линейного пространства  $L^2[a,b]$  ортогональных синус-косинусных функций, а коэффициенты  $S_n$  по (18.12) представляют собой проекции сигнала  $s(t)$  на эти базисные функции. Соответственно, сигнал  $s(t)$  в форме ряда Фурье (18.11) – это бесконечномерный вектор в пространстве  $L^2[a,b]$ , точка с координатами  $S_n$  по базисным осям пространства  $\exp(jn\Delta\omega t)$ .

Коэффициенты  $S_n$  в (18.12) отображают функцию  $s(t)$  в новое пространство единственным образом. Если функция  $s(t)$  непрерывна, то ряд (18.11) сходится равномерно к  $s(t)$ , при этом ошибка аппроксимации  $\|s(t)-s_N(t)\|$  функции  $s(t)$  с усечением ряда (18.11) до  $\pm N$  членов меньше ошибки аппроксимации любым другим рядом с тем же количеством членов. Если  $s(t)$  не является непрерывной (имеет разрывы), но конечна по энергии (квадратично интегрируема), то метрика  $\|s(t)-s_N(t)\|$  стремится к нулю при  $N \rightarrow \infty$ , при этом в точках разрыва сумма ряда стремится к  $(s(t^+)+s(t^-))/2$ .

Подынтегральную функцию экспоненты в выражении (18.12) с использованием тождества Эйлера

$$\exp(\pm j\omega t) = \cos(\omega t) \pm j \cdot \sin(\omega t)$$

можно разложить на косинусную и синусную составляющие и выразить комплексный спектр в виде действительной и мнимой части:

$$S_n = (1/T) \int_a^b s(t) [\cos(n\Delta\omega t) - j \sin(n\Delta\omega t)] dt = A_n - jB_n. \quad (18.13)$$

$$A_n \equiv A(n\Delta\omega) = (1/T) \int_a^b s(t) \cos(n\Delta\omega t) dt, \quad (18.14)$$

$$B_n \equiv B(n\Delta\omega) = (1/T) \int_a^b s(t) \sin(n\Delta\omega t) dt. \quad (18.15)$$

На рис. 18.10 приведен пример периодического сигнала (прямоугольный импульс на интервале (17.20), повторяющийся с



периодом  $T=40$ ) и форма действительной и мнимой части его спектра. Обратим внимание, что действительная часть спектра является четной относительно нуля функцией  $A(n\Delta\omega)=A(-n\Delta\omega)$ , так как при вычислении значений  $A(n\Delta\omega)$  по формуле (18.14) используется четная косинусная функция  $\cos(n\Delta\omega t) = \cos(-n\Delta\omega t)$ . Мнимая часть спектра является нечетной функцией  $B(n\Delta\omega)=-B(-n\Delta\omega)$ , так как для ее вычисления по (18.15) используется нечетная синусная функция  $\sin(n\Delta\omega t) = -\sin(-n\Delta\omega t)$ .

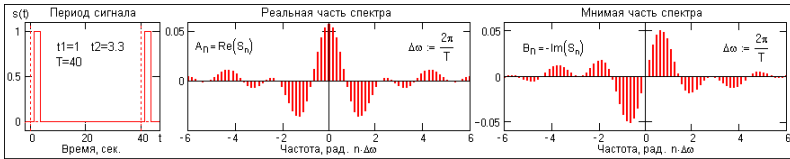


Рис. 18.10. Сигнал и его комплексный спектр.

Комплексные числа дискретной функции (18.13) могут быть представлены в виде модулей и аргументов комплексной экспоненты, что дает следующую форму записи комплексного спектра:

$$\begin{aligned}
 S_n &= R_n \exp(j\varphi_n), & (18.16) \\
 R_n^2 &\equiv R^2(n\Delta\omega) = A^2(n\Delta\omega) + B^2(n\Delta\omega), \\
 \varphi_n &\equiv \varphi(n\Delta\omega) = \arctg(-B(n\Delta\omega)/A(n\Delta\omega)).
 \end{aligned}$$

Модуль спектра  $R(n\Delta\omega)$  называют двусторонним спектром амплитуд или АЧХ - амплитудно-частотной характеристикой сигнала, а аргумент спектра (последовательность фазовых углов  $\varphi(n\Delta\omega)$ ) - двусторонним спектром фаз или ФЧХ - фазово-частотной характеристикой. Спектр амплитуд всегда представляет собой четную функцию:  $R(n\Delta\omega) = R(-n\Delta\omega)$ , а спектр фаз нечетную:  $\varphi(n\Delta\omega) = -\varphi(-n\Delta\omega)$ . Пример спектра в амплитудном и фазовом представлении для сигнала, показанного на рис. 18.10, приведен на рис. 18.11. При рассмотрении спектра фаз следует учитывать периодичность  $2\pi$  угловой частоты (при уменьшении фазового значения до величины менее  $-\pi$  происходит сброс значения  $-2\pi$ ).

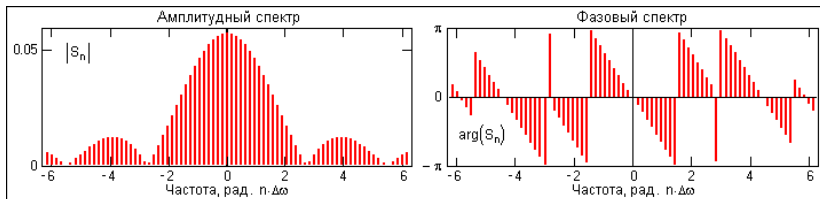


Рис. 18.11. Модуль и аргумент спектра.

Если функция  $s(t)$  является четной, то все значения  $B(n\Delta\omega)$  по (18.15) равны нулю, т.к. четные функции ортогональны синусным гармоникам и подынтегральное произведение  $s(t)\cdot\sin(n\Delta\omega t)$  дает нулевой интеграл. Следовательно, спектр функции будет представлен только вещественными коэффициентами. Напротив, при нечетности функции  $s(t)$  обнуляются все значения коэффициентов  $A(n\Delta\omega)$  (нечетные функции ортогональны косинусным гармоникам) и спектр является чисто мнимым. Этот фактор не зависит от выбора границ задания периода функции на числовой оси.

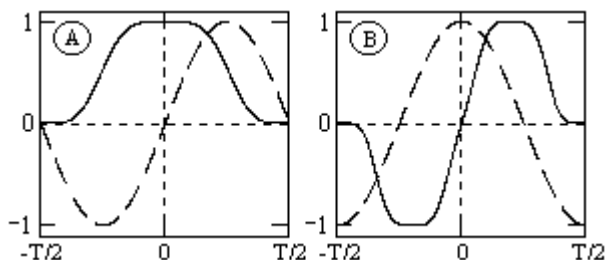


Рис. 18.12. Ортогональность функций.

На рис. 18.12(A) можно наглядно видеть ортогональность первой гармоники синуса и четной функции, а на рис. 18.12(B) соответственно косинуса и нечетной функции в пределах одного периода. Учитывая кратность частот последующих гармоник первой гармонике спектра, ортогональность сохраняется для всех гармоник ряда Фурье.

При  $n=0$  имеем  $B_0 = 0$ , и получаем постоянную составляющую сигнала:

$$S_0 \equiv A_0 \equiv R_0 \equiv (1/T) \int_a^b s(t) dt.$$

**Тригонометрическая форма рядов Фурье.** Объединяя в (18.11) комплексно сопряженные составляющие (члены ряда, симметричные относительно центрального члена ряда  $S_0$ ), можно перейти к ряду Фурье в тригонометрической форме:

$$s(t) = A_0 + 2 \sum_{n=1}^{\infty} (A_n \cos(n\Delta\omega t) + B_n \sin(n\Delta\omega t)), \quad (18.17)$$

$$s(t) = A_0 + \sum_{n=1}^{\infty} 2R_n \cos(n\Delta\omega t + \varphi_n). \quad (18.18)$$

Значения  $A_n$ ,  $B_n$  вычисляются по формулам (18.14-18.15),

значения  $R_n$  и  $\varphi_n$  – по формулам (18.16).

Ряд (18.17) представляют собой разложение периодического сигнала  $s(t)$  на сумму вещественных элементарных гармонических функций (косинусных и синусных) с весовыми коэффициентами, удвоенные значения которых (т.е. значения  $2A_n$ ,  $2B_n$ ) не что иное, как реальные амплитуды соответствующих гармонических колебаний с частотами  $n\Delta\omega$ . Совокупность амплитудных значений этих гармоник образует односторонний физически реальный (только для положительных частот  $n\Delta\omega$ ) спектр сигнала. Для сигнала на рис. 18.11, например, он полностью повторяет правую половину приведенных на рисунке спектров с удвоенными значениями амплитуд (за исключением значения  $A_0$  на нулевой частоте, которое, как это следует из (18.17), не удваивается). Но такое графическое отображение спектров используется довольно редко.

В технических приложениях более широкое применение для отображения физически реальных спектров находит формула (18.18). Спектр амплитуд косинусных гармоник  $2R_n$  при таком отображении называется амплитудно-частотным составом сигнала, а спектр фазовых углов гармоник – фазовой характеристикой сигнала. Форма спектров повторяет правую половину соответствующих двусторонних спектров (см. рис. 18.10) также с удвоенными значениями амплитуд. Для четных сигналов отсчеты фазового спектра могут принимать только значения 0 или  $\pi$ , для нечетных соответственно  $\pm\pi/2$ .

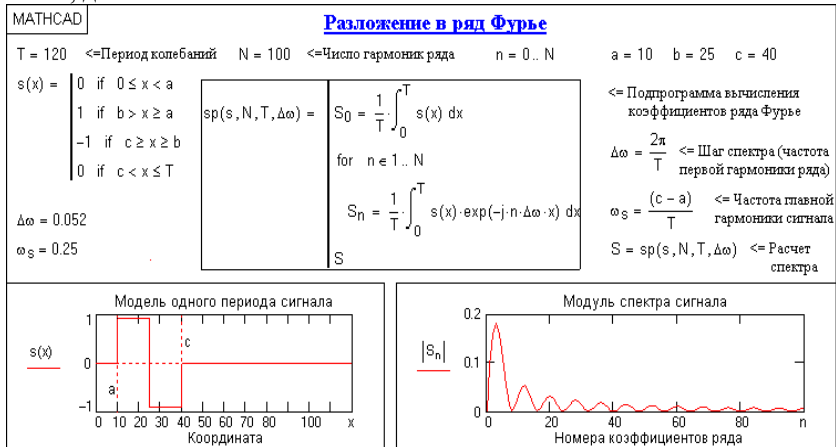


Рис. 18.13. Разложение сигнала в комплексный ряд Фурье.

На рис. 18.13 показано разложение в комплексный ряд Фурье модельного сигнала, выполненное в среде Mathcad. Модель сигнала задана с тремя разрывами первого рода (скачками). Любой скачок

функции содержит все частоты диапазона до бесконечности, в связи с чем ряд Фурье также бесконечен и очень медленно затухает. На рисунке приведены значения только первых 100 членов ряда, при этом график спектра сигнала, как это обычно принято на практике, построен в виде огибающей значений модулей коэффициентов ряда  $S_n$  и только по области положительных значений  $n$ .

Программа на рис. 18.14.продолжает программу рис. 18.13 и показывает реконструкцию сигнала по его спектру при ограничении числа членов ряда Фурье.

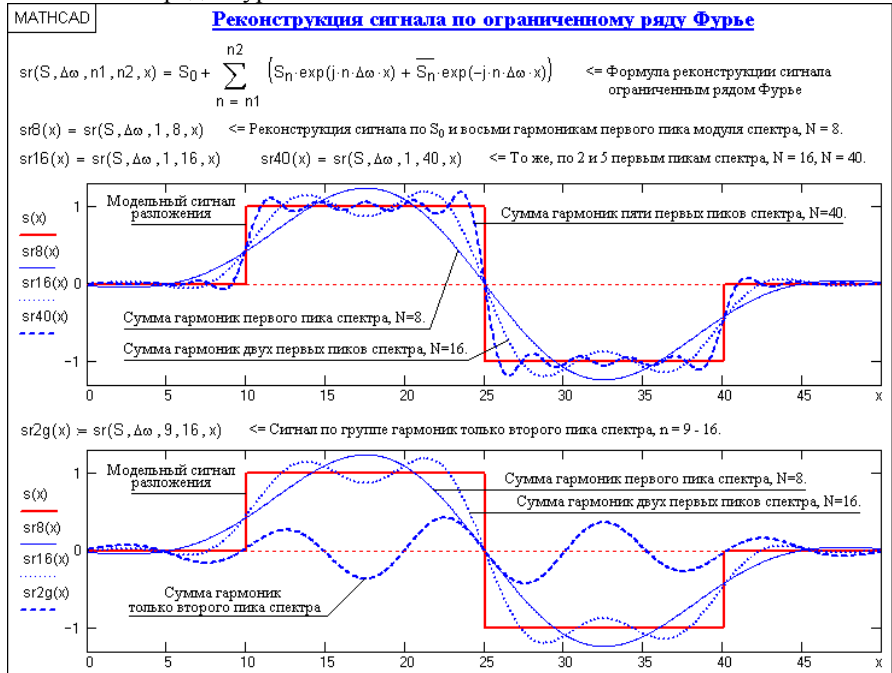


Рис. 18.14. Реконструкция сигнала (продолжение программы на рис. 18.13)

На верхнем графике рисунка приведен реконструированный сигнал при  $N = 8$  (гармоники первого пика спектра, центр которого соответствует главной гармонике сигнала и члену ряда  $n = \omega_s/\Delta\omega$ ),  $N = 16$  (гармоники двух первых пиков) и  $N=40$  (пять первых пиков спектра). Естественно, что чем больше членов ряда включено в реконструкцию, тем ближе реконструированный сигнал к форме исходного сигнала.

Принцип последовательного приближения к исходной форме наглядно виден на нижнем графике рисунка. На нем же можно видеть

и причины появления пульсаций на реконструкции скачков функций, которые носят название *эффекта Гиббса*. При изменении количества суммируемых членов ряда эффект Гиббса не исчезает. Не изменяется также относительная амплитуда пульсаций (по отношению к амплитуде скачка) и относительное затухание (по коэффициенту последовательного уменьшения амплитуды пульсаций по отношению к максимальному выбросу), изменяется только частота пульсаций, которая определяется частотой последних суммируемых гармоник.

Эффект Гиббса имеет место всегда при резких нарушениях монотонности функций. На скачках эффект максимален, во всех других случаях амплитуда пульсаций зависит от характера нарушения монотонности функции. Пример явления Гиббса для радиопульса приведен на рис. 18.15 (использована программа на рис. 18.13, точками показан реконструированный сигнал с увеличением масштаба в 10 раз).

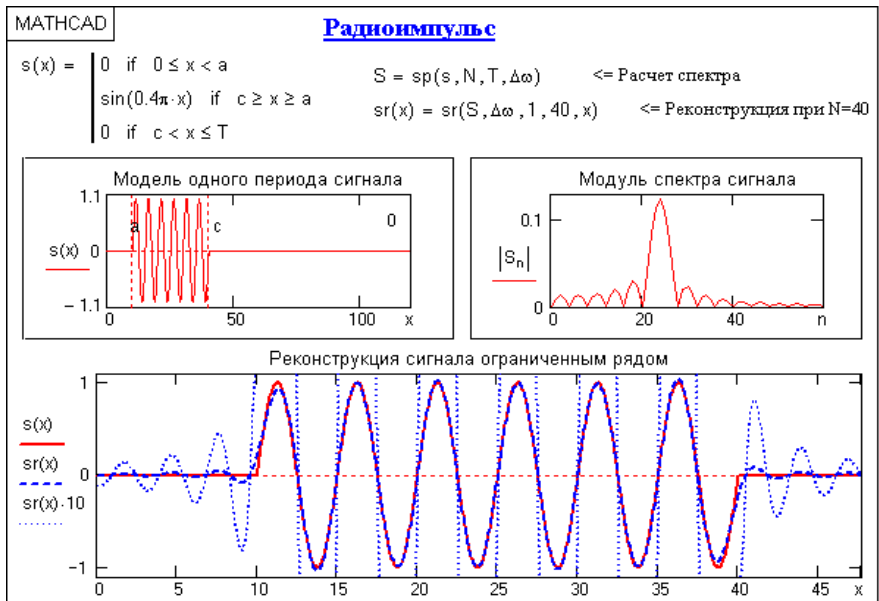


Рис. 18.15.

На рис. 18.16 приведен пример разложения в ряд Фурье одного периода  $T=(a,c)$  модельного периодического сигнала  $sq(x)$ , представленного информационным сигналом  $s(x)$  в сумме с шумовым сигналом. Спектр шумов близок к спектру белого шума (равномерное распределение энергии шумов по всем частотам спектра).

На спектре модельного сигнала достаточно четко выделяется

диапазон частот информационного сигнала. Реконструкция сигнала с ограничением ряда Фурье гармониками только информационного сигнала (сигнал  $sr5(x)$ ,  $N=5$ ) дает сглаженную форму сигнала по минимуму среднеквадратического расхождения с модельным сигналом для данного количества членов ряда, но только по периоду разложения (а, с), и наиболее точное приближение к информационному сигналу. При увеличении в реконструкции количества членов ряда Фурье восстановленный сигнал начинает приближаться к модельному сигналу, но только по данному периоду  $T=(a,c)$ , при этом расхождение с информационным сигналом увеличивается. Заметим, что спектр сигнала может определяться и по нескольким периодам сигнала, что повышает точность реконструкции информационного сигнала.

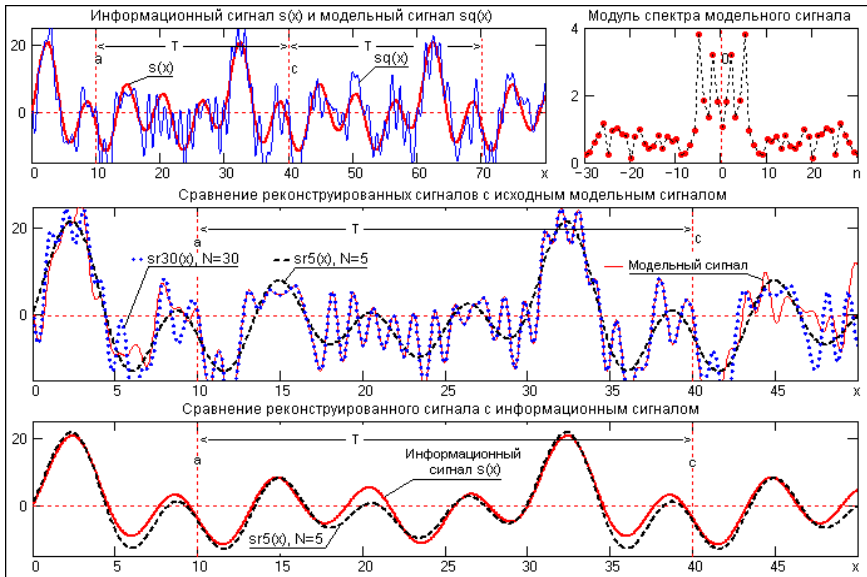


Рис. 18.16.

В ряд Фурье может разлагаться и произвольная непериодическая функция, заданная (ограниченная, вырезанная из другого сигнала, и т.п.) на интервале  $(a,b)$ , если нас не интересует ее поведение за пределами данного интервала. Однако следует помнить, что применение формул (18.11-18.17) автоматически означает периодическое продолжение данной функции за пределами заданного интервала (в обе стороны от него) с периодом  $T = b-a$ . Однако при этом на краях интервала может возникнуть явление Гиббса, если уровень сигнала на краях не совпадает и образуются скачки сигнала при его периодическом повторении, как это видно на рис. 18.17. При

разложении исходной функции в ограниченный ряд Фурье и его обработке в частотной области на самом деле при этом обрабатывается не исходная функция, а реконструированная из ограниченного ряда Фурье.

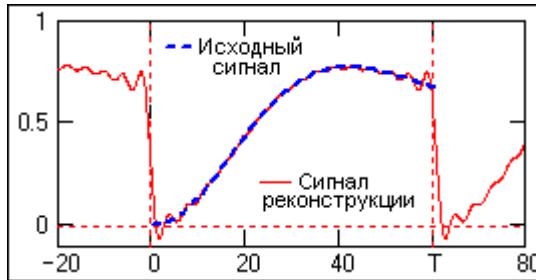


Рис. 18.17.

При усечении рядов Фурье определенное искажение функций существует всегда. Но при малой доле энергии отсекаемой части сигнала (при быстром затухании спектров функций) этот эффект может быть и мало заметен. На скачках и разрывах функций он проявляется наиболее ярко.

**Параметры эффекта Гиббса.** Большинство методов анализа и обработки сигналов представляют собой или имеют в своем составе операцию свертки сигналов с функцией оператора свертки. Как сигнал, так и оператор свертки, выполняющий определенную задачу обработки данных и реализующий определенную частотную функцию системы обработки, могут быть бесконечно большими. Практика же обработки на ЭВМ может иметь дело только с ограниченными множествами и данных, и коэффициентов оператора. В общем случае, эти ограниченные множества "вырезаются" из бесконечных множеств, а разложение в ряды Фурье, также ограниченные по размерам, является одной из самых распространенных операций обработки цифровых множеств. С учетом этого рассмотрим явление Гиббса более подробно, т.к. при любых ограничениях рядов Фурье оно всегда может весьма существенно сказаться на качестве и точности обработки сигналов.

Очевидно, что при усечении ряда Фурье (18.11) любой функции до конечного числа членов  $N$  мы будем иметь усеченный ряд Фурье:

$$s_N(x) = \sum_{n=-N}^N S(n) \exp(jxn\Delta\omega), \quad (18.19)$$

при этом происходит усечение спектральной характеристики функции до частоты  $n\Delta\omega$  и сходимость суммы остающихся членов ряда  $s_N(x)$  к исходной функции  $s(x)$  ухудшается в тем большей степени, чем меньше значение  $N$ . Особенно ярко это проявляется на крутых перепадах (разрывах, скачках) функций:

- крутизна перепадов "размывается", т.к. она не может быть больше, чем крутизна (в нулевой точке) последней сохраненной гармоники ряда (18.19);

- по обе стороны "размытых" перепадов появляются выбросы и затухающие осцилляции с частотой, равной частоте последнего сохраненного или первого отброшенного члена ряда (18.19).

Рассмотрим явление Гиббса на примере разложения в ряд Фурье функции единичного скачка  $s(x)$ , которая имеет разрыв величиной 1 в точке  $x = 0$ . Уравнение функции:

$$s(x) = -0.5 \text{ при } -T/2 \leq x < 0; \quad s(x) = 0.5 \text{ при } 0 \leq x \leq T/2.$$

Поскольку функция является нечетной, ее ряд Фурье не содержит косинусных членов, и коэффициенты ряда в односторонней тригонометрической форме определяются выражением (с учетом соотношения  $\Delta\omega = 2\pi/T$ ):

$$b_n = (2/T) \int_{-T/2}^{T/2} s(x) \sin(xn\Delta\omega) dx = (2/T) \int_0^{T/2} \sin(xn\Delta\omega) dx.$$

$$b_n = 2/(n\pi), \quad n\text{- нечетное,}$$

$$b_n = 0, \quad n\text{- четное.}$$

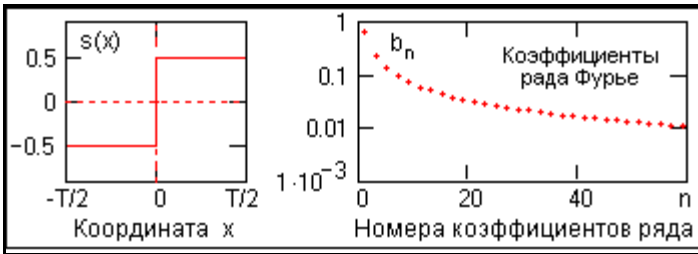


Рис. 18.18. Значения коэффициентов  $b_n$ .

Как видно на рис. 18.18, ряд коэффициентов  $b_n$  затухает очень медленно. Соответственно, медленно будет затухать и ряд Фурье функции  $s(x)$ :

$$s(x) = (2/\pi)[\sin x\Delta\omega + (1/3)\sin x3\Delta\omega + (1/5)\sin x5\Delta\omega + \dots].$$

$$s(x) = (2/\pi) \sum_{n=0}^{\infty} \sin[x(2n+1)\Delta\omega]/(2n+1). \quad (18.20)$$

Этот ряд при усечении до  $M$  нечетных членов можно записать в виде:



$$s(x) = (2\Delta\omega/\pi) \sum_{n=0}^M \int_0^x \cos(x(2n+1)\Delta\omega) dx =$$

$$(2\Delta\omega/\pi) \int_0^x \left[ \sum_{n=0}^M \cos(x(2n+1)\Delta\omega) \right] dx.$$

Сумма косинусного ряда равна  $\sin[2(M+1)x\Delta\omega]/(2\sin x\Delta\omega)$ . Отсюда:

$$s_M(x) = \frac{\Delta\omega}{\pi} \int_0^x \frac{\sin 2(M+1)x\Delta\omega}{\sin x\Delta\omega} dx. \quad (18.21)$$

Для определения местоположения максимумов и минимумов возникающих осцилляций функции, приравняем к нулю ее первую производную (подынтегральную функцию) выражения (18.21), при этом:

$$x_k = \pm k\pi/(2\Delta\omega(M+1)) = \pm kT/(4(M+1)), \quad k = 1, 2, \dots$$

Соответственно, амплитудные значения первых (максимальных) осцилляций функции приходится на точки  $x_{k=1} = \pm T/(4(M+1))$ , вторых (противоположных по полярности) - на точки  $x_{k=2} = \pm T/(2(M+1))$ . Период пульсаций равен

$$x_{k=3} - x_{k=1} \equiv 2x_{k=1} = \pm T/(2(M+1)),$$

т.е. на одном периоде задания сигнала появляется  $2(M+1)$  пульсация с частотой, обратным периоду и равной  $2(M+1)\Delta f$  – частоте последнего сохраненного в суммировании члена ряда Фурье. Функция пульсаций (при ее выделении) является нечетной относительно скачка. Соответственно, при скачке функции  $s(x)$  на точке периода  $T$  значения  $x_k$  являются значениями  $\Delta x_k$  относительно точки скачка. Амплитудные значения функции в точках  $x_1$  и  $x_2$  (при подстановках  $x_1$  и  $x_2$  верхним пределом в (18.21)) практически не зависят от количества членов ряда  $M$  и равны:

$$s_M(x_1) \approx 0.5 + 0.09, \quad s_M(x_2) \approx 0.5 - 0.05.$$

Амплитуда последующих осцилляций постепенно затухает.

Реконструкция скачка при трех значениях ряда приведена на рис. 18.19.

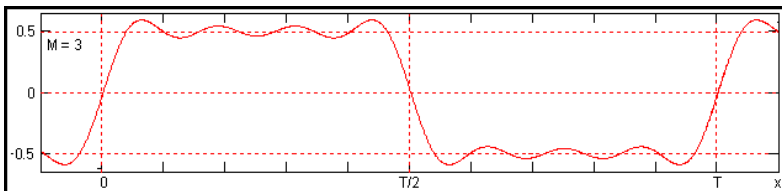


Рис. 18.19. Реконструкция скачка по ограниченному ряду Фурье при  $M=3$ .

Как и положено, функция продолжается периодически за пределами заданного интервала  $(-T/2, T/2)$ , при этом на границах периодов также образуются скачки. Скачки являются центрами возникающих осцилляций. Наложение осцилляций друг на друга в зависимости от расстояния между их центрами может как уменьшать амплитуду пульсаций, так и увеличивать.

Таким образом, для усеченных рядов Фурье предельные значения максимальных выбросов по обе стороны от скачка и следующих за ними обратных выбросов при единичной амплитуде разрыва функции достигают соответственно 9% и 5% значения амплитуды скачка. Кроме того, сам скачок функции из собственно скачка преобразуется в переходную зону, длина которой между точками максимальных выбросов по обе стороны скачка равна  $T/(2(M+1))$ , а по уровню исходных значений функции на скачке (в данном случае от -0.5 до 0.5) порядка  $(2/3)T/(2(M+1))$ . Это явление типично для всех функций с разрывами.

## **18.5. Непрерывные преобразования Фурье и Лапласа**

**Интеграл Фурье.** Спектры непериодических сигналов конечной длительности (финитных), зарегистрированных на интервале  $T$ , могут быть получены из уравнений для рядов Фурье как предельные значения функций суммирования при расширении периода  $T$  до бесконечности.

Зададим периодическую последовательность импульсов и разложим импульс на одном периоде  $T$  в ряд Фурье (формула 18.17). Не меняя положения импульса на интервале  $T$ , увеличим значение  $T$  в два раза (продлеваем интервал нулями), при этом выражение (18.17) для вычисления спектра остается без изменения, но по ней рассчитывается в 2 раза большее количество гармоник с уменьшением в 2 раза частоты первой гармоники и шага  $\Delta\omega=2\pi/T$ . Увеличение интервала  $T$  не влияет на результаты вычисления интеграла функции (18.17), т.к. интервал продления заполнен нулевыми значениями сигнала.

По существу, при увеличении периода  $T$  без изменения финитного сигнала форма спектра по оси частот остается без изменения, изменяется только шаг дискретизации спектра и, за счет множителя  $1/T$ , в 2 раза уменьшаются значения спектра. Новые гармоники располагаются в интервалах между гармониками первого ряда. Пример изменения спектра при увеличении периода  $T$  в 2 раза приведен на рис. 18.20.

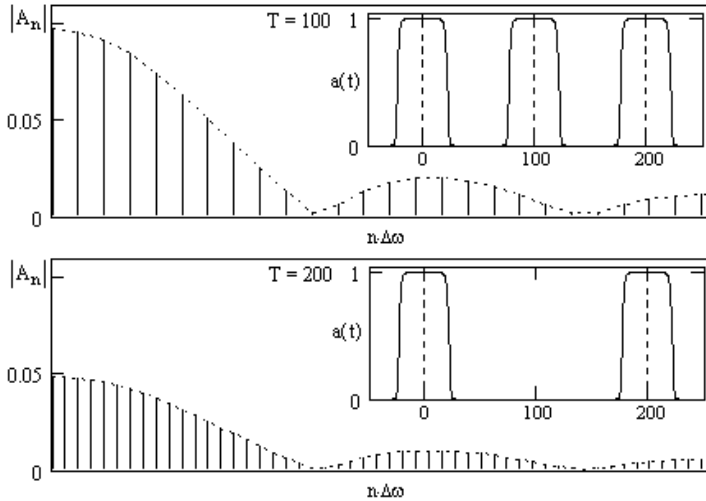


Рис. 18.20.

Процесс можно продолжить дальнейшим последовательным увеличением периода, при этом спектр будет приближаться к непрерывной функции. В пределе, при  $T \rightarrow \infty$ , периодическая последовательность импульсов заменяется одиночным финитным сигналом, расстояние между гармониками  $1/T = \Delta\omega/2\pi$  уменьшается до  $d\omega/2\pi$ , дискретные частоты  $n\Delta\omega$  при  $\Delta\omega \rightarrow 0$  обращаются в непрерывные текущие значения, а суммирование амплитудных значений заменятся интегрированием. При этом фазовый и амплитудный спектр становятся непрерывными, а сами значения спектра становятся бесконечно малыми ( $1/T = d\omega/2\pi \rightarrow 0$ ). Для исключения последнего уравнение для спектра нормируем на  $d\omega/2\pi$ :

$$\begin{aligned} S'(\omega) &= (d\omega/2\pi) \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt \rightarrow S'(\omega) 2\pi/d\omega = \\ &= \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt = S(\omega), \end{aligned}$$

где  $S(\omega)$  из значений спектра  $S'(\omega)$  превращается в плотность распределения значений спектра, и возвращаем нормировку при восстановлении сигнала по спектру:

$$s(t) = (d\omega/2\pi) \sum_{n=-\infty}^{\infty} S(\omega) \exp(jn\Delta\omega t) \rightarrow (1/2\pi) \int_{-\infty}^{\infty} S(\omega) \exp(j\omega t) d\omega.$$

Таким образом, интегральное преобразование Фурье

приобретает следующий вид:

$$s(t) = (1/2\pi) \int_{-\infty}^{\infty} S(\omega) \exp(j\omega t) d\omega, \quad (18.22)$$

$$S(\omega) = \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt. \quad (18.23)$$

Формулу (18.23) обычно называют формулой **прямого преобразования Фурье**, а формулу (18.22) – **обратного преобразования Фурье**. Этими выражениями устанавливается взаимно однозначная связь сигнала и его спектра, а точнее – плотности спектра сигнала в последовательной полосе малых (стремящихся к нулю) полосах частот. Эту величину называют **спектральной плотностью сигнала**. **Спектральные функции содержат ровно столько информации, сколько и исходный сигнал.**

При преобразовании сигнала в пространство гармонических частот и обратно формулы прямого и обратного преобразований Фурье тождественны за исключением знака аргументов экспоненты:

$$s(t) = \int_{-\infty}^{\infty} S(f) \exp(j2\pi ft) df, \quad (18.24)$$

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) dt. \quad (18.25)$$

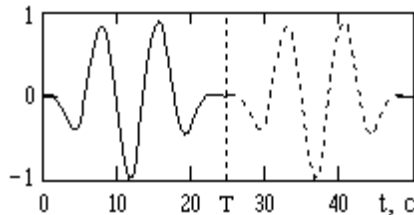


Рис. 18.21.

На рис. 18.21 сплошной кривой приведен пример непрерывного сигнала  $s(t)$ , энергия которого сосредоточена на конечном интервале  $T = (0, 25)$ . Если нас не интересует форма данного сигнала за пределами интервала  $T$ , то спектр сигнала в виде ряда Фурье можно определить по формуле (18.12). При обратном преобразовании Фурье по формуле (18.11), т.е. при восстановлении сигнала по его спектру, в интервале  $T$  будет восстановлен исходный сигнал  $s(t)$ . Но если интервал для восстановления будет задан больше интервала  $T$ , например равным  $0-2T$ , то за пределами этого интервала начнется периодическое повторение исходного сигнала, как это показано пунктиром на рис. 18.23. Если такой процесс нежелателен и за пределами интервала  $T$  должны быть сохранены нулевые значения

сигнала, то необходимо использовать интегральное преобразование Фурье (18.22, 18.23). При этом следует учитывать особенности интегрального преобразования.

Спектральная функция  $S(\omega)$  представляет собой комплексную спектральную плотность сигнала, непрерывную на частотном интервале от  $-\infty$  до  $\infty$ . Если  $s(t)$  – вещественная функция, то спектр этой функции является сопряжено симметричным относительно нулевой частоты

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

и содержит четную действительную и нечетную мнимую части:

$$S(\omega) = A(\omega) - jB(\omega), \quad (18.26)$$

$$A(\omega) = \int_{-\infty}^{\infty} s(t)\cos(\omega t) dt, \quad (18.27)$$

$$B(\omega) = \int_{-\infty}^{\infty} s(t)\sin(\omega t) dt. \quad (18.28)$$

Как и в случае рядов Фурье, вещественные четные функции имеют вещественный четный спектр, представленный только спектральной функцией  $A(\omega)$ , а вещественные нечетные – нечетный и только мнимый спектр, представленный спектральной функцией  $B(\omega)$ .

Пример спектральной функции  $S(f)$  для сигнала  $s(t)$  на рис. 18.21 приведен на рис. 18.22.

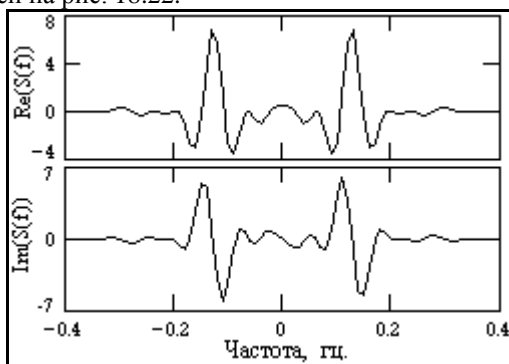


Рис. 18.22.

Как правило, графическое отображение спектральных функций выполняется в виде модуля и аргумента спектральной функции (амплитудного и фазового спектра), приведенных на рис. 18.23.

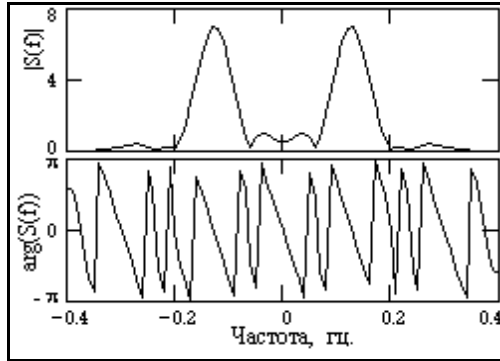


Рис. 18.23.

Такое представление аналогично (18.16):

$$R(\omega) = \sqrt{A^2(\omega) + B^2(\omega)}, \quad (18.29)$$

$$\varphi(\omega) = \arctg(-B(\omega)/A(\omega)), \quad (18.30)$$

но в отношении функции модуля также имеет смысл спектральной плотности модуля.

Заметим также, что сопряженная симметричность спектральной функции позволяет в формулах (18.22)-(18.23) менять местами знаки аргументов в экспонентах, при этом изменяется только знак мнимой части и аргумента спектра.

Еще раз подчеркнем различие между спектрами и спектральными функциями сигналов. При практическом использовании формулы (18.23) для вычисления спектральных функций конечных сигналов, заданных на определенном интервале  $T$ , пределы интегрирования обычно устанавливаются по границам интервала  $T$ , так как нет необходимости выполнять интегрирование в бесконечных пределах, если за пределами интервала  $T$  мы имеем нулевые (или незначимые) значения сигнала. Однако при сравнении формулы (18.23) с выражением (18.12) можно наглядно видеть, что значения интеграла (18.23) не нормируются на величину интервала  $T$ . Отсюда следует, что числовые отсчеты значений модуля функции  $S(\omega)$  для определенных значений  $\omega_i$  не являются амплитудными значениями соответствующих гармонических колебаний с частотой  $\omega_i$ . Значения  $S(\omega)$  по сравнению со значениями функции  $S(n\Delta\omega)$  по (18.12) при  $n\Delta\omega = \omega_i$  завышены на множитель  $T$ . Это можно объяснить тем, что обратное преобразование Фурье по (18.11) представляет собой прямое суммирование гармоник с соответствующими амплитудами колебаний, в то время как интегрирование по (18.22) представляет собой предельное суммирование значений  $S(\omega_i) \cdot d\omega_i$ , где  $d\omega = 2\pi/T$  (или, в

обычном частотном представлении,  $df = 1/T$  при  $T \Rightarrow \infty$ .

Что касается спектра фазовых углов, то значения по (18.30) и по (18.16) при  $n\Delta\omega = \omega_i$  полностью совпадают, так как их вычисление производится по отношению мнимой и действительной части спектра, наличие (или отсутствие) постоянного множителя в которых не меняет значение отношения.

**Тригонометрическая форма** интеграла Фурье (при объединении комплексно сопряженных частей спектральных функций):

$$s(t) = (1/2\pi) \int_{-\infty}^{\infty} [A(\omega)\cos(\omega t) + B(\omega)\sin(\omega t)] d\omega. \quad (18.31)$$

$$s(t) = (1/2\pi) \int_{-\infty}^{\infty} R(\omega)\cos(\omega t - \varphi(\omega)) d\omega. \quad (18.32)$$

Прямое и обратное преобразование Фурье подобны. Любая теорема, доказанная для прямого преобразования Фурье, справедлива и для обратного преобразования, и наоборот. Это непосредственно следует из выражений прямого и обратного преобразования Фурье, которые различаются только знаком в экспоненте. Особенно наглядно (см. рис. 18.24) это видно для четных сигналов (заданных функциями, симметричными относительно  $t=0$ ), для которых  $B(\omega)=0$  и, соответственно, фазовый спектр равен нулю:

$$s(t) = 2 \int_0^{\infty} S(f)\cos(2\pi ft)df, \quad S(f) = 2 \int_0^{\infty} s(t)\cos(2\pi ft)dt.$$

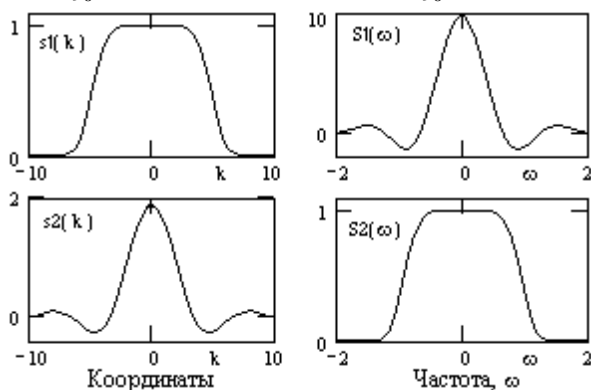


Рис. 18.24.

В математическом анализе для упрощения записей используют символическую форму обозначения преобразования Фурье:

$$s(t) \Leftrightarrow S(f), \quad s(t) \Leftrightarrow S(\omega),$$

где, в общем случае, как фурье-образ функции, так и она сама могут

быть комплексными.

Для физических сигналов и их достаточно корректных математических моделей преобразование Фурье, как правило, всегда существует. С чисто математических позиций сигналу  $s(t)$  можно сопоставить спектральную плотность  $S(\omega)$ , если существует интеграл:

$$\int_{-\infty}^{\infty} |s(t)| dt < \infty. \quad (18.33)$$

**Преобразование Лапласа.** Если условие (18.33) не выполняется, то определенные приближения спектральных плотностей вычисляются с использованием специальных методов, одним из которых является одностороннее преобразование Лапласа.

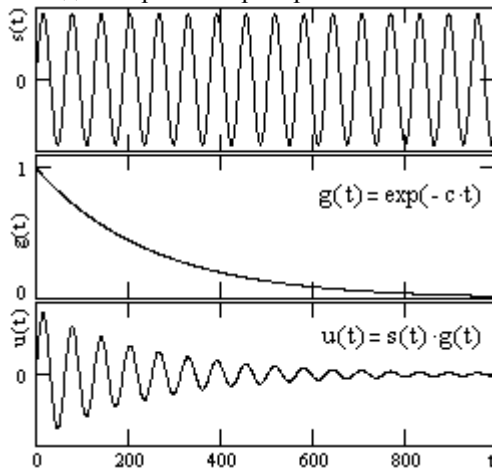


Рис. 18.25.

Допустим, что функция  $s(t)$  задана на интервале  $(0, \infty)$ , равна нулю при  $t < 0$ , а интеграл спектральной функции (18.23) расходится. Умножим  $s(t)$  на экспоненциальную функцию  $\exp(-\sigma t)$ , где  $\sigma$  - положительная константа, и выберем значение  $\sigma$  таким, чтобы произведение  $u(t) = s(t) \cdot \exp(-\sigma t)$  удовлетворяло условию абсолютной интегрируемости. Сущность данной операции хорошо видна на рис. 18.25 ( $\sigma=c$ ). Интегрируемость функции  $u(t)$  может быть установлена для любой функции  $s(t)$  соответствующим выбором коэффициента  $\sigma$ . При этом спектральная плотность функции  $u(t)$  может быть вычислена по формуле (18.23):

$$U(\omega, \sigma) = \int_0^{\infty} [s(t) \exp(-\sigma t)] \exp(-j\omega t) dt.$$

После объединения экспоненциальных функций это



выражение можно переписать следующим образом:

$$U(\sigma+j\omega) = \int_0^{\infty} s(t) \exp[-(\sigma+j\omega)t] dt. \quad (18.34)$$

Соответствующее обратное преобразование Фурье функции  $U(\sigma+j\omega)$ :

$$(1/2\pi) \int_{-\infty}^{\infty} U(\sigma+j\omega) \exp(j\omega t) d\omega = s(t) \exp(-\sigma t).$$

Для восстановления функции  $s(t)$  достаточно умножить обе части данного выражения на  $\exp(\sigma t)$ , объединить экспоненциальные множители под интегралом и заменить переменную интегрирования  $\omega$  на  $\sigma+j\omega$ :

$$s(t) = (1/2\pi j) \int_{\sigma-j\infty}^{\sigma+j\infty} S(\sigma+j\omega) \exp[(\sigma+j\omega)t] d(\sigma+j\omega). \quad (18.35)$$

Обозначим комплексную переменную  $\sigma+j\omega$  в выражениях (18.34, 18.35) через  $p$  (оператор Лапласа) и получим общепринятую форму прямого и обратного преобразования Лапласа:

$$S(p) = \int_0^{\infty} s(t) \exp[-pt] dt. \quad (18.36)$$

$$s(t) = (1/2\pi j) \int_{\sigma-j\infty}^{\sigma+j\infty} S(p) \exp(pt) dp. \quad (18.37)$$

Сигнальную функцию  $s(t)$  в преобразованиях Лапласа обычно называют *оригиналом*, а ее спектральную функцию  $S(p)$  - *изображением оригинала*. Пример спектральной функции Лапласа для оригинала - сложного и неограниченного во времени сигнала, состоящего из каузальной суммы трех гармоник, приведен на рис. 18.26. По спектральной функции Лапласа можно выделить эти три основных частоты сигнала и оценить соотношение их амплитуд. Ширина пиков спектральной функции при выделении "чистых" гармоник зависит от значения коэффициента  $\sigma$  и уменьшается при его уменьшении.

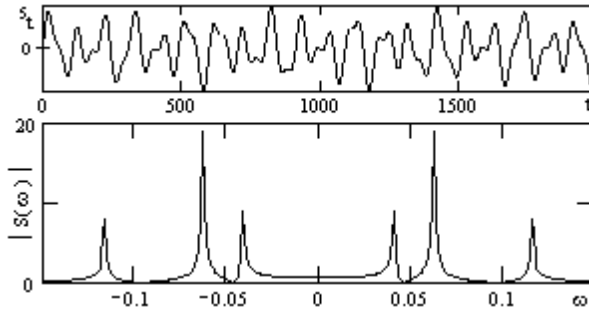


Рис. 18.26. Сигнал и его спектральная функция Лапласа при  $p=0.0005+j\omega$ .

Преобразование Лапласа справедливо только в области сходимости интеграла (18.34), которая определяется абсциссой абсолютной сходимости  $\sigma_0$  (при  $\sigma \geq \sigma_0$ ):

$$\int_0^{\infty} |s(t) \exp(-(\sigma+j\omega)t)| dt = \int_0^{\infty} |s(t)| |\exp(-j\omega t)| \exp(-\sigma t) dt = \int_0^{\infty} |s(t)| \exp(-\sigma t) dt < \infty.$$

Если вместо  $p$  в изображениях оригинала подставить переменную  $j\omega$ , то будут получены спектральные функции, полностью идентичные преобразованию Фурье каузальных функций (имеющих нулевые значения при  $t < 0$ ).

**Обобщенный ряд Фурье.** Тригонометрические функции не является единственно возможными функциями разложения сигналов. В общем случае разложение сигнала  $s(t)$  на интервале  $(a, b)$  в ряд вида  $\sum_{k=0}^N c_k \phi_k(t)$  может быть выполнено по произвольным функциям  $\phi_k(t)$ . При задании минимальной погрешности приближения

$$\Delta s = \int_a^b [s(t) - \sum_{k=0}^N c_k \phi_k(t)]^2 dt$$

коэффициенты  $c_k$  находятся из системы линейных уравнений:

$$\frac{\partial \Delta s}{\partial c_k} = \int_a^b \frac{\partial}{\partial c_k} [s(t) - \sum_{k=0}^N c_k \phi_k(t)]^2 dt = 0, \quad k = 0, 1, 2, \dots, N.$$

При линейной независимости функций  $\phi_k(t)$  данная система уравнений имеет единственное решение. Если все функции  $\phi_k(t)$  взаимно ортогональны и соответствующей нормировкой обеспечена их ортонормированность

$$\int_a^b \varphi_m(t) \varphi_n(t) dt = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases},$$

то процесс нахождения коэффициентов  $c_k$  оказывается наиболее простым:

$$c_k = \int_a^b s(t) \varphi_k(t) dt,$$

и для принятого значения  $N$  погрешность приближения  $\Delta s$  является минимальной. Если при  $N \rightarrow \infty$  имеет место  $\Delta s \rightarrow 0$ , система функций  $\varphi_k(t)$  называется базисной системой координат пространства сигналов  $L^2[a, b]$ . При этом имеет место равенство:

$$s(t) = \sum_{k=0}^{\infty} c_k \varphi_k(t).$$

Разложение по ортонормированной системе базисных функций называется *обобщенным рядом Фурье*, а набор коэффициентов  $c_k$  представляет собой спектр функции  $s(t)$  в соответствующем базисе. В зависимости от специфики решаемых задач применяются различные системы базисных функций. В частности, используются разложения по полиномам Лежандра, Чебышева, Лагерра, Эрмита, функциям Хаара и Уолша и т.п.

## 18.6. Свойства преобразований Фурье

Свойствами преобразований Фурье определяется взаимное соответствие трансформации сигналов и их спектров.

**1. Линейность.** Преобразование Фурье относится к числу линейных интегральных операций, т.е. спектр суммы сигналов равен сумме спектров этих сигналов.

$$\sum_n a_n s_n(t) \Leftrightarrow \sum_n a_n S_n(\omega). \quad (18.38)$$

Пример суммирования сигналов и его отображения в суммирования спектров приведен на рис. 18.27:

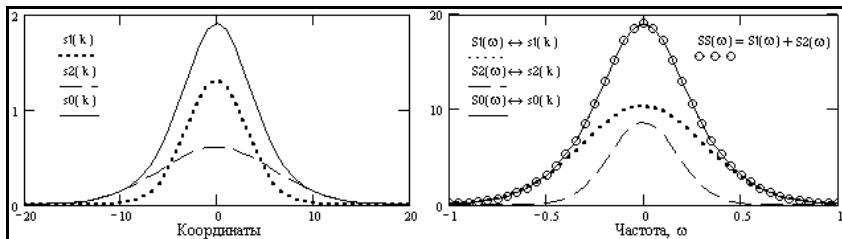


Рис. 18.27. Сигналы и их спектры.

$$s_0(k)=s_1(k)+s_2(k) \Leftrightarrow S_1(\omega)+S_2(\omega) = S_0(\omega).$$

Сигнал s(t)	Спектр S(ω)
Четный	Вещественный, четный
Нечетный	Мнимый, нечетный
Произвольный	Действительная часть – четная. Мнимая часть - нечетная

**2. Свойства четности** преобразования определяются косинусными (четными, действительными) и синусными (нечетными, мнимыми) частями разложения и подобием прямого и обратного преобразований. На рис. 18.28 приведены примеры, поясняющие свойства четности преобразования.

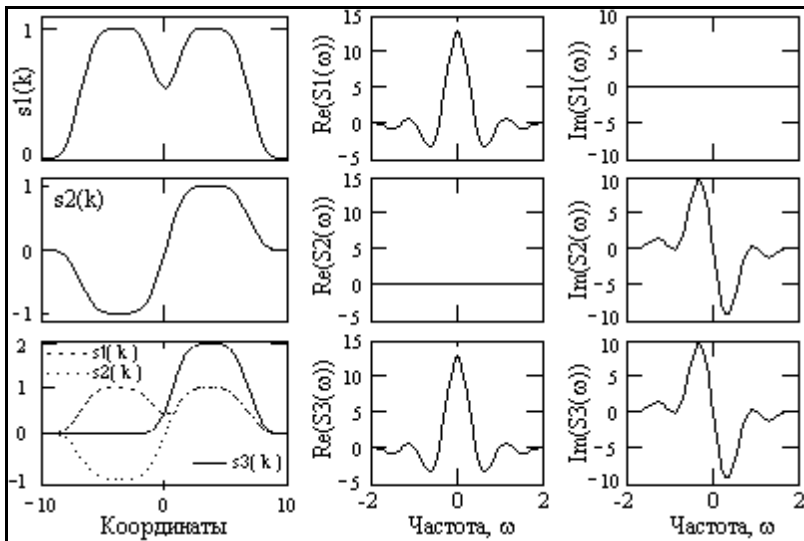


Рис. 18.28.

Сигнал  $s_1(k)$  является четным,  $s_1(k) = s_1(-k)$ , и имеет только вещественный четный спектр (мнимая часть спектральной функции представлена нулевыми значениями). Сигнал  $s_2(k) = -s_2(-k)$  нечетный и имеет мнимый нечетный спектр, а нулевыми значениями представлена его действительная часть. Сигнал  $s_3(k)$  образован суммой сигналов  $s_1(k)$  и  $s_2(k)$ . Соответственно, спектральная функция сигнала представлена и действительной четной частью (принадлежащей  $s_1(k)$ ), и мнимой нечетной частью (принадлежащей  $s_2(k)$ ). При обратном преобразовании Фурье отдельно действительной и мнимой части спектра  $S_3(\omega)$ , равно как и любых других комплексных спектров, будут

раздельно восстановлены четная и нечетная части исходного сигнала.

Заметим, что произвольный исходный сигнал может быть задан в одностороннем варианте (в интервале 0-T), но четная и нечетная части этого сигнала занимают интервал от -T до T, при этом на левой половине числовой оси (от -T до 0) эти два сигнала компенсируют друг друга, давая нулевые значения.

**3. Изменение аргумента функции** (сжатие или расширение сигнала) приводит к обратному изменению аргумента ее фурье-образа и обратно пропорциональному изменению его модуля. Действительно, если  $s(t) \leftrightarrow S(\omega)$ , то при изменении длительности сигнала с сохранением его формы (растяжении сигнала по временной оси), т.е. для сигнала с новым аргументом  $s(x)=s(at)$  при  $x=at$ , получаем:

$$\begin{aligned} s(at) &\Leftrightarrow \int_{-\infty}^{\infty} s(at)\exp(-j\omega t) dt = \\ &=(1/a) \int_{-\infty}^{\infty} s(x)\exp(-jx\omega/a) dx \quad s(at) \Leftrightarrow (1/a) S(\omega/a). \end{aligned} \quad (18.39)$$

Выражение (18.39) действительно при  $a>0$ . При  $a<0$  происходит зеркальный поворот сигнала относительно вертикальной оси, а замена переменной  $t=x/a$  вызывает перестановку пределов интегрирования и, соответственно, изменение знака спектра:

$$s(at) \Leftrightarrow -(1/a) S(\omega/a). \quad (18.40)$$

Обобщенная формула изменения аргумента:

$$s(at) \Leftrightarrow (1/|a|) S(\omega/a), \quad a \neq 0 \quad (18.41)$$

Если под аргументом функции и ее спектра понимать определенные физические единицы, например, время - частота, то отсюда следует: чем короче по своей длительности сигнал, тем шире по частоте его спектр, и наоборот. Это можно наглядно видеть на рис. 18.27. для сигналов  $s_1(k)$  и  $s_2(k)$  и их спектров  $S_1(\omega)$  и  $S_2(\omega)$ .

От изменения аргумента функций следует отличать изменение масштаба представления функций. Изменение масштаба аргументов изменяет только оцифровку числовых осей отображения сигналов и их спектров, но не изменяет самих сигналов и спектров. Так, при масштабе оси времен  $t=1$  секунда, масштаб оси частот  $f=1/t=1$  герц, а при  $t=1$  мксек  $f=1/t=1$  МГц ( $t=at$ ,  $f=1/at$ ,  $a=10^{-6}$ ).

**4. Теорема запаздывания.** Запаздывание (сдвиг, смещение) сигнала по аргументу функции на интервал  $t_0$  приводит к изменению фазочастотной функции спектра (фазового угла всех гармоник) на величину  $-\omega t_0$  без изменения модуля (амплитудной функции) спектра. Применяя замену переменной  $t-t_0 = x$ , получаем:

$$s(t-t_0) \Leftrightarrow \int_{-\infty}^{\infty} s(t-t_0) \exp(-j\omega t) dt = \int_{-\infty}^{\infty} s(x) \exp(-j\omega x) \exp(-j\omega t_0) dx = S(\omega) \exp(-j\omega t_0) \quad (18.42)$$

Совершенно очевидно, что амплитуды гармоник сигнала при его сдвиге изменяться не должны. С учетом того, что  $|\exp(-j\omega t_0)|=1$ , это следует и из (18.42):

$$|S(\omega) \exp(-j\omega t_0)| = |S(\omega)|.$$

Фазовый спектр сдвигается на  $-\omega t_0$  с линейной зависимостью от частоты:

$$S(\omega) \exp(-j\omega t_0) = R(\omega) \exp[j(\varphi(\omega))] \exp(-j\omega t_0) = R(\omega) \exp[j(\varphi(\omega) - \omega t_0)]. \quad (18.43)$$

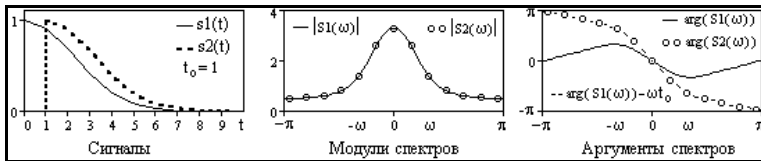


Рис. 18.29. Изменение спектра сигнала при его сдвиге.

Пример двух одинаковых сигналов, сдвинутых относительно друг друга на  $t_0=1$ , и соответствующих данным сигналам спектров приведен на рис. 18.29.

Аналогично нетрудно показать, что сдвиг спектра в частотной области на  $\omega_0$  вызывает умножение сигнала на  $\exp(j\omega_0 t)$ :

$$S(\omega - \omega_0) \Leftrightarrow s(t) \exp(j\omega_0 t),$$

что эквивалентно модуляции сигналом комплексной экспоненты во временной области.

**5. Преобразование производной** (дифференцирование сигнала):

$$\begin{aligned} s(t) = d[y(t)]/dt &= d\left[ \int_{-\infty}^{\infty} Y(\omega) \exp(j\omega t) d\omega \right] / dt = \\ &= \int_{-\infty}^{\infty} Y(\omega) [d(\exp(j\omega t))/dt] d\omega = \\ &= \int_{-\infty}^{\infty} j\omega Y(\omega) \exp(j\omega t) d\omega \Leftrightarrow j\omega Y(\omega). \end{aligned} \quad (18.44)$$

Таким образом, дифференцирование сигнала отображается в спектральной области простым умножением спектра сигнала на оператор дифференцирования сигнала в частотной области  $j\omega$ , что эквивалентно дифференцированию каждой гармоники спектра.

Умножение на  $j\omega$  приводит к обогащению спектра производной сигнала высокочастотными составляющими (по сравнению с исходным сигналом) и уничтожает составляющие с нулевой частотой.

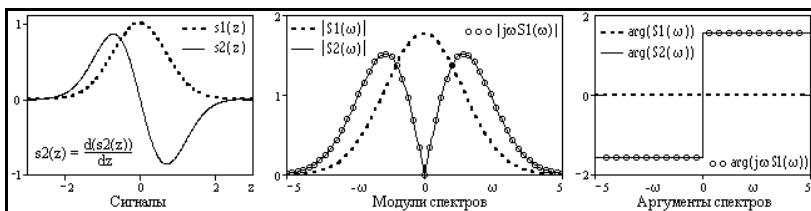


Рис. 18.30. Спектры сигнала и его производной.

Пример сигнала, его производной и соответствующих им спектров приведен на рис. 18.30. По изменению аргумента спектра (для четного исходного сигнала он был нулевым) можно видеть, что для всех гармоник спектра появляется сдвиг фаз на  $\pi/2$  ( $90^\circ$ ) для положительных частот, и на  $-\pi/2$  ( $-90^\circ$ ) для отрицательных частот.

В общем случае, для кратных производных:

$$d^n[y(t)]/dt^n = (j\omega)^n Y(\omega). \quad (18.45)$$

При дифференцировании спектра функции соответственно получаем:

$$d^n[S(\omega)]/d\omega^n = (-jt)^n s(t).$$

**6. Преобразование интеграла** сигнала в частотной области при известном спектре сигнала может быть получено из следующих простых соображений. Если имеет место

$$s(t) = d[y(t)]/dt \Leftrightarrow j\omega Y(\omega) = S(\omega),$$

то должна выполняться и обратная операция:

$$y(t) = \int s(t) dt \Leftrightarrow Y(\omega) = S(\omega)/j\omega.$$

Отсюда следует:

$$\int s(t)dt \Leftrightarrow (1/j\omega)S(\omega). \quad (18.46)$$

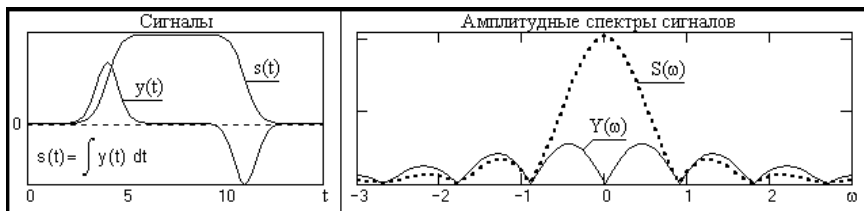


Рис. 18.31. Сигналы и амплитудные спектры сигналов.

Оператор интегрирования в частотной области  $(1/j\omega)$  при  $\omega > 1$

ослабляет в амплитудном спектре высокие частоты и при  $\omega < 1$  усиливает низкие. Фазовый спектр сигнала смещается на  $-90^\circ$  для положительных частот и на  $90^\circ$  для отрицательных. Пример модуля спектра сигнала и его интегральной функции приведены на рис. 18.31.

Формула (18.46) справедлива для сигналов с нулевой постоянной составляющей. При интегрировании сигналов с определенным значением постоянной составляющей  $C = \text{const}$  в правой части выражения (18.46) появляется дополнительное слагаемое преобразования Фурье постоянной составляющей  $C$ , которое представляет собой дельта-функцию на нулевой частоте с весовым коэффициентом, равным значению  $C$ :

$$(1/j\omega)S(\omega) + C \cdot \delta(0).$$

**7. Преобразование свертки** сигналов  $y(t) = s(t) * h(t)$ :

$$Y(\omega) = \int_{-\infty}^{\infty} y(t) \exp(-j\omega t) dt = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\tau) h(t-\tau) \exp(-j\omega t) dt d\tau.$$

$$Y(\omega) = \int_{-\infty}^{\infty} s(\tau) d\tau \int_{-\infty}^{\infty} h(t-\tau) \exp(-j\omega t) dt.$$

По теореме запаздывания (18.42):

$$\int_{-\infty}^{\infty} h(t-\tau) \exp(-j\omega t) dt = H(\omega) \exp(-j\omega \tau).$$

Отсюда:

$$Y(\omega) = \int_{-\infty}^{\infty} H(\omega) s(\tau) \exp(-j\omega \tau) d\tau = H(\omega) \cdot S(\omega).$$

$$s(t) * h(t) \Leftrightarrow S(\omega) H(\omega). \tag{18.47}$$

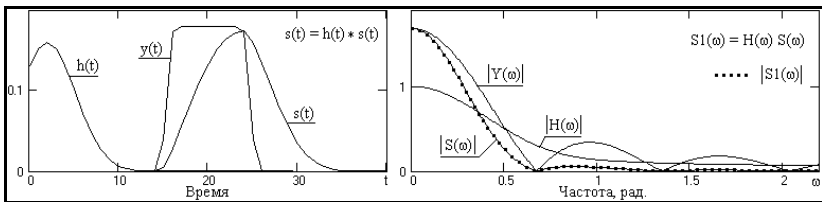


Рис. 18.32. Сигналы и амплитудные спектры сигналов.

Пример выполнения свертки в частотной области приведен на рис. 18.32. Отметим, что частотное представление  $H(\omega)$  импульсного отклика  $h(t)$  линейной системы (или соответствующей линейной операции) имеет смысл частотной передаточной функции системы и позволяет определить сигнал на выходе системы (в частотной форме представления) при задании произвольного сигнала (в частотной форме) на ее входе. По существу, функция  $H(\omega)$  представляет собой



распределение по частоте коэффициента пропускания частотных составляющих сигнала с входа на выход системы (операции).

Таким образом, *свертка функций в координатной форме отображается в частотном представлении произведением фурье-образов этих функций.*

Это положение имеет фундаментальное значение в практике обработки данных.

Любая линейная система обработки данных (информационных сигналов) реализует определенную операцию трансформации сигнала, т.е. выполняет операцию свертки входного сигнала  $s(t)$  с оператором системы  $h(\tau)$ . С использованием преобразования свертки эта операция может производиться как с динамической, так и с частотной формой представления сигналов. При этом обработка данных, представленных в цифровой форме, производится, как правило, в частотной области, т.к. может быть на несколько порядков выше по производительности, чем во временной области. Она представляет собой последовательность следующих операций.

1. Перевод сигнала в частотную область:  $s(t) \Leftrightarrow S(\omega)$ .

2. Умножение спектра обработанного сигнала на передаточную функцию системы:  $Y(\omega) = H(\omega) \cdot S(\omega)$ .

Передаточная функция системы определяется аналогичным преобразованием  $h(\tau) \Leftrightarrow H(\omega)$  или задается непосредственно в частотном представлении, что позволяет задавать передаточные функции сколь угодно сложной формы, в том числе с разрывами и скачками, для которых во временной области потребуются операторы  $h(\tau)$  с бесконечной импульсной характеристикой.

3. Перевод спектра обработанного сигнала во временную область:  $Y(\omega) \Leftrightarrow y(t)$ .

**8. Преобразование произведения** сигналов  $y(t) = s(t) \cdot h(t)$ :

$$\begin{aligned} Y(\omega) &= \int_{-\infty}^{\infty} s(t) h(t) \exp(-j\omega t) dt = \\ &= \int_{-\infty}^{\infty} s(t) \left[ (1/2\pi) \int_{-\infty}^{\infty} H(\omega') \exp(j\omega' t) d\omega' \right] dt = \\ &= (1/2\pi) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(t) H(\omega') \exp(-j(\omega - \omega')t) d\omega' dt = \\ &= (1/2\pi) \int_{-\infty}^{\infty} H(\omega') d\omega' \int_{-\infty}^{\infty} s(t) \exp(-j(\omega - \omega')t) dt = \\ &= (1/2\pi) \int_{-\infty}^{\infty} H(\omega') S(\omega - \omega') d\omega' = (1/2\pi) H(\omega) * S(\omega). \end{aligned} \quad (18.48)$$

Таким образом, *произведение функций в координатной форме отображается в частотном представлении сверткой фурье-образов этих функций*, с нормировочным множителем  $(1/2\pi)$ , учитывающем несимметричность прямого и обратного преобразования Фурье функций  $s(t)$  и  $h(t)$  при использовании угловых частот.

**9. Производная свертки** двух функций

$$s'(t) = d[x(t) * y(t)]/dt.$$

С использованием выражений (18.45) и (18.47), получаем:

$$s'(t) = j\omega [X(\omega) Y(\omega)] = (j\omega X(\omega)) Y(\omega) = X(\omega) (j\omega Y(\omega)).$$

$$s'(t) = x'(t) * y(t) = x(t) * y'(t).$$

Это выражение позволяет выполнять вычисление производной сигнала с одновременным сглаживанием весовой функцией, которая является производной сглаживающей функции (например, гауссиана).

**10. Спектры мощности.** Временная функция мощности сигнала в общей форме определяется выражением:

$$w(t) = s(t) s^*(t) = |s(t)|^2.$$

Спектральная плотность мощности, соответственно, равна преобразованию Фурье произведения  $s(t) \cdot s^*(t)$ , которое отобразится в спектральном представлении сверткой Фурье-образов этих функций:

$$W(f) = S(f) * S^*(f) = \int_{-\infty}^{\infty} S(f) S^*(f-v) dv. \quad (18.49)$$

Но для всех текущих значений частоты  $f$  интеграл в правой части этого выражения равен произведению  $S(f) \cdot S^*(f)$ , так как для всех значений сдвига  $v \neq 0$  в силу ортогональности гармоник  $S(f)$  и  $S^*(f-v)$  значения их произведения равны нулю. Отсюда:

$$W(f) = S(f) * S^*(f) = |S(f)|^2. \quad (18.50)$$

**Спектр мощности** - вещественная неотрицательная четная функция, которую очень часто называют *энергетическим спектром*. **Спектр мощности, как квадрат модуля спектра сигнала**, не содержит фазовой информации о частотных составляющих, а, следовательно, восстановление сигнала по спектру мощности невозможно. Это означает также, что сигналы с различными фазовыми характеристиками могут иметь одинаковые спектры мощности. В частности, сдвиг сигнала не отражается на его спектре мощности.

Для функций мощности взаимодействия сигналов в частотной области соответственно имеем частотные спектры мощности взаимодействия сигналов:

$$W_{xy}(f) = X(f) Y^*(f),$$

$$W_{yx}(f) = Y(f) X^*(f),$$

$$W_{xy}(f) = W_{yx}^*(f).$$

Функции мощности взаимодействия сигналов комплексные,

даже если обе функции  $x(t)$  и  $y(t)$  вещественны, при этом  $\text{Re}[W_{xy}(f)]$  - четная функция, а  $\text{Im}[W_{xy}(f)]$  - нечетная. Отсюда **полная энергия взаимодействия сигналов при интегрировании функций мощности взаимодействия определяется только реальной частью спектра:**

$$E_{xy} = (1/2\pi) \int_{-\infty}^{\infty} W_{xy}(\omega) d\omega = (1/\pi) \int_{-\infty}^{\infty} \text{Re}[W_{xy}] d\omega,$$

и всегда является **вещественным числом.**

**11. Равенство Парсеваля.** Полная энергия спектра сигнала:

$$E_s = \int_{-\infty}^{\infty} W(f) df = \int_{-\infty}^{\infty} |S(f)|^2 df. \quad (18.51)$$

Так как координатное и частотное представление по существу только разные математические отображения одного и того же сигнала, то равной должна быть и энергия сигнала в двух представлениях, откуда следует равенство Парсеваля:

$$\int_{-\infty}^{\infty} |s(t)|^2 dt = \int_{-\infty}^{\infty} |S(f)|^2 df,$$

т.е. **энергия сигнала равна интегралу модуля его частотного спектра - сумме энергий всех частотных составляющих сигнала.**

Аналогично для энергии взаимодействия сигналов:

$$\int_{-\infty}^{\infty} x(t) y^*(t) dt = \int_{-\infty}^{\infty} X(f) Y^*(f) df.$$

Из равенства Парсеваля следует инвариантность скалярного произведения сигналов и нормы относительно преобразования Фурье:

$$\langle x(t), y(t) \rangle = \langle X(f), Y(f) \rangle, \quad \|x(t)\|^2 = \|X(f)\|^2.$$

Не следует забывать, что при представлении спектров в круговых частотах (по  $\omega$ ) в правой части приведенных равенств должен стоять множитель  $1/2\pi$ .

## 18.7. Спектры некоторых сигналов

**1. Единичные импульсы.** Функция  $\delta(t)$ , центрированная относительно  $t = 0$ , значения которой по определению равны нулю при  $t \neq 0$ , а интеграл от  $-\infty$  до  $\infty$  равен 1, имеет равномерное спектральное распределение в бесконечной полосе частот от 0 до  $\infty$ :

$$\text{TF}[\delta(t)] = \int_{-\infty}^{\infty} \delta(t) \exp(-j\omega t) dt = 1. \quad (18.52)$$

Это следует и из свойства свертки функций, поскольку свертка функции с бесконечно коротким импульсом с единичной площадью не должна приводить к изменению функции:

$$s(t) * \delta(t) = s(t).$$

Выполняя преобразование Фурье правой и левой части данного выражения, имеем:

$$S(\omega) H(\omega) = S(\omega),$$

что может быть реализовано только при  $H(\omega) = 1$ .

Отсюда следует также, что дельта-функцию можно записать в виде обратного преобразования Фурье:

$$\delta(t) = (1/2\pi) \int_{-\infty}^{\infty} \exp(j\omega t) d\omega. \quad (18.53)$$

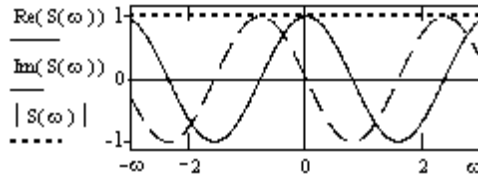


Рис. 18.33. Спектр функции  $\delta(t-2)$

С учетом теоремы запаздывания (18.42), для обобщенной функции Дирака соответственно имеем:

$$\delta(t-\tau) \leftrightarrow \exp(-j\omega\tau). \quad (18.54)$$

$$\delta(t-\tau) = (1/2\pi) \int_{-\infty}^{\infty} \exp(j\omega(t-\tau)) d\omega. \quad (18.55)$$

Пример спектра функции приведен на рис. 18.33.

Для единичного импульса с площадью, равной  $P$ :

$$P \cdot \delta(t) \leftrightarrow P.$$

Если спектром весовой дельта-функции является константа, то на основе дуальности преобразования Фурье спектром константы должна быть весовая дельта-функция в нуле частотной оси.

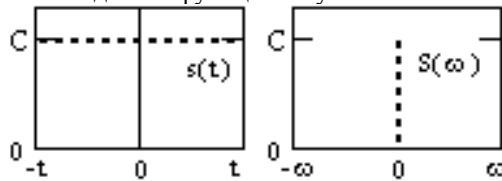


Рис. 18.34.

$$C \leftrightarrow C \cdot \delta(\omega).$$

Представить графически эту операцию для непрерывных функций невозможно. Но для дискретных спектральных функций с использованием весового импульса Кронекера она имеет вполне реальный смысл (рис. 18.34). С учетом дуальности преобразования Фурье, для  $\delta$ -функций в спектральной области соответственно имеем:

$$\delta(\omega-\omega_0) = \int_{-\infty}^{\infty} \exp(j(\omega-\omega_0)t) dt. \quad (18.56)$$

**2. Гребневая функция**  $\text{Ш}_T(t)$  представляет собой последовательность импульсов Дирака с периодом  $T=1/F$ , где  $F$ -частота следования импульсов:

$$\text{Ш}_T(t) = \sum_{k=-\infty}^{\infty} \delta(t-kT).$$

Спектр гребневой функции (с учетом теоремы запаздывания при  $\Delta f=1/T=F$ ) также представляет собой последовательность импульсов Дирака:

$$\text{Ш}_T(t) = (1/T) \sum_{k=-\infty}^{\infty} \exp(-2\pi j n \Delta f t) \Leftrightarrow (1/T) \sum_{k=-\infty}^{\infty} \delta(f-kF) = F \cdot \text{Ш}_F(f). \quad (18.57)$$

**3. Спектр прямоугольного импульса**  $\Pi_r(t)$  амплитудой  $U$  и длительностью  $\tau$  (рис. 18.35). При расположении начала координат по центру импульса:

$$\begin{aligned} \Pi_r(\omega) &= \int_{-\infty}^{\infty} \Pi_r(t) \exp(-j\omega t) dt = U \int_{-r/2}^{r/2} \exp(-j\omega t) dt, \\ \Pi_r(\omega) &= rU \sin(\omega r/2) / (\omega r/2) = rU \text{sinc}(\omega r/2). \quad (18.58) \end{aligned}$$

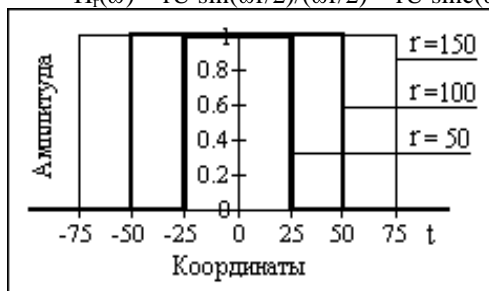


Рис. 18.35. П - импульсы

Вид функций  $\Pi_r(\omega)$  приведен на рис. 18.36.

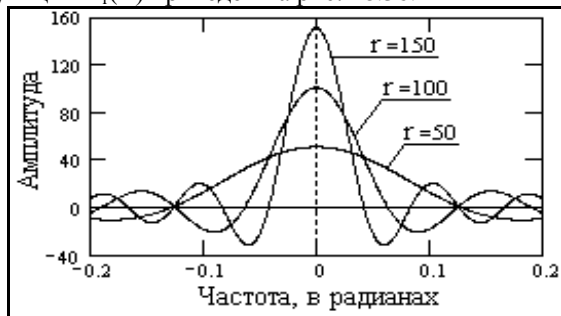


Рис. 18.36. Спектры П - импульсов.

Представлена только часть частотного диапазона, в остальной

части диапазона постепенно затухающие флюктуации спектров, которые, чисто теоретически, простираются до бесконечности.

Как и следовало ожидать, для вещественной и четной динамической П-функции спектр сигнала также является вещественной и четной функцией частоты. Спектр имеет лепестковый характер, и ширина лепестков (по пересечениям нулевой линии) обратно пропорциональна длительности импульсов и равна  $2\pi/r$ . Значение спектральной плотности на нулевой частоте равно площади импульсов.

Функция вида  $\sin(x)/x$  в анализе сигналов встречается довольно часто и имеет специальное обозначение:  $\text{sinc}(x) = \sin(x)/x$ . Она называется *интегральным синусом или функцией отсчетов*.

На рис. 18.37 приведены нормированные по площади спектры этих же импульсов.

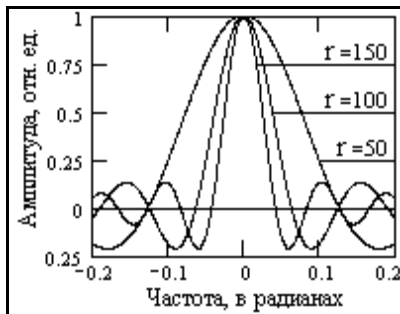


Рис. 18.37. Спектры П-импульсов

При сравнении спектров с рис. 18.35 можно наглядно видеть характер зависимости ширины спектров (по ширине главного максимума) от длительности импульсов. Нетрудно также заметить, что форма спектра П - импульсов остается практически постоянной и только "растягивается" по шкале частоты при уменьшении длительности импульсов. Чем шире сигнал, тем короче его спектр.

Если прямоугольный импульс начинается в момент времени  $t_0$ , то имеем:

$$\Pi(\omega) = U \int_{t_0}^{t_0+r} \exp(-j\omega t) dt = rU \text{sinc}(\omega r/2) \exp[-j\omega(t_0-r/2)]. \quad (18.59)$$

Это выражение может быть получено непосредственно из (18.58) с использованием теоремы сдвига. Вид функций  $\Pi(\omega)$  при  $r = 50$  и  $t_0 = 50$  приведен на рис. 18.38.

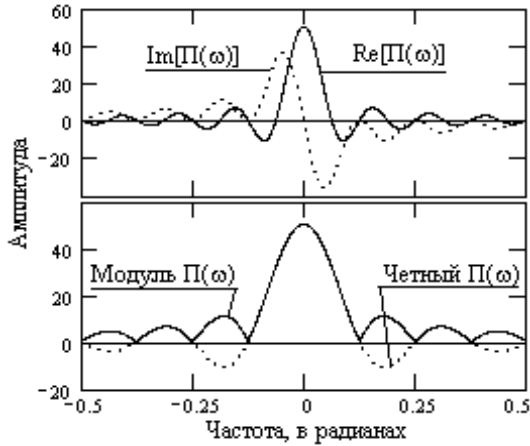


Рис. 18.38. Спектр задержанного П-импульса

Как видно на рисунке, спектр сигнала, несимметричного относительно  $t = 0$ , имеет две части: четную действительную  $A(\omega) = \text{Re}(\Pi(\omega))$ , и нечетную мнимую  $B(\omega) = \text{Im}(\Pi(\omega))$ . Модуль спектра  $R(\omega) = |\Pi(\omega)|$  всегда четный, имеет только положительные значения и полностью повторяет  $|\Pi_c(\omega)|$  четного импульса (при смещении начала координат в центр импульса).

При изменении величины сдвига импульса модуль спектра остается без изменений, т.к. амплитуда частотных составляющих сигнала зависит только от его формы и не меняется от места расположения сигнала на координатной оси. Сдвиг сигнала определяет его фазовый спектр, пример которого для задержанного П-импульса приведен на рис. 18.39.

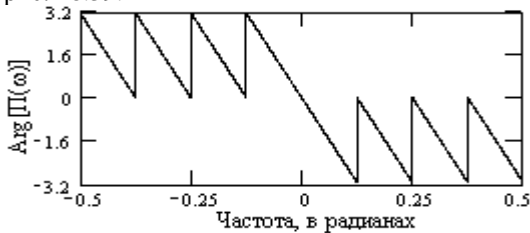


Рис. 18.39. Фазовый спектр задержанного П-импульса ( $t_0 = 50$ ,  $\tau = 50$ )

Заметим, что фактический фазовый спектр имеет непрерывный характер. Пилообразная форма кривых на рис. 18.39 объясняется периодическим сбросом действительных значений фазы сигнала на величину  $\pi$ .

Учитывая, что значения функций на отрицательных частотах

спектра всегда в одном и том же порядке однозначно соотносится со значениями на положительных частотах (четные функции  $A(\omega)$  и  $R(\omega)$ , нечетные функции  $B(\omega)$  и  $\varphi(\omega)$ ), в дальнейшем двусторонние спектры сигналов будем приводить только для области положительных частот.

Для количественной характеристики соотношения формы сигналов и их спектров применяют понятие *базы сигнала*, под которой понимают произведение эффективных значений длительности сигнала и ширины его спектра. Конкретное значение базы сигнала зависит от способа определения этих параметров. Для сигналов простой формы значение базы обычно составляет несколько единиц. Если для прямоугольного импульса эффективную ширину спектра принять по длительности центрального пика ( $2\pi/\tau$ ), то значение базы сигнала будет равно  $2\pi$ .

Длительности сигналов и ширина их спектров связаны *принципом неопределенности*, которым устанавливается, что значение их произведения (база сигналов) не может быть меньше 1. Этим устанавливается, что не может существовать коротких сигналов с узким спектром. Напротив, максимальное значение базы сигналов не ограничивается, и могут существовать *сигналы с большой базой*, имеющие как большую длительность, так и широкий спектр.

Если прямоугольные импульсы повторяются с периодом  $T$ , то соответственно при  $\Delta\omega = 1/T$  имеем:

$$P_r(k\Delta\omega) = (rU/T) \operatorname{sinc}(k\Delta\omega\tau/2) \exp(-jk\Delta\omega(t_0 - \tau/2)). \quad (18.60)$$

Как и положено, спектр периодического сигнала дискретен по  $\omega$ , а при снятии нормировки спектра на длительность периода (умножением на  $T$ ) огибающая спектра повторяет выражение (18.59).

**4. Треугольные импульсы** длительностью  $\tau$  по основанию с площадью, равной  $P$ , могут быть получены сверткой двух прямоугольных импульсов длительностью  $\tau/2$  с амплитудой  $2P/\tau$ , откуда:

$$s(t) = \Pi_{\tau/2}(t) * \Pi_{\tau/2}(t) \Leftrightarrow S(\omega) = \Pi_{\tau/2}(\omega) \Pi_{\tau/2}(\omega), \\ S(\omega) = P \operatorname{sinc}^2(\omega\tau/4).$$

Спектр треугольного импульса также имеет лепестковую структуру с шириной лепестков  $4\pi/\tau$ . Соответственно, база треугольного импульса равна  $4\pi$ . Спектральная функция за счет квадратирования интегрального синуса имеет только положительные значения. Аналогично можно получить и спектры трапеций (при разной длительности  $\Pi$ -импульсов).

Примеры импульсов и сопоставление формы их нормированных спектров (делением значений  $S(\omega)$  на площадь импульсов - значение  $S(0)$ ) приведены на рис. 18.40.



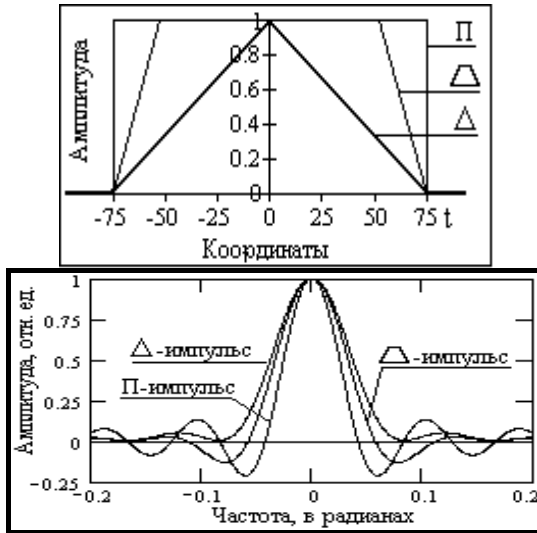


Рис. 18.40. Форма и спектры импульсов.

Заметим, что обратная операция – аппроксимация спектра сигнала произведением спектров простых сигналов с последующим переводом спектров в координатную область, позволяет представить сложный исходный сигнал в виде свертки более простых сигналов.

**5. Экспоненциальный импульс**  $s(t) = U \exp(-at), t \geq 0, a > 0$ .

Функция  $\exp(-at)$  только условно может быть названа импульсом, т.к. определена и при  $t \Rightarrow \infty$ , но при  $a > 0$  она достаточно быстро затухает. Преобразование Фурье экспоненциального импульса, как произвольного одностороннего сигнала, имеет действительную и мнимую части:

$$S(\omega) = U \int_0^{\infty} \exp(-(a+j\omega)t) dt = U/(a+j\omega). \quad (18.61)$$

Функция  $S(\omega)$  бесконечна по частоте. Форма импульса, модуль и аргумент его спектра (фазовая характеристика в градусах) приведены на рис. 18.41.

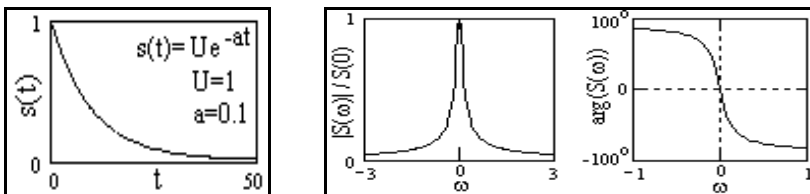


Рис. 18.41. Форма и спектр экспоненциального импульса.

**6. Функции Лапласа и Гаусса.** Для функции Лапласа (экспонента по модулю t) имеем следующее преобразование:

$$U \exp(-a|t|) \leftrightarrow 2aU/(a^2+\omega^2), a>0. \quad (18.62)$$

Форма функции (при U=1, a=0.1) и ее вещественный спектр (функция четная и представлена только действительной частью) приведены на рис. 18.42.

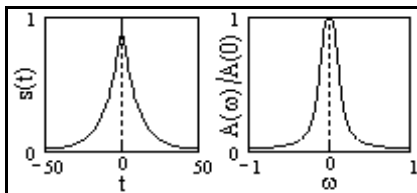


Рис. 18.42. Функция Лапласа

Преобразование для центрированной функции Гаусса:

$$U \exp(-pt^2) \leftrightarrow U \sqrt{\pi/p} \exp(-\omega^2/4p). \quad (18.63)$$

Спектр центрированной функции Гаусса - также функция Гаусса. Форма функции (при U=1, a=0.0.003) и ее вещественный спектр приведены на рис. 18.43.

Если эффективную длительность и ширину спектра гауссовых функций определять по уровню 1/e от максимума ( $\tau=2/a$ ,  $\Delta\omega=a$ ), то база сигнала равна 4.

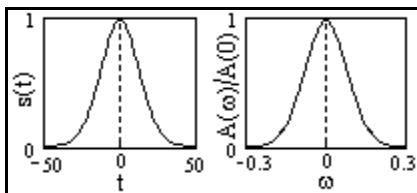


Рис. 18.43. Функция Гаусса.

Сравнивая на рисунках 18.42 и 18.43 функции Лапласа и Гаусса и их спектры (с учетом масштаба последних), нетрудно заметить, что чем более плавно изменяются значения сигнала (меньше его дифференциал), тем более низкочастотным является спектр сигнала.

**7. Гармонические колебания.** Одним из условий применения интегрального преобразования Фурье является абсолютная интегрируемость функций. Гармонические, а в общем случае и все периодические функции в пространстве  $R(-\infty, \infty)$ , не обладают условием абсолютной интегрируемости. Спектральные плотности таких сигналов обычно определяются с использованием  $\delta$ -

функций.

Допустим, имеем простейший периодический сигнал:

$$s(t) = A_o \cos \omega_o t.$$

Разложим сигнал по формуле Эйлера и выполним преобразование Фурье, не обращая внимания на неинтегрируемость функции:

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt = (A_o/2) \int_{-\infty}^{\infty} [\exp(j\omega_o t) + \exp(-j\omega_o t)] \exp(-j\omega t) dt = \\ &= (A_o/2) \int_{-\infty}^{\infty} \exp(-j(\omega - \omega_o)t) dt + (A_o/2) \int_{-\infty}^{\infty} \exp(j(\omega + \omega_o)t) dt. \end{aligned}$$

Но интегралы в этом выражении представляют собой  $\delta$ -функции в частотной области (см. формулу (18.56)) с бесконечно большими значениями спектральной плотности на частотах  $\pm\omega_o$ . Следовательно:

$$A_o \cos \omega_o t \Leftrightarrow (A_o/2) [\delta(\omega - \omega_o) + \delta(\omega + \omega_o)]. \quad (18.64)$$

При обратном преобразовании Фурье соответственно получаем:

$$(A_o/2) \delta(\omega - \omega_o) \Leftrightarrow (A_o/2) \exp(-j\omega_o t),$$

$$(A_o/2) \delta(\omega + \omega_o) \Leftrightarrow (A_o/2) \exp(j\omega_o t),$$

$$(A_o/2) [\delta(\omega - \omega_o) + \delta(\omega + \omega_o)] \Leftrightarrow \cos(\omega_o t).$$

Таким образом, спектральная плотность косинусоиды вещественна и представляет собой два импульса Дирака, расположенных симметрично относительно  $\omega = 0$  на частотах  $-\omega_o$  и  $\omega_o$  (рис. 18.44, с условной нормировкой по амплитуде).

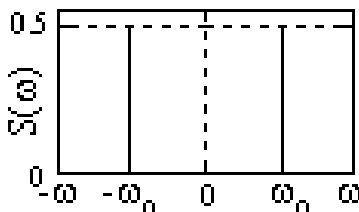


Рис. 18.44.

При наличии во входном сигнале определенного сдвига фазы  $(\omega_o + \varphi_o)$  выражение (18.64) дополняется соответствующими множителями:

$$A_o \cos (\omega_o + \varphi_o)t \Leftrightarrow (A_o/2) [\exp(j\varphi_o) \delta(\omega - \omega_o) + \exp(-j\varphi_o) \delta(\omega + \omega_o)]. \quad (18.65)$$

**8. Радиоимпульс.** Умножение сигнала на гармоническую функцию заполняет сигнал гармонической частотой и формирует радиоимпульс. Без учета начальной фазы гармоники:

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

Спектр радиоимпульса:

$$\begin{aligned} S(\omega) &= \int_{-\infty}^{\infty} u(t) \cos(\omega_0 t) \exp(-j\omega t) dt = \int_{-\infty}^{\infty} u(t) \frac{1}{2} [\exp(j\omega_0 t) + \exp(-j\omega_0 t)] \\ &\quad \exp(-j\omega t) dt = \\ &= \frac{1}{2} \int_{-\infty}^{\infty} u(t) \exp(j\omega_0 t) \exp(-j\omega t) dt + \frac{1}{2} \int_{-\infty}^{\infty} u(t) \exp(-j\omega_0 t) \exp(-j\omega t) dt = \\ &= \frac{1}{2} U(\omega) \exp(j\omega_0 t) + \frac{1}{2} U(\omega) \exp(-j\omega_0 t). \end{aligned} \quad (18.66)$$

Спектры сигналов обычно низкочастотные и сосредоточены в центре частотной оси. Частота гармоник заполнения, как правило, много больше максимальной частоты гармоник сигнала. Из (18.66) следует, что спектр сигнала раздваивается (с коэффициентом  $\frac{1}{2}$ ) и смещается влево и вправо по оси частот на частоты  $\pm\omega_0$ . Особенно наглядно это видно для четных сигналов и приведено на рис. 18.45.

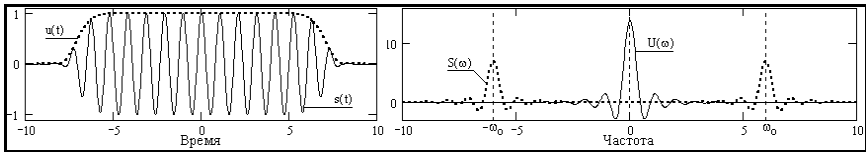


Рис. 18.45. Радиоимпульс и его амплитудный спектр.

## 18.8. Введение в энергетические спектры сигналов

Понятия мощности и энергии в теории сигналов не относятся к характеристикам каких-либо физических величин сигналов, а являются их количественными характеристиками, отражающими определенные свойства сигналов и динамику изменения их значений во времени, в пространстве или по любым другим аргументам.

Для произвольного, в общем случае комплексного, сигнала мгновенная мощность *по определению* равна квадрату функции его модуля, для вещественных сигналов - квадрату функции амплитуд. Энергия сигнала, также по определению, равна интегралу от мощности по всему интервалу существования или задания сигнала.

Энергия сигналов может быть конечной или бесконечной. Конечную энергию имеют финитные сигналы и сигналы, затухающие по своим значениям в пределах конечной длительности, которые не содержат дельта-функций и особых точек (разрывов второго рода и ветвей, уходящих в бесконечность). В противном случае их энергия равна бесконечности. Бесконечна также энергия периодических сигналов.

Частотное представление применяется не только для

спектрального анализа сигналов, но и для упрощения вычислений энергии сигналов и их корреляционных характеристик.

Как уже рассматривалось ранее, для произвольного сигнала  $s(t) = a(t) + jb(t)$ , где  $a(t)$  и  $b(t)$  - вещественные функции, мгновенная мощность сигнала (плотность распределения энергии) определяется выражением:

$$w(t) = s(t)s^*(t) = a^2(t) + b^2(t) = |s(t)|^2.$$

Энергия сигнала равна интегралу от мощности по всему интервалу существования сигнала. В пределе:

$$E_s = \int_{-\infty}^{\infty} w(t) dt = \int_{-\infty}^{\infty} |s(t)|^2 dt.$$

По существу, мгновенная мощность является плотностью мощности сигнала, так как измерения мощности возможны только через энергию, выделяемую на определенных интервалах ненулевой длины:

$$w(\tau) = (1/\Delta t) \int_{\tau - \Delta t/2}^{\tau + \Delta t/2} |s(t)|^2 dt.$$

Сигнал  $s(t)$  изучается, как правило, на определенном интервале  $T$  (для периодических сигналов - в пределах одного периода  $T$ ), при этом средняя мощность сигнала:

$$W_T(\tau) = (1/T) \int_{\tau}^{\tau+T} w(t) dt = (1/T) \int_{\tau}^{\tau+T} |s(t)|^2 dt.$$

Понятие средней мощности может быть распространено и на незатухающие сигналы, энергия которых бесконечно велика. В случае неограниченного интервала  $T$  строго корректное определение средней мощности сигнала производится по формуле:

$$W_s = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T w(t) dt.$$

Энергия и норма сигналов связаны соотношениями:

$$E_s = \|s(t)\|^2, \quad \|s\| = \sqrt{E_s}.$$

## **18.9. Энергетические спектры сигналов**

**Скалярное произведение сигналов.** Энергия суммы двух произвольных сигналов  $u(t)$  и  $v(t)$  определяется выражением:

$$E = \int_{-\infty}^{\infty} [u(t) + v(t)]^2 dt = E_u + E_v + 2 \int_{-\infty}^{\infty} u(t)v(t) dt. \quad (18.67)$$

Как следует из этого выражения, энергии сигналов, в отличие от самих сигналов, в общем случае не обладают свойством

аддитивности. Энергия суммарного сигнала  $u(t)+v(t)$ , кроме суммы энергий составляющих сигналов, содержит в себе и так называемую энергию взаимодействия сигналов или взаимную энергию:

$$E_{uv} = 2 \int_{-\infty}^{\infty} u(t)v(t) dt. \quad (18.68)$$

Интеграл выражения (18.68) для двух вещественных сигналов является фундаментальной характеристикой, пропорциональной взаимной энергии сигналов. Его называют *скалярным произведением* сигналов:

$$\Pi_{uv} = \langle u(t), v(t) \rangle = \int_{-\infty}^{\infty} u(t)v(t) dt = \|u\| \cdot \|v\| \cos \varphi, \quad (18.69)$$

Скалярное произведение обладает следующими свойствами:

5.  $\langle u, v \rangle \geq 0$ ;
6.  $\langle u, v \rangle = \langle v, u \rangle$ ;
7.  $\langle au, v \rangle = a \langle u, v \rangle$ , где  $a$  – вещественное число;
8.  $\langle u+v, a \rangle = \langle u, a \rangle + \langle v, a \rangle$ .

Линейное пространство сигналов с таким скалярным произведением называется *гильбертовым пространством*  $H$ . С учетом того, что  $\cos \varphi \leq 1$ , в гильбертовом пространстве справедливо неравенство Коши-Буняковского:

$$|\Pi_{uv}| \leq \|u\| \cdot \|v\|. \quad (18.70)$$

Для комплексного гильбертова пространства скалярное произведение также представляет собой вещественное число и вычисляется по формуле:

$$\Pi_{uv} = \int_{-\infty}^{\infty} u(t)v^*(t) dt \equiv \int_{-\infty}^{\infty} u^*(t)v(t) dt. \quad (18.71)$$

Из выражения (18.69) следует, что косинус угла между сигналами:

$$\cos \varphi = \Pi_{uv} / (\|u\| \cdot \|v\|). \quad (18.72)$$

При полной тождественности сигналов (равенстве амплитуд и временных координат) имеем  $\varphi=0$ ,  $\cos \varphi = 1$ , и скалярное произведение становится равным энергии сигналов:

$$\Pi_{uv} = \int_{-\infty}^{\infty} u(t)^2 dt \equiv \int_{-\infty}^{\infty} v(t)^2 dt \equiv \|u\|^2 \equiv \|v\|^2.$$

Дискретные сигналы обычно рассматриваются в *пространстве Евклида* (обозначение пространства -  $R_2$ ). Скалярное произведение двух сигналов в пространстве Евклида:

$$\Pi_{uv} = (u_k, v_k) = \sum_{k=1}^n u_k v_k,$$

где  $n$  - размерность пространства.

**Взаимный энергетический спектр.** Из очевидной однозначности энергии взаимодействия сигналов независимо от формы их математического представления (в динамической и частотной модели) следует выражение для скалярного произведения произвольных вещественных сигналов  $u(t)$  и  $v(t)$  через спектральные плотности сигналов  $U(\omega)$  и  $V(\omega)$  в комплексном гильбертовом пространстве:

$$\Pi_{uv} = (1/2\pi) \int_{-\infty}^{\infty} U(\omega)V^*(\omega) d\omega \equiv (1/2\pi) \int_{-\infty}^{\infty} U^*(\omega)V(\omega) d\omega. \quad (18.73)$$

Функции

$$W_{uv}(\omega) = U(\omega)V^*(\omega), \quad W_{vu}(\omega) = U^*(\omega)V(\omega), \quad W_{uv}(\omega) = W_{vu}^*(\omega), \quad (18.74)$$

для которых справедливо выражение (18.73), называется *взаимными энергетическими спектрами* вещественных сигналов, и являются функциями распределения плотности энергии взаимодействия сигналов (мощности взаимодействия) по частоте.

В общем случае, за исключением спектров четных функций, взаимные энергетические спектры также являются комплексными функциями:

$$U(\omega) = A_u(\omega) + j B_u(\omega), \quad V(\omega) = A_v(\omega) + j B_v(\omega).$$

$$W_{uv} = A_u A_v + B_u B_v + j (B_u A_v - A_u B_v) = \text{Re } W_{uv}(\omega) + j \text{Im } W_{uv}(\omega). \quad (18.75)$$

С учетом четности реальной части и нечетности мнимой части энергетических спектров, интеграл мнимой части выражения (18.75) равен нулю, а, следовательно, скалярное произведение сигналов всегда является вещественным и неотрицательным, как и энергия сигналов:

$$\Pi_{uv} = (1/2\pi) \int_{-\infty}^{\infty} W_{uv}(\omega) d\omega \equiv (1/\pi) \int_0^{\infty} \text{Re } W_{uv}(\omega) d\omega. \quad (18.76)$$



Рис. 18.46. Форма и энергетические спектры сигналов.

На рис. 18.46 приведена форма двух одинаковых сдвинутых во времени и частично перекрывающихся лапласовских импульсов  $u(t)$  и  $v(t)$ , а также суммарный импульс  $z(t)=u(t)+v(t)$ . Плотности энергии сигналов  $W(f)$  приведены в относительных единицах плотности энергии суммарного сигнала  $W_z(f)$  на нулевой частоте.

Как видно из графиков, плотности энергии сигналов являются вещественными неотрицательными функциями и содержат только реальные части. В отличие от них, плотность взаимной энергии сигналов является комплексной функцией, при этом модуль плотности по своим значениям на шкале частот соизмерим со средними значениями плотности энергии сигналов на этих частотах и не зависит от их взаимного расположения на временной оси. Для сигналов, одинаковых по форме, модуль взаимной плотности равен значениям плотности энергии сигналов.

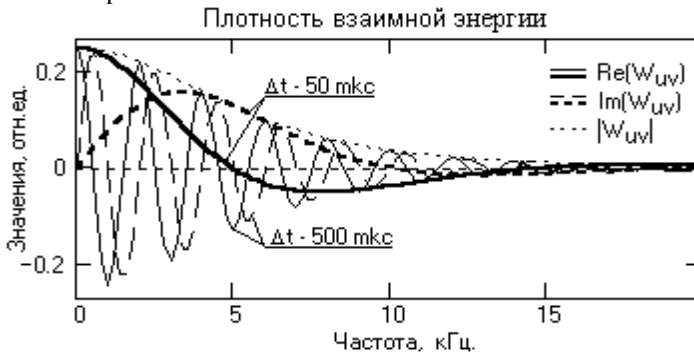


Рис. 18.47. Взаимные энергетические спектры сигналов.

На рис. 18.47 приведены плотности взаимной энергии тех же сигналов при разной величине временного сдвига  $\Delta t$  между сигналами. Однако при постоянном значении модуля взаимной энергии сигналов действительная и мнимая функции спектра мощности существенно изменяются при изменении сдвига между сигналами. При незначительной величине временного перекрытия сигналов частота осцилляций реальной и мнимой части плотности взаимной энергии достаточно велика, а относительный коэффициент затухания колебаний (уменьшение амплитудных значений от периода к периоду) достаточно мал. Соответственно, при вычислении скалярного произведения по формуле (18.76) положительные амплитудные значения осцилляций  $Re(W_{uv})$  практически полностью компенсируются отрицательными значениями и результирующий интеграл, а равно и энергия взаимодействия сигналов (удвоенное значение скалярного произведения), близка к нулевой (стремится к нулю по мере



увеличения сдвига между сигналами).

При увеличении степени взаимного перекрытия сигналов частота осцилляций плотности взаимной энергии уменьшается ( $\Delta t = 50 \text{ мкс}$  на рис. 18.47) и основным по энергии реальной части спектра становится центральный низкочастотный пик, площадь которого не компенсируется площадью последующей отрицательной полуволны осцилляции. Соответственно, возрастает и энергия взаимодействия сигналов. При полном перекрытии сигналов (при нулевом фазовом угле между сигналами) осцилляции исчезают, и энергия взаимодействия сигналов максимальна.

**Энергетический спектр сигнала.** Если функция  $s(t)$  имеет фурье-образ  $S(\omega)$ , то плотность мощности сигнала (*спектральная плотность энергии сигнала*) определяется выражением:

$$w(t) = s(t)s^*(t) = |s(t)|^2 \Leftrightarrow |S(\omega)|^2 = S(\omega)S^*(\omega) = W(\omega). \quad (18.77)$$

Спектр мощности  $W(\omega)$  - вещественная неотрицательная четная функция, которую обычно называют *энергетическим спектром*. Спектр мощности, как квадрат модуля спектральной плотности сигнала, не содержит фазовой информации о его частотных составляющих, а, следовательно, восстановление сигнала по спектру мощности невозможно. Это означает также, что сигналы с различными фазовыми характеристиками могут иметь одинаковые спектры мощности. В частности, сдвиг сигнала не отражается на его спектре мощности. Последнее позволяет получить выражение для энергетического спектра непосредственно из выражений (18.74). В пределе, для одинаковых сигналов  $u(t)$  и  $v(t)$  при сдвиге  $\Delta t \Rightarrow 0$ , мнимая часть спектра  $W_{uv}(\omega)$  стремится к нулевым значениям, а реальная часть - к значениям модуля спектра. При полном временном совмещении сигналов имеем:

$$W_{uv}(\omega) = U(\omega)V^*(\omega) = U(\omega)U^*(\omega) = |U(\omega)|^2 = W_u(\omega). \quad (18.78)$$

Соответственно, полная энергия сигнала:

$$E_u = \int_{-\infty}^{\infty} u(t)^2 dt = (1/2\pi) \int_{-\infty}^{\infty} W_u(t) dt = (1/2\pi) \int_{-\infty}^{\infty} |U(\omega)|^2 d\omega, \quad (18.79)$$

т.е. энергия сигнала равна интегралу квадрата модуля его частотного спектра - сумме энергии его частотных составляющих, и всегда является вещественной величиной.

Для произвольного сигнала  $s(t)$  равенство

$$\int_{-\infty}^{\infty} |s(t)|^2 dt = \int_{-\infty}^{\infty} |S(f)|^2 df$$

обычно называют равенством Парсеваля (в математике - теоремой Планшереля, в физике - формулой Релея). Равенство очевидно, так как координатное и частотное представления по существу только разные

математические отображения одного и того же сигнала. Аналогично для энергии взаимодействия двух сигналов:

$$\int_{-\infty}^{\infty} u(t) v^*(t) dt = \int_{-\infty}^{\infty} U(f) V^*(f) df.$$

Из равенства Парсеваля следует инвариантность скалярного произведения сигналов и нормы относительно преобразования Фурье:

$$\langle u(t), v(t) \rangle = \langle U(f), V(f) \rangle, \quad \|s(t)\|^2 = \|S(f)\|^2.$$

В целом ряде чисто практических задач регистрации и передачи сигналов энергетический спектр сигнала имеет весьма существенное значение.

Периодические сигналы переводятся в спектральную область в виде рядов Фурье. Запишем периодический сигнал с периодом  $T$  в виде ряда Фурье в комплексной форме:

$$s(t) = \sum_{k=-\infty}^{\infty} S_k \exp(j2\pi kt/T),$$

и вычислим среднюю мощность сигнала за один период:

$$W_T = (1/T) \int_0^T s^2(t) dt = (1/T) \sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} S_k S_m \int_0^T \exp(j2\pi(k+m)t/T) dt.$$

Интервал  $0-T$  содержит целое число периодов всех подынтегральных экспонент, и равен нулю, за исключением экспоненты при  $k = -m$ , для которой интеграл равен  $T$ . Соответственно, средняя мощность периодического сигнала равна сумме квадратов модулей коэффициентов его ряда Фурье:

$$W_T = \sum_{k=-\infty}^{\infty} |S_k|^2.$$

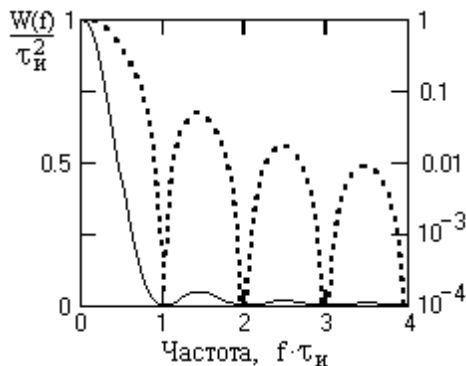


Рис. 18.48. Энергетический спектр прямоугольного импульса.

Как правило, спектры сигналов с крутыми фронтами (например, кодовых сигналов при передаче цифровых данных) являются многолепестковыми с постепенным затуханием энергии в последовательных лепестках. Пример нормированного энергетического спектра прямоугольного импульса длительностью  $\tau_n$  приведен на рис. 18.48. Спектры выполнены в линейном (сплошная линия) и логарифмическом (пунктир) масштабе по оси значений. Для четкого разделения лепестков функции спектров приведены по безразмерной частотной переменной  $f \cdot \tau_n$ . Интегрированием энергетического спектра по интервалам лепестков спектра нетрудно вычислить, что в пределах первого лепестка сосредоточено 90.2% энергии всего сигнала, в пределах второго – 4.8%, в пределах третьего – 1.7%, и т.д. Если форма сигналов в пункте их приема (детектирования) существенного значения не имеет, а регистрация сигналов идет на уровне статистических шумов, равномерно распределенных по всему частотному диапазону, то такие сигналы целесообразно пропускать через фильтр нижних частот с выделением только первого энергетического лепестка сигнала. Естественно, что при этом фронты регистрируемого сигнала будут сглажены. Но при расширении полосы пропускания фильтра на два или три лепестка энергия принимаемого сигнала будет увеличена соответственно на 4.8 или 6.5%, в то время как энергия шумов в 2 или 3 раза.

### **18.10. Введение в корреляцию сигналов**

Корреляция (correlation), и ее частный случай для центрированных сигналов – ковариация, является методом анализа сигналов. Приведем один из вариантов использования метода. Допустим, что имеется сигнал  $s(t)$ , в котором может быть (а может и не быть) некоторая последовательность  $x(t)$  конечной длины  $T$ , временное положение которой нас интересует. Для поиска этой последовательности в скользящем по сигналу  $s(t)$  временном окне длиной  $T$  вычисляются скалярные произведения сигналов  $s(t)$  и  $x(t)$ . Тем самым мы "прикладываем" искомый сигнал  $x(t)$  к сигналу  $s(t)$ , скользя по его аргументу, и по величине скалярного произведения оцениваем степень сходства сигналов в точках сравнения.

Корреляционный анализ дает возможность установить в сигналах (или в рядах цифровых данных сигналов) наличие определенной связи изменения значений сигналов по независимой переменной, то есть, когда большие значения одного сигнала (относительно средних значений сигнала) связаны с большими значениями другого сигнала (положительная корреляция), или, наоборот, малые значения одного сигнала связаны с большими

значениями другого (отрицательная корреляция), или данные двух сигналов никак не связаны (нулевая корреляция).

В функциональном пространстве сигналов эта степень связи может выражаться в нормированных единицах коэффициента корреляции, т.е. в косинусе угла между векторами сигналов, и, соответственно, будет принимать значения от 1 (полное совпадение сигналов) до -1 (полная противоположность) и не зависит от значения (масштаба) единиц измерений.

В варианте автокорреляции (autocorrelation) по аналогичной методике производится определение скалярного произведения сигнала  $s(t)$  с собственной копией, скользящей по аргументу. Автокорреляция позволяет оценить среднестатистическую зависимость текущих отсчетов сигнала от своих предыдущих и последующих значений (так называемый радиус корреляции значений сигнала), а также выявить в сигнале наличие периодически повторяющихся элементов.

Особое значение методы корреляции имеют при анализе случайных процессов для выявления неслучайных составляющих и оценки неслучайных параметров этих процессов.

Заметим, что в терминах "корреляция" и "ковариация" существует некоторая путаница. В математической литературе термин "ковариация" применяется к центрированным функциям, а "корреляция" – к произвольным. В технической литературе, и особенно в литературе по сигналам и методам их обработки, часто применяется прямо противоположная терминология. Принципиального значения это не имеет, но при знакомстве с литературными источниками стоит обращать внимание на принятое назначение данных терминов.

## **18.11. Автокорреляционные функции сигналов**

### **Понятие автокорреляционных функций сигналов.**

Автокорреляционная функция (АКФ, CF - correlation function) сигнала  $s(t)$ , конечного по энергии, является количественной интегральной характеристикой формы сигнала, выявления в сигнале характера и параметров взаимной временной связи отсчетов, что всегда имеет место для периодических сигналов, а также интервала и степени зависимости значений отсчетов в текущие моменты времени от предыстории текущего момента. АКФ определяется интегралом от произведения двух копий сигнала  $s(t)$ , сдвинутых относительно друг друга на время  $\tau$ :

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t) s(t+\tau) dt = \langle s(t), s(t+\tau) \rangle = \|s(t)\| \|s(t+\tau)\| \cos \varphi(\tau). \quad (18.80)$$

Как следует из этого выражения, АКФ является скалярным произведением сигнала и его копии в функциональной зависимости от

переменной величины значения сдвига  $\tau$ . Соответственно, АКФ имеет физическую размерность энергии, а при  $\tau=0$  значение АКФ непосредственно равно энергии сигнала и является максимально возможным (косинус угла взаимодействия сигнала с самим собой равен 1):

$$B_s(0) = \int_{-\infty}^{\infty} s(t)^2 dt = E_s.$$

АКФ относится к четным функциям, в чем нетрудно убедиться заменой переменной  $t = t - \tau$  в выражении (18.80):

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t-\tau) s(t) dt = B_s(-\tau).$$

Максимум АКФ, равный энергии сигнала при  $\tau=0$ , всегда положителен, а модуль АКФ при любом значении временного сдвига не превосходит энергии сигнала. Последнее прямо вытекает из свойств скалярного произведения (как и неравенство Коши-Буняковского):

$$\langle s(t), s(t+\tau) \rangle = \|s(t)\| \cdot \|s(t+\tau)\| \cdot \cos \varphi(\tau),$$

$$\cos \varphi(\tau) = 1 \text{ при } \tau = 0, \quad \langle s(t), s(t+\tau) \rangle = \|s(t)\| \cdot \|s(t)\| = E_s,$$

$$\cos \varphi(\tau) < 1 \text{ при } \tau \neq 0, \quad \langle s(t), s(t+\tau) \rangle = \|s(t)\| \cdot \|s(t+\tau)\| \cdot \cos \varphi(\tau) < E_s.$$

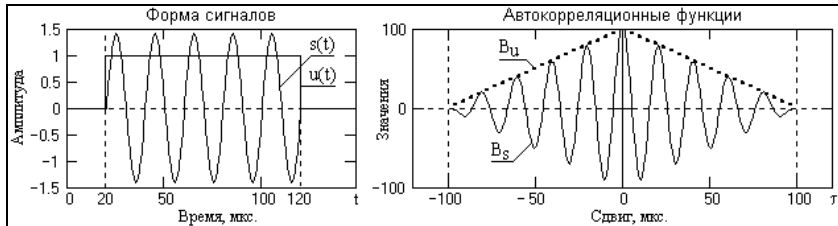


Рис. 18.49.

В качестве примера на рис. 18.49 приведены два сигнала – прямоугольный импульс и радиои́мпульс одинаковой длительности  $T$ , и соответствующие данным сигналам формы их АКФ. Амплитуда колебаний радиои́мпульса установлена равной  $\sqrt{T}$  амплитуды прямоугольного импульса, при этом энергии сигналов также будут одинаковыми, что подтверждается равными значениями центральных максимумов АКФ. При конечной длительности импульсов длительности АКФ также конечны, и равны удвоенным значениям длительности импульсов (при сдвиге копии конечного импульса на интервал его длительности как влево, так и вправо, произведение импульса со своей копией становится равным нулю). Частота колебаний АКФ радиои́мпульса равна частоте колебаний заполнения радиои́мпульса (боковые минимумы и максимумы АКФ возникают

каждый раз при последовательных сдвигах копии радиопульса на половину периода колебаний его заполнения).

С учетом четности, графическое представление АКФ обычно производится только для положительных значений  $\tau$ . На практике сигналы обычно задаются на интервале положительных значений аргументов от 0-T. Знак  $+\tau$  в выражении (18.80) означает, что при увеличении значений  $\tau$  копия сигнала  $s(t+\tau)$  сдвигается влево по оси  $t$  и уходит за 0. Для цифровых сигналов это требует соответствующего продления данных в область отрицательных значений аргумента. А так как при вычислениях интервал задания  $\tau$  обычно много меньше интервала задания сигнала, то более практичным является сдвиг копии сигнала влево по оси аргументов, т.е. применение в выражении (18.80) функции  $s(t-\tau)$  вместо  $s(t+\tau)$ .

$$B_s(\tau) = \int_{-\infty}^{\infty} s(t) s(t-\tau) dt. \quad (18.81)$$

Для финитных сигналов по мере увеличения значения величины сдвига  $\tau$  временное перекрытие сигнала с его копией уменьшается, а, соответственно, косинус угла взаимодействия и скалярное произведение в целом стремятся к нулю:

$$\lim_{|\tau| \rightarrow \infty} B_s(\tau) = 0.$$

АКФ, вычисленная по центрированному значению сигнала  $s(t)$ , представляет собой *автоковариационную* функцию сигнала:

$$C_s(\tau) = \int_{-\infty}^{\infty} [s(t)-\mu_s][s(t+\tau)-\mu_s] dt, \quad (18.82)$$

где  $\mu_s$  – среднее значение сигнала. Ковариационные функции связаны с корреляционными функциями достаточно простым соотношением:

$$C_s(\tau) = B_s(\tau) - \mu_s^2.$$

**АКФ сигналов, ограниченных во времени.** На практике обычно исследуются и анализируются сигналы, заданные на определенном интервале. Для сравнения АКФ сигналов, заданных на различных временных интервалах, практическое применение находит модификация АКФ с нормировкой на длину интервала. Так, например, при задании сигнала на интервале  $[a, b]$ :

$$B_s(\tau) = \frac{1}{b-a} \int_a^b s(t) s(t+\tau) dt. \quad (18.83)$$

АКФ может быть вычислена и для слабозатухающих сигналов с бесконечной энергией, как среднее значение скалярного произведения сигнала и его копии при устремлении интервала задания сигнала к бесконечности:

$$B_s(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s(t) s(t + \tau) dt . \quad (18.84)$$

АКФ по данным выражениям имеет физическую размерность мощности, и равна средней взаимной мощности сигнала и его копии в функциональной зависимости от сдвига копии.

**АКФ периодических сигналов.** Энергия периодических сигналов бесконечна, поэтому АКФ периодических сигналов вычисляется по одному периоду  $T$ , с усреднением скалярного произведения сигнала и его сдвинутой копии в пределах периода:

$$B_s(\tau) = (1/T) \int_0^T s(t) s(t-\tau) dt. \quad (18.85)$$

Математически более строгое выражение:

$$B_s(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T s(t) s(t - \tau) dt .$$

При  $\tau=0$  значение нормированной на период АКФ равно средней мощности сигналов в пределах периода. При этом АКФ периодических сигналов является периодической функцией с тем же периодом  $T$ . Так, для сигнала  $s(t) = A \cos(\omega_0 t + \varphi_0)$  при  $T=2\pi/\omega_0$  имеем:

$$B_s(\tau) = \frac{\omega_0}{2\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} A \cos(\omega_0 t + \varphi_0) A \cos(\omega_0(t-\tau) + \varphi_0) = (A^2/2) \cos(\omega_0 \tau). \quad (18.86)$$

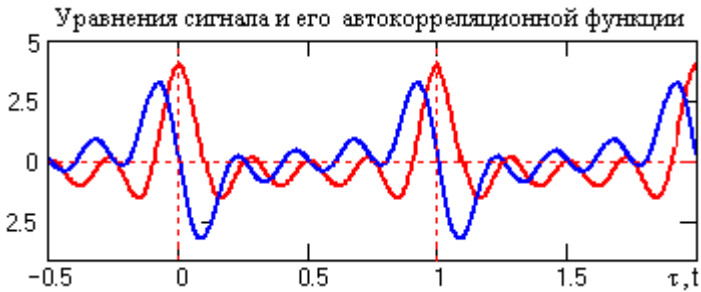


Рис. 18.50.

Полученный результат не зависит от начальной фазы гармонического сигнала, что характерно для любых периодических сигналов и является одним из свойств АКФ. С помощью функций автокорреляции можно проверять наличие периодических свойств в любых произвольных сигналах. Пример автокорреляционной функции периодического сигнала приведен на рис. 18.50.

**Функции автоковариации (ФАК)** вычисляются аналогично, по центрированным значениям сигнала. Замечательной

особенностью этих функций являются их простые соотношения с дисперсией  $\sigma_s^2$  сигналов (квадратом стандарта - среднего квадратического отклонения значений сигнала от среднего значения). Как известно, значение дисперсии равно средней мощности сигналов, откуда следует:

$$|C_s(\tau)| \leq \sigma_s^2, \quad C_s(0) = \sigma_s^2 \equiv \|s(t)\|^2. \quad (18.87)$$

Значения ФАК, нормированные на значение дисперсии, представляют собой функцию автокорреляционных коэффициентов:

$$\rho_s(\tau) = C_s(\tau)/C_s(0) = C_s(\tau)/\sigma_s^2 \equiv \cos \varphi(\tau). \quad (18.88)$$

Иногда эту функцию называют "истинной" автокорреляционной функцией. В силу нормировки ее значения не зависят от единиц (масштаба) представления значений сигнала  $s(t)$  и характеризуют степень линейной связи между значениями сигнала в зависимости от величины сдвига  $\tau$  между отсчетами сигнала. Значения  $\rho_s(\tau) \equiv \cos \varphi(\tau)$  могут изменяться от 1 (полная прямая корреляция отсчетов) до -1 (обратная корреляция).

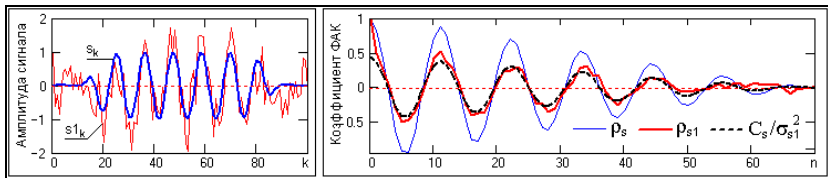


Рис. 18.51.

На рис. 18.51 приведен пример сигналов  $s(k)$  и  $s1(k) = s(k) + \text{шум}$  с соответствующими этим сигналам коэффициентами ФАК -  $\rho_s$  и  $\rho_{s1}$ . Как видно на графиках, ФАК уверенно выявила наличие периодических колебаний в сигналах. Шум в сигнале  $s1(k)$  понизил амплитуду периодических колебаний без изменения периода. Это подтверждает график кривой  $C_s/\sigma_{s1}^2$ , т.е. ФАК сигнала  $s(k)$  с нормировкой (для сопоставления) на значение дисперсии сигнала  $s1(k)$ , где наглядно можно видеть, что шумовые импульсы при полной статистической независимости своих отсчетов вызвали увеличение значения  $C_{s1}(0)$  по отношению к значению  $C_s(0)$  и несколько "размыли" функцию коэффициентов автоковариации. Это вызвано тем, что значение  $\rho_s(\tau)$  шумовых сигналов стремится к 1 при  $\tau \rightarrow 0$  и флуктуирует относительно нуля при  $\tau \neq 0$ , при этом амплитуды флуктуаций статистически независимы и зависят от количества выборок сигнала (стремятся к нулю при увеличении количества отсчетов).

**АКФ дискретных сигналов.** При интервале дискретизации данных  $\Delta t = \text{const}$  вычисление АКФ выполняется по интервалам  $\Delta \tau = \Delta t$



и обычно записывается, как дискретная функция номеров  $n$  сдвига отсчетов  $n\Delta t$ :

$$B_s(n\Delta t) = \Delta t \sum_{k=-\infty}^{\infty} s_k \cdot s_{k-n}. \quad (18.89)$$

Дискретные сигналы обычно задаются в виде числовых массивов определенной длины с нумерацией отсчетов  $k = 0, 1, \dots, K$  при  $\Delta t=1$ , а вычисление дискретной АКФ в единицах энергии выполняется в одностороннем варианте с учетом длины массивов. Если используется весь массив сигнала и число отсчетов АКФ равно числу отсчетов массива, то вычисление выполняется по формуле:

$$B_s(n) = \frac{K}{K-n} \sum_{k=0}^{K-n} s_k \cdot s_{k-n}. \quad (18.90)$$

Множитель  $K/(K-n)$  в данной функции является поправочным коэффициентом на постепенное уменьшение числа перемножаемых и суммируемых значений по мере увеличения сдвига  $n$ . Без этой поправки для нецентрированных сигналов в значениях АКФ появляется тренд суммирования средних значений. При измерениях в единицах мощности сигнала множитель  $K/(K-n)$  заменяется на множитель  $1/(K-n)$ . Формула (18.90) применяется довольно редко, в основном для детерминированных сигналов с небольшим числом отсчетов. Для случайных и зашумленных сигналов уменьшение знаменателя  $(K-n)$  и числа перемножаемых отсчетов по мере увеличения сдвига приводит к нарастанию статистических флуктуаций вычисления АКФ. Большую достоверность в этих условиях обеспечивает вычисление АКФ в единицах мощности сигнала по формуле:

$$B_s(n) = \frac{1}{K} \sum_{k=0}^{K-n} s_k \cdot s_{k-n}, \quad s_{k-n} = 0 \text{ при } k-n < 0, \quad (18.91)$$

т.е. с нормированием на постоянный множитель  $1/K$  и с продлением сигнала нулевыми значениями (в левую сторону при сдвигах  $k-n$  или в правую сторону при использовании сдвигов  $k+n$ ). Эта оценка является смещенной и имеет несколько меньшую дисперсию, чем по формуле (18.90). Разницу между нормировками по формулам (18.90) и (18.91) можно наглядно видеть на рис. 18.52.

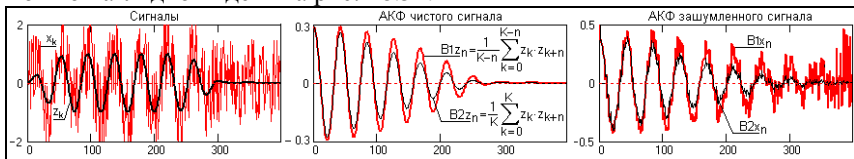


Рис. 18.52.

Формулу (18.91) можно рассматривать, как усреднение суммы произведений, т.е. как оценку математического ожидания:

$$B_s(n) = M\{s_k s_{k-n}\} \cong S_k S_{k-n} . \quad (18.92)$$

Практически, дискретная АКФ имеет такие же свойства, как и непрерывная АКФ. Она также является четной, а ее значение при  $n = 0$  равно энергии или мощности дискретного сигнала в зависимости от нормировки.

**АКФ зашумленных сигналов.** Зашумленный сигнал записывается в виде суммы  $v(k) = s(k)+q(k)$ . В общем случае, шум не обязательно должен иметь нулевое среднее значение, и нормированная по мощности автокорреляционная функция цифрового сигнала, содержащая  $N$  – отсчетов, записывается в следующем виде:

$$\begin{aligned} B_v(n) &= (1/N) \langle s(k)+q(k), s(k-n)+q(k-n) \rangle = \\ &= (1/N) [\langle s(k), s(k-n) \rangle + \langle s(k), q(k-n) \rangle + \langle q(k), s(k-n) \rangle + \langle q(k), q(k-n) \rangle] = \\ &= B_s(n) + M\{s_k q_{k-n}\} + M\{q_k s_{k-n}\} + M\{q_k q_{k-n}\} . \end{aligned}$$

$$B_v(n) = B_s(n) + S_k Q_{k-n} + Q_k S_{k-n} + Q_k Q_{k-n} . \quad (18.93)$$

При статистической независимости полезного сигнала  $s(k)$  и шума  $q(k)$  с учетом разложения математического ожидания

$$M\{s_k q_{k-n}\} = M\{s_k\} M\{q_{k-n}\} = S_k Q_k$$

может использоваться следующая формула:

$$B_v(n) = B_s(n) + 2 S_k Q_k + Q_k^2 . \quad (18.94)$$

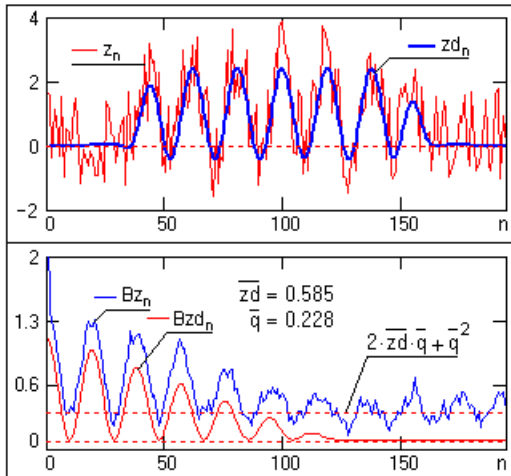


Рис. 18.53

Пример зашумленного сигнала и его АКФ в сопоставлении с

незашумленным сигналом приведен на рис. 18.53

Из формул (18.93) следует, что АКФ зашумленного сигнала состоит из АКФ сигнальной компоненты  $\frac{1}{2} \frac{1 - \cos 2\pi K t}{1 - \cos 2\pi K t + Q_k^{-2}}$  наложенной затухающей до значения  $2 S_k Q_k + Q_k^{-2}$  шумовой функцией. При больших значениях  $K$ , когда  $Q_k \rightarrow 0$ , имеет место  $V_v(n) \approx V_s(n)$ . Это дает возможность не только выделять по АКФ периодические сигналы, практически полностью скрытые в шуме (мощность шумов много больше мощности сигнала), но и с высокой точностью определять их период и форму в пределах периода, а для одночастотных гармонических сигналов – и их амплитуду с использованием выражения (18.86).

Таблица 18.1.

М	Сигнал Баркера	АКФ сигнала
2	1, -1	2, -1
3	1, 1, -1	3, 0, -1
4	1, 1, 1, -1	4, 1, 0, -1
	1, 1, -1, 1	4, -1, 0, 1
5	1, 1, 1, -1, 1	5, 0, 1, 0, 1
7	1, 1, 1, -1, -1, 1, -1	7, 0, -1, 0, -1, 0, -1
11	1, 1, 1, -1, -1, -1, 1, -1, -1, 1, -1	11, 0, -1, 0, -1, 0, -1, 0, -1, 0, -1
13	1, 1, 1, 1, 1, -1, -1, 1, 1, -1, 1, -1, 1	13, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1

**Кодовые сигналы** являются разновидностью дискретных сигналов. На определенном интервале кодового слова  $M \cdot \Delta t$  они могут иметь только два амплитудных значения: 0 и 1 или 1 и  $-1$ . При выделении кодов на существенном уровне шумов форма АКФ кодового слова имеет особое значение. С этой позиции наилучшими считаются такие коды, значения боковых лепестков АКФ которых минимальны по всей длине интервала кодового слова при максимальном значении центрального пика. К числу таких кодов относится код Баркера, приведенный в таблице 18.1. Как видно из таблицы, амплитуда центрального пика кода численно равна значению  $M$ , при этом амплитуда боковых осцилляций при  $n \neq 0$  не превышает 1.

## 18.12. Взаимные корреляционные функции сигналов

**Взаимная корреляционная функция** (ВКФ) разных сигналов (cross-correlation function, CCF) описывает как степень сходства формы двух сигналов, так и их взаимное расположение друг относительно друга по координате (независимой переменной).

Обобщая формулу (18.80) автокорреляционной функции на два различных сигнала  $s(t)$  и  $u(t)$ , получаем следующее скалярное произведение сигналов:

$$B_{su}(\tau) = \int_{-\infty}^{\infty} s(t) u(t+\tau) dt. \quad (18.95)$$

Взаимная корреляция сигналов характеризует определенную корреляцию явлений и физических процессов, отображаемых данными сигналами, и может служить мерой “устойчивости” данной взаимосвязи при отдельной обработке сигналов в различных устройствах. Для конечных по энергии сигналов ВКФ также конечна, при этом:

$$|B_{su}(\tau)| \leq \|s(t)\| \cdot \|u(t)\|,$$

что следует из неравенства Коши-Буняковского и независимости норм сигналов от сдвига по координатам.

При замене переменной  $t = t - \tau$  в формуле (18.95), получаем:

$$B_{su}(\tau) = \int_{-\infty}^{\infty} s(t-\tau) u(t) dt = \int_{-\infty}^{\infty} u(t) s(t-\tau) dt = B_{us}(-\tau).$$

Отсюда следует, что для ВКФ не выполняется условие четности,  $B_{su}(\tau) \neq B_{su}(-\tau)$ , и значения ВКФ не обязаны иметь максимум при  $\tau = 0$ .

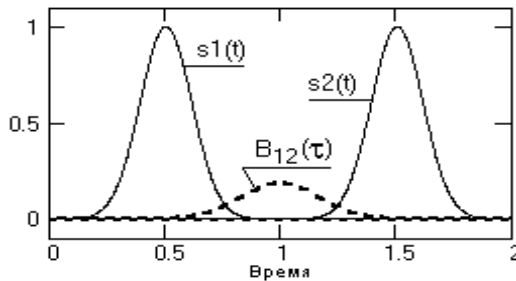


Рис. 18.54 Сигналы и ВКФ.

Это можно наглядно видеть на рис. 18.54 где заданы два одинаковых сигнала с центрами на точках 0.5 и 1.5. Вычисление по формуле (18.95) с постепенным увеличением значений  $\tau$  означает последовательные сдвиги сигнала  $s_2(t)$  влево по оси времени (для каждого значения  $s_1(t)$  для подынтегрального умножения берутся значения  $s_2(t+\tau)$ ). При  $\tau=0$  сигналы ортогональны и значение  $B_{12}(\tau)=0$ . Максимум  $B_{12}(\tau)$  будет наблюдаться при сдвиге сигнала  $s_2(t)$  влево на значение  $\tau=1$ , при котором происходит полное совмещение сигналов  $s_1(t)$  и  $s_2(t+\tau)$ .

Одни и те же значения ВКФ по формулам (18.95) и (18.96)

наблюдаются при одном и том же взаимном положении сигналов: при сдвиге на интервал  $\tau$  сигнала  $u(t)$  относительно  $s(t)$  вправо по оси ординат и сигнала  $s(t)$  относительно сигнала  $u(t)$  влево, т.е.  $B_{su}(\tau) = B_{us}(-\tau)$ .

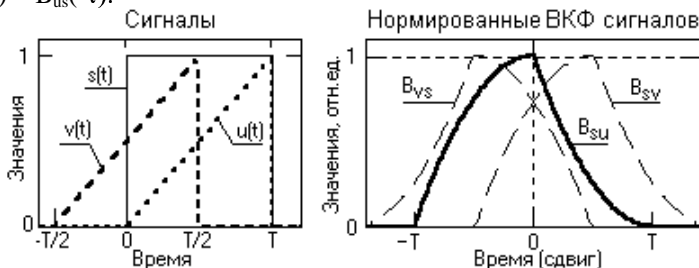


Рис. 18.55. Взаимнокорреляционные функции сигналов.

На рис. 18.55 приведены примеры ВКФ для прямоугольного сигнала  $s(t)$  и двух одинаковых треугольных сигналов  $u(t)$  и  $v(t)$ . Все сигналы имеют одинаковую длительность  $T$ , при этом сигнал  $v(t)$  сдвинут вперед на интервал  $T/2$ .

Сигналы  $s(t)$  и  $u(t)$  одинаковы по временному расположению и площадь "перекрывания" сигналов максимальна при  $\tau=0$ , что и фиксируется функцией  $B_{su}$ . Вместе с тем функция  $B_{su}$  резко асимметрична, так как при асимметричной форме сигнала  $u(t)$  для симметричной формы  $s(t)$  (относительно центра сигналов) площадь "перекрывания" сигналов изменяется по разному в зависимости от направления сдвига (знака  $\tau$  при увеличении значения  $\tau$  от нуля). При смещении исходного положения сигнала  $u(t)$  влево по оси ординат (на опережение сигнала  $s(t)$  - сигнал  $v(t)$ ) форма ВКФ остается без изменения и сдвигается вправо на такое же значение величины сдвига - функция  $B_{sv}$  на рис. 18.55. Если поменять местами выражения функций в (18.95), то новая функция  $B_{vs}$  будет зеркально повернутой относительно  $\tau=0$  функцией  $B_{sv}$ .

С учетом этих особенностей полное ВКФ вычисляется, как правило, отдельно для положительных и отрицательных запаздываний:

$$B_{su}(\tau) = \int_{-\infty}^{\infty} s(t) u(t+\tau) dt. \quad B_{us}(\tau) = \int_{-\infty}^{\infty} u(t) s(t+\tau) dt. \quad (18.96)$$

**Взаимная корреляция зашумленных сигналов.** Для двух зашумленных сигналов  $u(t)=s_1(t)+q_1(t)$  и  $v(t)=s_2(t)+q_2(t)$ , применяя методику вывода формул (18.93) с заменой копии сигнала  $s(t)$  на сигнал  $s_2(t)$ , нетрудно вывести формулу взаимной корреляции в следующем виде:

$$B_{uv}(\tau) = B_{s_1s_2}(\tau) + B_{s_1q_2}(\tau) + B_{q_1s_2}(\tau) + B_{q_1q_2}(\tau). \quad (18.97)$$

Последние три члена в правой части (18.97) затухают до нуля

при увеличении  $\tau$ . При больших интервалах задания сигналов выражение может быть записано в следующей форме:

$$B_{uv}(\tau) = B_{s_1s_2}(\tau) + \overline{s1(\tau)q2(\tau)} + \overline{q1(\tau)s2(\tau)} + \overline{q1(\tau)q2(\tau)}. \quad (18.98)$$

При нулевых средних значениях шумов и статистической независимости от сигналов имеет место:

$$B_{uv}(\tau) \rightarrow B_{s_1s_2}(\tau).$$

**ВКФ дискретных сигналов.** Все свойства ВКФ аналоговых сигналов действительны и для ВКФ дискретных сигналов, при этом для них действительны и особенности дискретных сигналов, изложенные выше для дискретных АКФ (формулы 18.89-18.92). В частности, при  $\Delta t = \text{const} = 1$  для сигналов  $x(k)$  и  $y(k)$  с числом отсчетов  $K$ :

$$B_{xy}(n) = \frac{K}{K-n} \sum_{k=0}^{K-n} x_k y_{k-n}. \quad (18.99)$$

При нормировании в единицах мощности:

$$B_{xy}(n) = \frac{1}{K} \sum_{k=0}^K x_k y_{k-n} \cong \overline{x_k y_{k-n}}. \quad (18.100)$$

**Оценка периодических сигналов в шуме.** Зашумленный сигнал можно оценить по взаимной корреляции с "эталонным" сигналом методом проб и ошибок с настройкой функции взаимной корреляции до максимального значения.

Для сигнала  $u(k) = s(k) + q(k)$  при статистической независимости шума и  $\overline{q_k} \rightarrow 0$  функция взаимной корреляции (18.97) с шаблоном сигнала  $p(k)$  при  $q_2(k) = 0$  принимает вид:

$$B_{up}(k) = \overline{B_{sp}(k) + B_{qp}(k)} = \overline{B_{sp}(k)} + \overline{q_k} \overline{p_k}.$$

А поскольку  $\overline{q_k} \rightarrow 0$  при увеличении  $N$ , то  $B_{up}(k) \rightarrow \overline{B_{sp}(k)}$ . Очевидно, что функция  $B_{up}(k)$  будет иметь максимум, когда  $p(k) = s(k)$ . Меняя форму шаблона  $p(k)$  и добиваясь максимизации функции  $B_{up}(k)$ , можно получить оценку  $s(k)$  в виде оптимальной формы  $p(k)$ .

**Функция взаимных корреляционных коэффициентов** (ВКФ) является количественным показателем степени сходства сигналов  $s(t)$  и  $u(t)$ . Аналогично функции автокорреляционных коэффициентов, она вычисляется через центрированные значения функций (для вычисления взаимной ковариации достаточно центрировать только одну из функций), и нормируется на произведение значений стандартов функций  $s(t)$  и  $v(t)$ :

$$\rho_{su}(\tau) = C_{su}(\tau) / \sigma_s \sigma_v. \quad (18.101)$$

Интервал изменения значений корреляционных

коэффициентов при сдвигах  $\tau$  может изменяться от  $-1$  (полная обратная корреляция) до  $1$  (полное сходство или стопроцентная корреляция). При сдвигах  $\tau$ , на которых наблюдаются нулевые значения  $\rho_{su}(\tau)$ , сигналы независимы друг от друга (некоррелированы). Коэффициент взаимной корреляции позволяет устанавливать наличие связи между сигналами вне зависимости от физических свойств сигналов и их величины.

При вычислении ВКФ зашумленных дискретных сигналов ограниченной длины с использованием формулы (18.99) имеется вероятность появления значений  $|\rho_{su}(n)| > 1$ .

Для периодических сигналов понятие ВКФ обычно не применяется, за исключением сигналов с одинаковым периодом, например, сигналов входа и выхода при изучении характеристик систем.

### **18.13. Спектральные плотности корреляционных функций**

**Спектральная плотность АКФ** может быть определена из следующих простых соображений.

В соответствии с выражением (18.80) АКФ представляет собой функцию скалярного произведения сигнала и его копии, сдвинутой на интервал  $\tau$ , при  $-\infty < \tau < \infty$ :

$$B_s(\tau) = \langle s(t), s(t-\tau) \rangle.$$

Скалярное произведение может быть определено через спектральные плотности сигнала и его копии, произведение которых представляет собой спектральную плотность взаимной мощности:

$$\langle s(t), s(t-\tau) \rangle = (1/2\pi) \int_{-\infty}^{\infty} S(\omega) S_{\tau}^*(\omega) d\omega.$$

Смещение сигнала по оси абсцисс на интервал  $\tau$  отображается в спектральном представлении умножением спектра сигнала на  $\exp(-j\omega\tau)$ , а для сопряженного спектра на множитель  $\exp(j\omega\tau)$ :

$$S_{\tau}^*(\omega) = S^*(\omega) \exp(j\omega\tau).$$

С учетом этого получаем:

$$B_s(\tau) = (1/2\pi) \int_{-\infty}^{\infty} S(\omega) S^*(\omega) \exp(j\omega\tau) d\omega = (1/2\pi) \int_{-\infty}^{\infty} |S(\omega)|^2 \exp(j\omega\tau) d\omega. \quad (18.102)$$

Но последнее выражение представляет собой обратное преобразование Фурье энергетического спектра сигнала (спектральной плотности энергии). Следовательно, энергетический спектр сигнала и его автокорреляционная функция связаны преобразованием Фурье:

$$B_s(\tau) \leftrightarrow |S(\omega)|^2 = W_s(\omega). \quad (18.103)$$

Таким образом, спектральная плотность АКФ есть не что иное, как спектральная плотность мощности сигнала, которая, в свою очередь, может определяться прямым преобразованием Фурье через АКФ:

$$|S(\omega)|^2 = \int_{-\infty}^{\infty} B_s(\tau) \exp(-j\omega\tau) d\tau. \quad (18.104)$$

Последние выражение накладывает определенные ограничения на форму АКФ и методику их ограничения по длительности.

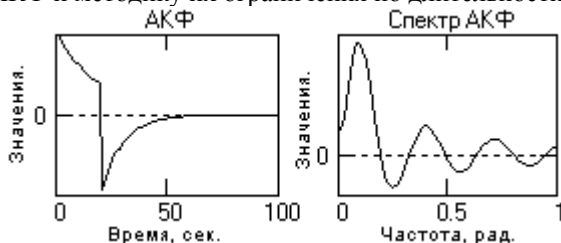


Рис. 18.56. Спектр несуществующей АКФ

Энергетический спектр сигналов всегда положителен, мощность сигналов не может быть отрицательной. Следовательно, АКФ не может иметь формы прямоугольного импульса, т.к. преобразование Фурье прямоугольного импульса – знакопеременный интегральный синус. На АКФ не должно быть и разрывов первого рода (скачков), т.к. с учетом четности АКФ любой симметричный скачек по координате  $\pm\tau$  порождает “разделение” АКФ на сумму определенной непрерывной функции и прямоугольного импульса длительностью  $2\tau$  с соответствующим появлением отрицательных значений в энергетическом спектре. Пример последнего приведен на рис. 18.56 (графики функций приведены, как принято для четных функций, только своей правой частью).

АКФ достаточно протяженных сигналов обычно ограничиваются по размерам (исследуются ограниченные интервалы корреляции данных от  $-T/2$  до  $T/2$ ). Однако усечение АКФ, это умножение АКФ на прямоугольный селекционирующий импульс длительностью  $T$ , что в частотной области отображается сверткой фактического спектра мощности со знакопеременной функцией интегрального синуса  $\text{sinc}(\omega T/2)$ . С одной стороны, это вызывает определенное сглаживание спектра мощности, что зачастую бывает полезным, например, при исследовании сигналов на значительном уровне шумов. Но, с другой стороны, может происходить и существенное занижение величины энергетических пиков, если в сигнале имеются какие-либо гармонические составляющие, а также появление отрицательных значений мощности на краевых частях



пиков и скачков. Пример проявления факторов приведен на рис. 18.57.



Рис. 18.57. Вычисление энергетического спектра сигнала по АКФ разной длины.

Как известно, спектры мощности сигналов не имеют фазовой характеристики и по ним невозможно восстановление сигналов. Следовательно, АКФ сигналов, как временное представление спектров мощности, также не имеет информации о фазовых характеристиках сигналов и восстановление сигналов по АКФ невозможно. Сигналы одной формы, сдвинутые во времени, имеют одинаковые АКФ. Больше того, сигналы разной формы могут иметь сходные АКФ, если имеют близкие спектры мощности.

Перепишем уравнение (18.104) в следующей форме

$$\int_{-\infty}^{\infty} s(t) s(t-\tau) dt = (1/2\pi) \int_{-\infty}^{\infty} S(\omega) S^*(\omega) \exp(j\omega\tau) d\omega,$$

и подставим в это выражение значение  $\tau=0$ . Полученное равенство хорошо известно и называется *равенством Парсеваля*

$$\int_{-\infty}^{\infty} s^2(t) dt = (1/2\pi) \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega.$$

Оно позволяет вычислять энергию сигнала, как по временной, так и по частотной области описания сигналов.

**Интервал корреляции сигнала** является числовым параметром оценки ширины АКФ и степени значимой корреляции значений сигнала по аргументу.

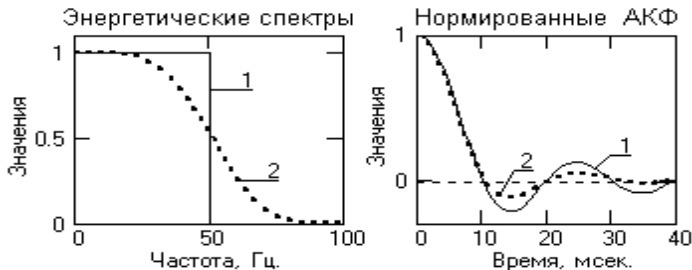


Рис. 18.58.

Если допустить, что сигнал  $s(t)$  имеет примерно равномерный энергетический спектр со значением  $W_0$  и с верхней граничной частотой до  $\omega_b$  (форма центрированного прямоугольного импульса, как, например, сигнал 1 на рис. 18.58 с  $f_b=50$  Гц в одностороннем представлении), то АКФ сигнала определится выражением:

$$B_s(\tau) = (W_0/\pi) \int_0^{\omega_b} \cos(\omega\tau) d\omega = (W_0\omega_b/\pi) \text{sinc}(\omega_b\tau)/(\omega_b\tau).$$

Интервалом корреляции сигнала  $\tau_k$  считается величина ширины центрального пика АКФ от максимума до первого пересечения нулевой линии. В данном случае для прямоугольного спектра с верхней граничной частотой  $\omega_b$  первое пересечение нуля соответствует  $\text{sinc}(\omega_b\tau) = 0$  при  $\omega_b\tau = \pi$ , откуда:

$$\tau_k = \pi/\omega_b = 1/2f_b. \quad (18.105)$$

Интервал корреляции тем меньше, чем выше верхняя граничная частота спектра сигнала. Для сигналов с плавным срезом по верхней граничной частоте роль параметра  $\omega_b$  играет средняя ширина спектра (сигнал 2 на рис. 18.58).

Спектральная плотность мощности статистических шумов при единичном измерении представляет собой случайную функцию  $W_q(\omega)$  со средним значением  $W_q(\omega) \Rightarrow \sigma_q^2$ , где  $\sigma_q^2$  – дисперсия шумов. В пределе, при равномерном спектральном распределении шумов от 0 до  $\infty$ , АКФ шумов стремится к значению  $B_q(\tau) \Rightarrow \sigma_q^2$  при  $\tau \Rightarrow 0$ ,  $B_q(\tau) \Rightarrow 0$  при  $\tau \neq 0$ , т.е. статистические шумы не коррелированы ( $\tau_k \Rightarrow 0$ ).

Практические вычисления АКФ финитных сигналов обычно ограничиваются интервалом сдвигов  $\tau = \{0, (3-5)\tau_k\}$ , в котором, как правило, сосредоточена основная информация по автокорреляции сигналов.

**Спектральная плотность ВКФ** может быть получена на основании тех же соображений, что и для АКФ, или непосредственно из формулы (18.102) заменой спектральной плотности сигнала  $S(\omega)$  на спектральную плотность второго сигнала  $U(\omega)$ :

$$B_{su}(\tau) = (1/2\pi) \int_{-\infty}^{\infty} S^*(\omega) U(\omega) \exp(j\omega\tau) d\omega. \quad (18.106)$$

Или, при смене порядка сигналов:

$$B_{us}(\tau) = (1/2\pi) \int_{-\infty}^{\infty} U^*(\omega) S(\omega) \exp(j\omega\tau) d\omega. \quad (18.107)$$

Произведение  $S^*(\omega)U(\omega)$  представляет собой взаимный энергетический спектр  $W_{su}(\omega)$  сигналов  $s(t)$  и  $u(t)$ . Соответственно,  $U^*(\omega)S(\omega)=W_{us}(\omega)$ . Следовательно, как и АКФ, взаимнокорреляционная функция и спектральная плотность взаимной

мощности сигналов связаны между собой преобразованиями Фурье:

$$B_{su}(\tau) \Leftrightarrow W_{su}(\omega) \equiv W_{us}^*(\omega). \quad (18.108)$$

$$B_{us}(\tau) \Leftrightarrow W_{us}(\omega) \equiv W_{su}^*(\omega). \quad (18.109)$$

В общем случае, за исключением спектров четных функций, из условия несоблюдения четности для функций ВКФ следует, что взаимные энергетические спектры являются комплексными функциями:

$$U(\omega) = A_u(\omega) + j B_u(\omega), \quad V(\omega) = A_v(\omega) + j B_v(\omega).$$

$$W_{uv} = A_u A_v + B_u B_v + j(B_u A_v - A_u B_v) = \text{Re } W_{uv}(\omega) + j \text{Im } W_{uv}(\omega),$$

и содержат определенную фазовую характеристику гармонических составляющих ВКФ, которой и формируется сдвиг максимума ВКФ.

На рис. 18.59 можно наглядно видеть особенности формирования ВКФ на примере двух одинаковых по форме сигналов, сдвинутых относительно друг друга.

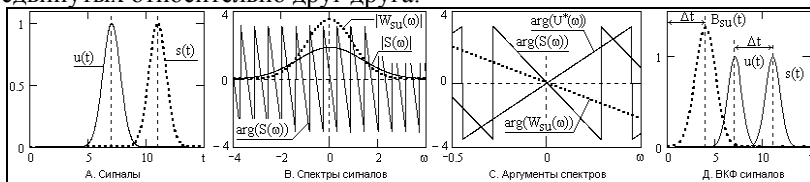


Рис. 18.59. Формирование ВКФ.

Форма сигналов и их взаимное расположение приведены на виде А. Модуль и аргумент спектра сигнала  $s(t)$  приведены на виде В. Модуль спектра  $u(t)$  тождественен модулю  $S(\omega)$ . На этом же виде приведен модуль спектра взаимной мощности сигналов  $S(\omega)U^*(\omega)$ . Как известно, при перемножении комплексных спектров модули спектров перемножаются, а фазовые углы складываются, при этом для сопряженного спектра  $U^*(\omega)$  фазовый угол меняет знак. Если первым в формуле вычисления ВКФ (18.95) стоит сигнал  $s(t)$ , а сигнал  $u(t-\tau)$  на оси ординат стоит впереди  $s(t)$ , то фазовые углы  $S(\omega)$  по мере увеличения частоты нарастают в сторону отрицательных значений углов (без учета периодического сброса значений на  $2\pi$ ), а фазовые углы  $U^*(\omega)$  по абсолютным значениям меньше фазовых углов  $s(t)$  и нарастают (за счет сопряжения) в сторону положительных значений. Результатом умножения спектров (как это видно на рис. 18.58, вид С) является вычитание из фазовых углов  $S(\omega)$  значений углов  $U^*(\omega)$ , при этом фазовые углы спектра  $S(\omega)U^*(\omega)$  остаются в области отрицательных значений, что обеспечивает сдвиг всей функции ВКФ (и ее пиковых значений) вправо от нуля по оси  $\tau$  на определенную величину (для одинаковых сигналов – на величину разности между сигналами по оси ординат). При смещении начального положения

сигнала  $u(t)$  в сторону сигнала  $s(t)$  фазовые углы  $S(\omega)U^*(\omega)$  уменьшаются, в пределе до нулевых значений при полном совмещении сигналов, при этом функция  $V_{su}(t)$  смещается к нулевым значениям  $\tau$ , в пределе до обращения в АКФ (для одинаковых сигналах  $s(t)$  и  $u(t)$ ).

Как известно для детерминированных сигналов, если спектры двух сигналов не перекрываются и, соответственно, взаимная энергия сигналов равна нулю, такие сигналы ортогональны друг другу. Связь энергетических спектров и корреляционных функций сигналов показывает еще одну сторону взаимодействия сигналов. Если спектры сигналов не перекрываются и их взаимный энергетический спектр равен нулю на всех частотах, то при любых временных сдвигах  $\tau$  друг относительно друга их ВКФ также равна нулю. А это означает, что такие сигналы являются некоррелированными. Это действительно как для детерминированных, так и для случайных сигналов и процессов.

**Вычисление корреляционных функций при помощи**

**БПФ** является, особенно для длинных числовых рядов, в десятки и сотни раз более быстрым методом, чем последовательными сдвигами во временной области при больших интервалах корреляции. Суть метода вытекает из формул (18.103) для АКФ и (18.108) для ВКФ. Учитывая, что АКФ можно рассматривать как частный случай ВКФ при одном и том же сигнале, процесс вычисления рассмотрим на примере ВКФ для сигналов  $x(k)$  и  $y(k)$  с числом отсчетов  $K$ . Он включает:

1. Вычисление БПФ спектров сигналов  $x(k) \rightarrow X(k)$  и  $y(k) \rightarrow Y(k)$ . При разном количестве отсчетов более короткий ряд дополняется нулями до размера большего ряда.
2. Вычисление спектров плотности мощности  $W_{xy}(k) = X^*(k) Y(k)$ .
3. Обратное БПФ  $W_{xy}(k) \rightarrow V_{xy}(k)$ .

Отметим некоторые особенности метода.

При обратном БПФ, как известно, вычисляется циклическая свертка функций  $x(k) \otimes y(k)$ . Если число отсчетов функций равно  $K$ , число комплексных отсчетов спектров функций также равно  $K$ , равно как и число отсчетов их произведения  $W_{xy}(k)$ . Соответственно, число отсчетов  $V_{xy}(k)$  при обратном БПФ также равно  $K$  и циклически повторяется с периодом, равным  $K$ . Между тем, при линейной свертке полных массивов сигналов по формуле (18.100) размер только одной половины ВКФ составляет  $K$  точек, а полный двусторонний размер составляет  $2K$  точек. Следовательно, при обратном БПФ с учетом циклическости свертки произойдет наложение на главный период ВКФ ее боковых периодов, как и при обычной циклической свертке двух функций.

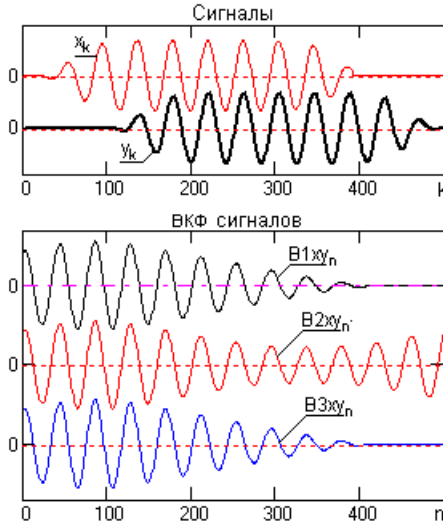


Рис. 18.60.

$V1$  – линейная свертка,  $V2$  – БПФ без продления сигналов нулями,  $V3$  – БПФ с продлением сигналов нулями.

На рис. 18.60 приведен пример двух сигналов и значения ВКФ, вычисленные линейной сверткой ( $V1x$ ) и циклической сверткой через БПФ ( $V2x$ ). Для исключения эффекта наложения боковых периодов необходимо дополнить сигналы нулями, в пределе, до удвоения количества отсчетов, при этом результат БПФ (график  $V3x$  на рисунке 18.60) полностью повторяет результат линейной свертки (с учетом нормировки на увеличение количества отсчетов).

На практике число нулей продления сигналов зависит от характера корреляционной функции. Минимальное количество нулей обычно принимается равным значимой информационной части функций, т.е. порядка (3-5) интервалов корреляции.

## 19. Дискретизация сингалов и преобразований

### 19.1. Введение в дискретизацию сингалов

В первой половине XX века при регистрации и обработке информации использовались, в основном, измерительные приборы и устройства аналогового типа, работающие в реальном масштабе

времени, при этом даже для величин, дискретных в силу своей природы, применялось преобразование дискретных сигналов в аналоговую форму. Положение изменилось с распространением микропроцессорной техники и ЭВМ. Цифровая регистрация и обработка информации оказалась более совершенной и точной, более универсальной, многофункциональной и гибкой. Мощь и простота цифровой обработки сигналов настолько преобладают над аналоговой, что преобразование аналоговых по природе сигналов в цифровую форму стало производственным стандартом.

**Под дискретизацией сигналов понимают преобразование функций непрерывных переменных в функции дискретных переменных, по которым исходные непрерывные функции могут быть восстановлены с заданной точностью.**

Роль дискретных отсчетов выполняют, как правило, квантованные значения функций в дискретной шкале координат.

**Под квантованием понимают преобразование непрерывной по значениям величины в величину с дискретной шкалой значений из конечного множества разрешенных, которые называют уровнями квантования.**

Если уровни квантования нумерованы, то результатом преобразования является число, которое может быть выражено в любой числовой системе. Округление с определенной разрядностью мгновенных значений непрерывной аналоговой величины с равномерным шагом по аргументу является простейшим случаем дискретизации и квантования сигналов при их преобразовании в цифровые сигналы.

Как правило, для производственных задач обработки данных обычно требуется значительно меньше информации, чем ее поступает от измерительных датчиков в виде непрерывного аналогового сигнала. При статистических флуктуациях измеряемых величин и конечной погрешности средств измерений точность регистрируемой информации также всегда ограничена определенными значениями. При этом рациональное выполнение дискретизации и квантования исходных данных дает возможность снизить затраты на хранение и обработку информации. Кроме того, использование цифровых сигналов позволяет применять методы кодирования информации с возможностью последующего обнаружения и исправления ошибок при обращении информации, а цифровая форма сигналов облегчает унификацию операций преобразования информации на всех этапах ее обращения.

## **19.2. Задачи дискретизации функций**

### **Сигналы и системы дискретного времени.**

Значения дискретного сигнала определены только при дискретных значениях времени или любой другой независимой переменной. Обычно его представляют в виде последовательности чисел:  $s(k) \equiv s(k\Delta t) \equiv s_k$ ,  $k = 0, 1, 2, \dots, K$ , где значениями чисел отображают значения сигнала в дискретные моменты времени. Значения интервала дискретизации обычно принято опускать, т.е. принимать равным  $\Delta t = 1$ , поскольку он является не более чем масштабным множителем по независимой переменной и при постоянном значении во всех параметрах и атрибутах обработки сигналов, включая сопряженные величины (например, масштаб частоты  $f = 1/\Delta t$ ), его физическая величина может вводиться в результаты на заключительной стадии обработки данных. По существу, при  $\Delta t = 1$  осуществляется нормирование сигналов и систем их обработки по независимой переменной.

**Система дискретного времени** – это алгоритм с входной последовательностью  $s(k)$  и выходной последовательностью  $y(k)$ , которая может быть линейной или нелинейной, инвариантной или изменяющейся во времени. Система дискретного времени линейна и инвариантна во времени (ЛИВ-система), если она подчиняется принципу суперпозиции (отклик на несколько входов равен сумме откликов на каждый вход в отдельности), а задержка (сдвиг) входного сигнала вызывает такую же задержку выходного сигнала. Вход и выход ЛИВ-систем связывает сверточная сумма:

$$y(k) = \sum_{n=-\infty}^{\infty} h(n) x(k-n),$$

где  $h(n)$  – дискретная импульсная характеристика (импульсный отклик) системы. Система устойчива, если выполняется условие:

$$\sum_{n=-\infty}^{\infty} |h(n)| < \infty.$$

Это условие справедливо всегда для систем с конечной импульсной характеристикой (КИХ-систем) без особых точек в своем составе, что характерно для нерекурсивных систем с ограниченным числом отсчетов (в общем случае,  $N_1 < n < N_2$ ), а также для систем с бесконечной импульсной характеристикой (БИХ-систем), если  $h(n) \rightarrow 0$  при  $n \rightarrow \infty$ , что должно выполняться для рекурсивных систем.

**Физически реализуемой** называется система, если ее импульсная характеристика существует только при  $n \geq 0$ .

**Принципы дискретизации.** Сущность дискретизации аналоговых сигналов заключается в том, что непрерывность во

времени аналоговой функции  $s(t)$  заменяется последовательностью коротких импульсов, амплитудные значения которых  $c_n$  определяются с помощью весовых функций, либо непосредственно выборками (отсчетами) мгновенных значений сигнала  $s(t)$  в моменты времени  $t_n$ . Представление сигнала  $s(t)$  на интервале  $T$  совокупностью дискретных значений  $c_n$  записывается в виде:

$$(c_1, c_2, \dots, c_N) = A[s(t)],$$

где  $A$  - оператор дискретизации. Запись операции восстановления сигнала  $s(t)$ :

$$s'(t) = B[(c_1, c_2, \dots, c_N)].$$

Выбор операторов  $A$  и  $B$  определяется требуемой точностью восстановления сигнала. Наиболее простыми являются линейные операторы. В общем случае:

$$c_n = \int_t q_n(t) s(t) dt, \quad (19.1)$$

где  $q_n(t)$  - система весовых функций.

Отсчеты в выражении (19.1) связаны с операцией интегрирования, что обеспечивает высокую помехоустойчивость дискретизации. Однако в силу сложности технической реализации "взвешенного" интегрирования, последнее используется достаточно редко, при высоких уровнях помех. Более широкое распространение получили методы, при которых сигнал  $s(t)$  заменяется совокупностью его мгновенных значений  $s(t_n)$  в моменты времени  $t_n$ . Роль весовых функций в этом случае выполняют гребневые (решетчатые) функции. Отрезок времени  $\Delta t$  между соседними отсчетами называют шагом дискретизации. Дискретизация называется равномерной с частотой  $F=1/\Delta t$ , если значение  $\Delta t$  постоянно по всему диапазону преобразования сигнала. При неравномерной дискретизации значение  $\Delta t$  между выборками может изменяться по определенной программе или в зависимости от изменения каких-либо параметров сигнала.

**Воспроизведение непрерывного сигнала** по выборкам может проводиться как на основе ортогональных, так и неортогональных базисных функций. Воспроизводящая функция  $s'(t)$  соответственно представляется аппроксимирующим полиномом:

$$s'(t) = \sum_n c_n v_n(t), \quad (19.2)$$

где  $v_n(t)$  - система базисных функций. Ортогональные базисные функции обеспечивают сходимость ряда к  $s(t)$  при  $n \rightarrow \infty$ .

**Оптимальными являются методы дискретизации, обеспечивающие минимальный числовой ряд при заданной погрешности воспроизведения сигнала.** При неортогональных базисных функциях используются, в основном, степенные



алгебраические полиномы вида:

$$s'(t) = \sum_{n=0}^N c_n t^n. \quad (19.3)$$

**Если значения аппроксимирующего полинома совпадают со значениями выборок в моменты их отсчета, то такой полином называют интерполирующим.** В качестве интерполирующих полиномов обычно используются многочлены Лагранжа. Для реализации интерполирующих полиномов необходима задержка сигнала на интервал дискретизации, что в системах реального времени требует определенных технических решений. В качестве экстраполирующих полиномов используют, как правило, многочлены Тейлора.

Естественным требованием к выбору частоты дискретизации является внесение минимальных искажений в динамику изменения сигнальных функций. Логично полагать, что искажения информации будут тем меньше, чем выше частота дискретизации  $F$ . С другой стороны также очевидно, что чем больше значение  $F$ , тем большим количеством цифровых данных будут отображаться сигналы, и тем большее время будет затрачиваться на их обработку. В оптимальном варианте значение частоты дискретизации сигнала  $F$  должно быть необходимым и достаточным для обработки информационного сигнала с заданной точностью, т.е. обеспечивающим допустимую погрешность восстановления аналоговой формы сигнала (среднеквадратическую в целом по интервалу сигнала, либо по максимальным отклонениям от истинной формы в характерных информационных точках сигналов).

### **19.3. Равномерная дискретизация**

**Спектр дискретного сигнала.** Допустим, что для обработки задается произвольный аналоговый сигнал  $s(t)$ , имеющий конечный и достаточно компактный фурье-образ  $S(f)$ . Равномерная дискретизация непрерывного сигнала  $s(t)$  с частотой  $F$  (шаг  $\Delta t = 1/F$ ) с математических позиций означает умножение функции  $s(t)$  на гребневую функцию  $\text{Ш}_{\Delta t}(t) = \sum_k \delta(t - k\Delta t)$  – непрерывную последовательность импульсов Кронекера:

$$s_{\Delta t}(t) = s(t) \cdot \text{Ш}_{\Delta t}(t) = s(t) \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \delta(t - k\Delta t). \quad (19.4)$$

С учетом известного преобразования Фурье гребневой функции

$$\text{Ш}_{\Delta t}(t) \Leftrightarrow (1/T) \sum_{n=-\infty}^{\infty} \delta(f-nF) = F \cdot \text{Ш}_F(f), \quad (19.5)$$

Фурье-образ дискретной функции  $s_{\Delta t}(t)$ :

$$S_F(f) = S(f) * F \cdot \text{Ш}_F(f). \quad (19.6)$$

Отсюда, для спектра дискретного сигнала имеем:

$$S_F(f) = F \cdot S(f) * \sum_{n=-\infty}^{\infty} \delta(f-nF) = F \sum_{n=-\infty}^{\infty} S(f-nF). \quad (19.7)$$

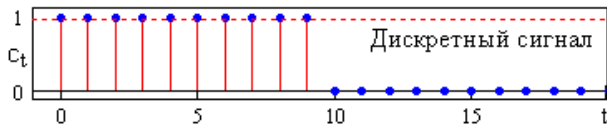
Из выражения следует, что спектр дискретного сигнала представляет собой *непрерывную периодическую функцию* с периодом  $F$ , совпадающую (при определенных условиях конечности спектра непрерывного сигнала) с функцией  $F \cdot S(f)$  непрерывного сигнала  $s(t)$  в пределах центрального периода от  $-f_N$  до  $f_N$ , где  $f_N = 1/2\Delta t = F/2$ . Частоту  $f_N$  (или для круговой частоты  $\omega_N = \pi/\Delta t$ ) называют частотой Найквиста. Центральный период функции  $S_F(f)$  называют главным частотным диапазоном.

Интуитивно понятно, что если спектр главного частотного диапазона с точностью до постоянного множителя совпадает со спектром непрерывного сигнала, то по этому спектру может быть восстановлена не только форма дискретного сигнала, но и форма исходного непрерывного сигнала. При этом шаг дискретизации и соответствующее ему значение частоты Найквиста должны иметь определяющее значение. Как правило, шаг дискретизации сигнала (шаг числовых массивов) условно принимают равным  $\Delta t=1$ , при этом главный частотный диапазон занимает интервал  $-0.5 \leq f \leq 0.5$ , или, в шкале угловых частот, соответственно  $-\pi \leq \omega \leq \pi$ .

Физическая сущность формирования спектров дискретных сигналов достаточно проста. Наиболее наглядно это можно увидеть, если воспользоваться программой Mathcad (см. рис. 19.1).

Сначала представим себе непрерывный сигнал постоянной единичной амплитуды  $s(t) = \text{const} = 1$  на произвольном интервале  $0-T$ , например, при  $T=100$ . Начнем дискретизировать сигнал с равномерным шагом  $\Delta t=1$ . Вычислим спектр первого дискретного отсчета  $c_0 = 1$ . При  $N=1$  сигнал является импульсом Кронекера, а, соответственно, модуль спектра отсчета  $c_0=1$  представляет собой непрерывное частотное распределение  $|C(\omega)| = \text{const}$  в диапазоне от  $-\infty$  до  $+\infty$  (показан только участок от  $-6\pi$  до  $+6\pi$  с нормировкой на  $N$  для наглядности сравнения спектров). Все частоты сигнала имеют нулевую фазу и при сложении взаимно компенсируются во всех временных точках за исключением точки  $t=0$ , в которой амплитуды частот суммируются, создавая единичный отсчет  $c_0$ .

$T := 100$     $t := 0..T$     $c_T := 0$     $N := 10$     $\leftarrow$  Кол-во задаваемых импульсов Кронекера  
 $n := 0..N-1$     $c_n := 1$



$\Omega := 20$     $\Delta\omega := \frac{\pi}{T}$     $\omega := -\Omega, -\Omega + \Delta\omega .. \Omega$     $\leftarrow$  Задание границ и шага расчета спектра

$$C(\omega) := \frac{1}{N} \sum_{t=0}^{N-1} c_t \cdot \exp(-j \cdot \omega \cdot t) \quad \leftarrow \text{Спектральная функция}$$

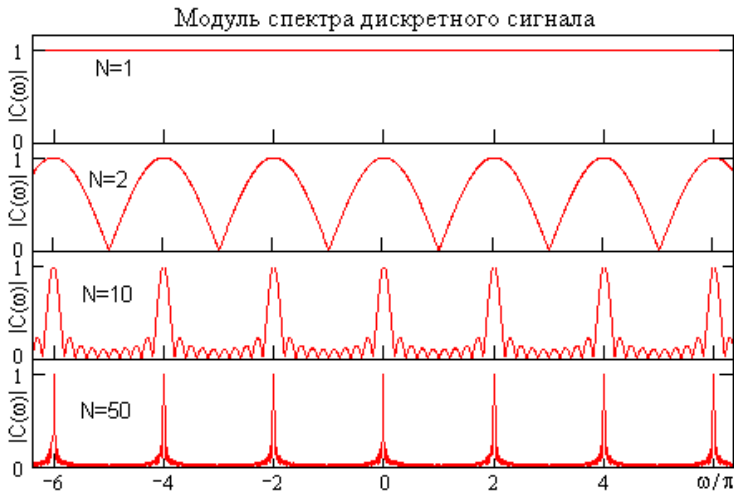


Рис. 19.1. Формирование спектра дискретного сигнала.

Добавим к сигналу второй дискретный отсчет  $c_1=1$  ( $N=2$ ). Если вычислить спектр только второго отсчета, то его модуль будет равен модулю первого отсчета (так как  $c_1=c_0$ ), но нулевые фазы гармоник этого спектра переместятся в точку  $t=1$ , т.е. относительно точки  $t=0$  фазы гармоник второго отсчета изменятся на  $-\omega\Delta t$  в соответствии с теоремой запаздывания преобразования Фурье. При сложении этих двух спектров первого и второго отсчета наблюдается интерференция частот и возникают пульсации частотного спектра с максимумами на частотах, кратных  $F=1/\Delta t$  или в угловых единицах  $2\pi/\Delta t$ , где фазы спектров первого и второго отсчетов совпадают и равны нулю. Форма модуля результирующего спектра при  $N=2$  приведена на рисунке.

При дальнейшем увеличении количества отсчетов периодичность совпадения нулевых фаз и положения максимумов сохраняется, а интерференция частот между максимумами усложняется, при этом ширина главных пиков по всему частотному диапазону спектра от минус до плюс бесконечности становится все уже. На рис. 19.1 приведены примеры спектров сигналов при  $N=10$  и  $N=50$ . В пределе, при двусторонней временной шкале  $\pm T \rightarrow \pm \infty$  и  $N \rightarrow \infty$ , гребневая функция из импульсов Кронекера во временной

области  $c_t \rightarrow \text{Ш}_{\Delta t}(t) = \sum_{k=-\infty}^{\infty} \delta(t-k\Delta t)$  превращается в идеальную

гребневую функцию  $(1/T) \sum_{n=-\infty}^{\infty} \delta(f-nF) = F \cdot \text{Ш}_F(f)$  в частотной области

(формула 19.5). Этот спектр непрерывен и физически реален в диапазоне частот от  $-\infty$  до  $+\infty$ .

Физический смысл интерференции частот остается тем же самым, если мы на произвольном интервале  $T$  зададим произвольный сигнал, например – синусоиду  $u(t) \Leftrightarrow U(f)$ , и выполним его дискретизацию, т.е. умножим сигнал на непрерывную последовательность импульсов Кронекера

$$c(t) \cdot u(t) \rightarrow u(t) \sum_{k=0}^K \delta(t-k\Delta t) = u(t) \cdot \text{Ш}_{\Delta t}(t).$$

А так как каждый дискретный отсчет в этом случае имеет свою определенную амплитуду и, соответственно, свой уровень амплитуд гармоник своего спектра, то сложение частот дает более сложную картину интерференции с расщеплением спектра общего сигнала всех дискретных отсчетов на две зеркальных составляющих относительно кратных частот  $2\pi/\Delta t$ .

Математически произведение двух функций во временной области отображается сверткой спектров этих функций в частотном представлении, т.е. сверткой спектра сигнала  $u(t)$  с частотной гребневой функцией спектра, порожденной временной гребневой функцией дискретизации  $u(t)\text{Ш}_{\Delta t}(t) \Leftrightarrow U(f) * F \cdot \text{Ш}_F(f)$ , откуда и следует формула (19.7). Пример дискретизации одного периода синусоиды приведен на рис. 19.2.

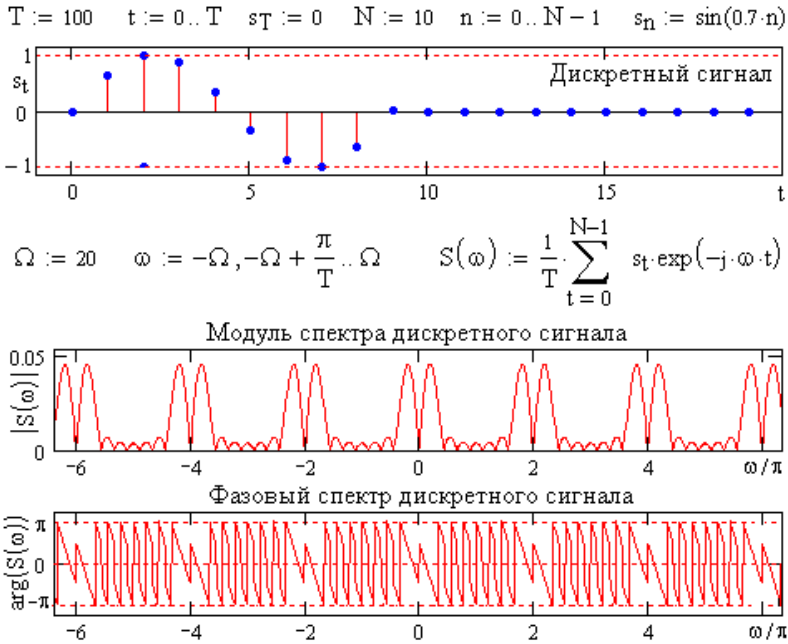


Рис. 19.2. Формирование спектра дискретного сигнала.

Вернемся к значению и роли частоты Найквиста при дискретизации сигналов.

На рис. 19.3 и 19.4. приведены примеры равномерной дискретизации аналоговых сигналов  $s_1(t) = \exp(-a|t|)$  и  $s_2(t) = \exp(-bt^2)$  (дискретные отсчеты нанесены кружками) и спектры этих дискретных сигналов.

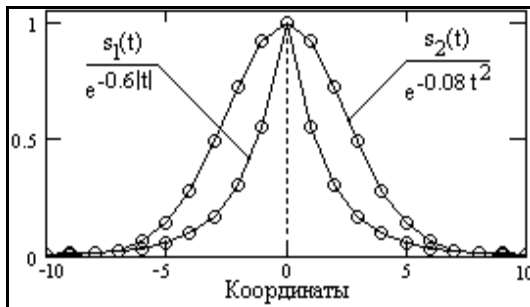


Рис. 19.3. Дискретные сигналы.

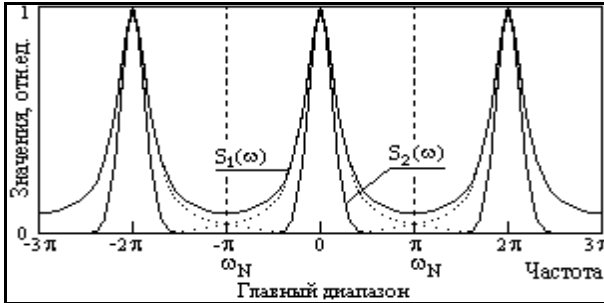


Рис. 19.4. Спектры дискретных сигналов.

Для того чтобы периодическое повторение спектра, вызванное дискретизацией аналогового сигнала, не изменяло спектр в главном частотном диапазоне (по отношению к спектру исходного аналогового сигнала), необходимо и достаточно, чтобы максимальные частотные составляющие  $f_{\max}$  в спектре аналогового сигнала не превышали частоты Найквиста ( $f_{\max} \leq f_N = F/2$ ). Это означает, что частота дискретизации сигнала должна быть минимум в два раза выше максимальной частотной составляющей в спектре сигнала:

$$F = 1/\Delta t \geq 2f_{\max}, \quad (19.8)$$

что обеспечивает выход спектра на нулевые значения на концах главного диапазона, как это имеет место для спектра  $S_2(\omega)$  на рис. 19.4.

Другими словами, на одном периоде колебаний с частотой  $f_{\max}$  должно быть минимум две точки отсчета. Это и понятно – по одной точке отсчета на периоде гармонического сигнала определение неизвестных параметров данной гармоники (амплитуда, фаза) невозможно.

Если условие (19.8) нарушается, искажения частотного спектра исходного аналогового сигнала неизбежны. На рис. 19.4 наглядно видно, что частота дискретизации для сигнала  $s_1(t)$  данному условию не удовлетворяет, спектры периодов перекрылись, и результирующий спектр дискретных отсчетов сигнала  $s_1(t)$  отличается от фактического спектра сигнала (фактический спектр и его периодические повторения в области перекрытия спектра главного частотного диапазона со спектрами боковых диапазонов показаны пунктиром). Аналоговый сигнал из спектра  $S_1(\omega)$  будет восстановлен с искажениями.

Характер возникающих искажений во временной области при нарушении условия (19.8) можно наглядно видеть на рис. 19.5.

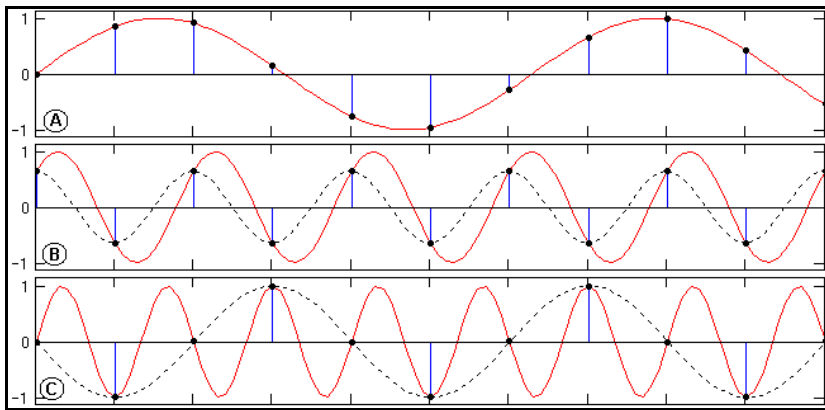


Рис. 19.5. Дискретизация гармоник с разной частотой.

На рисунке показаны три возможных варианта соотношения частот гармонических сигналов с постоянной частотой их дискретизации.

1. График А – частота гармонического сигнала меньше частоты Найквиста. Дискретным отсчетам может соответствовать только исходная гармоника, амплитуда, частота и фаза которой могут быть однозначно определены по любым трем последовательным точкам (три уравнения, три неизвестных).

2. График В – частота гармонического сигнала равна частоте Найквиста. Это означает периодическое повторение каждой пары последовательных отсчетов, а, следовательно, для решения имеется только два уравнения с тремя неизвестными с возможностью определения только частоты, и то при условии, что начальная фаза сигнала не совпадает с начальной фазой частоты дискретизации (в этом случае все отсчеты нулевые). Амплитуда и фаза сигнала определяются однозначно только при условии совпадения отсчетов с экстремумами гармоники.

3. График С – частота гармонического сигнала больше частоты Найквиста. Решение трех уравнений по трем последовательным точкам позволяет определить амплитуду гармоники, но дает искаженные значения частоты и фазы колебания (показано пунктиром). Это так называемый эффект *появления ложных (кажущихся) частот* (aliasing). Частоты гармонических колебаний выше частоты Найквиста как бы зеркально "отражаются" в главный частотный диапазон от его границ (на частоте Найквиста), что можно видеть на рис. 19.4 для действительного спектра сигнала  $S_1(\omega)$ , показанного точками. Этот эффект аналогичен всем известному эффекту обратного вращения

колес автомобиля (и любых других быстро вращающихся объектов) на экранах кино и телевизоров, когда скорость их вращения начинает превышать частоту смены кадров.

**Интерполяционный ряд Котельникова-Шеннона.**

Спектр дискретизированного сигнала (19.7) представляет собой сумму сдвинутых копий исходного аналогового сигнала с шагом сдвига, равным частоте дискретизации. Очевидно, что если спектры копий не перекрываются, то по центральной копии дискретного спектра можно восстановить исходный аналоговый сигнал с абсолютной точностью. Умножая функцию (19.6) на прямоугольную весовую функцию  $\Pi_F(f)$ , равную 1 в пределах главного частотного диапазона  $[-F/2, F/2]$  и нулю за его пределами, получаем непрерывный спектр в бесконечных по частоте границах, равный спектру  $F \cdot S(f)$  в пределах главного частотного диапазона:

$$F \cdot S(f) = F \cdot [S(f) * \text{Ш}_F(f)] \cdot \Pi_F(f). \quad (19.9)$$

Обратное преобразование Фурье такого спектра должно давать конечный и непрерывный сигнал. Произведем обратное преобразование обеих частей равенства (19.9):

$$F \cdot [S(f) * \text{Ш}_F(f)] \Leftrightarrow s_{\Delta t}(t), \quad \Pi_F(f) \Leftrightarrow F \cdot \text{sinc}(\pi Ft).$$

$$F \cdot s(t) = s_{\Delta t}(t) * F \cdot \text{sinc}(\pi Ft).$$

$$s(t) = \text{sinc}(\pi Ft) * \sum_{k=-\infty}^{\infty} s(k\Delta t) \delta(t-k\Delta t),$$

Дискретизированный сигнал  $s_{\Delta t}(t) = \sum_k s(k\Delta t) \delta(t-k\Delta t)$

представляет собой сумму последовательных весовых импульсов Кронекера, сдвинутых на интервал  $\Delta t$ , со значениями веса, равными значениям отсчетов функции  $s(t)$  в моменты  $k\Delta t$ . При прохождении такого сигнала через систему с импульсным откликом

$$h(t) = \text{sinc}(\pi Ft) = \sin(\pi Ft) / \pi Ft$$

каждый весовой импульс Кронекера возбудит на выходе соответствующую последовательную серию сдвинутых и масштабированных копий оператора фильтра. Отсюда, с учетом очевидного равенства

$$\delta(t-k\Delta t) * \text{sinc}(\pi Ft) = \text{sinc}[\pi F(t-k\Delta t)],$$

выходной сигнал будет представлять собой сумму сдвинутых весовых импульсных откликов системы, где значение веса определяется отсчетами дискретного сигнала:

$$s(t) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \text{sinc}[\pi F(t-k\Delta t)] = \sum_{k=-\infty}^{\infty} s(k\Delta t) \text{sinc}[\pi(t/\Delta t - k)]. \quad (19.10)$$



Эта конечная формула носит название интерполяционного ряда Котельникова-Шеннона. Из нее следует, что если наибольшая частота в спектре произвольной непрерывной функции  $s(t)$  не превышает частоты ее дискретизации, то она без потери точности может быть представлена в виде числовой последовательности дискретных значений  $s(k\Delta t)$ ,  $k = 0, 1, 2, \dots$ , и однозначно восстановлена по этой последовательности. В этом и состоит сущность теоремы отсчетов Котельникова. В зарубежной литературе она называется также теоремой Шеннона или теоремой дискретизации (sampling theorem).

По существу, ряд (19.10) представляет собой частный случай разложения сигнала в соответствии с формулой (19.2) по системе ортогональных функций интегрального синуса

$$v(t, k\Delta t) = \text{sinc}(\pi F(t - k\Delta t)) = \text{sinc}(\pi(t/\Delta t - k)),$$

образующих базис пространства сигналов  $s(t)$ . Для проверки ортогональности достаточно вычислить скалярное произведение базисных функций:

$$\int_{-\infty}^{\infty} v(t, n\Delta t) v(t, m\Delta t) dt = \begin{cases} \Delta t, & n = m \\ 0, & n \neq m \end{cases}.$$

Разложение (19.10) проще и понятнее, чем разложение в ряды Фурье, что можно видеть на рис. 19.6.

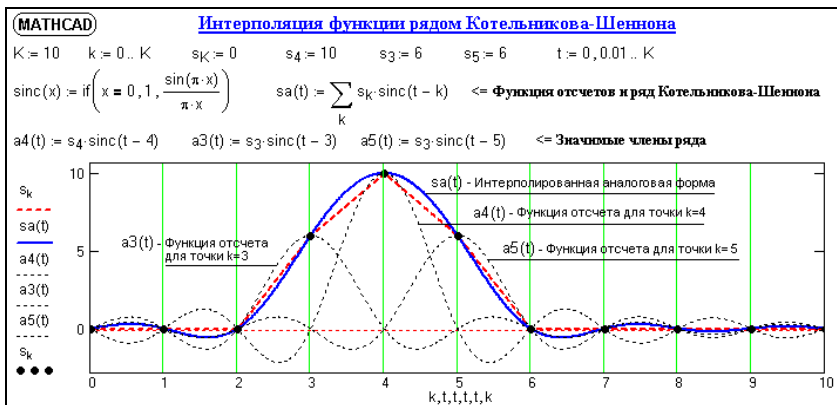


Рис. 19.6. Восстановление непрерывного сигнала по дискретным отсчетам.

Вес каждой функции отсчетов  $\text{sinc}[\pi F(t-k\Delta t)]$  формирует пиковое значение интегрального синуса в каждой текущей точке  $t = k\Delta t$ , равное значению сигнала  $s(k\Delta t)$ , при этом во всех остальных точках дискретных отсчетов  $\text{sinc}[\pi F(t-(k\pm j)\Delta t)]$ ,  $j = 1, 2, \dots$  значения интегрального синуса равны нулю. Ряд числовых значений интегрального синуса для дискретных значений  $t = n\Delta t$  при суммировании по  $k$  полностью эквивалентен гребневой функции:

$$\sum_{k=-\infty}^{\infty} \text{sinc}[\pi F(n\Delta t - k\Delta t)] \equiv \text{III}_{\Delta t}(t).$$

Однако, в отличие от гребневой функции, в интервале между дискретными отсчетами интегральный синус имеет не нулевые, а определенные осциллирующие значения. Суперпозицией этих значений по текущим значениям  $t$  от всех интегральных синусов, осцилляции которых доходят до данного значения  $t$ , и образуются значения аналогового сигнала в интервалах между отсчетами.

В принципе, функции отсчетов имеют бесконечные осцилляции, и восстанавливают аналоговый сигнал, бесконечный по аргументу. Амплитуда осцилляций функций отсчетов затухает достаточно медленно (см. рис. 19.7).

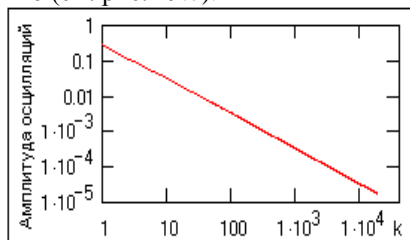


Рис. 19.7.

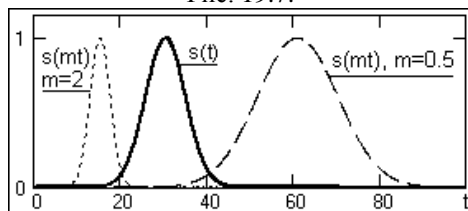


Рис. 19.8. Изменение масштаба при восстановлении аналоговой функции.

Однако на рис. 19.6 нетрудно заметить, что, в силу знакопеременности функций отсчетов по интервалам дискретизации, осцилляции восстанавливаемых кривых с финитным спектром

затухают достаточно быстро, и для данных без существенных выбросов и больших перепадов значений определяются, в основном, отсчетами, ближайшими к интерполируемому интервалу. Это позволяет ограничивать интервал суммирования в формуле (18.62) определенными окрестностями текущих точек интерполяции.

Ряд (19.10) позволяет простым введением масштабного множителя в аргумент интегрального синуса изменять представление сигнала на временной оси, растягивать или сжимать сигнал:

$$s(t) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \times \text{sinc}[\pi F(mt-k\Delta t)].$$

По аналогичной формуле может выполняться пересчет дискретных данных на другой интервал дискретизации:

$$s(n \cdot \Delta t_{\text{new}}) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \times \text{sinc}[\pi F(n \cdot \Delta t_{\text{new}} - k\Delta t)].$$

Примеры восстановления аналоговой формы произвольного финитного сигнала и изменения шага дискретизации данных приведены на рис. 19.9.

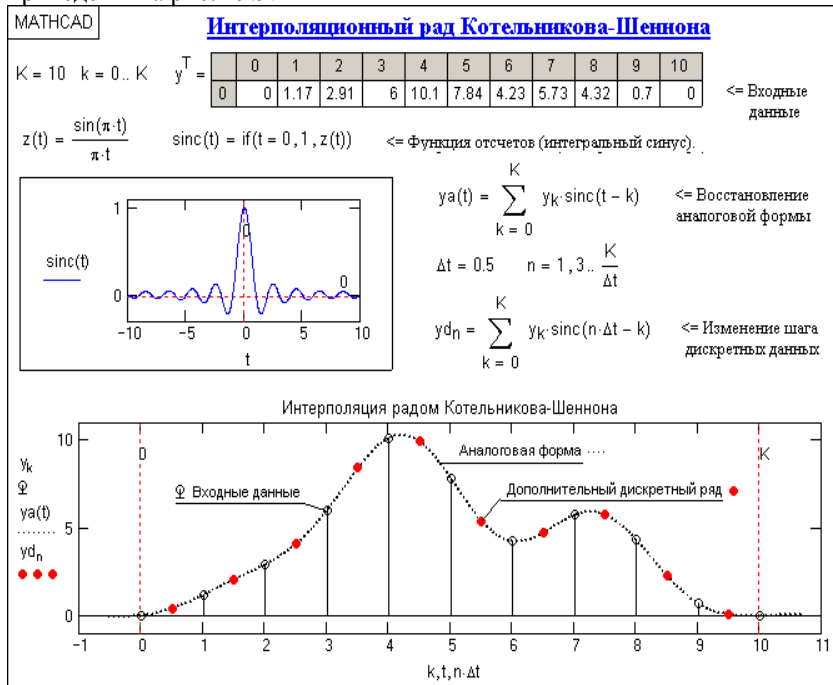


Рис. 19.9.

На рис. 19.10 приведено моделирование дискретизации аналогового сигнала, влияние наложение спектров боковых периодов на спектр главного диапазона дискретного сигнала и восстановление из этого спектра аналоговой формы сигнала.

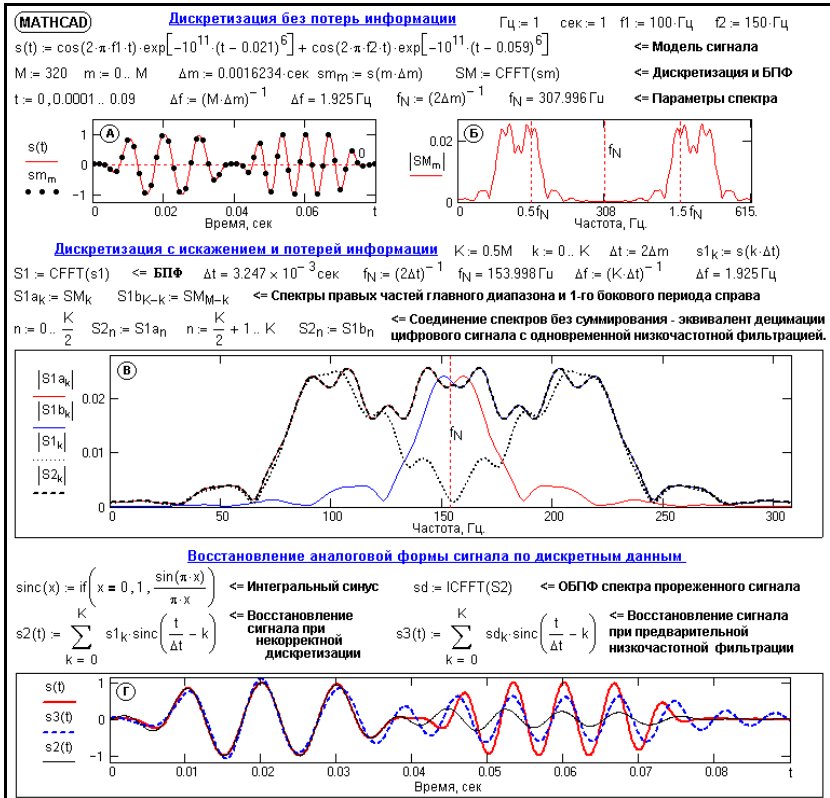


Рис. 19.10. Моделирование дискретизации аналогового сигнала.

Графики А и Б рисунка – модельный аналоговый сигнал, точки его дискретизации и модуль спектра дискретного сигнала. Вычисление спектра выполнено быстрым преобразованием Фурье (БПФ) и отображает, соответственно, частотный диапазон  $0-2f_N$ . Дискретизация выполнена корректно, с выполнением условия (19.8), о чем можно судить и по спектру дискретного сигнала (график Б, выход на незначимые значения к частоте Найквиста  $f_N$ ).

Кривая S1 на графике В – спектр модельного дискретного сигнала при нарушении условия (19.8). В данном случае это

произойдет при увеличении шага дискретизации в 2 раза, что вызовет уменьшение в 2 раза новой частоты Найквиста и перемещение границы главного диапазона на отметку  $0.5f_N$  на графике Б, при этом произойдет перекрытие спектров поддиапазонов. На графике приведены кривые S1a и S1b, которые являются отдельными спектрами правой половины главного диапазона без сложения со спектром правого бокового диапазона (интервал  $0-2f_N$ , где  $f_N$  – частота Найквиста новой дискретизации), и левой половины правого бокового диапазона на том же интервале  $0-2f_N$  без сложения со спектром главного диапазона. Хорошо видны «хвосты» спектров, выходящие за границы интервала Найквиста от центров диапазонов и заходящие в соседние диапазоны. Сложением этих спектров в интервале  $0-2f_N$  нетрудно убедиться, что полученный результат будет полностью соответствовать спектру S1 новой дискретизации исходного сигнала. Обратит внимание, что сложение спектров рядом расположенных диапазонов может вызывать не только увеличение высокочастотных составляющих (как это можно было видеть на рис. 19.4 – спектр S1), ни и их взаимную компенсацию, как имеет место для спектра S1 в данном случае (кривая точками на графике В).

Перекрытие спектров диапазонов вызовет искажение аналоговой формы сигнала, восстановленного из его дискретных отсчетов, что можно видеть на графике Г – кривая s2. В данном случае, при частичной взаимной компенсации перекрывающихся частей спектров, наиболее сильное искажение произошло во второй, высокочастотной части сигнала.

Дискретизируемые сигналы, как правило, содержат широкополосные шумы, высокочастотные составляющие которых неизбежно перекрываются при периодизации спектра, и увеличивают погрешность восстановления сигналов. Для исключения этого фактора перед проведением дискретизации должно быть обеспечено подавление всех частот выше частоты Найквиста, т.е. выполнена низкочастотная фильтрация сигнала. Если последнее не проведено, то при дискретизации целесообразно в 2-4 раза уменьшить интервал дискретизации относительно оптимального и первой операцией обработки данных выполнить низкочастотную цифровую фильтрацию, после чего можно провести децимацию данных.

Увеличение интервала дискретизации сигналов является довольно распространенной операцией при цифровой обработке данных, и не только при подготовке данных для хранения с целью сокращения их количества. При комплексной обработке данных различной природы интервалы дискретизации этих данных могут оказаться различными, и производится их приведение к одному

значению. Аналогичная операция выполняется, как правило, и при создании многослойных информационных пакетов. В таких случаях снижение частоты дискретизации каких-либо данных является вынужденной необходимостью даже с потерей части высокочастотных составляющих информации. Предварительное отфильтровывание отбрасываемых данных перед децимацией (для исключения их попадания в главный частотный диапазон и искажения основной информации) в этом случае является обязательным, особенно при достаточно высокой энергии этих составляющих сигнала. Пример такой децимации приведен на рис. 19.10 на графиках В и Г - спектр  $S_2(f)$  децимированных данных и аналоговый сигнал  $s_3(t)$ , восстановленный по дискретным отсчетам  $sd(k\Delta t) \leftrightarrow S_2(f)$ . Децимация выполнена непосредственно в частотной области путем смыкания на частотной части  $0-0.5f_N$  спектра  $SM(f)$  исходного сигнала  $sm(m\Delta m)$  с сопряженной частью на интервале  $1.5f_N- f_N$ , что сокращает новый интервал Найквиста в 2 раза и формирует спектр  $S_2(f)$ , соответствующий дискретному сигналу с увеличенным в два раза интервалом дискретизации данных с полностью подавленной частью спектральных составляющих от  $0.5f_N$  до  $1.5f_N$ . Такой метод может применяться для децимации (передискретизации) данных с любой кратностью.

**Дискретизация с усреднением.** Если дискретизация сигнала производится импульсами конечной ширины, то таким импульсам соответствуют средние значения сигнала на интервале длительности импульсов. При длительности импульсов  $\tau$  имеем:

$$s(k\Delta t) = (1/\tau) \int_{k\Delta t - \tau/2}^{k\Delta t + \tau/2} s(t) dt. \quad (19.11)$$

С использованием селектирующей и гребневой функций эта операция отображается следующим образом:

$$s_{\Delta t}(t) = (1/\tau)[s(t) * P_{\tau}(t)]Ш_{\Delta t}(t). \quad (19.12)$$

Соответственно спектр дискретной функции:

$$S_F(f) = [S(f) \cdot \text{sinc}(\pi f\tau)] * F \cdot Ш_F(f). \quad (19.13)$$

Отсюда следует, что при дискретизации с усреднением спектр  $S(f)$  заменяется спектром  $S(f) \cdot \text{sinc}(\pi f\tau)$ , периодическое продолжение которого и образует спектр дискретной функции. При обратном преобразовании Фурье и при использовании интерполяционной формулы Котельникова-Шеннона, вместо исходной функции  $s(t)$  получаем функцию  $s'(t) = s(t) * P_{\tau}(t)/\tau$ , что эквивалентно пропусканию сигнала через фильтр с откликом  $h(t) = P_{\tau}(t)/\tau$ , т.е. через низкочастотный сглаживающий фильтр "скользящего" среднего с окном  $\tau$ .

Допустим  $\tau = \lambda \Delta t$ ,  $\lambda \leq 1$ ,  $F = 2\alpha f_{\max}$ ,  $\alpha \geq 1$ . Для этих условий

частотная передаточная функция фильтра записывается в следующем виде:  $H(f) = \text{sinc}[(\pi\lambda/2\alpha)(f/f_{\max})]$ . Если потеря составляющих сигнала на всех частотах не должна превышать 3%, необходимо выполнить условие:  $\text{sinc}(\pi\lambda/2\alpha) \geq 0,97$ . При  $\alpha=1$  отсюда следует, что значение  $\lambda$  должно быть равно  $\lambda \leq 0,27$ , т.е. ширина импульса дискретизации может составлять до 27 % интервала дискретизации.

Отметим, что в выражении (19.11) значения отсчетов относятся к центру интервалов  $\tau$  импульсов дискретизации. Если отсчет будет относиться к концу интервалов  $\tau$ , что имеет место при обработке информации в режиме реального времени, то в выходной функции (19.12) появится сдвиг на интервал  $\tau/2$ , а в ее спектре соответственно сдвиг фаз на  $\omega\tau/2$  (в правой части выражения (19.13) добавится множитель  $\exp(-j\omega\tau/2)$ ).

**Дискретизация спектров.** Теоремы, доказанные для прямого преобразования Фурье, в такой же мере действительны и для обратного. При дискретизации спектра сигнала с шагом  $\Delta f$  динамическое представление сигнала также становится периодическим с периодом  $T=1/\Delta f$ . Для сохранения возможности точного восстановления сигнала в пределах главного периода (без наложения сигналов соседних периодов) частотный шаг дискретизации должен удовлетворять условию:

$$\Delta f \leq 1/T. \quad (19.14)$$

Попутно отметим, что для временной формы каузального сигнала главным периодом принимают интервал от 0 до  $T$ , хотя при обработке данных на ЭВМ это не имеет значения и главный период может устанавливаться от  $-T/2$  до  $T/2$ .

**Информационная тождественность динамической и частотной форм** дискретного представления сигнала непосредственно следует из теоремы Котельникова-Шеннона.

Основой любых преобразований при обработке данных обычно является финитный (конечный по длительности) сигнал, зарегистрированный на интервале  $0-T$  и состоящий из определенных частотных составляющих от 0 до  $f_{\max}$ . Оптимальная дискретизация аналогового сигнала без потери точности его восстановления, как рассмотрено выше, соответствует двум отсчетам на периоде максимальной частотной составляющей:

$$\Delta t = 1/2f_{\max}, \quad N_t = T/\Delta t. \quad (19.15)$$

где  $N_t$  – общее количество отсчетов на интервале  $T$  задания сигнала. Если сигнал зарегистрирован непосредственно в дискретной форме, то он автоматически ограничен по максимальной частоте, т.е. максимальные частоты в таком сигнале равны  $f_{\max} \leq 1/2\Delta t$ .

При переводе дискретного сигнала в частотную форму спектр сигнала непрерывен и периодичен с периодом  $1/\Delta t = 2f_N$ . Для оптимальной дискретизации по частоте без потери точности восстановления непрерывного спектра должны выполняться условия:

$$\Delta f = 1/T = 1/(\Delta t N_t), \quad f_N = 1/2\Delta t, \quad (19.16)$$

$$N_f = 2f_N/\Delta f = N_t. \quad (19.17)$$

Спектр сигнала подвергается каким-либо преобразованиям (обработке), как правило, только в главном частотном диапазоне и тем самым превращается в непериодический сигнал, существующий только в интервале  $2f_N$  (от  $-f_N$  до  $f_N$ ). Значения спектра за пределами главного диапазона по умолчанию полагаются равными нулю. При обратном переводе такого сигнала из частотной формы в динамическую сигнал также является непрерывным и периодическим с периодом  $1/\Delta f = T$ , при этом оптимальная дискретизация по координатам без потери точности восстановления непрерывной формы соответствует условиям:

$$\Delta t = 1/2f_N, \quad T = 1/\Delta f, \quad (19.18)$$

$$N_t = T/\Delta t = N_f. \quad (19.19)$$

При осуществлении преобразований  $s(k\Delta t) \Leftrightarrow S(n\Delta f)$ , равно как и  $S(n\Delta f) \Leftrightarrow s(k\Delta t)$ , условие  $N_f = N_t$  является необходимым и достаточным для полного сохранения информации при преобразованиях сигнала из одной формы представления в другую. Условия (19.15-19.19) задают оптимальность преобразований без потерь информации. Если исходный сигнал дискретизирован оптимально и представлен  $N$  отсчетами, то уменьшение количества отсчетов при преобразовании неизбежно приводит к определенным потерям информации.

Что касается увеличения числа отсчетов при преобразовании функций (уменьшение интервалов дискретизации), то оно всегда возможно, т.к. выходной сигнал преобразования финитных сигналов является непрерывной функцией и, в принципе, интервал дискретизации может быть установлен бесконечно малым. Однако увеличение числа отсчетов не увеличивает ни количества информации, заключенной в исходном сигнале, ни точности ее представления. По существу, такая операция полностью эквивалентна интерполяции исходного сигнала рядом Котельникова-Шеннона. Пример такой операции приведен на рис. 19.11.



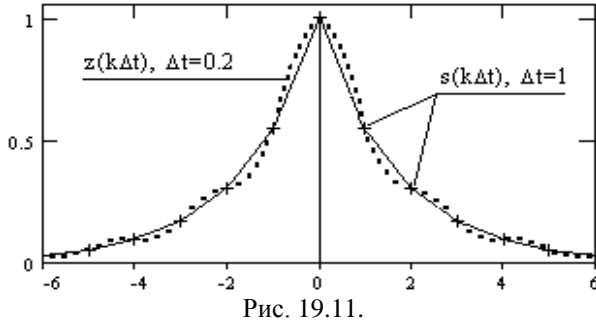


Рис. 19.11.

Отсчеты  $s(k\Delta t)$  и огибающая их кривая на рисунке 19.11 повторяют (в более детальном масштабе) сигнал  $s_1(t)$  на рис. 19.3, дискретизированный с шагом  $\Delta t=1$ . Как уже отмечалось, интервал дискретизации данного сигнала оказался завышенным, и спектр сигнала искажен (рис. 19.4). При выполнении операции  $s(k\Delta t) \Rightarrow S(n\Delta f)$  количество точек дискретизации спектра  $S(n\Delta f)$  было увеличено в 5 раз по отношению к количеству точек сигнала  $s(k\Delta t)$ , т.е.  $N_f = 5N_t$ . При обратном преобразовании  $S(n\Delta f) \Rightarrow z(k\Delta t)$ , были выполнены условия (19.18-19.19), при этом шаг дискретизации сигнала при его восстановлении оказался также в 5 раз меньше исходного ( $\Delta t = 0.2$ ). Результат можно видеть на рис. 19.11 (кривая  $z(k\Delta t)$ ). Абсолютно такой же результат дает и интерполяция сигнала  $s(k\Delta t)$  рядом Котельникова-Шеннона с переводом на шаг  $\Delta t = 0.2$ . Искажение аналогового сигнала закладывается при его дискретизации, если шаг дискретизации не удовлетворяет условию (19.8), и при любых дальнейших преобразованиях уже не может быть исправлено, т.к. информация о первоначальной форме аналогового сигнала при некорректной дискретизации утрачивается безвозвратно.

**Дискретизация усеченных сигналов.** При выполнении условия (19.8) для сигналов с ограниченным спектром аналоговая форма сигнала может быть восстановлена по дискретным отсчетам, если сигнал на интервале  $T$  его задания является финитным или, по крайней мере, настолько быстро затухающим, что отсчеты сигнала за пределами интервала  $T$  практически равны нулю. Задача дискретизации усложняется для медленно затухающих сигналов, сигналов бесконечной длительности и сигналов со спектром, неограниченным по частоте. Последнее имеет место, если в сигнале присутствуют разрывы и резкие скачки.

В общем случае, длительность сигнала и ширина его спектра не могут быть одновременно ограничены конечными интервалами. Если длительность сигнала ограничена и сигнал урезан в области не

нулевых значений, то спектр сигнала неограничен и наоборот. Однако обработка реальных сигналов возможна только с их ограничением, как по координатам, так и по ширине спектра. При этом в качестве оценки корректности ограничения сигналов используется энергетический критерий, согласно которому длительность сигнала  $T$  и практическую ширину спектра  $\Omega$  устанавливают такими, чтобы в них была сосредоточена подавляющая часть энергии сигнала. Это достигается при выполнении условий:

$$\int_0^T |s(t)|^2 dt = k \int_{-\infty}^{\infty} |s(t)|^2 dt, \quad (19.20)$$

$$\int_0^{\Omega} |S(\omega)|^2 d\omega = k \int_0^{\infty} |S(\omega)|^2 d\omega, \quad (19.21)$$

где  $k$ - коэффициент представительности (качества) задания сигнала, значение которого, в зависимости от целевых задач обработки сигналов, может устанавливаться от 0,9 до 0,99.

Допустим, что произвольный сигнал  $s(t)$  рассматривается в пределах конечного интервала  $[0, T]$  и принимается равным нулю за его пределами. Такой сигнал может быть получен умножением сигнала  $s(t)$  на прямоугольную весовую функцию  $\Pi_T(t)$ :

$$s_T(t) = s(t) \Pi_T(t).$$

Для спектра  $S_T(f)$  функции  $s_T(t)$  соответственно имеем:

$$S_T(f) = S(f) * T \cdot \text{sinc}(\pi f T). \quad (19.22)$$

Спектр  $S_T(f)$  неограничен, поскольку неограничен носитель функции  $\text{sinc}(\pi f T)$ . Отсюда следует, что частота дискретизации функции  $s_T(t)$  в принципе должна быть бесконечно большой, т.е. корректная дискретизация невозможна. На практике полагают, что спектр  $S_T(f)$  также определен в конечной области  $[-\Omega, \Omega]$ :

$$S'_T(f) = S_T(f) \cdot \Pi_{2\Omega}(f),$$

при этом вне этой области, по оценке Шеннона, для спектра  $S_T(f)$  справедлива формула:

$$|S_T(f)| \approx 1/\Omega T, \quad f \notin (-\Omega, \Omega). \quad (19.23)$$

Но усеченная часть спектра определяет разность значений между исходной функцией  $s_T(t)$  и функцией  $s'_T(t)$ , восстановленной по усеченному спектру  $S'_T(f)$ , т.к. отсеченных гармоник спектра будет недоставать для полного восстановления функции  $s_T(t)$ :

$$\varepsilon_T(t) = s_T(t) - s'_T(t).$$

Соответственно, оценка дисперсии погрешности аппроксимации определяется выражением:

$$\sigma^2 = \int_{-\infty}^{\infty} \varepsilon_T^2(t) dt \approx 1/\Omega T, \quad \sigma \approx 1/\sqrt{\Omega T}. \quad (19.24)$$

Эти выражения определяют порядок среднеквадратической погрешности аппроксимации, которая является интегральной по интервалу  $T$ , а не локальной разностью значений  $s_T(t) - s'_T(t)$ . Типичный вид погрешности аппроксимации усеченных сигналов приведен на рис. 19.12.

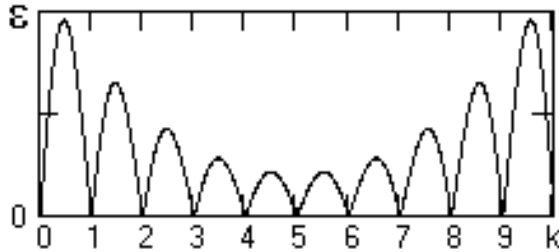


Рис. 19.12. Вид функции погрешности аппроксимации

В точках дискретизации погрешность равна нулю, максимальна на центрах интервалов дискретизации и нарастает при приближении к границам интервала  $T$ .

Физические данные обычно регистрируются по определенным интервалам  $T$  и, как правило, не выходят на нулевые значения на границах интервалов. В этом случае ограничение ширины спектра можно проводить по (19.24) с учетом допустимой среднеквадратической погрешности аппроксимации данных. Частота  $\Omega$  при усечении спектра может рассматриваться в качестве частоты Найквиста для сигнала  $s_T(t)$  при его дискретизации, что определяет частоту дискретизации не менее  $F=2\Omega$  и количество точек дискретизации не менее  $N=TF=2\Omega T$ .

В силу тождественности свойств прямого и обратного преобразования Фурье аналогичная методика может применяться и для оценки условий дискретизации спектров.

Таким образом, дискретизация усеченных сигналов возможна, однако при обработке усеченных сигналов необходимо проявлять осторожность и контролировать как значение среднеквадратической ошибки искажений, так и характер возникающих искажений сигнала и его спектра. Так, например, при усечении функции автокорреляции в спектре мощности сигнала могут появиться отрицательные значения, т.к. функция отсчетов  $\text{sinc}(\pi fT)$  в (19.22) является знакопеременной. Другой пример - проектирование частотных полосовых фильтров. При задании передаточной функции фильтра  $H(f)$  в частотной области в виде  $\Pi$ -образной функции  $H(f) = \Pi_r(f)$  обратное преобразование Фурье дает импульсный отклик фильтра  $h(t) \leftrightarrow H(f)$  бесконечно большой

длины. Усечение отклика  $h_T(t)=h(t)\Pi_T(t)$  вызывает изменение передаточной функции фильтра (явление Гиббса):

$$H_T(f) = \Pi_r(f) * \Pi_T(f) \Rightarrow \Pi_r(f) \cdot T \cdot \text{sinc}(\pi f T),$$

при этом по краям скачков  $\Pi$ -функции появляются затухающие флюктуации с амплитудой первого выброса до 9% от значений коэффициента передачи фильтра в полосе пропускания.

Так как частотный характер искажений, возникающих при усечении сигнала, определяется весовой функцией  $\Pi_T(t) \Leftrightarrow T \cdot \text{sinc}(\pi f T)$ , то допустимый уровень и форму искажения сигнала можно устанавливать не только подбором интервала  $T$ , но и применением других весовых функций. Так, для исключения появления отрицательных значений в спектрах мощности усечение функций автокорреляции целесообразно выполнять весовыми функциями, которые не имеют отрицательных значений в своих спектрах. Одной из таких функций является, например, треугольная весовая функция (окно Бартлетта).

**Соотношение спектров одиночного и периодического сигналов.**

Спектр  $S_T(f) = S(k\Delta f)$  периодического сигнала  $s_T(t)$  с периодом  $T$  дискретен ( $\Delta f = 1/T$ ). Спектр  $S(f)$  одиночного сигнала  $s(t)$ , заданного на интервале  $T$ , непрерывен и представляет собой спектральную плотность сигнала при  $T \Rightarrow \infty$ . Но периодический сигнал можно представить и в виде свертки одного периода с гребневой функцией Дирака:

$$s_T(t) = s(t) * \text{Ш}_T(t).$$

При переходе в частотную область получаем:

$$S_T(f) = (1/T) \cdot S(f) \cdot \text{Ш}_{1/T}(f) = S(k\Delta f),$$

$$S_T(f) = (1/T) \sum_{k=-\infty}^{\infty} S(f) \delta(f - k/T). \quad (19.25)$$

Отсюда следует, что спектр периодического сигнала представляет собой дискретизированный спектр одиночного сигнала, нормированный на длительность периода. С другой стороны, одиночный сигнал  $s(t)$  может быть получен из периодического сигнала  $s_T(t)$  умножением на селектирующий прямоугольный импульс  $\Pi_T(t)$ :

$$s(t) = s_T(t) \cdot \Pi_T(t).$$

Спектр одиночного сигнала:

$$S(f) = T \cdot S_T(f) * \Pi_T(f) = T \sum_{k=-\infty}^{\infty} S(k\Delta f) \cdot \text{sinc}[\pi T(f - k/T)], \quad (19.26)$$

т.е. непрерывный спектр одиночного сигнала однозначно устанавливается по спектру периодического сигнала (интерполяция рядом Котельникова-Шеннона в частотной области).

#### **19.4. Дискретизация по критерию наибольшего отклонения**

Задача абсолютно точного восстановления сигнала на практике обычно не ставится, в отличие от задачи минимального физического объема информации, при котором сохраняется возможность ее восстановления в непрерывной форме с определенным допустимым значением погрешности. Такая задача актуальна всегда, и особенно при дистанционных методах регистрации и обработки информации, передаче сигналов по каналам связи и при подготовке информации к длительному хранению. Одним из методов решения этой задачи является дискретизация сигналов по критерию наибольшего отклонения.

В процессе дискретизации по критерию наибольшего отклонения задается допустимое значение погрешности восстановления сигнала  $\sigma$ . При восстановлении сигнала непрерывная функция  $s(t)$  аппроксимируется, как правило, степенными полиномами  $n$ -го порядка. Погрешность восстановления функции  $s(t)$  полиномом  $s_a(t)$  определяется остаточным членом  $L(t)$ :

$$L(t) = s(t) - s_a(t) = \sigma(t).$$

Шаг дискретизации выбирается из условия обеспечения  $L(t) < \sigma$  по всему интервалу определения функции  $s(t)$ . Как правило, динамика функции  $s(t)$  может существенно изменяться в различные моменты времени по интервалу регистрации, при этом шаг дискретизации также может изменяться, при условии не превышения заданной погрешности на каждом шаге. При установленном значении  $\sigma$  уменьшение числа отсчетов обеспечивается повышением степени аппроксимирующего многочлена. На практике обычно ограничиваются ступенчатой, линейной и параболической аппроксимацией полиномами соответственно нулевой, первой и второй степеней.

В качестве интерполирующих многочленов используют многочлены Лагранжа. Для многочленов Лагранжа нулевой степени значение  $s_a(t)$  в момент времени  $t$  на интервале  $t_i < t < t_{i+1}$  между двумя последовательными отсчетами функции принимается равным отсчету  $s(t_{i+1})$ . Если восстановление сигнала  $s(t)$  проводить по двум отсчетам:  $s_a(t_i) = [s(t_{i+1}) - s(t_i)]/2$ , то при том же шаге дискретизации погрешность восстановления сигнала уменьшается вдвое. Но при использовании двух последовательных отсчетов лучше использовать многочлены Лагранжа первой степени, т.е. соединение двух последовательных отсчетов прямой линией, что дает еще большее уменьшение погрешности восстановления аналоговой формы сигнала.

В качестве экстраполирующих многочленов используется многочлены Тейлора. Для многочлена Тейлора нулевой степени условия восстановления сигнала практически не отличаются от многочлена Лагранжа, за исключением направления (от текущего зарегистрированного отсчета и вперед по  $t$ ). Для многочленов Тейлора более высоких степеней при восстановлении сигнала помимо отсчета  $s(t_i)$  используется также соответствующие значения производных в точке отсчета. Восстановление сигнала многочленами Тейлора происходит без задержки во времени. Однако при использовании многочленов выше нулевой степени для точного восстановления сигнала по сравнению с интерполяционными методами требуется в два раза более высокая частота дискретизации.

### **19.5. Адаптивная дискретизация**

Частота равномерной дискретизации информации рассчитывается по предельным значениям частотных характеристик сигналов. Адаптивная дискретизация ориентирована на динамические характеристики сигнала, что позволяет обеспечивать его восстановление при минимальном числе выборок. В основе принципов адаптивной дискретизации лежит слежение за текущей погрешностью восстановления сигнала. Наиболее широкое применение получили алгоритмы дискретизации с адаптацией по длине интервала аппроксимации. Сущность дискретизации заключается в последовательном наращивании интервала аппроксимации с непрерывным сравнением сигнала  $s(t)$  с воспроизводящей функцией  $s_a(t)$ . При достижении заданного значения  $\sigma$  наращивание интервала прекращается, и производится отсчет значения  $s(t_i)$ , т.е. дискретизация является неравномерной. Для воспроизведения сигналов нерегулярной дискретизации обычно используются степенные алгебраические полиномы нулевой и первой степени в интерполяционном или в экстраполяционном вариантах.

Наиболее простой является техника адаптивной дискретизации с использованием многочлена нулевой степени. На момент  $t_i$  начала каждого интервала аппроксимирующий полином  $s_a(t)$  принимается равным  $s(t_i)$ , вычисляется текущая разность  $L(t) = s(t) - s_a(t)$  и производится сравнение ее значения с заданным значением  $\sigma$ . При фиксировании равенства  $L(t) = \sigma$  производится очередной отсчет и начинается следующий интервал.

При использовании аппроксимирующего многочлена первой степени вычисляется значение  $s_a(t) = s(t_i) + s'(t_i)$ , где  $s'(t)$  - производная сигнала. Момент очередного отсчета определяется выполнением равенства  $s(t) - s(t_i) - s'(t_i) = \sigma$ . Следует иметь в виду, что данный алгоритм

неэффективен при наличии высокочастотных помех, к которым весьма чувствительна операция дифференцирования.

Самыми простыми способами восстановления сигналов при адаптивной дискретизации являются линейная и квадратичная интерполяции, которые выполняются по уравнениям:

$$f(x)_{\text{лин}} = a_0 + a_1x. \quad f(x)_{\text{кв}} = a_0 + a_1x + a_2x^2.$$

Эти уравнения являются частным случаем полиномиальной интерполяции с помощью аппроксимирующего полинома:

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n = \sum_{i=0}^n a_i \cdot x^i. \quad (19.27)$$

Для выполнения полиномиальной интерполяции достаточно по выражению (19.27) составить систему линейных уравнений для  $n$  последовательных отсчетов и определить  $n$  значений коэффициентов  $a_i$ . При *глобальной* интерполяции, по всем  $N$  точкам задания функции, степень полинома равна  $N-1$ . Глобальная интерполяция обычно выполняется для достаточно коротких (не более 8-10 отсчетов) массивов данных. Пример выполнения глобальной интерполяции приведен на рис. 19.13.

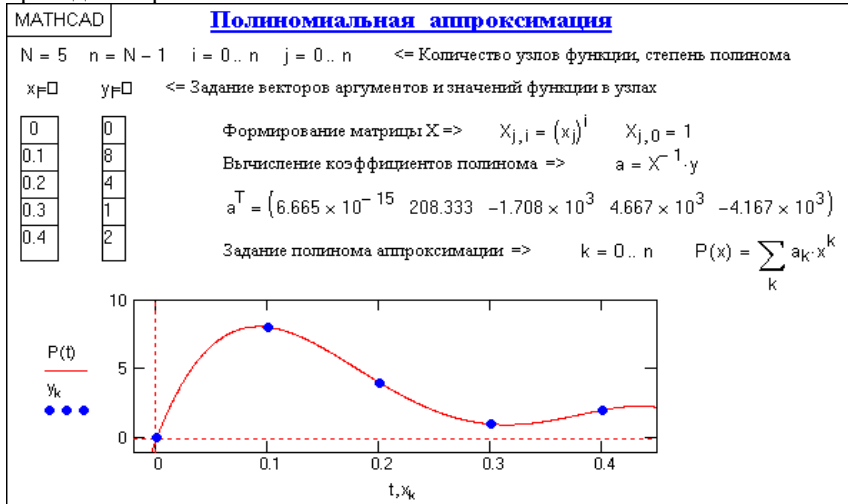


Рис. 19.13. Интерполяция данных.

Большие массивы данных интерполируются последовательными *локальными* частями или в скользящем по массиву данных окне интерполяции, как правило, с нечетным значением  $N$  и вычислением требуемых значений сигнала в определенном интервале центральной части окна.

Для практического использования более удобны формулы аппроксимации, не требующие предварительного определения коэффициентов аппроксимирующих полиномов. К числу таких формул относится интерполяционный многочлен по Лагранжу. При аппроксимации функции  $y(x)$  многочленом  $n$ -ой степени  $f(x)$ :

$$\begin{aligned}
 f(x) &= \\
 &= \frac{(x - x_1)(x - x_2)\dots(x - x_n)}{(x_0 - x_1)(x_0 - x_2)\dots(x_0 - x_n)} y_0 + \frac{(x - x_0)(x - x_2)\dots(x - x_n)}{(x_1 - x_0)(x_1 - x_2)\dots(x_1 - x_n)} y_1 + \dots \\
 &\dots + \frac{(x - x_0)(x - x_1)\dots(x - x_{n-1})}{(x_n - x_0)(x_n - x_1)\dots(x_n - x_{n-1})} y_n.
 \end{aligned} \tag{19.28}$$

Пример интерполяции по Лагранжу приведен на рис. 19.14.

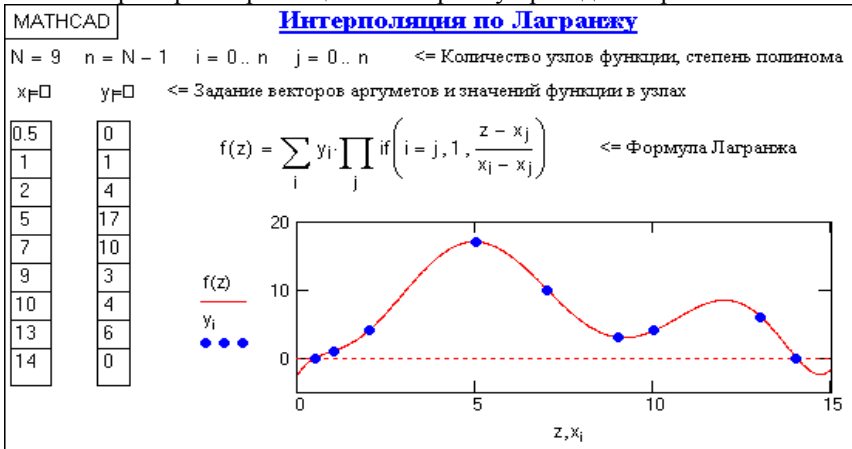


Рис. 19.14. Интерполяция по Лагранжу.

## 19.6. Квантование сигналов

Дискретизация аналоговых сигналов с преобразованием в цифровую форму связана с квантованием сигналов. Сущность квантования состоит в замене несчетного множества возможных значений функции, в общем случае случайных, конечным множеством цифровых отсчетов, и выполняется округлением мгновенных значений входной функции  $s(t_i)$  в моменты времени  $t_i$  до ближайших значений  $s_i(t_i) = n_i \Delta\sigma$ , где  $\Delta\sigma$  - шаг квантования шкалы цифровых отсчетов. Квантование с постоянным шагом  $\Delta\sigma$  называется *равномерным*. Математически операция квантования может быть выражена формулой:



$$s_i(t_i) = \left[ \frac{s(t_i)}{\Delta\sigma} + \frac{1}{2} \right] \cdot \Delta\sigma,$$

где скобки [...] означают целую часть значения в скобках.

При квантовании сигналов в большом динамическом диапазоне значений шаг квантования может быть и неравномерным, например, логарифмическим, т.е. пропорциональным логарифму значений входного сигнала. Установленный диапазон шкалы квантования от  $s_{\min}$  до  $s_{\max}$  и шаг квантования  $\Delta\sigma$  определяют число делений шкалы  $N_\sigma = (s_{\max} - s_{\min}) / \Delta\sigma$  и соответственно цифровую разрядность квантования. В результате дискретизации и квантования непрерывная функция  $s(t)$  заменяется числовой последовательностью  $\{s(k\Delta t)\}$ . Погрешность округления  $\varepsilon_i = s(t_i) - s_i(k\Delta t)$  заключена в пределах  $-\Delta\sigma/2 < \varepsilon < \Delta\sigma/2$  и называется **шумом квантования**. Требуемая точность квантования оценивается по влиянию возникающего шума квантования на последующую обработку сигналов.

При достаточно малом шаге квантования любое значение в его пределах можно считать равновероятным, при этом значения  $\varepsilon$  распределены по равномерному закону:

$$p(\varepsilon) = 1/\Delta\sigma, \quad -\Delta\sigma/2 \leq \varepsilon \leq \Delta\sigma/2.$$

Соответственно, дисперсия и среднее квадратическое значение шума квантования:

$$\varepsilon^2 = \Delta\sigma^2/12, \quad \bar{\varepsilon} \approx 0.3 \Delta\sigma. \quad (19.15)$$

При задании уровня шума квантования с использованием выражения (19.15) нетрудно определить допустимое значение шага квантования.

Входной сигнал содержит, как правило, аддитивную смесь собственно сигнала  $s(t)$  и помехи  $q(t)$  с дисперсией соответственно  $\sigma_q^2$ . Если помехи не коррелированы с сигналом, то после квантования суммарная дисперсия шумов:

$$\sigma^2 = \sigma_q^2 + \varepsilon^2.$$

На практике шаг квантования выбирают обычно таким, чтобы не происходило заметного изменения отношения сигнал/шум, т.е.  $\varepsilon^2 \ll \sigma_q^2$ .

## **19.7. Децимация и интерполяция данных**

**Децимацией** (прореживанием, сокращением) цифровых данных принято называть уплотнение данных с удалением избыточной информации. Последнее имеет место, если шаг дискретизации данных был установлен излишне подробным и  $f_N = 1/2\Delta t \gg f_{\max}$  сигнала. Информация высокочастотной части сигнала может быть ненужной, если основная энергия полезной части сигнала

заклучена в низкочастотной области. Децимация может потребоваться и в том случае, если массивы данных представлены с разным шагом дискретизации.

Децимации должна предшествовать низкочастотная фильтрация данных. Это связано с тем, что в процессе децимации шаг дискретизации  $\Delta t$  заменяется на новый шаг  $\Delta t' = p\Delta t$ , где  $p > 1$ , с соответствующим сжатием главного частотного диапазона, при этом появляется опасность отражения отбрасываемых частотных составляющих и высокочастотных шумов в главный диапазон (как и при неправильном выборе шага дискретизации). Точка отсечки низкочастотного фильтра устанавливается по новой частоте Найквиста:  $f_N' = 1/(2p\Delta t)$ .

Значение коэффициента  $p$  при децимации может быть произвольным, но, как правило, используются целочисленные значения, и децимация выливается в простое прореживание данных. При нецелочисленном значении  $p$  децимация может проводиться с использованием интерполяционного ряда Котельникова-Шеннона (равно как и любого другого интерполяционного многочлена) или преобразования Фурье. Последнее выполняется путем перевода сигнала в частотную форму и возвращением в координатную форму с новым шагом  $\Delta t' = p\Delta t$ , при этом низкочастотная фильтрация может производиться непосредственно в частотном диапазоне. Возможно также и прямое усечение главного частотного диапазона с  $N$  точек до  $N' = N/p$  с возвратом из нового частотного диапазона в координатную форму с количеством точек  $N'$ , но при этом следует учитывать последствия усечения спектральной функции (умножения на прямоугольное селектирующее окно) на форму восстанавливаемого по ней сигнала (свертка исходного сигнала с фурье-образом прямоугольного селектирующего окна).

Интерполяция данных отличается от децимации только значением коэффициента  $p < 1$ , с соответствующим увеличением частоты Найквиста, и не требует низкочастотной фильтрации.

Для децимации и интерполяции данных разработаны также специальные высокоскоростные методы и алгоритмы (цифровые фильтры) - экспандеры и компрессоры.

## **19.8. Введение в дискретные преобразования сигналов**

Цифровая обработка сигналов оперирует с дискретными преобразованиями сигналов и обрабатывающих данные сигналы систем. Математика дискретных преобразований зародилась в недрах

аналоговой математики еще в 18 веке в рамках теории рядов и их применения для интерполяции и аппроксимации функций, однако ускоренное развитие она получила в 20 веке после появления первых вычислительных машин. В принципе, в своих основных положениях математический аппарат дискретных преобразований подобен преобразованиям аналоговых сигналов и систем. Однако дискретность данных требует учета этого фактора, и его игнорирование может приводить к существенным ошибкам. Кроме того, ряд методов дискретной математики не имеет аналогов в аналитической математике.

### **19.9. Преобразование Фурье**

**Дискретное преобразование Фурье** может быть получено непосредственно из интегрального преобразования дискретизаций аргументов ( $t_k = k\Delta t$ ,  $f_n = n\Delta f$ ):

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp(-j2\pi ft) dt, \quad S(f_n) = \Delta t \sum_{k=-\infty}^{\infty} s(t_k) \exp(-f_n k \Delta t), \quad (19.29)$$

$$s(t) = \int_{-\infty}^{\infty} S(f) \exp(j2\pi ft) df, \quad s(t_k) = \Delta f \sum_{n=-\infty}^{\infty} S(f_n) \exp(j2\pi n \Delta f t_k). \quad (19.30)$$

Напомним, что дискретизация функции по времени приводит к периодизации ее спектра, а дискретизация спектра по частоте - к периодизации функции. Не следует также забывать, что значения (19.29) числового ряда  $S(f_n)$  являются дискретизацией непрерывной функции  $S'(f)$  спектра дискретной функции  $s(t_k)$ , равно как и значения (19.30) числового ряда  $s(t_k)$  являются дискретизацией непрерывной функции  $s'(t)$ , и при восстановлении этих непрерывных функций  $S'(f)$  и  $s'(t)$  по их дискретным отсчетам соответствие  $S'(f) = S(f)$  и  $s'(t) = s(t)$  гарантировано только при выполнении теоремы Котельникова-Шеннона. Для дискретных преобразований  $s(k\Delta t) \Leftrightarrow S(n\Delta f)$ , и функция, и ее спектр дискретны и периодичны, а числовые массивы их представления соответствуют заданию на главных периодах  $T = N\Delta t$  (от 0 до  $T$  или от  $-T/2$  до  $T/2$ ), и  $2f_N = N\Delta f$  (от  $-f_N$  до  $f_N$ ), где  $N$  – количество отсчетов, при этом:

$$\Delta f = 1/T = 1/(N\Delta t), \quad \Delta t = 1/2f_N = 1/(N\Delta f), \quad \Delta t \Delta f = 1/N, \quad N = Tf_N. \quad (19.31)$$

Соотношения (19.31) являются условиями информационной равноценности динамической и частотной форм представления дискретных сигналов. Другими словами: число отсчетов функции и ее спектра должны быть одинаковыми. Но каждый отсчет комплексного спектра представляется двумя вещественными числами и, соответственно, число отсчетов комплексного спектра в 2 раза больше

отсчетов функции? Это так. Однако представление спектра в комплексной форме - не более чем удобное математическое представление спектральной функции, реальные отсчеты которой образуются сложением двух сопряженных комплексных отсчетов, а полная информация о спектре функции в комплексной форме заключена только в одной его половине - отсчетах действительной и мнимой части комплексных чисел в частотном интервале от 0 до  $f_N$ , т.к. информация второй половины диапазона от 0 до  $-f_N$  является сопряженной с первой половиной и никакой дополнительной информации не несет.

При дискретном представлении сигналов аргумент  $t_k$  обычно проставляется номерами отсчетов  $k$  (по умолчанию  $\Delta t=1, k=0,1,\dots,N-1$ ), а преобразования Фурье выполняются по аргументу  $n$  (номер шага по частоте) на главных периодах. При значениях  $N$ , кратных 2:

$$S(f_n) \equiv S_n = \sum_{k=0}^{N-1} s_k \exp(-j2\pi kn/N), \quad n = -N/2, \dots, 0, \dots, N/2. \quad (19.32)$$

$$s(t_k) \equiv s_k = (1/N) \sum_{n=-N/2}^{N/2-1} S_n \exp(j2\pi kn/N), \quad k = 0, 1, \dots, N-1. \quad (19.33)$$

Главный период спектра в (19.32) для циклических частот от  $-0.5$  до  $0.5$ , для угловых частот от  $-\pi$  до  $\pi$ . При нечетном значении  $N$  границы главного периода по частоте (значения  $\pm f_N$ ) находятся на половину шага по частоте за отсчетами  $\pm(N/2)$  и, соответственно, верхний предел суммирования в (19.33) устанавливается равным  $N/2$ .

В вычислительных операциях на ЭВМ для исключения отрицательных частотных аргументов (отрицательных значений номеров  $n$ ) и использования идентичных алгоритмов прямого и обратного преобразования Фурье главный период спектра обычно принимается в интервале от 0 до  $2f_N$  ( $0 \leq n \leq N$ ), а суммирование в (19.33) производится соответственно от 0 до  $N-1$ . При этом следует учитывать, что комплексно сопряженным отсчетам  $S_n^*$  интервала  $(-N, 0)$  двустороннего спектра в интервале  $0-2f_N$  соответствуют отсчеты  $S_{N+1-n}$  (т.е. сопряженными отсчетами в интервале  $0-2f_N$  являются отсчеты  $S_n$  и  $S_{N+1-n}$ ).

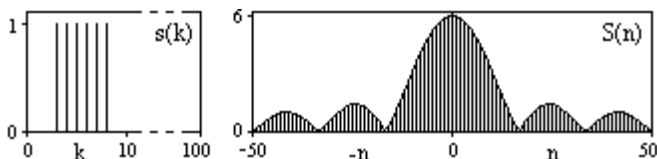


Рис. 19.16. Дискретный сигнал и модуль его спектра.

**Пример:** На интервале  $T = [0, 99]$ ,  $N=100$ , задан дискретный сигнал  $s(k) = \sum_{i=3}^8 \delta(k-i)$  - прямоугольный импульс с единичными значениями на точках  $k$  от 3 до 8. Форма сигнала и модуль его спектра в главном частотном диапазоне (вычисление по формуле  $S(n) = \sum_k s(k) \exp(-j2\pi kn/100)$  с шагом по частоте  $\Delta\omega=2\pi/100$ , приведены на рис. 19.16.

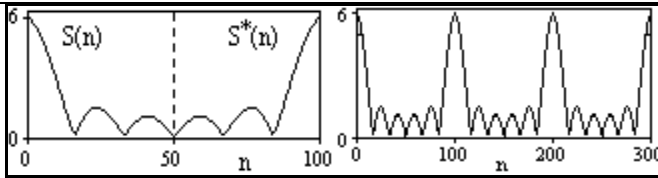


Рис. 19.17. Модуль спектра.      Рис. 19.18. Модуль спектра.

На рис. 19.7 приведена огибающая значений другой формы представления главного диапазона спектра. Независимо от формы представления спектр периодичен, в чем нетрудно убедиться, если вычислить значения спектра для большего интервала аргумента  $n$  с сохранением того же шага по частоте, как это показано на рис. 19.18 для огибающей значений спектра.

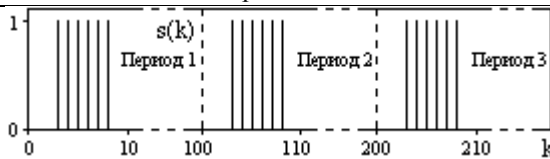


Рис. 19.19. Обратное преобразование Фурье.

На рис. 19.19. показано обратное преобразование Фурье для дискретного спектра, выполненное по формуле

$$s'(k) = (1/100) \sum_n S(n) \cdot \exp(j2\pi kn/100),$$

которое показывает периодизацию исходной функции  $s(k)$ , но главный период  $k=\{0,99\}$  этой функции полностью совпадает с исходным сигналом  $s(k)$ .

Преобразования (19.32-19.33) называют дискретными преобразованиями Фурье (ДПФ). Для ДПФ, в принципе, справедливы все свойства интегральных преобразований Фурье, однако при этом следует учитывать периодичность дискретных функций и спектров. Произведению спектров двух дискретных функций (при выполнении каких-либо операций при обработке сигналов в частотном представлении, как, например, фильтрации сигналов непосредственно в частотной форме) будет соответствовать свертка

периодизированных функций во временном представлении (и наоборот). Такая свертка называется циклической и ее результаты на концевых участках информационных интервалов могут существенно отличаться от свертки финитных дискретных функций (линейной свертки).

Из выражений ДПФ можно видеть, что для вычисления каждой гармоники нужно  $N$  операций комплексного умножения и сложения и соответственно  $N^2$  операций на полное выполнение ДПФ. При больших объемах массивов данных это может приводить к существенным временным затратам. Ускорение вычислений достигается при использовании быстрого преобразования Фурье.

**Быстрое преобразование Фурье** (БПФ, fast Fourier transform - FFT). Он базируется на том, что при вычислениях среди множителей (синусов и косинусов) есть много периодически повторяющихся значений (в силу периодичности функций). Алгоритм БПФ группирует слагаемые с одинаковыми множителями в пирамидальный алгоритм, значительно сокращая число умножений за счет исключения повторных вычислений. В результате быстрое действие БПФ в зависимости от  $N$  может в сотни раз превосходить быстрое действие стандартного алгоритма. При этом следует подчеркнуть, что алгоритм БПФ даже точнее стандартного, т.к. сокращая число операций, он приводит к меньшим ошибкам округления.

Допустим, что массив чисел  $s_k$  содержит  $N = 2^r$  отсчетов ( $r$  - целое). Разделим исходный массив на два первых промежуточных массива с четными и нечетными отсчетами:

$$s_k' = s_{2k}, \quad s_k'' = s_{2k+1}, \quad 0 \leq k \leq N/2-1.$$

Выполним ДПФ каждого массива с учетом того, что шаг функций равен 2 (при  $\Delta t=1$ ), а период промежуточных спектров будет соответственно равен  $N/2$ :

$$s_k' \Rightarrow S_n', \quad s_k'' \Rightarrow S_n'', \quad 0 \leq n \leq N/2-1.$$

Для получения одной половины искомого спектра  $S_n$  сложим полученные спектры с учетом теоремы запаздывания, т.к. отсчеты функции  $s_k''$  сдвинуты относительно  $s_k'$  на один шаг дискретизации:

$$S_n = S_n' + S_n'' \cdot \exp(-j2\pi n/N). \quad (19.34)$$

Вторая половина спектра, комплексно сопряженная с первой, с учетом периода повторения  $N/2$  промежуточных спектров определяется выражением:

$$S_{n+N/2} = S_n' + S_n'' \cdot \exp(-j2\pi(n+N/2)/N) = S_n' - S_n'' \cdot \exp(-j2\pi n/N). \quad (19.35)$$

Нетрудно видеть, что для вычисления полного спектра в данном случае потребуется  $N^2/4$  операций для вычисления промежуточных спектров плюс еще  $N$  операций комплексного

умножения и сложения, что создает ощутимый эффект по сравнению с ДПФ.

Но деление массивов на две части может быть применено и к первым промежуточным массивам, и ко вторым, и т.д. до тех пор, пока в массивах не останется по одному отсчету, фурье - преобразование которых равно самому отсчету. Тем самым, алгоритм преобразования превращается в пирамидальный алгоритм перестановок со сложением/вычитанием и с единичным умножением на значение  $\exp(-j2\pi n/N)$  соответствующего уровня пирамиды. Первый алгоритм БПФ на данном принципе (из множества модификаций, существующих в настоящее время) был разработан Кули-Тьюки в 1965 г. и позволил повысить скорость вычислений в  $N/r$  раз по сравнению с ДПФ. Чем больше  $N$ , тем больше эффект БПФ. Так, при  $N = 1024$  имеем  $r = 10$  и соответственно  $N/r \approx 100$ . Что касается условия по количеству точек  $N = 2^r$ , то оно рассматривается в варианте  $N_k \leq 2^r$ , где  $r$  - минимальное целое. Массив с  $N_k < 2^r$  дополняется до  $2^r$  нулями, что не изменяет форму спектра. Изменяется только шаг  $\Delta\omega$  по представлению спектра ( $\Delta\omega = 2\pi/2^r < 2\pi/N$ ), который несколько избыточен по адекватному представлению сигнала в частотной области. В настоящее время существуют и алгоритмы БПФ с другими основаниями и их комбинациями, при которых не требуется дополнения сигналов нулями до  $2^r$ .

Заметим, что в соответствии с (19.35) отсчеты, сопряженные с правой половиной главного частотного диапазона  $(0, \pi)$ , относятся не к диапазону  $(-\pi, 0)$ , а к диапазону  $(\pi, 2\pi)$ , что, учитывая периодичность спектра дискретных данных, значения не имеет. Т.е. выходной частотный диапазон БПФ равен  $(0, 2\pi)$ . Общее количество отсчетов комплексного спектра в этом условно главном диапазоне равно количеству точек исходного сигнала (с учетом нулевых точек при дополнении сигнала до  $N=2^r$ ). Алгоритм быстрого обратного преобразования (ОБПФ) тождественен алгоритму прямого БПФ.

***Алгоритмы прямого и обратного БПФ широко используются в современном программном обеспечении для анализа и обработки цифровых данных.*** Пример выполнения БПФ приведен на рис. 19.20.

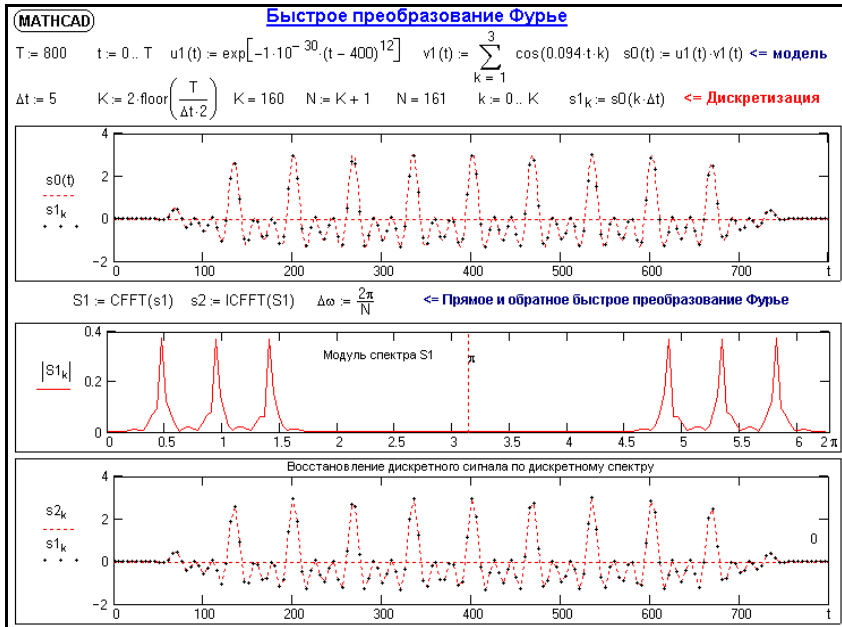


Рис. 19.20. Пример БПФ.

**Применение ДПФ.** Основная область использования ДПФ – спектральный анализ физических данных. При этом интерес обычно представляют только амплитуды отдельных гармоник, а не их фазы, и спектр отображается в виде графика зависимости амплитуды (модуля спектра) от частоты. Часто шкала амплитуд градуируется в децибелах. Децибелы - логарифмы отношения амплитудных значений. Например, разница на 20 дБ означает различие амплитуд в 10 раз, разница на 40 дБ - 100 раз. Различию амплитуд в 2 раза отвечает разница примерно в 6 дБ. Шкала частот также часто градуируется в логарифмическом масштабе.

Перед вычислением спектра из сигнала, как правило, вырезается отрезок сигнала. Число последовательных отсчетов отрезка для использования БПФ должно быть степенью двойки, если в программном обеспечении вычислительной системы не оговорена ее способность выполнять БПФ по произвольным числовым рядам. В противном случае числовой ряд дополняется нулями до необходимого размера, что не изменяет формы спектра и сказывается только на увеличении частотного разрешения по спектру.

При вычислении спектра возможен следующий нежелательный эффект. При разложении участка сигнала в ряд Фурье



мы тем самым принимаем этот участок за один период  $T$ , который периодически повторяется за пределами участка с фундаментальной частотой  $1/T$ . При ДПФ, а равно и при БПФ, вычисляется спектр именно такого периодического сигнала. При этом на границах периодов такая функция наверняка будет иметь разрывы или скачки, тем самым существенно искажая спектр. Для устранения этого эффекта применяются так называемые *весовые окна*, похожие на гауссиан, размер которых равен размеру участка. Анализируемый участок умножается на весовое окно, что плавно сводит сигнал на нет вблизи краев анализируемого участка и в значительной степени устраняют рассмотренные искажения спектра. Методика применения весовых окон подробно рассматривается в курсе цифровой обработки сигналов.

### 19.10. Преобразование Лапласа

Дискретное преобразование Лапласа (ДПЛ), как и ДПФ, может быть получено из интегрального преобразования дискретизаций аргументов ( $t_k = k\Delta t$ ,  $\omega_n = n\Delta\omega$ ):

$$Y(p) = \int_0^{\infty} y(t) \exp(-pt) dt, \quad Y(p_n) = \Delta t \sum_{k=0}^{\infty} y(t_k) \exp(-p_n t_k), \quad (19.36)$$

где  $p = \sigma + j\omega$  - комплексная частота,  $\sigma \geq 0$ .

$$y(t) = (1/2\pi j) \int_{\sigma - j\omega}^{\sigma + j\omega} Y(p) \exp(pt) dp. \quad y(t_k) = \Delta t \sum_{n=-\infty}^{\infty} Y(p_n) \exp(p_n t_k). \quad (19.37)$$

Функцию  $Y(p)$  называют изображением Лапласа функции  $y(t)$  - оригинала изображения. При  $\sigma=0$  преобразование Лапласа превращается в одностороннее преобразование Фурье, а для каузальных сигналов - в полную аналогию ПФ. Преобразование Лапласа применяется для спектрального анализа функций, не имеющих фурье-образов из-за расходимости интегралов Фурье:

$$\begin{aligned} Y(p) &= \int_0^{\infty} y(t) \exp(-\sigma t - j\omega t) dt = \int_0^{\infty} y(t) \exp(-\sigma t) \exp(-j\omega t) dt = \\ &= \int_0^{\infty} y'(t) \exp(-j\omega t) dt. \end{aligned}$$

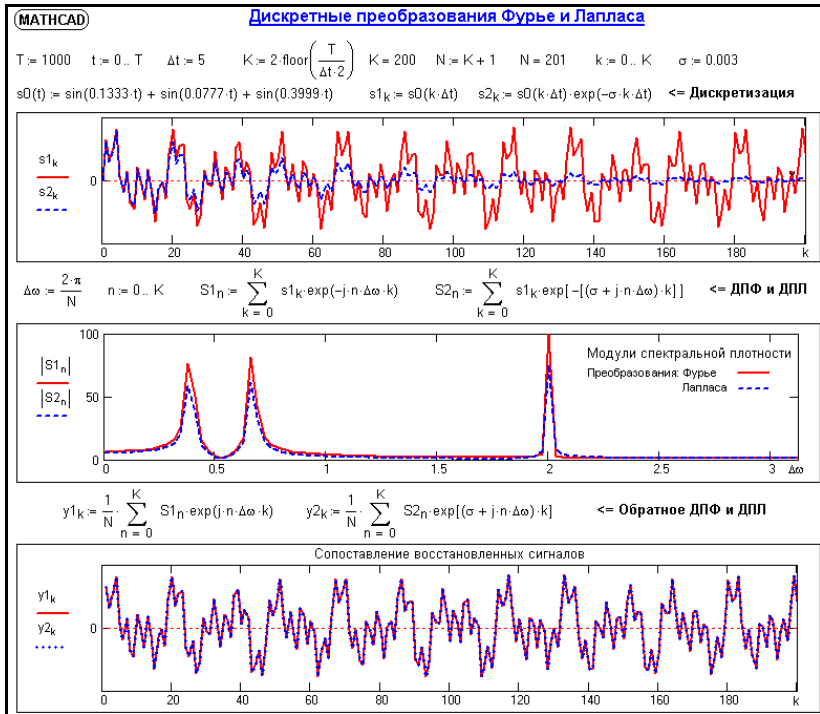


Рис. 19.21. Сопоставление преобразований Фурье и Лапласа.

Правый интеграл для каузальных сигналов представляет собой преобразование Фурье, при этом сигнал  $y'(t)$  за счет экспоненциального множителя  $\exp(-\sigma t)$  соответствующим выбором значения  $\sigma > 0$  превращается в затухающий и конечный по энергии. Все свойства и теоремы преобразований Фурье имеют соответствующие аналоги и для преобразований Лапласа.

Пример сопоставления преобразований Фурье и Лапласа приведен на рис. 19.21.

## 19.11. Z-преобразование сигналов

**Определение преобразования.** Распространенным способом анализа дискретных цифровых последовательностей является z-преобразование (z-transform).

Произвольной непрерывной функции  $s(t)$ , равномерно дискретизированной и отображенной отсчетами  $s_k = s(k\Delta t)$ , равно как и непосредственно дискретной функции, можно поставить в соответствие степенной полином по  $z$ , последовательными

коэффициентами которого являются значения  $s_k$ :

$$s_k = s(k\Delta t) \Leftrightarrow \text{TZ}[s(k\Delta t)] = \sum_{k=-\infty}^{\infty} s_k z^k = S(z). \quad (19.38)$$

где  $z = \sigma + j\omega = r \cdot \exp(-j\varphi)$  - произвольная комплексная переменная. Полином  $S(z)$  называют  $z$ -образом или  $z$ -изображением функции  $s(k\Delta t)$ . Преобразование имеет смысл для области тех значений  $z$ , в которой ряд  $S(z)$  сходится, т.е. сумма ряда представляет собой аналитическую функцию переменной  $z$ , не имеющую полюсов и особых точек.

---

**Пример:**  $s_k = \{1, 2, 0, -1, -2, -1, 0, 0\}$ .

$$S(z) = 1z^0 + 2z^1 + 0z^2 - 1z^3 - 2z^4 - 1z^5 + 0z^6 + 0z^7 = 1 + 2z - z^3 - 2z^4 - z^5.$$

---

Впервые  $z$ -преобразование введено в употребление П.Лапласом в 1779 и повторно "открыто" В.Гуревичем в 1947 году с изменением символики на  $z^{-1}$ . В настоящее время в технической литературе имеют место оба вида символики. На практическое использование преобразования это не влияет, т.к. смена знака только зеркально изменяет нумерацию членов полинома (относительно  $z^0$ ), числовое пространство которых в общем случае от  $-\infty$  до  $+\infty$ . В дальнейшем в качестве основной будем использовать символику положительных степеней  $z$ , давая пояснения по особенностям отрицательной символики, если таковая имеется.

По заданному или полученному в результате анализа какой-либо системы  $z$ -полиному однозначно восстанавливается соответствующая этому полиному функции путем идентификации коэффициентов степеней при  $z^k$  с  $k$ -отсчетами функции.

---

**Пример:**  $S(z) = 1 + 3z^2 + 8z^3 - 4z^6 - 2z^7 = 1z^0 + 0z^1 + 3z^2 + 8z^3 + 0z^4 + 0z^5 - 0z^6 - 2z^7$ .  
 $s_k = \{1, 0, 3, 8, 0, 0, -4, -2\}$ .

---

Смысл величины  $z$  в  $z$ -полиноме заключается в том, что она является оператором единичной задержки по координатам функции. Умножение  $z$ -образа сигнала  $s(k)$  на величину  $z^n$  означает задержку сигнала на  $n$  интервалов:  $z^n S(z) \Leftrightarrow s(k-n)$ .

$Z$ -образы с положительными степенями  $z$  соответствуют каузальным (физически реализуемым) процессам и системам, которые работают в реальном масштабе времени с текущими и "прошлыми" значениями сигналов. При обработке информации на ЭВМ каузальность сигналов не относится к числу ограничений и возможно использование отрицательных степеней  $z$ , соответствующих отсчетам сигналов "вперед", например, при синтезе симметричных операторов

фильтров, что позволяет производить обработку информации без внесения в сигнал фазовых искажений. При использовании символики  $z^{-1}$  "прошлым" значениям соответствуют значения с отрицательными степенями  $z$ , "будущим" – с положительными.

Основное достоинство  $z$ -преобразований заключается в простоте математических операций со степенными полиномами, что имеет немаловажное значение при расчетах цифровых фильтров и спектральном анализе.

**Примеры z-преобразования** часто встречающихся на практике дискретных сигналов.

**Импульсы Кронекера.** В общем случае, в произвольной точке числовой оси:

$$\delta(k-n)=1 \text{ при } k=n, \quad \delta(k-n)=0 \text{ при } k \neq n.$$

$$X_{\delta}(z) = \sum_{k=-\infty}^{\infty} \delta(k-n) z^k = z^n.$$

Для импульса Кронекера в нулевой точке соответственно  $X_{\delta}(z) = z^0 = 1$ .

**Функция Хевисайда** (единичный скачок).

$$x(k) = 0 \text{ при } k < 0, \quad x(k) = 1 \text{ при } k \geq 0.$$

$$X(z) = \sum_{k=0}^{\infty} z^k = z^k.$$

Ряд сходится при  $|z| < 1$ , при этом его сумма равна:

$$X(z) = 1/(1-z), \quad |z| < 1.$$

При использовании символики  $z^{-1}$ :

$$X(z) = 1/(1-z^{-1}) = z/(z-1), \quad |z| > 1.$$

**Экспоненциальная функция:**

$$x(k) = 0 \text{ при } k < 0, \quad x(k) = a^k \text{ при } k \geq 0.$$

$$X(z) = \sum_{k=-\infty}^{\infty} x(k) z^k = \sum_{k=0}^{\infty} a^k z^k = \sum_{k=0}^{\infty} (az)^k.$$

Как и в предыдущем случае, ряд сходится при  $|az| > 1$ , т.е. при  $|z| < |a|$ , при этом:

$$X(z) = 1/(1-az), \quad |z| < |a|.$$

**Связь с преобразованиями Фурье и Лапласа.** Запишем дискретный сигнал  $s_k$  в виде суммы весовых импульсов Кронекера:

$$s_k = s(k\Delta t) = \sum_{n=-\infty}^{\infty} s(n\Delta t) \delta(k\Delta t - n\Delta t).$$

Определим спектр сигнала по теореме запаздывания:

$$S(\omega) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \exp(-j\omega k\Delta t).$$

Выполним замену переменных,  $z = \exp(-j\omega\Delta t)$ , и получим:

$$S(\omega) = \sum_{k=-\infty}^{\infty} s(k\Delta t) \cdot z^k = S(z).$$

Отсюда следует, что дискретное преобразование Фурье является частным случаем  $z$ -преобразования при  $z = \exp(-j\omega\Delta t)$ . Аналогичной подстановкой  $z = \exp(-p)$  может осуществляться переход к дискретному преобразованию Лапласа. В общем виде:

$$S(\omega) = S(z), \quad z = \exp(-j\omega\Delta t); \quad S(p) = S(z), \quad z = \exp(-p\Delta t). \quad (19.39)$$

Обратное преобразование:

$$S(z) = S(\omega), \quad \omega = \ln z / \Delta t; \quad S(z) = S(p), \quad p = \ln z / \Delta t. \quad (19.40)$$

При отрицательной символике  $z$  связь между представлениями осуществляется соответственно подстановками  $z^{-1} = \exp(j\omega\Delta t)$  и  $z^{-1} = \exp(p)$ .

**Свойства  $z$ -преобразования.** Без углубления в теорию, можно констатировать, что все свойства ДПФ действительны и для  $z$ -преобразования. Отметим некоторые из них.

**Линейность:** Если  $S(k) = a \cdot x(k) + b \cdot y(k)$ , то  $S(z) = aX(z) + bY(z)$ . Соответственно,  $z$ -преобразование допустимо только для анализа линейных систем и сигналов, удовлетворяющих принципу суперпозиции.

**Задержка** на  $n$  тактов:  $y(k) = x(k-n)$ .

$$\begin{aligned} Y(z) &= \sum_{k=-\infty}^{\infty} y(k) \cdot z^k = \sum_{k=-\infty}^{\infty} x(k-n) \cdot z^k = z^n \sum_{k=-\infty}^{\infty} x(k-n) \cdot z^{k-n} = \\ &= z^n \sum_{m=-\infty}^{\infty} x(m) \cdot z^m = z^n X(z). \end{aligned}$$

Соответственно, умножение  $z$ -образа сигнала на множитель  $z^n$  вызывает сдвиг сигнала на  $n$  тактов дискретизации.

Для  $z$ -преобразования действительны все известные теоремы о спектрах. В частности, свертка двух сигналов отображается в  $z$ -области произведением их  $z$ -образов, и наоборот:

$$s(k) * h(k) \Leftrightarrow S(z)H(z), \quad s(k) \cdot h(k) \Leftrightarrow S(z) * H(z).$$

При  $z = \exp(-j\omega\Delta t)$   $z$ -преобразование представляет собой особую форму представления дискретных сигналов, при которой на полином  $S(z)$  можно ссылаться как на временную функцию (по значениям коэффициентов  $k\Delta t$ ), так и на функцию частотного спектра сигнала (по значениям аргумента  $\omega$ ).

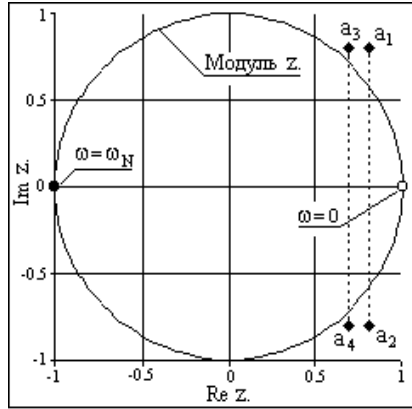


Рис. 19.22. Z - плоскость

**Отображение z-преобразования** выполняют на комплексной z-плоскости с  $\text{Re } z$  и  $\text{Im } z$  по осям координат (рис. 19.22). Спектральной оси частот  $\omega$  на z-плоскости соответствует окружность радиуса:

$$|z| = |\exp(-j\omega\Delta t)| = \sqrt{\cos^2(\omega\Delta t) + \sin^2(\omega\Delta t)} = 1.$$

Подстановка значения какой-либо частоты  $\omega$  в  $z = \exp(-j\omega\Delta t)$  отображается точкой на окружности. Частоте  $\omega = 0$  соответствует точка  $\text{Re } z = 1$  и  $\text{Im } z = 0$  на правой стороне оси абсцисс. При повышении частоты точка смещается по окружности против часовой стрелки, и занимает крайнее левое положение на частоте Найквиста

$$\omega_N = \pi/\Delta t \quad (\text{Re } z = -1, \text{Im } z = 0).$$

Отрицательные частоты спектра отображаются аналогично по часовой стрелке на нижней полуокружности. Точки  $\pm \omega_N$  совпадают, а при дальнейшем повышении или понижении частоты значения начинают повторяться в полном соответствии с периодичностью спектра дискретной функции. Проход по полной окружности соответствует одному периоду спектра, а любая гармоника спектра сигнала задается на плоскости двумя точками, симметричными относительно оси абсцисс.

Z-преобразование позволяет производить разложение сигналов и функций, например передаточных функций фильтров, на короткие составляющие операции свертки, для чего достаточно приравнять z-полином к нулю, найти его корни  $a_i$ , и переписать полином в виде произведения двучленов:

$$S(z) = a_0(z-a_1)(z-a_2)\dots,$$

где  $a_0$ -последний отсчет сигнала (коэффициент при старшей степени z).

Но произведению в  $z$ -области соответствует свертка в координатной области, и при обратном преобразовании двучлены  $(z-a_i)$  превращаются в двухточечные диполи  $\{-a_i, 1\}$ , а сигнал длиной  $N$  представляется сверткой  $(N-1)$  диполей:  $s_k = a_0\{-a_1, 1\} * \{-a_2, 1\} * \{-a_3, 1\} * \dots$

**Пример.**  $s_k = \{1.4464, -2.32, 3.37, -3, 1\}$ .

$$S(z) = z^4 - 3z^3 + 3.37z^2 - 2.32z + 1.4464. \quad a_0 = 1.$$

Корни полинома  $S(z)$ :  $a_1 = 0.8 + 0.8j$ ,  $a_2 = 0.8 - 0.8j$ ,  $a_3 = 0.7 + 0.8j$ ,  
 $a_4 = 0.7 - 0.8j$ ,

$$S(z) = (z - 0.8 - 0.8j)(z - 0.8 + 0.8j)(z - 0.7 - 0.8j)(z - 0.7 + 0.8j).$$

Корни полинома представлены на  $z$ -плоскости на рис. 19.22. Корни полинома комплексные и четыре двучлена в координатной области также будут комплексными. Но они являются сопряженными, и для получения вещественных функций следует перемножить сопряженные двучлены и получить биквадратные блоки:

$$S(z) = (z^2 - 1.4z + 1.13)(z^2 - 1.6z + 1.28).$$

При переходе в координатную область:

$$s_k = \{1.13, -1.4, 1\} * \{1.28, -1.6, 1\}.$$

Таким образом, исходный сигнал разложен на свертку двух трехчленных сигналов (функций).

**Аналитическая форма  $z$ -образов** существует для  $z$ -преобразований, если возможно свертывание степенного ряда в аналитическое выражение. Выше, в примерах  $z$ -преобразования, уже приводилось приведение к аналитической форме  $z$ -образов функции Хевисайда и экспоненциальной функции.

**Обратное  $z$ -преобразование** в общем случае производится интегрированием по произвольному замкнутому контуру  $C$ , расположенному в области сходимости и окружающему все особые точки (нули и полюсы)  $z$ -образа:

$$s_k = (1/2\pi j) \oint_C S(z) z^{k+1} dz$$

Способом, удобным для практического применения, является разложение рациональных  $S(z)$  на простые дроби. С учетом линейности преобразования:

$$S(z) = \sum_{n=1}^N a_n / (1 - b_n z) \Leftrightarrow \sum_{n=1}^N a_n (b_n)^k = s_k.$$

**Пример.**  $S(z) = 1/(1 - 5z + 6z^2) = 3/(1 - 3z) - 3/(1 - 2z) \Leftrightarrow 3 \cdot 3^k - 3 \cdot 2^k = s(k)$ .

При разложении функции  $S(z)$  по степеням  $z$  обратное  $z$ -преобразование не вызывает затруднений.

### 19.12. Дискретная свертка (конволюция)

Свертка – основной процесс в цифровой обработке сигналов. Поэтому важно уметь эффективно ее вычислять.

**Уравнение дискретной свертки** двух функций (сигналов) может быть получено непосредственно из интегрального уравнения свертки при замене интегрирования суммированием мгновенных значений функций с шагом  $\Delta t$ :

$$y(k\Delta t) = \Delta t \sum_n h(n\Delta t) s(k\Delta t - n\Delta t). \quad (19.41)$$

При выполнении дискретной свертки мы имеем дело с цифровыми массивами, при этом шаг дискретизации для массивов по физическому аргументу свертки должен быть равным и принимается за 1, а в качестве аргумента используется нумерация отсчетов в массивах:

$$y(k) = \sum_n h(n) s(k-n) \equiv \sum_n h_n s_{k-n} \equiv y_k. \quad (19.42)$$

$$y(k) = h(n) * s(k-n) \equiv s(k) * h(n) \equiv s_k * h_n.$$

**Техника свертки** приведена на рис. 19.23.

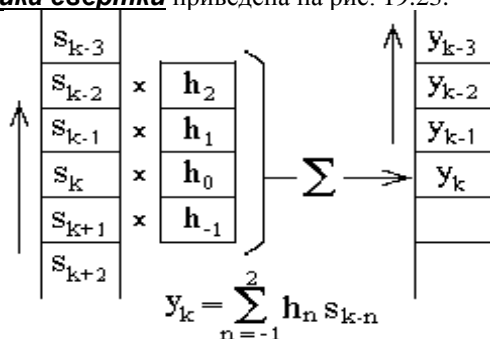


Рис. 19.23. Техника дискретной свертки.

Для вычисления свертки массив одной из функций ( $s_k$  - входного сигнала) располагается по ходу возрастания номеров. Массив второй функции ( $h_n$  - более короткой, оператор свертки), строится параллельно первому массиву в обратном порядке (по ходу уменьшения номеров, в режиме обратного времени). Для вычисления  $y_k$  значение  $h_0$  располагается против  $s_k$ , все значения  $s_{k-n}$  перемножаются с расположенными против них значениями  $h_n$  и суммируются. Результаты суммирования являются выходным значением функции  $y_k$ , после чего оператор  $h_n$  сдвигается на один номер  $k$  вперед (или функция  $s_k$  сдвигается ему навстречу) и вычисление повторяется для номера  $k+1$  и т.д.

В начальный момент свертки при вычислении значений  $y_k$  оператор  $h_n$ , построенный в режиме обратного времени, "зависает" для значений  $k-n$  при  $n > k$  против отсутствующих отсчетов входной



функции. "Зависание" исключают либо заданием начальных условий - дополнительных отсчетов, чаще всего нулевых или равных первому отсчету входной функции, либо началом свертки с отсчета входной функции  $k = n$  с соответствующим сокращением интервала выходной функции. Для операторов со значениями  $-n$  (вперед по времени) такой же момент может наступать и в конце входного массива.

**Пример.** Уравнение свертки:  $y_k = \sum_{n=0}^2 b_n x_{k-n} = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2}$ .

Значения оператора  $b_n$ :  $b_0 = 5$ ,  $b_1 = 3$ ,  $b_2 = 2$ . Входной сигнал:

$x_k = \{0, 1, 0, 0, 0\}$ , начальные условия:  $x_{-n} = 0$ .

Расчет выходного сигнала:

$$\begin{aligned} y_0 &= 5x_0 + 3x_{-1} + 2x_{-2} = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 = 0, \\ y_1 &= 5x_1 + 3x_0 + 2x_{-1} = 5 \cdot 1 + 3 \cdot 0 + 2 \cdot 0 = 5, \\ y_2 &= 5x_2 + 3x_1 + 2x_0 = 5 \cdot 0 + 3 \cdot 1 + 2 \cdot 0 = 3, \\ y_3 &= 5x_3 + 3x_2 + 2x_1 = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 1 = 2, \\ y_4 &= 5x_4 + 3x_3 + 2x_2 = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 = 0, \\ y_5 &= 5x_5 + 3x_4 + 2x_3 = 5 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 = 0 \end{aligned}$$

Выходной сигнал:  $y_k = \{0, 5, 3, 2, 0\}$

**Заметим:** свертка функции оператора с единичным входным сигналом представляет собой повторение функции оператора свертки на выходе.

На рис. 19.24 приведен пример выполнения дискретной свертки каузальным (односторонним) и четным (симметричным, двусторонним) оператором одного и того же сигнала.

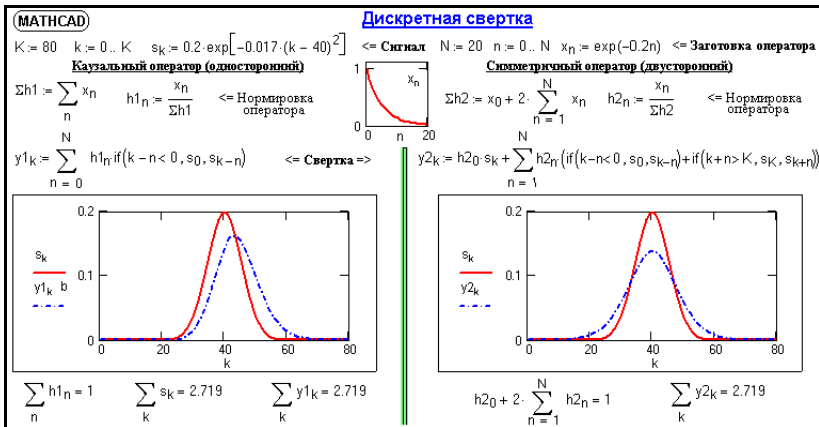


Рис. 19.24. Примеры выполнения дискретной свертки.

Прямое вычисление свертки требует  $K \cdot N$  умножений, где  $K$  – длина исходного сигнала, а  $N$  – длина ядра свертки. Как длина сигнала, так и длина ядра свертки может достигать нескольких тысяч точек, и число умножений становится огромным.

Для дискретной свертки действительны все свойства и теоремы интегральной свертки. В частности, свертка функций в координатной области отображается произведением их спектров в частотной области, а умножение в координатной области эквивалентно свертке в частотной области. Это значит, что для выполнения свертки двух сигналов можно перевести их в частотную область, перемножить их спектры, и перевести результат обратно во временную область, т.е. действовать по следующей схеме:

$$s(k) \Leftrightarrow S(\omega), \quad h(n) \Leftrightarrow H(\omega), \quad Y(\omega) = S(\omega) \cdot H(\omega), \quad Y(\omega) \Leftrightarrow y(k).$$

С появлением алгоритмов БПФ, позволяющих быстро вычислять преобразования Фурье, вычисление свертки через частотную область стало широко использоваться. При значительных размерах сигналов и длины ядра свертки такой подход позволяет в сотни раз сократить время вычисления свертки.

Выполнение произведения спектров может производиться только при одинаковой их длине, и оператор  $h(n)$  перед ДПФ необходимо дополнять нулями до размера функции  $s(k)$ .

Второй фактор, который следует принимать во внимание, это цикличность свертки при ее выполнении в спектральной области, обусловленная периодизацией дискретных функций. Перемножаемые спектры являются спектрами периодических функций, и результат на конечных интервалах может не совпадать с дискретной линейной сверткой, где условия продления интервалов (начальные условия) задаются, а не повторяют главный период.

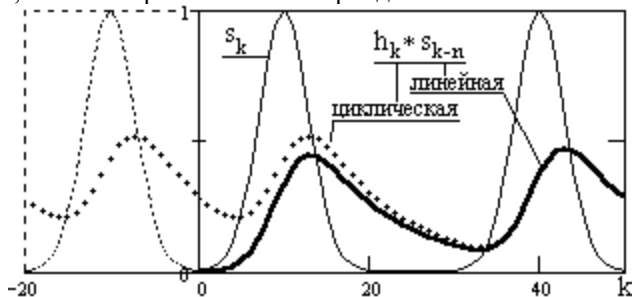


Рис. 19.25. Результаты двух видов свертки.

На рис. 19.25 приведены результаты свертки сигнала  $s_k$ , заданного на интервале  $k=(0-50)$ , с функцией  $h_n = a \cdot \exp(-a \cdot n)$ ,  $a = 0.1$ . Свертка, выполненная через ДПФ, в левой части интервала резко

отличается от линейной свертки. Характер искажения становится понятным, если дополнить главный интервал с левой стороны его периодическим продолжением (на рисунке показана часть левого бокового периода, свертка с которым заходит в главный период). Для операторов  $h_n$  со значениями  $n$ , вперед по положению, аналогичные искажения появятся и в правой части главного периода. Для устранения таких искажений сигнальная функция должна продлеваться нулями на размер оператора  $h(n)$ , что исключит наложение боковых периодов главной трассы функции.

При выполнении свертки через БПФ ощутимое повышение скорости вычислений появляется только при большой длине функций и операторов (например,  $M > 1000$ ,  $N > 100$ ). Следует также обращать внимание на разрядность результатов, т.к. перемножение чисел дает увеличение разрядности в 2 раза. При ограниченной разрядности числового представления с соответствующим округлением это может приводить к погрешностям суммирования.

В системах оперативной обработки данных часто возникает потребность вычислить свертку сигнала, поступающего на вход системы последовательными порциями (например, от датчиков скважинных приборов). В таких случаях применяется *секционная свертка*. Суть ее состоит в том, что каждая из этих частей сворачивается с ядром отдельно, а полученные части объединяются. Для объединения достаточно размещать их друг за другом с перекрытием в  $N-1$  точку ( $N$  – длина ядра свертки), и производить суммирование в местах перекрытия.

## **20. Случайные процессы и сигналы**

### **20.1. Введение в случайные процессы и сигналы**

Наряду с полезными информационными составляющими в реальных сигналах присутствуют помехи и шумы. К помехам обычно относят сигналы от других посторонних источников, "наводки" аппаратуры, влияние дестабилизирующих факторов на основной сигнал и т.п. Физическая природа помех, как правило, не случайна, и после соответствующего изучения может переводиться в разряд детерминированной помехи или исключаться из сигнала. К шумам относят случайные флуктуации сигнала, обусловленные природой его источника или устройств детектирования и формирования сигнала. При неизвестной природе помех они также могут относиться к числу случайных, если имеют случайное вероятностное распределение с

нулевым средним значением и дельта-подобную функцию автокорреляции.

Теория вероятностей рассматривает случайные величины и их характеристики в "статике". ***Задачи описания и изучения случайных сигналов "в динамике", как отображения случайных явлений, развивающихся во времени или по любой другой переменной, решает теория случайных процессов.***

В качестве универсальной координаты для распределения случайных величин по независимой переменной будем использовать, как правило, переменную "t" и трактовать ее, чисто для удобства, как временную координату. Распределения случайных величин во времени, а равно и сигналов их отображающих в любой математической форме, обычно называют случайными (или стохастическими) процессами. В технической литературе термины "случайный сигнал" и "случайный процесс" используются как синонимы.

В отличие от детерминированных сигналов значения случайных сигналов в произвольные моменты времени не могут быть вычислены. Они могут быть только предсказаны в определенном диапазоне значений с определенной вероятностью, меньшей единицы. Количественные характеристики случайных сигналов, позволяющие производить их оценку и сравнение, называют *статистическими*.

В процессе обработки и анализа физико-технических данных обычно приходится иметь дело с тремя типами сигналов, описываемых методами статистики. Во-первых, это информационные сигналы, отображающие физические процессы, вероятностные по своей природе, как, например, акты регистрации частиц ионизирующих излучения при распаде радионуклидов. Во-вторых, информационные сигналы, зависящие от определенных параметров физических процессов или объектов, значения которых заранее неизвестны, и которые обычно подлежат определению по данным информационным сигналам. И, в-третьих, это шумы и помехи, хаотически изменяющиеся во времени, которые сопутствуют информационным сигналам, но, как правило, статистически независимы от них как по своим значениям, так и по изменениям во времени. При обработке таких сигналов обычно ставятся задачи:

- обнаружение полезного сигнала,
- оценка параметров сигнала,
- выделение информационной части сигнала (очистка сигнала от шумов и помех),
- предсказание поведения сигнала на некотором последующем интервале (экстраполяция).

## 20.2. Случайные процессы и функции

Случайный процесс описывается статистическими характеристиками, называемыми моментами. Важнейшими характеристиками случайного процесса являются его стационарность, эргодичность и спектр мощности.

**Случайный процесс** в его математическом описании  $X(t)$  представляет собой функцию, которая отличается тем, что ее значения (действительные или комплексные) в произвольные моменты времени по координате  $t$  являются случайными. Строго с теоретических позиций, случайный процесс  $X(t)$  следует рассматривать как совокупность временных функций  $x_k(t)$ , имеющих определенную общую статистическую закономерность. При регистрации случайного процесса на определенном временном интервале осуществляется фиксирование единичной реализации  $x_k(t)$  из бесчисленного числа возможных реализаций процесса  $X(t)$ . Эта единичная реализация называется *выборочной функцией* случайного процесса  $X(t)$ . Отдельная выборочная функция не характеризует процесс в целом, но при определенных условиях по ней могут быть выполнены оценки статистических характеристик процесса. Примеры выборочных функций модельного случайного процесса  $X(t)$  приведены на рис. 20.1. В дальнейшем при рассмотрении различных параметров и характеристик случайных процессов для сопровождающих примеров будем использовать данную модель процесса.

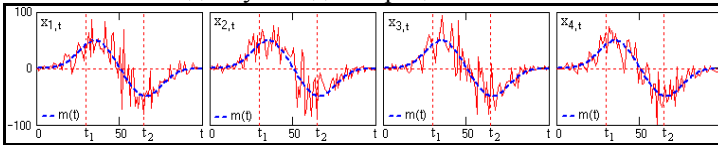


Рис. 20.1. Выборочные функции случайного процесса

### Функциональные характеристики случайного процесса.

С практической точки зрения выборочная функция является результатом отдельного эксперимента, после которого данную реализацию  $x_k(t)$  можно считать детерминированной функцией. Сам случайный процесс в целом должен анализироваться с позиции бесконечной совокупности таких реализаций, образующих *статистический ансамбль*. Полной статистической характеристикой процесса является  $N$ -мерная плотность вероятностей  $p(x_n; t_n)$ . Однако, как экспериментальное определение  $N$ -мерных плотностей вероятностей процессов, так и их использование в математическом анализе представляет значительные математические трудности.

Поэтому на практике обычно ограничиваются одно- и двумерной плотностью вероятностей процессов.

Допустим, что случайный процесс  $X(t)$  задан ансамблем реализаций  $\{x_1(t), x_2(t), \dots, x_k(t), \dots\}$ . В произвольный момент времени  $t_1$  зафиксируем значения всех реализаций  $\{x_1(t_1), x_2(t_1), \dots, x_k(t_1), \dots\}$ . Совокупность этих значений представляет собой случайную величину  $X(t_1)$  и является одномерным сечением случайного процесса  $X(t)$ . Примеры сечений случайного процесса  $X(t)$  по 100 выборкам  $x_k(t)$  (рис. 9.1.1) в точках  $t_1$  и  $t_2$  приведены на рис. 20.2.

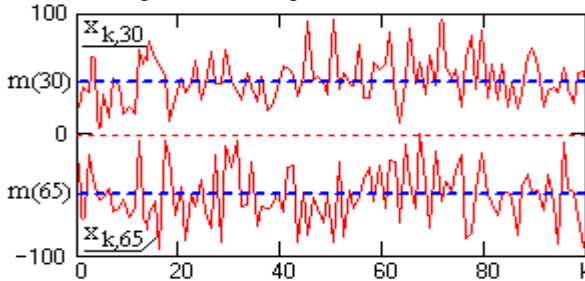


Рис. 20.2. Сечения случайного процесса  $X(t)$ .

**Одномерная функция распределения вероятностей**

$F(x, t_i)$  определяет вероятность того, что в момент времени  $t_i$  значение случайной величины  $X(t_i)$  не превысит значения  $x$ :

$$F(x, t_i) = P\{X(t_i) \leq x\}.$$

Очевидно, что в диапазоне значений вероятностей от 0 до 1 функция  $F(x, t)$  является неубывающей с предельными значениями  $F(-\infty, t)=0$  и  $F(\infty, t)=1$ . При известной функции  $F(x, t)$  вероятность того, что значение  $X(t_i)$  в выборках будет попадать в определенный интервал значений  $[a, b]$  определяется выражением:

$$P\{a < X(t_i) \leq b\} = F(b, t_i) - F(a, t_i).$$

**Одномерная плотность распределения вероятностей**  $p(x, t)$  случайного процесса  $X(t)$  определяет вероятность того, что случайная величина  $x(t)$  лежит в интервале  $\{x \leq x(t) \leq x+dx\}$ . Она характеризует распределение вероятностей реализации случайной величины  $X(t_i)$  в произвольный момент времени  $t_i$  и представляет собой производную от функции распределения вероятностей:

$$p(x, t_i) = dF(x, t_i)/dx. \tag{20.1}$$

Моменты времени  $t_i$  являются сечениями случайного процесса  $X(t)$  по пространству возможных состояний и плотность вероятностей  $p(x, t_i)$  представляет собой плотность вероятностей случайных величин  $X(t_i)$  данных сечений. Произведение  $p(x, t_i)dx$  равно вероятности

реализации случайной величины  $X(t_i)$  в бесконечно малом интервале  $dx$  в окрестности значения  $x$ , откуда следует, что плотность вероятностей также является неотрицательной величиной.

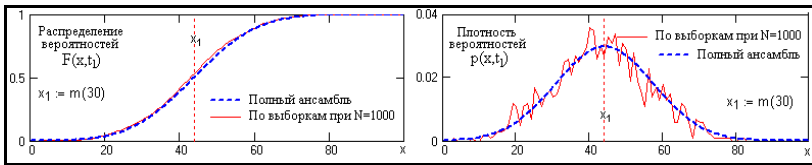


Рис. 20.3. Распределение вероятностей и плотность вероятностей сечения случайного процесса

На рис. 20.3 приведены примеры распределения вероятностей и плотности вероятностей сечения случайного процесса  $X(t)$  в точке  $t_1$  (рис. 20.1). Функции вероятностей определены по  $N=1000$  выборок дискретной модели случайного процесса и сопоставлены с теоретическими распределениями при  $N \rightarrow \infty$ .

При известной функции плотности вероятностей вероятность реализации значения  $X(t_i)$  в произвольном интервале значений  $[a, b]$  вычисляется по формуле:

$$P(a < X(t_i) \leq b) = \int_a^b p(x, t_i) dx.$$

Функция плотности вероятностей должна быть нормирована к 1, т.к. случайная величина обязана принимать какое-либо значение из числа возможных, образующих полное пространство случайных величин:

$$\int_{-\infty}^{\infty} p(x, t_i) dx = 1.$$

Плотность распределения вероятностей, соответственно, определяет функцию распределения вероятностей:

$$F(x, t_i) = \int_{-\infty}^x p(x, t_i) dx.$$

По известной плотности распределения вероятностей могут быть вычислены функции моментов случайного процесса, которые представляют собой математические ожидания соответствующих степеней (порядка) значений случайного процесса (начальные моменты) и значений флуктуационных составляющих процесса (центральные моменты, моменты относительно центров распределения случайных величин):

$$M\{x^n(t)\} = \int_{-\infty}^{\infty} x^n(t) p(x, t) dx,$$

$$M_0\{x^n(t)\} = M\{[x(t)-M\{x(t)\}]^n\} = \int_{-\infty}^{\infty} [x(t)-M\{x(t)\}]^n p(x, t) dx,$$

Функции моментов являются основными статистическими характеристиками случайного процесса. Они представляют собой неслучайные функции, но полностью и однозначно определяют случайный процесс, как и плотность распределения вероятностей, при определенном количестве порядков в зависимости от характера процесса. Минимальное число порядков, которое полностью определяет гауссово распределение плотности вероятностей, равно 2.

В практике анализа случайных процессов используются, в основном, начальные моменты первого порядка и центральные моменты второго порядка.

**Математическое ожидание** (mean value) является первым начальным моментом случайного процесса и представляет собой *статистическое усреднение* случайной величины  $X(t_i)$  в каком либо фиксированном сечении  $t_i$  случайного процесса. Соответственно, полная функция математического ожидания является теоретической оценкой среднего взвешенного значения случайного процесса по временной оси:

$$m_x(t) \equiv M\{X(t)\} \equiv \overline{x(t)} = \int_{-\infty}^{\infty} x p(x; t) dx, \quad (20.2)$$

Математическое ожидание  $m_x(t)$  представляет собой *неслучайную составляющую* случайного процесса  $X(t)$ . На рис. 20.1. и 20.2 неслучайные составляющие  $m(t)$  модели случайного процесса  $X(t)$  выделены пунктиром и соответствуют выборкам  $N \rightarrow \infty$ .

**Второй начальный момент** случайного процесса определяет его среднюю мощность:

$$w_x(t) \equiv M\{X^2(t)\} \equiv \overline{x(t)^2} = \int_{-\infty}^{\infty} x^2 p(x; t) dx, \quad (20.3)$$

**Функция дисперсии** (variance, function of a dispersion) случайного процесса. При анализе случайных процессов особый интерес представляет флуктуационная составляющая процесса, которая определяется разностью  $X(t)-m_x(t)$ . Функция дисперсии является теоретической оценкой среднего взвешенного значения разности  $X(t)-m_x(t)^2$ , т.е. является вторым центральным моментом процесса, и определяет мощность его флуктуационной составляющей:

$$D_x(t) = M\{[X(t)-m_x(t)]^2\} = M\{X^2(t)\} - m_x^2(t) = \int_{-\infty}^{\infty} [x_0(t)]^2 p(x; t) dx, \quad (20.4)$$

где  $x_0(t) = x(t)-m_x(t)$ .



**Функция среднего квадратического отклонения**

(standard deviation) служит амплитудной мерой разброса значений случайного процесса по временной оси относительно математического ожидания процесса:

$$\sigma_x(t) = \sqrt{D_x(t)}. \quad (20.5)$$

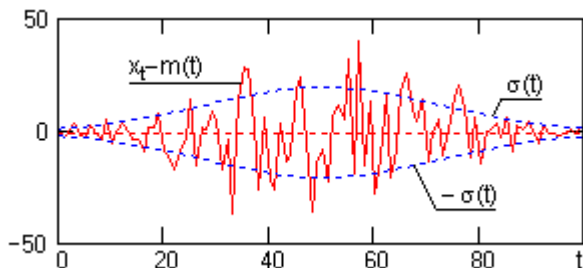


Рис. 20.4.

Учитывая последнее выражение, дисперсия случайной величины обычно обозначается индексом  $\sigma^2$ .

На рис. 9.1.4 приведен пример флюктуационной составляющей процесса  $X(t)$  (рис. 20.1) в одной из реализаций в сопоставлении со средним квадратическим отклонением  $\pm\sigma$  случайных величин от математического ожидания  $m(t)$ .

Одномерные законы плотности распределения вероятностей случайных процессов не несут каких-либо характеристик связи между значениями случайных величин для различных значений аргументов.

**Двумерная плотность распределения вероятностей**  $p(x_1, t_1; x_2, t_2)$  определяет вероятность совместной реализации значений случайных величин  $X(t_1)$  и  $X(t_2)$  в произвольные моменты времени  $t_1$  и  $t_2$ , что характеризует взаимосвязь случайного процесса в различные моменты времени и дает возможность определить характер изменения случайного процесса, т.е. динамику развития процесса во времени. Распределение описывает двумерную случайную величину  $\{X(t_i), X(t_j)\}$  в виде функции вероятности реализации случайной величины  $X(t_i)$  в бесконечно малом интервале  $dx_i$  в окрестностях  $x_i$  в момент времени  $t_i$  при условии, что в момент времени  $t_j$  значение  $X(t_j)$  будет реализовано в бесконечно малом интервале  $dx_j$  в окрестностях  $x_j$ :

$$p(x_1, t_1; x_2, t_2) = P\{x_1 \leq x(t_1) \leq x_1 + dx_1 \cap x_2 \leq x(t_2) \leq x_2 + dx_2\}.$$

С помощью двумерной плотности распределения вероятностей можно определить корреляционные функции процесса.

**Корреляционные функции случайных процессов.**

Характеристикой динамики изменения случайной величины  $X(t_i)$

является корреляционная функция, которая описывает случайный процесс в целом:

$$R_X(t_i, t_j) = M\{X(t_i) X(t_j)\}.$$

Корреляционная функция представляет собой статистически усредненное произведение значений случайного процесса  $X(t)$  в моменты времени  $t_i$  и  $t_j$  по всем значениям временных осей  $t_i$  и  $t_j$ , а, следовательно, тоже является двумерной функцией. В терминах теории вероятностей корреляционная функция является вторым начальным моментом случайного процесса.

На рис. 20.5 приведены примеры реализаций двух случайных процессов, которые характеризуются одной и той же функцией математического ожидания и дисперсии.

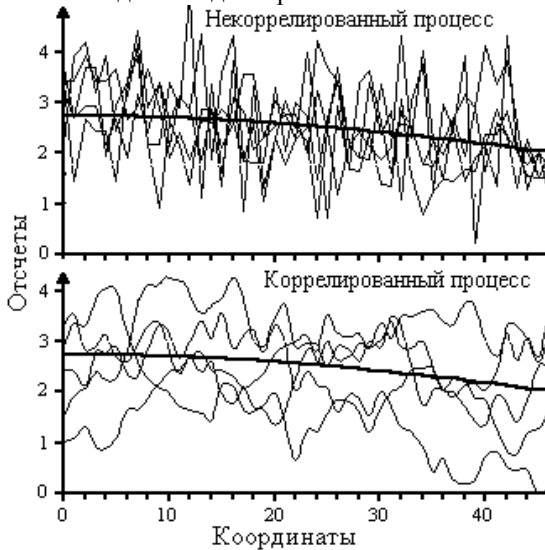


Рис. 20.5.

На рисунке видно, что хотя пространство состояний обоих процессов практически одно и то же, динамика развития процессов в реализациях существенно различается. Единичные реализации коррелированных процессов в произвольный момент времени могут быть такими же случайными, как и некоррелированных, а в пределе, во всех сечениях оба процесса могут иметь один и тот же закон распределения случайных величин. Однако динамика развития по координате  $t$  (или любой другой независимой переменной) единичной реализации коррелированного процесса по сравнению с некоррелированным является более плавной, а, следовательно, в коррелированном процессе имеется определенная связь между

последовательными значениями случайных величин. Оценка степени статистической зависимости мгновенных значений какого-либо процесса  $X(t)$  в произвольные моменты времени  $t_1$  и  $t_2$  и производится функцией корреляции. По всему пространству значений случайного процесса  $X(t)$  корреляционная функция определяется выражением:

$$R_X(t_i, t_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_i)x(t_j) p(x_i, t_j; x_i, t_j) dx_i dx_j, \quad (20.6)$$

При анализе случайных процессов второй момент времени  $t_j$  удобно задавать величиной сдвига  $\tau$  относительно первого момента, который при этом может быть задан в виде координатной переменной:

$$R_X(t, t+\tau) = M\{X(t)X(t+\tau)\}. \quad (20.7)$$

Функция, задаваемая этим выражением, обычно называется автокорреляционной функцией случайного процесса.

**Ковариационные функции.** Частным случаем корреляционной функции является функция автоковариации (ФАК), которая широко используется при анализе сигналов. Она представляет собой статистически усредненное произведение значений центрированной случайной функции  $X(t)-m_X(t)$  в моменты времени  $t_i$  и  $t_j$  и характеризует флюктуационную составляющую процесса:

$$K_X(t_i, t_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x(t_i)-m_X(t_i)) (x(t_j)-m_X(t_j)) p(x_i, t_j; x_i, t_j) dx_i dx_j, \quad (20.8)$$

В терминах теории вероятностей ковариационная функция является вторым центральным моментом случайного процесса. Для центрированных случайных процессов ФАК тождественна функции автокорреляции. При произвольных значениях  $m_X$  ковариационные и корреляционные функции связаны соотношением:

$$K_X(t, t+\tau) = R_X(t, t+\tau) - m_X^2(t).$$

Нормированная функция автоковариации (функция корреляционных коэффициентов):

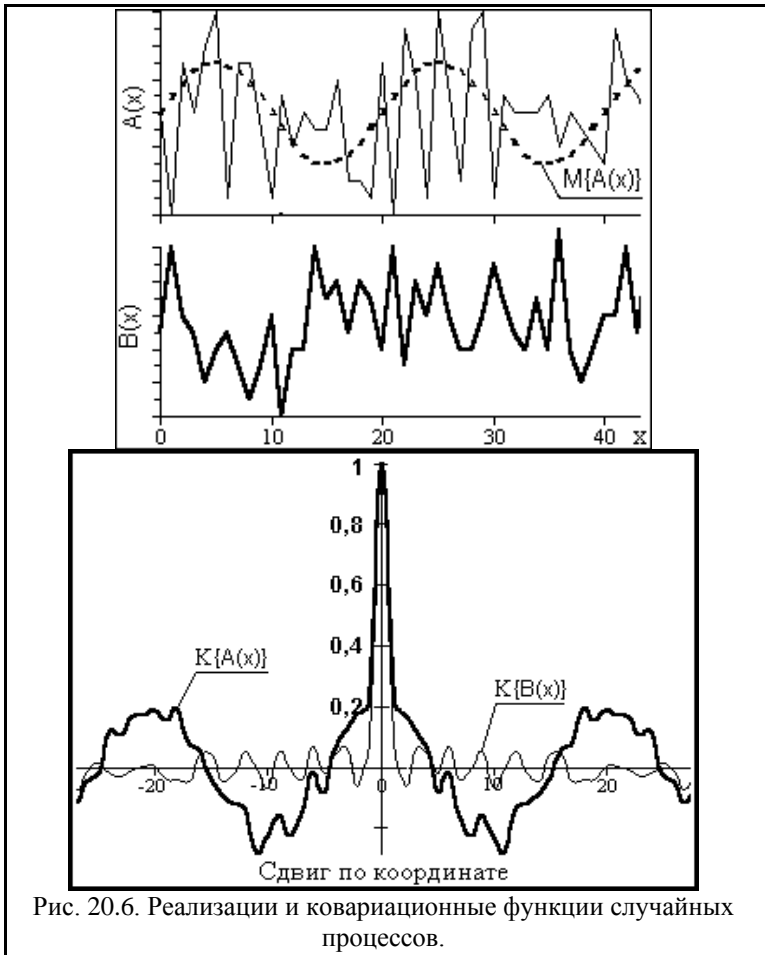
$$\rho_X(t, t+\tau) = K_X(t, t+\tau)/[\sigma(t)\sigma(t+\tau)]. \quad (20.9)$$

Функция корреляционных коэффициентов может принимать значения от +1 (полная статистическая корреляция случайных процессов на интервалах  $t$  и  $t+\tau$ ) до -1 (полная статистическая противоположность процессов на этих интервалах). Попутно отметим, что в математической статистике, а также довольно часто и в технической литературе, эту функцию называют функцией корреляции. При  $\tau = 0$  значение  $\rho_X$  равно 1, а ФАК вырождается в дисперсию случайного процесса:

$$K_X(t) = D_X(t).$$

Отсюда следует, что для случайных процессов и функций основными характеристиками являются функции математического

ожидания и корреляции (ковариации). Особой необходимости в отдельной функции дисперсии не имеется.



Примеры реализаций двух различных случайных процессов и их нормированных ковариационных функций приведены на рис. 20.6.

**Свойства функций автоковариации и автокорреляции.**

1. Максимум функций наблюдается при  $\tau = 0$ . Это очевидно, т.к. при  $\tau = 0$  вычисляется степень связи отсчетов с собой же, которая не может быть меньше связи разных отсчетов. Значение максимума функции корреляции равно средней мощности сигнала.

2. Функции автокорреляции и автоковариации являются четными:  $R_X(\tau) = R_X(-\tau)$ . Последнее также очевидно:

$$X(t)X(t+\tau) = X(t-\tau)X(t) \text{ при } t = t-\tau.$$

Говоря иначе, моменты двух случайных величин  $X(t_1)$  и  $X(t_2)$  не зависят от последовательности, в которой эти величины рассматриваются, и соответственно симметричны относительно своих аргументов:  $R_X(t_1, t_2) = R_X(t_2, t_1)$ .

3. При  $\tau \Rightarrow \infty$  значения ФАК для сигналов, конечных по энергии, стремятся к нулю, что прямо следует из физического смысла ФАК. Это позволяет ограничивать длину ФАК определенным максимальным значением  $\tau_{\max}$  - радиусом корреляции, за пределами которого отсчеты можно считать независимыми. Интегральной характеристикой времени корреляции случайных величин обычно считают *эффективный интервал корреляции*, определяемый по формуле:

$$T_k = \int_{-\infty}^{\infty} |\rho_X(\tau)| d\tau \equiv (1/K_X(0)) \int_{-\infty}^{\infty} |K_X(\tau)| d\tau. \quad (20.10)$$

Отсчеты (сечения) случайных функций, отстоящие друг от друга на расстояние большее  $T_k$ , при инженерных расчетах считают некоррелированными.

4. Если к случайной функции  $X(t)$  прибавить неслучайную функцию  $f(t)$ , то ковариационная функция не изменяется. Обозначим новую случайную функцию как  $Y(t)=X(t)+f(t)$ . Функция математического ожидания новой величины:  $\bar{y}(t) = \bar{x}(t) + f(t)$ . Отсюда следует, что  $Y(t) - \bar{y}(t) = X(t) - \bar{x}(t)$ , и соответственно

$$K_Y(t_1, t_2) = K_X(t_1, t_2).$$

5. Если случайную функцию  $X(t)$  умножить на неслучайную функцию  $f(t)$ , то ее корреляционная функция  $R_X(t_1, t_2)$  умножится на  $f(t_1) \cdot f(t_2)$ . Обоснование данного свойства проводится по методике, аналогичной предыдущему пункту.

6. При умножении функции случайного процесса на постоянное значение  $C$  значения ФАК увеличиваются в  $C^2$  раз.

**Взаимные моменты случайных процессов** второго порядка дают возможность оценить совместные свойства двух случайных процессов  $X(t)$  и  $Y(t)$  путем анализа произвольной пары выборочных функций  $x_k(t)$  и  $y_k(t)$ . Мера связи между двумя случайными процессами  $X(t)$  и  $Y(t)$  также устанавливается корреляционными функциями, а именно - функциями взаимной корреляции и взаимной ковариации. В общем случае, для произвольных фиксированных моментов времени  $t_1 = t$  и  $t_2 = t+\tau$ :

$$R_{XY}(t, t+\tau) = M\{X(t)Y(t+\tau)\}. \quad (20.11)$$

$$K_{XY}(t, t+\tau) = M\{(X(t)-m_x(t))(Y(t+\tau)-m_y(t+\tau))\}. \quad (20.12)$$

Взаимные функции являются произвольными функциями, не обладают свойствами четности или нечетности, и удовлетворяют следующим соотношениям:

$$R_{xy}(-\tau) = R_{yx}(\tau), \quad (20.13)$$
$$|R_{xy}(\tau)|^2 \leq R_x(0)R_y(0).$$

Если один из процессов центрированный, то имеет место  $R_{xy}(t) = K_{xy}(t)$ .

Нормированная взаимная ковариационная функция (коэффициент корреляции двух процессов) характеризует степень линейной зависимости между случайными процессами при данном сдвиге  $\tau$  одного процесса по отношению ко второму и определяется выражением:

$$\rho_{xy}(\tau) = K_{xy}(\tau)/(\sigma_x\sigma_y). \quad (20.14)$$

**Статистическая независимость случайных процессов** определяет отсутствие связи между значениями двух случайных величин  $X$  и  $Y$ . Это означает, что плотность вероятности одной случайной величины не зависит от того, какие значения принимает вторая случайная величина. Двумерная плотность вероятностей при этом должна представлять собой произведения одномерных плотностей вероятностей этих двух величин:

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

Это условие является обязательным условием статистической независимости случайных величин. В противном случае между случайными величинами может существовать определенная статистическая связь, как линейная, так и нелинейная. Мерой линейной статистической связи является коэффициент корреляции:

$$r_{xy} = [M\{X \cdot Y\} - M\{X\} \cdot M\{Y\}] / \sqrt{D\{X\}D\{Y\}}.$$

Значения  $r_{xy}$  могут изменяться в пределах от  $-1$  до  $+1$ . В частном случае, если случайные величины связаны линейным соотношением  $x=ay+b$ , коэффициент корреляции равен  $\pm 1$  в зависимости от знака константы  $a$ . Случайные величины некоррелированы при  $r_{xy}=0$ , при этом из выражения для  $r_{xy}$  следует:

$$M\{X \cdot Y\} = M\{X\} \cdot M\{Y\}.$$

Из статистической независимости величин следует их некоррелированность. Обратное не очевидно. Так, например, случайные величины  $x=\cos \varphi$  и  $y=\sin \varphi$ , где  $\varphi$  - случайная величина с равномерным распределением в интервале  $0 \dots 2\pi$ , имеют нулевой коэффициент корреляции, и вместе с тем их зависимость очевидна.

**Классификация случайных процессов.** Случайные процессы различают по степени однородности их протекания во

времени (по аргументу). Кроме моментов первого и второго порядка случайные процессы имеют моменты и более высоких порядков. По мере повышения порядка моментов вероятностная структура случайных процессов и их выборочных реализаций описывается все более детально. Однако практическая оценка этих моментов по выборкам ограничена, в основном, только стационарными случайными процессами.

**Стационарные процессы.** Процесс называют стационарным (более точно – слабо стационарным), если плотность вероятностей процесса не зависит от начала отсчета времени и если на интервале его существования выполняются условия постоянства математического ожидания и дисперсии, а корреляционная функция является функцией только разности аргументов  $\tau = t_2 - t_1$ , т.е.:

$$m_X(t_1) = m_X(t_2) = m_X = \text{const}, \quad (20.15)$$

$$D_X(t_1) = D_X(t_2) = D_X = \text{const},$$

$$R_X(t_1, t_1 + \tau) \equiv R_X(t_2 - \tau, t_2) = R_X(\tau) \equiv R_X(-\tau),$$

$$r_X(\tau) = R_X(\tau)/D_X, \quad r_X(0) = 1, \quad |r_X(\tau)| \leq 1, \quad r_X(-\tau) = r_X(\tau).$$

Последние выражения свидетельствует о четности корреляционной (а равно и ковариационной) функции и функции корреляционных коэффициентов. Из него вытекает также еще одно свойство смешанных моментов стационарных процессов:

$$|R_X(\tau)| \leq R_X(0), \quad |K_X(\tau)| \leq K_X(0) \equiv D_X.$$

Чем медленнее по мере увеличения значений  $\tau$  убывают функции  $R_X(\tau)$  и  $r_X(\tau)$ , тем больше интервал корреляции случайного процесса, и тем медленнее изменяются во времени его реализации.

Если от времени не зависят и моменты более высоких порядков (в частности, асимметрия и эксцесс), то такой процесс считается строго стационарным. В общем случае класс строго стационарных процессов входит в класс слабо стационарных. И только в случае гауссовых случайных процессов слабая стационарность автоматически влечет строгую, поскольку все характеристики этих процессов определяются средним значением и корреляционной функцией.

Стационарные случайные процессы наиболее часто встречаются при решении физических и технических задач. Теория стационарных случайных функций разработана наиболее полно. Случайные процессы, удовлетворяющие условиям стационарности на ограниченных, интересующих нас интервалах, также обычно рассматривают в классе стационарных и называют квазистационарными.

**Нестационарные процессы.** В общем случае значения функций математического ожидания, дисперсии и корреляции могут

быть зависимыми от момента времени  $t$ , т.е. изменяться во времени. Такие процессы составляют класс нестационарных процессов.

**Эргодические процессы.** Строго корректно характеристики случайных процессов оцениваются путем усреднения по ансамблю реализаций в определенные моменты времени (по сечениям процессов). Но большинство стационарных случайных процессов обладает **эргодическим свойством.** *Сущность его заключается в том, что по одной достаточно длинной реализации процесса можно судить обо всех его статистических свойствах так же, как по любому количеству реализаций. Другими словами, закон распределения случайных величин в таком процессе может быть одним и тем же как по сечению для ансамбля реализаций, так и по координате развития.* Такие процессы получили название **эргодических** (ergodic). Для эргодических процессов имеет место:

$$m_x(t) = M\{x(t)\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t) dt, \quad (20.16)$$

$$D_x(t) = M\{x(t) - m_x(t)\}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T (x(t) - m_x(t))^2 dt, \quad (20.17)$$

$$R_x(\tau) = M\{x(t)x(t+\tau)\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau) dt. \quad (20.18)$$

**Эргодичность является очень важным свойством случайных стационарных, и только стационарных процессов.** Математическое ожидание эргодического случайного процесса равно постоянной составляющей любой его реализации, а дисперсия является мощностью его флюктуационной составляющей. Так как определение функций производится по ограниченным статистическим данным одной реализации и является только определенным приближением к соответствующим фактическим функциям процессов, целесообразно называть эти функции статистическими. Свойства эргодичности могут проявляться только по отношению к двум первым моментам случайного процесса, что вполне достаточно для использования соответствующих методик исследования процессов. Практическая проверка эргодичности процесса обычно производится проверкой выполнения условия Слуцкого:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T K(\tau) d\tau = 0. \quad (20.19)$$

Если ковариационная функция процесса стремится к нулю при возрастании значения аргумента ( $\tau$ ), то процесс относится к числу эргодических, по крайней мере, относительно моментов первого и второго порядков.



**Пример.** Случайная функция задана выражением  $Z(t)=X(t)+Y$ , где  $X(t)$  - стационарная эргодичная функция,  $Y$  - случайная величина, некоррелированная с  $X(t)$ . Эргодична ли функция  $Z(t)$ ?

$$m_z(t) = m_x + m_y, \quad K_z(\tau) = K_x(\tau) + D_y.$$

Функция  $Z(t)$  стационарна, но не эргодична, так как при  $\tau \rightarrow \infty$  имеет место  $K_z(\tau) \Rightarrow D_y$ .

---

По формулам (20.16-20.18) можно вычислить моменты и для детерминированных процессов. Например, для периодической функции  $f(t)=a \sin \omega t$  автоковариационная функция описывается выражением:

$$K(\tau) = (a^2/2) \cos \omega \tau.$$

Соответственно, для произвольной периодической функции, представленной рядом Фурье (разложенной по рядам Фурье):

$$K(\tau) = (1/2) \sum_{n=0}^{\infty} a_n^2 \cos \omega_n \tau.$$

Таким образом, автоковариационная функция периодической функции также является периодической функцией от аргумента  $\tau$  - величины временного сдвига.

### **20.3. Функции спектральной плотности** **Каноническое разложение случайных функций.**

Введем понятие простейшей случайной функции, которая определяется выражением:

$$X(t) = X \cdot \varphi(t), \quad (20.20)$$

где  $X$  - обычная случайная величина,  $\varphi(t)$  - произвольная неслучайная функция. Математическое ожидание простейшей случайной функции:

$$m_x(t) = M\{X\varphi(t)\} = \varphi(t) \cdot M\{X\} = \varphi(t) \cdot m_x, \quad (20.21)$$

где  $m_x$  - математическое ожидание случайной величины  $X$ . При  $m_x = 0$  математическое ожидание  $m_x(t)$  также равно нулю для всех  $t$  и функция (20.20) в этом случае называется элементарной случайной функцией. Ковариационная функция элементарной случайной функции определится выражением:

$$K_x(t_1, t_2) = M\{X(t_1)X(t_2)\} = \varphi(t_1)\varphi(t_2) \cdot M\{X^2\} = \varphi(t_1)\varphi(t_2) \cdot D_x. \quad (20.22)$$

где  $D_x$  - дисперсия случайной величины  $X$ .

Центрированную случайную функцию  ${}^0X(t)$  можно представить суммой взаимно некоррелированных элементарных случайных функций:

$${}^0X(t) = \sum_{i=1}^M X_i \cdot \varphi_i(t), \quad (20.23)$$

Из взаимной некоррелированности элементарных случайных функций следует взаимная некоррелированность величин  $X_i$ . Математическое ожидание и ковариационная функция случайной функции  ${}^0X(t)$ :

$$M\{{}^0X(t)\} = M\left\{\sum_{i=1}^M X_i \cdot \varphi_i(t)\right\} = 0.$$

$$\begin{aligned} K_x(t_1, t_2) &= M\{{}^0X(t_1) {}^0X(t_2)\} = M\left\{\sum_{i,j} X_i \cdot \varphi_i(t_1) X_j \cdot \varphi_j(t_2)\right\} = \\ &= \sum_{i,j} \varphi_i(t_1) \varphi_j(t_2) M\{X_i X_j\}. \end{aligned}$$

В силу взаимной некоррелированности парных значений  $X_i X_j$  имеет место  $M\{X_i X_j\} = 0$  при  $i \neq j$ , и все члены суммы в последнем выражении равны нулю, за исключением значений при  $i = j$ , для которых  $M\{X_i X_j\} = M\{X_i^2\} = D_i$ . Отсюда:

$$K_x(t_1, t_2) = \sum_{i=1}^M \varphi_i(t_1) \varphi_i(t_2) D_i. \quad (20.24)$$

Произвольная нецентрированная случайная функция соответственно может быть представлена в виде

$$X(t) = m_x(t) + {}^0X(t) = m_x(t) + \sum_{i=1}^M X_i \cdot \varphi_i(t), \quad (20.25)$$

с математическим ожиданием  $m_x(t)$  и с той же самой ковариационной функцией (20.24) в силу свойств ковариационных функций, где  ${}^0X(t)$  - флуктуационная составляющая случайной функции  $X(t)$ . Выражение (20.25) и является каноническим разложением функции  $X(t)$ . Случайные величины  $X_i$  называются коэффициентами разложения, функции  $\varphi_i$  - координатными функциями разложения. При  $t_1 = t_2$  из (20.24) получаем функцию дисперсии случайной функции  $X(t)$ :

$$D_x(t) = \sum_{i=1}^M [\varphi_i(t)]^2 \cdot D_i. \quad (20.26)$$

Таким образом, зная каноническое разложение (20.25) функции  $X(t)$ , можно сразу определить каноническое разложение (20.24) ее ковариационной функции, и наоборот. Канонические разложения удобны для выполнения различных операций над случайными функциями. Это объясняется тем, что в разложении зависимость функции от аргумента  $t$  выражается через неслучайные функции  $\varphi_i(t)$ , а соответственно операции над функцией  $X(t)$  сводятся к соответствующим операциям математического анализа над координатными функциями  $\varphi_i(t)$ .

В качестве координатных функций разложения, как и при

анализе детерминированных сигналов, обычно используются гармонические синус-косинусные функции, а в общем случае комплексные экспоненциальные функции  $\exp(j\omega t)$ . С учетом последнего предварительно рассмотрим особенности представления случайных функций в комплексной форме.

**Комплексные случайные функции.** В общем случае случайный процесс может описываться комплексной случайной функцией:

$$Z(t) = X(t) + jY(t), \quad (20.27)$$

где  $X(t)$  и  $Y(t)$  - действительные случайные функции. Соответственно, математическое ожидание комплексной функции:

$$m_z(t) = m_x(t) + j \cdot m_y(t). \quad (20.28)$$

Заметим, что комплексное представление случайных функций не более чем удобная для анализа математическая форма их отображения, которая, с использованием выражений Эйлера, всегда может быть переведена в форму вещественных функций. Функции дисперсии, корреляции и ковариации должны представлять собой однозначные и неслучайные вещественные характеристики случайных процессов и функций, независимо от формы их математического представления. Это условие будет выполняться при использовании в выражениях моментов второго порядка операций умножения комплексных функций с комплексно сопряженными функциями. Так, выражение для вычисления корреляционной функции имеет следующий вид:

$$\begin{aligned} R_z(t_1, t_2) &= M\{Z(t_1) \cdot Z^*(t_2)\} = M\{[X(t_1) + jY(t_1)][X(t_2) - jY(t_2)]\} = \\ &= M\{X(t_1)X(t_2) + Y(t_1)Y(t_2) + j[Y(t_1)X(t_2) - X(t_1)Y(t_2)]\} = \\ &= R_x(t_1, t_2) + R_y(t_1, t_2) + j[R_{yx}(t_1, t_2) - R_{xy}(t_1, t_2)]. \end{aligned} \quad (20.29)$$

Если действительные и мнимые части комплексной функции некоррелированы, то  $R_{yx} = R_{xy} = 0$  и последний член выражения (20.29) также равен нулю.

Аналогичное выражение имеет место и для ковариационной функции. При  $t_1 = t_2 = t$  для функции дисперсии комплексной случайной величины имеем:

$$D_z(t) = M\{|Z(t) - m_z(t)|^2\} = D_x(t) + D_y(t), \quad (20.30)$$

Все приведенные выражения в общем случае могут использоваться для любых комплексных случайных функций с любым физическим смыслом переменной  $t$ .

**Финитное преобразование Фурье** случайных функций. По аналогии с функциями детерминированных сигналов, отдельно взятая на интервале  $0-T$  реализация  $x_k(t)$  стационарного случайного процесса  $^0X(t)$  может быть представлена в виде ряда Фурье:

$$x_k(t) = \sum_{i=-\infty}^{\infty} V_{x,k}(\omega_i) \exp(j\omega_i t), \quad (20.31)$$

$$V_{x,k}(\omega_i) = (1/T) \int_0^T x_k(t) \exp(-j\omega_i t) dt, \quad (20.32)$$

или, в односторонней тригонометрической форме:

$$x_k(t) = A_{x,k}(0) + 2 \sum_{i=1}^{\infty} (A_{x,k}(\omega_i) \cos(\omega_i t) + B_{x,k}(\omega_i) \sin(\omega_i t)), \quad (20.33)$$

$$A_{x,k}(\omega_i) = (1/T) \int_0^T x_k(t) \cos(\omega_i t) dt, \quad (20.34)$$

$$B_{x,k}(\omega_i) = (1/T) \int_0^T x_k(t) \sin(\omega_i t) dt. \quad (20.35)$$

где  $\omega_i = i \cdot \Delta\omega$  - частоты спектра,  $\Delta\omega = 2\pi/T$  - шаг по частоте. Выражения (20.32) обычно называют *спектральными характеристиками реализаций*. Из сравнения выражений (20.23) и (20.31) нетрудно сделать заключение, что выражения (20.31) относятся к числу канонических разложений случайных функций, при этом спектральная характеристика  $V_{x,k}(\omega)$ , а равно и ее составляющие  $A_{x,k}(\omega)$  и  $B_{x,k}(\omega)$ , также являются случайными функциями частоты - единичными реализациями случайных функций  $V_x(\omega)$ ,  $A_x(\omega)$  и  $B_x(\omega)$ . Соответственно, и частотное распределение амплитуд и фаз составляющих гармонических колебаний случайного процесса  ${}^0X(t)$  представляет собой случайные функции с соответствующими неслучайными функциями дисперсий.

Если функция  ${}^0X(t)$  является дискретной последовательностью случайных величин  ${}^0X(n \cdot \Delta t)$  в интервале по  $n$  от 0 до  $N$ , то, как это и положено для дискретных преобразований Фурье, расчет спектральных характеристик выполняется в главном частотном диапазоне (до частоты Найквиста  $\omega_N = \pi/\Delta t$ ), с заменой в выражениях (29.32) интегрирования на суммирование по  $n$  и с соответствующим изменением пределов суммирования в выражениях (20.31). Данное пояснение сохраняется и на все дальнейшие выкладки.

Спектральные характеристики единичных реализаций случайных процессов интереса, как правило, не представляют, и на практике используются довольно редко. Спектральная характеристика случайной функции  ${}^0X(t)$ , как ансамбля реализаций, может быть определена осреднением функций (20.31-32) по реализациям, в результате которого мы получим те же самые функции (20.31-32), только без индексов  $k$ . При этом, в силу центрированности стационарной случайной функции  ${}^0X(t)$ , мы должны иметь:

$$M\{X(t)\} = \sum_{i=-\infty}^{\infty} M\{V_x(\omega_i)\} \exp(j\omega_i t) = 0, \quad (20.36)$$

Последнее будет выполняться при условии  $M\{V_x(\omega_i)\} = 0$ , т.е. математическое ожидание значений спектральной характеристики центрированного стационарного случайного процесса должно быть равно нулю на всех частотах. Другими словами, спектральной характеристики центрированного стационарного случайного процесса не существует. Существуют только спектральные характеристики его отдельных реализаций, которые и используются, например, для моделирования этих реализаций.

Для произвольных нецентрированных случайных процессов  $X(t)$ , при записи последних в форме  $X(t) = m_x(t) + X(t)$ , будем соответственно иметь преобразование Фурье:

$$m_x(t) + X(t) \Leftrightarrow m_x(\omega) + V_x(\omega) = m_x(\omega),$$

т.е., по существу, функцию спектра (или спектральной плотности) неслучайной функции математического ожидания случайного процесса, естественно, в пределах той точности, которую может обеспечить выборочный ансамбль реализаций. Это лишний раз подтверждает отсутствие в спектрах случайных процессов какой-либо информации о флюктуационной составляющей процессов, и говорит о том, что фазы спектральных составляющих в реализациях процесса являются случайными и независимыми.

*С учетом вышеизложенного, под спектрами случайных процессов (или спектральной плотностью при интегральном преобразовании Фурье) повсеместно понимается не преобразования Фурье собственно случайных функций, а преобразования Фурье функций мощности случайных процессов, поскольку функции мощности не зависят от соотношения фаз спектральных составляющих процессов.*

### **Спектры мощности случайных функций**

определяются аналогично спектрам мощности детерминированных сигналов. Средняя мощность случайного процесса  $X(t)$ , зарегистрированного в процессе одной реализации на интервале  $0-T$ , с использованием равенства Парсеваля может быть вычислена по формуле:

$$W_T = \int_0^T [x^2(t)/T] dt = \int_{-\infty}^{\infty} [|X_T(f)|^2/T] df,$$

где  $X(f)$  – спектральная плотность единичной реализации  $x(t)$ . При увеличении интервала  $T$  энергия процесса на интервале неограниченно нарастает, а средняя мощность стремится к определенному пределу:

$$W = \int_{-\infty}^{\infty} \left[ \lim_{T \rightarrow \infty} \frac{1}{T} |X_T(f)|^2 \right] df,$$

где подынтегральная функция представляет собой спектральную плотность мощности данной реализации случайного процесса:

$$W(f) = \lim_{T \rightarrow \infty} \frac{1}{T} |X_T(f)|^2. \quad (20.37)$$

Очень часто это выражение называют просто спектром мощности. Плотность мощности является вещественной, неотрицательной и четной функцией частоты. В общем случае, плотность мощности необходимо усреднять по множеству реализаций, но для эргодических процессов допустимо усреднение по одной достаточно длительной реализации.

**Теорема Винера-Хинчина.** Рассмотрим сигнал  $q(t)$ , представляющий собой одну реализацию случайного стационарного эргодического процесса длительностью  $T$ . Для сигнала  $q(t)$  может быть определен спектр  $Q(\omega)$ . Если сдвинуть на  $\tau$  реализацию процесса, то получим спектр  $Q(\omega)\exp(j\omega\tau)$ . Для вещественных сигналов  $Q(\omega) = Q^*(\omega)$  равенство Парсеваля по энергии взаимодействия двух сигналов

$$\int_{-\infty}^{\infty} x(t) y^*(t) dt = \int_{-\infty}^{\infty} X(f) Y^*(f) df \quad (20.38)$$

может быть записано в следующей форме:

$$\int_{-\infty}^{\infty} q(t)q(t+\tau) dt = (1/2\pi) \int_{-\infty}^{\infty} Q(\omega)Q^*(\omega) \exp(j\omega\tau) d\omega. \quad (20.39)$$

Поделим обе части данного равенства на  $T$  и перейдем к пределу при  $T \Rightarrow \infty$ , при этом в его левой части мы увидим выражение для функции корреляции, а в правой части - преобразование Фурье спектра мощности сигнала:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T q(t)q(t+\tau) dt = \lim_{T \rightarrow \infty} \frac{1}{2\pi T} \int_{-\infty}^{\infty} |Q(\omega)|^2 \exp(j\omega\tau) d\omega, \quad (20.40)$$

$$R(\tau) = (1/2\pi) \int_{-\infty}^{\infty} W(\omega) \exp(j\omega\tau) d\omega. \quad (20.41)$$

Отсюда следует, что корреляционная функция случайного стационарного эргодического процесса представляет собой обратное преобразование Фурье его спектра мощности. Соответственно, для спектра мощности случайного процесса имеем прямое преобразование Фурье:

$$W(\omega) = \int_{-\infty}^{\infty} R(\tau) \exp(-j\omega\tau) d\tau. \quad (20.42)$$

В этом состоит суть теоремы Винера-Хинчина. Функции  $W(\omega)$  и  $R(\tau)$  являются вещественными и четными, а соответственно в тригонометрической форме:

$$R(\tau) = 2 \int_0^{\infty} W(f) \cos(2\pi f\tau) df, \quad W(f) = 2 \int_0^{\infty} R(\tau) \cos(2\pi f\tau) d\tau.$$

**Спектр ковариационных функций.** Так как ковариационные функции стационарных процессов являются частным случаем корреляционных функций, то эти выражения действительны и для ФАК, а, следовательно, преобразования Фурье ковариационных функций, являются спектрами мощности флюктуирующей составляющей процессов. С этих позиций дисперсия случайных процессов представляет собой среднюю мощность его флюктуаций

$$K(\tau=0) = \sigma^2 = (1/2\pi) \int_{-\infty}^{\infty} W(\omega) d\omega,$$

т.е., равна суммарной мощности всех его частотных составляющих процессов.

При представлении ковариационной функции на интервале  $0-T$ , шаг по спектру функции с учетом четности ковариационной функции устанавливается равным  $\Delta\omega = \pi/T$ ,  $\omega_i = i \cdot \Delta\omega$ , а спектр определяется обычно непосредственно по косинусам в односторонней форме:

$$K_x(\tau) = D_x(0)/2 + \sum_{i=1}^{\infty} D_x(\omega_i) \cos(\omega_i\tau), \quad (20.43)$$

$$D_x(\omega_i) = (2/T) \int_0^T K_x(\tau) \cos(\omega_i\tau) d\tau, \quad (20.44)$$

где  $D_x(\omega_i)$  в соответствии с (20.24) - дисперсии случайных величин  $V_x(\omega_i)$ , а равно и  $A_x(\omega_i)$  и  $B_x(\omega_i)$ , в разложениях (20.31). В комплексной форме, как обычно:

$$K_x(\tau) = \sum_{i=-\infty}^{\infty} D_x(\omega_i) \exp(j\omega_i\tau), \quad (20.45)$$

$$D_x(\omega_i) = (2/T) \int_0^T K_x(\tau) \exp(-j\omega_i\tau) d\tau, \quad (20.46)$$

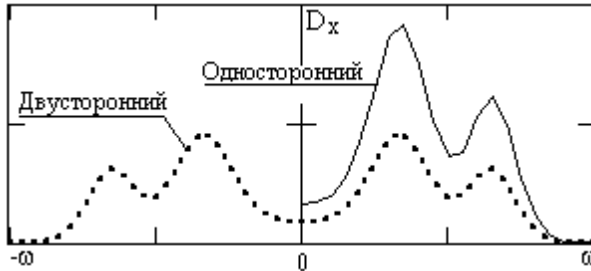


Рис. 20.7. Спектры случайных функций.

Спектры ковариационных функций всегда ограничены ( $D(\omega) \neq \infty$ ) и неотрицательны ( $D(\omega) \geq 0$ ), при двустороннем представлении всегда четные ( $D(-\omega) = D(\omega)$ ). Пример спектров в одно- и двустороннем представлении приведен на рис. 20.7.

Дисперсия стационарного случайного процесса  $X(t)$  может определяться по формуле (20.45) при  $\tau = 0$ :

$$D_x = \sum_{i=-\infty}^{\infty} D_x(\omega_i), \quad (20.47)$$

т.е. дисперсия стационарного случайного процесса равна сумме дисперсий всех случайных гармоник ее спектрального разложения.

**Эффективная ширина спектра мощности** является обобщенной характеристикой спектра случайного процесса и определяется по формуле:

$$B_k = \Delta\omega \cdot D_x / D_{\max}, \quad (20.47)$$

где  $D_{\max}$  – максимальное значение функции  $D_x(\omega_i)$ . Отметим, что ширина спектра является практической характеристикой случайного процесса, и вычисляется, как правило, для реальных частот по одностороннему спектру процесса.

При использовании предельного перехода  $T \Rightarrow \infty$  и соответственно интегралов Фурье в выражениях (20.45), двусторонние функции дисперсий  $D(\omega_i)$  заменяются функциями  $S(\omega)$ , а односторонние - функциями  $G(\omega)$ , которые называют соответственно дву- и односторонними функциями *спектральной плотности* случайных процессов. Такое же индексирование в научно-технической литературе применяют и для спектров корреляционных функций, а зачастую и для дискретных преобразований ковариационных функций вместо  $D(\omega_i)$ , хотя последнее применительно к ковариационным функциям более точно отражает физическую сущность величин. Но оно может считаться вполне приемлемым для сохранения общности математических описаний.

Эффективная ширина спектра для функций спектральной



плотности случайных процессов:

$$B_k = \int_0^\infty G_x(f) df / G_x(f)_{\max} = \int_0^\infty S_x(f) df / S_x(f)_{\max} = K_x(0) / S_x(f)_{\max}. \quad (20.48)$$

**Соотношение неопределенности** связывает эффективную ширину спектра  $B_k$  с эффективным интервалом ковариации  $T_k$ . Для его определения найдем произведение  $B_k T_k$  случайного процесса с использованием формул (20.10) и (20.48):

$$B_k T_k = 2 \int_0^\infty |K_x(\tau)| d\tau / S_x(f)_{\max}. \quad (20.49)$$

Оценка этого произведения и приводит к соотношению неопределенности:

$$B_k T_k \geq 1/2. \quad (20.50)$$

Следовательно, с уменьшением эффективной ширины спектра увеличивается эффективный интервал ковариации случайного процесса, и наоборот.

**Взаимные спектральные функции.** Статистическая связь двух случайных процессов  $X(t)$  и  $Y(t)$  оценивается по функциям взаимной ковариации  $K_{xy}(\tau)$  или  $K_{yx}(\tau)$ . Функции взаимной ковариации в общем случае являются произвольными, и соответственно функции взаимного спектра представляют собой комплексные выражения:

$$S_{xy}(\omega_i) = (1/T) \int_0^T K_{xy}(\tau) \exp(-j\omega_i \tau) d\tau, \quad (20.51)$$

при этом:

$$S_{xy}(-\omega) = S_{xy}^*(\omega) = S_{yx}(\omega).$$

Квадратурным аналогом нормированной взаимной ковариационной функции или функции коэффициентов ковариации двух процессов (20.14) в спектральной области является *функция когерентности*, которая определяется выражением:

$$\gamma_{xy}^2(\omega) = |S_{xy}(\omega)|^2 / (S_x(\omega) S_y(\omega)), \quad (20.52)$$

и для любых  $\omega$  удовлетворяет неравенствам

$$0 \leq \gamma_{xy}^2(\omega) \leq 1$$

Функция когерентности обычно используется при анализе линейных систем преобразования входной функции  $X(t)$  в выходную функцию  $Y(t)$  (рассмотрено ниже).

В заключение данного раздела еще раз отметим, что **спектральные плотности случайных процессов и спектры плотности мощности, это одно и то же понятие**. Оба термина используются достаточно широко в научно-технической литературе. Учитывая то обстоятельство, что понятие мощности по своему смыслу больше связано с энергетическими понятиями, а понятие спектральной

плотности - с анализом сигналов и систем, при дальнейшем рассмотрении случайных сигналов и процессов будем использовать, в основном, понятие спектральной плотности или (для дискретных величин) спектров случайных сигналов и процессов.

## **20.4. Преобразования случайных функций**

### **Системы преобразования случайных функций.**

Пусть имеется система преобразования с одним входом, на который поступает (подается) входная случайная функция  $X(t)$  - функция *воздействия* или *возбуждения*, и с одним выходом, с которого снимается выходная функция  $Z(t)$  - *отклик* или *выходная реакция* системы. Система осуществляет преобразование  $X(t) \Rightarrow Z(t)$  и описывается определенным *системным оператором* трансформации  $T$  - функцией, алгоритмом, набором правил преобразования входного сигнала в выходной. Обозначение операции преобразования:  $Z(t) = T[X(t)]$ . Символическое и полное отображение операции преобразования:

$$z(t) = h(\tau) \otimes x(t-\tau) = \int_0^{\infty} h(\tau) \cdot x(t-\tau) dt.$$

где  $h(\tau)$  - математическая функция импульсного отклика системы на единичное входное воздействие. Импульсный отклик определяет соответствующую частотную передаточную характеристику системы:  $h(\tau) \Leftrightarrow H(\omega)$ .

Для неслучайных (детерминированных) входных сигналов соотношение между выходными и входными сигналами всегда однозначно задается системным оператором. В случае реализации на входе системы случайного входного процесса (случайного сигнала) тоже существует однозначное соответствие процессов на выходе и входе системы, однако при этом одновременно происходит изменение статистических характеристик выходного сигнала (математического ожидания, дисперсии, ковариационной функции и пр.).

**Линейные и нелинейные системы** составляют два основных класса систем обработки сигналов. Термин линейности означает, что система преобразования сигналов должна иметь произвольную, но в обязательном порядке линейную связь между входным сигналом (возбуждением) и выходным сигналом (откликом). В нелинейных системах связь между входным и выходным сигналом определяется произвольным нелинейным законом.

**Основные системные операции** линейных систем, из которых могут быть сформированы любые линейные операторы преобразования, это *операции скалярного умножения, сдвига и*

**сложения сигналов:**

$$s(t) = c \times a(t), \quad s(t) = a(t-\Delta t), \quad s(t) = a(t)+b(t).$$

Для нелинейных систем выделим важный тип безынерционных операций нелинейной трансформации сигнала, результаты которой зависят только от его входных значений. К ним относятся, например, операции квадратирования и логарифмирования сигнала:

$$y(t) = [s(t)]^2, \quad y(t) = \log[s(t)].$$

**Система считается линейной**, если ее реакция на входные сигналы аддитивна (выполняется принцип суперпозиции сигналов) и однородна (выполняется принцип пропорционального подобия). Другими словами, **отклик линейной системы на взвешенную сумму входных сигналов должен быть равен взвешенной сумме откликов на отдельные входные сигналы независимо от их количества и для любых весовых коэффициентов, в том числе комплексных.**

Линейные системы могут быть неоднородными, если они осуществляют какое-либо линейное преобразование с прибавлением (вычитанием) заданной функции, т.е. операцию вида

$$Z(t) = T[X(t)] = T_0[X(t)] + f(t).$$

**Двухходовая система** описывается системным оператором  $T$ , который связывает два входных воздействия, соответственно  $X(t)$  и  $Y(t)$ , с выходной реакцией  $Z(t)$ . Система считается линейной, если принципы аддитивности и однородности выполняются для обоих входов. Двухходовая система может применяться, например, для суммирования двух случайных процессов с разными коэффициентами усиления их значений.

$$Z(t) = T[a(X_1(t)+X_2(t)), b(Y_1(t)+Y_2(t))] = a \cdot T[X_1(t), Y_1(t)] + b \cdot T[X_2(t), Y_2(t)].$$

**Связь выходных статистических функций с входными.** Для одноходовых систем при выполнении линейного преобразования  $Z(t) = T[X(t)]$  обычно ставится задача определения характеристик распределения  $Z(t)$  по известным характеристикам  $X(t)$ .

**Математическое ожидание выходного сигнала:**

$$m_z(t) = M\{Z(t)\} = M\{T[X(t)]\}.$$

Из теории линейных систем: *линейный оператор можно выносить за знак математического ожидания.* Отсюда следует:

$$m_z(t) = T[M\{X(t)\}] = T[m_x(t)], \quad (20.53)$$

т.е. для определения функции математического ожидания выходного сигнала  $Z(t)$  достаточно выполнить преобразование тем же системным оператором функции математического ожидания входного сигнала  $X(t)$ :

$$m_z(t) = h(\tau) \otimes m_x(t-\tau). \quad (20.54)$$

**Корреляционная функция выходного сигнала:**

$$R_z(t_1, t_2) = M\{Z(t_1)Z(t_2)\} = M\{T_1[X(t_1)] T_2[X(t_2)]\},$$

где  $T_1$  и  $T_2$  - один и тот же оператор  $T$  по переменным соответственно  $t_1$  и  $t_2$ , что позволяет вынести его за знак математического ожидания, сохраняя переменные:

$$R_z(t_1, t_2) = T_1 T_2 [M\{X(t_1)X(t_2)\}] = T_1 T_2 [R_x(t_1, t_2)], \quad (20.55)$$

т.е. **при известной функции корреляции входного сигнала функция корреляции выходного сигнала находится двойным преобразованием тем же оператором по двум аргументам.**

При определении функции  $R_z(\tau)$  следует учесть порядок преобразования. Для произведения выходных сигналов  $z(t)$  и  $z(t+\tau)$  линейной системы можно записать:

$$z(t) \cdot z(t+\tau) = \int_0^\infty \int_0^\infty h(\alpha)h(\beta) x(t-\alpha) x(t+\tau-\beta) d\alpha d\beta.$$

Если взять математические ожидания от обеих частей этого равенства, то, с учетом соотношения в подынтегральном выражении

$$M\{x(t-\alpha) x(t+\tau-\beta)\} = -R_x(t-\alpha-t+\tau+\beta) = R_x(\tau+\alpha-\beta),$$

получим:

$$R_z(\tau) = \int_0^\infty \int_0^\infty h(\alpha)h(\beta) R_x(\tau+\alpha-\beta) d\alpha d\beta \equiv R_x(\tau) \otimes h(\tau+\alpha) \otimes h(\tau-\beta). \quad (20.56)$$

Таким образом, функция корреляции выходного сигнала равна функции корреляции входного сигнала, свернутой дважды, в прямом и обратном направлении, с импульсным откликом системы, что сохраняет четность корреляционной функции выходного сигнала. Аналогичное заключение действительно и для ковариационных функций.

Заметим, что для свертки импульсных откликов, производя замену  $\tau-\beta = t$ , мы имеем равенство:

$$h(\tau+\alpha) \otimes h(\tau-\beta) = h(t+\alpha+\beta) \otimes h(t) = h(t) \otimes h(t+\gamma) = R_h(t),$$

где  $R_h(t)$  - функция корреляции импульсного отклика системы. Отсюда:

$$R_z(\tau) = R_x(\tau) \otimes R_h(\tau). \quad (20.57)$$

т.е. функция корреляции выходного сигнала равна свертке функции корреляции входного сигнала с функцией корреляции импульсного отклика системы. Это означает появление в случайном сигнале на выходе системы определенной ковариационной зависимости, вызванной инерционностью системы, причем радиус ковариации выходного сигнала обратно пропорционален верхней частоте, пропускаемой системой.

**Функции взаимной корреляции** входного и выходного сигналов определяются аналогично:

$$R_{zx}(t_1, t_2) = T_1[R_x(t_1, t_2)], \quad R_{xz}(t_1, t_2) = T_2[R_x(t_1, t_2)]. \quad (20.58)$$

Для функции  $R_{xz}$  входного и выходного сигналов имеем:

$$\int_0^{\infty} x(t) \cdot z(t+\tau) dt = \int_0^{\infty} \int_0^{\infty} h(\alpha) x(t) x(t+\tau-\alpha) d\alpha dt.$$

$$R_{xz}(\tau) = \int_0^{\infty} h(\alpha) R_x(\tau-\alpha) d\alpha \equiv R_x(\tau) \otimes h(\tau-\alpha). \quad (20.59)$$

т.е. функция взаимной корреляции входного и выходного сигналов равна свертке функции корреляции входного сигнала с функцией импульсного отклика системы.

Другая взаимно корреляционная функция  $R_{yx}$  может быть получена из соотношения:

$$R_{zx}(\tau) = R_{xz}(-\tau) \equiv R_x(\tau) \otimes h(\tau+\alpha). \quad (20.60)$$

Отметим, что для статистически независимых случайных величин при одностороннем импульсном отклике  $h(\tau) = 0$  при  $\tau < 0$  функция  $R_{xz}(\tau)$  также является односторонней, и равна 0 при  $\tau < 0$ , а функция  $R_{zx}$  соответственно равна 0 при  $\tau > 0$ .

**Спектральные соотношения**, которые характеризуют систему в целом по отношению к преобразованию случайных сигналов, это соотношения спектральных плотностей случайных сигналов (спектров мощности) на входе и выходе.

Применяя преобразование Фурье к выражениям (20.57), для спектра мощности выходного сигнала получаем:

$$S_z(f) = S_x(f) |H(f)|^2. \quad (20.61)$$

Спектр мощности случайного сигнала на выходе системы равен спектру мощности входного сигнала, умноженному на квадрат модуля частотной характеристики фильтра. С учетом четности ковариационных функций спектр мощности выходного сигнала также является четной действительной функцией и содержит только амплитудную характеристику системы.

Аналогично, для взаимного спектра мощности сигналов на основе выражений (20.59-20.60) имеем:

$$S_{xz}(f) = S_x(f) H(f), \quad S_{zx}(f) = S_x(f) H(-f). \quad (20.62)$$

Взаимный спектр сигналов при одностороннем импульсном отклике является комплексным, и содержит как амплитудную, так и фазовую характеристику системы.

Отметим, что с использованием выражения (20.62) можно производить определение частотной характеристики и импульсного отклика системы:

$$H(f) = S_{xz}/S_x \Leftrightarrow h(t).$$

**Дисперсия выходного сигнала** может быть определена с использованием формул (20.56-61) по функциям ковариации:

$$\sigma_z^2 = K_z(0) = \int_{-\infty}^{\infty} S_x(f) |H(f)|^2 df \equiv K_x(0) \int_0^{\infty} h^2(t) dt = \sigma_x^2 \int_0^{\infty} h^2(t) dt, \quad (20.63)$$

Если сигнал нецентрированный и значение дисперсии входного сигнала неизвестно, то по аналогичным формулам вычисляется сначала *средний квадрат* выходного сигнала или так называемая *средняя мощность сигнала*:

$$\overline{z^2} = \overline{z^2(t)} = R_z(0) \equiv \overline{x^2} \int_0^{\infty} h^2(t) dt \equiv \int_0^{\infty} S_x(f) |H(f)|^2 df. \quad (20.64)$$

Средняя мощность выходного сигнала равна средней мощности входного сигнала, умноженной на квадрат площади импульсной реакции системы (для цифровых систем - сумму квадратов коэффициентов импульсного отклика). Для центрированных случайных сигналов средняя мощность равна дисперсии сигналов. Для нецентрированных выходных сигналов:

$$\sigma_z^2 = \overline{z^2} - \bar{z}^2 \equiv (\overline{x^2} - \bar{x}^2) \int_0^{\infty} h^2(t) dt. \quad (20.65)$$

**Функция когерентности** дает оценку точности принятой линейной модели системы. Когерентность входного и выходного сигналов системы оценивается по формуле:

$$\gamma_{xz}^2(f) = |S_{xz}(f)|^2 / [S_x(f) \cdot S_z(f)]. \quad (20.66)$$

Если функции  $S_x(f)$  и  $S_z(f)$  отличны от нуля и не содержат дельта-функций, то для всех частот  $f$  значения функции когерентности заключены в интервале:

$$0 \leq \gamma_{xz}^2(f) \leq 1.$$

Для исключения дельта-функций на нулевой частоте определение функции когерентности производится по центрированным сигналам. Для линейных систем с постоянными параметрами функция когерентности равна 1, в чем нетрудно убедиться, если в формулу (20.66) подставить выражения  $S_{xz}$  и  $S_z$ , определенные через  $S_x$  в формулах (20.61-62). Для совершенно не связанных сигналов функция когерентности равна нулю. Промежуточные между 0 и 1 значения могут соответствовать трем ситуациям:

1. Система осуществляет преобразование  $x(t) \Rightarrow z(t)$ , но в измерениях этих сигналов или одного из них присутствует внешний шум. Так, например, в сигналах, зарегистрированных с ограничением по разрядности, появляется шум квантования (округления значений).
2. Система не является строго линейной. Это может наблюдаться, например, при определенном ограничении по разрядности вычислений в цифровых системах, при накоплении ошибки в рекурсивных системах и т.п.

3. Выходной сигнал  $z(t)$  помимо  $x(t)$  зависит еще от каких-то входных или внутренних системных процессов.

Величина  $1-\gamma_{xz}^2(f)$  задает долю среднего квадрата сигнала  $z(t)$  на частоте  $f$ , не связанную с сигналом  $x(t)$ .

Аналогично можно вычислить функцию когерентности двух реализаций  $x(t)$  и  $y(t)$ . Значения функции будут указывать на степень линейной зависимости одной реализации от другой, хотя это и не означает обязательности наличия какой-либо причинно-следственной связи между реализациями. Функция когерентности  $\gamma_{xy}$  сохраняется при точных однотипных линейных преобразованиях функций  $x(t)$  и  $y(t)$ , что позволяет производить ее определение без измерения самих величин  $x(t)$  и  $y(t)$ .

**Преобразования случайных функций.**

**Сложение случайных функций.** При сложении случайных функций, в общем случае, с произвольными постоянными коэффициентами  $a$  и  $b$ , и образовании случайной функции суммы  $Z(t) = a \cdot X(t) + b \cdot Y(t)$ , функция математического ожидания процесса  $Z(t)$ :

$$m_z(t) = M\{Z(t)\} = M\{aX(t)+bY(t)\} = a \cdot M\{X(t)\} + b \cdot M\{Y(t)\} = a \cdot m_x(t) + b \cdot m_y(t). \quad (20.67)$$

Корреляционная функция суммы вычисляется аналогично, и равна:

$$\begin{aligned} R_z(t_1, t_2) &= M\{Z(t_1) \cdot Z(t_2)\} = M\{[aX(t_1)+bY(t_1)][aX(t_2)+bY(t_2)]\} = \\ &= M\{a^2X(t_1)X(t_2)+b^2Y(t_1)Y(t_2)+ab[X(t_1)Y(t_2)+Y(t_1)X(t_2)]\} = \\ &= a^2R_x(t_1, t_2)+b^2R_y(t_1, t_2)+ab \cdot [R_{xy}(t_1, t_2)+R_{yx}(t_1, t_2)]. \end{aligned} \quad (20.68)$$

Для некоррелированных функций  $X(t)$  и  $Y(t)$  функции взаимной корреляции  $R_{xy}$  и  $R_{yx}$  обнуляются. Аналогичную форму записи имеют и ковариационные функции (как частный случай корреляционных функций при центрировании случайных процессов). Выражения легко обобщаются на сумму любого числа случайных функций. В частности, для корреляционной функции стационарной случайной функции  $Z(t) = \sum_i a_i X_i(t)$  при  $t_2 - t_1 = \tau$  имеем:

$$R_z(\tau) = \sum_i a_i^2 R_{xi}(\tau) + \sum_i \sum_{j, j \neq i} a_i a_j R_{xixj}(\tau). \quad (20.69)$$

При сложении случайной функции  $X(t)$  с неслучайной функцией  $y(t)$  математическое ожидание и корреляционная функция суммы  $Z(t) = X(t) + y(t)$  равны:

$$m_z(t) = m_x(t) + y(t), \quad R_z(t_1, t_2) = R_x(t_1, t_2). \quad (20.70)$$

При сложении случайной функции  $X(t)$  с некоррелированной случайной величиной  $Y$  математическое ожидание и корреляционная функция суммы  $Z(t) = X(t) + Y$ :

$$m_z(t) = m_x(t) + m_y, \quad R_z(t_1, t_2) = R_x(t_1, t_2) + D_y. \quad (20.71)$$

**Произведение случайной и неслучайной функций**  $X(t)$  и  $f(t)$ . Математическое ожидание и корреляционная функция выходного сигнала:

$$m_z(t) = M\{Z(t)\} = M\{f(t) \cdot X(t)\} = f(t) \cdot M\{X(t)\} = f(t) \cdot m_x(t). \quad (20.72)$$

$$R_z(t_1, t_2) = M\{f(t_1)X(t_1) f(t_2)X(t_2)\} = f(t_1)f(t_2)M\{X(t_1)X(t_2)\} = f(t_1)f(t_2) \cdot R_x(t_1, t_2). \quad (20.73)$$

Если  $f(t) = \text{const} = C$  и  $Z(t) = C \cdot X(t)$ , то соответственно имеем:

$$m_z(t) = C \cdot m_x(t), \quad R_z(t_1, t_2) = C^2 \cdot R_x(t_1, t_2). \quad (20.74)$$

**Производная от случайной функции**  $Z(t) = dX(t)/dt$ .

Если функция  $X(t)$  является непрерывной и дифференцируемой, то математическое ожидание производной:

$$m_z(t) = M\{Z(t)\} = M\{dX(t)/dt\} = d(M\{X(t)\})/dt = dm_x(t)/dt, \quad (20.75)$$

т.е. математическое ожидание производной от случайной функции равно производной от ее математического ожидания. Для корреляционной функции имеем:

$$R_z(t_1, t_2) = M\{(dX(t_1)/dt_1)(dX(t_2)/dt_2)\} = \frac{\partial^2}{\partial t_1 \partial t_2} M\{X(t_1)X(t_2)\} = \frac{\partial^2}{\partial t_1 \partial t_2} R_x(t_1, t_2), \quad (20.76)$$

т.е. корреляционная функция производной случайной функции равна второй смешанной частной производной от корреляционной функции исходной случайной функции.

**Интеграл от случайной функции**  $Z(t) = \int_0^t X(v)dv$ .

$$m_z(t) = M\{Z(t)\} = M\left\{\int_0^t X(v)dv\right\} = \int_0^t M\{X(v)\}dv = \int_0^t m_x(v)dv, \quad (20.77)$$

т.е. математическое ожидание интеграла от случайной функции равно интегралу от ее математического ожидания. Для корреляционной функции имеем:

$$R_z(t_1, t_2) = M\left\{\int_0^{t_1} X(t_1)dt_1 \int_0^{t_2} X(t_2)dt_2\right\} = M\left\{\int_0^{t_1} \int_0^{t_2} X(t_1)X(t_2)dt_1 dt_2\right\} = \int_0^{t_1} \int_0^{t_2} M\{X(t_1)X(t_2)\}dt_1 dt_2 = \int_0^{t_1} \int_0^{t_2} R_x(t_1, t_2)dt_1 dt_2, \quad (20.78)$$

т.е. корреляционная функция интеграла от случайной функции равна двойному интегралу от корреляционной функции исходной случайной функции.

**Преобразования стационарных случайных функций**

выполняются по вышеприведенным формулам и дают следующие результаты (вместо корреляционных функций приводятся ковариационные функции, которые обычно используются на



практике).

*Математическое ожидание* выходного сигнала  $Z(t)$  входной стационарной случайной функции  $X(t)$  по (20.54):

$$m_z = h(\tau) * m_x = m_x \int_{-\infty}^{\infty} h(\tau) d\tau, \quad (20.79)$$

Отсюда следует, что математическое ожидание выходных сигналов системы равно математическому ожиданию входных сигналов, умноженному на площадь (или сумму коэффициентов) импульсного отклика системы, т.е. на коэффициент усиления системой постоянной составляющей. Если система не пропускает постоянную составляющую сигналов (площадь или сумма коэффициентов импульсного отклика системы равна нулю), то случайный выходной сигнал всегда будет иметь нулевое математическое ожидание.

*Сумма двух стационарных случайных функций*  $X(t)$  и  $Y(t)$  дает стационарную случайную функцию  $Z(t)$ , при этом:

$$m_z = m_x + m_y, \quad D_z = D_x + D_y + 2K_{xy}(0). \quad (20.80)$$

$$K_z(t_1, t_2) = K_z(\tau) = K_x(\tau) + K_y(\tau) + K_{xy}(\tau) + K_{yx}(\tau). \quad (20.81)$$

*Сумма стационарной случайной и неслучайной функций*  $X(t)$  и  $y(t)$  нестационарна по математическому ожиданию:

$$m_z(t) = m_x + y(t), \quad K_z(\tau) = K_x(\tau). \quad (20.82)$$

*Произведение стационарной случайной и неслучайной функций*  $X(t)$  и  $y(t)$  - нестационарная случайная функция, так как:

$$m_z(t) = y(t) \cdot m_x, \quad D_z(t) = y^2(t) \cdot D_x. \quad (20.83)$$

$$K_z(t, \tau) = y(t)y(t+\tau)K_x(\tau). \quad (20.84)$$

*Производная от стационарной случайной функции* - стационарная случайная функция с математическим ожиданием  $m_z = 0$  и ковариационными функциями:

$$K_z(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} K_x(t_1 - t_2) = -\frac{d^2}{dt^2} K_x(\tau) = K_z(\tau). \quad (20.85)$$

$$K_{zx}(\tau) = d(K_x(\tau))/d\tau, \quad K_{xz}(\tau) = -d(K_x(\tau))/d\tau. \quad (20.86)$$

Из выражения (9.3.85) следует также, что для дифференцируемости  $X(t)$  необходимо, чтобы ее ковариационная функция была дважды дифференцируемой по  $t$ .

*Интеграл от стационарной случайной функции* - нестационарная случайная функция с математическим ожиданием

$m_z(t) = \int_0^t m_x(t) dt$  и функцией ковариации:

$$K_z(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} K_x(u_1 - u_2) du_1 du_2. \quad (20.87)$$

### 20.5. Модели случайных сигналов и помех

Наиболее распространенными моделями случайных сигналов и помех являются телеграфный сигнал, белый шум, гауссовый случайный процесс, гауссовый шум.

**Телеграфный сигнал** - это случайный процесс  $x_k(t)$ , представляющий собой последовательность прямоугольных положительных и отрицательных импульсов со случайными длительностями и детерминированными значениями амплитуд  $c$  и  $-c$ , причем перемены знака внутри любого интервала  $(t, t+\tau)$  происходят с интенсивностью  $\alpha$  в случайные моменты времени, и не зависят от процессов в смежных интервалах времени (рис. 20.8).

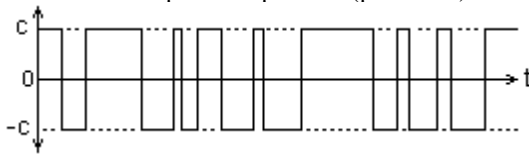


Рис. 20.8. Телеграфный сигнал.

Если считать случайной величиной телеграфного сигнала значение  $n$  - количество перемен знака внутри интервала  $\tau$ , то распределение вероятностей значений  $n$  будет описываться законом Пуассона:

$$P(n) = (\alpha|\tau|)^n \exp(-\alpha|\tau|)/n! \quad (20.88)$$

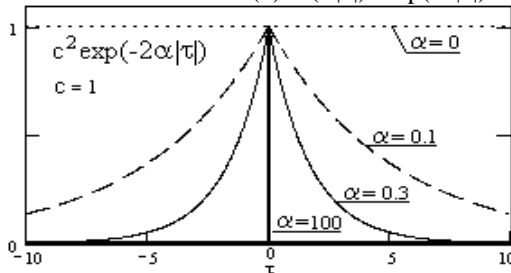


Рис. 20.9. Функция корреляции сигнала.

При вычислении корреляционной функции (рис. 20.9) телеграфного сигнала каждое отдельное произведение  $x_k(t)x_k(t+\tau)$  равно либо  $c^2$ , либо  $-c^2$  в зависимости от совпадения или несовпадения знаков  $x_k(t)$  и  $x_k(t+\tau)$ , причем вероятность  $c^2$  равна сумме вероятностей  $P(0)+P(2)+P(4)+\dots$ , а вероятность  $-c^2$  определяется соответственно суммой вероятностей  $P(1)+P(3)+P(5)+\dots$ .

Следовательно:

$$R_x(\tau) = c^2 \sum_{n=0}^{\infty} (-1)^n P(n) = c^2 \exp(-\alpha|\tau|) \sum_{n=0}^{\infty} (-1)^n (\alpha|\tau|)^n / n! = c^2 \exp(-2\alpha|\tau|). \quad (20.89)$$

Параметр  $\alpha$  полностью определяет ковариационные и спектральные свойства телеграфного сигнала. При  $\alpha \Rightarrow 0$  характеристики сигнала приближаются к характеристикам постоянной составляющей, при  $\alpha \Rightarrow \infty$  - к характеристикам белого шума.

Интервал ковариации сигнала:

$$T_k = 2 \int_0^{\infty} (R_x(\tau)/c^2) d\tau = 2/\alpha. \quad (20.90)$$

Отсюда следует, что чем больше  $\alpha$ , тем меньше время ковариации процесса. При  $\alpha \Rightarrow 0$   $T_k \Rightarrow \infty$  и процесс вырождается в детерминированный (стремится к постоянной составляющей).

При  $\alpha \Rightarrow \infty$   $T_k \Rightarrow 0$  и процесс вырождается в белый шум с некоррелированными отсчетами даже на соседних временных точках.

Двусторонняя спектральная плотность сигнала:

$$S_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) \exp(-j\omega\tau) d\tau = \alpha c^2 / (\alpha^2 + \omega^2). \quad (20.91)$$

Односторонняя спектральная плотность:

$$G_x(\omega) = 2\alpha c^2 / (\alpha^2 + \omega^2). \quad (20.92)$$

Ширина спектра телеграфного сигнала:

$$B_k = \int_0^{\infty} G_x(\omega) d\omega / G_x(0) \equiv \int_0^{\infty} S_x(\omega) d\omega / S_x(0) = \alpha\pi. \quad (20.93)$$

Отсюда следует, что спектр случайного процесса тем шире, чем меньше интервал ковариации процесса (рис. 20.10).

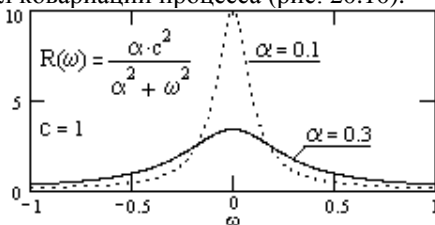


Рис. 20.10. Спектр сигнала.

**Белый шум** является стационарным случайным процессом  $q(t)$ , у которого автокорреляционная функция описывается дельта - функцией Дирака и, соответственно, спектральная плотность мощности не зависит от частоты и имеет постоянное значение  $W_q(f) = \sigma^2$ , равное дисперсии значений  $q(t)$ . Другими словами, все спектральные составляющие белого шума имеют одинаковую мощность (как белый цвет содержит все цвета видимого спектра). По существу, это идеализированный случайный процесс с бесконечной

энергией. Но в случае постоянства спектральной плотности мощности случайного процесса в конечном диапазоне частот введение такой идеализации позволяет разрабатывать достаточно легко реализуемые оптимальные методы фильтрации. Многие помехи в радиотехнике, в технике связи и в других отраслях, в том числе в информатике, рассматривают как белый шум, если эффективная ширина спектра сигналов  $B_s$  много меньше эффективной ширины спектра шумов  $B_q$

$$B_s/B_q \ll 1,$$

и спектральная плотность мощности шумов слабо изменяется в интервале спектра сигнала. Понятие "белый шум" определяет только спектральную характеристику случайного процесса, а, следовательно, под это понятие подпадают любые случайные процессы, имеющие равномерный энергетический спектр и различные законы распределения. Если частотный диапазон спектра, на котором рассматриваются сигналы и помехи, равен  $0-B$ , то спектральная плотность шума задается в виде:

$$W_q(f)=\sigma^2, 0 \leq f \leq B; W_q(f)=0, f > B, \quad (20.94)$$

при этом корреляционная функция шума определяется выражением:

$$R_q(\tau)=\sigma^2 B \cdot \sin(2\pi B\tau)/2\pi B\tau. \quad (20.95)$$

Эффективный интервал корреляции:

$$T_k = 2 \int_0^{\infty} |R_q(\tau)|d\tau /R_q(0). \quad (20.96)$$

Реальный интервал корреляции целесообразно определять по ширине главного максимума функции  $R_q(\tau)$  (значения  $\tau$  при первых пересечениях нулевой линии), в котором сосредоточена основная часть энергии шумов, при этом  $T_k = 1/B$  и  $BT_k = 1 > 1/2$ , т.е. соотношение неопределенности выполняется.

Как следует из всех этих выражений и наглядно видно на рис. 20.11, при ограничении частотного диапазона в шумах появляется определенная корреляция между значениями, и, чем меньше частотный диапазон шумов, тем больше их радиус корреляции.

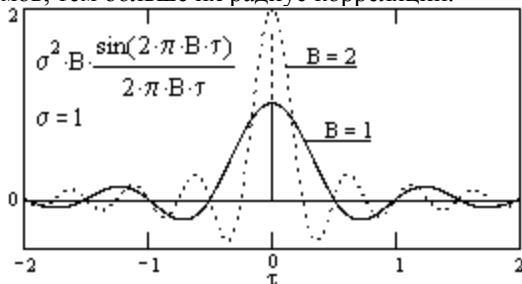


Рис. 20.11. Функции корреляции белого шума в частотном интервале  $0-B$ .

По существу, ограничение шумов определенным частотным диапазоном эквивалентно фильтрации белого шума частотным фильтром с соответствующей шириной полосы пропускания, при этом, корреляционная функция импульсного отклика фильтра свертывается с дельта – функцией белого шума.

Модель белого шума  $q(t)$  можно формировать как случайную по времени (аргументу) последовательность дельта - импульсов  $\delta(t_i)$  со случайными амплитудными значениями  $a_i$ :

$$q(t) = \sum_i a_i \delta(t-t_i), \quad (20.97)$$

которая удовлетворяет условиям статистической однородности: постоянное среднее число импульсов в единицу времени и статистическая независимость появления каждого импульса от предыдущих. Такой поток импульсов, который называют пуассоновским, является некоррелированным и имеет равномерный спектр плотности мощности:

$$W_q(\omega) = c^2 = N\sigma_a^2,$$

где  $N$  - число импульсов на интервале  $T$  реализации случайного процесса,  $\sigma_a^2$  - дисперсия амплитуд импульсов.

Спектральное описание белого шума оказывается удобным при учете влияния на него амплитудно-частотных характеристик различных устройств. Если на входе фильтра с импульсным откликом  $h(t)$  действует белый шум  $q(t)$ , то сигнал на выходе фильтра:

$$g(t) = h(t) \otimes q(t) = h(t) \otimes \sum_i a_i \delta(t-t_i) = \sum_i a_i h(t-t_i), \quad (20.98)$$

т.е. выходной сигнал будет представлять собой последовательность сигналов импульсной реакции фильтра  $h(t)$  с амплитудой  $a_i$ , при этом автокорреляционная функция и спектр мощности выходного потока также становятся подобными ФАК и спектру мощности импульсной реакции фильтра, и в первом приближении определяются выражениями:

$$R_g(\tau) \approx N \sigma_a^2 R_h(\tau) = c^2 R_h(\tau), \quad (20.99)$$

$$W_g(\omega) \approx N \sigma_a^2 |H(\omega)|^2 = c^2 |H(\omega)|^2. \quad (20.100)$$

Этот результат известен как теорема Кэмпбелла.

**Гауссовый шум** возникает при суммировании статистически независимых белых шумов и имеет следующую функцию корреляции:

$$R_x(\tau) = a \exp(-2\pi\sigma^2\tau^2). \quad (20.101)$$

Спектральная плотность шумов:

$$S_x(f) = (a/\sigma\sqrt{2\pi}) \exp(-f^2/2\sigma^2), \quad -\infty < f < \infty. \quad (20.102)$$

Эффективные шумовые ширина спектра и время ковариации:

$$B_k = \sigma\sqrt{2\pi}/2 = 1.25\sigma, \quad T_k = 1/\sigma\sqrt{2\pi} = 0.4/\sigma. \quad (20.103)$$

Соотношение неопределенности превращается в равенство:

$$B_k T_k = 1/2.$$

**Гауссовы случайные процессы** преобладают в практических задачах. Случайный процесс  $x(t)$  называется гауссовым, если для любого набора фиксированных моментов времени  $t_n$  случайные величины  $x(t_n)$  подчиняются многомерному нормальному распределению. Плотность вероятностей мгновенных значений  $x(t)$  эргодического гауссового процесса определяется выражением:

$$p(x) = (\sigma_x \sqrt{2\pi})^{-1} \exp(-(x-m_x)^2/2\sigma^2). \quad (20.104)$$

Среднее значение и его оценка по достаточно большому интервалу  $T$ :

$$m_x = \int_{-\infty}^{\infty} x p(x) dx, \quad m_x \approx (1/T) \int_0^T x(t) dt.$$

При нулевом среднем (или при центрировании функции  $x(t)$  для упрощения расчетов) дисперсия не зависит от переменной  $t$ , и равна:

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

Оценка дисперсии при больших значениях  $T$ :

$$\sigma_x^2 \approx (1/T) \int_0^T x^2(t) dt = \int_{-\infty}^{\infty} S_x(f) df = \int_0^{\infty} G_x(f) df. \quad (20.105)$$

Следовательно, плотность вероятностей гауссового процесса полностью характеризуется спектральной плотностью, по которой можно определить значение дисперсии процесса. На вид спектральных плотностей и соответствующих им ковариационных функций никаких ограничений не накладывается.

## **21. Стационарные линейные системы.**

### **21.1. Введение в стационарные линейные системы.**

Преобразование и обработка сигналов осуществляется в системах. Понятия сигнала и системы неразрывны, так как любой сигнал существует в какой-либо системе его обращения. Система обработки сигналов может быть реализована как в материальной форме (специальное устройство, измерительный прибор и т.п.), так и программно на ЭВМ или на любом другом вычислительном устройстве. Существуют и комплексные измерительно-вычислительные системы (ИВС), которые выполняют как регистрацию и первичную обработку сигналов непосредственно в материальной

форме их представления, так и преобразование сигналов в цифровую форму, и последующую программную обработку. Форма реализации систем существенного значения не имеет и определяет только их возможности при анализе и обработке сигналов. Основное внимание при рассмотрении данной темы будем уделять цифровым системам и дискретной математике их отображения и анализа, применяя аналитическую математику при рассмотрении общих вопросов, если последнее упрощает изложение и понимание теоретического материала.

## **21.2. Линейные системы**

**Общие понятия систем.** Безотносительно к назначению и исполнению система всегда имеет *вход*, на который подается входной сигнал или входное воздействие, в общем случае многомерное, и *выход*, с которого снимается обработанный выходной сигнал. Если устройство системы и внутренние операции преобразований принципиального значения не имеют, то система в целом может восприниматься как “черный ящик”, в формализованном виде. Формализованная система представляет собой определенный *системный оператор* (алгоритм) преобразования входного сигнала – *воздействия*  $s(t)$ , в сигнал на выходе системы  $y(t)$  – *отклик* или *выходную реакцию* системы. Символическое обозначение операции преобразования (трансформации):

$$y(t) = T[s(t)].$$

*Системный оператор*  $T$  - это правило (набор правил, алгоритм) преобразования сигнала  $s(t)$  в сигнал  $y(t)$ . Для общеизвестных операций преобразования сигналов применяются также расширенные символы операторов трансформации, где вторым символом и специальными индексами обозначается конкретный вид операции (как, например,  $TF$  - преобразование Фурье,  $TF^{-1}$  - обратное преобразование Фурье).

Входной сигнал системы может представлять собой  $m$  - мерный вектор ( $m$  входных сигналов), а выходной сигнал  $n$  - мерный вектор, при этом система будет иметь  $m$  входов и  $n$  выходов. Пример такой системы в геофизике: трехканальный гамма-спектрометр, на три входа решающего блока которого поступают сигналы от калиевого, радиевого и ториевого каналов спектрометра, а на три выхода выводятся сигналы содержащий калия, урана и тория, при этом системный оператор реализует алгоритм решения системы трех линейных уравнений с тремя неизвестными.

Для детерминированных входных сигналов соотношение между выходными и входными сигналами однозначно задается

системным оператором. В случае реализации на входе системы случайного входного процесса также существует однозначное соответствие процессов на выходе и входе системы, однако при этом одновременно происходит изменение статистических характеристик выходного сигнала (математического ожидания, дисперсии, корреляционной функции и пр.), которое также определяется системным оператором.

Для определения системы необходимо задать характер, тип и области допустимых величин входных и выходных сигналов. Как правило, системы выполняются на сигналы одного типа по входу/выходу и подразделяются на системы непрерывного времени (аналоговые или дискретные сигналы на входе и выходе) и цифровые системы. Совокупность системного оператора  $T$  и пространства сигналов образует математическую модель системы.

**Линейные системы.** Любые преобразования сигналов сопровождаются изменением их спектра и по характеру этих изменений разделяются на два вида: линейные и нелинейные. К нелинейным относят изменения, при которых в составе спектра сигналов появляются новые гармонические составляющие. При линейных изменениях сигналов изменяются амплитуды и/или начальные фазы гармонических составляющих спектра. Оба вида изменений могут происходить как с сохранением полезной информации, так и с ее искажением. Это зависит не только от характера изменения спектра сигналов, но и от спектрального состава самой полезной информации.

Линейные системы составляют основной класс систем обработки сигналов. Термин линейности означает, что система преобразования сигналов должна иметь произвольную, но в обязательном порядке линейную связь между входным сигналом (возбуждением) и выходным сигналом (откликом). В нелинейных системах связь между входным и выходным сигналом определяется произвольным нелинейным законом.

Система считается линейной, если в пределах установленной области входных и выходных сигналов ее реакция на входные сигналы аддитивна (выполняется принцип суперпозиции сигналов) и однородна (выполняется принцип пропорционального подобия).

Принцип *аддитивности* требует, чтобы реакция на сумму двух входных сигналов была равна сумме реакций на каждый сигнал в отдельности:

$$T[a(t)+b(t)] = T[a(t)]+T[b(t)].$$

Принцип *однородности* или пропорционального подобия требует сохранения однозначности масштаба преобразования при



любой амплитуде входного сигнала:

$$T[c \times a(t)] = c \times T[a(t)].$$

Другими словами, отклик линейной системы на взвешенную сумму входных сигналов должен быть равен взвешенной сумме откликов на отдельные входные сигналы независимо от их количества и для любых весовых коэффициентов, в том числе комплексных.

**Примеры.**

1. Система  $y(t) = a^2t$ .  $y(t_1) = a^2t_1$ ,  $y(t_2) = a^2t_2$ ,  $y(ct) = a^2ct$ .

$y(t_1+t_2) = a^2(t_1+t_2) = a^2t_1+a^2t_2 = y(t_1)+y(t_2)$ . Система аддитивна.

$cy(t) = ca^2t = a^2ct = y(ct)$ . Система однородна. И в целом линейна.

2. Система  $y(t) = at^2$ .  $y(t_1) = at_1^2$ ,  $y(t_2) = at_2^2$ ,  $y(ct) = a(ct)^2 = ac^2t^2$ .

$y(t_1+t_2) = a(t_1+t_2)^2 \neq y(t_1)+y(t_2) = at_1^2+at_2^2$ . Система не аддитивна.

$cy(t) = c at^2 \neq y(ct) = ac^2t^2$ . Система неоднородна. И в целом нелинейна.

При программной реализации линейных систем на ЭВМ особых затруднений с обеспечением линейности в разумных пределах значений входных и выходных сигналов, как правило, не возникает. При физической (аппаратной) реализации систем обработки данных диапазон входных и/или выходных сигналов, в котором обеспечивается линейность преобразования сигналов, всегда ограничен и должен быть специально оговорен в технической документации или методической инструкции.

**Основные системные операции.** К базовым линейным операциям, из которых могут быть сформированы любые линейные операторы преобразования, относятся операции скалярного умножения, сдвига и сложения сигналов:

$$y(t) = b \times x(t),$$

$$y(t) = x(t-\Delta t),$$

$$y(t) = a(t)+b(t).$$

Графическое отображение операций (цифровая форма) приведено на рис. 21.1.

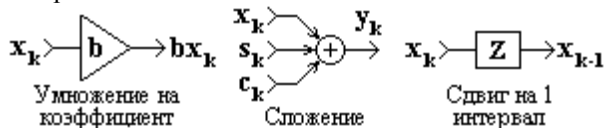


Рис. 21.1. Графика системных операций

Отметим, что операции сложения и умножения являются линейными только для аналоговых и дискретных сигналов. В случае

цифровых сигналов они линейны относительно самих цифровых сигналов, но если последние - результат операции амплитудно-цифрового преобразования, то сложение и умножение не может считаться линейным абсолютно точно по отношению к исходным сигналам.

Для систем, с размерностью 2 и более существует также еще одна базовая операция, которая называется операцией *пространственного маскирования*, которая может рассматриваться как обобщение скалярного умножения. Так, для двумерных систем:

$$z(x,y) = c(x,y) \cdot u(x,y),$$

где  $u(x,y)$  – двумерный входной сигнал,  $c(x,y)$  – пространственная маска постоянных (весовых) коэффициентов. Пространственное маскирование представляет собой поэлементное произведение значений сигнала с коэффициентами маски.

**Инвариантность систем к сдвигу.** Система называется инвариантной к сдвигу (инвариантной во времени, а равно и по любым другим аргументам), если сдвиг входного сигнала по аргументам вызывает соответствующий сдвиг выходного сигнала:

$$s(x,t) = T[a(x,t)], \quad T[a(x-\Delta x, t-\Delta t)] = s(x-\Delta x, t-\Delta t).$$

Линейность и инвариантность к сдвигу являются независимыми свойствами систем и не определяют друг друга. Так, например, операция квадратурования сигнала (возведения в квадрат всех значений сигнала) инвариантна к сдвигу, но нелинейна.

В теории анализа и обработки данных основное место занимают системы, линейные и инвариантные к сдвигу (ЛИС - системы). Они обладают достаточно широкими практическими возможностями при относительной простоте математического аппарата. В дальнейшем, если это специально не оговаривается, будем иметь в виду именно такие системы.

Преимущество, которое отдается ЛИС - системам в методах обработки информации, базируется на возможности разложения входного сигнала любой, сколь угодно сложной формы, на составляющие простейших форм, отклик системы на которые известен и хорошо изучен, с последующим вычислением выходного сигнала в виде суммы откликов на все составляющие входного сигнала. В качестве простейших форм разложения сигналов используются, как правило, единичные импульсы и гармонические составляющие. Первая применяется при представлении сигнала в динамической форме и использует преобразование свертки, вторая - частотное представление сигнала и преобразование Фурье.

Другой важной особенностью ЛИС - систем является то, что любые их комбинации также являются ЛИС - системами, а любую

сложную ЛИС - систему можно разложить на комбинации простых систем. Так, например, при последовательном (каскадном) соединении систем, когда выходной сигнал одной системы служит входным сигналом для второй и т.д., образуемая система в целом также является ЛИС - системой, если линейны и инвариантны к сдвигу все системы, в нее входящие, при этом по отношению к общей системной операции преобразования порядок соединения входящих в нее систем значения не имеет.

**Математическая модель системы** задает связь физических (программных) элементов, образующих техническое устройство, способное воспринимать внешнее воздействие  $s(t)$  и формировать выходную величину  $y(t)$ , определенным образом зависящую от воздействия  $x(t)$ . В аналитической форме эта связь в аналоговой одномерной линейной системе обычно выражается линейным дифференциальным уравнением:

$$\sum_{m=0}^M a_m \frac{d^m y(t)}{dt^m} = \sum_{n=0}^N b_n \frac{d^n s(t)}{dt^n} . \quad (21.1)$$

При нормировке к  $a_0 = 1$ , отсюда следует:

$$y(t) = \sum_{n=0}^N b_n \frac{d^n s(t)}{dt^n} - \sum_{m=1}^M a_m \frac{d^m y(t)}{dt^m} . \quad (21.2)$$

По существу, правой частью этого выражения в самой общей математической форме отображается содержание операции преобразования входного сигнала, т.е. задается оператор трансформации входного сигнала в выходной. Для однозначного решения уравнений (21.1) должны задаваться начальные условия: значения решения  $y(0)$  и его производных по времени в начальный (нулевой) момент времени. В физической системе эти значения определяются начальной энергией в элементах, способных ее накапливать. Входным воздействием может быть произвольный сигнал. Коэффициенты  $a_m$  и  $b_n$ , которые определяются составом и свойствами системы, должны быть известными.

Как правило, решение уравнения (21.1) складывается из суммы принужденной  $y_{пр}(t)$  и свободной  $y_{св}(t)$  составляющих:

$$y(t) = y_{пр}(t) + y_{св}(t).$$

В устойчивых системах  $y_{св}(t)$  определяется параметрами системы (реакция на "ударное" воздействие – единичный "дельта-импульс" или скачок) и с течением времени затухает. Если входное воздействие достаточно гладкое и не содержит скачков (разрывов первого рода), то по истечении определенного времени затухания  $y_{св}(t)$ ,

которое обычно называют временем переходного процесса, система переходит в установившийся динамический режим, а выходной сигнал  $y(t)$  определяется только значением  $y_{np}(t)$  и линейно связан с входным воздействием  $x(t)$ .

Аналогичная модель в цифровой системы описывается разностными уравнениями:

$$\sum_{m=0}^M a_m y((k-m)\Delta t) = \sum_{n=0}^N b_n s((k-n)\Delta t). \quad (21.3)$$

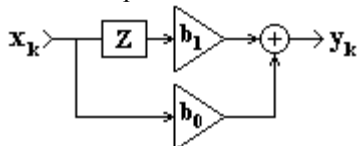
$$y(k\Delta t) = \sum_{n=0}^N b_n s((k-n)\Delta t) - \sum_{m=1}^M a_m y((k-m)\Delta t). \quad (21.4)$$

Последнее уравнение можно рассматривать как алгоритм последовательного вычисления значений  $y(k\Delta t)$ ,  $k = 0, 1, 2, \dots$ , по значениям входного сигнала  $s(k\Delta t)$  и предыдущих вычисленных значений  $y(k\Delta t)$  при известных значениях коэффициентов  $a_m$ ,  $b_n$  и с учетом задания определенных начальных условий - значений  $s(k\Delta t)$  и  $y(k\Delta t)$  при  $k < 0$ . Интервал дискретизации в цифровых последовательностях отсчетов обычно принимается равным 1, т.к. выполняет только роль масштабного множителя.

**Нерекурсивные цифровые системы.** При нулевых значениях коэффициентов  $a_m$  уравнение (21.4) переходит в уравнение дискретной свертки  $x(k)$  с оператором  $b_n$ :

$$y(k) = \sum_{n=0}^N b_n x(k-n). \quad (21.5)$$

При установленных значениях коэффициентов  $b_n$  значения выходных отсчетов свертки для любого аргумента  $k$  определяются текущим и "прошлыми" значениями входных отсчетов. Такая система называется нерекурсивной цифровой системой (НЦС). Пример простейшей НЦС приведен на рис. 2.2.



$$H(z) = b_0 + b_1 z. \quad y_k = b_0 x_k + b_1 x_{k-1}$$

Рис. 21.2. Пример НЦС.

Интервал суммирования по  $n$  получил название "окна" системы. Окно системы (21.5) составляет  $N+1$  точку, система является односторонней каузальной, причинно обусловленной текущими и "прошлыми" значениями входного сигнала, выходной сигнал не

опережает входного. Каузальная система может быть реализована аппаратно в реальном масштабе времени. При  $k < n$  проведение обработки входных данных возможно только при задании определенных начальных условий для точек  $x(-k)$ ,  $k=1,2,\dots,N$ . Как правило, в качестве начальных условий принимаются нулевые значения или значения отсчета  $x(0)$ . Применяется также четное или нечетное продление функции  $x(k)$  на интервал отрицательных значений  $k$ . Если при обработке данных начальные интервалы массивов  $x(k)$  существенного значения не имеют, то обработку можно начинать с отсчета  $k=N$ .

При обработке данных на ЭВМ ограничение по каузальности системного оператора снимается. В программном распоряжении системы могут находиться как "прошлые", так и "будущие" значения входных отсчетов, при этом уравнение (21.5) будет иметь вид:

$$y(k) = \sum_{n=-N'}^N b_n x(k-n). \quad (21.6)$$

При  $N' = N$  система называется двусторонней симметричной. Симметричные системы, в отличие от каузальных, не изменяют фазы обрабатываемых сигналов.

Техника выполнения свертки в координатной области не отличается от техники выполнения обычной дискретной свертки двух массивов данных.

Представим, что на одной полоске бумаги выписаны по порядку сверху вниз значения данных  $x(k)$ . На второй полоске бумаги находятся записанные в обратном порядке значения коэффициентов системы  $b_n$ . Для вычисления  $y(k)$  располагаем вторую полоску против первой таким образом, чтобы значение  $b_0$  совпало со значением  $x(k)$ , перемножаем все значения  $b_n$  с расположенными против них значениями  $x(k-n)$  и суммируем результаты перемножения. Результат суммирования является выходным значением сигнала  $y(k)$ . Сдвигаем окно системы - полоску коэффициентов  $b_k$  на один отсчет последовательности  $x(k)$  вниз (по порядку возрастания номеров  $k$ ) или массив  $x(k)$  сдвигаем на отсчет вверх и вычисляем аналогично следующее значение, и т.д.

Описанный процесс свертки в вещественной области массива данных  $x(k)$  с нерекурсивным оператором системы  $b_n$  (массивом весовых коэффициентов системы) обычно называют нерекурсивной цифровой фильтрацией данных, а саму систему, если она выполняет только данную операцию, нерекурсивным цифровым фильтром (НЦФ).

**Пример.** Сглаживание данных скользящим симметричным П-окном размером 5 отсчетов.

Уравнение НЦФ:  $y_k = \sum_{n=-2}^2 b_n x_{k-n}$ ,  $b_n = 0,2$ . Начальные условия - нулевые.

Входной сигнал:  $x_k = \{0,0,0,0,10,0,10,0,0,0,0\}$ .

Выходной сигнал:  $y_k = \{0,0,2,2,4, 4, 4,2,2,0,0\}$ .

Проверьте результат (выполните фильтрацию сигнала).

*Заметим: сумма коэффициентов  $b_n$  сглаживающего НЦФ равна 1, при этом сумма значений выходного сигнала равна сумме значений входного сигнала. Сигнал "размазался" по аргументу.*

---

**Рекурсивные цифровые системы.** Системы, которые описываются полным разностным уравнением (21.3), принято называть рекурсивными цифровыми системами (РЦС) или рекурсивными цифровыми фильтрами (РЦФ), так как в вычислении текущих значений выходного сигнала участвует не только входной сигнал, но и значения выходного сигнала, вычисленные в предшествующих циклах расчетов. С учетом последнего фактора рекурсивные системы называют системами с обратной связью. Пример рекурсивной системы приведен на рис. 21.3.

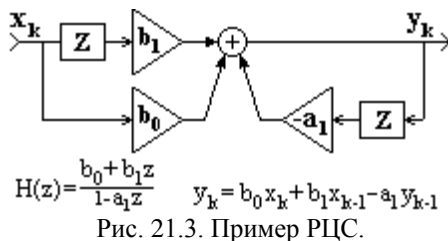


Рис. 21.3. Пример РЦС.

Полное окно рекурсивной системы состоит из двух составляющих: нерекурсивной части  $b_n$ , аналогичной окну нерекурсивной системы и ограниченной в работе текущими и "прошлыми" значениями входного сигнала (при реализации на ЭВМ возможно использование и "будущих" отсчетов сигнала), и рекурсивной части  $a_n$ , которая работает только с "прошлыми", ранее вычисленными значениями выходного сигнала. Техника вычислений для РЦС приведена на рис. 21.4.

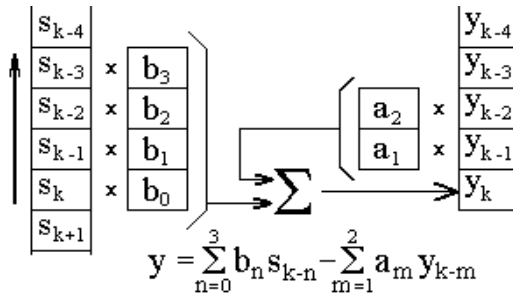


Рис. 21.4. Техника вычислений в РЦС.

**Пример.** Уравнение РЦС:  $y_k = b_0 x_k + a_1 y_{k-1}$ , при  $b_0 = a_1 = 0.5$ ,  $y_{-1} = 0$ .  
Входной сигнал:  $x_k = \{0, 10, 0, 0, 0, \dots\}$ .

Расчет выходного сигнала:

$$\begin{aligned} y_0 &= 0,5x_0 + 0,5y_{-1} = 0 & y_1 &= 0,5x_1 + 0,5y_0 = 5 \\ y_2 &= 0,5x_2 + 0,5y_1 = 2,5 & y_3 &= 0,5x_3 + 0,5y_2 = 1,25 \\ y_4 &= 0,5x_4 + 0,5y_3 = 0,625 & y_5 &= 0,5x_5 + 0,5y_4 = 0,3125 \text{ и т.д.} \end{aligned}$$

Выходной сигнал:  $y_k = \{0, 5, 2,5, 1,25, 0,625, 0,3125, 0,15625, \dots\}$

Из примера можно видеть, что реакция РЦС на конечный входной сигнал, в принципе, может иметь бесконечную длительность, в отличие от реакции НЦС, которая всегда ограничена количеством членов  $b_k$  (окном системы).

**Пример.** Уравнение РЦС:  $y_k = b_0 x_k + a_1 y_{k-1} + a_2 y_{k-2}$ , при  $b_0 = a_1 = a_2 = 1$ .  
 $x_k = \{0, 1, 0, 0, 0, \dots\}$ ,  $y_{-1} = y_{-2} = 0$ .

Расчет выходного сигнала:

$$\begin{aligned} y_0 &= x_0 + y_{-1} + y_{-2} = 0. & y_1 &= x_1 + y_0 + y_{-1} = 1. & y_2 &= x_2 + y_1 + y_0 = 1. \\ y_3 &= x_3 + y_2 + y_1 = 2. & y_4 &= x_4 + y_3 + y_2 = 3. & y_5 &= x_5 + y_4 + y_3 = 5. \\ y_6 &= x_6 + y_5 + y_4 = 8. & y_7 &= x_7 + y_6 + y_5 = 13. & & \text{и т.д.} \end{aligned}$$

**Заметим:** сумма коэффициентов обратной связи  $a_i$  больше 1 и выходной сигнал идет "в разнос".

**Пример.** Интегрирование данных по формуле трапеций. Уравнение РЦФ:  $y_k = (x_k + x_{k-1})/2 + y_{k-1}$ , начальные условия - нулевые. Входной сигнал:  $x_k = \{0, 0, 2, 2, 8, 0, 4, 4, 0, \dots\}$ . Выполните фильтрацию. Контроль:  $y_k = \{0, 0, 1, 3, 8, 12, 14, 18, 20, 20, \dots\}$ .

**Стационарные и нестационарные системы.** Система считается *стационарной* и имеет постоянные параметры, если ее свойства (математический алгоритм оператора преобразования) в пределах заданной точности не зависят от входного и выходного сигналов и не изменяются ни во времени, ни от каких-либо других

внешних факторов. Математически это означает задание системы уравнениями типа (21.3) с постоянными значениями коэффициентов  $a_j$  и  $b_i$  и реакция системы на какое-либо воздействие не зависит от времени (координат) его приложения. В противном случае система является нестационарной или *параметрической* (системой с переменными параметрами). Среди последних большое значение имеют так называемые адаптивные системы обработки данных. В этих системах производится, например, оценивание определенных параметров входных и выходных сигналов, по результатам сравнения которых осуществляется подстройка параметров преобразования (переходной характеристики системы) таким образом, чтобы обеспечить оптимальные по производительности условия обработки сигналов или минимизировать погрешность обработки.

### **21.3. Импульсная характеристика системы**

**Импульсный отклик системы.** По определению, импульсными характеристиками систем (второй широко используемый термин - импульсный отклик систем) называются функции  $h(t)$  для аналоговых и  $h(k\Delta t)$  для цифровых систем, которые являются реакцией (откликом) систем на единичные входные сигналы: дельта-функцию  $\delta(t)$  для аналоговых и импульс Кронекера  $\delta(k\Delta t)$  для цифровых систем, поступающие на вход систем соответственно при  $t=0$  и  $k=0$ . Эта реакция однозначно определяется оператором преобразования:

$$y(t) = T[\delta(t)] \equiv h(t). \quad (21.7)$$

$$y(k\Delta t) = T[\delta(k\Delta t)] \equiv h(k\Delta t). \quad (21.8)$$

Импульсный отклик аналоговой системы, как результат операции над дельта-функцией, в определенной мере представляет собой математическую абстракцию идеального преобразования. С практической точки зрения под импульсным откликом аналоговой системы можно понимать математическое отображение реакции системы на входной сигнал произвольной формы с площадью, равной 1, если длительность сигнала пренебрежимо мала по сравнению со временем реакции системы или с периодом ее собственных колебаний. Под временем (длиной) реакции системы обычно понимают интервал, на котором значения функции  $h(t)$  существенно отличаются от нуля после прекращения действия единичного сигнала на ее входе.

Для цифровых систем импульсный отклик однозначно определяется реакцией системы на импульс Кронекера  $\delta(k\Delta t)=1$  при  $k=0$ . Функцию импульсного отклика называют также *весовой функцией* системы.



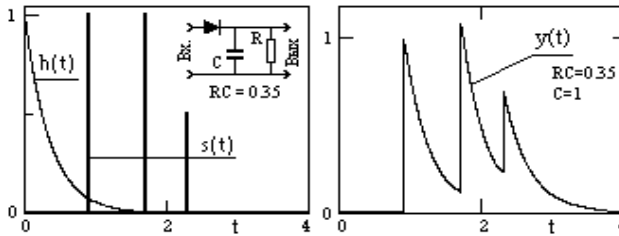


Рис. 21.5. Импульсный отклик системы  $h(t)$ ,

входной сигнал  $s(t)$  и выходная реакция системы  $y(t)$ .

На рисунке 21.5 приведен пример импульсного отклика  $h(t)$  интегрирующей RC-цепи. При подаче на вход RC-цепи импульса заряда  $\Delta q$  емкость  $C$  заряжается до напряжения  $V_0 = \Delta q/C$  и начинает разряжаться через сопротивление  $R$ , при этом напряжение на ней изменяется по закону  $v(t) = V_0 \cdot e^{-t/RC} = (\Delta q/C) \cdot e^{-t/RC}$ . Отсюда, отклик RC-цепи по выходному напряжению на входной сигнал  $\Delta q=1$ :  $h(t) = (1/C) \cdot e^{-t/RC}$ . По существу, импульсным откликом системы  $h(t)$  определяется доля входного сигнала, которая действует на выходе системы по истечении времени  $t$  после поступления сигнала на вход (запаздывающая реакция системы).

**Реакция системы на произвольный сигнал.** Если функция импульсного отклика системы известна, то, с учетом принципа суперпозиции сигналов в линейной системе, можно выполнить расчет реакции системы в любой произвольный момент времени на любое количество входных сигналов с любыми моментами времени их прихода путем суммирования запаздывающих реакций системы на эти входные сигналы, как это показано на рис. 21.5 для трех входных импульсов. В общем случае произвольный сигнал на входе системы может быть разложен в линейную последовательность взвешенных единичных импульсов:

$$y(t) = T[s(t)] \equiv T \left[ \int_{-\infty}^{\infty} s(\tau) \delta(t-\tau) d\tau \right]. \quad (21.9)$$

На основании принципа суперпозиции линейный оператор  $T$  может быть внесен под знак интеграла, т.к. последний представляет собой предельное значение суммы. При этом операция преобразования действует только по переменной  $t$ :

$$y(t) = \int_{-\infty}^{\infty} s(\tau) T[\delta(t-\tau)] d\tau = \int_{-\infty}^{\infty} s(\tau) h(t-\tau) d\tau. \quad (21.10)$$

Это выражение представляет собой интеграл Дюамеля или свертку (конволюцию) входного сигнала с импульсной характеристикой системы. Заменой переменных  $t-\tau = \tau$  можно

убедиться в том, что эта операция, как и положено свертке, коммутативна:

$$\int_{-\infty}^{\infty} s(\tau) h(t-\tau) d\tau \equiv \int_{-\infty}^{\infty} h(\tau) s(t-\tau) d\tau.$$

Аналогично, для дискретных сигналов:

$$y(k\Delta t) = \sum_{n=-\infty}^{\infty} s(n\Delta t) h(k\Delta t-n\Delta t) \equiv \sum_{n=-\infty}^{\infty} h(n\Delta t) s(k\Delta t-n\Delta t). \quad (21.11)$$

В символической форме математического представления:

$$y(t) = s(t) \otimes h(t-\tau) \equiv s(t-\tau) \otimes h(\tau) \equiv s(t) \otimes h(t).$$

В реальных физических системах импульсный отклик  $h(t)$  равен нулю при  $t < 0$  (реакция на выходе системы не может опережать входной сигнал) и, как правило, отличен от нуля только на определенном интервале  $\tau$ , по которому и ведется интегрирование или суммирование в выражениях свертки. При обработке данных на ЭВМ требований по односторонности импульсного отклика не предъявляется, равно как и по его размерам вперед и назад от нуля по координатам.

### **Усиление постоянной составляющей сигнала.**

Подадим на вход системы постоянный сигнал  $s(t) = A$ . При этом сигнал на выходе системы:

$$y(t) = \int_{-\infty}^{\infty} h(\tau) s(t-\tau) d\tau = A \int_{-\infty}^{\infty} h(\tau) d\tau = A \cdot K_{пс}, \quad (21.12)$$

т.е. площадь импульсного отклика (для цифровой системы соответственно сумма коэффициентов импульсного отклика) является коэффициентом  $K_{пс}$  усиления постоянной составляющей входного сигнала. Если при обработке сигналов должны изменяться только динамические характеристики их формы без изменения постоянной составляющей, а равно и различных постоянных уровней (фона, пьедесталов, региональных трендов и т.п.), то площадь импульсного отклика (сумма коэффициентов) должна нормироваться к единице.

На рис. 21.6 приведен пример выполнения свертки рассмотренной нами выше RC-цепью при нормированной к 1 площади импульсного отклика  $h(\tau)$ . Входной сигнал  $s(t)$  находится на постоянном фоновом значении, в данном случае - нулевом, при этом, как и следовало ожидать, площадь выходного сигнала  $y(t)$  равна площади входного сигнала.

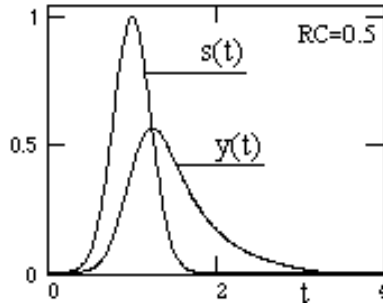


Рис. 21.6.

**Усиление шумов.** Критерием качества системы при использовании любого метода обработки информации можно считать выполнение целевого назначения с минимальным усилением шумов (максимальным их подавлением). Допустим, что система имеет нормированный к 1 импульсный отклик  $h(k)$ . Обозначим через  $\varepsilon(k)$  аддитивный шум с математическим ожиданием  $M\{\varepsilon(k)\} = \bar{\varepsilon} = 0$  и дисперсией  $D_\varepsilon = \sigma^2$ , который в сумме с сигналом поступает на вход системы. Значения  $\varepsilon(k)$  статистически независимы и некоррелированы с сигналом. С учетом помехи во входном сигнале значение сигнала на выходе системы:

$$y(k) = \sum_n h(n)[x(k-n) + \varepsilon(k-n)].$$

Математическое ожидание значений выходного сигнала:

$$M\{y(k)\} = \sum_n h(n)[x(k-n) + M\{\varepsilon(k-n)\}] = \sum_n h(n) x(k-n).$$

Вычислим дисперсию распределения отсчетов выходного сигнала:

$$D\{y(k)\} = M\{[\sum_n h(n)[x(k-n) + \varepsilon(k-n)] - M\{y(k)\}]^2\} = M\{[\sum_n h(n) \varepsilon(k-n)]^2\}.$$

Если правую часть последнего выражения представить в виде

$$M\{[\sum_n h(n) \varepsilon(k-n)] \cdot [\sum_m h(m) \varepsilon(k-m)]\},$$

то в этом выражении математические ожидания всех членов произведения с сомножителями  $\varepsilon(n)\varepsilon(m)$  при  $n \neq m$  равны 0 в силу статистической независимости значений шума. Остаются только члены с  $n = m$ , т.е.:

$$M\{\sum_n h^2(n) \varepsilon^2(n)\} = \sum_n h^2(n) M\{\varepsilon^2(n)\} = D_\varepsilon \sum_n h^2(n) = \sigma^2 \sum_n h^2(n). \quad (21.13)$$

Отсюда следует, что сумма квадратов значений нормированного импульсного отклика системы представляет собой

коэффициент усиления аддитивных шумов во входном сигнале.

---

**Пример.** Сглаживающий фильтр:  $y(k) = 0.2 \sum_{n=-2}^2 x(k-n)$ .

Коэффициент усиления шумов:  $5 (0.2^2) = 0.2$ . Дисперсия шумов уменьшается в  $1/0.2 = 5$  раз.

---

Для систем с  $m$  входами и  $n$  выходами аналогично определяются парциальные импульсные отклики  $h_{ij}(t)$ ,  $i = \{1, 2, \dots, n\}$ ,  $j = \{1, 2, \dots, m\}$ , каждым из которых отображается сигнал на  $i$ -м выходе при поступлении сигнала  $\delta(t)$  на  $j$ -й вход. Полная совокупность импульсных откликов образует матрицу:

$$\bar{\mathbf{h}}(t) = \begin{vmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,m} \\ h_{2,1} & h_{2,2} & \dots & h_{2,m} \\ \cdot & \cdot & \cdot & \cdot \\ h_{n,1} & h_{n,2} & \dots & h_{n,m} \end{vmatrix},$$

а выражение свертки приобретает вид:

$$\bar{\mathbf{U}}_{\text{вых}}(t) = \int_{-\infty}^{\infty} \bar{\mathbf{h}}(t-\tau) \bar{\mathbf{u}}_{\text{вх}}(\tau) d\tau.$$

**Определение импульсной реакции** требуется, как правило, для рекурсивных систем, так как импульсная реакция для НЦС специального определения не требует:

$$h(k) = \sum_{n=-N}^N b(n)\delta(k-n) \equiv b(k).$$

Если выражение для системы известно в общей форме (21.3), определение импульсной реакции производится подстановкой в уравнение системы импульса Кронекера с координатой  $k = 0$  при нулевых начальных условиях, при этом сигнал на выходе системы будет представлять собой импульсную реакцию системы:  $y(k) \equiv h(k)$ .

---

**Пример.** Уравнение РЦС:  $y_k = x_k + 0.5y_{k-1}$ . Входной сигнал:  $x_k = \delta_0 = \{1, 0, 0, 0, \dots\}$ .

Расчет выходного сигнала при нулевых начальных условиях:

$$\begin{aligned} y_0 &= x_0 + 0.5 y_{-1} = 1 + 0 = 1 = h_0. & y_1 &= x_1 + 0.5 y_0 = 0 + 0.5 = 0.5 = h_1. \\ y_2 &= x_2 + 0.5 y_1 = 0 + 0.25 = 0.25 = h_2. & y_3 &= x_3 + 0.5 y_2 = 0 + 0.125 = h_3. \\ y_4 &= x_4 + 0.5 y_3 = 0 + 0.0625 = h_4, & & \text{т.д.} \end{aligned}$$

Импульсный отклик системы:

$$h_k = \{1, 0.5, 0.25, 0.125, \dots\} \equiv (0.5)^k, \quad k = 0, 1, 2, \dots$$

---

Определение импульсной реакции физических систем обычно производится подачей на вход систем ступенчатой функции  $u_0(k) = 1$  при  $k \geq 0$ , и  $u_0(k) = 0$  при  $k < 0$ :

$$g(k) = \sum_{n=0}^N h(n) u_0(k-n) = \sum_{n=0}^N h(n).$$

$$h(k) = g(k) - g(k-1), \quad k=0,1,2,\dots$$

Функция  $g(k)$  получила название переходной характеристики системы (перехода из одного статического состояния в другое).

## **21.4. Передаточные функции цифровых систем**

**Z-преобразование** является удобным методом решения разностных уравнений линейных систем. Применяя z-преобразование к обеим частям равенства (21.3), получаем:

$$Y(z) \sum_{m=0}^M a_m z^m = X(z) \sum_{n=0}^N b_n z^n, \quad (21.14)$$

где  $X(z), Y(z)$  - соответствующие z-образы входного и выходного сигнала. Из этого выражения, полагая  $a_0 = 1$ , получаем в общей форме функцию связи входа и выхода системы - уравнение *передаточной функции* системы (или *системной функции*) в z-области:

$$H(z) = Y(z)/X(z) = \sum_{n=0}^N b_n z^n / (1 + \sum_{m=1}^M a_m z^m). \quad (21.15)$$

Для нерекурсивных систем при  $a_m = 0$ :

$$H(z) = \sum_{n=0}^N b_n z^n. \quad (21.16)$$

При подаче на вход системы единичного импульса Кронекера  $\delta_0$ , имеющего z-образ  $\delta(z) = z^0 = 1$ , сигнал на выходе системы будет представлять собой импульсную реакцию системы  $y(k) = h(k)$ , при этом:

$$H(z) \equiv Y(z) = Y(z)/\delta(z) = TZ[y(k)] = \sum_{k=0}^{\infty} h(k)z^k, \quad (21.17)$$

т.е. передаточная функция системы является z-образом ее импульсной реакции. При обратном z-преобразовании передаточной функции соответственно получаем импульсную характеристику системы:

$$h(k) \Leftrightarrow H(z). \quad (21.18)$$

Если функция  $H(z)$  представлена конечным степенным полиномом, как это обычно имеет место для НЦС, то обратное z-преобразование элементарно. Передаточная функция РЦС также может быть представлена степенным полиномом прямым делением числителя

на знаменатель правой части выражения (21.15), однако результат при этом может оказаться как конечным, так и бесконечным, т.е. система может иметь либо конечную, либо бесконечную импульсную характеристику. Системы с бесконечной импульсной характеристикой получили название БИХ-систем, с конечной импульсной характеристикой соответственно КИХ-систем. Нерекурсивные системы всегда имеют конечную импульсную характеристику, т.к. длительность импульсной реакции НЦС определяется окном фильтра.

**Пример.** Передаточная функция рекурсивной системы:  
 $H(z) = (1-z^5)/(1-z)$ .

Делением числителя на знаменатель получаем:  
 $H(z) = 1+z+z^2+z^3+z^4$ .  $H(z) \Leftrightarrow h(n) = \{1,1,1,1,1\}$ .

Система имеет конечную импульсную характеристику.

**Пример.** Передаточная функция:  $H(z)=1/(1-2z)$ . Методом обратного преобразования:  $h(n) = 2^n$ .

Система имеет бесконечную импульсную характеристику.

---

**Устойчивость систем.** Любая практическая система должна быть *устойчивой*, т.е. для сигналов, конечных по энергии или средней мощности, выходные сигналы также должны быть конечными по этим параметрам. Система называется устойчивой, если при любых начальных условиях реакция системы на любое ограниченное воздействие также ограничена.

Для конечного по энергии входного сигнала, можно записать:

$$|y(t)| \leq \int_{-\infty}^{\infty} |h(\tau)||x(t-\tau)| d\tau = A \int_{-\infty}^{\infty} |h(\tau)| d\tau.$$

Отсюда следует условие, при котором выходной сигнал также будет ограниченным:

$$\int_{-\infty}^{\infty} |h(\tau)| d\tau < \infty, \quad (21.19)$$

т.е. необходимым и достаточным условием устойчивости системы является абсолютная сходимость ее импульсной характеристики, или, для цифровых систем, абсолютная суммируемость импульсного отклика:

$$\sum_n |h(n)| < \infty. \quad (21.20)$$

Анализ устойчивости может быть проведен по передаточной функции. В устойчивой системе значение  $H(z)$  должно быть конечным во всех точках  $z$ -плоскости, где  $|z| \leq 1$ , а, следовательно, передаточная функция не должна иметь особых точек (полосов) при  $z \leq 1$  (на и внутри единичного круга на  $z$ -плоскости). Полосы определяются

корнями многочлена знаменателя передаточной функции  $H(z)$ .

**Пример.**

Передаточная функция РЦФ:  $H(z)=1/(1+1.25z)$ .

Корни знаменателя:  $z = -0,8$ .  $|z| < 1$ .

Фильтр неустойчив.

Передаточная функция:  $H(z) = (1-z^3)/(1-0.6z+0.25z^2)$ .

Корни:  $z_{1,2} = 1.2+1.6j$ .  $|z| = 2 > 1$ .

Фильтр устойчив.

Приведенный критерий устойчивости относится к несократимой дроби, т.к. в противном случае возможна компенсация полюса нулем передаточной функции и следует проверить наличие однозначных нулей и полюсов.

Проверка на устойчивость требуется только для рекурсивных цифровых фильтров (систем с обратной связью), нерекурсивные системы всегда устойчивы.

## 21.5. Частотные характеристики систем

**Передаточная функция.** Для линейных систем, принимая

в качестве сигнала на входе системы собственную функцию  $x(k\Delta t) = B(\omega)\exp(j\omega k\Delta t)$ , мы вправе ожидать на выходе системы сигнал  $y(k\Delta t) = A(\omega)\exp(j\omega k\Delta t)$ . Подставляя эти выражения в разностное уравнение системы (21.3), получаем:

$$\sum_{m=0}^M a_m A(\omega)\exp(j\omega k\Delta t - j\omega m\Delta t) = \sum_{n=0}^N b_n B(\omega)\exp(j\omega k\Delta t - j\omega n\Delta t).$$

$$A(\omega)\exp(j\omega k\Delta t) \sum_{m=0}^M a_m \exp(-j\omega m\Delta t) = B(\omega)\exp(j\omega k\Delta t) \sum_{n=0}^N b_n \exp(-j\omega n\Delta t).$$

$$A(\omega) \sum_{m=0}^M a_m \exp(-j\omega m\Delta t) = B(\omega) \sum_{n=0}^N b_n \exp(-j\omega n\Delta t). \quad (21.21)$$

Отсюда, частотная передаточная функция системы (частотная характеристика при нормировке к  $a_0=1$ ):

$$H(\omega) = A(\omega)/B(\omega) = \sum_{n=0}^N b_n \exp(-j\omega n\Delta t) / [1 + \sum_{m=1}^M a_m \exp(-j\omega m\Delta t)]. \quad (21.22)$$

Нетрудно убедиться, что подстановкой  $z=\exp(-j\omega\Delta t)$  в выражение передаточной функции  $H(z)$  (21.15) может быть получено абсолютно такое же выражение для частотной характеристики, т.е.:

$$H(\omega) \equiv H(z) \text{ при } z = \exp(-j\omega\Delta t).$$

При обратном преобразовании  $H(z)$  во временную область с использованием выражений (21.17-18) отсюда следует также, что

частотная характеристика системы представляет собой Фурье-образ ее импульсной реакции, и наоборот. При  $\Delta t = 1$ :

$$H(\omega) = \sum_{n=-\infty}^{\infty} h(n) \exp(-j\omega n), \quad (21.23)$$

$$h(n) = (1/2\pi) \int_{-\pi}^{\pi} H(\omega) \exp(j\omega n) d\omega. \quad (21.24)$$

В общем случае  $H(\omega)$  является комплексной функцией, модуль которой  $R(\omega)$  называется амплитудно-частотной характеристикой системы (АЧХ), а аргумент  $\varphi(\omega)$  - фазочастотной характеристикой (ФЧХ).

$$A(\omega) = |H(\omega)| = \sqrt{\text{Re}^2 H(\omega) + \text{Im}^2 H(\omega)}.$$

$$\varphi(\omega) = \text{arctg}(\text{Im } H(\omega)/\text{Re } H(\omega)).$$

Физический смысл частотной характеристики системы достаточно прост. Произвольный сигнал на входе системы может рассматриваться в виде суммы гармонических составляющих с различным набором амплитуд и начальных фазовых углов. Амплитудно-частотной характеристикой системы устанавливаются коэффициенты усиления системой (коэффициенты передачи) этих частотных составляющих, а фазочастотной характеристикой - сдвиг фаз этих частотных составляющих в выходном сигнале относительно начальных фаз во входном сигнале.

**Основные свойства** частотных характеристик систем:

1. Частотные характеристики являются непрерывными функциями частоты.

2. При дискретизации данных по интервалам  $\Delta t$  функция  $H(\omega)$  является периодической. Период функции  $H(\omega)$  равен частоте дискретизации входных данных  $F=1/\Delta t$ . Первый низкочастотный период (по аргументу  $\omega$  от  $-\pi/\Delta t$  до  $\pi/\Delta t$ , по  $f$  от  $-1/2\Delta t$  до  $1/2\Delta t$ ) называется главным частотным диапазоном передачи сигнала. Граничные частоты главного частотного диапазона соответствуют частоте Найквиста  $\pm\omega_N$ ,  $\omega_N = \pi/\Delta t$ . Частота Найквиста определяет предел частотной разрешающей способности системы по обработке данных.

3. Для систем с вещественными коэффициентами импульсной реакции  $h(n\Delta t)$  функция АЧХ является четной, а функция ФЧХ - нечетной. С учетом этого частотные характеристики систем обычно задаются только на интервале положительных частот  $0-\omega_N$  главного частотного диапазона. Значения функций на интервале отрицательных частот являются комплексно сопряженными со значениями на



интервале положительных частот.

## **21.6. Реакция систем на случайные сигналы**

Если сигнал на входе линейной системы является детерминированным, то, при известных параметрах системы, его соотношение с выходным сигналом является однозначным. Таким же однозначным является соотношение процессов на входе и выходе и для случайных сигналов, однако в силу природы последних явное представление, как входного сигнала, так и отклика системы, не представляется возможным. Для описания отклика системы необходимо использовать статистический подход. При рассмотрении материала данного раздела ограничимся только физически реализуемыми системами с односторонним импульсным откликом  $h(t)$  ( $h(t)=0$  при  $t<0$ ) и соответствующей частотной характеристикой  $H(f)$ . Если параметры входного сигнала специально не оговариваются, то по умолчанию принимается, что на вход системы поступает реализация случайного стационарного процесса  $x(t)$  с нулевым средним и вызывает сигнал  $y(t)$  на выходе системы.

**Квазидетерминированный сигнал** в какой-то мере позволяет оценить сохранение однозначности преобразования системой случайных сигналов.

Допустим, что система имеет импульсный отклик  $h(t) = \exp(-at)$ ,  $t \geq 0$ . Квазидетерминированный случайный сигнал стационарен, не обладает свойством эргодичности, но может быть описан в явной математической форме. Зададим сигнал на входе системы следующего вида:

$$x(t) = A + \cos(2t+\varphi),$$

где  $A$  и  $\varphi$  - взаимно независимые случайные величины, причем  $\varphi$  равномерно распределена в интервале  $[0, 2\pi]$ . Выходной сигнал определится выражением:

$$\begin{aligned} y(t) &= h(t) * x(t-\tau) \equiv \int_0^{\infty} h(\tau) \cdot x(t-\tau) dt = \\ &= A/3 + [3 \cos(2t+\varphi) + 2 \sin(2t+\varphi)]/13. \end{aligned}$$

Из выражения следует, что выходной сигнал системы также является случайным процессом и содержит те же самые случайные параметры, что и входной сигнал, а, следовательно, для него также могут быть определены статистические характеристики.

**Математическое ожидание** произвольного случайного стационарного сигнала  $x(t)$  на выходе линейной системы определится выражением:

$$\bar{y} = M\{y(t)\} = \int_0^{\infty} M\{x(t-\tau)\} \cdot h(\tau) d\tau = \bar{x} \int_0^{\infty} h(\tau) d\tau = \bar{x} K_{\text{пс}}. \quad (21.25)$$

Отсюда следует, что математическое ожидание выходных сигналов системы равно математическому ожиданию входных сигналов, умноженному на коэффициент усиления системой постоянной составляющей. Если система не пропускает постоянную составляющую сигналов (площадь или сумма коэффициентов импульсного отклика системы равна нулю), то случайный выходной сигнал всегда будет иметь нулевое математическое ожидание.

**Корреляционные соотношения.** Для произведения выходных сигналов  $y(t)$  и  $y(t+\tau)$  линейной системы можно записать:

$$y(t) \cdot y(t+\tau) = \int_0^{\infty} \int_0^{\infty} h(\alpha)h(\beta) x(t-\alpha)x(t+\tau-\beta) d\alpha d\beta.$$

Если взять математические ожидания от обеих частей этого равенства, то, с учетом соотношения в подынтегральном выражении

$$M\{x(t-\alpha) x(t+\tau-\beta)\} = -R_x(t-\alpha-t+\beta) = R_x(\tau+\alpha-\beta),$$

получим:

$$R_y(\tau) = \int_0^{\infty} \int_0^{\infty} h(\alpha)h(\beta) R_x(\tau+\alpha-\beta) d\alpha d\beta \equiv R_x(\tau) * h(\tau+\alpha) * h(\tau-\beta). \quad (21.26)$$

Таким образом, функция автокорреляции (АКФ) выходного сигнала равна АКФ входного сигнала, свернутой дважды, в прямом и обратном направлении, с импульсным откликом системы, что сохраняет четность АКФ выходного сигнала. Для нецентрированных процессов аналогичное заключение действительно и для ковариационных функций.

Заметим, что для свертки импульсных откликов, производя замену  $\tau-\beta = t$ , мы имеем равенство:

$$h(\tau+\alpha) * h(\tau-\beta) = h(t+\alpha+\beta) * h(t) = h(t) * h(t+\gamma) = K_h(t),$$

где  $K_h(t)$  - функция ковариации импульсного отклика системы. Отсюда:

$$R_y(\tau) = R_x(\tau) * K_h(\tau). \quad (21.27)$$

Это означает появление в случайном сигнале на выходе системы определенной корреляционной зависимости, вызванной инерционностью системы, причем радиус корреляции выходного сигнала обратно пропорционален верхней частоте, пропускаемой системой.

Для взаимной корреляционной функции (ВКФ)  $R_{xy}$  входного и выходного сигналов соответственно имеем:

$$x(t) \cdot y(t+\tau) = \int_0^{\infty} h(\alpha) x(t)y(t+\tau-\alpha) d\alpha.$$

$$R_{xy}(\tau) = \int_0^{\infty} h(\alpha) R_x(\tau-\alpha) d\alpha \equiv R_x(\tau) * h(\tau-\alpha). \quad (21.28)$$

т.е. функция взаимной корреляции входного и выходного сигналов равна свертке АКФ входного сигнала с функцией импульсного отклика системы - фильтрации АКФ сигнала этим же фильтром. Заключение действительно и для функций ковариации.

Другая взаимно корреляционная функция  $R_{yx}$  может быть получена из соотношения:

$$R_{yx}(\tau) = R_{xy}(-\tau) \equiv R_x(\tau) * h(\tau+\alpha). \quad (21.29)$$

Отметим, что для статистически независимых случайных величин при одностороннем импульсном отклике  $h(\tau)=0$  при  $\tau < 0$  функция  $R_{xy}(\tau)$  также является односторонней и равна 0 при  $\tau < 0$ , а функция  $R_{yx}$  соответственно равна 0 при  $\tau > 0$ .

**Спектральные соотношения**, которые характеризуют систему в целом по отношению к преобразованию случайных сигналов, это соотношения частотных плотностей распределения мощности случайных процессов на входе и выходе, которые для краткости обычно называют спектральными плотностями процессов (сигналов) или спектрами мощности.

Применяя преобразование Фурье к выражениям (21.26), для спектра мощности выходного сигнала получаем:

$$W_y(f) = W_x(f) |H(f)|^2. \quad (21.30)$$

Спектр мощности сигнала на выходе системы равен спектру мощности входного сигнала, умноженному на квадрат модуля частотной характеристики фильтра. С учетом четности корреляционных функций спектр мощности выходного сигнала также является четной действительной функцией и не имеет фазовой характеристики процесса.

Аналогично, для взаимного спектра мощности на основе выражений (21.28-29):

$$\begin{aligned} W_{xy}(f) &= W_x(f) H(f). \\ W_{yx}(f) &= W_x(f) H(-f). \end{aligned} \quad (21.31)$$

Взаимный спектр мощности при одностороннем импульсном отклике является комплексным и содержит как амплитудную, так и фазовую характеристику процесса.

Отметим, что с использованием выражения (21.31) можно производить определение частотной характеристики и импульсного отклика системы:

$$H(f) = W_{xy}/W_x \Leftrightarrow h(t).$$

**Дисперсия выходного сигнала** может быть определена с использованием формул (21.26-30):

$$\sigma_y^2 = R_y(0) = \int_{-\infty}^{\infty} W_x(f) |H(f)|^2 df \equiv R_x(0) \int_0^{\infty} h^2(t) dt = \sigma_x^2 \int_0^{\infty} h^2(t) dt, \quad (21.32)$$

что полностью соответствует полученной ранее формуле (21.30) для цифровой системы.

Если сигнал нецентрированный и значение дисперсии входного сигнала неизвестно, то по аналогичным формулам вычисляется сначала *средний квадрат* выходного сигнала или так называемая *средняя мощность сигнала*:

$$\overline{y^2} = \overline{y^2(t)} = R_y(0) \equiv \overline{x^2} \int_0^{\infty} h^2(t) dt \equiv \int_0^{\infty} W_x(f) |H(f)|^2 df, \quad (21.33)$$

Вывод: средняя мощность выходного сигнала равна средней мощности входного сигнала, умноженной на модуль квадрата площади импульсной реакции системы (для цифровых систем - сумму квадратов коэффициентов импульсного отклика). Для центрированных случайных сигналов средняя мощность равна дисперсии сигналов. Для нецентрированных выходных сигналов:

$$\sigma_y^2 = \overline{y^2} - \overline{y}^2 \equiv (\overline{x^2} - \overline{x}^2) \int_0^{\infty} h^2(t) dt. \quad (21.35.)$$

**Функция когерентности** дает оценку точности принятой линейной модели системы. Когерентность входного и выходного сигналов системы оценивается по формуле:

$$\gamma_{xy}^2(f) = |W_{xy}(f)|^2 / [W_x(f) \cdot W_y(f)]. \quad (21.36)$$

Если функции  $W_x(f)$  и  $W_y(f)$  отличны от нуля и не содержат дельта-функций, то для всех частот  $f$  значения функции когерентности заключены в интервале:

$$0 \leq \gamma_{xy}^2(f) \leq 1.$$

Для исключения дельта-функций на нулевой частоте (постоянная составляющая сигнала) определение функции когерентности производится по центрированным сигналам. Для линейных систем с постоянными параметрами функция когерентности равна 1, в чем нетрудно убедиться, если в формулу (21.35) подставить выражения  $W_{xy}$  и  $W_y$ , определенные через  $W_x$  в формулах (21.35-36). Для совершенно не связанных сигналов функция когерентности равна нулю. Промежуточные между 0 и 1 значения могут соответствовать трем ситуациям:

1. Система осуществляет преобразование  $x(t) \Rightarrow y(t)$ , но в измерениях этих сигналов или одного из них присутствует внешний шум. Так, например, в сигналах, зарегистрированных с ограничением по разрядности, появляется шум квантования (округления значений).
2. Система не является строго линейной. Это может

наблюдаться, например, при определенном ограничении по разрядности вычислений в цифровых системах, при накоплении ошибки в рекурсивных системах и т.п.

3. Выходной сигнал  $y(t)$  помимо  $x(t)$  зависит еще от каких-то входных или внутренних системных процессов.

Величина  $1-\gamma_{xy}^2(f)$  задает долю среднего квадрата сигнала  $y(t)$  на частоте  $f$ , не связанную с сигналом  $x(t)$ .

## 21.7. Структурные схемы систем

**Структурные схемы.** Алгоритмы обработки сигналов в системах наглядно отображаются в виде структурных схем. Базовые элементы схем и примеры построения структурных схем приводились ранее на рис. 21.1-3. Как правило, структурные схемы соответствуют программной реализации систем и будут рассматриваться ниже применительно к цифровым системам, но не определяют аппаратной реализации в специальных радиотехнических устройствах, которая может существенно отличаться от программной реализации, особенно для аналоговых систем.

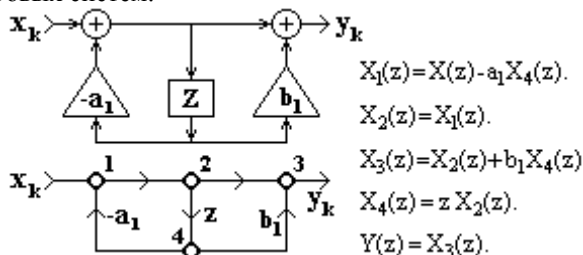


Рис. 21.7. Структурная схема и граф системы.

**Графы систем.** Наряду со структурной схемой система может быть представлена в виде графа, который отображает диаграмму прохождения сигналов и состоит из направленных ветвей и узлов.

Пример структурной схемы системы с передаточной функцией  $H(z) = (1 + b_1 z) / (1 + a_1 z)$  и графа, ей соответствующего, приведен на рисунке 21.7. С каждым  $i$  узлом графа связано значение сигнала  $x_i(k)$  или его образа  $X_i(z)$ , которые определяются суммой всех сигналов или их  $z$ -образов входящих в узел ветвей. В каждой  $ij$ -ветви (из узла  $i$  в узел  $j$ ) происходит преобразование сигнала в соответствии с передаточной функцией ветви, например, задержка сигнала или умножение на коэффициент.

**Соединения систем.** Различают следующие соединения систем (рис. 21.8-10).

1. **Последовательное соединение.** Выходной сигнал

предшествующей системы является входным для последующей. Эквивалентная передаточная функция общей системы равна произведению передаточных функций систем, в нее входящих:

$$H(z) = H_1(z) \cdot H_2(z) \cdot \dots \cdot H_N(z).$$

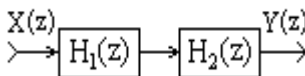


Рис. 21.8.

**2. Параллельное соединение.** Сигнал подается на входы всех параллельно соединенных систем одновременно, выходные сигналы систем суммируются. Эквивалентная передаточная функция общей системы равна сумме передаточных функций систем, в нее входящих:

$$H(z) = H_1(z) + H_2(z) + \dots + H_N(z).$$

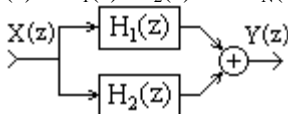


Рис. 21.9.

**3. Соединение обратной связи.** Выходной сигнал первой системы подается на выход системы и одновременно на вход системы обратной связи, выходной сигнал которой суммируется, со знаком плюс или минус в зависимости от вида связи (отрицательной или положительной), с входным сигналом первой системы. Эквивалентная передаточная функция общей системы:  $H(z) = H_1(z)/(1 \pm H_1(z)H_2(z))$ .

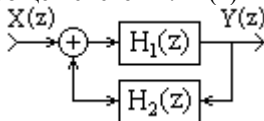


Рис. 21.10

**Схемы реализации систем.** По принципам структурной реализации систем различают следующие схемы:

1. **Прямая форма.** Реализуется непосредственно по разностному уравнению

$$y_k = \sum_{n=0}^N b_n x_{k-n} - \sum_{m=1}^M a_m y_{k-m},$$

или по передаточной функции

$$H(z) = \sum_{n=0}^N b_n z^n / (1 + \sum_{m=1}^M a_m z^m).$$

Пример прямой системы приведен на рис. 21.11.

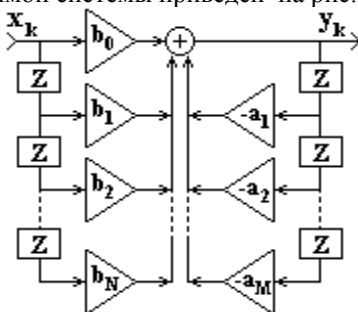


Рис. 21.11.

2. **Прямая каноническая форма** содержит минимальное число элементов задержки. Передаточную функцию РЦС можно представить в виде:

$$H(z) = Y(z)/X(z) = H_1(z)H_2(z),$$

$$H_1(z) = V(z)/X(z) = 1/(1 + \sum_{m=1}^M a_m z^m), \quad H_2(z) = Y(z)/V(z) = \sum_{n=0}^N b_n z^n.$$

Отсюда:

$$v(k) = x(k) - \sum_{m=1}^M a_m v(k-m), \quad (21.37)$$

$$y(k) = \sum_{n=0}^N b_n v(k-n). \quad (21.38)$$

В разностных уравнениях (21.37-38) осуществляется только задержка сигналов  $v(k)$ . Граф реализации РЦС приведен на рисунке 21.12.

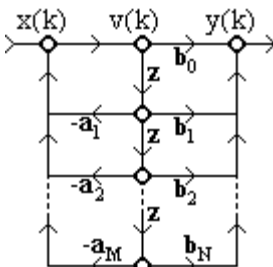


Рис. 21.12.

3. **Каскадная (последовательная) форма.** Соответствует представлению передаточной функции в виде произведения:

$$H(z) = \prod_{i=1}^k H_i(z).$$

$H_i(z)$  - составляющие функции типа  $(1-r_i z)/(1-p_i z)$  при представлении  $H(z)$  в факторизованной форме, где  $r_i$  и  $p_i$  - нули и полюсы функции  $H(z)$ . В качестве функций  $H_i(z)$  обычно используются передаточные функции биквадратных блоков - фильтров второго порядка:

$$H_i(z) = (b_{0i} + b_{1i} \cdot z + b_{2i} \cdot z^2) / (1 + a_{1i} \cdot z + a_{2i} \cdot z^2).$$

4. **Параллельная форма.** Используется много реже и соответствует представлению передаточной функции в виде суммы биквадратных блоков или более простых функций.

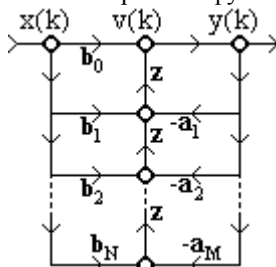


Рис. 21.13.

**Обращенные формы.** В теории линейных сигнальных графов существуют процедуры преобразования исходных графов с сохранением передаточных функций. Одна из таких процедур - обращение графов, которое выполняется путем изменения направления всех ветвей цепи, при этом вход и выход графа также меняются местами. Для ряда систем такая транспозиция позволяет реализовать более эффективные алгоритмы обработки данных. Пример обращения графа прямой канонической формы рекурсивной системы (с перестроением на расположение входа с левой стороны) приведен на рис. 21.13.

## 21.8. Нерекурсивная фильтрация сигналов

Предмет фильтрации сигналов является основной областью применения нерекурсивных систем. Под фильтрацией будем понимать любое преобразование, при котором во входных сигналах целенаправленно изменяются определенные соотношения (динамические или частотные) между различными компонентами этих сигналов.

К операциям фильтрации информации традиционно относят операции сглаживания, прогнозирования, дифференцирования,



интегрирования и разделения сигналов, а также выделение информационных (полезных) сигналов и подавление шумов (помех). Основным методом фильтрации данных являются частотная селекция сигналов. Системы, выполняющие операции фильтрации, обычно называют фильтрами.

Операция свертки сигнала с ядром (оператором) фильтра отображается в частотной области умножением спектра сигнала на частотный образ оператора фильтра. Тем самым, мы получаем возможность целенаправленно изменять спектр сигнала. Например, в звукозаписи изменение спектра сигнала позволяет очищать запись от шумов, компенсировать искажения сигнала различными устройствами звукозаписи, менять тембры инструментов. При обработке изображений фильтрация позволяет применять к изображению разные эффекты: размытие, подчеркивание границ, тиснение и пр. В самых различных областях фильтрация часто служит для разделения сигналов, смешанных в один, очищения сигнала от шумов, а также является составным компонентом многих других, более сложных процессов.

**Частотные характеристики фильтров.** В общем случае, фильтр меняет в спектре сигнала и амплитуды, и фазы гармоник. Однако фильтры можно проектировать так, чтобы они или не меняли фазу сигнала, или сдвигали все гармоники сигнала по времени на одну и ту же величину (сдвигали во времени весь сигнал). Такие фильтры называют *фильтрами с линейной фазой*.

Основное свойство фильтра – его *амплитудно-частотная* и *фазочастотная характеристики* (АЧХ и ФЧХ). Они показывают, какое влияние фильтр оказывает на амплитуду и фазу различных гармоник обрабатываемого сигнала. Если фильтр имеет линейную фазу, то рассматривается только АЧХ фильтра. Обычно частотная характеристика изображается в виде графика зависимости амплитуды от частоты (в децибелах). Так, если фильтр пропускает все сигналы в какой-либо полосе частот без изменения (коэффициент передачи равен 1), то это отображается значением 0 дБ. Подавление каких либо частот отображается со знаком минус (в 2 раза  $\Rightarrow$  -6 дБ, в 10 раз  $\Rightarrow$  -20 дБ, в 100 раз  $\Rightarrow$  -40 дБ), а усиление, соответственно, со знаком плюс. Пример частотной характеристики фильтра приведен на рис. 21.14.

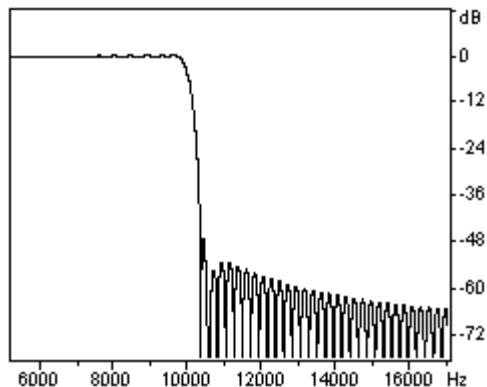


Рис. 21.14. Пример АЧХ фильтра.

В зависимости от общего вида частотной характеристики можно выделяют следующие распространенные типы фильтров: НЧ-фильтры (*low-pass filters*), ВЧ-фильтры (*high-pass filters*), полосовые фильтры, которые пропускают (*band-pass filters*) или подавляют (*band-reject filters*) сигнал только в определенной частотной полосе. Существуют и другие типы фильтров с более сложными частотными характеристиками.

Обычно в задачах фильтрации сигнала задается требуемая частотная характеристика фильтра. Построить в точности заданный фильтр обычно бывает не так просто. Тогда строится фильтр, близкий по характеристикам к заданному.

Например, невозможно построить *идеальный фильтр низких частот*, который пропускает без изменения все сигналы ниже определенной частоты (в *полосе пропускания, pass band*) и полностью подавляет все сигналы выше этой частоты (в *полосе подавления, stop band*). Такой фильтр реализуется оператором бесконечно большого размера. Реальные фильтры низких частот обладают плавным переходом от полосы пропускания (0 дБ с максимально возможными отклонениями обычно не более  $\pm 0.5$  дБ) к полосе подавления, где сигнал подавляется, как правило, более чем в 100-1000 раз (рис. 21.14). Крутизна спада и значения подавления после спада определяются конкретными требованиями к фильтру.

В пределе, частотная характеристика фильтра может задаваться произвольной кривой. Например, фильтр может иметь несколько частотных полос пропускания с разными коэффициентами усиления, разделенных полосами с разными коэффициентами подавления.

**Расчет операторов нерекурсивных фильтров.**

Существует много способов построения фильтров с заданной частотной характеристикой. Наиболее универсальный и, в то же время, наиболее простой – проектирование фильтров путем обратного преобразования Фурье требуемой частотной характеристики фильтра. Рассмотрим этот способ подробнее на примере чисто технической задачи формирования сигналов с высоким временным разрешением на выходе детекторов излучения.

Сигналы регистрации ионизирующих частиц формируются интегрированием заряда, в который трансформируется энергия частиц при ее поглощении в рабочей среде детектора. Идентификация энергии частиц по зарегистрированным сигналам обеспечивается при полном сборе заряда на выходе детектора, при этом постоянная времени RC интегрирования заряда на выходной RC-цепи детектора должна быть на порядок больше времени преобразования энергии в детекторе.

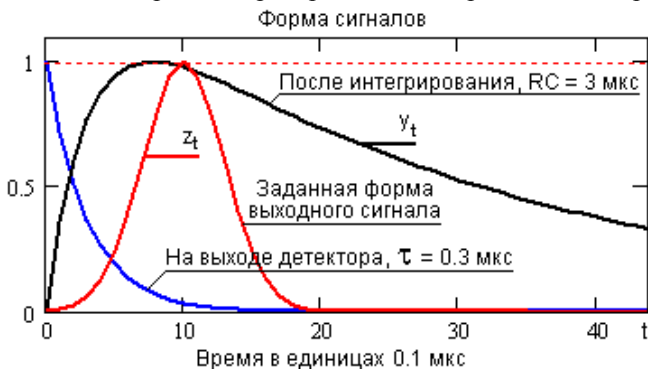


Рис. 21.15.

На рис. 21.15 приведен пример формы выходных сигналов детектора до и после сбора зарядов при постоянной  $\tau = 0.3$  мкс выхода заряда и  $RC = 3$  мкс интегрирования заряда. Амплитуды сигналов нормированы по максимуму для наглядности сравнения.

Для исключения ошибок измерений при возможных наложениях последующих сигналов на спад предыдущих применяются различные способы укорочения сигналов  $y(t)$  и быстрого восстановления нулевой линии. Применение нерекурсивного формирующего фильтра позволяет решить эти задачи просто и эффективно.

Допустим, что необходимо превратить выходной сигнал  $y(t)$  в короткий и удобный для измерения амплитудных значений сигнал  $z(t)$ , форма которого приведена на рис. 21.15. Это можно выполнить операцией свертки:

$$z(t) = h(t) * y(t),$$

для чего требуется определить соответствующий оператор преобразования  $h(t)$ . Переведем левую и правую часть уравнения свертки в спектральную область:

$$Z(f) = H(f) \cdot Y(f). \quad (21.39)$$

Отсюда, частотная характеристика и функция требуемого оператора свертки определяются выражениями:

$$H(f) = Z(f) / Y(f). \quad (21.40)$$

$$H(f) \leftrightarrow h(t). \quad (21.41)$$

Возможность реализации оператора  $h(t)$  определяется устойчивостью решения уравнения (21.40) и зависит от частотных характеристик сигналов  $y(t)$  и  $z(t)$ . К числу обязательных условий реализации следует отнести отсутствие полюсов функции  $Y(f)$  для исключения деления на ноль, и более быстрое затухание функции  $Z(f)$  по сравнению с функцией  $Y(f)$ . В качестве сигнала  $z(t)$  целесообразно задавать функцию Гаусса такой ширины (на половине своей высоты), спектр которой  $Z(f)$  по своей основной значимой части соизмерим со спектром сигнала  $Y(f)$ . Чем меньше ширина функции Гаусса, тем лучше будет выполняться сжатие сигнала, однако при чрезмерном сжатии и, соответственно, увеличении высокочастотных составляющих в  $Z(f)$ , оператору  $h(t)$  придется осуществлять существенный подъем высокочастотных составляющих сигнала  $y(t)$ , что приведет к увеличению коэффициента усиления фильтром дисперсии помех. С учетом каузальности работы формирующих фильтров выходные сигналы должны задаваться за пределами фронта сигнала  $y(t)$ , а экстремумы сигналов – за пределами экстремума входного сигнала.

В силу линейности фильтра и принципа суперпозиции зарядов на интегрирующей емкости, в любой текущей временной точке оператор фильтра реагирует на разряд интегрирующей емкости, как на сумму разрядов от всех предыдущих импульсов. И если по разряду одного импульса система будет точно настроена на нулевую линию, то она будет сохранять нулевую линию независимо от количества и времени прихода импульсов, а равно и от наложения импульсов друг на друга.

Форма сигнала  $y(t)$  в первом приближении соответствует выражению:

$$y(t) = \exp(-t/RC) - \exp(-t/\tau). \quad (21.42)$$

Основная информация сигнала заключена в его фронте и в спектральной области формирует высокочастотные составляющие сигнала. На рис. 21.16 приведены модули спектров  $Y(f)$  и  $Z(f)$  сигналов

$y(t)$  и  $z(t)$ , и частотные передаточные функции оператора фильтра  $Z(f)/Y(f)$ . Для наглядного представления их формы все модули нормированы к 1 по максимальным значениям.

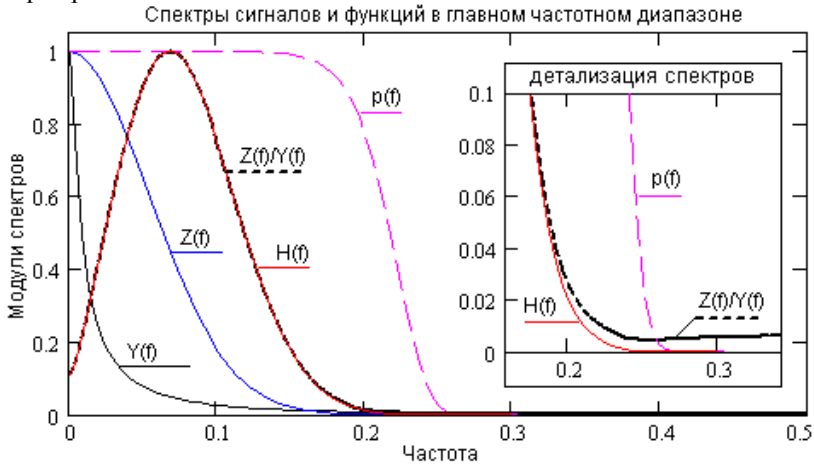


Рис. 21.16.

Значимую часть спектра формируемого оператора целесообразно выделить умножением спектра оператора на весовую функцию  $p(f)$ , равную 1 в пределах значимой части отношения  $Z(f)/Y(f)$ , и плавно спадающую к нулю за его границами (окно детализации на рисунке). Одновременно это подавит высокочастотные шумы в сигнале  $y(t)$ , появление которых в регистрируемом сигнале неизбежно в силу природы ионизирующего излучения. В результате мы получим спектральную функцию  $H(f)$  требуемого оператора преобразования сигналов.

Временная функция оператора фильтра  $h(t)$  вычисляется обратным преобразованием Фурье функции  $H(f)$ . Для начала длина оператора устанавливается соизмеримой с длиной входного сигнала (оператор  $hc(t)$  на рис. 21.17). В данном случае оператор фильтра является конечным, достаточно быстро затухает, и может быть ограничен до величины  $h(t)$ .

В общем случае, для ирезанных и скачкообразных функций  $H(f)$ , оператор фильтра может затухать достаточно медленно и может потребоваться его усечение до определенных конечных размеров с применением весовых функций.

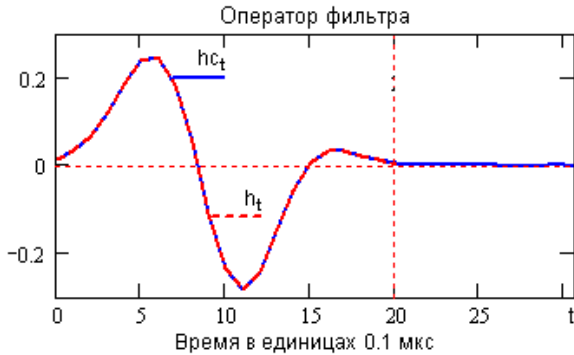


Рис. 21.17.

По заданной частотной характеристике  $H(f)$  может быть синтезирован аналоговый фильтр, но настройка такого фильтра под конкретные параметры детектора будет представлять собой достаточно трудоемкую операцию. Больше возможностей в этом отношении представляют дискретные методы преобразования сигналов. Ограничение размеров дискретных операторов определяется допустимой погрешностью реконструкции заданной формы сигналов и точностью формирования нулевой линии при заданном временном разрешении. Работоспособность фильтра проверяется сверткой оператора с сигналом  $y(t)$ . На рис. 21.18 приведено сопоставление заданной формы сигнала  $z(t)$  и формы сигнала  $zh(t)$  на выходе фильтра при подаче на его вход сигнала  $y(t)$ . Фильтр был смоделирован в цифровой форме с шагом  $\Delta t = 0.1$  мкс с размером окна  $N = 20$ . Оператор фильтра имел коэффициент передачи постоянной составляющей  $K_{пс} = 0.2$  и коэффициент усиления дисперсии шумов 0.85.

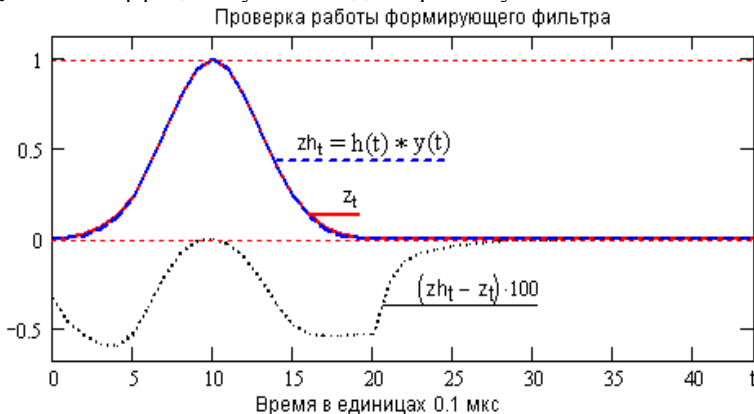


Рис. 21.18.

На рисунке приведена также абсолютная погрешность реконструкции сигнала  $\Delta z(t)=zh(t)-z(t)$ , увеличенная для наглядности в 100 раз. Погрешность реконструкции флангов сигнала при амплитудных измерениях значения не имеет. Гораздо большее значение имеет быстрое и точное восстановление нулевой линии после формирования выходного сигнала, которое определяет погрешность измерения амплитуды последующего сигнала, наложенного на разряд интегрирующей емкости от предыдущего. Компенсация погрешностей, возникающих за счет усечения спектров и ограничения размеров самого оператора, может проводиться коррекцией значения последнего члена оператора (выход на нулевую линию), а коррекцией значения первого члена оператора может настраиваться точное соотношение 1:1 амплитуд входных и выходного сигналов.

Реконструкция сигналов, приведенная на рис. 21.18, обеспечила временное разрешение измерений 3-4 мкс при точности амплитудных измерений 0.1%. На рис. 21.19 приведена проверка временного разрешения фильтра на модели сигнала в виде трех последовательных импульсов одинаковой и достаточно большой амплитуды, следующих через 4 мкс друг за другом и вызывающих значительное смещение напряжения на интегрирующей емкости RC-цепи. Пунктиром на рисунке приведено восстановление нулевой линии на выходе фильтра с увеличением масштаба в 100 раз.

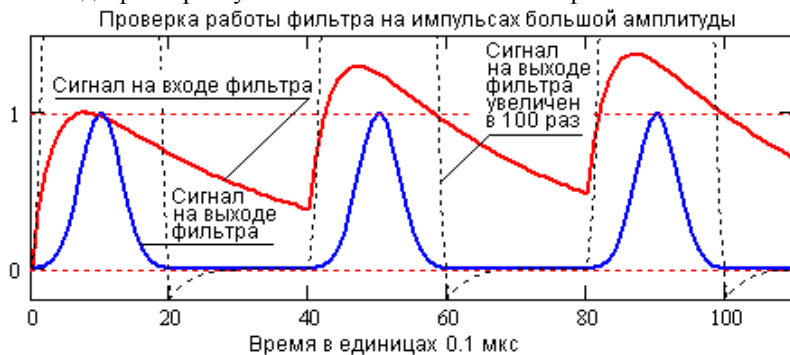


Рис. 21.19. Проверка временного разрешения фильтра.

Точность реконструкции сигналов при разных значениях амплитуд входных сигналов можно видеть на рис. 21.20.

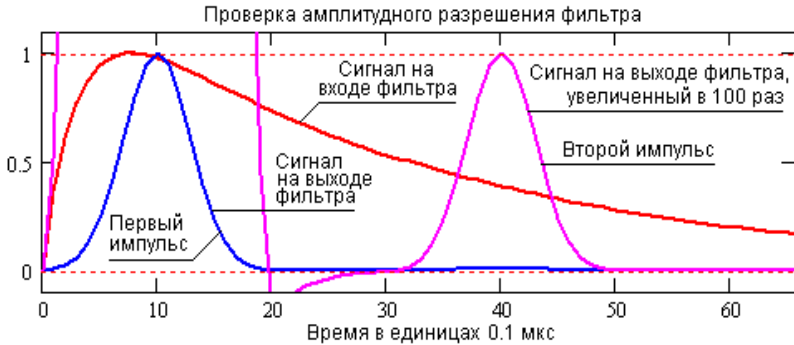


Рис. 21.20.

На модели установлено отношение амплитуд 100:1 первого и второго импульсов. Второй импульс, наложенный на спад первого импульса, практически не отражается на его форме. Фильтр уверенно разделяет эти импульсы с ошибкой формирования амплитуды второго импульса не более 1%.

## 21.9. Преобразование Хартли (основные сведения)

Преобразование Хартли является аналогом преобразования Фурье и может применяться для спектрального анализа, фильтрации и обработки сигналов. Название преобразование получило по имени Р.Хартли, опубликовавшего в 1942 г. статью о паре интегральных преобразований - прямом и обратном, использующих введенную им функцию  $\text{cas}\theta = \cos\theta + \sin\theta$ . Преобразование оставались в забвении до 80-х годов прошлого века.

Обращение к преобразованию Хартли было обусловлено ситуацией, сложившейся в ряде методов обработки информации, использующих вещественные последовательности данных, обработку которых желательно осуществлять в области вещественных чисел. В отличие от преобразования Фурье, отображающего вещественные функции в комплексную область и несимметричного по комплексной переменной, преобразование Хартли осуществляет преобразования только в вещественной области, отображая вещественные сигналы  $s(t)$  в вещественные  $S(\omega)$ . Прямое и обратное преобразование Хартли взаимно симметричны. Большой вклад в развитие преобразования внес Р. Брейсуэлл, разработавший основы теории непрерывного и



дискретного преобразования Хартли, а также один из вариантов его быстрого преобразования. Применение преобразования перспективно для обработки изображений.

**Определение преобразования.** Преобразование Хартли задается парой формул:

$$\text{Sh}(\omega) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} s(t) \text{cas } \omega t \, dt \quad (21.43)$$

$$s(t) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \text{Sh}(\omega) \text{cas } \omega t \, d\omega \quad (21.44)$$

где функция cas представляет собой сумму косинуса и синуса одного аргумента:

$$\text{cas } \omega t = \cos \omega t + \sin \omega t. \quad (21.45)$$

Множители  $1/\sqrt{2\pi}$  обусловлены применением в формулах аргумента  $\omega$ . Они могут заменяться одним множителем  $1/2\pi$  только в формуле (21.44), но это нарушает симметричность прямого и обратного преобразования. При необходимости применения симметричных алгоритмов в формулах можно использовать аргумент циклической частоты:

$$\text{Sh}(f) = \int_{-\infty}^{\infty} s(t) \text{cas } 2\pi f t \, dt \quad (21.46)$$

$$s(t) = \int_{-\infty}^{\infty} \text{Sh}(f) \text{cas } 2\pi f t \, df \quad (21.47)$$

На первый взгляд в формулах отсутствуют существенные отличия от интегральных преобразований Фурье, однако на практике эти различия могут быть достаточно ощутимыми, что определяется вещественным характером функции  $\text{Sh}(f)$ .

**Связь преобразований Фурье и Хартли.**

Допустим, имеется произвольная функция  $s(t) \xrightarrow{\text{ТН}} \text{Sh}(f)$ ,

$s(t) \xrightarrow{\text{ТФ}} S(f)$ , где  $\text{Sh}(f)$  и  $S(f)$  – результаты преобразования Хартли и Фурье (Хартли- и Фурье-образы  $s(t)$ ). Любая функция  $y(x)$  может быть представлена в виде суммы четной и нечетной компонент, и однозначно по ним восстановлена. Четная компонента определяется как полусумма функции  $y(x)$  и ее зер-

кального изображения  $y(-x)$ , нечетная компонента определяется как полуразность этих функций и обладает свойством антисимметрии, т.е.  $y(-x) = -y(x)$ . Запишем для преобразования Хартли:

$$\text{Sh}(f) = \int_{-\infty}^{\infty} s(t) \cos 2\pi ft \, dt = \int_{-\infty}^{\infty} s(t) \cos 2\pi ft \, dt + \int_{-\infty}^{\infty} s(t) \sin 2\pi ft \, dt.$$

$$\text{Sh}(f) = \text{Sh}_{\text{sym}}(f) + \text{Sh}_{\text{asym}}(f),$$

$$\text{Sh}_{\text{sym}}(f) = [\text{Sh}(f) + \text{Sh}(-f)]/2 = \int_{-\infty}^{\infty} s(t) \cos 2\pi ft \, dt, \quad (21.48)$$

$$\text{Sh}_{\text{asym}}(f) = [\text{Sh}(f) - \text{Sh}(-f)]/2 = \int_{-\infty}^{\infty} s(t) \sin 2\pi ft \, dt, \quad (21.49)$$

где  $\text{Sh}_{\text{sym}}(f)$  и  $\text{Sh}_{\text{asym}}(f)$  – четная и нечетная компоненты  $\text{Sh}(f)$ . С другой стороны, для преобразования Фурье имеем:

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp(-2\pi ft) \, dt = A(f) - j B(f).$$

$$A(f) = \int_{-\infty}^{\infty} s(t) \cos 2\pi ft \, dt, \text{ - четная вещественная часть спектра,}$$

$$B(f) = \int_{-\infty}^{\infty} s(t) \sin 2\pi ft \, dt, \text{ - нечетная мнимая часть спектра.}$$

Сравнивая эти две группы выражений, нетрудно сделать выводы по формулам связи преобразований Фурье и Хартли:

$$S(f) = \text{Sh}_{\text{sym}}(f) - j \text{Sh}_{\text{asym}}(f), \quad (21.50)$$

$$\text{Sh}(f) = A(f) - B(f). \quad (21.51)$$

Таким образом, преобразование Фурье равно разности четной составляющей преобразования Хартли и нечетной составляющей, умноженной на  $j$ , а преобразование Хартли определяется как разность вещественной и мнимой составляющих преобразования Фурье.

Пример прямого и обратного преобразования Хартли для простого сигнала (гауссиан с быстро затухающей частотной характеристикой) и сопоставления четной и нечетной частей его спектра с действительной и мнимой частью спектра Фурье приведен на рис. 21.21.

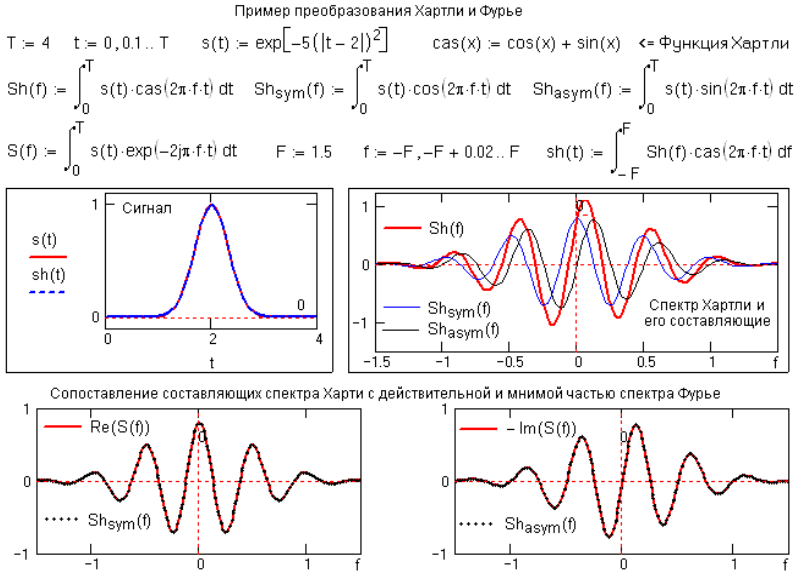


Рис. 21.21.

Как можно видеть на рисунке, спектр Хартли  $\text{Sh}(f)$  даже такого простого сигнала, как гауссиан, выглядит достаточно сложно, равно как и его четная и нечетная составляющие, и мало пригоден для визуального анализа. Отметим, однако, что действительная и мнимая части преобразования Фурье, хотя и имеют определенный физический смысл, как амплитудные распределения косинусных и синусных колебаний, в качественном анализе сигналов также используются достаточно редко. Гораздо большее практическое значение для анализа имеют модуль и фаза спектра (амплитудно-частотная и фазово-частотная характеристика) и спектр плотности мощности сигнала. Отметим также, что для преобразования Хартли не характерна избыточность преобразования Фурье, т.к. в общем случае вещественные функции Хартли эквивалентны сопряженным комплексным функциям Фурье.

**Энергетический и фазовый спектры.**

Энергетический спектр преобразования Фурье (спектр плотности мощности сигнала) задается выражением:

$$W(f) = A^2(f) + B^2(f).$$

Частотный спектр плотности мощности сигнала не должен

зависеть от формы математического представления спектра. Из этих соображений для вычисления энергетического спектра по спектру Хартли следует:

$$\begin{aligned} Wh(f) &= (Sh_{sym}(f))^2 + (Sh_{asym}(f))^2 = [Sh(f) + Sh(-f)]^2/4 + [Sh(f) - Sh(-f)]^2/4, \\ Wh(f) &= [Sh^2(f) + Sh^2(-f)]/2. \end{aligned} \quad (21.52)$$

Аналогично, вычисление фазовой частотной характеристики сигнала:

$$\begin{aligned} \varphi(f) &= \text{argtg}(B(f)/A(f)) = \text{argtg}(-Sh_{asym}(f)/Sh_{sym}(f)), \\ \varphi(f) &= \text{argtg}(-[Sh(f) - Sh(-f)]/[Sh(f) + Sh(-f)]). \end{aligned} \quad (21.53)$$

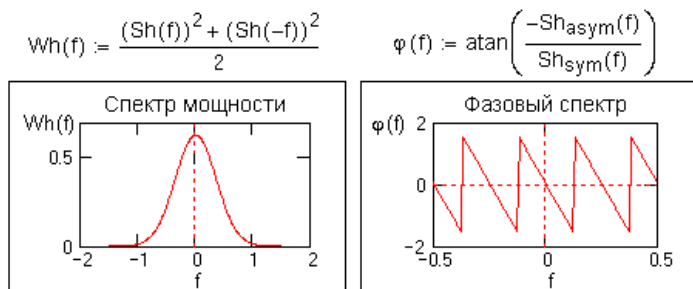


Рис. 21.22.

На рис. 21.22. приведено продолжение вычислений для сигнала на рис. 21.21 и показаны частотные функции спектра плотности мощности и фазовой характеристики сигнала.

С учетом всех вышеприведенных выражений преобразование Хартли может рассматриваться как гладкая вещественная форма представления спектра вещественного сигнала.

### 21.10. Свойства преобразования

**Линейность.** Преобразование Хартли относится к числу линейных интегральных операций, т.е. спектр суммы сигналов равен сумме спектров этих сигналов.

$$\sum_n a_n s_n(t) \xrightarrow{TH} \sum_n a_n Sh_n(\omega). \quad (21.54)$$

Пример суммирования сигналов и его отображения в суммирования спектров приведен на рис. 21.23:

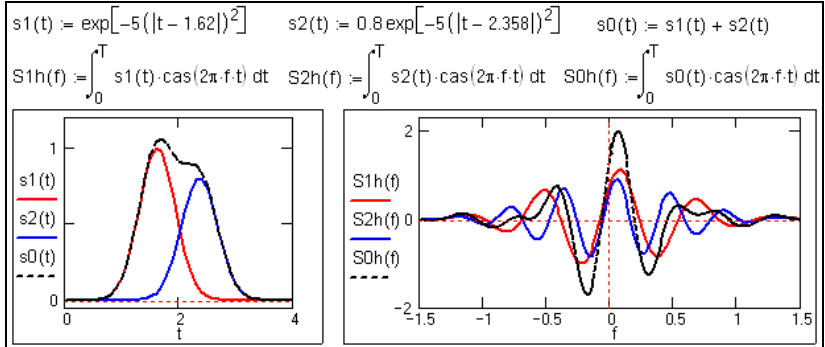


Рис. 21.23. Сигналы и их спектры.  $s_0(t)=s_1(t)+s_2(t) \Leftrightarrow S_1h(f)+S_2h(f) = S_0h(f)$ .

**Четность и нечетность спектральных функций.**

Свойства четности преобразования Фурье распространяются и на преобразование Хартли. Для четных сигналов равен нулю интеграл (21.49) и в спектре Хартли отсутствует нечетная составляющая (функция  $Sh(f)$  – четная). Для нечетных сигналов равен нулю интеграл (21.48) и спектральная функция нечетная. Это можно наглядно видеть на рис. 21.24.

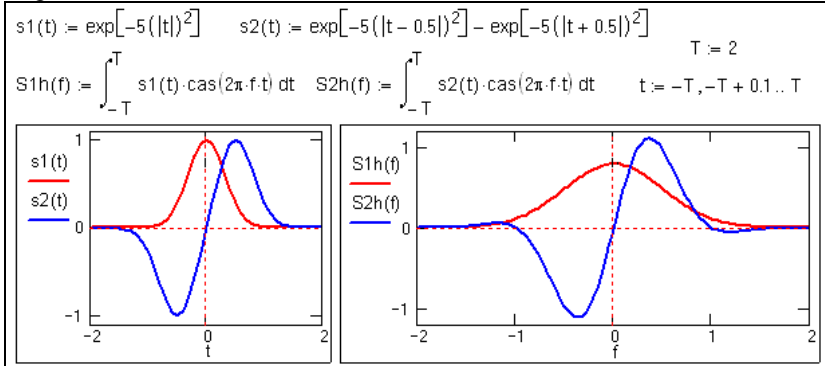


Рис. 21.24. Преобразование Хартли четных и нечетных функций.

Энергетические и фазовые частотные спектры, как и для преобразования Фурье, также являются четными и нечетными соответственно.

**Изменение аргумента функции** (сжатие или расширение сигнала) приводит к обратному изменению аргумента ее хартли-образа и обратно пропорциональному изменению его модуля (рис. 21.25).

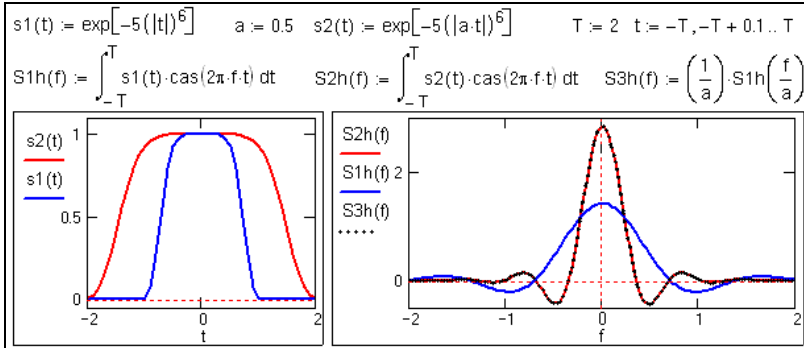


Рис. 21.25.

Действительно, если  $s(t) \xrightarrow{TH} Sh(f)$ , то при изменении длительности сигнала с сохранением его формы (растяжении сигнала по временной оси), для сигнала с новым аргументом  $s(x) = s(at)$  при  $x=at$ , получаем:

$$s(at) \xrightarrow{TH} \int_{-\infty}^{\infty} s(at) \cos(2\pi ft) dt = (1/a) \int_{-\infty}^{\infty} s(x) \cos(2\pi fx/a) dx$$

$$s(at) \xrightarrow{TH} (1/a) Sh(f/a). \quad (21.55)$$

Выражение (21.55) действительно при  $a>0$ . При  $a<0$  происходит зеркальный поворот сигнала относительно вертикальной оси, а замена переменной  $t=x/a$  вызывает перестановку пределов интегрирования и, соответственно, изменение знака спектра:

$$s(at) \xrightarrow{TH} -(1/a) Sh(f/a). \quad (21.56)$$

Обобщенная формула изменения аргумента:

$$s(at) \xrightarrow{TH} (1/|a|) Sh(f/a), \quad a \neq 0 \quad (21.57)$$

Если под аргументом функции и ее спектра понимать определенные физические единицы, например, время - частота, то отсюда следует: чем короче по своей длительности сигнал, тем шире по частоте его спектр, и наоборот, что полностью аналогично преобразованию Фурье.

**Теорема запаздывания.** Применяя замену переменной  $t-t_0=x$ , после тригонометрических преобразований (выполнить самостоятельно) получаем (рис. 21.24):

$$s(t-t_0) \xrightarrow{TH} \int_{-\infty}^{\infty} s(t-t_0) \cos(2\pi ft) dt = \cos(2\pi ft_0) Sh(f) + \sin(2\pi ft_0) Sh(-f). \quad (21.58)$$

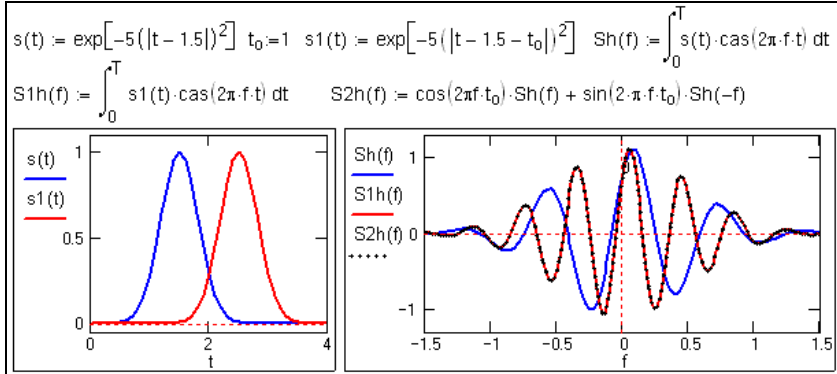


Рис. 21.26. Изменение спектра сигнала при сдвиге.

Очевидно, что амплитуды гармоник сигнала (модуль спектра) при его сдвиге изменяться не должны. Запаздывание (смещение) сигнала по аргументу функции на интервал  $t_0$ , как и для преобразования Фурье, приводит к изменению фазово-частотной функции спектра (фазового угла всех гармоник) на величину  $-2\pi f t_0$ . Если известна фазовая характеристика  $\varphi(f)$  спектра сигнала, то для определения фазовой функции сдвинутого сигнала достаточно выполнить вычисление  $\varphi(f, t_0) = \varphi(f) - 2\pi f t_0$  (рис. 21.27).

$$\varphi(f) := \operatorname{atan}\left(\frac{-Sh(f) + Sh(-f)}{Sh(f) + Sh(-f)}\right) \quad \varphi_1(f) := \operatorname{atan}\left(\frac{-S1h(f) + S1h(-f)}{S1h(f) + S1h(-f)}\right)$$

$$\varphi_2(f) := \varphi(f) - 2\pi \cdot f \cdot t_0 \quad a(f) := \tan(\varphi_1(f)) \quad \varphi_2(f) := \operatorname{atan}(a(f))$$

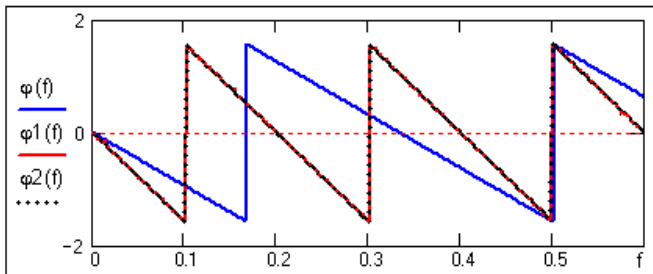


Рис. 21.27. Фазовая функция спектра.

**Преобразование производной** (дифференцирование сигнала):

$$s'(t) = d[s(t)]/dt = d\left[\int_{-\infty}^{\infty} Sh(f) \cos(2\pi f t) df\right]/dt = \int_{-\infty}^{\infty} Sh(f) [d(\cos(2\pi f t))/dt] df =$$

$$= - \int_{-\infty}^{\infty} 2\pi f \operatorname{Sh}(-f) \operatorname{cas}(2\pi f t) df \xrightarrow{\text{TH}} -2\pi f \operatorname{Sh}(-f). \quad (21.59)$$

Таким образом, дифференцирование сигнала отображается в спектральной области простым умножением зеркального изображения спектра сигнала на *оператор дифференцирования сигнала в частотной области*  $-2\pi f$ . Умножение на  $2\pi f$  приводит к обогащению спектра производной сигнала высокочастотными составляющими (по сравнению с исходным сигналом) и уничтожает составляющие с нулевой частотой.

Пример сигнала, его производной и соответствующих им спектров приведен на рис. 21.28.

$$s_1(t) := \exp[-3(|t-2|)^2] \quad s_2(t) := \frac{d}{dt}s_1(t) \quad S_1h(f) := \int_0^T s_1(t) \cdot \operatorname{cas}(2\pi \cdot f \cdot t) dt$$

$$S_2h(f) := \int_0^T s_2(t) \cdot \operatorname{cas}(2\pi \cdot f \cdot t) dt \quad S_3h(f) := -2\pi \cdot f \cdot S_1h(-f)$$

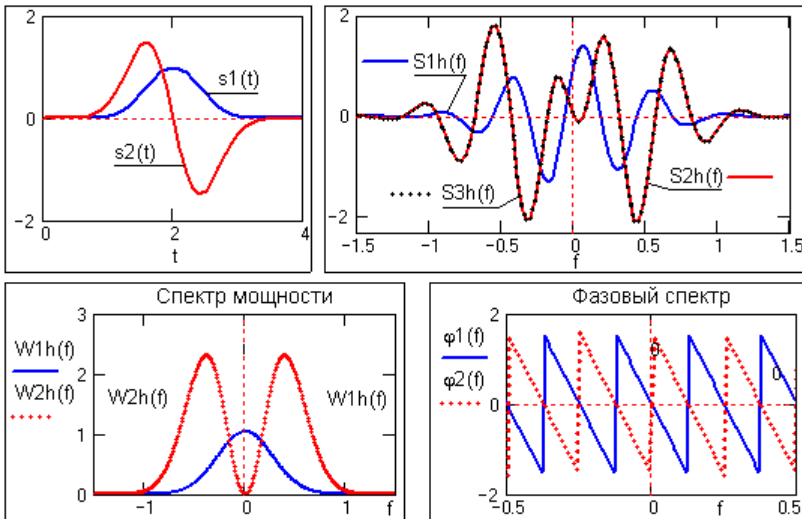


Рис. 21.28. Спектры сигнала и его производной

По изменению аргумента спектра можно видеть, что для всех гармоник спектра появляется сдвиг фаз на  $\pi/2$  ( $90^\circ$ ) для положительных частот, и на  $-\pi/2$  ( $-90^\circ$ ) для отрицательных частот.

Аналогично могут быть получены выражения для производных более высокого порядка. В частности, для второй производной:

$$d^2[s(t)]/dt^2 = -(2\pi f)^2 \operatorname{Sh}(f).$$



**Преобразование интеграла** сигнала в частотной области при известном спектре сигнала может быть получено из следующих соображений. Если имеет место  $y(t) = d[s(t)]/dt \xrightarrow{TH} -2\pi f \text{Sh}(-f) = Yh(f)$ , то должна выполняться и обратная операция:

$$s(t) = \int y(t)dt \xrightarrow{TH} (1/2\pi f) Yh(-f). \quad (21.60)$$

$$s(t) := \int_0^t s2(t) dt \quad \text{Sh}(f) := \int_0^T s(t) \cdot \text{cas}(2\pi \cdot f \cdot t) dt \quad \text{Sh}(f) := \frac{S2h(-f)}{2\pi f} \quad F := 1.50001$$

$$f := -F, -F + 0.01 .. F$$

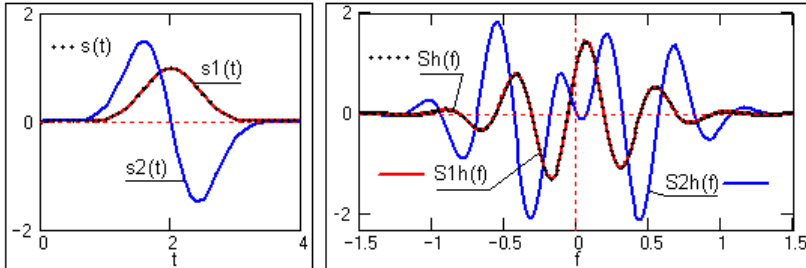


Рис. 21.29. Спектры сигнала и его интеграла

На рис. 21.29 выполнено интегрирование сигнала  $s2(t) = d[s(t)]/dt$ , дифференцирование которого показано на рис. 21.28, т.е. восстановление сигнала  $s(t)$ . Как и в преобразовании Фурье, оператор интегрирования  $(1/2\pi f)$  в частотной области  $f > 1$  ослабляет гармоники высоких частот, а при  $f < 1$  усиливает низкие частоты. Фазовый спектр сигнала смещается на  $-90^\circ$  для положительных частот и на  $90^\circ$  для отрицательных. При  $f=0$  в выражении (21.60) имеется особая точка (деление на ноль), вычисление значения в которой должно выполняться путем предельного перехода ( $f \rightarrow 0$ ).

**Преобразование свертки и произведения** сигналов у Хартли в общем случае выглядит сложнее, чем в преобразовании Фурье. Для свертки сигналов имеем:

$$s(t) * u(t) \xrightarrow{TH} 0.5 [\text{Sh}(f)Uh(f) - \text{Sh}(-f)Uh(-f) + \text{Sh}(f)Uh(-f) + \text{Sh}(-f)Uh(f)]. \quad (21.61)$$

Для произведения сигналов:

$$s(t) u(t) \xrightarrow{TH} 0.5 [\text{Sh}(f)*Uh(f) - \text{Sh}(-f)*Uh(-f) + \text{Sh}(f)*Uh(-f) + \text{Sh}(-f)*Uh(f)]. \quad (21.62)$$

Однако это не более чем видимость, т.к. выполняя свертку через преобразование Фурье  $s(t)*u(t) \xrightarrow{TF} S(f)U(f)$  мы производим перемножение двух комплексных функций  $A(f)-jB(f)$  с

соответствующим суммированием также произведений четырех членов. С позиций обработки и фильтрации данных, существенное значение имеет тот фактор, что если хотя бы одна из функций, входящих в формулу свертки, является либо четной, либо нечетной, то формулы прямых преобразований Хартли и Фурье для свертки полностью совпадают. Как правило, операторы фильтров для исключения сдвига фазы обрабатываемых данных выполняют симметричными, при этом преобразование Хартли имеет вид:

$$h(t) * s(t) \xrightarrow{\text{TH}} \text{Hh}(f) \text{Sh}(f). \quad (21.63)$$

Если оператор фильтра асимметричный (нечетный), то формула приобретает вид:

$$h(t) * s(t) \xrightarrow{\text{TH}} \text{Hh}(f) \text{Sh}(-f). \quad (21.64)$$

**Преобразование функции корреляции.** В частотной области преобразование Хартли автокорреляционной функции, как и преобразование Фурье, представляет собой спектральную плотность мощности сигнала:

$$W_s(\tau) = \langle s(t), s(t+\tau) \rangle \xrightarrow{\text{TH}} 0.5 [\text{Sh}^2(f) + \text{Sh}^2(-f)] = \text{Wh}_s(f). \quad (21.65)$$

Учитывая четность автокорреляционной функции и спектра мощности, оно практически ничем не отличается от преобразования Фурье, за исключением алгоритма вычислений. На рис. 21.30 приведен пример вычисления автокорреляционной функции сигнала с использованием преобразования Хартли.

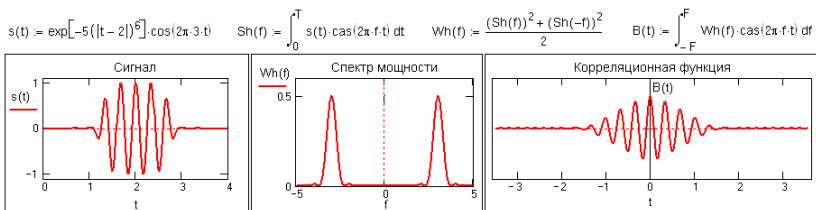


Рис. 21.30. Вычисление корреляционной функции сигнала.

**Двумерное преобразование Хартли.** Для двумерной функции  $s(x, y)$  двумерное преобразование Хартли задается выражениями:

$$\text{Sh}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x, y) \text{cas}[2\pi(ux+vy)] dx dy, \quad (21.66)$$

$$s(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Sh}(u, v) \text{cas}[2\pi(ux+vy)] dudv, \quad (21.67)$$

Связь двумерного преобразования Хартли с преобразованием

Фурье легко устанавливается аналогично одномерному преобразованию:

$$\text{Sh}(u, v) = \text{Re}(S(u, v)) - \text{Im}(S(u, v)). \quad (21.68)$$

Соответственно, для двумерного ПХ сохраняются свойства четности и нечетности одномерного преобразования. Если функция  $s(x, y)$  обладает свойством круговой симметрии, то ее двумерное преобразование Хартли совпадает с двумерным преобразованием Фурье.

Основные свойства двумерного преобразования:

Подобие:

$$s(ax, by) \xrightarrow{\text{TH}} (1/|ab|) \text{Sh}(u/a, v/b). \quad (21.69)$$

Сдвиг:

$$s(x-a, y-b) \xrightarrow{\text{TH}} \cos(2\pi(au+bv))\text{Sh}(u, v) + \sin(2\pi(au+bv))\text{Sh}(-u, -v). \quad (21.70)$$

Модуляция:

$$s(x, y) \cos(2\pi(u_0x+v_0y)) \xrightarrow{\text{TH}} 0.5 [\text{Sh}(u-u_0, v-v_0) + \text{Sh}(u+u_0, v+v_0)]. \quad (21.71)$$

Корреляция:

$$V(\tau_x, \tau_y) \xrightarrow{\text{TH}} T [\text{Sh}^2(u, v) + \text{Sh}^2(-u, -v)]. \quad (21.72)$$

Произведение при разделении переменных:

$$s(x)h(x) \xrightarrow{\text{TH}} 0.5 [\text{Sh}(u, v)\text{Hh}(u, v) - \text{Sh}(-u, -v)\text{Hh}(-u, -v) + \text{Sh}(u, v)\text{Hh}(-u, -v) + \text{Sh}(-u, -v)\text{Hh}(u, v)]. \quad (21.73)$$

Свертка:

$$s(x, y)*h(x, y) \xrightarrow{\text{TH}} 0.5T [\text{Sh}(u, v)\text{Hh}(u, v) - \text{Sh}(-u, -v)\text{Hh}(-u, -v) + \text{Sh}(u, v)\text{Hh}(-u, -v) + \text{Sh}(-u, -v)\text{Hh}(u, v)]. \quad (21.74)$$

## 21.11. Дискретное преобразование Хартли

**Формулы преобразования.** Как обычно, для представления функций с равномерной дискретизацией примем обозначение  $s(n\Delta t)$ , или  $s(n) \equiv s_n$  при  $\Delta t = 1$ . Количество отсчетов функции равно  $N$ , интервал задания  $[0, N-1]$ . Прямое и обратное преобразование Хартли:

$$\text{Sh}(m\Delta f) = (1/N) \sum_{n=0}^{N-1} s(n\Delta t) \text{cas}(2\pi m\Delta f n\Delta t), \quad n = 0 \dots N-1, \quad (21.75)$$

$$s(n\Delta t) = \sum_{m=0}^{M-1} \text{Sh}(m\Delta t) \text{cas}(2\pi m\Delta f n\Delta t), \quad m = 0 \dots M-1. \quad (21.76)$$

Здесь:  $M$  – количество отсчетов спектральной функции с шагом дискретизации по частоте  $\Delta f$ . Дискретизация сигнала вызывает периодизацию частотных функций. Частота Найквиста для главного частотного диапазона  $f_N=1/2\Delta t$ , главный частотный диапазон вычислений от  $0$  до  $2f_N$ , оптимальный и достаточный шаг частотной

дискретизации для сохранения всей сигнальной информации и восстановления сигнала при обратном преобразовании без погрешности  $\Delta f = 1/N$ . Пример преобразования приведен на рис. 21.31.

$$N = 100 \quad n = 0..N-1 \quad M := N \quad m = 0..M-1 \quad \Delta f := \frac{1}{N} \quad s_n := \exp[-10^{-8}(|n-40|)^6] \cdot \cos(2\pi \cdot 0.08 \cdot n)$$

$$Sh_m := \frac{1}{N} \cdot \sum_{n=0}^{N-1} s_n \cdot \cos(2\pi \cdot m \cdot \Delta f \cdot n) \quad sh_n := \sum_{m=0}^{M-1} Sh_m \cdot \cos(2\pi \cdot m \cdot \Delta f \cdot n) \quad \Delta f1 := \Delta f \cdot 0.2 \quad Sh1_m := \frac{1}{N} \cdot \sum_{n=0}^{N-1} s_n \cdot \cos(2\pi \cdot m \cdot \Delta f1 \cdot n)$$

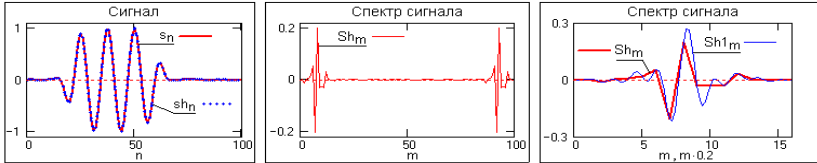


Рис. 21.31. Дискретное преобразование Хартли.

**Спектры числовых рядов.** При обработке произвольных рядов данных значение  $\Delta t$  по умолчанию равно 1 и формулы (21.75-21.76) применяются в следующем виде:

$$Sh(m) = (1/N) \sum_{n=0}^{N-1} s(n) \cos(2\pi m n/N), \quad (21.77)$$

$$s(n) = \sum_{m=0}^{M-1} Sh(m) \cos(2\pi m n/N). \quad (21.78)$$

Спектр сигналов является непрерывной функцией. Оптимальная дискретизация спектра при малом количестве данных может существенно исказить форму спектра, и для визуального просмотра шаг дискретизации может быть уменьшен, как это показано на рис. 21.31 для спектра  $Sh1$ . Однако при этом следует учитывать, что при уменьшении шага спектра в  $k$ -раз для восстановления сигнала из спектра требуется выполнять расчет с увеличением в  $k$  раз количества точек спектра  $M$ .

$$N := 100 \quad n = 0..N-1 \quad \Delta f := \frac{1}{N} \quad s_n := \exp[-10^{-8}(|n-40|)^6]$$

$$k := \frac{-N}{2} .. \frac{N}{2} \quad Sh(k) := \frac{1}{N} \cdot \sum_{n=0}^{N-1} s_n \cdot \cos(2\pi \cdot k \cdot \Delta f \cdot n)$$

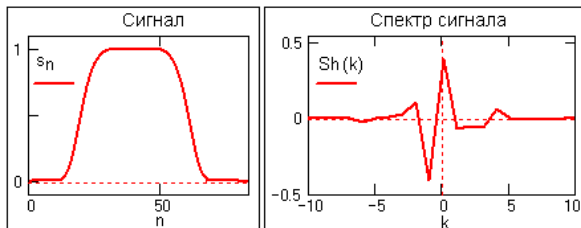


Рис. 21.32.

Это объясняется тем, что спектр Хартли в главном диапазоне не имеет избыточности, как это характерно для преобразования Фурье (комплексная сопряженность). В диапазоне главного периода  $[-M/2, M/2]$ , спектр Хартли в общем случае представляет собой произвольную функцию (рис. 21.32). При смещении расчетного диапазона спектра в интервал  $[0, M-1]$ , что выполняется для ускорения расчетов и исключения отрицательного индексирования отсчетов, восстановление сигнала по интервалу  $[0, M/2]$  невозможно, за исключением четных и нечетных функций, где информационная избыточность заложена в задании самой функции.

**Свойства дискретного преобразования** по своей сущности аналогичны свойствам непрерывного преобразования. С учетом смещения расчетного главного диапазона спектра в область положительных индексов на половину периода, значениям  $Sh(-m)$  соответствуют значения  $Sh(M-m)$  и, соответственно, несколько изменяются формулы свойств преобразования. Из чисто практических соображений построения алгоритмов расчетов заметим, что при  $m=0$  значениям  $Sh(M-m)$  должно соответствовать значение  $Sh(0)$ . Для сохранения общности алгоритмов это выполняется расчетом спектров в формулах (21.75) не до индекса  $M-1$  (при  $M=N$ ), а до индекса  $M$  (с сохранением предела суммирования  $M-1$  в формулах обратного преобразования).

Вычисление четной и нечетной составляющих:

$$Sh_{sym}(m) = [Sh(m)+Sh(M-m)]/2, \quad (21.79)$$

$$Sh_{asym}(m) = [Sh(m) - Sh(M-m)]/2. \quad (21.80)$$

Связь с преобразованием Фурье:

$$S(m) = Sh_{sym}(m) - j Sh_{asym}(M-m), \quad (21.81)$$

Энергетический и фазовый спектры:

$$Wh(m) = [Sh^2(m)+Sh^2(M-n)]/2. \quad (21.82)$$

$$\varphi(m) = \operatorname{arctg}(-[Sh(m) - Sh(M-n)] / [Sh(m)+Sh(M-n)]). \quad (21.83)$$

Сдвиг сигнала. Смещение на полпериода нумерации отсчетов вызывает изменение знака синусного члена:

$$s(n-n_0) \xleftarrow{\text{TH}} = \cos(2\pi mn_0/N) Sh(m) - \sin(2\pi mn_0/N) Sh(M-m). \quad (21.84)$$

Преобразование производной:

$$s'(n) = \xleftarrow{\text{TH}} (2\pi m/N) Sh(M-m). \quad (21.85)$$

Функция корреляции:

$$B_s(k) = \langle s(n), s(n+k) \rangle \xleftarrow{\text{TH}} 0.5 [Sh^2(m) + Sh^2(M-m)] = Wh_s(m). \quad (21.86)$$

Преобразование свертки:

$$s(n) * u(n) \xrightarrow{TH} 0.5N[Sh(m)Uh(m) - Sh(M-m)Uh(M-m) + Sh(m)Uh(M-m) + Sh(M-m)Uh(m)]. \quad (21.87)$$

**Цифровая фильтрация методом свертки.**

Выполнение свертки через преобразование Хартли полностью аналогично циклической свертке через преобразование Фурье:

$$s(n) = x(n) * y(n) \xrightarrow{TH} Xh(n)Yh(n) = Sh(n) \xrightarrow{TH} s(n). \quad (21.88)$$

$$\begin{aligned}
 N &:= 100 \quad n := 0..N-1 \quad M := N \quad m := 0..M \quad \Delta f := N^{-1} \\
 s_n &:= \exp[-10^{-8}(|n-40|)^6] \quad h_n := \exp\left(\frac{-n}{10}\right) \quad a := \sum_n h_n \quad h_n := \frac{h_n}{a} \\
 Sh_m &:= \frac{1}{N} \sum_{n=0}^{N-1} s_n \cdot \cos(2\pi \cdot m \cdot \Delta f \cdot n) \quad Hh_m := \frac{1}{N} \sum_{n=0}^{N-1} h_n \cdot \cos(2\pi \cdot m \cdot \Delta f \cdot n) \\
 S1h_m &:= 0.5 \cdot N \cdot (Sh_m \cdot Hh_m - Sh_{M-m} \cdot Hh_{M-m} + Sh_m \cdot Hh_{M-m} + Sh_{M-m} \cdot Hh_m) \\
 s1_n &:= \sum_{i=0}^{N-1} h_i \cdot \text{if}(n-i < 0, 0, s_{n-i}) \quad s2_n := \sum_{m=0}^{M-1} S1h_m \cdot \cos(2\pi \cdot m \cdot \Delta f \cdot n)
 \end{aligned}$$

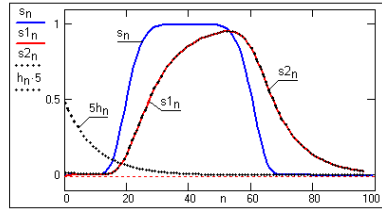


Рис. 21.33. Свертка функций во временной и спектральной области.

Как и для ДПФ, при фильтрации сигналов с выполнением свертки через частотную область ДПХ для исключения наложений от боковых периодов на сигнал главного периода интервал задания сигнала следует продлевать (нулями, значениями фона или тренда) на длину оператора фильтра, для симметричных операторов – с обеих сторон интервала. После выполнения этой операции длина самого оператора фильтра, которая обычно много меньше длины сигнала, также должна продлеваться до количества задания отсчетов сигнала с учетом его продления, т.к. умножение спектров сигнала и фильтра требует равного количества отсчетов их спектрального представления.

**Двумерная дискретная фильтрация.** Спектр Хартли функции  $s(k, n)$ , представленной в виде матрицы размера  $K \times N$  ( $k=k\Delta x$ ,  $n=n\Delta y$ ), имеет вид вещественной матрицы  $Sh(m, p)$  также размером  $K \times N$ , при этом прямые и обратные преобразования матриц записываются в виде:

$$Sh(m, p) = (1/KN) \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} s(k, n) \cos(2\pi kn/K + 2\pi kn/N), \quad (21.89)$$

$$s(k, n) = (1/KN) \sum_{m=0}^{K-1} \sum_{p=0}^{N-1} Sh(m, h) \cos(2\pi kp/K + 2\pi kp/N). \quad (21.90)$$

## **22. Многомерные сигналы и системы**

Обработка многомерных сигналов, используя в частных случаях методы обработки одномерных сигналов, имеет и существенные особенности. Это объясняется тремя факторами. Во-первых, математические методы описания многомерных систем далеки от совершенства и завершенности. Во-вторых, при решении многомерных задач используется значительно больший объем данных. И в третьих, многомерные системы обладают большим числом степеней свободы и, соответственно, значительно большей гибкостью. Так, например, при дискретизации информации в одномерном случае устанавливается только частота отсчетов, а в многомерном не только частота, но и форма раstra дискретизации. С другой стороны, многомерные полиномы разлагаются на множители только в частном случае, а, следовательно, многие одномерные методы не обобщаются на случай многомерных задач.

Ниже будут рассматриваться сигналы и системы с размерностью два и более, при этом основное внимание будет уделяться двумерным задачам, имеющим широкое распространение в практике. Повышение размерности выше двух не приводит к качественным отличиям от двумерных случаев, кроме повышения сложности вычислений.

Многомерная информация в своем абсолютном большинстве, это дискретная информация в цифровой форме – многомерные массивы данных. Многомерные непрерывные функции используются только в чисто теоретических исследованиях. Даже двумерных данных, непрерывных (аналоговых) по обоим аргументам практически не существует. С учетом этого ниже рассматриваются, в основном, многомерные сигналы в дискретной форме.

### **22.1. Двумерные и многомерные сигналы**

**Понятие многомерного сигнала.** Многомерные сигналы представляют собой функции  $P$  независимых переменных при  $P > 1$ . В общем случае, сигнал может быть непрерывным, дискретным или смешанным. Понятия непрерывности и дискретности аналогичны одномерным сигналам. Что касается смешанного сигнала, то это многомерный сигнал, который описывается функцией некоторого количества непрерывных и некоторого количества дискретных переменных. Пример смешанного двумерного сигнала: ансамбль непрерывных сигналов, изменяющихся во времени ( $t$  - вторая переменная), снимаемых с набора сейсмических приемников сейсмострассы (номера датчиков - первая переменная).

В общем случае, двумерный непрерывный сигнал представляет собой функцию, значения которой зависят от двух независимых переменных (аргументов, координат):

$$s(x,y) = \sin(x^2+y^2), \quad -\infty < x,y < \infty \quad (22.1)$$

График функции (в пределах одного периода) приведен на рис. 22.1.

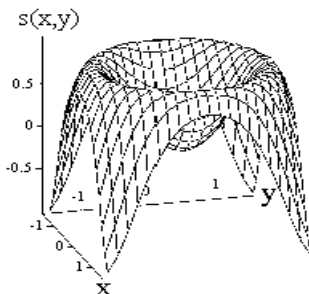


Рис. 22.1.

Двумерный дискретный сигнал (цифровой массив) - это функция, определенная на совокупности пар числовых значений координат с определенным шагом дискретизации  $\Delta x$  и  $\Delta y$ . В общем случае, при различной физической размерности аргументов  $x$  и  $y$ , значения  $\Delta x$  и  $\Delta y$  не равны друг другу:

$$s_{n,m} = s(n\Delta x, m\Delta y), \quad -\infty < n,m < \infty. \quad (22.2)$$

Элемент последовательности  $s_{n,m}$  представляет собой отсчет двумерной функции  $s$  в координатной точке  $(x=n\Delta x, y=m\Delta y)$ , где значения  $x$  и  $y$  – независимые переменные (аргументы) функции. Для числовых массивов значения шага дискретизации по аргументам также могут приниматься равными 1 (независимо от размерности) и использоваться аргументация  $s(n,m) \equiv s_{n,m}$ . Результаты геофизических съемок какого-либо одного геофизического параметра по поверхности земли относятся к двумерным функциям: дискретным - если это отсчеты в отдельных точках по определенной координатной сети  $(x,y)$ , или смешанным - если это непрерывная регистрация данных по профилям (например - мощности экспозиционной дозы гамма излучения горных пород при аэросъемке). Но в настоящее время геофизические съемки относятся даже не к двумерным, а к многомерным функциям, так как регистрируется, как правило, сразу несколько физических параметров геологических сред. Так, например, при спектрометрической съемке естественной радиоактивности горных пород регистрируется содержание в горных породах урана, тория и калия, в гравиразведке - трехкоординатный вектор силы тяжести, и т.п.



Если на какой-либо площади проведена съемка нескольких видов геофизики, то их результаты также могут рассматриваться в совокупности, как многомерная функция физических параметров данной геологической среды.

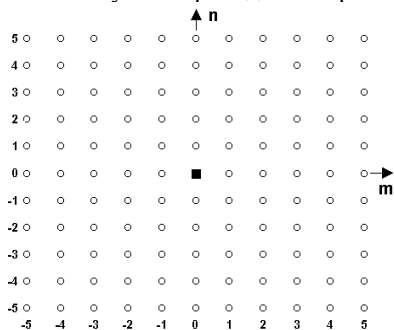
По определениям (22.1) двумерные функции и сигналы, равно как и многомерные, имеют бесконечную протяженность по координатам. На практике мы всегда имеем дело с конечными координатами наших данных. Учитывая это, будем считать, что значения наших сигналов за пределами определенных координат равны нулю.

Отметим некоторые двумерные последовательности (функции, сигналы), имеющие специальные названия.

**Двумерный единичный импульс**  $\delta(n\Delta x, m\Delta y) = \delta_{n,m}$  или единичный отсчет:

$$\begin{aligned} \delta_{n,m} &= 1, \text{ при } n = m = 0. \\ &= 0, \text{ при } n \neq 0, m \neq 0. \\ \delta_{n,m} &= \delta_n \delta_m, \end{aligned}$$

где  $\delta_n, \delta_m$  - одномерные единичные импульсы (импульсы Кронекера) по координатам  $n$  и  $m$ . Стилизованное графическое представление двумерного единичного импульса приведено на рис. 22.2.



Кружки - отсчеты 0, квадраты - отсчеты 1.

Рис. 22.2.

Произвольное расположение двумерного единичного импульса по координатам  $n_1, m_1$  соответственно записывается в виде:  $\delta((n-n_1)\Delta x, (m-m_1)\Delta y) = \delta_{n-n_1, m-m_1}$ . Попутно напомним, что математическая запись импульса Кронекера обозначает не единичный отсчет, а функцию, определяющую место положения единичного отсчета и нулевые значения по остальным координатам (аргументам).

**Двумерный линейный импульс** представляет собой последовательность единичных отсчетов по одной координате:  $s(n,m) = \delta(n)$  или  $s(n,m) = \delta(m)$ . На рис. 22.3 приведены два двумерных

линейных импульса, первый - по координате  $m = 0$ :  $s(n,m) = \delta(m)$ , и второй импульс по координате  $n = 2$ :  $s(n,m) = \delta(n-2)$ .

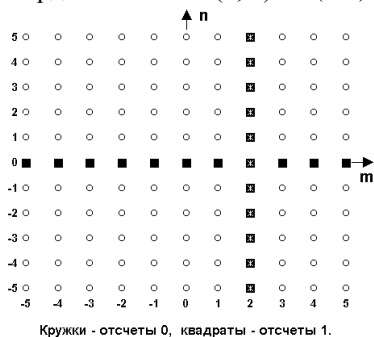


Рис. 22.3.

Очевидно, что для  $P$ -мерных случаев точно таким же образом могут быть определены  $P$ -мерные единичные импульсы,  $P$ -мерные линейные импульсы,  $P$ -мерные площадные импульсы и т.д., хотя понятие импульса, заимствованное из теории одномерных сигналов, здесь несколько не к месту.

**Двумерная единичная ступенька**  $u(n,m)$ , представленная на рис. 22.4, определяется выражением:

$$u(n,m) = 1, \text{ при } n \geq 0 \text{ и } m \geq 0, \\ = 0, \text{ в остальных случаях.} \\ u(n,m) = u(n) u(m),$$

где  $u(j)$  представляют собой единичные ступеньки соответственно по координатам  $n$  и  $m$ :  $u(j)=1$  при  $j \geq 0$ ,  $u(j)=0$  при  $j < 0$ . Двумерная единичная ступенька отлична от нуля в одном квадранте  $(n,m)$ -плоскости.

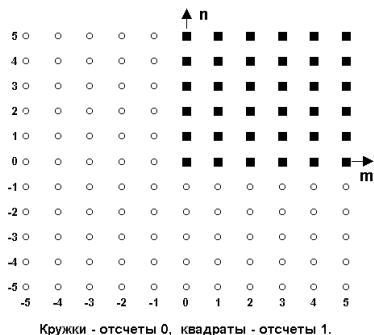


Рис. 22.4.

**Экспоненциальная последовательность:**

$s(n,m) = a^n b^m, -\infty < n,m < \infty$ , где  $a$  и  $b$  в общем случае комплексные

числа. При  $a = \exp(j\omega_1)$ ,  $b = \exp(j\omega_2)$ ,  $|a|=1$ ,  $|b|=1$ :

$$s(n,m) = \exp(jn\omega_1 + jm\omega_2) = \cos(n\omega_1 + m\omega_2) + j\sin(n\omega_1 + m\omega_2).$$

Экспоненциальные последовательности, как и в одномерном случае, являются собственными функциями двумерных линейных систем, инвариантных к сдвигу.

**Разделимые последовательности.**

Разделимой называют последовательность, которую можно представить в виде произведения одномерных последовательностей. Так, для двумерной разделимой последовательности:

$$s(n,m) = s(n) s(m).$$

Разделение возможно для немногих практических сигналов. Однако любое двумерное множество с конечным числом ненулевых отсчетов разлагается на конечную сумму разделимых последовательностей:

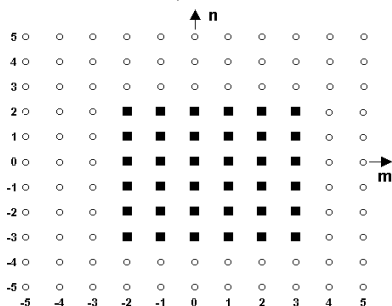
$$s(n,m) = \sum_{i=1}^N s_{i,n}(n) s_{i,m}(m),$$

где  $N$  - число ненулевых строк или столбцов массива. Для этого достаточно выразить  $s(n,m)$  в виде суммы отдельных строк:

$$s(n,m) = \sum_{i=1}^N s(n,i) \delta(m-i). \tag{22.3}$$

**Конечные последовательности.**

Важным классом сигналов являются последовательности конечной протяженности, для которых сигнал равен нулю вне определенной области, называемой *опорной областью* сигнала. На рис.22.5 условно представлена двумерная последовательность конечной протяженности, значения которой отличны от нулевых только внутри ограниченной прямоугольной области  $-3 \leq n \leq 2$ ,  $-2 \leq m \leq 5$ .



Кружки - отсчеты 0, квадраты - отсчеты 1.

Рис. 22.5.

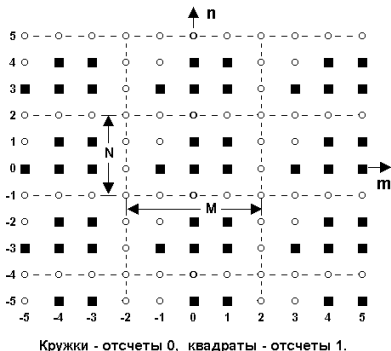
Опорная область сигнала может быть произвольной формы и выходить за пределы сигнала, частично включая нулевые отсчеты.

Отсчеты за пределами опорной области считаются равными нулю.

**Периодические последовательности.** Двумерные последовательности могут быть периодическими, регулярно повторяющимися в пространстве. Последовательность, удовлетворяющая условиям:

$$\begin{aligned} s(n,m+M) &= s(n,m), \\ s(n+N,m) &= s(n,m), \end{aligned} \quad (22.4)$$

обладает периодичностью в двух направлениях, по  $n$  и по  $m$ . Значения  $M$  и  $N$  называют интервалами периодичности сигнала соответственно по координатам  $m$  и  $n$  (*горизонтальными* и *вертикальными* интервалами периодичности). Прямоугольная форма области периода (пример на рис.22.6) наиболее удобна при обработке данных, но не является единственно возможной.



Кружки - отсчеты 0, квадраты - отсчеты 1.

Рис. 22.6.

Для двумерных последовательностей условия (22.4) могут рассматриваться как частный случай общих условий периодичности:

$$s(n+N_1, m+M_1) = s(n,m), \quad (22.5)$$

$$s(n+N_2, m+M_2) = s(n,m),$$

$$D = N_1M_2 - N_2M_1 \neq 0.$$

Упорядоченные пары  $(N_1, M_1)$  и  $(N_2, M_2)$  представляют собой смещения от отсчетов одного периода к соответствующим отсчетам других периодов и могут рассматриваться как векторы  $\mathbf{N}$  и  $\mathbf{M}$ , которые образуют области периодов в форме параллелограмма. Линейная независимость векторов обеспечивается при ненулевом определителе  $D$ , а количество отсчетов в пределах периода равно  $|D|$ . Пример периодической последовательности с векторами  $(4,4)$  и  $(3,-5)$  приведен на рис. 22.7.

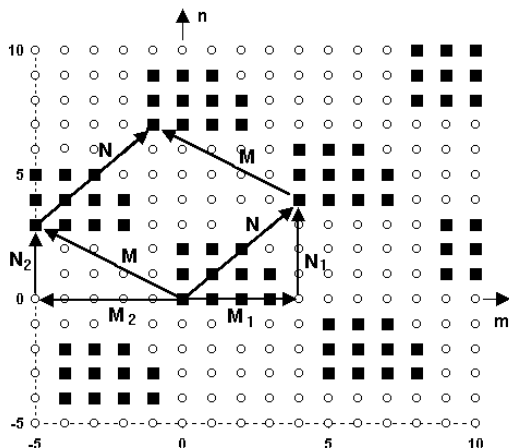


Рис. 22.7.

Понятие периодичности можно обобщить на многомерные сигналы.  $P$ -мерный сигнал  $s(\bar{n})$  будет представлять собой  $P$ -мерную периодическую последовательность, если существует  $P$  линейно независимых  $P$ -мерных целочисленных  $\mathbf{N}$ -векторов периодичности, с которыми выполняется условие:

$$s(\bar{n}) = s(\bar{n} + \bar{\mathbf{N}}_i), \quad i = 1, 2, 3, \dots, P. \quad \bar{\mathbf{P}}$$

Столбцы векторов  $\mathbf{N}_i$  образуют матрицу периодичности  $\mathbf{N}$  размером  $P \times P$ . Векторы периодичности матрицы линейно независимы при наличии у матрицы ненулевого определителя. Абсолютное значение определителя равно числу отсчетов в периоде. Последовательность  $s(\bar{n})$  прямоугольно периодична для случаев диагональной матрицы  $\mathbf{N}$ . Если функция  $s(\bar{n})$  периодична с матрицей периодичности  $\mathbf{N}$ , то для любого целочисленного вектора  $\mathbf{P}$  имеет место  $s(\bar{n} + \mathbf{P} \mathbf{N}) = s(\bar{n})$ , и матрица  $\mathbf{P} \mathbf{N}$  также будет матрицей периодичности для  $s(\bar{n})$ . Отсюда следует, что любая многомерная периодическая последовательность имеет не единственную матрицу периодичности.

## 22.2. Двумерные системы

Системы осуществляют преобразование сигналов. Формализованная система - это оператор (операция) отображения входного сигнала на выходной:  $z(x,y) = T[s(x,y)]$ .

**Базовыми операциями** в системах, комбинациями которых осуществляются преобразования, являются операции скалярного

умножения, сдвига и сложения:

$$\begin{aligned} z(n,m) &= c s(n,m), \\ z(n,m) &= s(n-N,m-M), \\ z(n,m) &= s(n,m)+u(n,m). \end{aligned}$$

Используя базовые операции, любую двумерную последовательность можно разложить на сумму взвешенных двумерных единичных импульсов:

$$s(n,m) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} s(i,j) \delta(n-i,m-j). \quad (22.6)$$

Обобщением скалярного умножения является пространственное маскирование:

$$z(n,m) = c_{n,m} s(n,m). \quad (22.7)$$

Правая часть равенства (22.7) представляет собой поэлементное произведение входного сигнала на совокупность чисел  $c_{n,m}$ .

Кроме линейных операций в системах используются также безынерционные нелинейные преобразования с независимым нелинейным воздействием на значения отсчетов входной последовательности. Пример операции - возведение в квадрат:

$$z_{n,m} = (s_{n,m})^2.$$

**Линейные системы.** Система считается линейной при выполнении двух условий:

1. Пропорциональное изменение входного сигнала вызывает пропорциональное изменение выходного сигнала.
2. Суммарный сигнал двух входных последовательностей дает суммарный сигнал двух соответствующих выходных последовательностей.

Другими словами, если оператор  $T[s(x,y)]$  описывает линейную систему и имеет место

$$z(x,y) = T[s(x,y)], \quad q(x,y) = T[u(x,y)],$$

то

$$T[as(x,y)+bu(x,y)] = az(x,y)+bq(x,y).$$

Линейные системы подчиняются принципу суперпозиции сигналов.

В выражении (22.6) значения  $s(i, j)$  можно рассматривать как скалярные множители для соответствующих единичных импульсов. Применяя оператор преобразования  $T[.]$  к левой и правой части (22.6), получаем:

$$T[s(n,m)] = y(n,m) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} s(i,j) T[\delta(n-i,m-j)],$$

$$z(n,m) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} s(i,j) h_{ij}(n,m), \quad (22.8)$$

где  $h_{ij}(n,m)$  - отклик системы в точке  $(n,m)$  на единичный импульс в точке  $(i,j)$ . Если импульсный отклик  $h_{ij}(n,m)$  определен для всех точек  $(i,j)$ , то отклик системы на произвольный многомерный сигнал, как и для одномерных систем, находится с помощью суперпозиции.

**Инвариантность к сдвигу.** Система инвариантна к сдвигу, если сдвиг входной последовательности приводит к такому же сдвигу выходной последовательности:

$$T[s(n-N,m-M)] = z(n-N,m-M).$$

Линейность и инвариантность к сдвигу являются независимыми свойствами системы. Так, пространственное маскирование линейно, но не инвариантно к сдвигу, а безынерционные операторы нелинейны, но инвариантны к сдвигу.

В дальнейшем ограничимся рассмотрением только систем, широко распространенных при решении практических задач - линейных и инвариантных к сдвигу (ЛИС-системы).

**Импульсный отклик** на произвольно расположенный входной импульс, как следует из выражения (22.8), описывается выражением:

$$h_{ij}(n,m) = T[\delta(n-n_i, m-m_j)].$$

Для частного случая  $i = j = 0$  имеем:

$$h_0(n,m) = T[\delta(n,m)].$$

Используя принцип инвариантности к сдвигу, получим:

$$h_{ij}(n,m) = h_0(n-i, m-j) = h(n-i, m-j), \quad (22.9)$$

т.е. импульсный отклик на произвольно расположенный входной импульс равен сдвинутому импульсному отклику на входной импульс, расположенный в начале координат.

**Двумерная свертка.** Подставляя (22.9) в выражение (22.8), получаем:

$$z(n,m) = \sum_i \sum_j s(i,j) h(n-i, m-j). \quad (22.10)$$

Двумерная дискретная свертка (22.10), является аналогом одномерной дискретной свертки. При замене переменных  $n-i = k$ ,  $m-j=l$ , получим:

$$z(n,m) = \sum_k \sum_l h(k,l) s(n-k, m-l), \quad (22.11)$$

т.е. двумерная свертка коммутативна, как и одномерная. В такой же мере она обладает свойством ассоциативности по отношению к последовательности операций свертки нескольких функций (результат не зависит от порядка свертки) и свойством дистрибутивности по отношению к операции свертки с суммой функций (результат аналогичен сумме сверток с каждой функцией). Эти свойства

определяют и основное свойство двумерных (и многомерных) линейных систем при их параллельном и/или последовательном соединении – результирующая система также является линейной.

Для упрощения символьного аппарата двумерную свертку обозначают индексом (\*\*):

$$z(n,m) = h(k,l) ** s(n-k,m-l).$$

При обобщении этого выражения на многомерные системы, в векторной форме:

$$z(\bar{n}) = h(\bar{k}) ** s(\bar{n} - \bar{k}).$$

**Разделимые системы.** Если импульсный отклик системы может быть разделен:

$$h(k,l) = h(k) h(l), \tag{22.12}$$

то выражение (22.11) принимает вид:

$$z(n,m) = \sum_k h(k) \sum_l h(l) s(n-k,m-l), \tag{22.13}$$

или:

$$z(n,m) = \sum_k h(k) g(n-k,m), \quad g(n-k,m) = \sum_l h(l) s(n-k,m-l).$$

Массив  $g(n,m)$  вычисляется одномерной сверткой столбцов массива  $s(n,m)$  при  $n = \text{const}$  (сечения массива по координатам  $n$ ) с откликом  $h(l)$ , с последующим вычислением выходного массива  $z(n,m)$  одномерной сверткой строк  $g(n,m)$  при  $m = \text{const}$  с откликом  $h(k)$ . Результат не изменится, если сначала выполнять свертку по строкам, а затем по столбцам. Система с откликом вида (22.12) называется ***разделимой***. Отметим, что в разделимой системе входной и выходной сигнал не обязаны быть разделимыми. Аналогичные разделимые системы могут существовать и в многомерном варианте.

**Устойчивость систем.** Интерес для практики представляют только устойчивые системы, обеспечивающие определенный конечный результат системной операции на конечные входные сигналы. Необходимым и достаточным условием устойчивости системы является абсолютная суммируемость ее импульсного отклика:  $\sum_k \sum_l |h(k,l)| < \infty$ .

**Специальные двумерные системы.** На практике используются также системы с несколькими входами и/или выходами.

Допустим, система имеет  $i$ -входы и  $j$ -выходы, линейна и инвариантна к сдвигу по переменной  $t$ . Если на  $i$ -вход системы поступает одномерный единичный импульс  $\delta_i(t)$  при нулевых сигналах на остальных входах, то  $j$ -выходные сигналы будут импульсным откликом системы  $h_{ij}(t)$ . При известном полном ансамбле значений  $h_{ij}$  для всех  $i$ -входов, для произвольной комбинации входных сигналов  $s_i(t)$  сигнал на  $j$ -выходе будет определяться выражением:

$$z_j(t) = \sum_i \sum_k h_{ij}(k) s_i(t-k). \tag{22.14}$$



### 22.3. Частотные характеристики сигналов и систем

**Частотный отклик системы.** Допустим, что двумерная ЛИС-система имеет импульсный отклик  $h(k\Delta x, l\Delta y)$ . Подадим на вход системы сигнал вида комплексной синусоиды:

$$s(n, m) = \exp(jn\Delta x \cdot \omega_x + jm\Delta y \cdot \omega_y),$$

где  $\omega_x$  и  $\omega_y$  – значения частоты сигнала соответственно по координатам  $x$  и  $y$ . Принимая  $\Delta x = 1$ ,  $\Delta y = 1$  и выполняя двумерную свертку (22.10), получаем:

$$\begin{aligned} z(n, m) &= h(k, l) \exp[j \cdot \omega_x \cdot (n-k) + j \cdot \omega_y \cdot (m-l)] = \\ &= \exp(jn \cdot \omega_x + jm \cdot \omega_y) h(k, l) \exp(-jk \cdot \omega_x - jl \cdot \omega_y) = H(\omega_x, \omega_y) \exp(jn \cdot \omega_x + jm \cdot \omega_y). \end{aligned}$$

$$H(\omega_x, \omega_y) = h(k, l) \exp(-jk \cdot \omega_x - jl \cdot \omega_y). \quad (22.15)$$

Таким образом, выходной сигнал представляет собой комплексную синусоиду с теми же значениями частоты, что и у входного сигнала, с изменением амплитуды и фазы за счет комплексного множителя  $H(\omega_x, \omega_y)$ , который носит название частотного отклика (частотной характеристики) системы. Для дискретных сигналов частотный отклик периодичен с периодом  $2\pi$  по обоим частотным переменным:

$$H(\omega_x + 2\pi k, \omega_y + 2\pi l) = H(\omega_x, \omega_y).$$

При разделимости импульсного отклика частотный отклик многомерных систем также является разделимой функцией:

$$\begin{aligned} h(k, l) &= q(k)g(l) \Leftrightarrow Q(\omega_x)G(\omega_y) = H(\omega_x, \omega_y) \\ Q(\omega_x) &= \sum_k q(k) \exp(-jk \cdot \omega_x). \\ G(\omega_y) &= \sum_l g(l) \exp(-jl \cdot \omega_y). \end{aligned}$$

#### **Пример расчета частотного отклика системы.**

Определить частотную характеристику системы с импульсным откликом:

$$h(0, 0) = 0.25, \quad h(0, \pm 1) = 0.125, \quad h(\pm 1, 0) = 0.125, \quad h(\pm 1, \pm 1) = 0.0625.$$

$$\begin{aligned} \text{Частотный отклик: } H(\omega_x, \omega_y) &= \sum_{n=-1}^1 \sum_{m=-1}^1 h(n, m) \exp(-jn\omega_x - jm\omega_y) = \\ &= 0.25 + 0.125[\exp(-j\omega_x) + \exp(j\omega_x) + \exp(-j\omega_y) + \exp(j\omega_y)] + \\ &\quad + 0.0625[\exp(-j\omega_x - j\omega_y) + \exp(-j\omega_x + j\omega_y) + \exp(j\omega_x - j\omega_y) + \exp(j\omega_x + j\omega_y)] = \\ &= 0.25(1 + \cos \omega_x)(1 + \cos \omega_y). \end{aligned}$$

Система является примером двумерного фильтра нижних частот. Частотный отклик системы на плоскости  $(\omega_x, \omega_y)$ , приведенный на рис. 22.8, имеет осевую симметрию с коэффициентом передачи 1 в центре  $(\omega_x=0, \omega_y=0)$  со спадом до нуля при  $\omega_x = \pm \pi$  и  $\omega_y = \pm \pi$ .

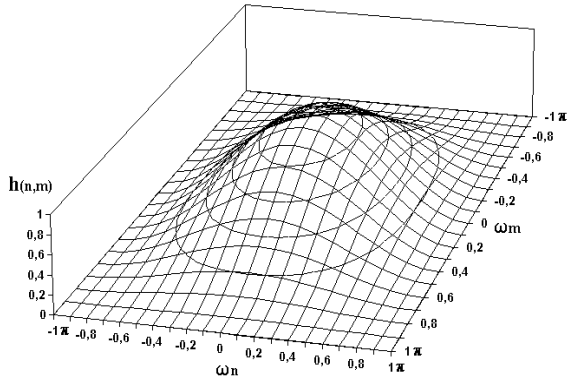


Рис. 22.8. Частотная характеристика ФНЧ.

**Импульсный отклик системы.** Выражение (22.15) описывает разложение функции  $H(\omega_x, \omega_y)$  в двумерный ряд Фурье с коэффициентами разложения в виде отсчетов импульсного отклика  $h(k,l)$ , т.е. прямое преобразование Фурье. Очевидно, что обратным преобразованием Фурье с интегрированием в пределах одного периода из частотного отклика  $H(\omega_x, \omega_y)$  можно получить импульсный отклик системы:

$$h(k,l) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(\omega_x, \omega_y) \exp(jk\omega_x + jl\omega_y) d\omega_x d\omega_y. \quad (22.16)$$

---

**Пример расчета импульсного отклика фильтра.**

Определить импульсный отклик идеального фильтра низких частот с прямоугольной частотной характеристикой вида:  $H(\omega_x, \omega_y) = 1$  при  $|\omega_x| \leq a < \pi$ ,  $|\omega_y| \leq b < \pi$ ;  $H(\omega_x, \omega_y) = 0$  в остальных случаях.

Импульсный отклик: 
$$h(k,l) = \frac{1}{4\pi^2} \int_{-a}^a \int_{-b}^b \exp(jk\omega_x + jl\omega_y) d\omega_x d\omega_y.$$

Система разделима: 
$$h(k,l) = \frac{1}{2\pi} \exp(jk\omega_x) d\omega_x \cdot \frac{1}{2\pi} \exp(jl\omega_y) d\omega_y = \frac{\sin ak}{\pi k} \cdot \frac{\sin bl}{\pi l}.$$

---

**Пример расчета неразделимого импульсного отклика.**

Определить импульсный отклик идеального кругового фильтра нижних частот:

$$H(\omega_x, \omega_y) = 1 \text{ при } \omega_x^2 + \omega_y^2 < R^2 < \pi^2; \quad H(\omega_x, \omega_y) = 0 \text{ в остальных случаях.}$$

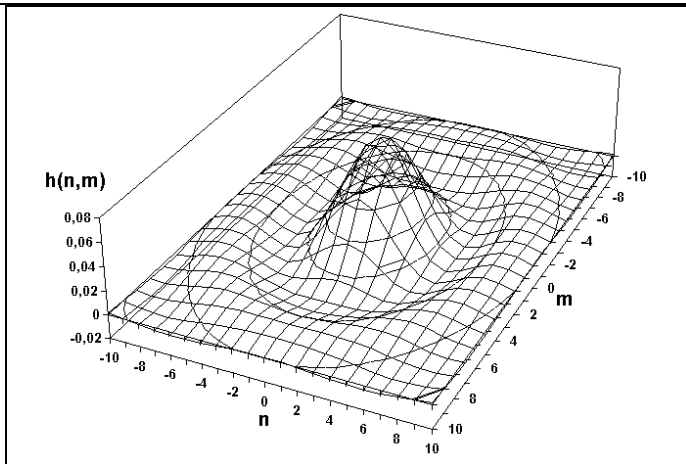
Вычисления по круговой области целесообразно выполнять в полярных координатах:  $\omega = \sqrt{\omega_x^2 + \omega_y^2}$ ,

$\varphi = \text{arctg}(\omega_y/\omega_x)$ ,  $\phi = \text{arctg}(m/n)$ , при этом выражение 22.16 переписывается в следующем виде:

$$\begin{aligned} h(n,m) &= \frac{1}{4\pi^2} \int_0^R \int_0^{2\pi} \omega \exp[j\omega \cos(\phi-\varphi)] d\varphi d\omega = \\ &= \frac{1}{2\pi} \int_0^R \omega J_0(\omega) d\omega = (R/2\pi) J_1(R) / , \end{aligned}$$

где  $J_0(\dots)$ ,  $J_1(\dots)$ - функции Бесселя 1-го рода 0-го и 1-го порядков соответственно.

На рис. 22.9 приведена пространственная форма импульсного отклика фильтра, расчет которой проведен при  $R=1$  с ограничением по  $N=10$  и  $M=10$ , и сечения отклика по координате  $m$ .



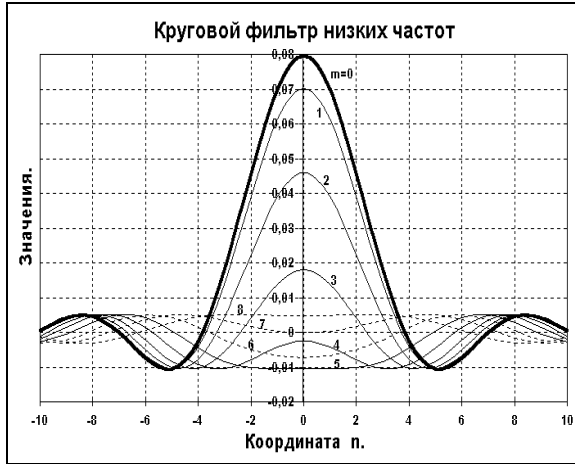


Рис. 22.9. Круговой низкочастотный фильтр (справа - сечения по координате m).

**Свойства двумерного преобразования Фурье.**

Вышеприведенные преобразования импульсного отклика в частотный отклик и наоборот представляют собой двумерные дискретные преобразования Фурье с прямоугольным растром дискретизации информации, эквивалентные одномерным преобразованиям. На двумерные преобразования с прямоугольным растром переносятся и другие свойства одномерных систем. В частности:

1. Фурье-преобразования сигналов.

$$S(\omega_x, \omega_y) = \sum_n \sum_m s(n, m) \exp(-jn\omega_x - jm\omega_y). \tag{22.17}$$

$$s(n, m) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} S(\omega_x, \omega_y) \exp(jn\omega_x + jm\omega_y) d\omega_x d\omega_y. \tag{22.18}$$

2. Теорема о свертке.

$$z(n, m) = h(n, m) ** s(n, m) \Leftrightarrow H(\omega_x, \omega_y) S(\omega_x, \omega_y) = Z(\omega_x, \omega_y).$$

$$z(n, m) = c(n, m) s(n, m) \Leftrightarrow C(\omega_x, \omega_y) ** S(\omega_x, \omega_y) = Z(\omega_x, \omega_y).$$

3. Основные свойства Фурье-преобразования.

1) Линейность (в том числе для любых комплексных чисел a и b):

$$a \cdot s(n, m) + b \cdot z(n, m) \Leftrightarrow aS(\omega_x, \omega_y) + bZ(\omega_x, \omega_y).$$

2) Пространственный сдвиг:

$$s(n-N, m-M) \Leftrightarrow S(\omega_x, \omega_y) \exp(-jN\omega_x - jM\omega_y).$$

3) Дифференцирование:

$$dS(\omega_x, \omega_y)/d\omega_x \Leftrightarrow -jn s(n, m),$$

$$dS(\omega_x, \omega_y)/d\omega_y \Leftrightarrow -jm s(n, m),$$

$$d^2S(\omega_x, \omega_y)/(d\omega_x d\omega_y) \Leftrightarrow -nm s(n, m).$$

4) Комплексное сопряжение:

$$x^*(n,m) \leftrightarrow S^*(-\omega_x, -\omega_y).$$

Вещественная и мнимая части Фурье-образов последовательностей  $s(n,m)$ :

$$S(\omega_x, \omega_y) = S^*(-\omega_x, -\omega_y).$$

$$\text{Re} [S(\omega_x, \omega_y)] = \text{Re} [S(-\omega_x, -\omega_y)].$$

$$\text{Im} [S(\omega_x, \omega_y)] = -\text{Im} [S(-\omega_x, -\omega_y)].$$

5) Теорема Парсеваля:

$$\sum_n \sum_m s(n,m) s^*(n,m) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} S(\omega_x, \omega_y) S^*(\omega_x, \omega_y) d\omega_x d\omega_y.$$

В частности, при  $s(n,m) = s(n,m)$ :

$$\sum_n \sum_m |s(n,m)|^2 = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |S(\omega_x, \omega_y)|^2 d\omega_x d\omega_y,$$

где левая часть уравнения представляет собой полную энергию дискретного сигнала  $s(n,m)$ , а функция  $|S(\omega_n, \omega_m)|^2$  - спектральную плотность энергии сигнала.

## **22.4. Дискретизация двумерных сигналов**

**Прямоугольный растр дискретизации.** Из способов обобщения одномерной периодической дискретизации на двумерный случай наиболее простым является периодическая дискретизация в прямоугольных координатах:

$$s(n,m) = s_a(n\Delta x, m\Delta y),$$

где  $\Delta x$  и  $\Delta y$  - горизонтальный и вертикальный интервалы дискретизации двумерного непрерывного сигнала  $s_a(x,y)$  с непрерывными координатами  $x$  и  $y$ . Ниже значения  $\Delta x$  и  $\Delta y$ , как и в одномерном случае, принимаются равными 1.

Дискретизация двумерного, а в общем случае и многомерного сигнала, также приводит к периодизации его спектра и наоборот. Сохраняется также и условие информационной равноценности координатной и частотной представлений дискретного сигнала при равном количестве точек дискретизации в главных диапазонах сигнала. Для прямоугольной дискретизации связь фурье-преобразований непрерывного и дискретного сигналов устанавливается аналогично одномерной дискретизации.

**Интегральные преобразования Фурье** аналоговых сигналов в непрерывной шкале частот  $\Omega_x$  и  $\Omega_y$ :

$$S_a(\Omega_x, \Omega_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s_a(x,y) \exp(-j\Omega_x x - j\Omega_y y) dx dy. \quad (22.19)$$

$$s_a(x,y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_a(\Omega_x, \Omega_y) \exp(j\Omega_x x + j\Omega_y y) d\Omega_x d\Omega_y. \quad (22.20)$$

**Дискретные преобразования Фурье:**

$$S(k,l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} s(n,m) \exp(-jn2\pi k/N - jm2\pi l/M), \quad (22.21)$$

$$S(k,l) = \sum_{n=0}^{N-1} \exp(-jn2\pi k/N) \sum_{m=0}^{M-1} s(n,m) \exp(-jm2\pi l/M), \quad (22.22)$$

$$s(n,m) = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} S(k,l) \exp(-jn2\pi k/N - jm2\pi l/M). \quad (22.23)$$

$$s(n,m) = \sum_{k=0}^{N-1} \exp(-jn2\pi k/N) \sum_{l=0}^{M-1} S(k,l) \exp(-jm2\pi l/M). \quad (22.24)$$

Выражения (22.22) и (22.24) показывают, что двумерное ДПФ по прямоугольному растру дискретизации данных может вычисляться с помощью одномерных последовательных ДПФ. Вторые суммы выражений являются одномерными ДПФ сечений функций  $s(n,m)$  и  $S(k,l)$  по линиям  $n$  и  $k$  соответственно, а первые - одномерными ДПФ вычисленных функций в сечениях по  $m$  и  $l$ . Другими словами, исходные матрицы значений  $s(n,m)$  и  $S(k,l)$  пересчитываются сначала в промежуточные матрицы с ДПФ по строкам (или по столбцам), а промежуточные - в окончательные с ДПФ по столбцам (или соответственно по строкам).

**Интерполяционный ряд восстановления двумерного сигнала.** Если непрерывный сигнал  $s_a(x,y)$  является сигналом с ограниченным спектром, а периоды дискретизации выбраны достаточно малыми и спектры соседних периодов не перекрываются:

$$S_a(\Omega_x, \Omega_y) = 0 \quad \text{при } |\Omega_x| \geq \pi/\Delta x, \quad |\Omega_y| \geq \pi/\Delta y,$$

то, как и в одномерном случае, сигнал  $s_a(x,y)$  может быть восстановлен по дискретному сигналу с использованием двумерного аналога ряда Котельникова-Шеннона:

$$s_a(x,y) = \sum_n \sum_m s(n,m) \frac{\sin \pi(x/\Delta x - n)}{\pi(x/\Delta x - n)} \frac{\sin \pi(y/\Delta y - m)}{\pi(y/\Delta y - m)}. \quad (22.25)$$

Сигнал с неограниченным спектром также может быть дискретизирован, однако в этом случае имеет место наложение спектров в смежных периодах, при этом высокие частоты, большие частоты Найквиста, будут "маскироваться", как и в одномерном случае, под низкие частоты главного периода. Эффект "отражения" от границ периода дает еще более сложную картину вследствие

интерференции частот, отраженных по разным координатам.

**Произвольный растр дискретизации.** Понятие прямоугольной дискретизации обобщается на произвольный растр дискретизации с линейно независимыми векторами  $\mathbf{v}_1 = (v_{11}, v_{21})^T$  и  $\mathbf{v}_2 = (v_{12}, v_{22})^T$ , где T - индекс транспонирования (рис. 22.10).

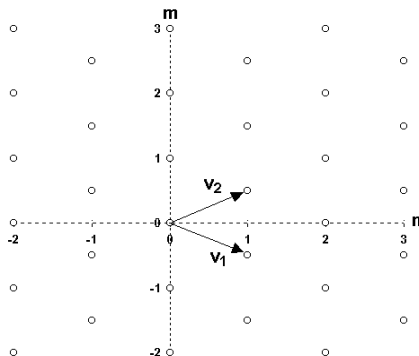


Рис. 22.10.

Координаты двумерного периодического множества отсчетов на плоскости (x,y):

$$\begin{aligned} x &= v_{11}n + v_{12}m, \\ y &= v_{21}n + v_{22}m. \end{aligned}$$

С использованием векторных обозначений:

$$\bar{\mathbf{z}} = \bar{\mathbf{V}} \bar{\mathbf{n}}$$

где  $\bar{\mathbf{z}} = (x,y)^T$ ,  $\bar{\mathbf{n}} = (n,m)^T$ ,  $\bar{\mathbf{V}} = (\mathbf{v}_1 | \mathbf{v}_2)$  - матрица дискретизации. Определитель матрицы  $\bar{\mathbf{V}}$  не равен нулю, если вектора  $\mathbf{v}_1$  и  $\mathbf{v}_2$  линейно независимы. При дискретизации непрерывного сигнала  $s_a(x,y)$  матрицей  $\bar{\mathbf{V}}$  формируется дискретный сигнал:

$$s(\bar{\mathbf{n}}) \Leftarrow s_a(\bar{\mathbf{n}} \bar{\mathbf{V}}).$$

**Двумерное интегральное преобразование Фурье**

непрерывного сигнала по непрерывному вектору  $\bar{\Omega} = (\Omega_1, \Omega_2)^T$ :

$$S_a(\bar{\Omega}) = \int_{-\infty}^{\infty} s_a(\bar{\mathbf{z}}) \exp(-j \bar{\Omega}^T \bar{\mathbf{z}}) d\bar{\mathbf{z}}, \quad (22.26)$$

$$s_a(\bar{\mathbf{z}}) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} S_a(\bar{\Omega}) \exp(j \bar{\Omega}^T \bar{\mathbf{z}}) d\bar{\Omega}, \quad (22.27)$$

Данные интегралы являются двойными, поскольку дифференциалы  $d\bar{\mathbf{z}}$  и  $d\bar{\Omega}$  являются векторами.

**Преобразование Фурье дискретного сигнала:**

$$S(\bar{\omega}) = \sum_n s(\bar{\mathbf{n}}) \exp(-j \bar{\omega}^T \bar{\mathbf{n}}), \quad (22.28)$$

$$s(\bar{\mathbf{n}}) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} S(\bar{\omega}) \exp(j \bar{\omega}^T \bar{\mathbf{n}}) d\bar{\omega}. \quad (22.29)$$

где:  $\bar{\omega} = (\omega_x, \omega_y)^T$ .

Выражение  $s(\bar{\mathbf{n}})$  может быть получено дискретизацией выражения  $s_a(\bar{\mathbf{z}})$  (13.4.7):

$$s(\bar{\mathbf{n}}) = s_a(\bar{\mathbf{V}} \bar{\mathbf{n}}) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} S_a(\bar{\Omega}) \exp(j \bar{\Omega}^T \bar{\mathbf{V}} \bar{\mathbf{n}}) d\bar{\Omega}.$$

После подстановки в это выражение значения  $\bar{\omega} = \bar{\Omega} \bar{\mathbf{V}}^T$ , получаем:

$$s(\bar{\mathbf{n}}) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \frac{1}{|\det \bar{\mathbf{V}}|} S_a(\bar{\omega} / \bar{\mathbf{V}}^T) \exp(j \bar{\omega}^T \bar{\mathbf{n}}) d\bar{\omega}.$$

Или, с учетом периодичности по квадратным областям плоскости  $\bar{\omega}$ :

$$s(\bar{\mathbf{n}}) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \frac{1}{|\det \bar{\mathbf{V}}|} \sum_{\bar{\mathbf{k}}} S_a((\bar{\omega} - 2\pi \bar{\mathbf{k}}) / \bar{\mathbf{V}}^T) \exp(j \bar{\omega}^T \bar{\mathbf{n}}) d\bar{\omega}, \quad (22.30)$$

где  $\bar{\mathbf{k}}$  - вектор целочисленных значений периодов дискретизированной функции по осям  $\omega_x$  и  $\omega_y$ . Сравнивая последнее выражение с выражением (22.29), получаем:

$$S(\bar{\omega}) = \frac{1}{|\det \bar{\mathbf{V}}|} \sum_{\bar{\mathbf{k}}} S_a((\bar{\omega} - 2\pi \bar{\mathbf{k}}) / \bar{\mathbf{V}}^T),$$

$$S(\bar{\Omega} \bar{\mathbf{V}}^T) = \frac{1}{|\det \bar{\mathbf{V}}|} \sum_{\bar{\mathbf{k}}} S_a(\bar{\Omega} - \bar{\mathbf{U}} \bar{\mathbf{k}}), \quad (22.31)$$

где  $\bar{\mathbf{U}}$  - матрица периодичности:

$$\bar{\mathbf{U}}^T \bar{\mathbf{V}} = 2\pi \bar{\mathbf{I}}, \quad (22.32)$$

которой задаются два линейно независимых вектора периодичности спектра,  $\bar{\mathbf{I}}$  - единичная матрица  $2 \times 2$ . Выражение (22.31) определяет связь между преобразованиями Фурье дискретных и аналоговых сигналов.

Как и в одномерном случае, интервалы дискретизации  $\Delta x$  и  $\Delta y$  определяют главный период двумерного спектра соответственно по осям  $\omega_x$  и  $\omega_y$  и частоты Найквиста:  $\omega_{xN} = \pi/\Delta x$  и  $\omega_{yN} = \pi/\Delta y$ . Спектр дискретного сигнала также является периодическим продолжением



спектра аналогового сигнала. Для исключения искажений спектра (наложения спектров боковых периодов на главный период) предельные частоты сигнала должны быть меньше частот Найквиста.

На рис. 22.11 приведен пример центральной части спектра дискретного сигнала при  $\Delta x=1$  и  $\Delta y=1$ .

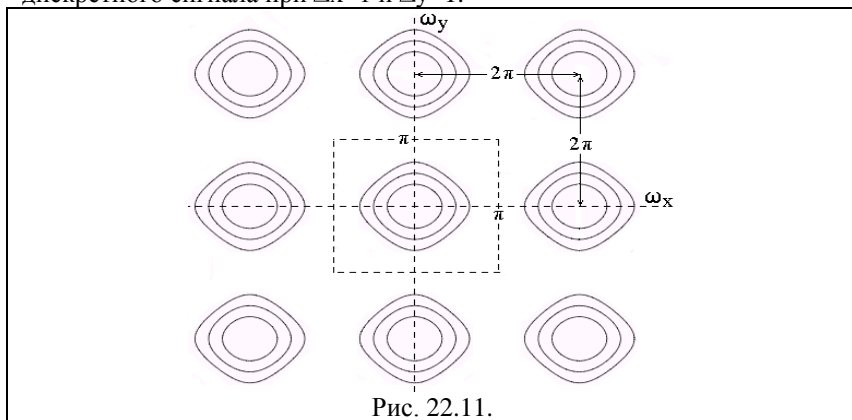


Рис. 22.11.

В случае прямоугольной дискретизации:

$$\bar{\mathbf{V}} = \begin{bmatrix} \Delta x & 0 \\ 0 & \Delta y \end{bmatrix}, \quad \det \bar{\mathbf{V}} = \Delta x \Delta y, \quad (22.33)$$

$$\bar{\mathbf{U}} = \begin{bmatrix} 2\pi/\Delta x & 0 \\ 0 & 2\pi/\Delta y \end{bmatrix}. \quad (22.34)$$

**Интерполяция дискретных сигналов.** Для сигнала с ограниченным спектром изменением матрицы дискретизации  $\bar{\mathbf{V}}$  можно подобрать матрицу периодичности  $\bar{\mathbf{U}}$  таким образом, чтобы в правой части выражения (22.31) не было перекрытия спектров. Тогда для значений по точкам  $\bar{\Omega} \bar{\mathbf{V}}^T$  области  $S$  главного периода спектра выражение (22.31) упрощается:

$$S(\bar{\Omega} \bar{\mathbf{V}}^T) = S_a(\bar{\Omega}) / |\det \bar{\mathbf{V}}|. \quad (22.35)$$

$$S_a(\bar{\Omega}) = |\det \bar{\mathbf{V}}| S(\bar{\Omega} \bar{\mathbf{V}}^T) = |\det \bar{\mathbf{V}}| S(\bar{\omega}), \quad \bar{\Omega} \in S. \quad (22.36)$$

Из выражения (22.36) следует, что при корректной дискретизации непрерывной двумерной функции ее спектр с точностью до нормировочного множителя  $|\det \bar{\mathbf{V}}|$  может быть восстановлен по спектру дискретной функции. Соответственно, выполнив обратное преобразование Фурье левой и правой части

равенства (22.36), получим уравнение восстановления непрерывной функции по ее дискретному варианту (многомерный аналог интерполяционного ряда Котельникова-Шеннона):

$$s_a(\bar{\mathbf{z}}) = \frac{|\det \bar{\mathbf{V}}|}{4\pi^2} \sum_{\bar{\mathbf{n}}} s(\bar{\mathbf{n}}) \exp(j \bar{\mathbf{\Omega}}^T (\bar{\mathbf{z}} - \bar{\mathbf{V}} \bar{\mathbf{n}})) d\bar{\mathbf{\Omega}}.$$

$$s_a(\bar{\mathbf{z}}) = \sum_{\bar{\mathbf{n}}} s(\bar{\mathbf{n}}) f(\bar{\mathbf{z}} - \bar{\mathbf{V}} \bar{\mathbf{n}}), \quad (22.37)$$

где  $f(\cdot)$  – интерполяционная функция:

$$f(\bar{\mathbf{z}} - \bar{\mathbf{V}} \bar{\mathbf{n}}) = \frac{|\det \bar{\mathbf{V}}|}{4\pi^2} \exp(j \bar{\mathbf{\Omega}}^T (\bar{\mathbf{z}} - \bar{\mathbf{V}} \bar{\mathbf{n}})) d\bar{\mathbf{\Omega}}. \quad (22.38)$$

Все приведенные векторные уравнения могут быть обобщены на  $P$ -мерные функции с заменой константы  $4\pi^2$  там, где она встречается, на  $(2\pi)^P$ .

**Прямоугольный и гексагональный растры дискретизации.** В принципе, сигнал с ограниченным спектром можно представить по различным растрам дискретизации. Выбор растра обычно производят из условия минимальной плотности отсчетов на плоскости, т.е. минимизацией величины  $|\det \bar{\mathbf{V}}|$ , при котором обеспечивается отсутствие наложений для частот анализируемых сигналов. На практике для двумерных сигналов используют, как правило, только два варианта растров дискретизации - прямоугольный и гексагональный. Прямоугольному варианту соответствуют диагональные матрицы дискретизации и периодичности (22.33-34). Для гексагональной дискретизации, пример которой приведен на рис. 22.10, в частном случае при  $\Delta t = \Delta x \sqrt{3}$  каждый отсчет располагается на равном расстоянии от шести ближайших отсчетов, при этом матрицы дискретизации:

$$\bar{\mathbf{V}} = \begin{bmatrix} 1 & 1 \\ \sqrt{3} & -\sqrt{3} \end{bmatrix}, \quad \bar{\mathbf{U}} = \begin{bmatrix} 2\pi & 2\pi \\ 2\pi/\sqrt{3} & -2\pi/\sqrt{3} \end{bmatrix}.$$

Допустим, имеем сигнал с частотным спектром, ограниченным круговой областью частот  $\omega_r$ :

$$S_a(\omega_x, \omega_y) = 0 \text{ при } \omega_x^2 + \omega_y^2 > \omega_r^2.$$

Круговая область частот вписывается без перекрытий в квадрат со стороной  $2\omega_r$  или в шестиугольник со стороной  $2\omega_r/\sqrt{3}$ . Матрицы дискретизации:

$$\bar{\mathbf{V}}_{\text{пр}} = \begin{bmatrix} \pi/\omega_t & 0 \\ 0 & \pi/\omega_t \end{bmatrix}, \quad \det \bar{\mathbf{V}} = \pi^2/\omega_r^2,$$

$$\bar{\mathbf{V}}_{\text{гекс}} = \begin{bmatrix} \pi/(\omega_r \sqrt{3}) & \pi/(\omega_r \sqrt{3}) \\ \pi/\omega_r & -\pi/\omega_r \end{bmatrix}, \quad \det \bar{\mathbf{V}} = 2\pi^2/(\omega_r^2 \sqrt{3}).$$

Поскольку плотность отсчетов пропорциональна  $1/|\det \bar{\mathbf{V}}|$ , то отсюда следует, что для представления одного и того же сигнала гексагональный растр дискретизации требует на 13.4% меньше отсчетов по сравнению с прямоугольным. Эффективность "гексагональной" матрицы возрастает при увеличении размерности сигнала. Так, при 4-мерном сигнале для "гексагональной" матрицы требуется в 2 раза меньше отсчетов, чем для "прямоугольной".

## **22.5. Многомерный спектральный анализ**

**Периодические последовательности.** Двумерная последовательность  $s_n(n,m)$  прямоугольно периодична, если

$$s_n(n,m) = s(n,n+M) = s(n+N,m)$$

для всех  $(n,m)$  при целочисленных значениях  $N$  и  $M$ . Минимальные значения  $N$  и  $M$ , при которых выполняется данное равенство, называют горизонтальным и вертикальным периодами функции  $s_n(\cdot)$ , которыми ограничивается прямоугольная область  $R_{N,M}$  главного периода, содержащая  $NM$  независимых отсчетов:  $0 \leq n \leq N-1, 0 \leq m \leq M-1$ .

Последовательность  $s_n(n,m)$  с периодами  $N$  и  $M$  можно представить в виде конечной суммы (ряда Фурье) комплексных синусоид с кратными частотами:

$$s_n(n,m) = (1/NM) \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} S_n(k,l) \exp(j2\pi nk/N + j2\pi ml/M). \quad (22.39)$$

$$S_n(k,l) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} s_n(n,m) \exp(-j2\pi nk/N - j2\pi ml/M). \quad (22.40)$$

**Пример.** Разложить в ряд Фурье периодический сигнал:

$$s_n(n,m) = \delta(n,m), \quad 0 \leq n \leq 4, 0 \leq m \leq 2$$

(единичные импульсы с периодом:  $N = 5, M = 3$ ).

$$S_n(k,l) = \sum_{n=0}^4 \sum_{m=0}^2 \delta(n,m) \exp(-j2\pi nk/5 - j2\pi ml/3) = 1 \text{ для всех } k \text{ и } l,$$

т.е. равномерная частотная характеристика в главном диапазоне. Сам сигнал может быть записан в виде двумерного ряда Фурье:

$$s_n(n,m) = (1/20) \sum_{k=0}^4 \sum_{l=0}^2 \exp(j2\pi nk/5 + j2\pi ml/3).$$

**Конечные последовательности.** Если  $s(n,m)$  представляет собой последовательность конечной протяженности, имеющей опорную область  $R_{N,M}$ , то периодическую последовательность  $s_n(n,m)$  с главным периодом  $R_{N,M}$  можно сформировать периодическим продолжением  $s(n,m)$ :

$$\begin{aligned} s_n(n,m) &= s(n-aN,m-bM), \\ s(n,m) &= s_n(n,m), \quad (n,m) < R_{N,M}. \\ &= 0, \text{ в остальных случаях.} \end{aligned}$$

Отсюда следует, что любой финитный сигнал может быть полностью определен своим периодическим продолжением и опорной областью.

Аналогично можно записать и для частотной области:

$$\begin{aligned} S_n(k,l) &= \sum_a \sum_b S(k-aN,l-bM). \\ S(k,l) &= S_n(k,l), \quad 0 \leq k \leq N, \quad 0 \leq l \leq M \\ &= 0, \quad \text{в остальных случаях.} \end{aligned}$$

Отсюда значения  $s(n,m)$  и  $S(k,l)$  можно вычислить с использованием выражений (22.39-40) путем последовательности операций:

$$s(n,m) \Rightarrow s_n(n,m) \Leftrightarrow S_n(k,l) \Rightarrow S(k,l).$$

Практически это означает, что для получения ДПФ последовательности конечной протяженности достаточно из выражений (22.39-40) для рядов Фурье убрать знак периодичности, при этом следует помнить, что вычисление отсчетов  $s(\dots)$  вне опорной области приведет к вычислению значений не отсчетов  $s(\dots)$ , а отсчетов  $s_n(\dots)$  периодического продолжения сигнала  $s(\dots)$ .

Таким образом:

1. Дискретизация сигнала в пространственной области вызывает периодизацию частотного спектра сигнала.
2. Дискретизация частотного спектра сигнала вызывает периодизацию его пространственного представления.
3. Прямое и обратное ДПФ сигнала ограниченной протяженности автоматически означает периодизацию как его спектра, так и его пространственного представления.
4. Сигналы, ограниченные в пространстве, можно точно отобразить отсчетами их фурье-преобразования.
5. Частотное представление сигнала с ограниченным спектром обратным фурье-преобразованием может быть точно переведено в пространственную область.
6. Ограниченность как пространственного сигнала, так и его спектра является обязательным условием корректного ДПФ, т.к. в противном случае периодизация сигнала может привести к искажению

его спектрального и пространственного представления.

**Многомерные последовательности.** Определение ДПФ для Р-мерной последовательности с опорной областью  $R_P = \{ \bar{n} : 0 \leq n_i \leq N_i - 1, i=1,2,3, \dots, P \}$  производится введением диагональной матрицы значений  $N_i$ :

$$\bar{N} =$$

при этом Р-мерное ДПФ записывается в виде:

$$S(\bar{k}) = s(\bar{n}) \exp(-j \bar{k}^T 2\pi \bar{n} / \bar{N}). \quad (22.41)$$

$$s(\bar{n}) = \frac{1}{|\det \bar{N}|} S(\bar{k}) \exp(j \bar{k}^T 2\pi \bar{n} / \bar{N}). \quad (22.42)$$

Кратко рассмотрим особенности многомерных ДПФ (на примере двумерных последовательностей).

ДПФ суммы двух последовательностей с опорной областью на  $R_{N,M}$  равно сумме их ДПФ:

$$as(n,m) + bz(n,m) \Leftrightarrow aS(k,l) + bZ(k,l),$$

но при этом все ДПФ должны быть одного размера и этот размер должен быть достаточным, чтобы включить всю опорную область суммарной последовательности  $as(n,m) + bz(n,m)$ . Практически это означает, что  $x(\cdot)$  и  $z(\cdot)$  должны иметь одну и ту же опорную область. Опорная область каждой последовательности при необходимости дополняется нулями.

Операция свертки двух функций в пространственной области отображается операцией умножения фурье-образов функций в частотной области, однако при этом линейная свертка полных пространственных сигналов при ее вычислении через ДПФ в силу периодического продолжения пространственных функций переходит в циклическую свертку (как и для одномерных сигналов). Результат свертки зависит от периодов  $N$  и  $M$ .

Допустим, что  $s(n,m)$  имеет опорную область  $R_{P1,P2}$ , а  $h(n,m)$  -  $R_{Q1,Q2}$ . Результат линейной свертки:

$$s(n,m) = \sum_k \sum_l h(k,l) s(n-k,m-l).$$

Опорная область последовательности  $s(n,m)$ :

$$0 \leq n \leq P1+Q1-1, \quad 0 \leq m \leq P2+Q2-1.$$

Следовательно, наложения периодов результата свертки не произойдет и циклическая свертка в главном частотном диапазоне будет равна линейной свертке при опорной области ДПФ:

$$N \geq P1+Q1-1, \quad M \geq P2+Q2-1.$$

**Пример.**

Заданы последовательности (начало координат в нижнем левом углу):  $s(n,m) =$  ,  $h(n,m) =$  .

Вычислить свертку  $z(n,m) = s(n,m) ** h(n,m)$ .

ДПФ размера  $2 \times 2$  для  $s(n,m)$ .

$$S(0,0) = s(0,0) + s(1,0) + s(0,1) + s(1,1) = 2+1+1+0 = 4$$

$$S(1,0) = s(0,0) - s(1,0) + s(0,1) - s(1,1) = 2 -1+1 -0 = 2$$

$$S(0,1) = s(0,0) + s(1,0) - s(0,1) - s(1,1) = 2+1 -1 -0 = 2$$

$$S(1,1) = s(0,0) - s(1,0) - s(0,1) + s(1,1) = 2 -1 -1+0 = 0$$

После аналогичного вычисления  $H(k,l)$  и перемножения

$$S(k,l) = S(k,l) H(k,l):$$

$$S(k,l) = , H(k,l) = , Z(k,l) = .$$

После обратного ДПФ размера  $2 \times 2$  получим результат циклической свертки:  $z(n,m) =$  .

Дополним опорные области  $s(\cdot)$  и  $h(\cdot)$  до размера  $4 \times 4$  (для исключения искажения спектра, в принципе, достаточен размер  $3 \times 3$ ), и повторим вычисления:

$$s(n,m) = , h(n,m) = .$$

$$S(k,l) = , H(k,l) = .$$

$$S(k,l) = . s(n,m) = .$$

Сравнение данного результата с ДПФ размером  $2 \times 2$  позволяет наглядно видеть эффект цикличности свертки.

---

В настоящее время имеются разнообразные и весьма эффективные алгоритмы ДПФ. Для прямого вычисления  $P$ -мерного ДПФ требуется  $(N_1 \times N_2 \times \dots \times N_P)^2$  операций умножения и сложения. Для многомерного ДПФ, как и для одномерного, существуют алгоритмы быстрых преобразований Фурье. Простейший из них в двумерном ДПФ - разбиение на строки и столбцы, который мы уже рассматривали. Аналогично,  $P$ -мерное ДПФ может заменяться  $P$ -операциями одномерных ДПФ, при этом общее количество операций умножения и сложения сокращается.

## **23. Телекоммуникации и связь**

### **23.1. Основные сведения**

Телекоммуникационные системы и системы сетевой связи начали широко развиваться в процессе информационно-технологической революции и компьютеризации всех сфер хозяйственной деятельности человечества. Их основная задача – объединение компьютеров и других устройств в сетевые ассоциации для коллективного использования вычислительных, информационных и других ресурсов.

По территориальной принадлежности различают локальные сети (предприятий и организаций), региональные (районы и города) и глобальные. В глобальных сетях выделяют сети доступа (соединяющие близко расположенные друг к другу узлы) и магистральные сети (объединяющие узлы друг с другом).

Любая работоспособная сеть включает в себя определенную сеть каналов связи (первичная сеть) и коммуникационное оборудование. Характерные понятия сети:

*Протокол* - набор правил взаимодействия, определяющих способ кодирования информации и передачи служебных данных.

*Интерфейс* - устройство для приема-передачи информации и набор методов взаимодействия приемо-передающих устройств, в том числе между программными компонентами.

*Маршрутизация* - определение маршрута передачи сообщения в сложных сетях.

К основным задачам формирования и обработки сигналов на этапе сетевой приемо-передачи можно отнести:

*Кодирование* – изменение формы представления информации с целью ее передачи, хранения или обработки.

*Физическое кодирование* – представление данных в виде изменений во времени какой-либо физической величины – напряжения, амплитуды или фазы гармонических колебаний.

*Мультиплексирование* – использование одного канала связи для передачи данных нескольких абонентов.

**Территориальная сеть связи (ТКС)** – это географически распределенная сеть передачи данных (СПД), обеспечивающая оперативный и надежный обмен информацией между абонентами сети. Общепринятое название обменной информации - *сообщение*. Главные показатели эффективности ТКС - верность и время доставки информации. Они зависят от пропускной способности каналов связи, числа и способов соединения каналов связи между абонентами, протоколов информационного обмена, и ряда других факторов.

В ТКС используются телефонные, телеграфные, телевизионные, спутниковые сети связи. В качестве линий связи применяются кабельные линии связи (от простейших телефонных до специальных коаксиальных и волоконно-оптических), радиорелейные линии связи, и радиолинии. Среди кабельных линий

связи наилучшие показатели имеют световоды. Они имеют высокую пропускную способность передачи данных (сотни мегабит в секунду), и нечувствительны к внешним электромагнитным полям при отсутствии собственных электромагнитных излучений.

**Линия связи** состоит из физической среды, по которой передаются информационные сигналы, и аппаратуры передачи данных. Синоним термина - канал связи. Основной тип сигналов – электрические и электромагнитные.

В зависимости от среды передачи данных линии связи разделяются на проводные (воздушные), кабельные (в том числе волоконно-оптические) и радиоканалы наземной и спутниковой связи.

Проводные линии связи представляют собой открытые провода без изоляции и экранов. Они имеют низкую помехозащищенность и используются, в основном, для передачи телефонных и телеграфных сигналов.

Кабели представляют собой несколько проводников, заключенных в экранирующую и изолирующую оплетки. В компьютерных сетях применяются три типа кабелей: кабели на основе скрученных пар медных проводов, коаксиальные и волоконно-оптические кабели.

Радиоканалы имеют передатчики и приемники радиоволн, и отличаются друг от друга частотным диапазоном, который определяет дальность радиосвязи. Для компьютерной связи используются диапазоны УКВ и СВЧ, но для организации каналов в этих диапазонах необходима прямая видимость между передатчиком и приемником, или ретрансляция.

**Разновидности каналов связи.** Выделяют три основных разновидности каналов связи:

- симплексный - при связи приемника с передатчиком по одному каналу, с односторонней передачей информации (например, в телевизионной и радиовещательной сетях);
- полудуплексный - когда два узла связи соединены одним каналом, по которому информация передается попеременно то в одном направлении, то в противоположном (в информационно-справочных и запросно-ответных системах);
- дуплексный - позволяет передавать данные одновременно в двух направлениях за счет использования четырехпроводной линии



связи (два провода для передачи, два других – для приема данных), или двух полос частот.

Для повышения достоверности передачи данных основной канал может снабжаться дополнительным (обратным) каналом небольшой пропускной способности (на 1-2 порядка меньше основного), который используется для передачи служебной информации. По этому каналу передаются, например, сигналы подтверждения приема блоков данных и запросы на повторную передачу блоков при обнаружении ошибок.

Различают также выделенные (некоммутируемые) и коммутируемые на время передачи информации каналы связи. При использовании выделенных каналов связи приемопередающая аппаратура узлов связи постоянно соединена между собой. Этим обеспечивается высокая степень готовности системы к передаче информации, более высокое качество связи. Для коммутируемых каналов связи, создаваемых только на время передачи фиксированного объема информации, характерна небольшая стоимость, однако при этом имеют место потери времени на установление связи между абонентами (блокировки по приоритету и очередности).

**Передача информации по каналам связи.** В сетях ТКС информация передается в аналоговой форме. Это единственно возможный способ передачи информации по каналам связи. При передаче цифровых данных выполняется цифро-аналоговое и аналого-цифровое преобразование (ЦАП и АЦП) модемами на выходных и входных узлах связи. Для безошибочной передачи цифровых данных работа принимающего модема должна быть синхронизирована с работой передающего. Для этого используются два вида передачи данных: асинхронная и синхронная.

**Асинхронная передача** реализуется по символно - ориентированной схеме. Каждая передаваемая последовательность состоит из стартового бита, за которым следуют информационные символы, и завершается стоповым битом. Асинхронный режим передачи используется для низкоскоростных устройств и устройств, у которых отсутствует буфер.

**Синхронная передача** применяется для высокоскоростной передачи данных. При символно - ориентированной синхронной передаче блоку передаваемых символов предшествует один или несколько синхронизирующих символов. При побитно - ориентированной синхронной передаче в

передаваемый блок данных перед сообщением включается *флаг* - специальная битовая последовательность.

В зависимости от вида передачи данных используются соответственно синхронные и асинхронные модемы. Синхронная передача может проводиться только синхронными модемами, асинхронная может выполняться и с помощью синхронных модемов.

Применяется и гибридная схема передачи – ***изохронная***. Каждый символ в ней сопровождается стартовым и стоповым битами, а работа передающего и приемного модемов синхронизируется с помощью интервалов между передаваемыми символами.

Пересылка данных в ТКС осуществляется последовательной передачей битов сообщения от источника к пункту назначения. Физически информационные биты передаются в виде модулированных или импульсно-кодовых электрических сигналов, которые зачастую называют цифровыми. *Модулированные* сигналы менее чувствительны к искажениям, обусловленным затуханием в передающей среде. *Импульсно-кодовые* сигналы могут иметь одно или конечный набор значений в пределах определенного тактового интервала. Как правило, для импульсно-кодовой передачи используется двухуровневый сигнал.

**Системы телеобработки информации** являются специализированными системами телекоммуникаций. Основная цель систем телеобработки данных (СТД) – обеспечить прием данных непосредственно с мест их получения и выдачу результатов обработки к местам использования. При этом нет необходимости в промежуточных носителях данных, повышается оперативность взаимодействия с ЭВМ, повышается скорость и эффективность работы системы, для которой производится обработка данных. Телеобработка позволяет использовать мощные ЭВМ с большими базами данных. С помощью линий связи к таким ЭВМ может подключаться значительное число пользователей, что обеспечивает высокий уровень загрузки и использования ЭВМ.

Значительная протяженность линий связи затрудняет возможность обмена отдельными сигналами между ЭВМ и оконечным оборудованием. Поэтому взаимодействие ЭВМ и оконечного оборудования организуется с помощью сообщений – блоков данных, передаваемых в виде единого целого. Сообщения имеют специальную структуру, обеспечивающую представление в них наряду с собственно

данными служебной информации, необходимой для идентификации сообщения и защиты данных от искажений. Возможность взаимодействия абонентов с ЭВМ только посредством сообщений вносит определенную специфику в организацию программного обеспечения телеобработки.

## **23.2. Каналы связи**

**Аппаратура линий связи.** Формирование сигналов для передачи по линии связи осуществляется аппаратурой передачи данных (Data Circuit terminating Equipment). Примерами ДСЕ являются модемы. Подготовка данных для передачи осуществляется оконечным оборудованием данных (Data Terminal Equipment). Для усиления сигнала, проходящего через линию, а также для организации совместного использования линий связи (мультиплексирования и коммутации) может использоваться дополнительное оборудование.

Хотя сами сигналы в линиях связи всегда являются аналоговыми (непрерывными или кусочно-непрерывными), в зависимости от способа передачи данных линии связи делятся на цифровые и аналоговые. В цифровых линиях данные представляются сигналами, имеющими конечное число состояний, т.е. информация заключена в значениях сигнала в определенные моменты времени, причем сигнал может принимать конечное число значений. В аналоговых линиях используются сигналы с непрерывным диапазоном своих значений.

При передаче данных в аналоговой форме сигналы имеют более узкий спектр, поэтому их используют в линиях связи с узкой полосой пропускания, например в телефонных сетях. Цифровые сигналы имеют высокую скорость передачи данных, но более широкий спектр.

**Характеристики линий связи.** Основные характеристики канала связи – пропускная способность и достоверность передачи данных. *Пропускная способность канала* (количество информации, передаваемое в единицу времени) оценивается числом бит данных, передаваемых по каналу за единицу времени (в бит/с). *Достоверность передачи данных* оценивается по интенсивности битовых ошибок (Bit Error Rate), определяемой вероятностью искажения передаваемого бита данных. Величина BER для каналов связи без дополнительной защиты от ошибок составляет  $10^{-4}$  –  $10^{-6}$ . Основная причина искажений – воздействие помех на линию связи. Помехи, как правило, носят импульсный характер и имеют тенденцию к группированию – образованию пачек помех, искажающих сразу группу соседних бит в передаваемых данных.

Пропускная способность канала связи определяется полосой частот и помехоустойчивостью канала. Полоса частот, в которой амплитудно-частотная характеристика (АЧХ) линии связи имеет значение не ниже заданного (например, по уровню 0.5) называется *полосой пропускания*. Полоса частот  $\Delta F = f_v - f_n$ , где  $f_n$  и  $f_v$  – нижняя и верхняя границы частот, определяет диапазон частот, эффективно передаваемых по линии. Полоса частот зависит от типа линии и ее протяженности. Проводные линии связи имеют полосу частот примерно 10 кГц, кабельные – 100 кГц, коаксиальные – 100 МГц, радиорелейные – 1000 МГц, волоконно-оптические – 100 ГГц. Коротковолновая радиосвязь для передачи данных использует диапазон частот от 3 до 30 МГц.

Помехоустойчивость линии зависит от мощности помех, создаваемых в линии внешней средой или возникающих в самой линии. Обычно для уменьшения помех проводники экранируют или скручивают. Наименее помехоустойчивыми являются радиолнии, хорошей помехоустойчивостью обладают кабельные линии, отличной – волоконно-оптические линии, не восприимчивые к электромагнитному излучению.

**Пропускная способность канала** зависит от ширины полосы частот линии связи и отношения мощностей сигнала и шума. Математически эта связь описывается формулами Шенона (14.2.1) и Найквиста (14.2.2). Максимальная пропускная способность канала, построенного на основе линии с полосой частот  $\Delta F$  и отношением сигнал-шум  $P_c/P_{ш}$ , составляет (бит в секунду):

$$C_{\max} = \Delta F \log_2(1 + P_c/P_{ш}). \quad (23.1)$$

Значение  $(1 + P_c/P_{ш})$  определяет число уровней сигнала, которое может быть воспринято приемником. Так, если отношение  $P_c/P_{ш} > 3$ , то единичный сигнал может переносить четыре значения, т. е. 2 бита информации.

$$C = F \log_2(M), \quad (23.2)$$

где  $M$  – число различных состояний сигнала.

Практическая пропускная способность канала зависит от способа физического кодирования информации. При этом стремятся выбрать способ кодирования таким, чтобы максимально использовать возможности линии. Возможность применения на линии того или иного способа кодирования определяется полосой пропускания и затуханием сигнала.

*Затухание* - это отношение амплитуд входного и выходного сигналов на заданной частоте для определенной длины линии связи, выражается в децибелах и вычисляется по формуле

$$A = 20 \log_{10}(A_{\text{вых}}/A_{\text{вх}}).$$

При передаче данных широко используются двоичные сигналы, принимающие значения 0 и 1. Минимальная длительность такта, с которой могут передаваться сигналы по каналу с полосой частот  $\Delta F$ , равна  $T_{\min} = 1/(2\Delta F)$ . Если вероятность искажения символов 0 и 1 из-за помех одинакова и равна  $p$ , то число двоичных символов, которые можно безошибочно передать по каналу в секунду:

$$C = 2\Delta F[1 + p \log_2 p + (1-p) \log_2(1-p)]. \quad (23.3)$$

Это выражение определяет пропускную способность двоичного канала. Величина в квадратных скобках определяет долю двоичных символов, которые передаются по каналу с частотой  $2\Delta F$  без искажений. Если помехи отсутствуют, вероятность искажения символа  $p=0$  и пропускная способность  $C=2\Delta F$ . Если вероятность искажения  $p=0,5$ , то пропускная способность  $C=0$ . Если по каналу передается сообщение длиной  $n$  двоичных символов, то вероятность появления в нем  $m$  ошибок

$$P(n, m) = C_n^m p^m (1-p)^{n-m}.$$

Телефонный канал имеет полосу пропускания 3,1 кГц (диапазон частот 0,3-3,4 кГц). Коммутируемый телефонный канал обеспечивает скорость передачи данных  $C=1200$  бит/с, некоммутируемый – до 9600 бит/с.

Стандартизированы следующие скорости передачи данных по каналам связи: 200, 300, 600, 1200, 2400, 4800, 9600, 12000, 24000, 48000 и 96000 бит/с. Каналы с пропускной способностью до 300 бит/с называются низкоскоростными, от 600 до 4800 бит/с – среднескоростными, и с большей пропускной способностью – высокоскоростными.

Пропускная способность зависит не только от физического кодирования, но и от предварительного логического кодирования, которое заключается в предварительной подготовке данных, влияющей на ширину спектра итогового сигнала.

**Типы кабелей.** В компьютерных сетях применяются кабели, соответствующие стандартам. Наиболее употребительным является международный стандарт ISO/IEC 11801. В стандартах должны быть регламентированы следующие основные характеристики кабелей:

- Затухание (в дБ/м).
- Устойчивость кабеля к внутренним источникам помех. Если в кабеле проложены две пары проводов, одна из которых используется для передачи, а другая для приема, то возникают перекрестные наводки. Для оценки перекрестных наводок используется показатель NEXT (Near End Cross Talk), рассчитываемый по формуле  $20\log_{10}(A_{\text{ввых}}/A_{\text{вх}})$ , где  $A_{\text{ввых}}$  и  $A_{\text{вх}}$  –

амплитуды выходного и наведенного входного сигналов.

- Импеданс (волновое сопротивление) – эффективное входное сопротивление кабеля для переменного тока. При изменении напряжения реакция системы зависит не только от нагрузки на удаленном конце кабеля, а от эффективного сопротивления, которое определяется погонной емкостью и индуктивностью кабеля. Если нагрузка на кабеле не совпадает с его волновым сопротивлением (кабель рассогласован), то в кабеле возникают отражения сигнала от нагрузки (вплоть до возникновения стоячей волны при большом сопротивлении нагрузки), что приводит к невозможности передачи данных. Для обеспечения согласования необходимо, чтобы кабели в сети и концевые нагрузки имели одинаковое волновое сопротивление.
- Уровень внешнего электромагнитного излучения в проводнике, характеризующий помехозащищенность кабеля, то есть степень ослабления внешних помех от различных источников (линий электропередачи, средств связи, оргтехники и бытовой техники, электромоторов).

Наиболее широкое применение находят следующие типы кабелей.

**Неэкранированная витая пара UTP** является наиболее употребительной в локальных сетях и подразделяется по категориям. Кабели категории 3 и 4 имеют рабочий диапазон до 16 и 20 МГц, предназначены для передачи данных со скоростью до 10 и 15 Мбит/с. Кабели категории 5 - наиболее распространенный вид, ориентированы на рабочий диапазон до 100 МГц. Кабели категорий 6 и 7 имеют рабочий диапазон 200 и 600 МГц и лучшие характеристики затухания и помехозащищенности, но используются редко из-за своей дороговизны. Волновое сопротивление кабелей витой пары составляет 100 Ом. Все кабели выпускаются в 4-парном исполнении.

**Экранированная витая пара STP** имеет лучшие характеристики по сравнению с неэкранированной. Основным стандартом, определяющим параметры кабелей данного типа, является стандарт фирмы IBM, в котором кабели разделены на девять типов.

**Коаксиальные кабели** широко используются не только в компьютерных сетях, но и для передачи ВЧ телевизионных сигналов.

Кабель RG-8 и RG-11 – «толстый» коаксиальный кабель, имеет волновое сопротивление 50 Ом и внешний диаметр 2.5 см. Это дорогой кабель с высокими характеристиками.

Кабели RG-58/U (сплошной тонкий проводник), RG-58 A/U (многожильный проводник) и RG-58 C/U, RG-59 – тонкие коаксиальные кабели с волновым сопротивлением 50 Ом.

**Волоконно-оптические кабели** (ВОК) состоят из центрального проводника света (волокна), окруженного другим проводником – оболочкой. Оболочка обладает меньшим показателем преломления, чем сердцевина, поэтому излучение не выходит за пределы волокна.

Различают одномодовое волокно (дорогое, очень тонкого диаметра), с полосой пропускания сотни гигагерц, и многомодовое волокно, с более широким сердечником и меньшей полосой пропускания (500-800 МГц). В многомодовом волокне из-за относительно больших размеров электромагнитная волна высокой частоты может распространяться в нескольких режимах (модах), с разными скоростями, что приводит к искажениям передачи информации. Поэтому верхняя граничная частота такого волокна ограничена нижней частотой возникновения высших мод.

В качестве источников света в ВОК используют светодиоды и полупроводниковые лазеры. Для передачи информации используется свет с длиной волны 850-1300 нм.

**Способы передачи данных.** Для передачи данных используются способы с максимальным использованием свойств каналов по скорости и достоверности передачи данных.

Данные первоначально предоставляются последовательностью прямоугольных импульсов. Для их передачи без искажения требуется полоса частот от нуля до бесконечности. Реальные каналы имеют конечную полосу частот, с которой необходимо согласовать передаваемые сигналы. Согласование обеспечивается, во-первых, путем модуляции – переноса сигнала на заданную полосу частот и, во-вторых, путем кодирования – преобразовании данных в вид, позволяющий обнаруживать и исправлять ошибки, возникающие из-за помех в канале связи.

При использовании высокочастотных проводных и кабельных линий, полоса частот которых начинается практически от нуля, сигналы можно передавать в их естественном виде без модуляции. Каналы, работающие без модуляции, называются телеграфными и обеспечивают передачу данных со скоростью, как правило, 50-200 бит/с.

Если канал имеет ограниченную полосу частот (например, радиоканал), перенос сигнала в заданную полосу производится посредством модуляции. В этом случае между оконечным оборудованием данных, работающим с двоичными сигналами, и каналом устанавливается модем – модулятор и демодулятор.

*Модулятор* перемещает спектр первичного сигнала в окрестность несущей частоты  $f_0$ .

*Демодулятор* выполняет над сигналом обратное преобразование, формируя из модулированного сигнала импульсный двоичный сигнал.

Способы модуляции подразделяются на *аналоговые* и *дискретные*.

**Аналоговая модуляция** применяется для передачи дискретных данных по каналам с узкой полосой частот, например, по каналу тональной частоты телефонных сетей. Информация кодируется изменением амплитуды, частоты или фазы гармонического сигнала несущей частоты.

*Амплитудная модуляция (АМ)* выполняется модуляцией амплитуды несущей частоты информационным (модулирующим) сигналом. Амплитудную модуляцию кодовых последовательностей называют амплитудной манипуляцией (АМн). В зависимости от значения передаваемого двоичного элемента (бита) несущий сигнал может принимать одно из двух состояний: "включено" (наличие сигнала) – при передаче «1», либо "выключено" (отсутствие сигнала) – при передаче «0». Двоичную АМн называют *on-off* манипуляцией. В большинстве случаев амплитудную модуляцию объединяют с фазовой модуляцией (квадратурная модуляция), что повышает помехоустойчивость передачи данных и обеспечивает большую скорость передачи.

*Частотная модуляция (ЧМ)*. При ЧМ под воздействием модулирующих сигналов изменяется частота несущего колебания. В случае частотной манипуляции (ЧМн) несущая частота выбирается из дискретного набора значений частот в соответствии с передаваемым битом или битовой последовательностью. В общем случае, для ЧМн может использоваться  $M=2^n$  частот, каждая из которых представляет соответствующую  $n$ -битную последовательность. Так, в двоичной ЧМн используются две несущих частоты  $f_{01}$  и  $f_{02}$ , одна для двоичного нуля, другая – для двоичной единицы. Этот способ модуляции не требует сложных схем в модемах и обычно применяется в низкоскоростных модемах, работающих на скоростях 300 или 1200 бит/с. При квадратурной частотной манипуляции применяются четыре частоты для передачи битных последовательностей: 00, 01; 10; 11.

*Фазовая модуляция (ФМ)*. При ФМ в соответствии с последовательностью передаваемых информационных бит, изменяется фаза несущего синусоидального колебания  $f_0$ . В случае фазовой манипуляции (ФМн) фаза несущей частоты может принимать одно из нескольких дискретных значений. В двоичной схеме ФМн (биты 0 и 1) значение фазы несущей может равняться либо  $0^\circ$ , либо  $180^\circ$ , в одном периоде несущего сигнала передается один бит информации. Для



передачи двойных битов 00, 01, 10 и 11 в одном периоде сигнала используется квадратурная ФМн, при которой фаза может принимать одно из четырех значений соответственно:  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ ,  $315^\circ$  или  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ .

При декодировании сигналов с ФМн фаза сигнала в приемнике сравнивается с фазой несущего колебания, что усложняет демодуляцию. На практике используется *относительная фазовая манипуляция* (ОФМн), при которой каждому информационному биту ставится в соответствие не абсолютное значение фазы, а ее изменение относительно предыдущего значения.

В скоростных модемах используются комбинированные методы модуляции, например, амплитудная в сочетании с фазовой. Применяемые коды имеют достаточно большое число состояний, например 4 различных уровня амплитуды в сочетании с восьмью возможными значениями сдвига фазы (квадратурная амплитудная модуляция). Часть из этих состояний являются запрещенными, и за счет этой избыточности реализуется коррекция ошибок.

**Дискретная (цифровая) модуляция** применяются для преобразования аналоговых сигналов, например речевых, в цифровые. Для этих целей наиболее широко используются амплитудно-импульсная, кодово-импульсная и время-импульсная модуляция.

Кодирование передаваемых данных производится в основном для повышения помехоустойчивости данных. Так, первичные коды символов могут быть представлены в помехозащищенной форме – с использованием кодов Хемминга, обеспечивающих обнаружение и исправление ошибок в передаваемых данных. В последнее время функция повышения достоверности передаваемых данных возлагается на оконечное оборудование данных и обеспечивается за счет введения информационной избыточности в передаваемые сообщения.

При цифровом кодировании дискретной информации применяют потенциальные и импульсные коды. В потенциальных кодах для представления логических единиц и нулей используется только значение потенциала сигнала. Импульсные коды позволяют представить двоичные данные либо импульсами определенной полярности, либо перепадом потенциала определенного направления.

При выборе метода цифрового кодирования к нему предъявляют следующие требования:

- Наименьшая ширина спектра результирующего сигнала. Узкий спектр сигналов позволяет добиваться высокой скорости передачи данных. Кроме того, часто к спектру сигнала предъявляется

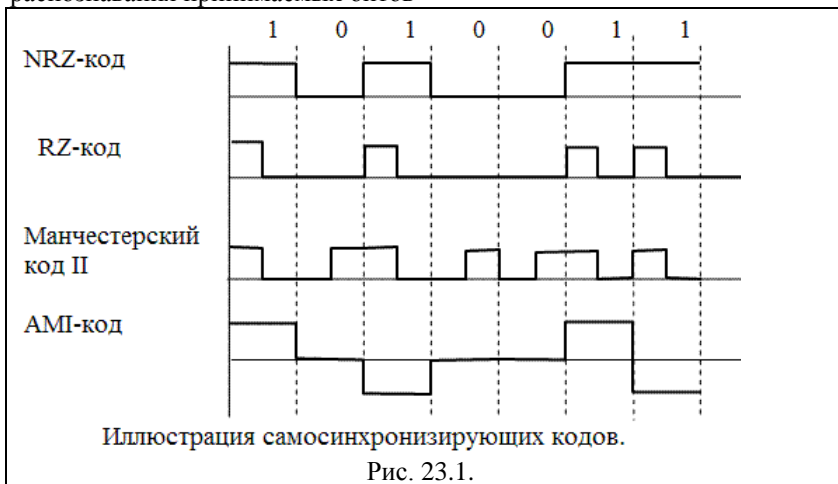
требование отсутствия постоянной составляющей.

- Возможность синхронизации между передатчиком и приемником. Как правило, в сетях применяются самосинхронизирующиеся коды, сигналы которых позволяют передатчику автоматически определять тактовую частоту передачи информационных битов (например, по резким перепадам сигналов).

- Возможность распознавания ошибок. Распознавание и коррекция ошибок реализуется средствами логического кодирования, или используется избыточность физических кодов.

**Способы цифрового кодирования данных.** В самосинхронизирующихся кодах каждый переход уровня сигнала от высокого к низкому уровню или наоборот используется для подстройки приемника. Лучшими считаются коды, которые обеспечивают переход уровня сигнала не менее одного раза в течение интервала времени на приеме одного информационного бита. Наиболее распространенными являются:

*Потенциальный код без возвращения к нулю* (NRZ - Non Return to Zero) – 0 и 1 кодируются различными уровнями сигнала. Это наиболее простой способ кодирования, но имеет постоянную составляющую в спектре. При передаче длинных серий одноименных битов (единиц или нулей) уровень сигнала остается неизменным для каждой серии, что снижает качество синхронизации и надежность распознавания принимаемых битов



*Потенциальный код с возвращением к нулю* (RZ - Return to Zero) – код, аналогичный NRZ, с возвращением к нулю на середине каждого тактового интервала. Имеет большее число переходов уровня

сигнала, чем сигнал в коде NRZ.

*Биполярное кодирование с альтернативной инверсией (AMI)* – 0 кодируется нулевым потенциалом, а 1 – положительным или отрицательным ненулевым, причем потенциал каждой следующей единицы противоположен по знаку предыдущей. Спектр кода не содержит постоянной составляющей. Используется три уровня сигналов, что требует увеличения мощности передатчика. Обладает хорошими синхронизирующими свойствами при передаче серий единиц и прост в реализации. Недостатком кода является ограничение на плотность нулей в потоке данных, поскольку длинные последовательности нулей ведут к потере синхронизации.

*Манчестерский код (PE - Phase Encode, фазовое кодирование)* – наиболее популярный код, применяемый в локальных сетях. При манчестерском кодировании информация передается перепадами потенциала, происходящими в середине такта. Единица кодируется перепадом от низкого уровня к высокому, а ноль наоборот. В начале каждого такта может происходить, а может и не происходить служебный перепад (он происходит, если в предыдущем такте передаваемый бит имел то же значение, что и в текущем.). Манчестерский код обеспечивает изменение уровня сигнала при представлении каждого бита и не имеет постоянной составляющей.

**Логическое кодирование** – это предварительное изменение передаваемой информации с целью сделать ее более удобной для передачи. Например, путем логического кодирования избавляются от длинных последовательностей нулей и единиц.

*Избыточное кодирование* – вариант логического кодирования, при котором к основной информации добавляется дополнительная. При этом передаваемая последовательность бит разбивается на порции, называемые символами, и каждый символ заменяется на новый, имеющий большее количество бит, чем исходный. Например, логический код 4В/5В заменяет исходные символы длиной 4 бита на символы длиной 5 бит. При этом возникает 16 «запрещенных» символов, прием которых свидетельствует об ошибке связи и появляется возможность исправления ошибок передачи.

**Методы синхронизации.** Применение самосинхронизирующихся кодов обеспечивает побитовую синхронизацию, но для обнаружения начала и окончания потока бит и его разделения на байты (побайтовая синхронизация) необходимы специальные методы.

В асинхронных протоколах для низкоскоростного взаимодействия с периферийными устройствами каждый байт (или группа бит фиксированной длины) предваряются специальным

набором старт-бит и завершаются стоп-битами.

В синхронных протоколах данные передаются большими блоками произвольной длины (кадрами). Кадр предваряется преамбулой - фиксированной последовательностью бит. После преамбулы начало и конец кадра обозначаются стартовыми и стоповыми ограничителями (флагами). Для того чтобы при возникновении внутри поля данных последовательности бит, совпадающей со стоповым ограничителем, не нарушался порядок приема кадра, существуют специальные методы, например, передача в заголовке кадра (имеющего фиксированную длину) сведений о длине поля данных.

### **Методы обнаружения искажений информации**

основаны на передаче в составе кадра избыточной информации, по которой можно судить о достоверности принятых данных. Эту служебную информацию называют последовательностью контроля кадра – Frame Check Sequence (FCS). Она вычисляется как функция от основной информации. Принимающая сторона заново вычисляет эту функцию, и совпадение результатов считается критерием правильности приема.

Метод контроля по паритету заключается в том, что к блоку информации добавляется один бит – 1, если число единиц в блоке нечетное, и 0, если четное. Он позволяет обнаруживать только нечетное число ошибок при передаче.

В методе циклического избыточного контроля (CRC) в качестве контрольной информации используется остаток от деления двоичного числа, которое представляют собой исходные данные, на известный делитель. Как правило, делитель выбирается таким, чтобы остаток составлял 2 или 4 байта. При приеме остаток прибавляется к принятым данным и полученное число делится на тот же делитель. Равенство итогового остатка нулю свидетельствует о правильности приема.

Восстановление потерянных кадров проводится путем их повторной передачи. При передаче с установлением логического соединения каждый блок данных нумеруется и для подтверждения его получения приемник посылает передатчику специальный блок данных - положительную или отрицательную квитанцию. Если передатчик не получил квитанции или получил отрицательную квитанцию, он проводит повторную передачу.

**Импульсно – кодовая передача данных** выполняется напрямую без модуляции, в первичной полосе частот, путем изменения уровней сигналов, несущих информацию, с определенной тактовой частотой. Например, если в ЭВМ цифровые

данные представляются сигналами уровнями 5 вольт - для кода «1» и 0,2 вольта - для кода «0», то при передаче этих данных в линию связи уровни сигналов преобразуются соответственно, например, в +12 и в -12 вольт на одном тактовом интервале. Такое кодирование осуществляется, в частности, с помощью асинхронных последовательных адаптеров RS-232-C при передаче цифровых данных от одного компьютера к другому на небольшие расстояния. Импульсно-кодовая передача является узкополосной, тактовая частота является, по существу, несущей частотой передачи.

### **23.3. Информационные характеристики источника дискретных сообщений**

Опираясь на формализованное описание сигналов и введенную меру количества информации, рассмотрим информационные характеристики источников сообщений и каналов связи, позволяющие установить пути повышения эффективности систем передачи информации, и, в частности, определить условия, при которых можно достигнуть максимальной скорости передачи сообщений по каналу связи как в отсутствие, так и при наличии помех.

Источники сообщений и каналы связи в системах передачи отличаются большим разнообразием по своей структуре и физической природе. Используются механические, акустические, оптические, электрические и радиоканалы. Для выяснения общих закономерностей необходимо абстрагироваться от их конкретного физического воплощения и оперировать формализованными понятиями источника сообщения и канала связи.

Источник дискретных сообщений формирует дискретные последовательности из ограниченного числа элементарных сообщений. На выходе источника непрерывных сообщений образуются непрерывные сообщения. Источник сообщений в теории информации полностью определяется статистическими данными о формируемых им сообщениях.

Если канал используется для передачи дискретных сообщений, он называется *дискретным каналом*. *Непрерывным* будем называть канал, предназначенный для передачи непрерывных сообщений.

Так как в процессе передачи дискретных сообщений модулятором в соответствии с поступающей последовательностью символов осуществляется изменение информативного параметра непрерывного (чаще всего высокочастотного) сигнала, генерируемого передатчиком в линию связи, то часть дискретного канала от выхода модулятора до входа демодулятора (рис. 23.2) является непрерывным каналом связи.

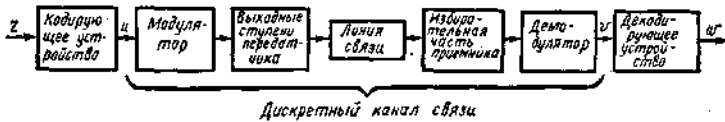


Рис. 23.2

Включив в состав этого канала на передающей стороне модулятор, а на приемной — демодулятор, получим дискретный канал, характеризующийся на входе множеством символов  $u$ , а на выходе — множеством символов  $v$ .

Если вредным действием помех в канале можно пренебречь, то для анализа используется модель в виде *идеализированного канала*, называемого *каналом без помех*. В идеальном канале каждому сообщению на входе однозначно соответствует определенное сообщение на выходе и наоборот.

Когда требования к достоверности велики и пренебрежение неоднозначностью связи между сообщениями  $z$  и  $w$  недопустимо, используется более сложная модель — *канал с помехами*.

Канал считается заданным, если известны статистические данные о сообщениях на его входе и выходе и ограничения, накладываемые на входные сообщения физическими характеристиками канала. Канал прямой передачи (от источника сообщений к их получателю), дополненный обратным каналом, например, для запроса повторной передачи в случае обнаружения ошибки, называют *каналом с обратной связью*.

**Модели источника дискретных сообщений.** Ранее речь шла о средней неопределенности и среднем количестве информации, приходящихся на одно состояние источника сообщений. Математической моделью множества возможных реализаций источника была дискретная или непрерывная случайная величина.

На практике, однако, нас чаще всего интересует не одно конкретное состояние источника, а дискретные или непрерывные последовательности состояний, реализуемых источником за длительный промежуток времени, например телеграммы, видеосюжеты и т. п. Для описания таких сообщений используются математические модели в виде дискретных и непрерывных случайных процессов.

Для построения модели необходимо знать объем  $l$  алфавита знаков  $(z_1, z_2, \dots, z_l)$ , из которых источником формируются сообщения, и вероятности создания им отдельных знаков с учетом возможной взаимосвязи между ними.

При доказательстве основных положений теории информации Шенноном использовалась модель, называемая *эргодическим источником сообщений*. Предполагается, что создаваемые им сообщения математически можно представить в виде эргодической случайной последовательности. Такая последовательность, как известно, удовлетворяет условиям стационарности и эргодичности. Первое означает, что вероятности отдельных знаков и их сочетаний не зависят от расположения последних по длине сообщения. Из второго следует, что статистические закономерности, полученные при исследовании одного достаточно длинного сообщения с вероятностью, близкой к единице, справедливы для всех сообщений, создаваемых источником. Из статистических характеристик в данном случае нас интересует средняя неопределенность в расчете на один знак последовательности.

Стационарный источник сообщений, выбирающий каждый знак формируемой последовательности независимо от других знаков, всегда является эргодическим. Его также называют источником без памяти.

На практике, однако, чаще встречаются источники, у которых вероятность выбора одного знака сообщения зависит от того, какие знаки были выбраны источником до этого (источники с памятью). Поскольку такая связь, как правило, распространяется на ограниченное число предыдущих знаков, для описания функционирования источника целесообразно использовать цепи Маркова.

Цепь Маркова порядка  $n$  характеризует последовательность событий, вероятности которых зависят от того, какие  $n$  событий предшествовали данному. Эти  $n$  конкретных событий определяют состояние источника, в котором он находится при выдаче очередного знака. При объеме алфавита знаков  $l$  число  $R$  различных состояний источника не превышает  $l^n$ . Обозначим эти состояния через  $S_1 \dots S_q \dots S_R$ , а вероятности выбора в состоянии  $S_q$  знака  $z_i$ , — через  $p_q(z_i)$ . При определении вероятности  $p_q(z_i)$  естественно предположить, что к моменту выдачи источником очередного знака известны все знаки, созданные им ранее, а следовательно, и то, в каком состоянии находится источник.

Если источник находится в состоянии  $S_q$ , его частная энтропия  $H(S_q)$  определяется соотношением

$$H(S_q) = - \sum_{i=1}^l p_q(z_i) \log p_q(z_i). \quad (23.4)$$

Усредняя случайную величину  $H(S_q)$  по всем возможным состояниям  $q = 1, 2, \dots, R$ , получаем энтропию источника сообщений:

$$H(Z) = - \sum_{q=1}^R p(S_q) \sum_{i=1}^l p_q(z_i) \log p_q(z_i), \quad (23.5)$$

где  $p(S_q)$  — вероятность того, что источник сообщений находится в состоянии  $S_q$ .

Величина  $H(Z)$  характеризует неопределенность, приходящуюся в среднем на один знак, выдаваемый источником сообщений.

Определим энтропию источника сообщений для нескольких частных случаев.

Если статистические связи между знаками полностью отсутствуют, то после выбора источником знака  $z_i$  его состояние не меняется ( $R = 1$ ). Следовательно,  $p(S_1)=1$ , и для энтропии источника сообщений справедливо выражение:

$$H(Z) = - \sum_{i=1}^l p(z_i) \log p(z_i).$$

Когда корреляционные связи наблюдаются только между двумя знаками (простая цепь Маркова), максимальное число различных состояний источника равно объему алфавита. Следовательно,  $R = l$  и  $p_q(z_i) = p(z_i/z_q)$ , где  $q = 1, 2, \dots, l$ . При этом выражение (23.5) принимает вид

$$H(Z) = - \sum_{q=1}^l p(z_q) \sum_{i=1}^l p(z_i/z_q) \log p(z_i/z_q). \quad (23.6)$$

При наличии корреляционной связи между тремя знаками состояния источника определяются двумя предшествующими знаками. Поэтому для произвольного состояния источника  $S_q$  удобно дать обозначение с двумя индексами  $S_{k,h}$ , где  $k = 1, 2, \dots, l$  и  $h = 1, 2, \dots, l$ .

Тогда

$$p(S_q) = p(S_{k,h}) = p(z_k, z_h) \text{ и } p_q(z_i) = p(z_i/z_k z_h).$$

Подставляя эти значения в (4.2), находим

$$H(Z) = - \sum_{k=1}^l \sum_{h=1}^l p(z_k z_h) \sum_{i=1}^l p(z_i/z_k z_h) \times \log p(z_i/z_k z_h). \quad (23.7)$$

Аналогично можно получить выражения для энтропии источника сообщений и при более протяженной корреляционной связи между знаками.

**Пример 1.** Определить, является ли эргодическим стационарный дискретный источник сообщений, алфавит которого состоит из четырех знаков  $z_1, z_2, z_3$  и  $z_4$ , причем безусловные вероятности выбора знаков одинаковы [ $p(z_1)=p(z_2)=p(z_3)=p(z_4) = 1/4$ ], а условные вероятности  $p(z_i/z_q)$  заданы табл. 23.1.



Таблица 23.1

	$z_1$	$z_1$	$z_2$	$z_3$	$z_4$
$z_1$					
$z_2$	$1/3$	$1/3$	$1/3$	$0$	$0$
$z_3$	$1/3$	$1/3$	$1/3$	$0$	$0$
$z_4$	$0$	$0$	$0$	$0$	$1$

Анализ табл. 23.1 показывает, что источник имеет два режима работы. С вероятностью, равной  $3/4$ , первым будет выбран один из знаков  $z_1, z_2$  или  $z_3$  и источник начнет формировать последовательность с равновероятным появлением знаков. Если же первым будет выбран знак  $z_4$  (вероятность такого случая равна  $1/4$ ), то генерируется последовательность, содержащая только знаки  $z_4$ .

Усреднение по ансамблю предполагает наличие множества однотипных источников, примерно три четверти из которых будет работать в первом режиме, а остальные — во втором. При этом в соответствии с (23.6) энтропия источника

$$H(Z) = -\frac{3}{4} \log_2 \frac{1}{3} - \frac{1}{4} \log_2 1 = 1,19 \text{ дв. ед.}$$

Среднее по последовательности (времени) вычисляется с использованием конкретной последовательности и поэтому зависит от режима функционирования источника. В первом режиме неопределенность, приходящаяся на один знак достаточно длинной последовательности (энтропия последовательности), равна 1,586 дв. ед., а во втором — нулю.

Поскольку энтропии формируемых последовательностей не совпадают с энтропией источника, он не является эргодическим.

Отметим, однако, что любой стационарный источник сообщений может быть представлен совокупностью нескольких эргодических источников, различающихся режимами работы.

**Свойства эргодических последовательностей знаков.**

Характер последовательностей, формируемых реальным источником сообщений, зависит от существующих ограничений на выбор знаков. Они выражаются в том, что вероятности реализации знаков различны и между ними существуют корреляционные связи. Эти ограничения приводят к тому, что вероятности формируемых последовательностей существенно различаются.

Пусть, например, эргодический источник без памяти последовательно выдает знаки  $z_1, z_2, z_3$  в соответствии с вероятностями 0,1; 0,3; 0,6. Тогда в образованной им достаточно длинной последовательности знаков мы ожидаем встретить в среднем на один знак  $z_1$  три знака  $z_2$  и шесть знаков  $z_3$ . Однако при ограниченном числе знаков в последовательности существуют вероятности того, что она будет содержать;

- только знаки  $z_1$  (либо  $z_2$ , либо  $z_3$ );
- только знаки  $z_1$  и один знак  $z_2$  или  $z_3$ ;
- только знаки  $z_2$  и один знак  $z_1$  или  $z_3$ ;
- только знаки  $z_3$  и один знак  $z_1$  или  $z_2$
- только знаки  $z_1$  и два знака  $z_2$  или  $z_3$  и т. д.

С увеличением числа знаков вероятности появления таких последовательностей уменьшаются.

Фундаментальные свойства длинных последовательностей знаков, создаваемых эргодическим источником сообщений, отражает следующая теорема:

*как бы ни малы были два числа  $\delta > 0$  и  $\mu > 0$  при достаточно большом  $N$ , все последовательности могут быть разбиты на две группы.*

Одну группу составляет подавляющее большинство последовательностей, каждая из которых имеет настолько ничтожную вероятность, что даже суммарная вероятность всех таких последовательностей очень мала и при достаточно большом  $N$  будет меньше сколь угодно малого числа  $\delta$ . Эти последовательности называют *нетипичными*.

Вторая группа включает *типичные* последовательности, которые при достаточно большом  $N$  отличаются тем, что вероятности их появления практически одинаковы, причем вероятность  $p$  любой такой последовательности удовлетворяет неравенству

$$|\log(1/p)/N - H(Z)| < \mu, \quad (23.8)$$

где  $H(Z)$  — энтропия источника сообщений.

Соотношение (23.8) называют также свойством *асимптотической равномерности* длинных последовательностей. Рассмотрим его подробнее.

Поскольку при  $N \rightarrow \infty$  источник сообщений с вероятностью, сколь угодно близкой к единице, выдает только типичные последовательности, принимаемое во внимание число последовательностей равно  $1/p$ . Неопределенность создания каждой такой последовательности с учетом их равновероятности составляет  $\log(1/p)$ . Тогда величина  $\log(1/p)/N$  представляет собой

неопределенность, приходящуюся в среднем на один знак. Конечно, эта величина практически не должна отличаться от энтропии источника, что и констатируется соотношением (23.8).

Ограничимся доказательством теоремы для простейшего случая эргодического источника без памяти. Оно непосредственно вытекает из закона больших чисел, в соответствии с которым в длинной последовательности из  $N$  элементов алфавита  $l$  ( $z_1, z_2, \dots, z_l$ ), имеющих вероятности появления  $p_1, p_2, \dots, p_b$  содержится  $Np_1$  элементов  $z_1, Np_2$  элементов  $z_2$  и т. д.

Тогда вероятность  $p$  реализации любой типичной последовательности близка к величине

$$p = p_1^{Np_1} p_2^{Np_2} \dots p_l^{Np_l}. \quad (23.9)$$

Логарифмируя правую и левую части выражения (23.9), получаем

$$\log p = N \sum_{i=1}^N p_i \log p_i,$$

откуда (при очень больших  $N$ )

$$\log (1/p)/N = H(Z).$$

Для общего случая теорема доказывается с привлечением цепей Маркова.

Покажем теперь, что за исключением случая равновероятного и независимого выбора букв источником, когда нетипичные последовательности отсутствуют, типичные последовательности при достаточно большом  $N$  составляют незначительную долю от общего числа возможных последовательностей.

При объеме алфавита источника  $l$  и количестве знаков в последовательности  $N$  число всех возможных последовательностей

$$n_1 = l^N = 2^{N \log_2 l}. \quad (23.10)$$

Принимая во внимание соотношение (23.8), число типичных последовательностей  $n_2$  можно записать в виде

$$n_2 = 2^{NH(Z)}. \quad (23.11)$$

Тогда

$$n_1/n_2 = 2^{N(\log_2 l - H(Z))}.$$

Так как  $H(Z) < \log_2 l$ , то  $n_2 \ll n_1$  и неравенство усиливается с увеличением  $N$ .

К. Шеннон показал, что рассмотренные свойства длинных последовательностей могут служить основанием для осуществления эффективного кодирования информации.

**Пример 2.** Оценить, какую долю общего числа возможных последовательностей следует учитывать в практических расчетах, если эргодический источник характеризуется параметрами:

$$l=16, \quad H(Z) = 3,5 \text{ дв. ед.}, \quad a = N = 50.$$

В соответствии с (23.10) и (23.11) имеем

$$n_1 = 16^{50} = 2^{200}; \quad n_2 = 2^{50 \cdot 3,5} = 2^{175},$$

откуда

$$n_2/n_1 = 2^{175}/2^{200} = 1/2^{25} \approx 1/30 \cdot 10^6.$$

Следовательно, к типичным последовательностям относится только одна тридцатимиллионная доля всех возможных реализаций!

**Избыточность.** Следствием ограничений на выбор источником знаков является также недоиспользование их как переносчиков информации. Известная априорная информация о вероятностях выбора отдельных знаков и их сочетаний приводит к уменьшению средней неопределенности выбора источником знака, а следовательно, и переносимого им количества информации. При равновероятном и некоррелированном выборе ту же информационную нагрузку на знак можно обеспечить, используя алфавит меньшего объема. В связи с этим говорят об *избыточности алфавита*  $l$  источника сообщений или просто об избыточности источника.

Мерой избыточности служит величина  $D$ , показывающая, насколько хорошо используются знаки данного источника:

$$D = [H_{\max}(Z) - H(Z)]/[H_{\max}(Z)], \quad (23.12)$$

где  $H_{\max}(Z)$  — максимально возможная энтропия, равная  $\log l$ ;  $H(Z)$  — энтропия источника.

Если избыточность источника равна нулю, то формируемые им сообщения оптимальны в смысле наибольшего количества переносимой информации. Для передачи определенного количества информации  $I$  при отсутствии помех в этом случае необходимо  $kI = I[H_{\max}(Z)]$  знаков.

Поскольку энтропия сообщений, формируемых реальным источником, обладающим избыточностью, меньше максимальной, то для передачи того же количества информации  $I$  знаков требуется больше, а именно:  $k_2 = I/H(Z) > k_1$ . Поэтому говорят также об избыточности знаков в сообщении или просто об избыточности сообщения, характеризуя ее тем же параметром  $D$ :

$$D = (k_2 - k_1)/k_2 = [H_{\max}(Z) - H(Z)]/[H_{\max}(Z)].$$

Избыточность нельзя рассматривать как признак несовершенства источника сообщений. Обычно она является следствием его физических свойств. Ограничения, существующие в любом естественном языке, связаны, например, с особенностями артикуляции, не позволяющими формировать слова, состоящие из произвольных сочетаний букв.

Последствия от наличия избыточности сообщений неоднозначны. С одной стороны, избыточные сообщения требуют дополнительных затрат на передачу, например, увеличения длительности передач или расширения практической ширины спектра канала связи, что нежелательно. С другой стороны, при использовании сообщений, подчиняющихся априорно известным ограничениям, появляется возможность обнаружения и исправления ошибок, которые приводят к нарушению этих ограничений. Следовательно, наличие избыточности способствует повышению помехоустойчивости сообщений. Высокая избыточность большинства естественных языков обеспечивает, например, надежное общение людей даже при наличии у них акцентов и дефектов речи.

Однако при обмене информацией в автоматических системах естественная избыточность подлежит устранению. Это объясняется тем, что алгоритмы обнаружения и исправления ошибок, базирующихся на статистических закономерностях функционирования источника, оказываются слишком сложными для реализации их техническими средствами. В случае необходимости для повышения помехоустойчивости затем вводится «рациональная» избыточность, позволяющая обеспечить обнаружение и исправление наиболее вероятных и опасных по последствиям ошибок простыми техническими средствами. При низком уровне помех в канале связи устранение избыточности приводит к увеличению скорости передачи информации и может дать значительный экономический эффект.

**Пример 3.** Определить возможный эффект от устранения избыточности при передаче текста на русском языке.

Максимальная энтропия текста на русском языке (с учетом пренебрежения при передаче различиями в буквах е и ё, ь и Ъ) установлена ранее и равна 5 дв. ед. Там же определена энтропия с учетом неравномерного распределения вероятностей появления отдельных букв (4,42 дв. ед). Имея сведения о переходных вероятностях и исходя из предположения, что текст представляет собой простую цепь Маркова, можно установить, что энтропия уменьшается до 3,52 дв. ед. Учет всех ограничений в языке, включая связи между словами, позволяет оценить минимальную величину энтропии значением 1,5 дв. ед. Таким образом, избыточность русского языка составляет

$$D = [H_{\max}(Z) - H(Z)] / H_{\max}(Z) = (5 - 1,5) / 5 = 0,7.$$

Это означает, что каналы связи, построенные без учета ограничений, существующих в языке, и способные передавать равновероятные буквы, следующие друг за другом в любых сочетаниях, при передаче информации без помех текстом на русском языке используется всего

на 30 %. Полное устранение избыточности позволило бы повысить эффективность их использования более чем в 3 раза!

*Производительность источника дискретных сообщений.* Под *производительностью источника сообщений* подразумевают количество информации, вырабатываемое источником в единицу времени. Эту характеристику источника называют также скоростью создания сообщений или потоком входной информации. Поскольку возможное воздействие помех на источник сообщений принято учитывать эквивалентным изменением характеристик модели канала связи, то производительность источника сообщений равна энтропии источника, приходящейся на единицу времени.

Длительность выдачи знаков источником в каждом из состояний в общем случае может быть различной. Обозначим длительность выдачи знака  $z_i$ , формируемого источником в состоянии  $S_q$  через  $\tau_{qz_i}$ . Тогда средняя длительность выдачи источником одного знака

$$\tau_n = \sum_{q=1}^R p(S_q) \sum_{i=1}^I p_q(z_i) \tau_{qz_i}. \quad (23.13)$$

Производительность источника  $\bar{I}(z)$  теперь можно выразить формулой

$$\bar{I}(Z) = H(Z)/\tau_n. \quad (23.14)$$

Как следует из (23.13), повышение производительности источника возможно не только за счет увеличения энтропии, но и за счет снижения средней длительности формирования знака. Длительность знаков желательно выбирать обратно пропорциональными вероятностям их появления.

Если длительность выдачи знака не зависит от состояния источника, для всех знаков одинакова и равна  $\tau$ , то  $\tau_n = \tau$ . Выражение для  $\bar{I}(Z)$  принимает вид

$$\bar{I}(Z) = H(Z)/\tau. \quad (23.15)$$

Наибольшая производительность источника в этом случае достигается при максимальной энтропии.

### **23.4. Информационные характеристики дискретных каналов связи**

**Модели дискретных каналов.** *Дискретным каналом* называют совокупность средств, предназначенных для передачи дискретных сигналов. Такие каналы широко используются, например, при передаче данных, в телеграфии, радиолокации.

Дискретные сообщения, состоящие из последовательности знаков алфавита источника сообщений (первичного алфавита)  $z_1, z_2, \dots, z_n$ , преобразуются в кодирующем устройстве в последовательности символов. Объем  $m$  алфавита символов (вторичного алфавита)  $u_1, u_2, \dots, u_m$ , как правило, меньше объема  $l$  алфавита знаков, но они могут и совпадать.

Материальным воплощением символа является элементарный сигнал, получаемый в процессе манипуляции — дискретного изменения определенного параметра переносчика информации. Элементарные сигналы формируются с учетом физических ограничений, накладываемых конкретной линией связи. В результате манипуляции каждой последовательности символов ставится в соответствие сложный сигнал. Множество сложных сигналов конечно. Они различаются числом, составом и взаимным расположением элементарных сигналов.

Термины «элементарный сигнал» и «символ», так же как «сложный сигнал» и «последовательность символов», в дальнейшем будут использоваться как синонимы.

Информационная модель канала с помехами задается множеством символов на его входе и выходе и описанием вероятностных свойств передачи отдельных символов. В общем случае канал может иметь множество состояний и переходить из одного состояния в другое как с течением времени, так и в зависимости от последовательности передаваемых символов.

В каждом состоянии канал характеризуется матрицей условных вероятностей  $p(v_j/u_i)$  того, что переданный символ  $u_i$  будет воспринят на выходе как символ  $v_j$ . Значения вероятностей в реальных каналах зависят от многих различных факторов: свойств сигналов, являющихся физическими носителями символов (энергия, вид модуляции и т.д.), характера и интенсивности воздействующих на канал помех, способа определения сигнала на приемной стороне.

При наличии зависимости переходных вероятностей канала от времени, что характерно практически для всех реальных каналов, он называется нестационарным каналом связи. Если эта зависимость незначительна, используется модель в виде стационарного канала, переходные вероятности которого не зависят от времени. Нестационарный канал может быть представлен рядом стационарных каналов, соответствующих различным интервалам времени.

Канал называется с «памятью» (с последствием), если переходные вероятности в данном состоянии канала зависят от его предыдущих состояний. Если переходные вероятности постоянны, т.е.

канал имеет только одно состояние, он называется *стационарным каналом без памяти*. Под  $k$ -ичным каналом подразумевается канал связи, у которого число различных символов на входе и выходе одинаково и равно  $k$ .

*Стационарный дискретный двоичный канал без памяти* однозначно определяется четырьмя условными вероятностями:  $p(0/0)$ ,  $p(1/0)$ ,  $p(0/1)$ ,  $p(1/1)$ . Такую модель канала принято изображать в виде графа, представленного на рис. 23.3, где  $p(0/0)$  и  $p(1/1)$  — вероятности неискаженной передачи символов, а  $p(0/1)$  и  $p(1/0)$  — вероятности искажения (трансформация) символов 0 и 1 соответственно.

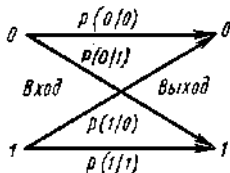


Рис. 23.3

Если вероятности искажения символов можно принять равными, т.е.  $p(0/1) \approx p(1/0) = q$ , то такой канал называют *двоичным симметричным каналом* [при  $p(0/1) \neq p(1/0)$  канал называется *несимметричным*]. Символы на его выходе правильно принимают с вероятностью  $p$  и неправильно — с вероятностью  $1 - p = q$ . Математическая модель упрощается.

Именно этот канал исследовался наиболее интенсивно не столько в силу своей практической значимости (многие реальные каналы описываются им весьма приближенно), сколько в силу простоты математического описания.

Важнейшие результаты, полученные для двоичного симметричного канала, распространены на более широкие классы каналов.

Следует отметить еще одну модель канала, которая приобрела большое значение. Это дискретный канал со стиранием. Для него характерно, что алфавит выходных символов отличается от алфавита входных символов. На входе, как и ранее, символы 0 и 1, а на выходе канала фиксируются состояния, при которых сигнал с равным основанием может быть отнесен как к единице, так и к нулю. На месте такого символа не ставится ни нуль, ни единица: состояние отмечается дополнительным символом стирания  $S$ . При декодировании значительно легче исправить такие символы, чем ошибочно определенные.



На рис. 23.3 приведены модели стирающего канала при отсутствии (рис. 23.3, а) и при наличии (рис. 23.3,б) трансформации символов.

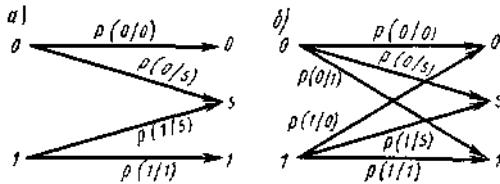


Рис. 23.3

**Скорость передачи информации по дискретному каналу.**

Характеризуя дискретный канал связи, используют два понятия скорости передачи: технической и информационной.

Под *технической скоростью передачи*  $V_T$ , называемой также скоростью манипуляции, подразумевают число элементарных сигналов (символов), передаваемых по каналу в единицу времени. Она зависит от свойств линии связи и быстродействия аппаратуры канала.

С учетом возможных различий в длительностях символов скорость

$$V_T = 1/\tau_{cp}, \tag{23.16}$$

где  $\tau_{cp}$  — среднее значение длительности символа.

При одинаковой продолжительности  $\tau$  всех передаваемых символов  $\tau_{cp} = \tau$ .

Единицей измерения технической скорости служит *бод* — скорость, при которой за одну секунду передается один символ.

*Информационная скорость*, или *скорость передачи информации*, определяется средним количеством информации, которое передается по каналу в единицу времени. Она зависит как от характеристик данного канала связи, таких, как объем алфавита используемых символов, техническая скорость их передачи, статистические свойства помех в линии, так и от вероятностей поступающих на вход символов и их статистической взаимосвязи. При известной скорости манипуляции  $V_T$  скорость передачи информации по каналу  $I(V,U)$  задается соотношением

$$\bar{I}(V,U) = V_T I(V,U), \tag{23.17}$$

где  $I(V,U)$  — среднее количество информации, переносимое одним символом.

**Пропускная способность дискретного канала без помех.** Для теории и практики важно выяснить, до какого предела и каким путем можно повысить скорость передачи информации по конкретному каналу связи. Предельные возможности канала по передаче информации характеризуются его пропускной способностью.

Пропускная способность канала  $C_d$  равна той максимальной скорости передачи информации по данному каналу, которой можно достигнуть при самых совершенных способах передачи и приема:

$$C_d = \max I(V,U) = \max V_T I(V,U). \quad (23.18)$$

При заданном алфавите символов и фиксированных основных характеристиках канала (например, полосе частот, средней и пиковой мощности передатчика) остальные характеристики должны быть выбраны такими, чтобы обеспечить наибольшую скорость передачи по нему элементарных сигналов, т. е. обеспечить максимальное значение  $V_T$ . Максимум среднего количества информации, приходящейся на один символ принятого сигнала  $I(V,U)$ , определяется на множестве распределений вероятностей между символами  $u_1 \dots u_i \dots u_m$ .

Пропускная способность канала, как и скорость передачи информации по каналу, измеряется числом двоичных единиц информации в секунду (дв. ед./с).

Так как в отсутствие помех имеет место взаимно-однозначное соответствие между множеством символов  $\{v\}$  на выходе канала и  $\{u\}$  на его входе, то  $I(V,U)=I(U,V)=H\{U\}$ . Максимум возможного количества информации на символ равен  $\log m$ , где  $m$  — объем алфавита символов, откуда пропускная способность дискретного канала без помех

$$C_d = V_T \log m. \quad (23.19)$$

Следовательно, для увеличения скорости передачи информации по дискретному каналу без помех и приближения ее к пропускной способности канала последовательность букв сообщения должна подвергнуться такому преобразованию в кодере, при котором различные символы в его выходной последовательности появлялись бы по возможности равновероятно, а статистические связи между ними отсутствовали бы.

Доказано, что это выполнимо для любой эргодической последовательности букв, если кодирование осуществлять блоками такой длины, при которой справедлива теорема об их асимптотической равновероятности.

Расширение объема алфавита символов  $m$  приводит к повышению пропускной способности канала (рис. 23.4), однако возрастает и сложность технической реализации.

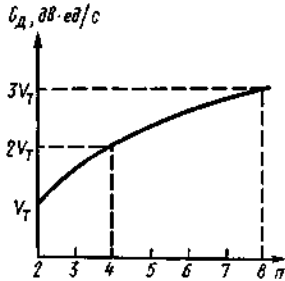


Рис 23.4

**Пропускная способность дискретного канала с помехами.** При наличии помех соответствие между множествами символов на входе и выходе канала связи перестает быть однозначным. Среднее количество информации  $I(V, U)$ , передаваемое по каналу одним символом, определяется в этом случае соотношением

$$I(VU) = H(V) - H_v(V) = H(U) - H_v(U). \quad (23.20)$$

Если статистические связи между символами отсутствуют, энтропия сигнала на выходе линии связи равна

$$H(V) = - \sum_{i=1}^m p(v_i) \log p(v_i). \quad (23.21)$$

При наличии статистической связи энтропию определяют с использованием цепей Маркова. Поскольку алгоритм такого определения ясен и нет необходимости усложнять изложение громоздкими формулами, ограничимся здесь только случаем отсутствия связей.

Апостериорная энтропия характеризует уменьшение количества переданной информации вследствие возникновения ошибок. Она зависит как от статистических свойств последовательностей символов, поступающих на вход канала связи, так и от совокупности переходных вероятностей, отражающих вредное действие помехи.

Если объем алфавита входных символов  $u$  равен  $m_1$ , а выходных символов  $v - m_2$ , то

$$H_v(V) = - \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} p(v_j u_i) \log p(v_j / u_i), \quad (23.22)$$

Подставив выражения (23.21) и (23.22) в (23.20) и проведя несложные преобразования, получим

$$I(V, U) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} p(v_j u_i) \log \frac{p(v_j u_i)}{p(v_j) p(u_i)}. \quad (23.23)$$

Скорость передачи информации по каналу с помехами

$$\bar{I}(V,U) = V_T \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} p(v_j u_i) \log \frac{p(v_j u_i)}{p(v_j)p(u_i)}. \quad (23.24)$$

Считая скорость манипуляции  $V_T$  предельно допустимой при заданных технических характеристиках канала, величину  $I(V,U)$  можно максимизировать, изменяя статистические свойства последовательностей символов на входе канала посредством преобразователя (кодера канала). Получаемое при этом предельное значение  $C_d$  скорости передачи информации по каналу называют *пропускной способностью* дискретного канала связи с помехами:

$$C_d = \max_{p\{u\}} V_T \bar{I}(VU), \quad (23.25)$$

где  $p\{u\}$  — множество возможных распределений вероятностей входных сигналов.

Важно подчеркнуть, что при наличии помех пропускная способность канала определяет наибольшее количество информации в единицу времени, которое может быть передано со сколь угодно малой вероятностью ошибки.

К пропускной способности канала связи с помехами можно приблизиться, кодируя эргодическую последовательность букв источника сообщений блоками такой длины, при которой справедлива теорема об асимптотической равновероятности длинных последовательностей.

Произвольно малая вероятность ошибки оказывается достижимой только в пределе, когда длина блоков становится бесконечной.

При удлинении кодируемых блоков возрастает сложность технической реализации кодирующих и декодирующих устройств и задержка в передаче сообщений, обусловленная необходимостью накопления требуемого числа букв в блоке. В рамках допустимых усложнений на практике при кодировании могут преследоваться две цели: либо при заданной скорости передачи информации стремятся обеспечить минимальную ошибку, либо при заданной достоверности — скорость передачи, приближающуюся к пропускной способности канала.

Предельные возможности канала никогда не используются полностью. Степень его загрузки характеризуется *коэффициентом использования канала*

$$\lambda = \bar{I}(Z)/C_d, \quad (23.26)$$

где  $\bar{I}(Z)$  — производительность источника сообщений;  $C_d$  — пропускная способность канала связи.

Поскольку нормальное функционирование канала возможно, как показано далее, при изменении производительности источника в пределах  $0 \leq \bar{I}(Z) \leq C_n$ ,  $\lambda$  теоретически может изменяться в пределах от 0 до 1.

**Пример 4.** Определить пропускную способность двоичного симметричного канала (ДСК) со скоростью манипуляции  $V_T$  в предположении независимости передаваемых символов.

Запишем соотношение (23.22) в следующем виде:

$$H_U(V) = - \sum_{i=1}^2 p(u_i) \sum_{j=1}^2 p(v_j/u_i) \log_2 p(v_j/u_i).$$

Воспользовавшись обозначениями на графе (рис. 23.5), можем записать

$$\begin{aligned} H_U(V) &= -p(0)[(1-p) \log_2 (1-p) + p \log_2 p] - \\ &\quad - p(1)[p \log_2 p + (1-p) \log_2 (1-p)] = \\ &= -[p(0) + p(1)][p \log_2 p + (1-p) \log_2 (1-p)]. \end{aligned}$$

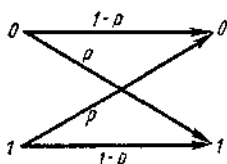


Рис. 23.5

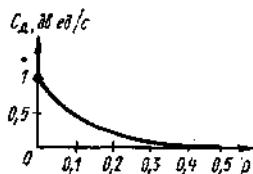


Рис. 23.6

Так как  $p(0) + p(1) = 1$ , то

$$H_U(V) = -p \log_2 p - (1-p) \log_2 (1-p).$$

Величина  $H_U(V)$  не зависит от вероятностей входных символов, что является следствием симметрии канала. Следовательно, пропускная способность

$$C_n = V_T [\max H(V) + p \log_2 p + (1-p) \log_2 (1-p)].$$

Максимум  $H(V)$  достигается при равенстве вероятностей появления символов, он равен 1. Отсюда

$$C_n = V_T [1 + p \log_2 p + (1-p) \log_2 (1-p)]. \quad (23.26)$$

График зависимости пропускной способности ДСК от  $p$  показан на рис. 23.6. При увеличении вероятности трансформации символа с 0 до  $1/2$   $C_n(p)$  уменьшается от 1 до 0. Если  $p=0$ , то шум в канале отсутствует и его пропускная способность равна 1. При  $p = 1/2$  канал бесполезен, так как значения символов на приемной стороне с равным успехом можно устанавливать по результатам подбрасывания монеты (герб — 1, решетка — 0). Пропускная способность канала при этом равна нулю.

### **23.5. Информационные характеристики источника непрерывных сообщений**

**Эпсилон-производительность непрерывного источника сообщений.** Под конкретным непрерывным сообщением  $z_j(t)$  подразумевают некоторую реализацию случайного процесса длительностью  $T$ . Источник непрерывных сообщений характеризуется ансамблем его реализаций. Наиболее плодотворной оказалась модель непрерывного сообщения в виде эргодического случайного процесса.

Для определения производительности источника непрерывных сообщений воспользуемся подходом и результатами предыдущих изложений, где определена  $\varepsilon$ -энтропия случайной величины.

Под  $\varepsilon$ -производительностью источника непрерывных сообщений  $H_\varepsilon(z)$  понимают минимальное количество информации, которое необходимо создать источнику в единицу времени, чтобы любую реализацию  $z_j(t)$  можно было воспроизвести с заданной вероятностью  $\varepsilon$ .

Допустим, что  $z_T(t)$  воспроизводится реализацией  $u_T(t)$ . Наблюдаемые реализации следует рассматривать, как сигналы, обладающие ограниченным, хотя возможно и достаточно широким спектром  $F$ .

При достаточно большой длительности  $T$  как  $z_T(t)$ , так и  $u_T(t)$  могут быть представлены  $N$ -мерными ( $N = 2FT$ ) векторами ( $z_1, z_2, \dots, z_N$ ) и ( $u_1, u_2, \dots, u_N$ ), координатами которых являются отсчеты. Ансамбли сообщений  $[z_T(t)]$  и воспроизводящих сигналов  $[u_T(t)]$  характеризуют при этом  $N$ -мерными случайными векторами  $\mathbf{Z}$  и  $\mathbf{U}$ , составляющими которых являются соответственно случайные величины  $Z_1, Z_2, \dots, Z_N$  и  $U_1, U_2, \dots, U_N$ . Статистическое описание каждого из ансамблей задается  $N$ -мерными плотностями распределения вероятностей  $p(\mathbf{Z}) = p(z_1, z_2, \dots, z_N)$  и  $p(\mathbf{U}) = p(u_1, u_2, \dots, u_N)$ . Связь между ансамблями отражают условные плотности распределений

$$p_U(\mathbf{Z}) = p(z_1, z_2, \dots, z_N / u_1, u_2, \dots, u_N)$$

и

$$p_Z(\mathbf{U}) = p(u_1, u_2, \dots, u_N / z_1, z_2, \dots, z_N),$$

а также совместная плотность распределения вероятностей

$$p(\mathbf{Z}, \mathbf{U}) = p(z_1, z_2, \dots, z_N; u_1, u_2, \dots, u_N).$$

Распространяя формулу (23.23) на  $N$ -мерные случайные векторы  $\mathbf{Z}$  и  $\mathbf{U}$  для количества информации одного из них относительно второго, получим

$$I(\mathbf{Z}, \mathbf{U}) = \int \int_{\mathbf{Z}\mathbf{U}} p(\mathbf{Z}, \mathbf{U}) \log \frac{p(\mathbf{Z}, \mathbf{U})}{p(\mathbf{Z})p(\mathbf{U})} d\mathbf{Z}d\mathbf{U}, \quad (23.28)$$

где интегралы являются  $N$ -мерными.

Используем, как и ранее, среднеквадратический критерий верности  $\theta(\mathbf{Z}, \mathbf{U})$ , который в рассматриваемом случае имеет вид

$$\theta(\mathbf{Z}, \mathbf{U}) = \int \int_{\mathbf{Z}\mathbf{U}} p(\mathbf{Z})p_{\mathbf{Z}}(\mathbf{U})p(\mathbf{Z}, \mathbf{U})d\mathbf{Z}d\mathbf{U},$$

где  $p(\mathbf{Z}, \mathbf{U})$  представляет собой квадрат расстояния  $l(\mathbf{Z}, \mathbf{U})$  в  $N$ -мерном евклидовом пространстве.

Количество информации, приходящееся в среднем на один отсчет дискретизованных сигналов  $Z_T(t)$  и  $U_T(t)$ , определяется выражением

$$I(\mathbf{Z}, \mathbf{U}) = \frac{1}{N} \int \int_{\mathbf{Z}\mathbf{U}} p(\mathbf{Z}, \mathbf{U}) \log \frac{p(\mathbf{Z}, \mathbf{U})}{p(\mathbf{Z})p(\mathbf{U})} d\mathbf{Z}d\mathbf{U}.$$

Тогда в соответствии с определением для  $\varepsilon$ -пропорциональности источника непрерывных сообщений  $H_\varepsilon(\mathbf{Z})$  запишем

$$H_\varepsilon(\mathbf{Z}) = \nu \min_{\{p(\mathbf{Z}, \mathbf{U})\}} \bar{I}(\mathbf{Z}, \mathbf{U}) \quad (23.29)$$

при выполнении условия  $\theta(\mathbf{Z}, \mathbf{U}) \leq \varepsilon^2$ .

Величина  $\nu$  характеризует скорость формирования источником отсчетов ( $\nu = 2F$ ).

**Пример 5.** Определить  $\varepsilon$ -производительность источника, формирующего со скоростью  $\nu_1$  некоррелированные отсчеты стационарного нормального случайного сигнала с дисперсией  $\sigma^2$ .

Воспользовавшись полученным в (13.92) значением  $\varepsilon$ -энтропии для нормально распределенной случайной величины, найдем

$$\overline{H_\varepsilon(\mathbf{Z})} = \nu_1 H_\varepsilon(\mathbf{Z}) = \frac{\nu_1}{2} \log_2 \frac{\sigma^2}{\varepsilon^2} \frac{\text{дв. ед.}}{\text{с}}.$$

Возможности воспроизведения любого сообщения  $Z_T(t)$  с заданной верностью можно дать геометрическое толкование. Поскольку все реализации эргодического процесса достаточно большой длительности являются типичными и обладают практически одной и той же средней мощностью, концы соответствующих им векторов в  $N$ -мерном пространстве сообщений составляют непрерывное множество точек, равноудаленных от начала координат (гиперсферу).

Конечное подмножество воспроизводящих сигналов  $u_T(t)$  размещается в центрах непересекающихся правильных сферических  $N$ -угольников ( $\varepsilon$ -областей), на которое гиперсфера разбивается без промежутков. Размеры  $\varepsilon$ -областей определены заданной верностью воспроизведения сообщений. Если источником реализуется сообщение  $z_T^*(t)$ , конец вектора которого должен попасть в  $\varepsilon$ -область сигнала  $u_T^*(t)$ , то воспроизводится сигнал  $u_T^*(t)$ .

Следует отметить, что заданная верность воспроизведения будет достигнута с вероятностью, близкой к единице, только при достаточно большой длительности сообщений, когда погрешностью от замены непрерывных реализаций последовательностями отсчетов можно будет пренебречь. Для уменьшения указанной погрешности при ограниченной длительности сообщений  $T$  необходимо увеличивать число отсчетов  $N$ . В пределе при  $N \rightarrow \infty$  получим непрерывные реализации.

В вычислении  $\varepsilon$ -производительности источника и геометрическом толковании возможности воспроизведения сообщений с заданной верностью принципиально ничего не изменяется. Следует лишь учесть, что  $N$ -мерное евклидово пространство сообщений становится гильбертовым и мерой близости двух сигналов должно быть расстояние в этом пространстве.

## **23.6. Информационные непрерывных каналов связи**

**Модели непрерывных каналов связи.** Каналы, используемые для передачи непрерывных сигналов, принято называть непрерывными. Такие каналы до сих пор находят широкое применение, например, в технике телефонной связи, радиовещании.

Реальные непрерывные каналы представляют собой сложные инерционные нелинейные объекты, характеристики которых случайным образом изменяются во времени.

Для анализа таких каналов разработаны математические модели различных уровней сложности и степени адекватности реальным каналам. Модели, получившие наиболее широкое распространение, — это разновидности гауссова канала.

Под *гауссовым каналом* понимают математическую модель реального канала, построенную при следующих допущениях:

- 1) основные физические параметры канала являются известными детерминированными величинами;
- 2) полоса пропускания канала ограничена частотой  $F_k$  герц;
- 3) в канале действует аддитивный гауссовый белый шум — аддитивная флюктуационная помеха ограниченной мощности с равномерным частотным спектром и нормальным распределением амплитуд.

Предполагается также, что по каналу передаются сигналы с постоянной средней мощностью, статистические связи между сигналами и шумом отсутствуют, ширина спектра сигнала и помехи ограничена полосой пропускания канала.



При рассмотрении информационных характеристик канала (скорости передачи, пропускной способности, коэффициента использования) основное внимание будет уделено гауссовому каналу.

**Скорость передачи информации по непрерывному каналу.** Скорость передачи информации по непрерывному каналу — это количество информации, которое передается в среднем принятыми непрерывными сигналами  $v(t)$ , относительно переданных  $u(t)$  в единицу времени.

Поскольку полоса пропускания канала всегда ограничена, непрерывные сообщения на достаточно продолжительном интервале времени  $T$  с некоторой погрешностью могут быть представлены последовательностями отсчетов. С учетом наличия корреляционных связей между отсчетами и конечной верности воспроизведения, обусловленной воздействием помехи, для средней скорости  $\tilde{I}(\mathbf{VU})$  передачи информации дискретизованным сигналом получаем

$$\tilde{I}(\mathbf{VU}) = I(\mathbf{VU})/T, \quad (23.30)$$

где  $I(\mathbf{VU})$  определяется выражением, аналогичным (23.28). По мере увеличения длительности  $T$  эта скорость возрастает, так как при каждом новом отсчете реализации уточняются. В пределе при  $T \rightarrow \infty$   $N$ -мерные распределения становятся бесконечномерными и выражение (23.30) будет определять скорость передачи информации по непрерывному каналу:

$$\bar{I}(\mathbf{VU}) = \lim_{T \rightarrow \infty} \tilde{I}(\mathbf{VU}).$$

Переход к пределу при  $T \rightarrow \infty$  также означает усреднение скорости по всем возможным сигналам.

Степень вредного воздействия помехи с известными статистическими свойствами на различные ансамбли входных сигналов различна. Вследствие этого различны и значения скорости передачи информации.

**Пропускная способность непрерывного канала связи.** Максимально возможную скорость  $C_n$  передачи информации по непрерывному каналу с известными техническими характеристиками называют *пропускной способностью непрерывного канала*:

$$C_n = \max_{\{p(u)\}} \bar{I}(\mathbf{VU}), \quad (23.31)$$

где максимум находят по всем возможным ансамблям входных сигналов.

Определим скорость передачи информации по гауссову каналу.

Пусть по гауссову каналу передается непрерывный сигнал  $u_T(t)$  из ансамбля  $[u_T(t)]$  со средней мощностью  $P_u$ , равной дисперсии  $\sigma_u^2$ . На выходе канала получим сигнал  $v_T(t)$  из ансамбля  $[v_T(t)]$ , искаженный гауссовой помехой  $\xi(t)$ , среднюю мощность которой обозначим  $P_\xi$  ( $P_\xi = \sigma_\xi^2$ ).

Будем считать, что длительность  $T$  сигнала  $u_T(t)$  достаточно велика, чтобы с приемлемой погрешностью можно было заменить  $u_T(t)$  и  $v_T(t)$  последовательностями отсчетов, взятых через интервалы  $\Delta t = 1/(2F_K)$ , где  $F_K$  — полоса пропускания канала.

В соответствии с (23.20) выражение для среднего количества информации, передаваемой сигналом  $v_T(t)$ , принимает вид

$$I(\mathbf{V}, \mathbf{U}) = H(\mathbf{V}) - H_U(\mathbf{V}), \quad (23.32)$$

где  $H(\mathbf{V})$  и  $H_U(\mathbf{V})$  — априорная и апостериорная энтропии  $N$ -мерного случайного вектора  $\mathbf{V}$ , составляющими которого являются случайные величины  $V_1, V_2, \dots, V_N$ .

Поскольку помеха в канале аддитивна и статистически не связана с входным сигналом, справедливо равенство

$$H_U(\mathbf{V}) = H_U(\mathbf{U} + \mathbf{\Xi}) = H(\mathbf{\Xi}). \quad (23.33)$$

Величина  $H(\mathbf{\Xi})$  в (23.33) представляет собой энтропию  $N$ -мерного случайного вектора помехи  $\mathbf{\Xi}$ , составляющими которого являются случайные величины  $\Xi_1, \Xi_2, \dots, \Xi_N$ .

Учитывая, что значения белого шума в моменты отсчетов будут некоррелированными, запишем

$$H(\mathbf{\Xi}) = 2F_K T h(\xi), \quad (23.34)$$

где  $h(\xi)$  — дифференциальная энтропия одного отсчетного значения помехи.

Для помехи с нормальным распределением и дисперсией  $\sigma_\xi^2$  она составит (13.76)

$$h(\xi) = \log \sigma_\xi \sqrt{2\pi e} = \frac{1}{2} \log 2\pi e P_\xi. \quad (23.35)$$

Будем считать, что отсчетные значения входных функций  $u_T(t)$  независимы. При воздействии на них независимых значений помехи отсчетные значения выходных сигналов  $V_T(t)$  также независимы.

Тогда  $H(\mathbf{V})$  можно выразить через дифференциальную энтропию  $h(V)$  одного отсчета выходного сигнала:

$$H(\mathbf{V}) = 2F_K T h(V). \quad (23.36)$$

Подставив (23.35) и (23.36) в (23.32), получим

$$I(\mathbf{V}, \mathbf{U}) = 2F_K T \left[ h(V) - \frac{1}{2} \log 2\pi e \sigma_\xi^2 \right]. \quad (23.37)$$

Соответственно скорость передачи информации по непрерывному каналу связи

$$\bar{I}(\mathbf{V}\mathbf{U}) = 2F_{\kappa} \left[ h(V) - \frac{1}{2} \log 2\pi e \sigma_{\xi}^2 \right]. \quad (23.38)$$

Определим теперь пропускную способность гауссова канала.

Найдем ансамбль входных сигналов, при котором обеспечивается максимальное значение  $h(V)$  в выражении (23.38).

Так как выходные сигналы образуются в результате суммирования входных сигналов и помехи, средние мощности которых ограничены, то и средняя мощность выходных сигналов ограничена. Для таких сигналов, как было установлено, наибольшее значение  $h(V)$  достигается при распределении  $V$  по нормальному закону. Известно также, что сумма двух нормально распределенных случайных величин имеет такую же функцию распределения с суммарной дисперсией. Отсюда следует, что при нормально распределенной помехе  $\xi$ , выходной сигнал  $V$  будет распределен по нормальному закону лишь при нормально распределенном входном сигнале  $u$ .

Наибольшее значение энтропии  $h(V)$ , а следовательно, и максимальная скорость передачи информации могут быть достигнуты при использовании нормальных центрированных случайных сигналов. Центрированность сигнала при заданной средней мощности соответствует максимальному значению дисперсии.

Они также должны иметь широкий и равномерный энергетический спектр, поскольку только в этом случае можно говорить о независимости отсчетов и использовать соотношение

$$U(t) = \sum_{-\infty}^{\infty} U(n\Delta t) \frac{\sin \omega_n(t - n\Delta t)}{\omega_n(t - n\Delta t)},$$

где  $\Delta t = \pi/\omega_n = 1/(2F_n)$ .

Таким образом, для более полного использования возможностей канала передаваемый сигнал должен обладать свойствами помехи, т. е. должен быть шумоподобным.

Максимальная величина дифференциальной энтропии

$$h(V) = -\log 2\pi e(\sigma_u^2 + \sigma_{\xi}^2) = \frac{1}{2} \log 2\pi e(P_u + P_{\xi}). \quad (23.39)$$

Подставляя (23.39) в (23.38), получаем выражение для пропускной способности гауссова канала:

$$C_{\pi} = F_{\kappa} \left[ \log 2\pi e(P_u + P_{\xi}) - \log 2\pi e P_{\xi} \right] = F_{\kappa} \log(1 + P_u/P_{\xi}). \quad (23.40)$$

Выясним, как зависит пропускная способность гауссова канала от ширины полосы пропускания  $F_{\kappa}$ .

Из выражения (23.40) следует, что эта зависимость нелинейна, поскольку  $F_K$  также влияет на мощность помехи. Учитывая равномерность энергетического спектра белого шума, представим его мощность  $P_\xi$  через удельную мощность  $P_0$  на единицу частоты.

Выражение (23.40) примет вид

$$C_n = F_K \log_2 [1 + P_u / (P_0 F_K)] .$$

Рост пропускной способности канала при неограниченном расширении его полосы пропускания ограничен пределом  $C_M$ :

$$C_M = \lim_{F_K \rightarrow \infty} C_n = \lim_{F_K \rightarrow \infty} \frac{\log_2 [1 + P_u / (P_0 F_K)]}{1/F_K} .$$

Обозначив  $\gamma = 1/F_K$ , по правилу Лопитала определяем предел  $C_n$  при  $\gamma \rightarrow 0$ :

$$C_M = 1,443 P_u / P_0 .$$

О характере зависимости  $C_n = f(F_K)$  можно судить по графику, представленному на рис. 23.7.

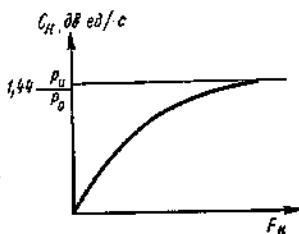


Рис. 23.7

### **23.7. Согласование физических характеристик сигнала и канала**

Конкретный канал связи обладает определенными физическими параметрами, от которых зависит возможность передачи по нему тех или иных сигналов. Независимо от назначения непрерывного канала его можно характеризовать тремя основными параметрами: временем, в течение которого он предоставляется для передачи сигнала  $T_K$ , шириной полосы пропускания сигнала  $F_K$  и допустимым превышением сигнала над помехой в канале  $H_K$ . Превышение  $H_K$  характеризуется разностью максимально допустимого сигнала в канале  $P_{u \max}$  и уровня помех  $P_\xi$  (в логарифмическом масштабе). Для проводных каналов превышение в основном определяется пробивным напряжением и уровнем перекрестных помех, для радиоканалов — возможностями выявления сигнала на соответствующих расстояниях.

Произведение указанных основных параметров канала связи принято называть объемом (емкостью) канала и обозначать  $V_K$ :

$$V_K = T_K F_K H_K. \quad (23.41)$$

При оценке возможностей передачи сигнала по каналу с заданными физическими характеристиками также ограничиваются рассмотрением трех основных параметров сигнала: его длительности  $T_c$ , ширины спектра  $F_c$  и превышения над помехой  $H_c$ , причем

$$H_c = \log(P_u/P_\xi), \quad (23.42)$$

где  $P_u$  — средняя мощность передаваемого сигнала;  $P_\xi$  — средняя мощность помехи в канале.

Превышение  $H_c$  связано с возможностями передатчика и дальностью передачи. Чем больше  $H_c$ , тем меньше вероятность ошибочного приема. Аналогично объему канала вводится понятие объема (емкости)  $V_c$  передаваемого сигнала:

$$V_c = T_c F_c H_c. \quad (23.42)$$

Как объем сигнала, так и объем канала могут быть представлены в трехмерном пространстве с соответствующими координатами  $T, F, H$  (рис. 23.8).

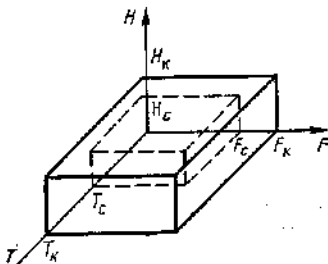


Рис. 23.8

Необходимым условием принципиальной возможности неискаженной передачи сигнала по данному каналу является выполнение соотношения

$$V_c \leq V_K. \quad (23.44)$$

При этом, однако, могут потребоваться преобразования для обеспечения достаточных условий передачи, а именно:

$$T_c \leq T_k, F_c \leq F_k, H_c \leq H_k.$$

(23.44)

Когда канал имеет меньшую полосу пропускания, чем практическая ширина спектра, подлежащего передаче сигнала, последнюю можно уменьшить за счет увеличения длительности сигнала. Объем сигнала при этом сохраняется неизменными. Практически такое преобразование можно осуществить, например, посредством записи сигнала на магнитную ленту с высокой скоростью и последующего воспроизведения со скоростью, при которой ширина его спектра равна полосе пропускания канала.

Если, наоборот, широкополосный канал предоставляется на время меньшее длительности сигнала, то согласование осуществляется за счет расширения спектра сигнала. Для реализации также может использоваться накопитель на магнитной ленте, однако в данном случае скорость воспроизведения должна быть выше скорости записи.

При низком допустимом уровне превышения сигнала в канале преобразование заключается в уменьшении уровня превышения передаваемого сигнала с одновременным увеличением его длительности путем многократного повторения передачи. Возможны и другие виды преобразования.

Рассмотрим, какова связь между объемом канала и количеством информации, которое можно получить о передаваемом по этому каналу сигнале.

В соответствии с выражением (23.40) предельное количество информации, которое может быть передано по каналу связи за время  $T_k$ ,

$$I_{\max}(V, U) = T_k F_k \log(1 + P_u/P_{\xi}).$$

Отсюда следует, что если  $P_u/P_{\xi} \gg 1$ , то при условии обеспечения посредством преобразования сигнала полного использования физических возможностей канала максимальное количество информации, которое можно получить о сигнале, близко к емкости канала:

$$I_{\max}(V, U) = V_k = T_k F_k \log(P_{u \max}/P_{\xi}).$$

### **23.8. Согласование статистических свойств источника сообщений и канала связи**

Согласование статистических свойств и отражающих их информационных характеристик источника сообщений и канала связи проводится с целью улучшения качества системы передачи. Оценка качества осуществляется по трем основным показателям: достоверности, средней скорости передачи и сложности технической реализации системы, определяющей ее стоимость и надежность. Хотя с точки зрения практики сложность технической реализации может иметь решающее значение, при определении предельных возможностей системы целесообразно ограничиться только первыми двумя показателями.

Достоверность дискретного канала обычно оценивается значением вероятности ошибочного приема одного символа (элементарного сигнала). В случае передачи непрерывных сообщений о достоверности судят по значению среднеквадратической ошибки при воспроизведении сообщения

$$M[\varepsilon^2] = M[(w(t) - z(t))^2],$$

где  $w(t)$  — сообщение, поступающее с выхода канала;  $z(t)$  — сообщение на входе канала.

Достоверность характеризует помехоустойчивость информационной системы.

Под скоростью передачи подразумевают среднее количество информации, передаваемое по каналу в единицу времени. Именно эта (а не техническая) скорость формирования символов подлежит согласованию с пропускной способностью канала.

Скорость передачи информации характеризует эффективность системы.

Если высоких требований в отношении скорости передачи и достоверности к системе передачи не предъявляется, то согласование статистических (информационных) характеристик источника сообщений и канала связи не является принципиально необходимым.

При преобразовании сообщений в сигналы в этом случае могут преследоваться две основные цели. Одна из них заключается в том, чтобы преобразовать сообщения в такую систему символов (код), чтобы она обеспечивала простоту и надежность аппаратурной реализации информационных устройств и приемлемую их эффективность: простоту аппаратуры различения элементарных сигналов, соответствующих отдельным символам, приемлемое время

при их передаче, простоту выполнения в этой системе арифметических и логических действий. Техническая реализация процесса кодирования в таком простейшем виде при непрерывном входном сигнале осуществляется аналого-цифровыми преобразователями.

Другой целью преобразования сообщений является защита их от несанкционированного доступа. Такое преобразование называют *шифрованием*. Оно может проводиться как на уровне знаков, так и на уровне символов.

В случае отсутствия необходимости в статистическом согласовании источника сообщений с каналом связи вопросы повышения качества функционирования системы решаются для дискретного канала от входа модулятора до выхода демодулятора.

Считается, что символы на вход модулятора поступают равновероятно и статистические связи между ними отсутствуют. Из множества сигналов, удовлетворяющих заданным ограничениям по мощности и полосе частот, для отображения символов отбираются такие, которые в предположении воздействия на них аддитивного гауссова шума обеспечивают наибольшую достоверность приема каждого отдельного символа. Одновременно определяется и структура оптимального приемника. Наиболее полно эти вопросы рассмотрены для случая двоичного канала ( $m = 2$ ).

Увеличение эффективности и помехоустойчивости системы передачи информации, как показал Шеннон, возможно за счет введения в канал связи кодирующего, а следовательно, и декодирующего устройств, цель которых состоит в статистическом согласовании свойств источника сообщений и канала связи.

Доказанными им теоремами обосновано существование оптимального способа кодирования, при котором достигается скорость передачи информации, сколь угодно близкая к пропускной способности данного канала связи.

***Под способом кодирования при этом подразумевается совокупность операций по преобразованию сообщений в сигналы и обратного преобразования смеси сигнала с помехами в сообщения, включая операции в части канала «модулятор-демодулятор».***

К сожалению, указанные теоремы не дают конструктивных рекомендаций относительно путей реализации оптимального способа кодирования. Определить соответствующую совокупность операций, а следовательно, и структуру оптимальной системы связи пока не удалось даже при ряде допущений, существенно упрощающих модели каналов. Для упрощения задачи переходят к оптимизации системы по частям путем нахождения наилучшего кода при условии оптимально спроектированной части канала «модулятор-демодулятор».



Выяснилась также целесообразность разделения процедур кодирования, обусловленных статистическими свойствами источника сообщений, и процедур кодирования, зависящих от статистических свойств канала связи. Такое разделение способствует лучшему пониманию существа процессов преобразования. С практической точки зрения оно ценно тем, что позволяет реализовать как кодирующее, так и декодирующее устройства из двух фактически независимых блоков: кодера КИ и декодера ДКИ источника и кодера КК и декодера ДКК канала. Уточненная структурная схема системы передачи дискретных сообщений показана на рис. 23.9.

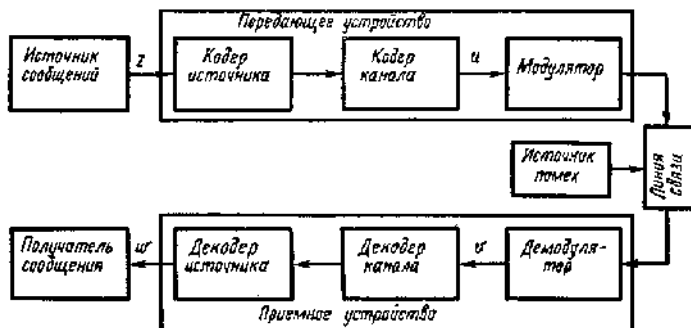


Рис. 23 9

Рассмотрим теперь особенности статистического согласования различных источников сообщений и каналов связи.

Предположим, что дискретные сообщения, поступающие с источника, обладают избыточностью, а вредным действием помех в канале можно пренебречь, что будет близко к реальности при отношении сигнал/помеха, значительно превышающем единицу. В этом случае учитывать проблему обеспечения помехоустойчивости нет необходимости и остается задача повышения эффективности.

В основной теореме Шеннона о кодировании для дискретного канала без помех утверждается, что посредством преобразования сообщений в статистически независимые и равновероятные символы можно повысить скорость передачи вплоть до пропускной способности этого канала (подробно о теореме и методах кодирования говорится в книге 5 настоящей работы).

Техническая реализация указанной возможности осуществляется кодером источника, обеспечивающим такое кодирование, при котором за счет устранения избыточности снижается среднее число символов, требующихся для выражения знака сообщения. При отсутствии помех

это непосредственно дает выигрыш во времени передачи (или в объеме запоминающего устройства), что повышает эффективность системы. Поэтому такое кодирование получило название эффективного или оптимального.

При наличии помех в канале оно позволяет преобразовать входную информацию в последовательность символов, наилучшим образом (в смысле максимального сжатия) подготовленную для дальнейших преобразований.

При статистическом согласовании источника, формирующего дискретные сообщения, не обладающие избыточностью, с каналом, подверженным действию помехи, использование кодера источника не имеет смысла. Однако для повышения достоверности передачи сообщений при минимальном сокращении скорости передачи по каналу дополнительную избыточность необходимо ввести так, чтобы она максимально способствовала устранению вредного действия помехи с определенными статистическими свойствами.

Из теоремы Шеннона о кодировании для дискретного канала с помехами следует неожиданное и фундаментальное заключение о том, что помехи в канале не накладывают ограничений на достоверность передачи. Ограничение накладывается только на скорость передачи, при которой может быть достигнута сколь угодно высокая достоверность. Она не должна превышать пропускной способности дискретного канала с помехами. Количество избыточной информации, необходимое для обеспечения достоверной передачи безыбыточных сообщений, невелико и равно потерям информации в канале, обусловленным действием помехи.

Техническая реализация возможности существенного повышения достоверности передачи осуществляется кодером и декодером канала. Такое кодирование получило название помехоустойчивого. Подробному рассмотрению указанной теоремы и методов помехоустойчивого кодирования посвящена книга 5 настоящей работы.

В общем случае, когда источник формирует сообщения, обладающие избыточностью, и требуется передавать их по каналу с помехами, целесообразно ввести в канал как кодер (и декодер) источника, так и кодер (и декодер) канала.

Целесообразность устранения избыточности сообщений методами эффективного кодирования с последующим перекодированием помехоустойчивым кодом обусловлена тем, что избыточность источника сообщения в большинстве случаев не согласована со статистическими закономерностями помехи в канале связи и поэтому не может быть полностью использована для повышения достоверности

принимаемого сообщения, тогда как обычно можно подобрать подходящий помехоустойчивый код. Кроме того, избыточность источника сообщений часто является следствием весьма сложных вероятностных зависимостей и позволяет обнаруживать и исправлять ошибки только после декодирования всего сообщения, пользуясь сложнейшими алгоритмами и интуицией.

Передача непрерывных сообщений по каналу без помех не рассматривается, поскольку в этом теоретическом случае проблема связи вообще не возникает. Одним импульсом, амплитуда которого на приемной стороне воспринимается с неограниченной точностью, может быть передано бесконечно большое количество информации, что с точки зрения практики абсурдно.

Несколько подробнее остановимся на статистическом согласовании источника непрерывных сообщений с непрерывным каналом, подверженным действию помех. Предельные возможности системы передачи в этом случае определяются следующей теоремой Шеннона:

*если  $\varepsilon$ -производительность  $\bar{H}_\varepsilon(Z)$  источника непрерывных сообщений не превышает пропускной способности непрерывного канала  $C_\infty$ , то существует такой способ передачи, при котором с вероятностью, сколь угодно близкой к единице, любое полученное сообщение будет отличаться от переданного только в пределах принятой оценки верности воспроизведения.*

Утверждается также, что при  $H_c(Z) > C_\infty$  такую передачу никаким способом обеспечить невозможно.

Не доказывая теорему, поясним возможность осуществления указанного в ней способа передачи, используя геометрическую форму представления сигналов.

Если сообщения должны воспроизводиться с определенной верностью, то из бесконечного множества непрерывных сообщений длительностью  $T$  передавать необходимо только счетное подмножество воспроизводящих сообщений.

Процесс кодирования в этом случае заключается в отождествлении полученного от источника сообщения с ближайшим воспроизводящим и сопоставлением ему конкретного сигнала из множества разрешенных сигналов, специально подобранных для передачи с учетом действующей в канале помехи.

При декодировании полученный сигнал отождествляется с ближайшим разрешенным и ставится в соответствие воспроизводящему сообщению. Ошибки не произойдет, если конец вектора принятого сигнала в гильбертовом пространстве попадет в собственную область соответствующего разрешенного сигнала,

размеры которой зависят от средней мощности помехи. Это накладывает ограничения на расстояния между концами векторов разрешенных сигналов. Таким образом, на поверхности гиперсферы, соответствующей определенному уровню средней мощности передаваемых сигналов, можно разместить только ограниченное число разрешенных сигналов. Оно и определяет предельную скорость передачи информации с обеспечением заданного уровня верности.

Поскольку обычно допускается возможность появления любого значения помехи, вероятность воспроизведения другого разрешенного сигнала остается конечной. Однако доказано, что она стремится к нулю при неограниченном увеличении длительности передаваемых сигналов.

### **23.9. Вычисление количества информации при передаче сообщений по дискретному каналу связи с шумами**

Мы уже говорили, что для некоторых идеализированных условий (отсутствие потерь, отсутствие взаимозависимости) количество информации при передаче сообщений определяется как произведение числа сообщений  $k$  на энтропию, приходящуюся на одно сообщение:

$$I = kH. \quad (23.46)$$

Для равновероятных независимых сообщений

$$H_1 = \log_2 m_1 = n \log_2 m_2 \text{ бит/символ}, \quad (23.47)$$

где  $m_1$  и  $m_2$  соответственно количество качественных признаков первичного и вторичного алфавитов.

Для неравновероятных, независимых сообщений

$$H_2(A) = - \sum_i p(a_i) \log_2 p(a_i) \text{ бит/символ}. \quad (23.48)$$

Для неравновероятных взаимозависимых сообщений энтропия может быть подсчитана при помощи следующих выражений:

при описании канала связи со стороны источника

$$H_3(B/A) = - \sum_i \sum_j p(a_i) p(b_j/a_i) \log_2 p(b_j/a_i) \text{ бит/символ}, \quad (23.49)$$

или

$$H_4(A/B) = - \sum_i \sum_j p(b_j) p(a_i/b_j) \log_2 p(a_i/b_j) \text{ бит/символ}; \quad (23.50)$$

при описании канала связи со стороны приемника сообщений

$$H_3(A, B) = H(B, A) = - \sum_i \sum_j p(a_i, b_j) \log_2 p(a_i/b_j) \text{ бит/два символа. (23.51)}$$

Нельзя утверждать, что количество информации всегда может быть определено путем непосредственного умножения числа переданных сообщений  $k$  на  $H_1 \div H_3$ . Для вычисления энтропии источника или приемника используются преимущественно выражения (23.47) и (23.48). Выражения (23.49—23.51) используются для вычисления энтропии систем, между элементами которых наблюдается взаимозависимость, а также для определения потерь в каналах связи. В последнем случае количество информации вычисляется как произведение числа переданных сообщений  $k$  на сумму или разность соответствующих энтропий.

Задача определения информационных потерь при передаче информации по каналам связи с шумами является одной из центральных задач теории информации, так как практически не существует системы передачи без аппаратурных помех или помех в каналах связи. Уровень помех может быть более или менее опасным по сравнению с уровнем передаваемого сигнала, но действие помех всегда следует учитывать.

Для удобства исследования помехи всегда считают сосредоточенными в линии связи, математическое описание которой задается в виде вероятностных характеристик сигнала на передающем и приемном концах. Графически влиянием помех может быть проиллюстрировано рис. 23.10.

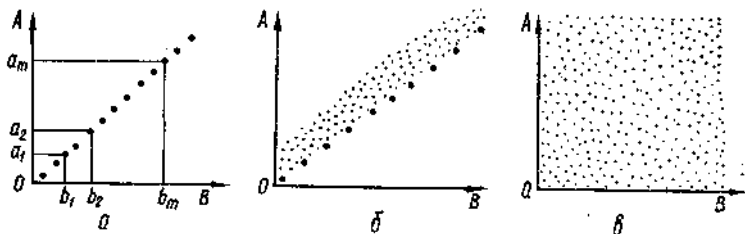


Рис. 23.10. Графическое представление различных уровней помех в каналах связи

Предположим, передаются два равновероятных сигнала  $a_1$  и  $a_2$ , т. е.  $p(a_1) = p(a_2) = 0,5$ . Если в канале связи нет помех (рис. 23.10, а), то на приемном конце мы получим сигналы  $b_1$  и  $b_2$ , причем  $p(b_1) = p(b_2)$ . Наличие помех в канале связи вводит неоднозначность и в ряде случаев может привести к тому, что при передаче  $a_1$  мы примем не  $b_1$ , а  $b_2$  (вместо сигнала 0 — сигнал 1, вместо сигнала отрицательной полярности — сигнал положительной полярности и т. д.), т. е.  $p(b_1)$  не

всегда равна  $p(b_2)$ . Например, если  $b_2$  — сигнал положительной полярности и помеха имеет положительную полярность (рис. 23.10, б), то  $p(b_1) \ll p(b_2)$ . Чем выше уровень помех, тем меньше будет связь, между статистическими характеристиками, описывающими функциональное взаимодействие между  $a_1, b_1$  и  $a_2, b_2$ , которые характеризуются условными вероятностями  $p(b_1/a_1)$  и  $p(b_2/a_2)$ , и больше будут вероятности ложных переходов  $p(b_2/a_1)$  и  $p(b_1/a_2)$ . Предельный случай — отсутствие полной статистической зависимости (рис. 23.10, в).

В общем случае, если передаются  $m$  сигналов  $A$  и ожидается получить  $m$  сигналов  $B$ , влияние помех в канале связи полностью описывается матрицами (13.27), (13.32).

Условная энтропия  $H(A/B)$  определяет количество недостающей информации на приемном конце в результате действия помех и выражает не только соответствие принятой буквы  $b_j$ , переданной  $a_i$ , но и какой-то другой переданной букве ансамбля сообщений, составленного из первичного алфавита. Например, принят сигнал  $b_j$ . В канале связи без помех с уверенностью можно было бы сказать, что был послан соответствующий ему символ  $a_j$ . Но наличие помех лишает нас полной уверенности в этом. Мы все еще предполагаем, что был послан сигнал  $a_i$ , но уже допускаем возможность, что мог быть послан и другой сигнал. Полученная информация меньше той, которую мы получили бы при отсутствии помех, на величину этой неуверенности, неопределенности, неоднозначности, неэквивалентности. Вот эта потеря информации и характеризуется распределением условных вероятностей вида  $p(a_i/b_j)$ .

Если будет принят символ  $b_j$ , то количество потерянной информации, неопределенность состояния на входе теперь будет выражаться энтропией распределения условных вероятностей  $j$ -го столбца канальной матрицы

$$H(A/b_j) = - \sum_{i=1}^m p(a_i/b_j) \log p(a_i/b_j).$$

При этом каждый столбец матрицы должен удовлетворять условию

$$p(a_1/b_j) + p(a_2/b_j) + \dots + p(a_i/b_j) + \dots + p(a_m/b_j) = 1. \tag{23.52}$$

Для определения среднего количества потерянной информации необходимо взять среднее значение условной энтропии  $H(a_i/b_j)$ :

$$H(A/B) = - \sum_j \sum_i p(b_j) p(a_i/b_j) \log p(a_i/b_j). \tag{23.53}$$

Используя свойство симметрии энтропии объединения, можно записать

$$H(A) + H(B/A) = H(B) + H(A/B). \quad (23.54)$$

Очевидно, что знак равенства не нарушается, если (23.54) запишем в виде

$$H(A) - H(A/B) = H(B) - H(B/A). \quad (23.55)$$

Обе половины равенства (23.55) могут быть использованы для подсчета количества информации, передаваемой по каналу связи с помехами. При этом, если вход канала  $A$ , а выход  $B$ , то матрица, описывающая канал связи, имеет вид (13.27) и используется левая часть равенства (2355), если вход  $B$  и выход  $A$ , то матрица имеет вид (13.32) и используется правая часть равенства (2355), естественно, подразумевается, что в первом случае заданы безусловные вероятности источника, а во втором — приемника сообщений.

В общем случае, когда был передан ансамбль сообщений с энтропией  $H(A)$  и получен ансамбль сообщений с энтропией  $H(B)$ , при наличии помех количество принятой информации

$$I(B, A) = H(A) - H(A/B). \quad (23.56)$$

Другими словами, количество информации, содержащееся в ансамбле принятых сообщений  $B$  относительно ансамбля переданных сообщений  $A$ , равно энтропии передаваемых сообщений  $H(A)$  минус потеря информации  $H(A/B)$ , вызванная действием помех. Величину  $H(A/B)$  определяют по формуле (23.53), а распределение частных условных вероятностей задается каналной матрицей.

**Пример.** Канал связи описан следующей матрицей:

$$p(b/a) = \begin{vmatrix} 0,98 & 0,01 & 0,01 \\ 0,1 & 0,75 & 0,15 \\ 0,2 & 0,3 & 0,5 \end{vmatrix}$$

Вычислить среднее количество информации, которое переносится одним символом сообщения, если вероятности появления символов источника сообщений равны  $p(a_1) = 0,7$ ;  $p(a_2) = 0,2$ ;  $p(a_3) = 0,1$ . Чему равны информационные потери при передаче сообщения из 400 символов алфавита  $a_1, a_2, a_3$ . Чему равно количество принятой информации?

**Решение:** 1) Энтропия источника сообщений

$$\begin{aligned} H(A) &= - \sum_{i=1}^m p(a_i) \log_2 p(a_i) = - (0,7 \log_2 0,7 + 0,2 \log_2 0,2 + 0,1 \log_2 0,1) = \\ &= 0,3602 + 0,4644 + 0,3322 = 1,1568 \text{ бит/символ.} \end{aligned}$$

2) Общая условная энтропия

$$H(B/A) = - \sum_{i=1}^m \sum_{j=1}^m p(a_i) p(b_j/a_i) \log_2 p(b_j/a_i) = [0,7 (0,98 \log_2 0,98 + 2 \cdot 0,01 \log_2 0,01) + 0,2 (0,75 \log_2 0,75 + 0,1 \log_2 0,1 + 0,15 \log_2 0,15) + 0,1 (0,2 \log_2 0,2 + 0,3 \log_2 0,3 + 0,5 \log_2 0,5)] = 0,465 \text{ бит/символ.}$$

3) Потери в канале связи

$$\Delta I = kH(B/A) = 400 \cdot 0,465 = 186 \text{ бит.}$$

4) Энтропия приемника

$$H(B) = - \sum_{j=1}^m p(b_j) \log_2 p(b_j);$$

$$p(b_1) = \sum_i p(a_i) p(b_1/a_i) = p(a_1) p(b_1/a_1) + p(a_2) p(b_1/a_2) + p(a_3) p(b_1/a_3) = 0,7 \cdot 0,98 + 0,2 \cdot 0,1 + 0,1 \cdot 0,2 = 0,726;$$

$$p(b_2) = p(a_1) p(b_2/a_1) + p(a_2) p(b_2/a_2) + p(a_3) p(b_2/a_3) = 0,7 \cdot 0,01 + 0,2 \cdot 0,75 + 0,1 \cdot 0,3 = 0,187;$$

$$p(b_3) = p(a_1) p(b_3/a_1) + p(a_2) p(b_3/a_2) + p(a_3) p(b_3/a_3) = 0,7 \cdot 0,01 + 0,2 \cdot 0,15 + 0,1 \cdot 0,5 = 0,087;$$

$$p(b_1) + p(b_2) + p(b_3) = 0,726 + 0,187 + 0,087 = 1. \quad \text{т. е.} \quad \sum_j p(b_j) = 1,$$

$$H(B) = - (0,726 \log_2 0,726 + 0,187 \log_2 0,187 + 0,087 \log_2 0,087) = 1,095 \text{ бит/символ.}$$

5) Среднее количество взаимной информации на 400 сообщений

$$I(A, B) = k [H(B) - H(B/A)] = kH(B) - \Delta I = 252 \text{ бит.}$$

Если в рассмотренном примере символам  $a_1, a_2, a_3$  на выходе источника сообщений будут соответствовать частоты передаваемых сигналов  $f_1, f_2, f_3$ , то по виду матрицы, описывающей канал связи, можно установить, что наилучшее прохождение в данном канале связи имеет сигнал частотой  $f_1$ , соответствующей символу  $a_1$  и наихудшее прохождение имеет сигнал частотой  $f_3$ , соответствующей символу  $a_3$ . Поэтому частота  $h$  присвоена символу с наибольшей вероятностью появления в сообщении. Если бы  $f_1$  присвоили символу  $a_3$ , а  $f_3$  символу  $a_1$ , то информационные потери в канале связи увеличились бы значительно, так как в этом случае  $H(B/A)$  была бы равна

$$H(B/A) = - [0,1 (0,98 \log_2 0,98 + 2 \cdot 0,01 \log_2 0,01) + 0,2 (0,75 \log_2 0,75 + 0,1 \log_2 0,1 + 0,15 \log_2 0,15) + 0,7 (0,2 \log_2 0,2 + 0,3 \log_2 0,3 + 0,5 \log_2 0,5)] = 1,25 \text{ бит/символ,}$$

т. е. потери увеличились более чем в 2,5 раза.

Если помехи отсутствуют или их уровень настолько низок, что они не в состоянии уничтожить сигнал или имитировать полезный сигнал в отсутствие передачи, то при передаче  $a_1$  мы будем твердо уверены, что получим  $b_1$ . Статистические характеристики, описывающие функци-



ональную связь событий  $A$  и  $B$ , жестко связаны, условная вероятность  $p(b_j/a_i) = 1$ , а условная энтропия

$$H(A/B) = 0.$$

В этом случае количество информации, содержащееся в принятом ансамбле сообщений  $B$ , равно энтропии передаваемых сообщений ансамбля  $A$ .

При очень высоком уровне помех любой из принятых сигналов  $b_j$  может соответствовать любому переданному сигналу  $a_i$ , статистическая связь между переданными и принятыми сигналами отсутствует. В этом случае вероятности  $p(a_i)$  и  $p(b_j)$  есть вероятности независимых событий. Известно, что вероятность совместного появления независимых событий равна произведению вероятностей этих событий. Следовательно,

$$p(a_i, b_j) = p(a_i) p(b_j) = p(b_j) p(a_i) = p(b_j, a_i). \quad (23.57)$$

С другой стороны, согласно принципу умножения вероятностей, вероятность совместного появления двух событий может быть представлена через условную вероятность появления одного из них, умноженную на условную вероятность другого относительно первого, т. е.

$$p(a_i, b_j) = p(a_i) p(b_j/a_i) = p(b_j) p(a_i/b_j). \quad (23.58)$$

Сопоставляя выражения (23.57) и (23.58), находим, что при отсутствии статистической связи между  $a_i$  и  $b_j$

$$p(p_j/a_i) = p(b_j); \quad (23.59)$$

$$p(a_i/b_j) = p(a_i).$$

Это вполне объяснимо, так как корреляция отсутствует и принятое  $b_j$  не означает, что было передано  $a_i$ , а переданное  $a_i$  не означает, что будет принято  $b_j$ .

Подставляя значения  $p(a_i/b_j)$  для случая отсутствия корреляции (23.59) в формулу (23.55), находим

$$\begin{aligned} H(A/B) &= - \sum_{i,j} p(b_j) p(a_i) \log p(a_i) = - \sum_j \sum_i p(b_j, a_i) \log p(a_i) = \\ &= \sum_i p(a_i) \log p(a_i) = H(A), \end{aligned}$$

так как

$$\sum_j p(b_j, a_i) = p(a_i).$$

Следовательно, для случая, когда уровень помех настолько высок, что полностью отсутствует статистическая связь между переданными и принятыми сигналами, условная энтропия равна безусловной, а количество информации, содержащееся в  $B$  относительно  $A$ , равно нулю:

$$I(A, B) = H(A) - H(A/B) = 0.$$

Информационные характеристики реальных каналов связи лежат между двумя предельными случаями — когда помехи полностью отсутствуют и когда уровень помех настолько высок, что любому из принятых сигналов может соответствовать любой переданный. Другими словами, несмотря на то что часть информации поражается помехами, между принятыми и переданными сообщениями существует статистическая взаимосвязь. Это позволяет описывать информационные характеристики реальных каналов связи при помощи энтропии объединения статистически зависимых событий.

Использовать энтропию объединения для вычисления среднего количества принятой информации удобно, если канал связи с помехами описан при помощи матрицы совместных вероятностей (23.59).

Свойство симметрии энтропии объединения позволяет записать

$$H(A, B) = H(B) + H(A/B) = H(A) + H(B/A), \quad (23.60)$$

отсюда

$$H(A/B) = H(A, B) - H(B); \quad (23.61)$$

$$H(B/A) = H(A, B) - H(A). \quad (23.62)$$

Подставляя (23.61) или (23.62) в соответствующую часть равенства (23.56), получим выражение для подсчета среднего количества информации при передаче сообщений по каналу связи с помехами непосредственно из матриц объединения

$$I(B, A) = H(A) + H(B) - H(B, A). \quad (23.63)$$

**Пример.** Канал связи с помехами описан матрицей

$$p(A, B) = \begin{vmatrix} 0,1 & 0,1 & 0 \\ 0 & 0,2 & 0,1 \\ 0 & 0,2 & 0,3 \end{vmatrix}.$$

Определить  $I(B, A)$ .

**Решение:** 1) Находим безусловные вероятности вида  $p(a_i)$  и  $p(b_j)$ :

$$p(a_i) = \sum_j p(a_i, b_j); \quad p(a_1) = 0,2; \quad p(a_2) = 0,3; \quad p(a_3) = 0,5;$$

$$p(b_j) = \sum_i p(a_i, b_j); \quad p(b_1) = 0,1; \quad p(b_2) = 0,5; \quad p(b_3) = 0,4.$$

2) Энтропия источника и приемника сообщений

$$H(A) = - \sum_i p(a_i) \log_2 p(a_i) = -(0,2 \log_2 0,2 + 0,3 \log_2 0,3 + 0,5 \log_2 0,5) = \\ = 0,4644 + 0,5211 + 0,5 = 1,4855 \text{ бит/символ.}$$

$$H(B) = - \sum_j p(b_j) \log_2 p(b_j) = -(0,1 \log_2 0,1 + 0,5 \log_2 0,5 + 0,4 \log_2 0,4) = \\ = 0,3321 + 0,5 + 0,5287 = 1,3608 \text{ бит/символ.}$$

3) Энтропия объединения

$$H(A, B) = - \sum_i \sum_j p(a_i, b_j) \log_2 p(a_i, b_j) = -(3 \cdot 0,1 \log_2 0,1 + 2 \cdot 0,2 \log_2 0,2 + \\ + 0,3 \log_2 0,3) = 0,9963 + 0,9288 + 0,5211 = 2,4462 \text{ бит/два символа.}$$

4) Среднее количество информации на сообщение

$$I(A, B) = H(A) + H(B) - H(B, A) = 1,4855 + 1,3608 - 2,4462 = 0,4 \text{ бит.}$$

Исследуем подробнее выражение (23.63). Для этого значения безусловных энтропии и энтропии объединения запишем в виде

$$H(A) = - \sum_i p(a_i) \log p(a_i);$$

$$H(B) = - \sum_j p(b_j) \log p(b_j);$$

$$H(A, B) = - \sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j) = \\ = - \sum_i \sum_j p(a_i) p(b_j/a_i) \log p(a_i) p(b_j/a_i),$$

и подставим в выражение (23.63), тогда

$$I(B, A) = - \sum_i p(a_i) \log p(a_i) - \sum_j p(b_j) \log p(b_j) + \\ + \sum_i \sum_j p(a_i) p(b_j/a_i) \log p(a_i) p(b_j/a_i).$$

Поскольку

$$p(a_i) = \sum_j p(a_i, b_j); \quad p(b_j) = \sum_i p(a_i, b_j), \quad \text{а } p(a_i, b_j) = \\ = p(a_i) p(b_j/a_i) = p(b_j) p(a_i/b_j),$$

то выражение для  $I(A, B)$  может быть записано в виде:

$$I(B, A) = - \sum_i \sum_j p(b_j/a_i) p(a_i) \log p(a_i) -$$

$$- \sum_i \sum_j p(a_i/b_j) p(b_j) \log p(b_j) + \sum_i \sum_j p(a_i) p(b_j/a_i) \log p(a_i) p(b_j/a_i).$$

Но

$$\sum_i \sum_j p(a_i) p(b_j/a_i) = \sum_i \sum_j p(b_j) p(a_i/b_j),$$

что позволяет записать

$$\begin{aligned}
 I(B, A) &= I(A, B) = - \sum_i \sum_j p(a_i) p(b_j/a_i) \log p(a_i) - \\
 &- \sum_i \sum_j p(a_i) p(b_j/a_i) \log p(b_j) + \sum_i \sum_j p(a_i) p(b_j/a_i) \log p(a_i) p(b_j/a_i).
 \end{aligned}
 \tag{23.64}$$

Тогда, используя свойство логарифмов, согласно которому

$$-k \log_a x - k \log_a y + k \log_a z = k \log_a \frac{z}{xy},$$

выражение (23.64) запишем в виде

$$\begin{aligned}
 I(B, A) &= \sum_i \sum_j p(a_i) p(b_j/a_i) \log \frac{p(a_i) p(b_j/a_i)}{p(a_i) p(b_j)} = \\
 &= \sum_i \sum_j p(a_i) p(b_j/a_i) \log \frac{p(b_j/a_i)}{p(b_j)}.
 \end{aligned}
 \tag{23.65}$$

Так как  $I(B, A) = I(A, B)$ , то справедливо и следующее выражение:

$$I(B, A) = \sum_i \sum_j p(b_j) p(a_i/b_j) \log \frac{p(a_i/b_j)}{p(a_i)}. \tag{23.66}$$

Выражения (23.65) и (23.66) позволяют определить среднее количество информации, содержащееся в принятом ансамбле сообщений  $B$  относительно переданного ансамбля сообщений  $A$  в условиях действия помех. Используя формулу (23.58), находим

$$\begin{aligned}
 I(A, B) &= I(B, A) = \sum_i \sum_j p(a_i, b_j) \log \frac{p(b_j/a_i)}{p(b_j)} = \\
 &= \sum_i \sum_j p(b_j, a_i) \log \frac{p(a_i/b_j)}{p(a_i)} = \sum_i \sum_j p(a_i, b_j) \log \frac{p(a_i, b_j)}{p(a_i) p(b_j)}.
 \end{aligned}
 \tag{23.67}$$

Для практических вычислений выражения (23.65÷67) можно применять в виде

$$I(B, A) = \sum_j \sum_i p(a_i) [p(a_i/b_j) \log p(a_i/b_j) - p(a_i/b_j) \log p(a_i)];$$

$$I(A, B) = \sum_i \sum_j p(b_j) [p(b_j/a_i) \log p(b_j/a_i) - p(b_j/a_i) \log p(b_j)];$$

$$\begin{aligned}
 I(A, B) &= I(B, A) = \sum_i \sum_j p(a_i, b_j) \log p(a_i, b_j) - \\
 &- \sum_i \sum_j p(a_i, b_j) \log p(a_i) p(b_j).
 \end{aligned}$$

Выражения (23.65) и (23.66) тем удобны, что для определения количества информации в принятом ансамбле  $B$  относительно переданного ансамбля  $A$  можно не прибегать к вычислению вероятности совместных событий  $A$  и  $B$ . Выражение (23.67) иллюстрирует возможность выражения среднего количества информации как через энтропию источника сообщений, так и через энтропию адресата. Так как  $I(A, B) = I(B, A)$ , то можно сказать, что количество информации, содержащееся в  $B$  относительно  $A$ , равно количеству информации, которое содержится в  $A$  относительно  $B$ . Как видим, количество информации, является характеристикой как источника сообщений  $A$ , так и адресата  $B$ . Количество информации характеризует взаимосвязь между источником сообщений и адресатом и является мерой соответствия принятых сигналов переданным.

**Выводы:** 1. При помощи матриц (13.27, 13.32) может быть подсчитана средняя условная энтропия для произвольного количества качественных признаков, а следовательно, и количество информации, передаваемой по симметричному каналу с шумами при любом основании кода. При помощи матрицы (13.43) могут быть подсчитаны все информационные характеристики канала связи.

2. Для уменьшения информационных потерь в канале связи символам источника сообщений, имеющим, наибольшую вероятность, следует присваивать качественные признаки, которые, согласно матрице, описывающей данный канал связи, имеют наименьшую частную условную энтропию.

3. При отсутствии помех энтропия приемника сообщений всегда равна энтропии источника, условная энтропия равна нулю, а взаимная энтропия (энтропия объединения) равна удвоенной энтропии источника сообщений и имеет размерность бит/два символа.

4. Количество информации может быть определено как со стороны источника сообщений, так и со стороны адресата. Являясь отражением одного объекта другим и мерой соответствия состояний объектов на передающем и приемном концах, информация обладает свойством симметрии.

## **23.10. Взаимная информация между произвольным числом дискретных и непрерывных ансамблей**

Теперь, когда усвоены основные понятия теории информации, можно перенести полученные результаты на более общие случаи и

отвлечься от привычных ансамблей источника и приемника сообщений.

Рассмотрим два дискретных выборочных пространства  $A$  и  $B$  (определение выборочного пространства дано в п. 13.4). Элементы пространства  $A$  обозначим через  $a_i$ , а пространства  $B$  — через  $b_j$ . Пространство, в котором только одна точка соответствует произвольным парам точек  $a_i, b_j$ , является *произведением пространств*  $A$  и  $B$ . Произведение двух пространств есть фигура двумерного пространства, в которой по столбцам располагаются точки пространства  $A$ , а по строкам — точки пространства  $B$  (рис. 23.11, а).

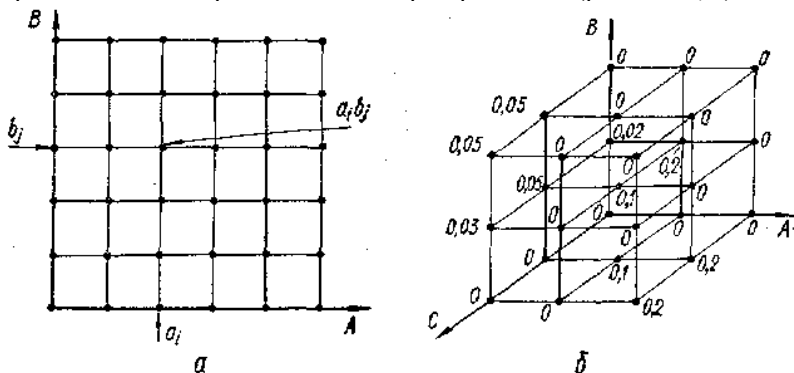


Рис. 23.11. Произведение пространств:  
а — двух —  $A$  и  $B$ ; б — трех —  $A$ ,  $B$  и  $C$

*Вероятностное выборочное пространство* задается значениями вероятностей состояний каждого элемента ансамбля. При этом ансамбль и его вероятностная мера могут представлять выборочное пространство, описывающее как выход некоторого абстрактного источника сообщений (или вход приемника), так и состояния элементов произвольной системы, например механической. Вероятность частного события дискретного ансамбля равна сумме вероятностей элементов выборочного пространства, связанных с этим событием. Другими словами, для ансамблей  $A$  и  $B$

$$p(a) = \sum_j p(a, b);$$

$$p(b) = \sum_i p(a, b),$$

а условные вероятности

$$p(b/a) = \frac{p(a, b)}{p(a)}; \quad (23.68)$$

$$p(a/b) = \frac{p(a, b)}{p(b)}.$$

В случае взаимодействия трех ансамблей  $A$ ,  $B$  и  $C$  произведение пространства  $ABC$  будет трехмерным и каждая из его точек будет образована в результате пересечения трех плоскостей и иметь координаты  $a_i, b_j, c_k$ . В *трехмерном* вероятностном выборочном пространстве: безусловные вероятности

$$\begin{aligned} p(a) &= \sum_j p(a, b) = \sum_k p(a, c) = \sum_k \sum_j p(a, b, c); \\ p(b) &= \sum_k p(b, c) = \sum_i p(b, a) = \sum_i \sum_k p(a, b, c); \quad (23.69) \\ p(c) &= \sum_i p(c, a) = \sum_j p(c, b) = \sum_i \sum_j p(a, b, c); \end{aligned}$$

условные вероятности

$$p(a/bc) = \frac{p(a, b, c)}{p(b, c)}; \quad p(b/ac) = \frac{p(a, b, c)}{p(a, c)}; \quad p(c/ba) = \frac{p(a, b, c)}{p(b, a)},$$

используя (23.68) и (23.69), запишем

$$\begin{aligned} p(a, b, c) &= p(a/bc) p(b, c) = p(a/bc) p(b) p(c/b) = p(b/ac) p(a, c) = \\ &= p(b/ac) p(a) p(c/a) = p(c/ba) p(a, c) = p(c/ba) p(c) p(a/c); \end{aligned}$$

частные условные энтропии второго порядка

$$H(A/bc) = - \sum_i p(a/bc) \log_2 p(a/bc) \text{ бит/символ};$$

$$H(B/ac) = - \sum_j p(b/ac) \log_2 p(b/ac) \text{ бит/символ};$$

$$H(C/ab) = - \sum_k p(c/ba) \log_2 p(c/ba) \text{ бит/символ};$$

общие условные энтропии второго порядка

$$H(A/BC) = - \sum_j \sum_k \sum_i p(b_j) p(c_k) p(a/bc) \log_2 p(a/bc) \text{ бит/символ};$$

$$H(B/AC) = - \sum_i \sum_k \sum_j p(a_i) p(c_k) p(b/ac) \log_2 p(b/ac) \text{ бит/символ};$$

$$H(C/AB) = - \sum_i \sum_j \sum_k p(a_i) p(b_j) p(c/ba) \log_2 p(c/ba) \text{ бит/символ};$$

взаимная энтропия

$$H(A, B, C) = - \sum_i \sum_j \sum_k p(a, b, c) \log_2 p(a, b, c) \text{ бит/три символа.}$$

$$\begin{aligned} H(A, B, C) &= - \sum_{i,j,k} p(a, b, c) \log p(a, b, c) = \\ &= - \sum_{i,j,k} p(a, b, c) \log p(b) p(c/b) p(a/bc) = - \left[ \sum_{i,j,k} p(a, b, c) \log p(b) + \right. \\ &\quad \left. + \sum_{i,j,k} p(a, b, c) \log p(c/b) + \sum_{i,j,k} p(a, b, c) \log p(a/bc) \right]. \end{aligned} \tag{23.70}$$

Рассмотрим отдельно каждое слагаемое выражения (23.70):

$$\begin{aligned} 1) \sum_{ik} p(a, b, c) &= p(b); \quad \sum_j p(b) \log p(b) = H(B); \\ &\quad \sum_{i,j,k} p(a, b, c) \log p(b) = H(B); \\ 2) \sum p(a, b, c) &= p(b, c) = p(b) p(c/b); \\ \sum_{i,k} p(b) p(c/b) \log p(c/b) &= H(C/B); \quad \sum_{i,j,k} p(a, b, c) \log p(c/b) = H(C/B); \\ 3) p(a, b, c) &= p(b) p(c/b) p(a/bc); \quad p(b) p(c/b) = p(bc); \\ \sum_k p(b, c) &= p(b); \quad \sum_i p(a/bc) \log p(a/bc) = H(a/bc); \quad \sum_j p(b, c) = p(c); \\ &\quad \sum_{i,j,k} p(b) p(c) p(a/bc) \log p(a/bc) = H(A/BC); \\ &\quad \sum_{i,j,k} p(a, b, c) \log p(a/bc) = H(A/BC); \\ H(A, B, C) &= H(B) + H(C/B) + H(A/BC). \end{aligned} \tag{23.71}$$

Подставляя в (23.70) значения  $p(a, b, c) = p(a) p(c/a) p(b/ac)$  и  $p(a, b, c) = p(c) p(a/c) p(c/ba)$ , а также учитывая, что  $H(A) + H(B/A) = H(B, A)$ ;  $H(B) + H(C/B) = H(B, C)$  и так далее, получим

$$\begin{aligned} H(A, B, C) &= H(A) + H(B/A) + H(C/AB) = H(A, B) + H(C/AB); \\ H(A, B, C) &= H(A) + H(C/A) + H(B/AC) = H(A, C) + H(B/AC); \\ H(A, B, C) &= H(B) + H(A/B) + H(C/BA) = H(B, A) + H(C/BA); \\ H(A, B, C) &= H(B) + H(C/B) + H(A/BC) = H(B, C) + H(A/BC); \\ H(A, B, C) &= H(C) + H(A/C) + H(B/CA) = H(C, A) + H(B/CA); \\ H(A, B, C) &= H(C) + H(B/C) + H(A/CB) = H(C, B) + H(A/CB). \end{aligned} \tag{23.72}$$

Используя (23.72), запишем

$$\begin{aligned} H(A) + H(B/A) + H(C/AB) &= H(B) + H(C/B) + H(A/BC) = \\ &= H(C) + H(A/C) + H(B/AC). \end{aligned}$$

Используя возможность переноса слагаемых с обратным знаком в разные части равенства, получим



$$\begin{aligned}
 H(A) - H(A/C) - H(B/AC) &= H(B) - H(B/A) - H(BIAB) = \\
 &= H(C) - H(C/B) - H(A/BC). \quad (23.73)
 \end{aligned}$$

Любая из частей равенства (23.73) может быть использована для вычисления среднего количества условной информации для случая трех взаимосвязанных вероятностных выборочных пространств.

**Пример.** Взаимодействие вероятностных пространств  $A, B, C$  представлено фигурой трехмерного пространства, которую будем называть куб вероятностей (рис. 23.11, б).

Найти информационные характеристики каждого из ансамблей  $A, B, C$ .

**Решение:** 1) Суммируя (сжимая) элементы куба вероятностей по  $A$ , находим для произведения пространства  $BC$

$$\sum_i p(b, c) = p(c)$$

$$p(b, c) = \left| \begin{array}{ccc|c} 0 & 0,22 & 0 & 0,22 \\ 0,05 & 0,15 & 0,3 & 0,5 \\ 0,05 & 0,03 & 0,2 & 0,28 \end{array} \right|$$

$$\sum_k p(b, c) = p(b) \left| \begin{array}{ccc|c} 0,1 & 0,4 & 0,5 & \end{array} \right|$$

2) Суммируя (сжимая) элементы куба вероятностей по  $B$ , находим для произведения пространств  $AC$

$$\sum_i p(a, c) = p(c)$$

$$p(a, c) = \left| \begin{array}{ccc|c} 0,02 & 0,2 & 0 & 0,22 \\ 0,1 & 0,2 & 0,2 & 0,5 \\ 0,08 & 0 & 0,2 & 0,28 \end{array} \right|$$

$$\sum_k p(a, c) = p(a) \left| \begin{array}{ccc|c} 0,2 & 0,4 & 0,4 & \end{array} \right|$$

3) Суммируя (сжимая) элементы куба вероятностей по  $C$ , находим для произведения пространств

$$\sum_i p(b, a) = p(b)$$

$$p(b, a) = \left| \begin{array}{ccc|c} 0,1 & 0 & 0 & 0,1 \\ 0,1 & 0,3 & 0 & 0,4 \\ 0 & 0,1 & 0,4 & 0,5 \end{array} \right|$$

$$\sum_k p(b, a) = p(a) \left| \begin{array}{ccc|c} 0,2 & 0,4 & 0,4 & \end{array} \right|$$

Условные вероятности для двух взаимодействующих ансамблей можно найти по данным матрицам, используя схемы рис. 13.12, 13.15.

Условные вероятности для трех взаимодействующих ансамблей  $A, B$  и  $C$  находятся следующим образом:

$p(a_0/b_0c_0) = p(a_0, b_0, c_0)/p(b_0, c_0);$ $p(a_0/b_0c_1) = p(a_0, b_0, c_1)/p(b_0, c_1);$ $p(a_0/b_0c_2) = p(a_0, b_0, c_2)/p(b_0, c_2);$ $p(a_1/b_1c_0) = p(a_1, b_1, c_0)/p(b_1, c_0);$ $p(a_1/b_1c_1) = p(a_1, b_1, c_1)/p(b_1, c_1);$ $p(a_1/b_1c_2) = p(a_1, b_1, c_2)/p(b_1, c_2);$ $p(a_2/b_2c_0) = p(a_2, b_2, c_0)/p(b_2, c_0);$ $p(a_2/b_2c_1) = p(a_2, b_2, c_1)/p(b_2, c_1);$ $p(a_2/b_2c_2) = p(a_2, b_2, c_2)/p(b_2, c_2);$ $p(c_0/b_0a_0) = p(c_0, b_0, a_0)/p(b_0, a_0);$ $p(c_0/b_0a_1) = p(c_0, b_0, a_1)/p(b_0, a_1);$ $p(c_0/b_0a_2) = p(c_0, b_0, a_2)/p(b_0, a_2);$	$p(b_0/a_0c_0) = p(b_0, a_0, c_0)/p(a_0, c_0);$ $p(b_0/a_0c_1) = p(b_0, a_0, c_1)/p(a_0, c_1);$ $p(b_0/a_0c_2) = p(b_0, a_0, c_2)/p(a_0, c_2);$ $p(b_1/a_1c_0) = p(b_1, a_1, c_0)/p(a_1, c_0);$ $p(b_1/a_1c_1) = p(b_1, a_1, c_1)/p(a_1, c_1);$ $p(b_1/a_1c_2) = p(b_1, a_1, c_2)/p(a_1, c_2);$ $p(b_2/a_2c_0) = p(b_2, a_2, c_0)/p(a_2, c_0);$ $p(b_2/a_2c_1) = p(b_2, a_2, c_1)/p(a_2, c_1);$ $p(b_2/a_2c_2) = p(b_2, a_2, c_2)/p(a_2, c_2);$ $p(c_1/b_1a_0) = p(c_1, b_1, a_0)/p(b_1, a_0);$ $p(c_1/b_1a_1) = p(c_1, b_1, a_1)/p(b_1, a_1);$ $p(c_1/b_1a_2) = p(c_1, b_1, a_2)/p(b_1, a_2);$ $p(c_2/b_2a_0) = p(c_2, b_2, a_0)/p(b_2, a_0);$ $p(c_2/b_2a_1) = p(c_2, b_2, a_1)/p(b_2, a_1);$ $p(c_2/b_2a_2) = p(c_2, b_2, a_2)/p(b_2, a_2).$
--	---

Вероятности совместных событий вида  $p(a, b, c)$  находятся непосредственно из куба вероятностей, в данном случае они заданы

$p(a_0, b_0, c_0) = 0;$	$p(a_0, b_1, c_0) = 0,02;$	$p(a_0, b_2, c_0) = 0;$
$p(a_1, b_0, c_0) = 0;$	$p(a_1, b_1, c_0) = 0,2;$	$p(a_1, b_2, c_0) = 0;$
$p(a_2, b_0, c_0) = 0;$	$p(a_2, b_1, c_0) = 0;$	$p(a_2, b_2, c_0) = 0;$
$p(a_0, b_0, c_1) = 0;$	$p(a_0, b_1, c_1) = 0,05;$	$p(a_0, b_2, c_1) = 0,05;$
$p(a_1, b_0, c_1) = 0,1;$	$p(a_1, b_1, c_1) = 0,1;$	$p(a_1, b_2, c_1) = 0;$
$p(a_2, b_0, c_1) = 0,2;$	$p(a_2, b_1, c_1) = 0;$	$p(a_2, b_2, c_1) = 0;$
$p(a_0, b_0, c_2) = 0;$	$p(a_0, b_1, c_2) = 0,03;$	$p(a_0, b_2, c_2) = 0,05;$
$p(a_1, b_0, c_2) = 0;$	$p(a_1, b_1, c_2) = 0;$	$p(a_1, b_2, c_2) = 0;$
$p(a_2, b_0, c_2) = 0,2;$	$p(a_2, b_1, c_2) = 0;$	$p(a_2, b_2, c_2) = 0;$

$$\sum_{i,j,k} p(a, b, c) = 1.$$

Зная вероятности вида  $p(a_1); p(a_1/a_2); p(a_1/a_2a_3); p(a, b, c)$ , нетрудно найти остальные информационные характеристики.

Для *четырёхмерного* вероятностного выборочного пространства безусловные вероятности

$$\begin{aligned}
 p(a) &= \sum_k p(a, c) = \sum_j p(a, b) = \sum_i p(a, d) = \sum_i \sum_k p(a, b, c) = \\
 &= \sum_k \sum_j p(a, c, d) = \sum_j \sum_i p(a, b, d) = \sum_i \sum_k \sum_j p(a, b, c, d); \\
 p(b) &= \sum_i p(b, a) = \sum_k p(b, c) = \sum_l p(b, d) = \dots = \\
 &= \sum_l \sum_k \sum_i p(a, b, c, d); \\
 p(c) &= \sum_j p(c, d) = \sum_i p(c, a) = \sum_l p(c, d) = \dots = \\
 &= \sum_j \sum_i \sum_l p(a, b, c, d); \\
 p(d) &= \sum_i p(d, a) = \sum_j p(d, b) = \sum_k p(d, c) = \dots = \\
 &= \sum_i \sum_j \sum_k p(a, b, c, d);
 \end{aligned}$$

взаимная энтропия

$$H(A, B, C, D) = - \sum_i \sum_j \sum_k \sum_l p(a, b, c, d) \log_2 p(a, b, c, d) \text{ (бит/четыре символа)}.$$

В общем случае для  $N$ -мерного вероятностного выборочного пространства, описанного совместным распределением

$$p(a_1, a_2, \dots, a_N), \quad p(a) = \sum_{i_1, \dots, i_{N-1}} p(a_1, a_2, \dots, a_N),$$

$$p(a_k, \dots, a_N/a_1, \dots, a_{k-1}) = \frac{p(a_1, \dots, a_N)}{p(a_1, \dots, a_{k-1})} \quad (k \leq n).$$

**Теорема.** *Взаимная энтропия дискретного ансамбля  $N$ -мерного вероятностного выборочного пространства равна сумме его безусловной энтропии и энтропии 1-го, 2-го, ...,  $N - 1$ -го порядка взаимодействующих с ним ансамблей этого же пространства.*

**Доказательство.** Вероятности совместных событий вида  $p(a_1, a_2, \dots, a_N)$  обладают свойством иерархической мультипликативности

$$\begin{aligned}
 p(a_1, a_2, \dots, a_N) &= p(a_1) p(a_2/a_1) p(a_3/a_1, a_2) \dots \times \\
 &\times p(a_N/a_1, a_2, \dots, a_{N-1}).
 \end{aligned} \tag{23.74}$$

Взаимная энтропия  $N$ -мерного вероятностного выборочного пространства

$$H(A_1, A_2, \dots, A_N) = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} p(a_1, a_2, \dots, a_N) \times \\ \times \log p(a_1, a_2, \dots, a_N). \quad (23.75)$$

Подставляя значение (23.74) в логарифмическую часть выражения (23.75) и учитывая (23.71), получим

$$H(A_1, A_2, \dots, A_N) = H(A_1) + H(A_2/A_1) + H(A_3/A_1, A_2) + \dots + \\ + H(A_N/A_1, A_2, \dots, A_{N-1}). \quad (23.76)$$

Произведения таких ансамблей могут быть представлены фигурами  $N$ -мерного пространства. Каждая точка такого вероятностного пространства образуется в результате пересечения  $N$  плоскостей.

**Пример.** Определить среднюю взаимную энтропию для ансамблей произведения вероятностных пространств  $A, B, C$ , представленного кубом вероятностей (рис. 23.11, б).

**Решение.** 1) Используя результаты, полученные в предыдущем примере, находим безусловные энтропии  $H(A)$ ,  $H(B)$  и  $H(C)$ :

$$H(A) = - \sum_i p(a_i) \log_2 p(a_i) = -(0,2 \log_2 0,2 + 2 \cdot 0,4 \log_2 0,4) = \\ = 0,4643 + 2 \cdot 0,5287 = 1,5217 \text{ бит/символ};$$

$$H(B) = - \sum_j p(b_j) \log_2 p(b_j) = -(0,1 \log_2 0,1 + 0,4 \log_2 0,4 + 0,5 \log_2 0,5) = \\ = 0,3321 + 0,5287 + 0,5 = 1,3608 \text{ бит/символ};$$

$$H(C) = - \sum_k p(c_k) \log_2 p(c_k) = -(0,22 \log_2 0,22 + 0,5 \log_2 0,5 + 0,28 \log_2 0,28) = \\ = 0,4805 + 0,5 + 0,5142 = 1,4947 \text{ бит/символ}.$$

2) Взаимная энтропия для трех ансамблей

$$H(A, B, C) = - \sum_i \sum_j \sum_k p(a_i, b_j, c_k) \log_2 p(a_i, b_j, c_k) = \\ = -(3 \cdot 0,05 \log_2 0,05 + 0,02 \log_2 0,02 + 0,03 \log_2 0,03 + 2 \cdot 0,1 \log_2 0,1 + \\ + 3 \cdot 0,2 \log_2 0,2) = 3 \cdot 0,2160 + 0,1128 + 0,1517 + 2 \cdot 0,3321 + 3 \cdot 0,4643 = \\ = 2,9696 \text{ бит/три символа}.$$

3) Взаимная энтропия для двух ансамблей

$$H(B, C) = H(C, B) = -(0,22 \log_2 0,22 + 2 \cdot 0,05 \log_2 0,05 + 0,15 \log_2 0,15 + \\ + 0,3 \log_2 0,3 + 0,03 \log_2 0,03 + 0,2 \log_2 0,2) = 0,4806 + 0,2161 + 0,2161 + \\ + 0,4105 + 0,5211 + 0,1517 + 0,4644 = 2,4605 \text{ бит/два символа};$$

$$H(A, C) = H(C, A) = -(0,02 \log_2 0,02 + 4 \cdot 0,2 \log_2 0,2 + 0,1 \log_2 0,1 + \\ + 0,08 \log_2 0,08) = 0,1129 + 1,8576 + 0,3322 + 0,2915 = \\ = 2,5942 \text{ бит/два символа};$$

$$H(B, A) = H(A, B) = -(3 \cdot 0,1 \log_2 0,1 + 0,3 \log_2 0,3 + 0,4 \log_2 0,4) = \\ = 0,9966 + 0,5211 + 0,5288 = 2,0463 \text{ бит/два символа}.$$

4) Условная энтропия первого порядка

$$H(A|C) = H(A, C) - H(C) = 2,5942 - 1,4947 = 1,0995 \text{ бит/символ};$$

$$H(A|B) = H(A, B) - H(B) = 2,0465 - 1,3608 = 0,6857 \text{ бит/символ};$$

$$H(B|A) = H(A, B) - H(A) = 2,0465 - 1,5217 = 0,5217 \text{ бит/символ};$$

$$H(C|A) = H(A, C) - H(A) = 2,5942 - 1,5217 = 1,0725 \text{ бит/символ};$$

$$H(C|B) = H(B, C) - H(B) = 2,4606 - 1,3608 = 1,0997 \text{ бит/символ}.$$

5) Условная энтропия второго порядка

$$H(C|AB) = H(C|BA) = H(A, B, C) - H(A, B) = 0,9231 \text{ бит/символ};$$

$$H(B|AC) = H(B|CA) = H(A, B, C) - H(A, C) = 0,3754 \text{ бит/символ};$$

$$H(A|BC) = H(A|CB) = H(A, B, C) - H(B, C) = 0,5091 \text{ бит/символ}.$$

*Проверка.* Используя (23.72) и подставляя значения энтропии первого и второго порядка, имеем

$$H(A, B, C) = H(A) + H(B|A) + H(C|AB) = 1,5217 + 0,5248 + 0,9231 = \\ = 2,9696 \text{ бит/три символа};$$

$$H(A, B, C) = H(B) + H(C|B) + H(A|BC) = 1,3608 + 1,0997 + 0,5091 = \\ = 2,9696 \text{ бит/три символа};$$

$$H(A, B, C) = H(C) + H(A|C) + H(B|AC) = 1,4947 + 1,0995 + 0,3754 = \\ = 2,9696 \text{ бит/три символа, что подтверждает правильность произведенных расчетов}.$$

Информационные характеристики элементов взаимосвязанных вероятностных пространств оцениваются по характеру измерения априорной и апостериорной вероятностей их элементов. При этом под априорной вероятностью подразумевается вероятность нахождения элементов дискретных пространств в том или ином исходном состоянии перед началом каждого опыта. Например, вероятность сообщения быть переданным до того, как на приемном конце о нем получены какие-либо сведения, будет априорной вероятностью, а вероятность того, что действительно было передано данное сообщение, после того как оно было получено, будет апостериорной вероятностью.

Предположим, нас интересует степень достоверности того, что при приеме  $b_i$  было передано  $a_i$ . Информация, устанавливающая степень достоверности интересующего нас факта, сводится к изменению

априорной вероятности  $p(a_i)$  до апостериорной вероятности  $p(a_i/b_j)$ . Количественной мерой этого изменения является логарифм отношения апостериорной вероятности к априорной.

Количество информации, содержащееся в событии  $b_j$  относительно события  $a_i$ ,

$$I(a_i, b_j) = \log \frac{p(a_i/b_j)}{p(a_i)}$$

Умножим числитель и знаменатель на  $b_j$ , тогда

$$\begin{aligned} I(a_i, b_j) &= \log \frac{p(b_j) p(a_i/b_j)}{p(b_j) p(a_i)} = \log \frac{p(a_i, b_j)}{p(b_j) p(a_i)} = \\ &= \log \frac{p(a_i) p(b_j/a_i)}{p(b_j) p(a_i)} = \log \frac{p(b_j/a_i)}{p(b_j)} = I(b_j, a_i), \end{aligned}$$

что лишний раз подчеркивает свойство симметрии информации. Благодаря именно этому свойству информация  $I(a_i, b_j)$  называется взаимной информацией между событиями  $a_i$  и  $b_j$ . Взаимная информация есть мера статистической связи между  $a_i$  и  $b_j$ . При полной статистической связи между  $a_i$  и  $b_j$  вероятность  $p(a_i/b_j) = 1$ , а

$$I(a_i, b_j) = \log \frac{1}{p(a_i)} = -\log p(a_i). \quad (23.77)$$

Выражение (23.77) представляет количество *собственной информации*, содержащейся в величине  $a_i$ .

При этом сама величина  $a_i$  может состоять из элементов, положение каждого из которых относительно других элементов имеет определенную информацию. При этом сумма собственных информационных элементов, представляющих величину  $a_i$ , равна собственной информации самой величины  $a_i$ .

**Пример.** Определить количество собственной информации в каждом символе принятого сообщения  $A_5$  и в сообщении в целом, если априорные вероятности появления сообщений на выходе источника заданы табл. 23.2:

*Таблица 23.2. К примеру*

Сообщения	Кодовые слова	Вероятности сообщений			
		$p^0$	$p^1$	$p^2$	$p^3$
$A_1$	0 0 0	1/2	0	0	0
$A_2$	0 0 1	1/4	0	0	0
$A_3$	0 1 0	1/8	0	0	0
$A_4$	0 1 1	1/16	0	0	0
$A_5$	1 0 0	1/16	1/3	1/2	1
$A_6$	1 0 1	1/16	1/3	1/2	0
$A_7$	1 1 0	1/32	1/6	0	0
$A_8$	1 1 1	1/32	1/6	0	0



$$I(a_1) = I(A_5, a_1) = \log \frac{1/3}{1/16} = \log 16/3.$$

Количество собственной информации во второй принятой кодовой посылке  $a_2$  сообщения  $A_5$  равно количеству добавочной информации, вносимой  $a_2$  при известном  $a_1$ .

$$I(a_2) = I(A_5; a_2/a_1) = \log \frac{1/2}{1/3} = \log 3/2.$$

Количество собственной информации в третьей принятой кодовой посылке равно количеству добавочной информации, вносимой  $a_3$  при известных  $a_1$  и  $a_2$ ,

$$I(a_3) = I(A_5; a_3/a_1a_2) = \log \frac{1}{1/2} = \log 2.$$

Количество собственной информации во всем сообщении  $A_5$  равно сумме собственных информации отдельных символов, составляющих кодовое слово, соответствующее  $A_5$ , и равно количеству взаимной информации каждой дискретной кодовой посылки  $a_1$ ;  $a_2$ ;  $a_3$  относительно сообщения  $A_5$ :

$$I(A_5) = I(A_5, a_1a_2a_3) = \log_2 \frac{16}{3} + \log_2 3/2 + \log_2 2 = 4 \text{ бит},$$

т. е.

$$I(A_5) = \log_2 1/p_5^0 = \log_2 1/16 = 4 \text{ бит},$$

что возможно в том случае, когда условные вероятности  $p(b/a) = 1$ , т. е. *при отсутствии помех кодовые слова всегда однозначно определяют сообщение.*

Нетрудно убедиться в том, что количество информации при передаче  $A_1$  равнялось бы 1 бит  $[I(A_1) = \log_2 \frac{1}{1/3}]$ , при передаче  $A_8$  равнялось бы 5 бит  $[I(A_8) = \log_2 1/32]$ .

Среднее же количество информации на сообщение, создаваемое данным источником,

$$\begin{aligned} H(A) &= \sum_i p_i \log_2 \frac{1}{p(a_i)} = 1/2 \log_2 2 + 1/8 \log_2 8 + \dots + 1/32 \log_2 1/32 = \\ &= 0,5 + 2 \cdot 0,75 + 2 \cdot 15 \cdot 625 = 2,3125 \text{ бит}. \end{aligned}$$

Как видим, среднее значение собственной информации элементов ансамбля есть энтропия ансамбля

$$H(A) = \sum_i p(a_i) \log \frac{1}{p(a_i)} = - \sum_i p(a_i) \log p(a_i).$$

Аналогично, *условная собственная информация*, содержащаяся в событии  $a_i$  при условии появления  $b_j$



$$I(a/b) = \log \frac{1}{p(a_i/b_j)} = -\log p(a_i/b_j).$$

Математическое ожидание этой величины на ансамбле  $AB$  есть общая условная энтропия

$$\begin{aligned} H(A/B) &= -\sum_i \sum_j p(a_i, b_j) \log p(a_i/b_j) = \\ &= -\sum_j \sum_i p(b_j) p(a_i/b_j) \log p(a_i/b_j). \end{aligned}$$

Взаимная информация, как и всякая случайная величина, имеет среднее значение, дисперсию, моменты любого порядка и производящую функцию моментов.

Среднее значение взаимной информации получается в результате усреднения частных значений по всему произведению ансамблей. Так для ансамблей  $A$  и  $B$

$$\begin{aligned} I(A, B) &= \sum_i \sum_j p(a_i, b_j) \log \frac{p(a_i/b_j)}{p(a_i)} = \\ &= \sum_i \sum_j p(b_j, a_i) \log \frac{p(b_j/a_i)}{p(b_j)} = \sum_i \sum_j p(a_i, b_j) \log \frac{p(a_i/b_j)}{p(a_i) p(b_j)}. \end{aligned} \tag{23.78}$$

Выражение (23.78) можно записать в виде

$$I(A, B) = \sum_i \sum_j p(a_i, b_j) \log p(a_i/b_j) - \sum_i \sum_j p(a_i, b_j) \log p(a_i).$$

Первое слагаемое

$$\sum_i \sum_j p(a_i, b_j) \log p(a_i/b_j) = \sum_j \sum_i p(b_j) p(a_i/b_j) \log p(a_i/b_j) = -H(A/B).$$

Второе слагаемое

$$-\sum_i \sum_j p(a_i, b_j) \log p(a_i) = -\sum_i p(a_i) \log p(a_i) = -H(A),$$

так как

$$\sum_j p(a_i, b_j) = p(a_i).$$

Теперь (23.78) можно записать в виде

$$I(A, B) = H(A) - H(A/B). \tag{23.79}$$

Выражение (23.79) лишний раз подтверждает, что количество информации служит мерой уменьшения неопределенности.

Какой бы ни был применен подход к изменению количества информации — статистический, алгоритмический или любой другой — количество информации исчисляется как разность, как уменьшение

разнообразия, как уменьшение некоторой исходной неопределенности — энтропии.

С точки зрения вещественно-энергетических процессов информация служит мерой уменьшения шумов, хаоса, энтропии. П. Бриллюэн рассматривает информацию как меру упорядоченности и называет ее нэгэнтропией, подчеркивая этим связь информации с уменьшением исходной неопределенности.

С точки зрения теории познания информацию можно трактовать как уменьшение незнания, как увеличение знаний. В этом смысле между информацией и знанием — прямая связь.

Определяя информационные характеристики вероятностных пространств, мы имели в виду дискретные пространства. Однако существует обширный класс задач, которые не могут быть решены при помощи безоговорочного использования выражений, полученных для дискретных пространств. К таким задачам относятся определение информационных характеристик физических каналов связи, передающих радиосигналы, задачи, в которых событие на входе и выходе исследуемого объекта есть временная функция в некотором непрерывном интервале и т. д.

Если квантование непрерывных сигналов происходит с частотой более высокой либо равной минимуму, определенному теоремой Котельникова, то при передаче таких сигналов информационных потерь не будет (влияние помех в данном случае не рассматриваем). Если взаимную информацию между точками непрерывного пространства рассматривать как предел, к которому стремится взаимная информация между конечными областями, стягивающимися к этим точкам, то для вычисления взаимной информации можно использовать выражения для дискретных пространств, заменив в них вероятности на соответствующие плотности распределения вероятностей.

Плотности вероятностей для совместного ансамбля  $AB$  непрерывных пространств  $A$  и  $B$  задаются равенствами

$$p_A(a_i) = \int_{-\infty}^{\infty} p_{AB}(a_i, b_j) db_j,$$
$$p_B(b_j) = \int_{-\infty}^{\infty} p_{AB}(a_i, b_j) da_i.$$

*Среднее значение взаимной информации*

$$\begin{aligned}
 I(A, B) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(a, b) \log \frac{p(a, b)}{p(a) p(b)} da db = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(a, b) \log \frac{p(a/b)}{p(a)} da db = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(a, b) \log \frac{p(b/a)}{p(b)} da db.
 \end{aligned}
 \tag{23.80}$$

Если  $a, b$  и  $c$  — случайные величины действительных конечных выборочных пространств  $A, B$  и  $C$ , а  $p(a, b, c)$  их совместная плотность вероятности, тогда *взаимная информация между  $a$  и  $b$  при заданном  $c$*

$$I(a, b/c) = \log \frac{p(a, b/c)}{p(a/c) p(b/c)}. \tag{23.81}$$

Средняя условная взаимная информация может рассматриваться как предел дробления на все меньшие части величины  $I(a, b/c)$  по осям  $a, b, c$  и описываться равенством

$$I(A, B/C) = \iiint p(a, b, c) \log \frac{p(a, b/c)}{p(a/c) p(b/c)} da db dc. \tag{23.82}$$

*Энтропия непрерывных ансамблей*, также как и для дискретных ансамблей, может рассматриваться как среднее значение собственной информации соответствующего ансамбля.

Так для ансамбля  $A$  с плотностью вероятностей  $p(a)$  энтропия

$$H(A) = - \int_{-\infty}^{\infty} p(a) \log p(a) da.$$

Условная энтропия

$$H(A/B) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(a, b) \log p(b/a) da db.$$

По аналогии с (23.56) и (23.63)  $I(B, A)$  для непрерывных ансамблей  $A$  и  $B$  вычисляется как

$$I(B, A) = H(A) - H(A/B) = H(B) - H(B/A) = H(A) + H(B) - H(B, A).$$

Эта величина может принимать как положительные, так и отрицательные значения и не может быть строго интерпретирована как средняя собственная информация.

**Выводы:** 1. Ансамбль и его вероятностная мера могут представлять выборочное пространство, описывающее состояние элементов произвольной системы.

2. Произведение  $N$  выборочных пространств есть фигура  $N$ -мерного пространства.

3. Для произведения вероятностных выборочных пространств значение взаимной энтропии растет по мере увеличения числа взаимодей-

ствующих ансамблей, т. е. взаимная энтропия обладает свойством иерархической аддитивности.

4. При вычислении взаимной информации между точками непрерывных пространств могут быть использованы соответствующие выражения для дискретных пространств, если в них вместо вероятностей подставить соответствующие плотности распределения вероятностей.

### 23.11. Импульсно-модулированные сигналы

В импульсной модуляции в качестве носителя модулированных сигналов используются последовательности импульсов, как правило – прямоугольных. В беспроводных системах передачи данных (в радиосвязи) эти последовательности заполняются высокочастотными колебаниями, создавая тем самым двойную модуляцию. Как правило, эти виды модуляции применяются при передаче дискретизированных данных. Для прямоугольных импульсов наиболее широко используются амплитудно-импульсная (АИМ) и широтно-импульсная (ШИМ) модуляция.

#### Амплитудно-импульсная модуляция (АИМ)

заключается в изменении приращения амплитуды импульсов пропорционально функции управляющего сигнала при постоянной длительности импульсов и периоде их следования:

$$U(t) = U_0 + k \cdot s(t), \quad \tau_n = \text{const}, \quad T = \text{const}. \quad (23.83)$$

Спектр АИМ рассмотрим на примере модулирования однотонального сигнала  $s(t)$ , приведенного на рис. 23.12.

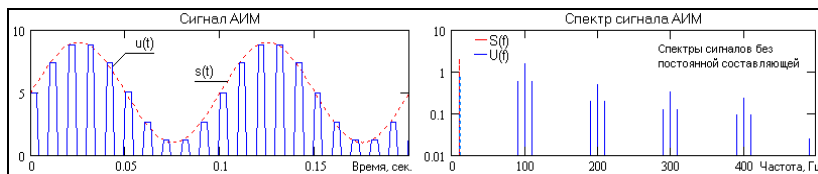


Рис. 23.12.

Напишем уравнение модулированного сигнала в следующей форме:

$$u(t) = (1+M \cos \Omega t) \cdot f(t), \quad (23.84)$$

где  $f(t)$  – периодическая последовательность прямоугольных импульсов с частотой  $\omega_0$ , которую можно аппроксимировать рядом Фурье (без учета фазы):

$$f(t) = U_0 + \sum_{n=1}^{\infty} U_n \cos n\omega_0 t. \quad (23.85)$$

Подставляя (23.85) в (23.84), получаем:

$$u(t) = (1+M \cos \Omega t)U_0 + \sum_{n=1}^{\infty} U_n \cos n\omega_0 t \cdot (1+M \cos \Omega t)$$

$$u(t) = U_0 + U_0 M \cos \Omega t + \sum_{n=1}^{\infty} U_n \cos n\omega_0 t +$$

$$+ 0.5M \sum_{n=1}^{\infty} U_n \cos (n\omega_0 + \Omega)t + 0.5M \sum_{n=1}^{\infty} U_n \cos (n\omega_0 - \Omega)t. \quad (23.86)$$

Форма спектра, в начальной части спектрального диапазона, приведена на рис. 23.12. В целом, спектр бесконечен, что определяется бесконечностью спектра прямоугольных импульсов. Около каждой гармоники  $n\omega_0$  спектра прямоугольных импульсов появляются боковые составляющие  $n\omega_0 \pm \Omega$ , соответствующие спектру моделирующей функции (при многотональном сигнале – боковые полосы спектров). При дополнительном высокочастотном заполнении импульсов весь спектр смещается в область высоких частот на частоту заполнения.

**Широтно-импульсная модуляция** (ШИМ, в английской терминологии pulse width modulation, PWM), которую иногда называют *модуляцией по длительности импульсов (ДИМ)*, заключается в управлении длительностью импульсов пропорционально функции управляющего сигнала при постоянной амплитуде импульсов и периоде следования по фронту импульсов:

$$\tau(t) = t_0 + k \cdot s(t), \quad U = \text{const}, \quad T = \text{const}. \quad (23.87)$$

Рассмотрим выполнение ШИМ в простейшем варианте на примере гармонического колебания, приведенного на рис. 23.13.

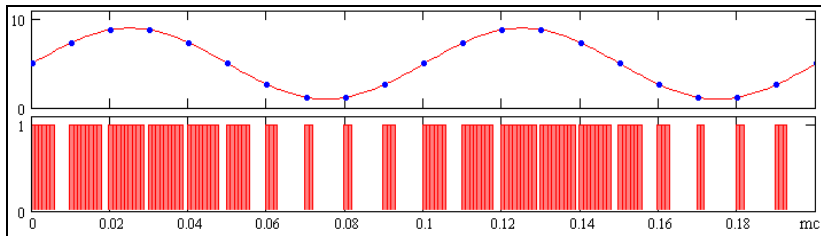


Рис. 23.13. Широтно-импульсная модуляция.

Передаваемая кривая дискретизируется, при этом имеет значение, как интервал дискретизации, так и количество уровней

квантования. При передаче данных прямоугольные импульсы начинаются в моменты дискретных отсчетов данных, а длительность импульсов устанавливается пропорциональной значению отсчетов, при этом максимальная длительность импульсов не должна превышать интервала дискретизации данных. Пример сформированных импульсов приведен на рис. 23.13 непосредственно под дискретизированной гармоникой, при этом число уровней квантования гармоники принято равным 8.

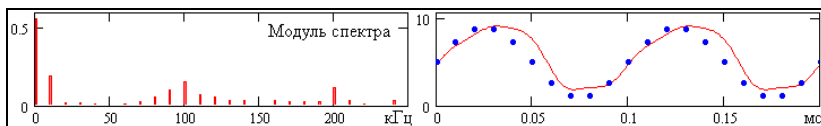


Рис. 23.14. Спектр ШИМ – сигнала.

Рис. 23.15. Восстановленный сигнал.

На рис. 23.14 приведен спектр сформированного сигнала ШИМ. В начальной части спектра он содержит постоянную составляющую среднего уровня сигнала и пик частоты гармоники, закодированной в ШИМ – сигнале. Если выделить из спектра эти две составляющие, то восстанавливается исходный сигнал с погрешностью квантования, приведенный на рис. 23.15. Естественно, что при малом числе уровней квантования погрешность восстановления исходного гармонического сигнала очень велика.

Попутно заметим, что широтно-импульсная модуляция с последующим выделением постоянной составляющей может весьма эффективно использоваться (и используется) для слежения за средним уровнем сигнала и автоматического регулирования его динамического диапазона, как, например, в системах установки громкости звука и яркости цветов и изображения в целом в современных телевизионных установках.

**Временная импульсная модуляция** (ВИМ) представляет собой девиацию импульсов по временной оси по закону модулирующего сигнала, и по существу аналогична угловой модуляции гармонической несущей. Она также может быть фазовой (ФИМ) или частотной (ЧИМ).

**Кодово-импульсная модуляция** заключается в том, что в точках дискретизации модулирующего сигнала производится квантование его значений и кодирование квантованных значений, как правило, в двоичной системе исчисления. Кодированные значения затем передаются при помощи соответствующей кодовой последовательности стандартных символов.

## 23.12. Модуляция символьных и кодов данных

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

Символьные последовательности являются дискретными квантованными сигналами, которые обычно передаются следующим образом. Каждому из возможных символов устанавливается определенный набор параметров несущего колебания, которые поддерживаются постоянными на интервале  $T$  до прихода следующего символа. Это означает преобразование последовательности чисел в ступенчатый сигнал (кусочно-постоянная интерполяция) который используется в качестве модулирующего сигнала. Соответственно, параметры несущего колебания, на которые переносится ступенчатый сигнал, также меняются скачкообразно. Такой способ модуляции несущей называется *манипуляцией* (keying), и может выполняться с использованием всех рассмотренных методов модулирования.

**Амплитудно-манипулированные сигналы** простейшего типа представляют собой последовательности радиоимпульсов, разделенные паузами. Такие сигналы используются в радиотелеграфии и в системах передачи дискретных данных. Форма огибающей радиоимпульсов в общем случае может быть произвольной, паузы могут отличаться по длительности от радиоимпульсов.

На рис. 23.16. приведен пример амплитудно-манипулированного сигнала:

$$u(t) = U_m \cdot \cos 2\pi f_0 t,$$

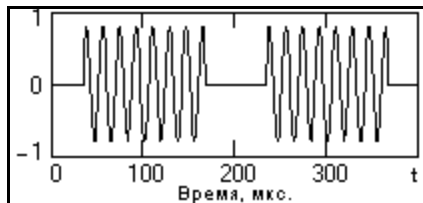


Рис. 23.16. АМП-сигнал.

с прямоугольной П-формой огибающей. Соответственно, в частотной области спектр АМП – сигнала образуется сверткой спектра огибающей функции (в данном случае – спектра прямоугольного

импульса) со спектром косинусного колебания (дельта - функции на частоте  $f_0$ ). Модуль спектральной плотности сигнала приведен на рис. 23.17

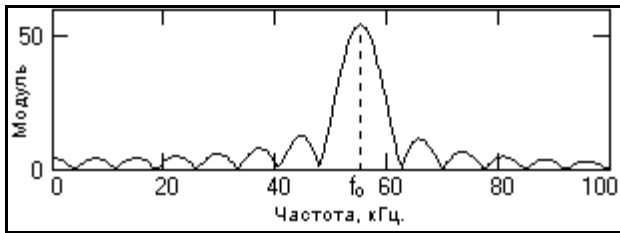


Рис. 23.17. Модуль спектра АМП-сигнала.

Спектр прямоугольного импульса довольно слабо затухает и простирается неограниченно далеко, а поэтому его использование в качестве огибающей АМП - сигнала не рекомендуется, хотя и является наиболее простым по техническому исполнению.

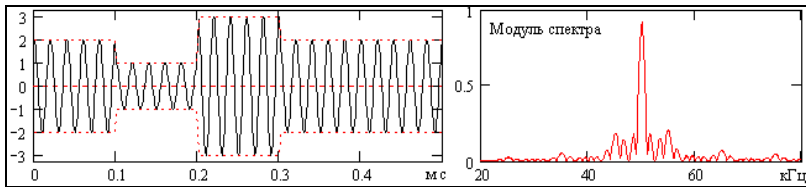


Рис. 23.18.

Рис. 23.19.

На рис. 23.18. приведен пример формы классического АМП сигнала при передаче нескольких символов, каждому из которых соответствует индивидуальная амплитуда несущей частоты при постоянной длительности интервалов послылки. Модуль спектра сигнала приведен на рис. 23.19 и тоже имеет достаточно большую ширину значимой части спектра вокруг несущей частоты.

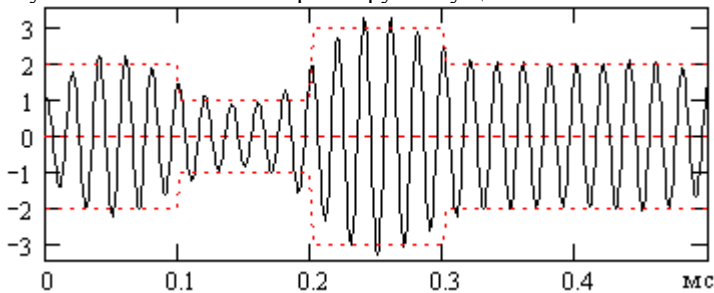


Рис. 23.20.



Естественно, что при передаче данных частотный диапазон канала передачи данных ограничивается значимой частью спектра, ширина которого устанавливается по допустимой степени искажения приемных сигналов. Степень искажения сигналов существенно зависит от длительности посылок. Пример искажения вышеприведенного сигнала при ограничении спектра интервалом 40-60 кГц приведен на рис. 23.20.

**Угловая манипуляция**, как правило, использует частотные методы модулирования, в которых каждому возможному значению передаваемого символа сопоставляется индивидуальное значение частоты гармонической несущей. При этом в точках сопряжения интервалов посылок могут происходить скачки напряжения, с соответствующим усложнением спектра модулированного сигнала. Самый простой способ – синусоидальное начало несущей на каждом интервале с кратным количеством периодов несущей в посылке. При более сложных способах, независимых от точного сопряжения несущих частот с интервалами посылок, осуществляется управление скоростью изменения фазы несущих на границах посылок.

Демодуляция сигналов осуществляется корреляционными методами. Сущность методов – вычисление взаимной корреляции между принимаемым сигналом и набором опорных частот, используемых при модулировании, с идентификацией символов по максимумам взаимной корреляции.

Для повышения помехоустойчивости передачи данных желательно, чтобы разносимвольные посылки были некоррелированы. Если для бинарных символов 0 и 1 принять частоты посылок равными

$$s_0(t) = \cos \omega_0(t), \quad s_1(t) = \cos \omega_1(t),$$

то их ВКФ при нулевом временном сдвиге определится выражением:

$$B_{01}(0) = \int_0^T s_0(t)s_1(t) dt = \frac{1}{2}(\sin(\omega_1 + \omega_0)T)/(\omega_1 + \omega_0) + \frac{1}{2}(\sin(\omega_1 - \omega_0)T)/(\omega_1 - \omega_0).$$

При  $(\omega_1 + \omega_0)T \gg 1$  первым слагаемым можно пренебречь, оно много меньше второго. А второе слагаемое обращается в нуль при  $(\omega_1 - \omega_0)T = \pi k$ , где  $k = 1, 2, \dots$  – целое число. Отсюда, минимальное значение между частотами манипуляции для некоррелированных посылок определяется выражениями:

$$\Delta\omega_{\min} = \pi/T, \quad \Delta f_{\min} = 1/2T = f_T/2,$$

где  $f_T$  – символьная скорость.

Фазовая манипуляция применяется значительно реже, в связи со значительными сложностями измерения абсолютных значений начальных фаз в посылках. Проще определяется относительный

фазовый сдвиг в соседних посылках, поэтому обычно используется фазоразностная манипуляция.

### **23.13. Введение в модулированные сигналы**

Сигналы от измерительных датчиков и любых других источников информации передаются по линиям связи к приемникам - измерительным приборам, в измерительно-вычислительные системы регистрации и обработки данных, в любые другие центры накопления и хранения данных. Как правило, информационные сигналы являются низкочастотными и ограниченными по ширине спектра. Каналы связи, напротив, являются высокочастотными, широкополосными и рассчитаны на передачу сигналов от множества источников одновременно с частотным разделением каналов. Перенос спектра сигналов из низкочастотной области в выделенную для их передачи область высоких частот выполняется операцией *модуляции*.

Допустим, что низкочастотный сигнал, подлежащий передаче по каналу связи, задается функцией  $s(t)$ . В канале связи для передачи данного сигнала выделяется определенный диапазон высоких частот. На входе канала связи в специальном передающем устройстве формируется вспомогательный, как правило, непрерывный во времени периодический высокочастотный сигнал  $u(t)=f(t; a_1, a_2, \dots a_m)$ . Совокупность параметров  $a_i$  определяет форму вспомогательного сигнала. Значения параметров  $a_i$  в отсутствие модуляции являются величинами постоянными. Если на один из этих параметров перенести сигнал  $s(t)$ , т.е. сделать его значение пропорционально зависимым от значения  $s(t)$  во времени (или по любой другой независимой переменной), то форма сигнала  $u(t)$  приобретает новое свойство. Она несет информацию, тождественную информации в сигнале  $s(t)$ . Поэтому сигнал  $u(t)$  называют несущим сигналом, *несущим колебанием* или просто *несущей* (carrier), а процесс переноса информации на параметры несущего сигнала – его *модуляцией* (modulation). Информационный сигнал  $s(t)$  называют *модулирующим* (modulating signal), результат модуляции – *модулированным сигналом* (modulated signal). Обратную операцию выделения модулирующего сигнала из модулированного колебания называют *демодуляцией* (demodulation).

Основным видом несущих сигналов являются гармонические колебания:

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

которые имеют три свободных параметра:  $U$ ,  $\omega$  и  $\varphi$ . В зависимости от того, на какой из данных параметров переносится информация,

различают *амплитудную* (АМ), *частотную* (ЧМ) или *фазовую* (ФМ) *модуляцию* несущего сигнала. Частотная и фазовая модуляция взаимосвязаны, поскольку изменяют аргумент функции косинуса, и их обычно объединяют под общим названием - *угловая* модуляция (angle modulation). В каналах передачи цифровой информации получила также распространение *квадратурная* модуляция, при которой одновременно изменяются амплитуда и фаза несущих колебаний.

При использовании в качестве несущих сигналов периодических последовательностей импульсов свободными параметрами модуляции могут быть амплитуда, длительность, частота следования импульсов и фаза (положение импульса относительно определенной точки тактового интервала). Это дает четыре основных вида импульсной модуляции: АИМ, ДИМ, ЧИМ и ФИМ.

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

## **23.14. Амплитудная модуляция**

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

АМ соответствует переносу информации  $s(t) \Rightarrow U(t)$  при постоянных значениях параметров несущей частоты  $\omega$  и фазы  $\varphi$ . АМ – сигнал представляет собой произведение информационной огибающей  $U(t)$  и гармонического колебания ее заполнения. Форма записи амплитудно-модулированного сигнала:

$$u(t) = U(t) \cdot \cos(\omega_0 t + \varphi_0), \quad (23.88)$$

$$U(t) = U_m \cdot [1 + M \cdot s(t)], \quad (23.89)$$

где  $U_m$  – постоянная амплитуда несущего колебания при отсутствии модулирующего сигнала  $s(t)$ ,  $M$  – коэффициент амплитудной модуляции.

Значение  $M$  характеризует *глубину* амплитудной модуляции. В простейшем случае, если модулирующий сигнал представлен одночастотным гармоническим колебанием с амплитудой  $S_0$ , то коэффициент модуляции равен отношению амплитуд модулирующего и несущего колебания  $M = S_0 / U_m$ . Значение  $M$  должно находиться в

пределах от 0 до 1 для всех гармоник модулирующего сигнала. При значении  $M < 1$  форма огибающей несущего колебания полностью повторяет форму модулирующего сигнала  $s(t)$ , что можно видеть на рис. 23.21. Малую глубину модуляции  $M \ll 1$  для основных гармоник модулирующего сигнала применять нецелесообразно, т.к. при этом мощность передаваемого информационного сигнала будет много меньше мощности несущего колебания и мощность передатчика будет использоваться неэкономично.

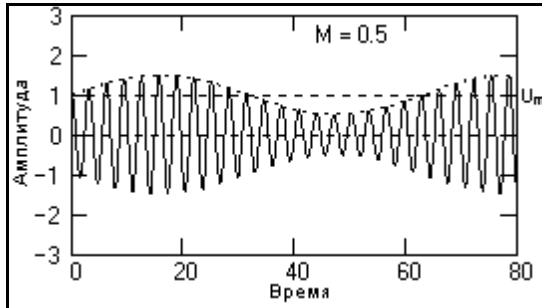


Рис. 23.21. Модулированный сигнал.

На рис. 23.22 приведен пример *глубокой модуляции*, при которой значение  $M$  стремится к 1.

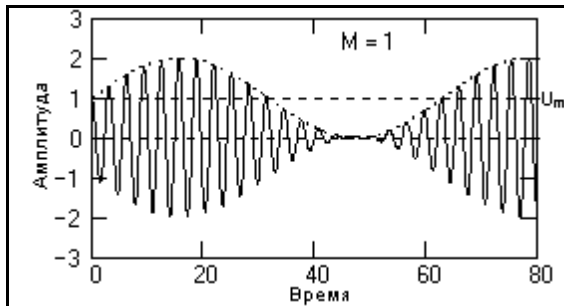


Рис. 23.22. Глубокая модуляция

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

При  $M > 1$  возникает так называемая *перемодуляция*, пример

которой приведен на рис. 23.23.

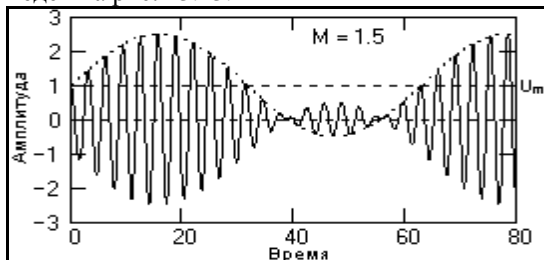


Рис. 23.23. Перемодуляция сигнала

Форма огибающей при перемодуляции искажается относительно формы модулирующего сигнала, и после демодуляции, если применяются ее простейшие методы, информация может быть искажена.

**Однотональная модуляция.**

Простейшая форма модулированного сигнала создается при модуляции несущего сигнала гармоническим колебанием с одной частотой  $\Omega$ :

$$u(t) = U_m [1 + M \cdot \cos \Omega t] \cdot \cos \omega_0 t. \quad (23.90)$$

Значения начальных фазовых углов несущего и модулирующего колебания для упрощения выражений будем принимать равными нулю, если они не имеет принципиального значения. С учетом формулы

$$\cos(x) \cdot \cos(y) = (1/2)[\cos(x+y) + \cos(x-y)],$$

из выражения (23.90) получаем:

$$u(t) = U_m \cos \omega_0 t + (U_m M / 2) \cos[(\omega_0 + \Omega)t] + (U_m M / 2) \cos[(\omega_0 - \Omega)t]. \quad (23.91)$$

Отсюда следует, что модулирующее колебание с частотой  $\Omega$  перемещается в область частоты  $\omega_0$  и расщепляется на два колебания, симметричные относительно частоты  $\omega_0$ , с частотами соответственно  $(\omega_0 + \Omega)$  – верхняя боковая частота, и  $(\omega_0 - \Omega)$  – нижняя боковая частота (рис. 23.24 для сигнала, приведенного на рис. 23.21).

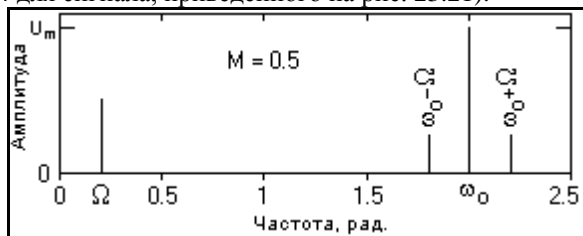


Рис. 23.24. Физические спектры сигналов.

Амплитуды колебаний на боковых частотах равны друг другу, и при 100%-ной модуляции равны половине амплитуды колебаний несущей частоты. Если получить уравнение (23.96) с учетом начальных фаз несущей и модулирующей частоты, то правило изменения фаз аналогично изменению частоты: начальная фаза модулирующего колебания для верхней боковой частоты складывается с начальной фазой несущей, для нижней – вычитаются из фазы несущей. Физическая ширина спектра модулированного сигнала в два раза больше ширины спектра сигнала модуляции.

**Энергия однотонового АМ-сигнала.** Обозначим раздельными индексами (*нес*- несущая, *вб*- верхняя боковая, *нб*- нижняя боковая) составляющие колебания в левой части выражения (23.91) однотонового АМ-сигнала и определим функцию его мгновенной мощности:

$$u(t) = u_{\text{нес}}(t) + u_{\text{вб}}(t) + u_{\text{нб}}(t). \\ p(t) = u_{\text{нес}}^2(t) + u_{\text{вб}}^2(t) + u_{\text{нб}}^2(t) + 2u_{\text{нес}}(t)u_{\text{вб}}(t) + 2u_{\text{нес}}(t)u_{\text{нб}}(t) + 2u_{\text{вб}}(t)u_{\text{нб}}(t). \quad (23.92)$$

Для определения средней мощности сигнала выполним усреднение функции  $p(t)$ :

$$P_u = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p(t) dt$$

Все взаимные мощности модулированного сигнала при усреднении становятся равными нулю (спектры не перекрываются), при этом:

$$P_u = P_{\text{нес}} + P_{\text{вб}} + P_{\text{нб}} = U_m^2/2 + (U_m M)^2/4. \quad (23.93)$$

Доля мощности боковых частот в единицах мощности несущей частоты:

$$(P_{\text{вб}} + P_{\text{нб}})/P_{\text{нес}} = M^2/2, \quad (23.94)$$

т.е. не превышает 50% даже при 100%-ной модуляции.

Под полезной мощностью модулированных сигналов понимают мощность боковых частот, несущих информацию. Коэффициент полезного действия модуляции определяется отношением мощности боковых частот к общей мощности модулированного сигнала:

$$\eta_{\text{АМ}} = (U_m^2 M^2/4) / P_u = M^2/(M^2+2). \quad (23.95)$$

Как можно видеть на рис. 23.25, даже при  $M=1$  КПД амплитудной модуляции составляет только 33%, а при практическом использовании обычно меньше 20%.

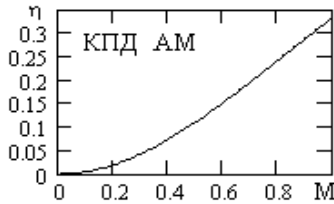


Рис. 23.25.

Для модулированных сигналов применяют также понятие пиковой мощности  $P_{\max}$ . Значение пиковой мощности для однотонового АМ-сигнала:

$$P_{\max} = U_m^2 (1+M)^2.$$

**Многотональный модулирующий сигнал** имеет произвольный спектральный состав. Математическая модель такого сигнала может быть аппроксимирована тригонометрической суммой гармонических составляющих, в пределе бесконечной:

$$s(t, n) = \sum_n a_n \cos(\Omega_n t + \Phi_n), \quad (23.96)$$

где значения амплитуд  $a_n$  и начальных фаз  $\Phi_n$  упорядоченной возрастающей последовательности гармоник  $\Omega_n$  произвольны. Подставляя (23.96) в (23.89) и заменяя произведения  $M \cdot a_n$  парциальными (частичными) коэффициентами модуляции  $M_n = M \cdot a_n$ , получим обобщенное уравнение амплитудно-модулированного сигнала и его физического спектра:

$$u(t) = U_m [1 + \sum_n M_n \cos(\Omega_n t + \Phi_n)] \cdot \cos \omega_0 t. \quad (23.97)$$

$$u(t) = U_m \cos \omega_0 t + (U_m/2) \sum_n M_n \cos[(\omega_0 + \Omega_n)t + \Phi_n] + \sum_n M_n \cos[(\omega_0 - \Omega_n)t - \Phi_n].$$

На рис. 23.26 приведен схематический пример амплитудных спектров модулирующего и АМ-сигналов при многотональной модуляции.

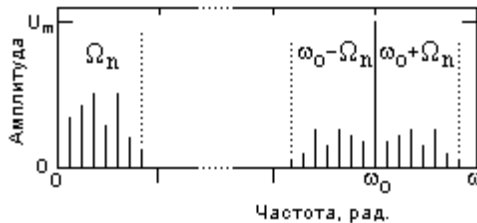


Рис. 23.26. Многотональная модуляция.

Он также содержит полосы верхних и нижних боковых частот

относительно несущей частоты  $\omega_0$ , являющихся прямой и зеркальной масштабными копиями модулирующего сигнала. Полная ширина спектра АМ-сигнала равна удвоенной ширине спектра модулирующего сигнала.

---

**Пример.** Частотный диапазон одного километра каротажного кабеля 0-200 кГц. Частотный диапазон измерительных датчиков скважинного прибора 0-5 кГц. От какого количества датчиков одновременно может передаваться информация по данному каротажному кабелю?

Минимальная несущая частота должна быть на порядок выше максимальной частоты модулирующего сигнала, т.е. порядка 50 кГц. Для передачи сигнала от одного датчика потребуется полоса частот  $2 \cdot 5 = 10$  кГц плюс пустой защитный интервал для исключения перекрестных помех порядка 1 кГц, т.е. 11 кГц. Общее количество каналов передачи информации:  $(200-50-5)/11 = 13$  каналов.

---

В соответствии огибающей модулированного сигнала форме модулирующего сообщения нетрудно убедиться вычислением модуля аналитического сигнала  $z(t) = u(t) + \tilde{u}(t)$  (см. раздел "Аналитические сигналы").

При

$$u(t) = U_m [1 + \sum_n M_n \cdot s(t, n)] \cos \omega_0(t),$$

квадратурное дополнение сигнала определяется преобразованием Гильберта и равно

$$\tilde{u}(t) = U_m [1 + \sum_n M_n \cdot s(t, n)] \sin \omega_0(t).$$

Огибающая сигнала:

$$\begin{aligned} |z(t)| &= \sqrt{u^2(t) + \tilde{u}^2(t)} = \\ &= \sqrt{U_m^2 [1 + \sum_n M_n s(t, n)]^2 (\cos^2 \omega_0(t) + \sin^2 \omega_0(t))} = \\ &= U_m [1 + \sum_n M_n \cdot s(t, n)]. \end{aligned} \quad (23.98)$$

**Автокорреляционная функция** АМ-сигналов:

$$B_u(\tau) = \int_{-\infty}^{\infty} u(t) u(t-\tau) dt. \quad (23.99)$$

Постоянная фаза сигнала не влияет на форму АКФ. При



$u(t)=U(t) \cdot \cos \omega_0 t$  получаем:

$$\cos \omega_0 t \cdot \cos \omega_0(t-\tau) = 0.5 \cos \omega_0 \tau + 0.5 \cos \omega_0(2t-\tau).$$

$$B_u(\tau) = \frac{\cos \omega_0 \tau}{2} \int_{-\infty}^{\infty} U(t)U(t-\tau) dt + 0.5 \int_{-\infty}^{\infty} U(t)U(t-\tau) \cos \omega_0(2t-\tau) dt. \quad (23.100)$$

Второй интеграл в формуле АКФ существенно меньше первого (произведение медленно меняющейся функции  $U(t)U(t-\tau)$  и сильно осциллирующего члена с частотой  $2\omega_0$ ) и им можно пренебречь. Первый интеграл представляет собой АКФ сигнала  $U(t)$ . Отсюда:

$$B_u(\tau) \cong \frac{\cos \omega_0 \tau}{2} B_U(\tau). \quad (23.101)$$

Полная энергия сигнала за счет усреднения по высокочастотным колебаниям:

$$B_u(0) \cong (1/2) B_U(0).$$

При бесконечной энергии сигнала:

$$B_u(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T u(t) u(t-\tau) dt. \quad (23.102)$$

На рис. 23.27 приведена типичная форма автокорреляционных функций однотонального модулированного сигнала при  $M=1$  и  $U_m=1$ .

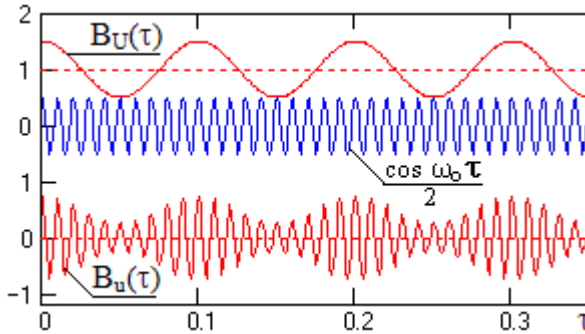


Рис. 23.27.

**Демодуляция АМ-сигналов** может выполняться несколькими способами.

Самый простой способ – двухполупериодное детектирование (вычисление модуля сигнала) с последующим сглаживанием однополярных полупериодов несущей фильтром низких частот.

На рис. 23.28 приведен пример изменения однотонального

амплитдно-модулированного сигнала и его физического спектра при детектировании (в реальной односторонней шкале частот и в реальной шкале амплитудных значений гармоник колебаний). Параметры представленного сигнала: несущая частота 30 Гц, частота модуляции 3 Гц, коэффициент модуляции  $M=1$ .

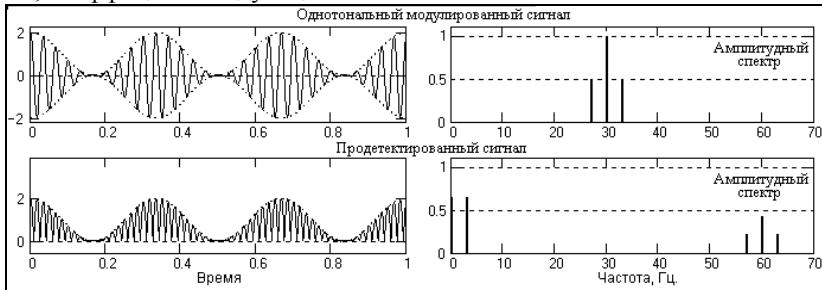


Рис. 23.28. Изменение однотонального модулированного сигнала при детектировании

Как видно на рисунке, при детектировании спектр модулированного сигнала становится однополярным, переходит на основную несущую частоту  $2\omega_0$  и уменьшается по энергии. Основная часть энергии (более  $4/5$ ) трансформируется в область низких частот и распределяется между постоянной составляющей и выделенной гармоникой сигнала модуляции в зависимости от значения коэффициента модуляции  $M$ . При  $M=1$  энергии равны, при  $M=0$  (в отсутствие сигнала модуляции) вся энергия переходит на постоянную составляющую.

Кроме этих составляющих в спектре появляются также 2-я, 3-я и более высокие гармоники детектированного модулированного сигнала (т.е. на частотах  $4\omega_0 \pm \Omega$ ,  $6\omega_0 \pm \Omega$ , и т.д.), которые не показаны на рисунке. Энергия второй гармоники не превышает 2%, а остальных и вовсе незначительна. Демодуляторы сигнала выделяют после детектирования только низкочастотный информационный сигнал и подавляют все остальные частоты, включая постоянную составляющую (низкочастотный фильтр с подавлением постоянной составляющей).

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

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

$$y(t) = u(t) \cos \omega_0 t = U(t) \cos \omega_0 t \cdot \cos \omega_0 t = \frac{1}{2} U(t) + \frac{1}{2} U(t) \cos 2\omega_0 t. \quad (23.103)$$

Как следует из этого выражения, сигнал разделяется на два слагаемых, первое из которых повторяет исходный модулирующий сигнал, а второе повторяет модулированный сигнал на удвоенной несущей частоте  $2\omega_0$ .

На рис. 23.29 приведено визуальное сопоставление двухполупериодного и синхронного детектирования, которое наглядно показывает практически полное подобие процессов.

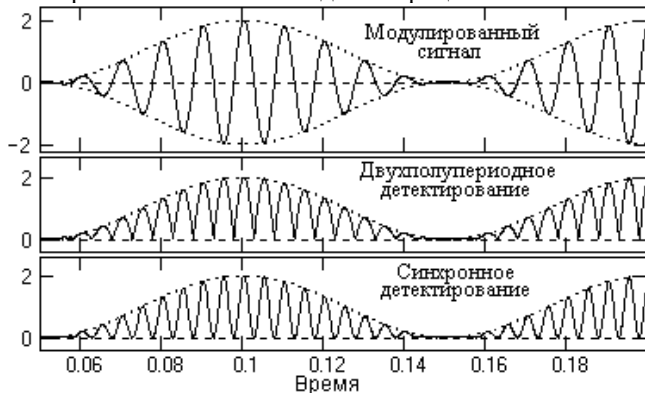


Рис. 23.29.

Но форма новой несущей при синхронном детектировании является чистой гармоникой, в отличие от двухполупериодного детектирования.

Физический амплитудный спектр сигналов после демодуляции однозначно соотносится со спектром входного модулированного сигнала: амплитуды гармоник модулированного сигнала на частоте  $2\omega_0$  в два раза меньше амплитуд входного сигнала, постоянная составляющая равна амплитуде несущей частоты  $\omega_0$  и не зависит от глубины модуляции, амплитуда информационного демодулированного сигнала в 2 раза меньше амплитуды исходного модулирующего сигнала.

Замечательной особенностью синхронного детектирования является полная независимость от глубины модуляции, т.е. коэффициент модуляции сигнала может быть больше 1. Пример синхронного детектирования перемодулированного сигнала приведен на рис. 23.30.

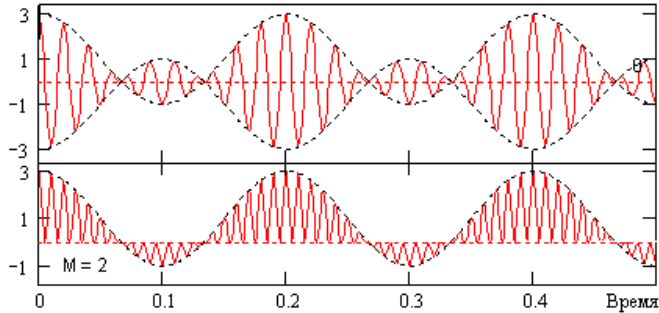


Рис. 23.30.

Однако при синхронном детектировании требуется точное совпадение фаз и частот опорного колебания демодулятора и несущей гармоники АМ-сигнала. При сдвиге фазы опорного колебания на  $\Delta\varphi$  относительно несущей частоты выходной сигнал демодулятора оказывается умноженным на косинус фазовой ошибки:

$$y(t) = U(t) \cos \omega_0 t \cdot \cos(\omega_0 t - \Delta\varphi) = \frac{1}{2} U(t) \cos(-\Delta\varphi) + \frac{1}{2} U(t) \cos(2\omega_0 t - \Delta\varphi),$$
 и амплитуда сигнала занижается, а при  $\Delta\varphi = \pi/2$  становится равной нулю.

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

$$y(t) = U(t) \cos \omega_0 t \cdot \cos(\omega_0 t - \Delta\omega) = \frac{1}{2} U(t) \cos(-\Delta\omega t) + \frac{1}{2} U(t) \cos((2\omega_0 - \Delta\omega)t),$$
 при этом выходной сигнал демодулятора начинает пульсировать с частотой биений (beat)  $\Delta\omega$ .

Для синхронизации опорного колебания с несущей частотой сигнала в составе демодуляторов используются следующие системы фазовой автоподстройки опорной частоты.

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

$$u(t) = U_m M \cdot \cos \Omega t \cdot \cos \omega_0 t = (U_m M / 2) \{ \cos[(\omega_0 + \Omega)t] + \cos[(\omega_0 - \Omega)t] \}, \quad (23.104)$$

т.е. однотональный модулирующий сигнал переносится на биения двух высоких частот. Пример сигнала с балансной модуляцией приведен на рис. 23.31. Амплитудный спектр сигнала подобен приведенному на

рис. 23.24 с отсутствующей несущей частотой  $\omega_0$ . Аналогично, многотональный балансно - модулированный сигнал имеет две симметричные относительно частоты  $\omega_0$  группы верхних и нижних боковых колебаний:

$$u(t) = (U_m/2) \left\{ \sum_{n=1}^N M_n \cos[(\omega_0 + \Omega_n)t + \Phi_n] + \sum_{n=1}^N M_n \cos[(\omega_0 + \Omega_n)t - \Phi_n] \right\}. \quad (23.105)$$

Подавление несущей частоты определяется следующим. При переходе огибающей биений  $U(t)$  через нуль фаза несущей частоты высокочастотного заполнения скачком изменяется на  $180^\circ$ , поскольку функция косинуса огибающей имеет разные знаки слева и справа от нуля. При этом в высокодобротной системе (с малыми потерями энергии), настроенной на частоту  $\omega_0$ , колебания, возбужденные одним периодом биений, гасятся последующим периодом.

Однако балансная модуляция не получила широкого распространения в связи с трудностями, возникающими при демодуляции сигналов. В принципе, синхронное детектирование позволяет выполнять демодуляцию без каких-либо проблем, но при условии известной несущей частоты сигнала и точной фазовой синхронизации опорной частоты с несущей. Но во входном сигнале АМ-ПН несущая частота отсутствует. Для снятия этой трудности обычно применяют неполное подавление несущей и оставляют в модулированном сигнале определенный "остаток" несущей (пилот-сигнал), который и используется для фазочастотной автосинхронизации при демодуляции.

**Однополосная амплитудная модуляция.** При идентичности информации в группах верхних и нижних боковых частот нет необходимости в их одновременной передаче. Одна из них перед подачей сигнала в канал связи может быть удалена, чем достигается двукратное сокращение полосы занимаемых сигналом частот. Уравнение сигнала с одной боковой полосой (ОБП – сигнал, single side band - SSB) может быть получено непосредственно из 23.98. Для верхней (знаки '+' во втором слагаемом) или нижней (знаки '-') боковой полосы:

$$u(t) = U_m \cos(\omega_0 t + \varphi_0) + (U_m/2) \sum_{n=1}^N M_n \cos[(\omega_0 \pm \Omega_n)t + \varphi_0 \pm \Phi_n]. \quad (23.106)$$

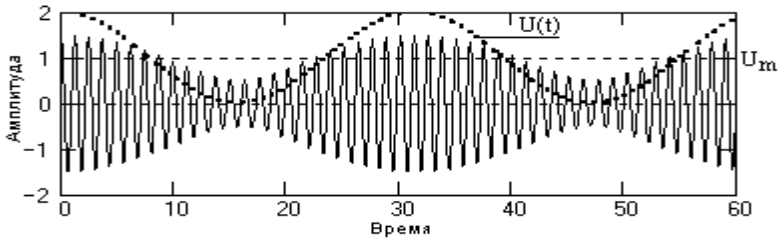


Рис. 23.32. Однополосная амплитудная модуляция.

Внешняя форма сигнала ОБП (пример на рис. 23.32 при однотоновой модуляции) сходна с обычным АМ – сигналом, но ее огибающая, как это можно заметить, отличается от огибающей  $U(t)$ , заданной при модуляции при  $M = 1$  (показана пунктиром).

Для демодуляции ОБП – сигнала может использоваться как двухполупериодное, так и синхронное детектирование, со всеми особенностями, присущими этим методам. Результаты демодуляции отличаются от демодуляции АМ – сигналов только в 2 раза меньшей амплитудой выходных сигналов.

При однополосной модуляции также возможно подавление несущей частоты (полное или частичное), что позволяет полнее использовать мощность передатчика.

**Полярная модуляция** решает конкретную техническую задачу – передачу двух сигналов одновременно, что требуется, например, в стереовещании или при передаче стереоснимков. Рассмотрим это на примере стереосигналов.

В системе стереовещания необходимо передавать два сигнала  $s_1(t)$  и  $s_2(t)$  одновременно (левый и правый каналы) при условии совмещения с монофоническими приемниками. Для выполнения этого условия создается специальный модулирующий сигнал. Процесс создания сигнала поясняется на рис. 23.33, где в качестве канальных сигналов приняты моногармоники  $s_1$  и  $s_2$ .

Специальный модулирующий сигнал формируется из двух сигналов - монофонического и разностного. Монофонический сигнал образуется суммой сигналов в каналах, разностный – разностью сигналов:

$$s_{\text{mono}}(t) = s_1(t) + s_2(t),$$

$$s_{\text{diff}}(t) = s_1(t) - s_2(t),$$

что позволяет восстанавливать исходные сигналы каналов:

$$s_1(t) = (s_{\text{mono}}(t) + s_{\text{diff}}(t)) / 2,$$

$$s_2(t) = (s_{\text{mono}}(t) - s_{\text{diff}}(t)) / 2.$$

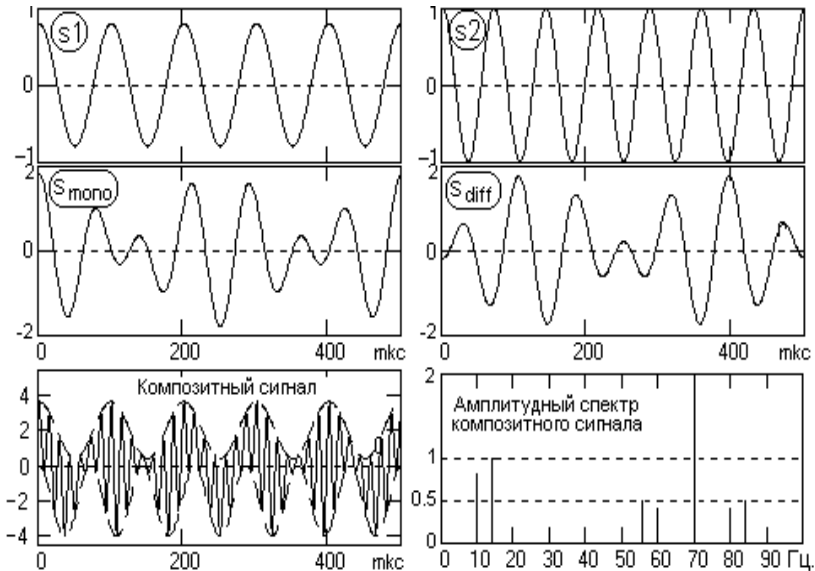


Рис. 23.33. Полярная модуляция.

Монофонический сигнал является основным и не изменяется по частоте, что позволяет принимать его монофоническим приемникам. Для одновременной передачи разностного сигнала монофонический сигнал суммируется с *поднесущей* частотой  $\omega_{sc}$  (subcarrier), которая располагается за звуковым диапазоном частот монофонических приемников (в области ультразвука), и модулируется разностным сигналом (с установкой коэффициента модуляции значением смещения  $A_0$ ):

$$s(t) = s_{mono}(t) + (A_0 + s_{diff}(t)) \cdot \cos \omega_{sc} t.$$

Полученный сигнал называют *композитным стереосигналом*. Именно он используется в качестве модулирующего сигнала для любого метода модуляции, в том числе и для угловой модуляции, которая будет рассматриваться ниже. Как видно на рис. 23.33, верхняя и нижняя огибающие композитного сигнала с точностью до постоянной составляющей соответствуют первому и второму сигналу стереоканалов, что позволяет достаточно просто выделять эти сигналы на приемной стороне. На практике поднесущую частоту композитного сигнала обычно частично или целиком подавляют. Подавление поднесущей выполняется изменением значения смещения  $A_0 \rightarrow 0$ , при этом разностный сигнал переходит в режим перемодуляции, а динамический диапазон амплитуд композитного сигнала уменьшается в два раза.

### 23.15. Сигналы с угловой модуляцией

При угловой модуляции (angle modulation) в несущем гармоническом колебании  $u(t) = U_m \cos(\omega t + \varphi)$  значение амплитуды колебаний  $U_m$  остается постоянным, а информация  $s(t)$  переносится либо на частоту  $\omega$ , либо на фазовый угол  $\varphi$ . И в том, и в другом случае текущее значение фазового угла гармонического колебания  $u(t)$  определяет аргумент  $\psi(t) = \omega t + \varphi$ , который называют *полной фазой* колебания.

**Фазовая модуляция** (ФМ, phase modulation - PM). При фазовой модуляции значение фазового угла  $\varphi(t)$  несущей частоты колебаний  $\omega_0$  пропорционально амплитуде модулирующего сигнала  $s(t)$ . Уравнение ФМ – сигнала:

$$u(t) = U_m \cos[\omega_0 t + \varphi(t)], \quad \varphi(t) = \beta s(t). \quad (23.107)$$

Коэффициент пропорциональности  $\beta$  называется индексом фазовой модуляции. Полная фаза колебаний несущей в текущие моменты времени соответственно определяется выражением:

$$\psi(t) = \omega_0 t + \beta s(t).$$

Пример однотонального ФМ – сигнала приведен на рис. 23.34.

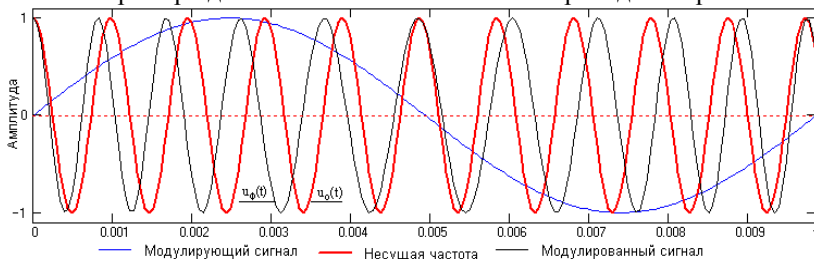


Рис. 23.34. Фазомодулированный сигнал.

При  $s(t) = 0$ , ФМ – сигнал является простым гармоническим колебанием и показан функцией  $u_0(t)$ . С увеличением значений  $s(t)$  полная фаза колебаний  $\psi(t)$  нарастает быстрее и опережает линейное нарастание  $\omega_0 t$ . Соответственно, при уменьшении значений  $s(t)$  скорость роста полной фазы во времени спадает. В моменты экстремальных значений  $s(t)$  абсолютное значение фазового сдвига  $\Delta\psi$  между ФМ – сигналом и значением  $\omega_0 t$  немодулированного колебания также является максимальным и носит название *девиации фазы*.

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

$$\omega(t) = \psi(t)/dt = \omega_0 + \beta ds(t)/dt.$$



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

$$\psi(t) = \int_0^t \omega(t) dt + \varphi_0,$$

где  $\varphi_0 = \text{const}$  – произвольная постоянная интегрирования.

**Частотная модуляция** (ЧМ, frequency modulation - FM) характеризуется линейной связью модулирующего сигнала с мгновенной частотой колебаний, при которой мгновенная частота колебаний образуется сложением частоты высокочастотного несущего колебания  $\omega_0$  со значением амплитуды модулирующего сигнала с определенным коэффициентом пропорциональности  $\Delta\omega$  - девиацией частоты:

$$\omega(t) = \omega_0 + \Delta\omega \cdot s(t). \quad (23.108)$$

Соответственно, полная фаза колебаний:

$$\psi(t) = \omega_0 t + \Delta\omega \int_0^t s(t) dt + \varphi_0,$$

Уравнение ЧМ – сигнала:

$$u(t) = U_m \cos(\omega_0 t + \Delta\omega \int_0^t s(t) dt + \varphi_0). \quad (23.109)$$

Частотная и фазовая модуляция взаимосвязаны. Если изменяется начальная фаза колебания, изменяется и мгновенная частота, и наоборот. По этой причине их и объединяют под общим названием **угловой модуляции** (УМ). По форме колебаний с угловой модуляцией невозможно определить, к какому виду модуляции относится данное колебание, к ФМ или ЧМ, а при достаточно гладких функциях  $s(t)$  формы сигналов ФМ и ЧМ вообще практически не отличаются.

**Однотональная угловая модуляция.** Рассмотрим гармонический модулирующий сигнал с постоянной частотой колебаний  $\Omega$ . Начальная фаза ФМ колебаний:

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

где  $\beta$  - индекс угловой модуляции (modulation index), которым задается интенсивность колебаний начальной фазы. Полная фаза модулированного сигнала с учетом несущей частоты  $\omega_0$ :

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

Уравнение модулированного сигнала:

$$u(t) = U_m \cos(\omega_0 t + \beta \sin \Omega t). \quad (23.110)$$

Мгновенная частота колебаний:

$$\omega(t) = d\psi(t)/dt = \omega_0 + \beta\Omega \cos \Omega t.$$

Как следует из этих формул, и начальная фаза, и мгновенная

частота изменяется по гармоническому закону. Максимальное отклонение от среднего значения  $\omega_0$  характеризует девиацию частоты (frequency deviation) при ФМ модуляции и равно  $\omega_d = \beta\Omega = \Delta\omega$ . Отсюда, индекс угловой модуляции равен отношению девиации частоты к частоте модулирующего сигнала:

$$\beta = \omega_d/\Omega. \quad (23.111)$$

Для ЧМ колебаний начальная фаза сигнала определяется выражением:

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

а мгновенная частота колебаний выражением:

$$\omega(t) = \omega_0 + \beta\Omega \cos \Omega t.$$

Соответственно, полная фаза и уравнение модулированного сигнала:

$$\psi(t) = d\omega(t)/dt = \omega_0 t + \beta \cos \Omega t,$$

$$u(t) = U_m \cos (\psi(t)).$$

Различия между частотной и фазовой модуляцией проявляются при изменении частоты  $\Omega$  модулирующего сигнала.

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

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

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

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

### **Спектры сигналов с угловой модуляцией.**

Формулу (23.110) однотоновой модуляции можно преобразовать к виду:

$$u(t) = U_m \cos(\beta \cdot \sin(\Omega t)) \cos(\omega_0 t) - U_m \sin(\beta \cdot \sin(\Omega t)) \sin(\omega_0 t). \quad (23.112)$$

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

$$\cos(\beta \cdot \sin \Omega t) \approx 1, \quad \sin(\beta \cdot \sin \Omega t) \approx \beta \cdot \sin \omega_0 t.$$

При их использовании в (23.112), получаем:

$$u(t) \approx U_m \cos \omega_0 t + (\beta U_m/2) \cos[(\omega_0 + \Omega)t] + (-\beta U_m/2) \cos[(\omega_0 - \Omega)t]. \quad (23.113)$$

Сравнение данного выражения с формулой АМ – сигнала (23.91) позволяет сделать вывод, что амплитудные спектры однотоновых ФМ и ЧМ сигналов при  $\beta \ll 1$  практически аналогичны АМ сигналам и также содержат верхнюю и нижнюю боковые частоты  $\omega_0 + \Omega$  и  $\omega_0 - \Omega$ . Различие заключается только в смене знака амплитуды нижней боковой частоты на минус, т.е. в дополнительном фазовом сдвиге нижней боковой частоты на  $180^\circ$  относительно верхней боковой

частоты. Соответственно, гармонические АМ сигналы могут быть трансформированы в ЧМ сигналы изменением на  $180^\circ$  начальной фазы одной из боковых полос. Заметим также, что при малых значениях индекса  $\beta$  основная мощность сигнала (как и в АМ) приходится на несущую частоту.

Математическая модель однотоновых ЧМ и ФМ сигналов с любым значением индекса модуляции  $\beta$  в общем случае получается разложением функции (23.110) в следующий ряд:

$$u(t) = U_m \sum_{k=-\infty}^{\infty} J_k(\beta) \cos[(\omega_0 + k\Omega)t],$$

где  $J_k(\beta)$  – функция Бесселя  $k$ -го индекса от аргумента  $\beta$ . Из этого уравнения следует, что спектр сигнала содержит бесконечное число составляющих - нижних и верхних боковых колебаний, с частотами  $\omega_0 \pm k\Omega$ , которые соответствуют гармоникам частоты модуляции, и с амплитудами, пропорциональными значениям  $J_k(\beta)$ . Амплитуды пяти первых гармоник и несущей частоты при  $U_m=1$  в зависимости от индекса модуляции приведены на рис. 23.35.

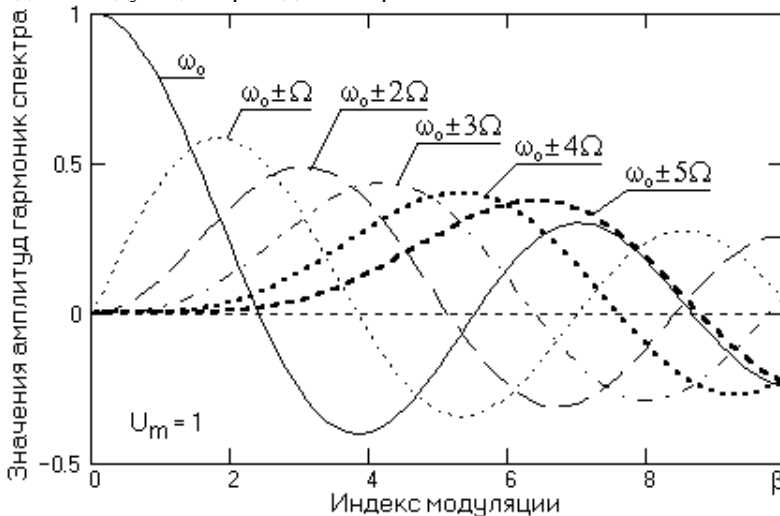
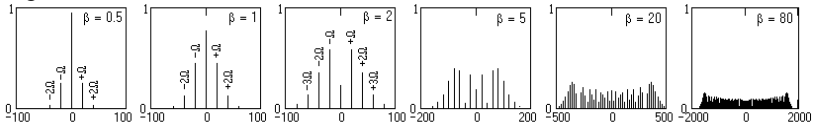


Рис. 23.35. Амплитуды гармоник сигналов с угловой модуляцией.

При малой величине индекса  $\beta$  значимые амплитудные значения имеют только первые гармоники. С ростом величины  $\beta$  количество значимых боковых составляющих увеличивается, а энергия сигнала перераспределяется на боковые составляющие. Функции Бесселя имеют колебательный характер, поэтому спектр при удалении

от несущей частоты  $\omega_0$  спадает немонотонно. На рис. 23.35 можно также видеть, что при определенных значениях индекса модуляции (2.405, 5.52, 8.654 и т.д.) несущая частота  $\omega_0$  в спектре сигнала полностью отсутствует. Форма амплитудный спектров модулированных сигналов при разных индексах модуляции приведена на рис. 23.36.



15.2.3. Модули спектров ЧМ сигнала при разных индексах модуляции (несущая частота 2500 Гц, гармоника модуляции 25 Гц, шкала частот в Гц относительно несущей)

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

$$P_{\text{практ}} = 2(\beta+1)\Omega, \quad (23.114)$$

т.е. спектральными составляющими с номерами  $k > (\beta+1)$  пренебрегают. Формирование реальных сигналов, как правило, выполняется при  $\beta \gg 1$ , при этом эффективная ширина спектра равна удвоенной девиации частоты:

$$P_{\text{практ}} \approx 2\beta\Omega = 2\omega_d. \quad (23.115)$$

Отсюда следует, что по сравнению с АМ – сигналами, полоса частот которых равна  $2\Omega$ , для передачи сигналов с угловой модуляцией требуется полоса частот, в  $\beta$  раз большая. С другой стороны, именно широкополосность ЧМ и ФМ сигналов обеспечивает их большую помехоустойчивость по сравнению с АМ сигналами.

**Сигналы с многотональной угловой модуляцией**

отличаются еще большей сложностью спектрального состава. В их спектре присутствуют не только боковые частоты с гармониками частот модулирующего сигнала, но и боковые комбинационные частоты типа  $\omega_0 \pm \Omega_1 \pm \Omega_2 \pm \dots \Omega_i$ , со всеми возможными комбинациями частот модулирующего сигнала  $\Omega_i$ . При непрерывном спектре модулирующего сигнала спектры ЧМ и ФМ сигналов также становятся непрерывными.

**Демодуляция УМ – сигналов** много сложнее демодуляции сигналов АМ.

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

$$u_a(t) = u(t) + j u_h(t),$$

где  $u_h(t)$  – аналитически сопряженный сигнал или квадратурное дополнение сигнала  $u(t)$ , которое вычисляется сверткой сигнала  $u(t)$  с оператором Гильберта ( $1/\pi$ ):

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

Полная фаза колебаний представляет собой аргумент аналитического сигнала:

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

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

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

При частотной модуляции фазовая функция дифференцируется с вычитанием из результата значения частоты  $\omega_0$ :

$$\varphi(t) = d\psi(t)/dt - \omega_0.$$

В принципе, данный метод может применяться и в реальном масштабе времени, но с определенной степенью приближения, поскольку оператор Гильберта слабо затухает.

При демодуляции в реальном масштабе времени используется квадратурная обработка, при которой входной сигнал умножается на два опорных колебания со сдвигом фазы между колебаниями в  $90^\circ$ :

$$u_1(t) = u(t) \cos \omega_0 t = U_m \cos(\omega_0 t + \varphi(t)) \cos \omega_0 t = \frac{1}{2} U_m \cos \varphi(t) + \frac{1}{2} \cos(2\omega_0 t + \varphi(t)),$$

$$u_2(t) = u(t) \sin \omega_0 t = U_m \cos(\omega_0 t + \varphi(t)) \sin \omega_0 t = -\frac{1}{2} U_m \sin \varphi(t) + \frac{1}{2} \sin(2\omega_0 t + \varphi(t)).$$

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

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

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

**Квадратурная модуляция** позволяет модулировать несущую частоту одновременно двумя сигналами путем модуляции амплитуды несущей одним сигналом, и фазы несущей другим сигналом. Уравнение результирующих колебаний амплитудно-фазовой модуляции:

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

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

колебаний.

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

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

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

Полученный сигнал называют квадратурным (quadrature), а способ модуляции - *квадратурной* модуляцией (КАМ).

Спектр квадратурного сигнала может быть получен непосредственно по уравнению балансной модуляции (23.104) для суммы двух сигналов:

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

Демодуляция квадратурного сигнала соответственно выполняется умножением на два опорных колебания, сдвинутых относительно друг друга на  $90^\circ$ :

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

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

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

### **Пример моделирования квадратурной модуляции в системе Mathcad.**

Моделирование выполняется в дискретной форме.

$N := 2999$     $n := 0 .. N$     $\Delta t := 0.001$    'Интервал и шаг дискретизации (в сек).

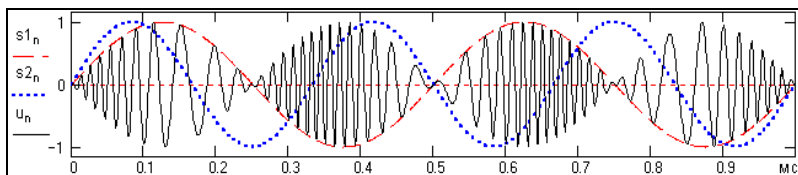
$f_0 := 50$     $f_1 := 2$     $f_2 := 3$    'Частоты в Гц несущей, первого и второго сигналов.

$s_{1n} := \sin(2 \cdot \pi \cdot f_1 \cdot n \cdot \Delta t)$    'Первый модулирующий сигнал (моногармоника с амплитудой 1).

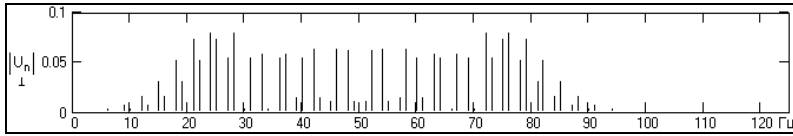
$s_{2n} := \sin(2 \cdot \pi \cdot f_2 \cdot n \cdot \Delta t)$    'Второй модулирующий сигнал (моногармоника с амплитудой 1).

$\beta := 10$     $\varphi_n := \beta \cdot s_{2n}$    'Перенос информации  $s_{2n}$  на фазу

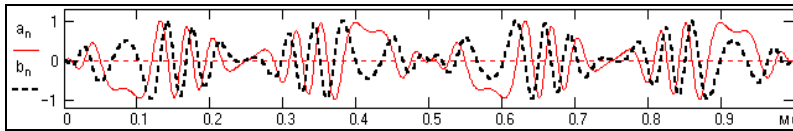
$u_n := s_{1n} \cdot \cos(2 \cdot \pi \cdot f_0 \cdot n \cdot \Delta t + \varphi_n)$    'Амплитудно-фазовая модуляция



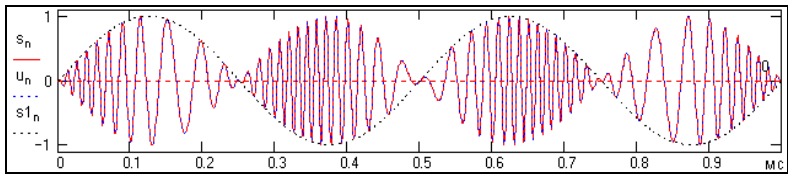
$U := \text{CFFT}(u) \quad \Delta f := 1/[(N+1)\cdot\Delta t] \quad \text{'БПФ и шаг по частоте}$



$a_n := s1_n \cdot \cos(\varphi_n) \quad b_n := s1_n \cdot \sin(\varphi_n) \quad \text{'Формирование модулирующих сигналов}$

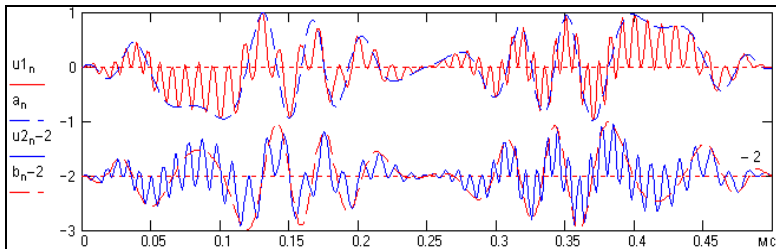


$s_n := a_n \cdot \cos(2 \cdot \pi \cdot f_0 \cdot n \cdot \Delta t) + b_n \cdot \sin(2 \cdot \pi \cdot f_0 \cdot n \cdot \Delta t) \quad \text{'Квадратурный сигнал.}$   
 Сравнением с сигналом  $u_n$  нетрудно убедиться в их идентичности,  $a_n$ , следовательно, идентичны и их спектры.



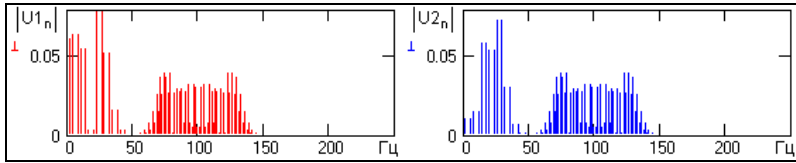
**Демодуляция квадратурного сигнала.**

$u1_n := s_n \cdot \cos(2 \cdot \pi \cdot f_0 \cdot n \cdot \Delta t) \quad \text{'Раздельная синхронная демодуляция}$   
 сигналов  $a_n$  и  $b_n$ . Графики  $u2_n := s_n \cdot \sin(2 \cdot \pi \cdot f_0 \cdot n \cdot \Delta t) \quad \text{'сигналов } u2_n \text{ и}$   
 $b_n$  смещены на -2 для представления в одном поле.

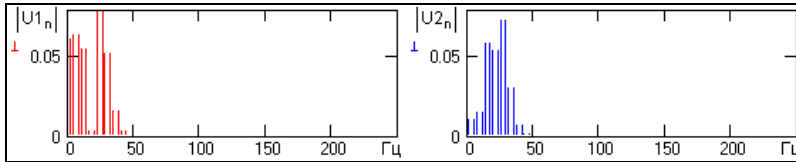


$U1 := \text{CFFT}(u1) \quad U2 := \text{CFFT}(u2) \quad \text{'Спектры сигналов, БПФ.}$

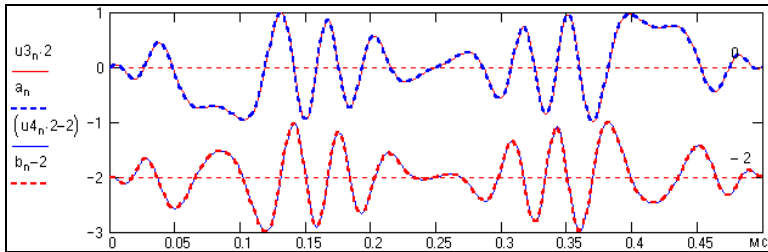
---



$M := 50/\Delta f$     $m := M .. N+1-M$     $U1_m := 0$     $U2_m := 0$    'Удаление высоких частот (после 50 Гц).



$u3 := \text{ICFFT}(U1)$     $u4 := \text{ICFFT}(U2)$    'ОБПФ оставшихся низких частот спектра. На графиках 'амплитуды сигналов  $u3_n$  и  $u4_n$  увеличены в 2 раза 'для сопоставления с исходными сигналами  $a_n$  и  $b_n$ .



### 23.16. Внутримпульсная частотная модуляция

Сигнал с внутримпульсной частотной модуляцией – это радиоимпульс, высокочастотное заполнение которого имеет переменную частоту.

**ЛЧМ – сигналы.** Если закон изменения мгновенной частоты заполнения имеет линейный характер, то такие сигналы носят название ЛЧМ – сигналов (линейная частотная модуляция). Наиболее широкое применение они получили в радиолокации. Пример ЛЧМ – сигнала с огибающей прямоугольной формы приведен на рис. 23.37.



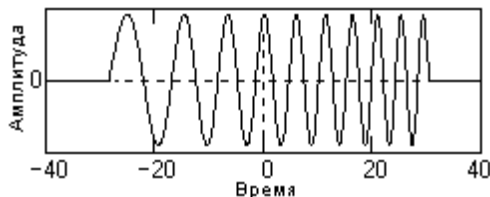


Рис. 23.37. ЛЧМ – сигнал.

ЛЧМ – сигналы имеют одно замечательное свойство. Если сигнал подать на частотно-зависимую линию задержки, время задержки сигнала которой велико на малых частотах (в начальной части ЛЧМ – сигнала) и уменьшается по мере нарастания частоты в ЛЧМ – сигнале, то на выходе такой линии происходит "сжатие" сигнала в один период высокочастотного колебания путем суммирования амплитудных значений всех периодов сигнала. При этом происходит увеличение амплитуды выходного сигнала и уменьшение статистических шумов, так как суммируемые одновременно по этим же периодам шумы не коррелированы.

Для модели радиоимпульса с прямоугольной огибающей примем его длительность равной  $\tau_{и}$ , а точку  $t = 0$  поместим в центр радиоимпульса. Допустим также, что частота заполнения линейно нарастает от начала импульса к его концу со скоростью  $\mu$  ( $c^{-2}$ ), при этом:

$$\omega(\tau) = \omega_0 + \mu\tau. \quad (23.116)$$

Девияция частоты за время длительности импульса и полная фаза сигнала:

$$\Delta\omega = \mu \cdot \tau_{и}. \quad (23.117)$$

$$\psi(t) = \omega_0 t + \mu t^2/2. \quad (23.118)$$

Уравнение ЛЧМ – сигнала:

$$u(t) = \begin{cases} 0, & t < -\tau_{и}/2 \\ U_m \cos(\omega_0 t + \mu t^2/2), & -\tau_{и}/2 \leq t \leq \tau_{и}/2 \\ 0, & t > \tau_{и}/2 \end{cases} \quad (23.119)$$

**Спектр прямоугольного ЛЧМ – сигнала** вычисляется через преобразование Фурье. Девияция частоты за время длительности импульса по сравнению с несущей частотой обычно мала ( $\Delta\omega \ll \omega_0$ ) и форма спектра зависит от так называемой базы импульса:

$$B = \Delta\omega \cdot \tau_{и} = \mu \cdot \tau_{и}^2. \quad (23.120)$$

На рис. 23.38 приведен пример формы спектральной плотности ЛЧМ – сигнала при малом значении базы в области несущей частоты

сигнала.

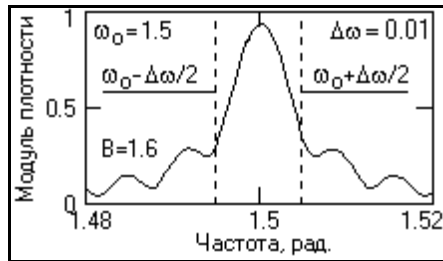


Рис.23.38. Спектр ЛЧМ- сигнала.

На практике значение базы сигналов обычно много больше 1. Увеличение базы сопровождается расширением полосы спектра  $\Delta\omega$ , при этом в пределах этой полосы модуль спектральной плотности практически постоянен и равен  $U_m \cdot \sqrt{\pi/2\mu}$ . Пример спектра приведен на рис. 23.39.

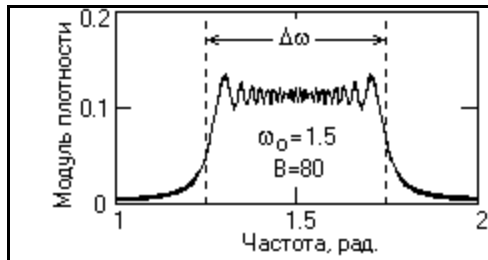


Рис. 23.39. Спектр при  $B \gg 1$ .

### 23.17. Импульсно-модулированные сигналы

В импульсной модуляции в качестве носителя модулированных сигналов используются последовательности импульсов, как правило – прямоугольных. В беспроводных системах передачи данных (в радиосвязи) эти последовательности заполняются высокочастотными колебаниями, создавая тем самым двойную модуляцию. Как правило, эти виды модуляции применяются при передаче дискретизированных данных. Для прямоугольных импульсов наиболее широко используются амплитудно-импульсная (АИМ) и широтно-импульсная (ШИМ) модуляция.

**Амплитудно-импульсная модуляция** (АИМ)

заключается в изменении приращения амплитуды импульсов пропорционально функции управляющего сигнала при постоянной длительности импульсов и периоде их следования:

$$U(t) = U_o + k \cdot s(t), \quad \tau_n = \text{const}, \quad T = \text{const}. \quad (23.121)$$

Спектр АИМ рассмотрим на примере модулирования однотонального сигнала  $s(t)$ , приведенного на рис. 23.40.

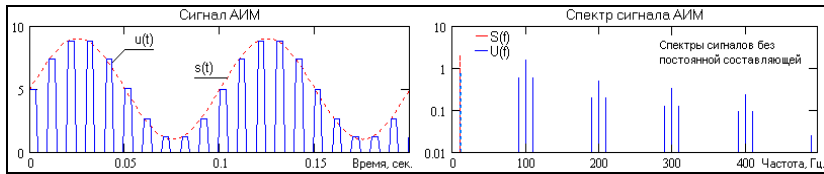


Рис. 23.40.

Напишем уравнение модулированного сигнала в следующей форме:

$$u(t) = (1+M \cos \Omega t) \cdot f(t), \quad (23.122)$$

где  $f(t)$  – периодическая последовательность прямоугольных импульсов с частотой  $\omega_o$ , которую можно аппроксимировать рядом Фурье (без учета фазы):

$$f(t) = U_o + \sum_{n=1}^{\infty} U_n \cos n\omega_o t. \quad (23.123)$$

Подставляя (23.123) в (23.122), получаем:

$$u(t) = (1+M \cos \Omega t)U_o + \sum_{n=1}^{\infty} U_n \cos n\omega_o t \cdot (1+M \cos \Omega t)$$

$$u(t) = U_o + U_o M \cos \Omega t + \sum_{n=1}^{\infty} U_n \cos n\omega_o t +$$

$$+ 0.5M \sum_{n=1}^{\infty} U_n \cos (n\omega_o + \Omega)t + 0.5M \sum_{n=1}^{\infty} U_n \cos (n\omega_o - \Omega)t. \quad (23.124)$$

Форма спектра, в начальной части спектрального диапазона, приведена на рис. 23.40. В целом, спектр бесконечен, что определяется бесконечностью спектра прямоугольных импульсов. Около каждой гармоники  $n\omega_o$  спектра прямоугольных импульсов появляются боковые составляющие  $n\omega_o \pm \Omega$ , соответствующие спектру моделирующей функции (при многотональном сигнале – боковые полосы спектров). При дополнительном высокочастотном заполнении импульсов весь спектр смещается в область высоких частот на частоту заполнения.

**Широтно-импульсная модуляция** (ШИМ, в английской терминологии pulse width modulation, PWM), которую иногда называют *модуляцией по длительности импульсов (ДИМ)*, заключается в управлении длительностью импульсов пропорционально функции управляющего сигнала при постоянной амплитуде импульсов и периоде следования по фронту импульсов:

$$\tau(t) = t_0 + k \cdot s(t), \quad U = \text{const}, \quad T = \text{const}. \quad (23.125)$$

Рассмотрим выполнение ШИМ в простейшем варианте на примере гармонического колебания, приведенного на рис. 23.41.

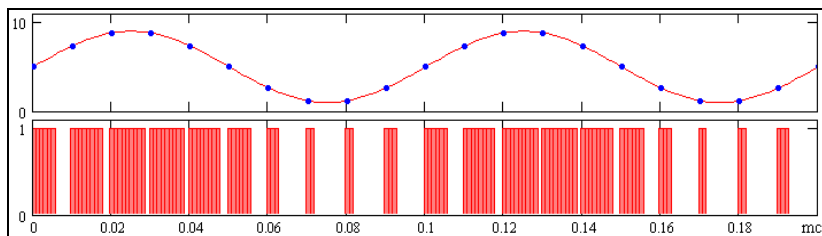


Рис. 23.41. Широтно-импульсная модуляция.

Передаваемая кривая дискретизируется, при этом имеет значение, как интервал дискретизации, так и количество уровней квантования. При передаче данных прямоугольные импульсы начинаются в моменты дискретных отсчетов данных, а длительность импульсов устанавливается пропорциональной значению отсчетов, при этом максимальная длительность импульсов не должна превышать интервала дискретизации данных. Пример сформированных импульсов приведен на рис. 23.41 непосредственно под дискретизированной гармоникой, при этом число уровней квантования гармоники принято равным 8.

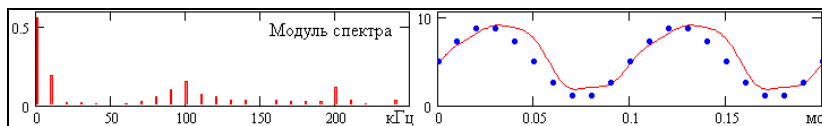


Рис. 23.42. Спектр ШИМ – сигнала.

Рис. 23.43. Восстановленный сигнал.

На рис. 23.42 приведен спектр сформированного сигнала ШИМ. В начальной части спектра он содержит постоянную составляющую среднего уровня сигнала и пик частоты гармоники, закодированной в ШИМ – сигнале. Если выделить из спектра эти две составляющие, то восстанавливается исходный сигнал с погрешностью квантования, приведенный на рис. 23.43. Естественно, что при малом числе уровней квантования погрешность восстановления исходного

гармонического сигнала очень велика.

Попутно заметим, что широтно-импульсная модуляция с последующим выделением постоянной составляющей может весьма эффективно использоваться (и используется) для слежения за средним уровнем сигнала и автоматического регулирования его динамического диапазона, как, например, в системах установки громкости звука и яркости цветов и изображения в целом в современных телевизионных установках.

**Временная импульсная модуляция** (ВИМ) представляет собой девиацию импульсов по временной оси по закону модулирующего сигнала, и по существу аналогична угловой модуляции гармонической несущей. Она также может быть фазовой (ФИМ) или частотной (ЧИМ).

**Кодово-импульсная модуляция** заключается в том, что в точках дискретизации модулирующего сигнала производится квантование его значений и кодирование квантованных значений, как правило, в двоичной системе исчисления. Кодированные значения затем передаются при помощи соответствующей кодовой последовательности стандартных символов.

## **23.18. Модуляция символьных и кодовых данных**

В настоящее время информация передается по каналам связи в основном в цифровой форме. Числа при передаче с периодом  $T$  поступают от источника информации и называются *символами* (symbol), а частота передачи символов – *символьной скоростью* (symbol rate)  $f_T=1/T$ .

Символьные последовательности являются дискретными квантованными сигналами, которые формируются следующим образом. Весь диапазон сигнала  $s(t)$  делится на  $Q_s$  разрешенных уровней с некоторым шагом  $\Delta q$ . Сигнал  $s(t)$  дискретизируется с равномерным шагом, а мгновенные значения отсчетов сигнала округляются до ближайшего разрешенного уровня  $Q_s(t)$ . Полученный сигнал называется квантованным АИМ (КАИМ). Значения сигнала  $Q_s(t)$  отличаются от  $s(t)$  на так называемый шум квантования, которым определяется погрешность восстановления исходного сигнала. С увеличением числа уровней квантования шум квантования уменьшается. Наличие шума является недостатком цифровых методов передачи, однако она открывает и новые возможности передачи. В частности, зная всю шкалу разрешенных уровней на приеме, можно "очистить" сигнал от внешних помех, если их уровень меньше  $0,5\Delta q$ .

Каждому из возможных символов  $Q_s$  устанавливается определенный набор параметров несущего колебания, которые поддерживаются постоянными на интервале  $T$  до прихода следующего символа. Это означает преобразование последовательности чисел в ступенчатый сигнал (кусочно-постоянная интерполяция) который используется в качестве модулирующего сигнала. Соответственно, параметры несущего колебания, на которые переносится ступенчатый сигнал, также меняются скачкообразно. Такой способ модуляции несущей называется манипуляцией (keying), и может выполняться с использованием всех рассмотренных методов модулирования.

**Амплитудно-манипулированные сигналы** простейшего типа представляют собой последовательности радиоимпульсов, разделенные паузами. Такие сигналы используются в радиотелеграфии и в системах передачи дискретных данных. Форма огибающей радиоимпульсов в общем случае может быть произвольной, паузы могут отличаться по длительности от радиоимпульсов. На рис. 23.44. приведен пример амплитудно-манипулированного сигнала:

$$u(t) = U_m \cdot \cos 2\pi f_0 t,$$

с прямоугольной П-формой огибающей.

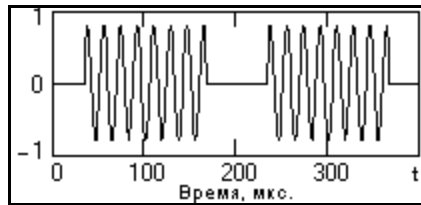


Рис. 23.44. АМП-сигнал.

Соответственно, в частотной области спектр АМП – сигнала образуется сверткой спектра огибающей функции (в данном случае – спектра прямоугольного импульса) со спектром косинусного колебания (дельта - функции на частоте  $f_0$ ). Модуль спектральной плотности сигнала приведен на рис. 23.45.

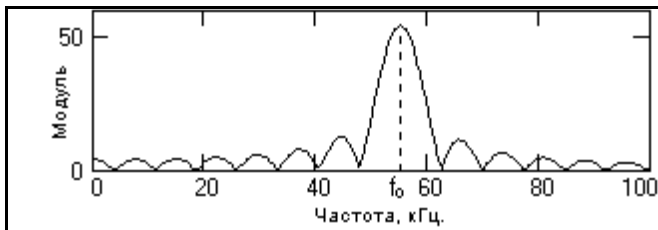


Рис. 23.45. Модуль спектра АМП-сигнала.

Спектр прямоугольного импульса довольно слабо затухает и простирается неограниченно далеко, а поэтому его использование в качестве огибающей АМП - сигнала не рекомендуется, хотя и является наиболее простым по техническому исполнению.

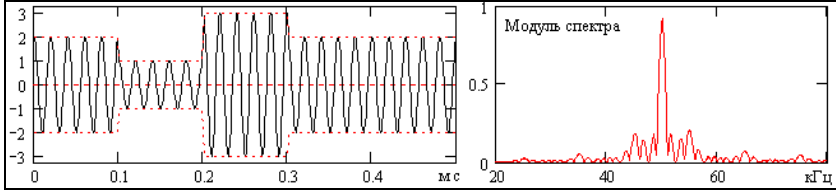


Рис. 23.46.

Рис. 23.47.

На рис. 23.46. приведен пример формы классического АМП сигнала при передаче нескольких символов, каждому из которых соответствует индивидуальная амплитуда несущей частоты при постоянной длительности интервалов посылки. Модуль спектра сигнала приведен на рис. 23.47 и тоже имеет достаточно большую ширину значимой части спектра вокруг несущей частоты.

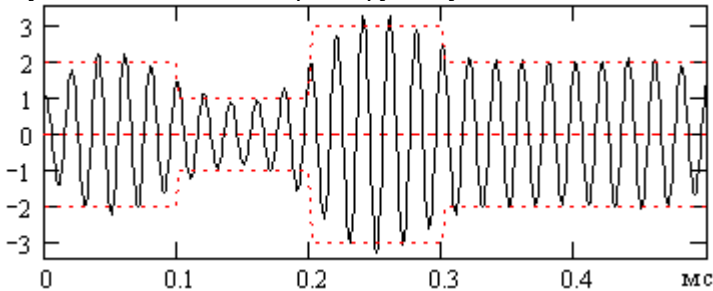


Рис. 23.48.

Естественно, что при передаче данных частотный диапазон канала передачи данных ограничивается значимой частью спектра, ширина которого устанавливается по допустимой степени искажения приемных сигналов. Степень искажения сигналов существенно зависит от длительности посылок. Пример искажения вышеприведенного сигнала при ограничении спектра интервалом 40-60 кГц приведен на рис. 23.48.

**Угловая манипуляция**, как правило, использует частотные методы модулирования, в которых каждому возможному значению передаваемого символа сопоставляется индивидуальное значение частоты гармонической несущей. При этом в точках сопряжения интервалов посылок могут происходить скачки напряжения, с соответствующим усложнением спектра модулированного сигнала.

Самый простой способ – синусоидальное начало несущей на каждом интервале с кратным количеством периодов несущей в посылке. При более сложных способах, независимых от точного сопряжения несущих частот с интервалами посылок, осуществляется управление скоростью изменения фазы несущих на границах посылок.

Демодуляция сигналов осуществляется корреляционными методами. Сущность методов – вычисление взаимной корреляции между принимаемым сигналом и набором опорных частот, используемых при модулировании, с идентификацией символов по максимумам взаимной корреляции.

Для повышения помехоустойчивости передачи данных желательно, чтобы разносимвольные посылки были некоррелированы. Если для бинарных символов 0 и 1 принять частоты посылок равными

$$s_0(t) = \cos \omega_o(t), \quad s_1(t) = \cos \omega_1(t),$$

то их ВКФ при нулевом временном сдвиге определится выражением:

$$B_{01}(0) = \int_0^T s_0(t)s_1(t)dt = \frac{1}{2} \frac{\sin(\omega_1 + \omega_o)T}{(\omega_1 + \omega_o)} + \frac{1}{2} \frac{\sin(\omega_1 - \omega_o)T}{(\omega_1 - \omega_o)}.$$

При  $(\omega_1 + \omega_o)T \gg 1$  первым слагаемым можно пренебречь, оно много меньше второго. А второе слагаемое обращается в нуль при  $(\omega_1 + \omega_o)T = \pi k$ , где  $k = 1, 2, \dots$  – целое число. Отсюда, минимальное значение между частотами манипуляции для некоррелированных посылок определяется выражениями:

$$\Delta\omega_{\min} = \pi/T, \quad \Delta f_{\min} = 1/2T = f_T/2,$$

где  $f_T$  – символьная скорость.

Фазовая манипуляция применяется значительно реже, в связи со значительными сложностями измерения абсолютных значений начальных фаз в посылках. Проще определяется относительный фазовый сдвиг в соседних посылках, поэтому обычно используется фазоразностная манипуляция.

### **23.19. Аналитические сигналы**

Аналитический сигнал – это один из способов комплексного представления сигнала, который применяется при анализе сигналов и систем их обработки. Он позволяет ввести в анализ понятия огибающей и мгновенной частоты сигнала.

**Комплексное представление вещественных сигналов.** При математическом анализе вместо вещественных сигналов с целью упрощения математического аппарата преобразования данных удобно использовать эквивалентное комплексное представление сигналов. Так, например, в теории



электрических цепей вещественная запись синусоидального напряжения

$$u(t) = U_0 \cos(\omega_0 t + \varphi)$$

заменяется комплексной формой записи:

$$\dot{u}(t) = \dot{U}_0(t) \exp(j \omega_0 t), \quad \dot{U}_0 = U_0 \exp(j \varphi),$$

при этом

$$u(t) = \text{Re } \dot{u}(t).$$

Произвольный сигнал  $s(t)$ , заданный на определенном участке временной оси (как конечном, так и бесконечном), имеет комплексную двустороннюю спектральную плотность  $S(\omega)$ . При обратном преобразовании Фурье реально и мнимой части спектра  $S(\omega)$  сигнал  $s(t)$  разделяется на четную и нечетную составляющие, которые являются двусторонними относительно  $t = 0$ , и суммирование которых полностью восстанавливает исходный сигнал. На рис. 23.49 приведен пример сигнала (А), его комплексного спектра (В) и получения четной и нечетной части сигнала из реальной и мнимой части спектра (С).

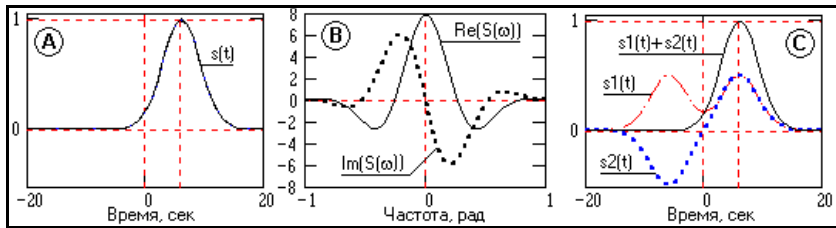


Рис. 23.49. Сигнал, спектральная плотность сигнала, четная и нечетная составляющие.

**Аналитический сигнал.** Можно выполнить обратное преобразование Фурье и в другой форме - отдельно для положительных и отрицательных частот спектра:

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^0 S(\omega) \cdot \exp(j\omega t) d\omega + \frac{1}{2\pi} \int_0^{\infty} S(\omega) \cdot \exp(j\omega t) d\omega. \quad (23.126)$$

Информация в комплексном спектре сигнала является избыточной. В силу комплексной сопряженности полную информацию о сигнале  $s(t)$  содержит как левая (отрицательные частоты), так и правая (положительные частоты) часть спектра  $S(\omega)$ . *Аналитическим сигналом*, отображающим вещественный сигнал  $s(t)$ , называют второй интеграл выражения (23.126), нормированный на  $\pi$ , т.е. обратное преобразование Фурье спектра сигнала  $s(t)$  только по положительным частотам:

$$z_s(t) = (1/\pi) \int_0^{\infty} S(\omega) \exp(j\omega t). \quad (23.127)$$

Дуальность свойств преобразования Фурье определяет, что аналитический сигнал  $z_s(t)$ , полученный из односторонней спектральной функции, является комплексным и может быть представлен в виде:

$$z_s(t) = \text{Re } z(t) + j \cdot \text{Im } z(t). \quad (23.128)$$

Аналогичное преобразование первого интеграла выражения (23.126) дает сигнал  $z_s^*(t)$ , комплексно сопряженный с сигналом  $z(t)$ :

$$z_s^*(t) = \text{Re } z(t) - j \cdot \text{Im } z(t),$$

что наглядно видно на рис. 23.50 при восстановлении сигналов по односторонним частям спектра, приведенного на рис. 23.49-В.

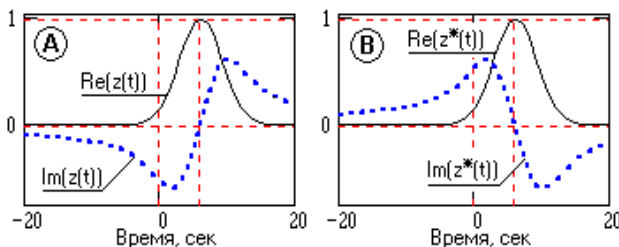


Рис. 23.50. Сигналы  $z(t)$  и  $z^*(t)$ .

По рисунку 23.50 можно видеть, что при сложении функций  $z_s(t)$  и  $z_s^*(t)$  мнимые части функций взаимно компенсируются, а вещественные части, с учетом нормировки только на  $\pi$ , а не на  $2\pi$ , как в (23.126), в сумме дают полный исходный сигнал  $s(t)$ :

$$\begin{aligned} [z_s(t) + z_s^*(t)]/2 &= \text{Re } z(t) = \frac{1}{\pi} \int_0^{\infty} S(\omega) \cdot \exp(j\omega t) d\omega = \\ &= (1/2\pi) \int_{-\infty}^{\infty} S(\omega) \cos \omega t dt = s(t). \end{aligned}$$

Отсюда, реальная часть аналитического сигнала  $z_s(t)$  равна самому сигналу  $s(t)$ .

Для выявления характера мнимой части сигнала  $z_s(t)$  выполним перевод всех членов функции (23.128) в спектральную область с отдельным представлением по положительным и отрицательным частотам (индексами  $-$  и  $+$ ) реальных и мнимых частей спектра:

$Z_s(\omega) = A_-(\omega) + A_+(\omega) + jB_-(\omega) + jB_+(\omega) + j[A'_-(\omega) + A'_+(\omega) + jB'_-(\omega) + jB'_+(\omega)]$ , где индексами  $A'$  и  $B'$  обозначены функции преобразования  $\text{Im}(z(t))$ . В

этом выражении функции в левой части спектра (по отрицательным частотам) должны взаимно компенсировать друг друга согласно определению аналитического сигнала (23.127), т.е.:

$$B'(\omega) = A_-(\omega), \quad A'(\omega) = -B_-(\omega).$$

Отсюда, с учетом четности вещественных  $A'(\omega)$  и нечетности мнимых  $B'(\omega)$  функций спектра, следуют равенства:

$$B'_+(\omega) = -A_+(\omega), \quad A'_+(\omega) = B_+(\omega).$$

Но эти четыре равенства есть не что иное, как преобразование Гильберта в частотной области спектра функции  $\text{Re } z(t) \Leftrightarrow A(\omega) + jB(\omega)$  в спектр функции  $A'(\omega) + jB'(\omega) \Leftrightarrow \text{Im } z(t)$  умножением на сигнатурную функцию  $-j \cdot \text{sgn}(\omega)$ . Следовательно, мнимая часть аналитического сигнала  $z_s(t)$  является аналитически сопряженной с его действительной частью  $\text{Re } z(t) = s(t)$  через преобразование Гильберта. Эта часть аналитического сигнала получила название *квадратурного дополнения* сигнала  $s(t)$ :

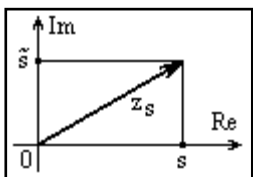
$$\text{Im } z(t) = \tilde{s}(t) = \text{TH}[s(t)] = s(t) * \text{hb}(t), \quad (23.129)$$

$$\text{hb}(t) = 1/(\pi t),$$

$$z_s(t) = s(t) + j \cdot \tilde{s}(t). \quad (23.130)$$

где индексом  $\tilde{s}(t)$  обозначен сигнал, *аналитически сопряженный* с сигналом  $s(t)$ ,  $\text{hb}(t)$  – оператор Гильберта.

Таким образом, квадратурное дополнение сигнала  $s(t)$  представляет собой свертку сигнала  $s(t)$  с оператором  $1/(\pi t)$  и может быть выполнено линейной системой с постоянными параметрами:



$$\tilde{s}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau, \quad (23.131)$$

Аналитический сигнал зависит от действительного аргумента, является однозначным и дифференцируемым. На комплексной плоскости он отображается вектором, модуль и фазовый угол которого изменяются от аргумента, а проекция сигнала на вещественную ось равна значению исходного сигнала  $s(t)$ . Какой-либо новой информации аналитический сигнал не несет, так как получен линейным преобразованием из исходного сигнала и представляет собой его новую математическую модель.

**Спектральная плотность аналитического сигнала**, если он сформирован непосредственно во временной области, определяется обычным преобразованием Фурье:

$$Z_s(\omega) = \int_{-\infty}^{\infty} z_s(t) \exp(-j\omega t) dt.$$

Эта функция, с учетом определения аналитического сигнала по

выражению (23.127), должна быть отлична от нуля только в области положительных частот, где ее значения (в силу нормировки на  $\pi$ , а не на  $2\pi$ ) должны быть равны удвоенным значениям спектральной плотности сигнала  $s(t)$ :

$$Z_s(\omega) = \begin{cases} 0, & \omega < 0, \\ 2 \cdot S(\omega), & \omega > 0. \end{cases} \quad (23.132)$$

С другой стороны, при непосредственном преобразовании Фурье левой и правой части формулы (23.130) аналитического сигнала  $z_s(t)$ , получаем:

$$Z_s(\omega) = S(\omega) + j \tilde{S}(\omega). \quad (23.133)$$

Данное выражение действительно для всей частотной оси (от  $-\infty$  до  $+\infty$ ) и должно быть равно выражению (23.132). А это означает, что левая часть спектра (23.133) (отрицательные частоты  $\omega$ ) должна быть обращена в ноль, аналогично формированию каузальной функции из ее четной и нечетной части. Это может быть выполнено следующим образом.

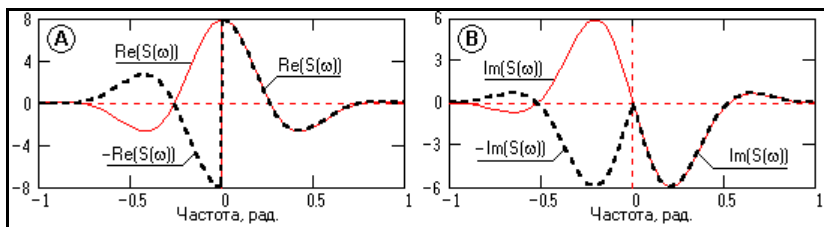


Рис. 23.51.

Если левые части спектра сигнала  $S(\omega)$  умножить на  $-1$ , обнулить реальную часть на частоте  $\omega=0$  и оставить без изменения правые части спектра, то будут получены функции, показанные пунктиром на рис. 23.51), которые дают нули в левой части спектра при сложении с исходной функцией  $S(\omega)$  и увеличивают в 2 раза правые части спектра. Такая операция может быть выполнена умножением спектра  $S(\omega)$  на сигнатурную функцию  $\text{sgn}(\omega)$ :

$$\text{sgn}(\omega) = \begin{cases} -1, & \omega < 0, \\ 0, & \omega = 0, \\ 1, & \omega > 0. \end{cases} \quad (23.134)$$

Однако при этом реальная часть новой функции  $\text{sgn}(\omega) \cdot S(\omega)$ , как это можно видеть на рис. 23.51, становится нечетной, а мнимая часть четной, что не соответствует статусу спектральных функций. Для

восстановления статуса полученный результат нужно дополнительно умножить на  $-j$ . Применяя для левой и правой части частотных аргументов индексирование соответственно  $\omega_l$  и  $\omega_r$ , можно записать подробные выражения для спектров:

$$S(\omega) = \text{Re } S(\omega_l) + j \cdot \text{Im}(\omega_l) + \text{Re } S(\omega_r) + j \cdot \text{Im}(\omega_r),$$

$$\tilde{S}(\omega) = j \cdot \text{Re } S(\omega_l) - \text{Im}(\omega_l) - j \cdot \text{Re } S(\omega_r) + \text{Im}(\omega_r).$$

При умножении квадратурной функции  $\tilde{S}(\omega)$  на  $j$  (для выражения в (23.133)):

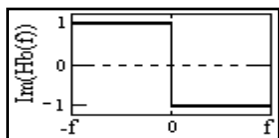
$$j \cdot \tilde{S}(\omega) = -\text{Re } S(\omega_l) - j \cdot \text{Im}(\omega_l) + \text{Re } S(\omega_r) + j \cdot \text{Im}(\omega_r).$$

Отсюда нетрудно видеть результат:

$$Z_s(\omega) = S(\omega) + j \tilde{S}(\omega) = 2 \cdot \text{Re } S(\omega_r) + j \cdot 2 \cdot \text{Im}(\omega_r) = 2 \cdot S(\omega_r),$$

что полностью соответствует выражению (23.132). В краткой форме:

$$\tilde{S}(\omega) = \begin{cases} jS(\omega), & \omega < 0, \\ 0, & \omega = 0, \\ -jS(\omega), & \omega > 0. \end{cases} \quad \tilde{S}(\omega) = -j \cdot \text{sgn}(\omega) \cdot S(\omega), \quad (23.135)$$



$$\text{Hb}(\omega) = -j \cdot \text{sgn}(\omega) = \begin{cases} j, & \omega < 0 \\ 0, & \omega = 0 \\ -j, & \omega > 0 \end{cases} \quad (23.136)$$

Таким образом, спектральная плотность  $\tilde{S}(\omega)$  аналитически сопряженного сигнала  $\tilde{s}(t)$  образуется из спектра  $S(\omega)$  исходного сигнала  $s(t)$  умножением на функцию  $-j \cdot \text{sgn}(\omega)$ . Это обеспечивает при суммировании  $S(\omega) + j \tilde{S}(\omega)$  удвоение амплитуд частотных составляющих в области положительных частот и их взаимную компенсацию в области отрицательных частот.

Из выражения (23.135) в спектральной области непосредственно следует соответствующая связь функций  $s(t)$  и  $\tilde{s}(t)$  во временной области:

$$\tilde{s}(t) = s(t) * \text{hb}(t), \quad (23.137)$$

$$s(t) = -\tilde{s}(t) * \text{hb}(t). \quad (23.138)$$

где  $\text{hb}(t) = \text{TF}[-j \cdot \text{sgn}(\omega)] = 1/(\pi t)$  – обратное преобразование Фурье функции  $-j \cdot \text{sgn}(\omega)$ .

Пример преобразования сигнала  $x(t)$  оператором Гильберта для формирования аналитического сигнала  $z_x(t) = x(t) + j \cdot \tilde{x}(t)$  приведен

на рис. 23.52.

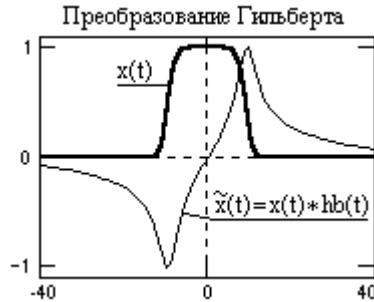


Рис. 23.52.

Частотную характеристику оператора Гильберта (23.136) можно записать и в следующем виде:

$$H_b(\omega) = |H_b(\omega)| \cdot \exp(j\varphi_h(\omega)), \text{ где } |H_b(\omega)| = 1.$$

$$H_b(\omega) = -j \cdot \text{sgn}(\omega) = \begin{cases} \exp(j(\pi/2)), & \omega > 0 \\ 0, & \omega = 0 \\ \exp(j(\pi/2)), & \omega < 0 \end{cases}, \quad (23.139)$$

Если спектр функции  $x(t)$  также представить в форме

$$S(\omega) = |S(\omega)| \cdot \exp(j\varphi_s(\omega)),$$

то выражение (23.135) преобразуется к следующей форме:

$$\tilde{S}(\omega) = |S(\omega)| \cdot \exp(j\varphi_s(\omega)) \cdot \exp(j\varphi_h(\omega)) = |S(\omega)| \cdot \exp[j(\varphi_s(\omega) + \varphi_h(\omega))], \quad (23.140)$$

т.е. модуль  $|S(\omega)|$  - амплитудный спектр сигнала  $\tilde{x}(t)$  как результат преобразования Гильберта сигнала  $s(t)$ , не изменяется и остается равным амплитудному спектру сигнала  $s(t)$ . Фазовый спектр сигнала  $\tilde{x}(t)$  (начальные фазовые углы всех гармонических составляющих сигнала) сдвигается на  $-90^\circ$  при  $\omega > 0$  и на  $90^\circ$  при  $\omega < 0$  относительно фазового спектра сигнала  $s(t)$ . Но такой фазовый сдвиг означает не что иное, как превращение косинусных гармоник в синусные, а синусных в косинусные. Это можно наглядно видеть на единичной гармонике. Так, если  $x(t) = \cos(2\pi f_0 t)$ , то имеем следующее преобразование Гильберта через частотную область:

$$\tilde{x}(t) = \text{TH}[x(t)] \Leftrightarrow \text{TF}[\text{TH}[x(t)]] = -j \text{sgn}(f) \cdot [\delta(f+f_0) + \delta(f-f_0)]/2.$$

$$\tilde{X}(f) = -j \cdot [-\delta(f+f_0) + \delta(f-f_0)]/2 = j \cdot [\delta(f+f_0) - \delta(f-f_0)]/2.$$

Но последнее уравнение - спектр синусоиды. При обратном преобразовании Фурье:

$$\tilde{x}(t) = \text{TF}^{-1}[\tilde{X}(f)] = \sin(2\pi f_0 t).$$

При  $x(t) = \sin(2\pi f_0 t)$  аналогичная операция дает  $\tilde{x}(t) = -\cos(2\pi f_0 t)$ . Знак минус демонстрирует отставание (запаздывание) выходного сигнала преобразования, как операции свертки, от входного сигнала. Для гармонических сигналов любой частоты с любой начальной фазой это запаздывание составляет четверть периода колебаний. На рис. 23.53 этот сдвиг на четверть периода для единичной гармонической составляющей (несущей частоты радиоимпульса) виден достаточно наглядно.

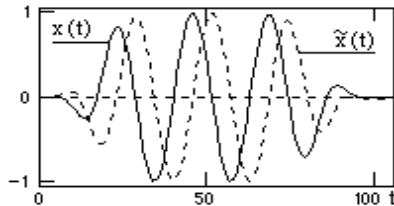


Рис. 23.53.

Таким образом, аналитический сигнал, по существу, представляет собой сумму двух ортогональных сигналов, все гармонические составляющие которых сдвинуты по фазе на  $90^\circ$  друг относительно друга.

### **23.20. Примеры применения аналитических сигналов**

**Огибающая и мгновенная фаза сигналов.** Допустим, что имеем зарегистрированный радиоимпульсный сигнал  $x(t)$  с несущей частотой  $\omega_0$ , который содержит определенную информацию, заключенную в огибающей сигнала  $u(t)$  и его фазе  $\varphi(t)$ :

$$x(t) = u(t) \cos(\omega_0 t + \varphi(t)). \quad (23.141)$$

Требуется выделить информационные составляющие сигнала. Запишем выражение (23.141) в другой форме:

$$x(t) = a(t) \cdot \cos(\omega_0 t) + b(t) \cdot \sin(\omega_0 t), \quad (23.142)$$

где функции  $a(t)$  и  $b(t)$  называются низкочастотными квадратурными составляющими сигнала  $x(t)$ :

$$\begin{aligned} a(t) &= u(t) \cos \varphi t, & b(t) &= u(t) \sin \varphi t. \\ u(t) &= \sqrt{a^2(t) + b^2(t)}, & \text{tg } \varphi(t) &= b(t)/a(t). \end{aligned}$$

С использованием преобразования Гильберта из сигнала  $x(t)$  можно сформировать аналитически сопряженный сигнал  $\tilde{x}(t)$ . Математическую форму сигнала  $\tilde{x}(t)$  получим из выражения (23.142) с учетом свойства модуляции преобразования Гильберта:

$$\tilde{x}(t) = a(t) \cdot \sin(\omega_0 t) - b(t) \cdot \cos(\omega_0 t).$$

$$z(t) = x(t) + j \cdot \tilde{X}(t).$$

Квадрат модуля сигнала  $z(t)$ :

$$|z(t)|^2 = x^2(t) + \tilde{X}^2(t) = a^2(t)[\cos^2(\omega_0 t) + \sin^2(\omega_0 t)] + b^2(t)[\cos^2(\omega_0 t) + \sin^2(\omega_0 t)] = u^2(t).$$

Отсюда, огибающая  $u(t)$  и мгновенная фаза  $\phi(t)$  сигнала  $x(t)$ :

$$u(t) = \sqrt{x^2(t) + \tilde{X}^2(t)}. \quad (23.143)$$

$$\phi(t) = \omega_0 t + \varphi(t) = \arctg[\tilde{X}(t)/x(t)]. \quad (23.144)$$

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

**Мгновенная частота** сигнала определяется по скорости изменения мгновенной фазы:

$$d\phi(t)/dt = \frac{\tilde{X}'(t) \cdot x(t) - x'(t) \cdot \tilde{X}(t)}{\tilde{X}^2(t) + x^2(t)}. \quad (23.145)$$

Для амплитудно-модулированных сигналов с одной несущей частотой эти результаты достаточно очевидны (см. рис. 23.53).

Но выражения (23.143-23.145), полученные из общих соображений, остаются действительными и для любых произвольных сигналов.

На рис. 23.54. представлен сигнал, сложенный двумя гармониками:

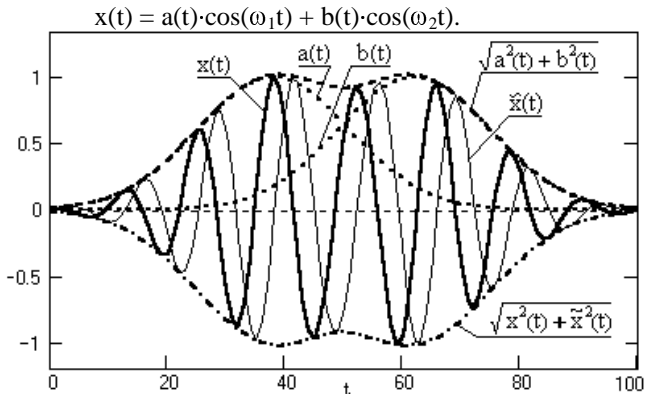


Рис. 23.54

Квадратурное дополнение и аналитический сигнал:

$$\tilde{X}(t) = a(t) \cdot \sin(\omega_1 t) + b(t) \cdot \sin(\omega_2 t).$$

$$z(t) = x(t) + j \cdot \tilde{X}(t).$$

Огибающая такого сигнала, как это можно видеть на рисунке 23.54, должна вычисляться по формуле (23.143). При этом для данного сигнала получаем:



$$u(t) = \sqrt{a^2(t) + b^2(t) + 2 \cdot a(t) \cdot b(t) \cdot \cos((\omega_2 - \omega_1)t)},$$

что может существенно отличаться от функции  $\sqrt{a^2(t) + b^2(t)}$ .

Мгновенная фаза сигнала, график которой приведен на рис. 23.55, зависит от времени нелинейно:

$$\phi(t) = \arg \operatorname{tg} \frac{a(t) \cdot \sin(\omega_1 t) + b(t) \cdot \sin(\omega_2 t)}{a(t) \cdot \cos(\omega_1 t) + b(t) \cdot \cos(\omega_2 t)}.$$

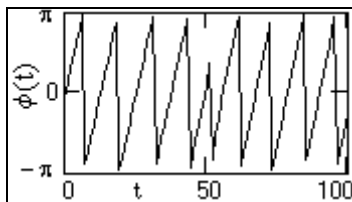


Рис. 23.55

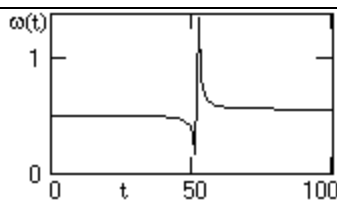


Рис. 23.56

Мгновенная частота сигнала (рис. 23.144) также имеет нелинейную зависимость от времени, причем ее значения могут существенно превышать даже суммарное значение частот, составляющих сигнал:

$$\omega(t) = \frac{\omega_1 a^2(t) + \omega_2 b^2(t) + a(t)b(t) \cdot (\omega_1 + \omega_2) \cdot \cos((\omega_2 - \omega_1)t)}{a^2(t) + b^2(t) + 2 \cdot a(t)b(t) \cdot \cos((\omega_2 - \omega_1)t)}.$$

Аналогичная методика определения огибающих, мгновенных значений фазы и частоты применяется и для анализа случайных процессов.

**Огибающие модулированных сигналов.** В качестве примера применения огибающих рассмотрим связь форм относительно узкополосных радиосигналов с формой модулирующих сообщений.

**Амплитудная модуляция.** Уравнение модулированного сигнала:

$$x(t) = U_o \cdot [1 + m \cdot s(t)] \cdot \cos \omega_o t, \quad s(t) \leq 1, \quad m \leq 1$$

Квадратурное дополнение и аналитический сигнал:

$$\tilde{x}(t) = U_o \cdot [1 + m \cdot s(t)] \cdot \sin \omega_o t, \quad z_x(t) = x(t) + j\tilde{x}(t).$$

Огибающая сигнала  $x(t)$ :

$$u(t) = |z_x(t)| = U_o \cdot [1 + m \cdot s(t)],$$

т.е. точно повторяет форму модулирующего сообщения (см. рис. 23.57)

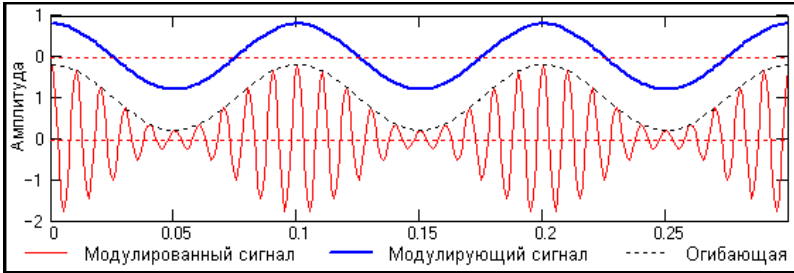


Рис. 23.57. Амплитудная модуляция.

**Балансная модуляция.** Уравнение модулированного сигнала, приведенного на рис. 23.58:



Рис. 23.58. Балансная модуляция.

$$x(t) = U_o \cdot s(t) \cdot \cos \omega_o t,$$

Квадратурное дополнение, аналитический сигнал, огибающая сигнала  $x(t)$ :

$$\tilde{x}(t) = U_o \cdot s(t) \cdot \sin \omega_o t, \quad z_x(t) = x(t) + j\tilde{x}(t), \quad u(t) = |z_x(t)| = U_o \cdot |s(t)|.$$

Огибающая сигнала  $x(t)$  существенно отличается от модулирующего сообщения, но связана с ним простым соотношением.

**Анализ каузальных систем.** Каузальная (физически осуществимая) линейная система задается односторонним импульсным откликом  $h(t)$ ,  $t \geq 0$ , и имеет частотную характеристику  $H(f)$ :

$$H(f) = X(f) - jY(f),$$

Осуществим обратное преобразование Фурье для всех частей выражения раздельно:

$$h(t) = x(t) + y(t),$$

$$x(t) = \int_{-\infty}^{\infty} X(f) \cos(2\pi ft) df,$$

$$y(t) = \int_{-\infty}^{\infty} Y(f) \sin(2\pi ft) df,$$

где  $x(t)$  и  $y(t)$  - четная и нечетная части функции  $h(t)$ . Нечетная функция  $y(t)$  в каузальной системе однозначно связана с четной функцией  $x(t)$ :

$$y(t) = \operatorname{sgn}(t) \cdot x(t). \quad (23.146)$$

Осуществляя обратное преобразование Фурье обеих частей равенства (23.146) при известном преобразовании сигнатурной функции ( $\operatorname{sgn}(t) \Leftrightarrow -j/(\pi f)$ ), получаем:

$$\operatorname{TF}[y(t)] = (-j/\pi f) * X(f) = (-j/\pi) \int_{-\infty}^{\infty} [X(u)/(f-u)] du.$$

Отсюда:

$$Y(f) = (1/\pi) \int_{-\infty}^{\infty} [X(u)/(f-u)] du = \operatorname{TH}[X(f)],$$

т.е. мнимая часть спектра импульсного отклика каузальной системы (и любой каузальной функции) является преобразованием Гильберта действительной части спектра. Соответственно, уравнение для определения действительной компоненты спектра по мнимой части:

$$X(f) = -\operatorname{TH}[Y(f)] = -(1/\pi) \int_{-\infty}^{\infty} [Y(u)/(f-u)] dv.$$

### **23.21. Преобразование Гильберта**

Преобразование Гильберта для любого произвольного сигнала представляет собой идеальный широкополосный фазовращатель, который осуществляет поворот начальных фаз всех частотных составляющих сигнала на угол, равный  $90^\circ$  (сдвиг на  $\pi/2$ ). Применение преобразования Гильберта позволяет выполнять квадратурную модуляцию сигналов, в каждой текущей координате модулированных сигналов производить определение огибающей и мгновенной фазы и частоты сигналов, выполнять анализ каузальных систем обработки сигналов.

**Определение преобразования.** Прямое преобразование Гильберта произвольной действительной функции  $x(t)$ ,  $-\infty < t < \infty$ , результат которого будем отображать со знаком тильды над индексом исходной функции, задается сверткой  $x(t)$  с функцией  $hb(t) = 1/(\pi t)$ :

$$\begin{aligned} \tilde{x}(t) &= \operatorname{TH}[x(t)] = x(t) * (1/\pi t), \\ \tilde{x}(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau. \end{aligned} \quad (23.147)$$

Функция  $1/(t-\tau)$  называется *ядром преобразования* Гильберта. Обратное преобразование Гильберта определяется выражением:

$$x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\tilde{X}(\tau)}{t-\tau} d\tau \quad (23.148)$$

Интегралы преобразования имеет особую точку  $a = t-\tau \Rightarrow 0$ , в которой при вычислении используется их главное значение по Коши:

$$\lim_{a \rightarrow 0} \left[ \int_{-\infty}^{t-a} \dots + \int_{t+a}^{\infty} \dots \right].$$

Оператор Гильберта определен по аргументу от  $-\infty$  до  $\infty$  и имеет полюс в точке  $t=0$  с разрывом значений от  $-\infty$  до  $\infty$ . Основной участок формы оператора Гильберта и пример преобразования сигнала приведены на рис. 23.59. Функции  $x(t)$  и  $\tilde{X}(t)$  обычно называют сопряженными по Гильберту.

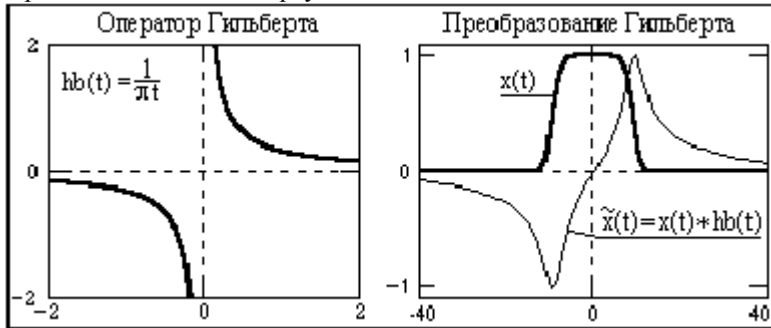


Рис. 23.59.

**Спектральная характеристика преобразования.**

Выполним преобразование Фурье функции (17.1.1). В общей форме:

$$\tilde{X}(f) = \text{TF}[\tilde{x}(t)] = X(f) \cdot \text{Hb}(f), \quad (23.149)$$

$$\tilde{X}(f) = \int_{-\infty}^{\infty} \tilde{x}(t) \exp(-j2\pi ft) dt. \quad (23.150)$$

Заметим, что произведение  $X(f) \cdot \text{Hb}(f)$  не является преобразованием Гильберта спектральной функции  $X(f)$ . Это не более чем преобразование Фурье свертки функций:  $x(t) * \text{hb}(t) \Leftrightarrow X(f) \cdot \text{Hb}(f)$ , которое позволяет вычислить результат преобразования Гильберта во временной области через частотную область:

$$\tilde{x}(t) = \int_{-\infty}^{\infty} \tilde{X}(f) \cdot \exp(j2\pi ft) df = \int_{-\infty}^{\infty} X(f) \cdot \text{Hb}(f) \cdot \exp(j2\pi ft) df .$$

Функция  $\text{hb}(t)=1/\pi t$  является нечетной, а спектр этой функции, представленный только мнимой частью, является обратной

сигнатурной функцией (рис. 23.60):

$$\text{Hb}(f) = \text{TF}[1/\pi t] = -j \cdot \text{sgn}(f) = \begin{cases} -j, & f > 0 \\ 0, & f = 0 \\ j, & f < 0 \end{cases} \quad (23.151)$$

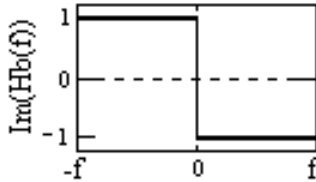


Рис. 23.60.

Соответственно, формулы (23.147) задают преобразование сигнала  $x(t)$  системой, частотная передаточная характеристика которой отображается функцией  $-j \cdot \text{sgn}(f)$ . Фурье-образ функции  $\tilde{x}(t)$ :

$$\tilde{X}(f) = -j \text{sgn}(f) \cdot X(f). \quad (23.152)$$

**Изменение спектра сигналов** при выполнении преобразования Гильберта. На рис. 23.61 приведено преобразование радиоимпульсного сигнала  $x(t)$  с несущей частотой  $f_0$  в сигнал  $\tilde{x}(t)$  во временной области непосредственно через операцию свертки по (23.147).

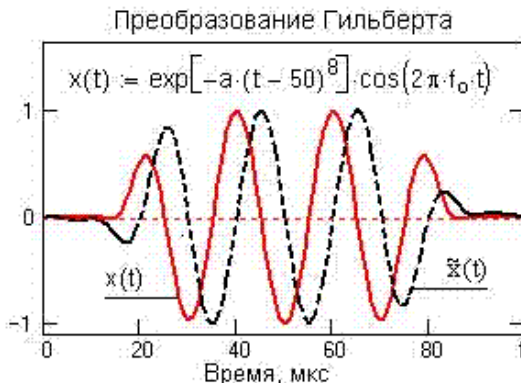


Рис. 23.61.

Сигнал  $x(t)$  является односторонним каузальным. Спектр сигнала содержит реальную и мнимую составляющие, т.е. может быть записан в виде  $X(\omega) = \text{Re}(X(\omega)) + j \cdot \text{Im}(X(\omega))$ . Эти составляющие для сигнала

$x(t)$  показаны непрерывными кривыми на рис. 23.62.

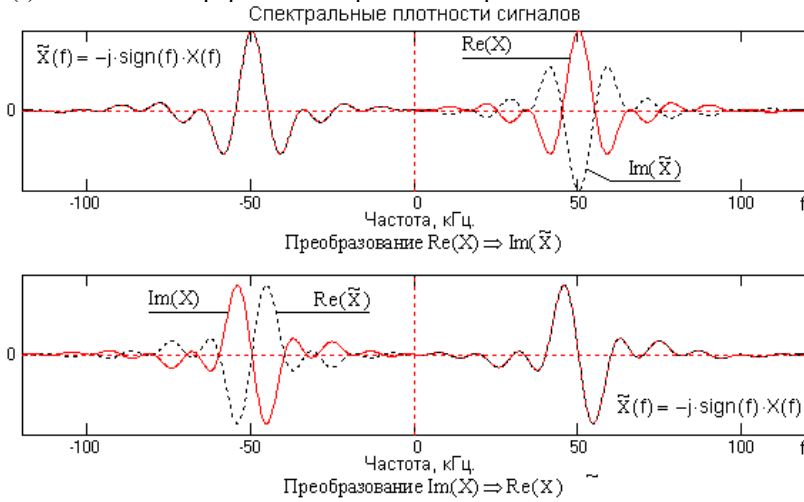


Рис. 32.62.

При выполнении преобразования (23.152) реальная и мнимая части спектра  $X(\omega)$  умножаются на  $-j \cdot \text{sgn}(\omega)$ . Функция  $\text{Re}(X(\omega))$  умножается на 1 при  $\omega < 0$ , на 0 при  $\omega = 0$  и на  $-1$  при  $\omega > 0$ , и тем самым превращается в нечетную мнимую часть  $\text{Im}(\tilde{X}(\omega))$  спектра  $\tilde{X}(\omega)$  функции  $\tilde{x}(t)$ , показанную пунктиром. Это означает, что все косинусные гармоники сигнала, которым соответствует реальная часть спектра сигнала, превращаются в синусные гармоники.

Аналогично на функцию  $-j \cdot \text{sgn}(\omega)$  умножается и мнимая функция  $j \cdot \text{Im}(X(\omega))$ , при этом сигнатурная функция инвертируется ( $-j \cdot j = 1$ ), что меняет знак левой части функции  $\text{Im}(X(\omega))$  – области отрицательных частот, и превращает ее в реальную четную часть  $\text{Re}(\tilde{X}(\omega))$  спектра  $\tilde{X}(\omega)$ . Синусные гармоники спектра сигнала превращаются в косинусные гармоники.

Угол, на который изменяется фаза гармоник, можно определить из следующих соображений. Частотную характеристику  $H_b(f)$  (23.151) можно записать в следующем виде:

$$H_b(f) = |H_b(f)| \cdot \exp(j\phi_b(f)), \text{ где } |H_b(f)| = 1.$$

$$H_b(f) = -j \cdot \text{sgn}(f) = \begin{cases} \exp(j(\pi/2)), & f > 0 \\ 0, & f = 0, \\ \exp(j(\pi/2)), & f < 0 \end{cases} \quad (23.153)$$

Если спектр функции  $x(t)$  также представить в виде

$$X(f) = |X(f)| \cdot \exp(j\varphi_x(f)),$$

то выражение (23.149) преобразуется к следующей форме:

$$\tilde{X}(f) = |X(f)| \cdot \exp(j\varphi_x(f)) \cdot \exp(j\varphi_h(f)) = |X(f)| \cdot \exp[j(\varphi_x(f) + \varphi_h(f))], \quad (23.154)$$

т.е. амплитудный спектр сигнала  $\tilde{X}(t)$  – как результат преобразования Гильберта сигнала  $x(t)$ , не изменяется и остается равным амплитудному спектру сигнала  $x(t)$ . Фазовый спектр сигнала  $\tilde{X}(t)$  (начальные фазовые углы всех гармонических составляющих сигнала) сдвигается на  $-90^\circ$  при  $f > 0$  и на  $90^\circ$  при  $f < 0$  относительно фазового спектра сигнала  $x(t)$ . Но такой фазовый сдвиг и означает не что иное, как превращение косинусных гармоник в синусные, а синусных в косинусные. Последнее нетрудно проверить на единичной гармонике.

Если  $x(t) = \cos(2\pi f_0 t)$ , то имеем следующее преобразование Гильберта через частотную область:

$$\tilde{X}(f) = H[x(t)] \Leftrightarrow TF[H[x(t)]] = -j \operatorname{sgn}(f) \cdot [\delta(f+f_0) + \delta(f-f_0)]/2. \quad (23.155)$$

$$\tilde{X}(f) = -j \cdot [-\delta(f+f_0) + \delta(f-f_0)]/2 = j \cdot [\delta(f+f_0) - \delta(f-f_0)]/2. \quad (23.156)$$

Но последнее уравнение - спектр синусоиды. При обратном преобразовании Фурье:

$$\tilde{X}(t) = TF^{-1}[\tilde{X}(f)] = \sin(2\pi f_0 t). \quad (23.157)$$

При  $x(t) = \sin(2\pi f_0 t)$  аналогичная операция дает  $\tilde{X}(t) = -\cos(2\pi f_0 t)$ . Знак минус демонстрирует отставание (запаздывание) выходного сигнала преобразования, как операции свертки, от входного сигнала. Для гармонических сигналов любой частоты с любой начальной фазой это запаздывание составляет четверть периода колебаний. На рис. 23.61 этот сдвиг на четверть периода для единичной гармонической составляющей (несущей частоты радиоимпульса) виден достаточно наглядно. Таким образом, преобразование Гильберта, по существу, представляет собой идеальный фазовращатель, осуществляющий фазовый сдвиг на  $90^\circ$  всех частотных составляющих сигналов одновременно.

Сдвиг фазы спектров сигналов  $x(t)$  на  $\pi/2$  определяет изменение четности и самих сигналов: четный  $x(t) \Leftrightarrow$  нечетный  $\tilde{X}(t)$ , и наоборот.

**Спектры каузальных функций.** Каузальная (физически осуществимая) линейная система (равно как и произвольная причинно обусловленная функция) задается односторонним импульсным откликом (выражением)  $h(t)$ ,  $t \geq 0$ , и имеет частотную характеристику  $H(f)$ :

$$H(f) = A(f) - j \cdot B(f),$$

где  $A(f)$  и  $B(f)$  - действительная (четная) и мнимая (нечетная) части

частотной характеристики. Осуществим обратное преобразование Фурье для всех частей этого выражения:

$$h(t) = a(t) + b(t),$$

$$a(t) = \int_{-\infty}^{\infty} A(f) \cos(2\pi ft) df, \quad b(t) = \int_{-\infty}^{\infty} B(f) \sin(2\pi ft) df,$$

где  $a(t)$  и  $b(t)$  - соответственно четная и нечетная части импульсного отклика  $h(t)$ . Условие каузальности для импульсного отклика ( $h(t) = 0$  при  $t < 0$ ) будет выполнено, если при  $t < 0$  функции  $a(t)$  и  $b(t)$  компенсируют друг друга. Тогда общее условие каузальности, как можно наглядно видеть на рис. 23.63, с учетом нечетности функции  $b(t)$ , запишется в следующем виде:

$$\begin{aligned} b(t) &= -a(t), & b(t) &= -h(t)/2, & a(t) &= h(t)/2, & t < 0, \\ b(t) &= 0, & a(t) &= a(0) = h(0), & t &= 0, \\ b(t) &= a(t) & &= h(t)/2, & t > 0. \end{aligned} \tag{23.158}$$

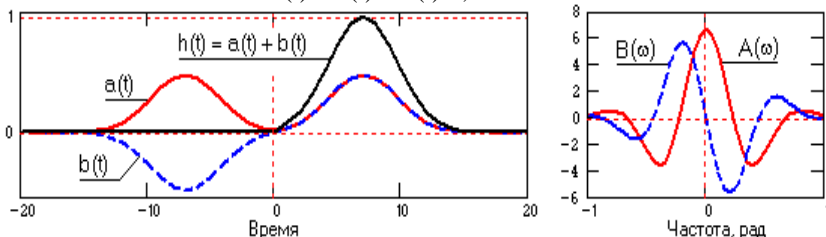


Рис. 23.63. Параметры каузальной функции.

Из этих условий следует, что нечетная функция  $b(t)$  в каузальной системе однозначно связана с четной функцией:

$$b(t) = \text{sgn}(t) \cdot a(t), \tag{23.159}$$

Осуществляя преобразование Фурье обеих частей данного равенства при известном преобразовании сигнатурной функции ( $\text{sgn}(t) \Leftrightarrow j/\pi f$ ), получаем:

$$\text{Im}(H(f)) = (j/\pi f) * A(f),$$

или, с учетом знака мнимой части:

$$B(f) = -(1/\pi f) * A(f) = -(1/\pi) \int_{-\infty}^{\infty} [A(v)/(f-v)] dv. \tag{23.160}$$

Аналогично определяется и действительная компонента спектра по мнимой части:

$$A(f) = (1/\pi f) * B(f) = (1/\pi) \int_{-\infty}^{\infty} [B(v)/(f-v)] dv. \tag{23.161}$$

Таким образом, реальная и мнимая части спектра физически осуществимых (односторонних) систем, а равно и произвольных каузальных сигналов, связаны парой преобразований Гильберта. Они



позволяют производить определение любой, действительной или мнимой, части частотной характеристики каузальной функции путем свертки другой ее части с функцией  $1/\pi f$ .

### 23.22. Свойства преобразования Гильберта

Для любых произвольных функций  $x(t)$  и  $y(t)$ , имеющих Фурье – образы  $X(\omega)$ ,  $Y(\omega)$  и преобразования Гильберта  $\tilde{x}(t) = \text{TH}[x(t)]$  и  $\tilde{y}(t) = \text{TH}[y(t)]$ , действительны следующие свойства:

**Линейность.**  $\text{TH}[a \cdot x(t) + b \cdot y(t)] = a \cdot \tilde{x}(t) + b \cdot \tilde{y}(t)$  при любых постоянных значениях коэффициентов  $a$  и  $b$  для любых произвольных функций  $x(t)$  и  $y(t)$ .

**Сдвиг.**  $\text{TH}[x(t-a)] = \tilde{x}(t-a)$ .

**Преобразование константы,**  $a$  в силу линейности преобразования, и постоянной составляющей сигнала, равно нулю. Это прямо следует из нечетности ядра преобразования Гильберта. Отсюда следует, что при преобразовании Гильберта из квадратурной составляющей исключается постоянная составляющая.

**Свойство четности и нечетности** определяется сдвигом всех гармоник сигнала на  $\pi/2$ , при этом четные сигналы  $x(t)$  дают нечетные сигналы  $\tilde{x}(t)$ , и наоборот. Это действительно и для произвольных сигналов относительно их четных и нечетных частей.

**Последовательное двойное преобразование** Гильберта возвращает исходную функцию с обратным знаком

$$\text{TH}[\text{TH}[x(t)]] = \text{TH}[\tilde{x}(t)] = -x(t).$$

Это определяется тем, что при двойном преобразовании фазы всех гармоники сигнала сдвигаются на  $\pi$ , что изменяет знак их гармоник. Однако в силу исключения из сигнала при первом преобразовании постоянной составляющей, при двойном преобразовании сигнал  $x(t)$  восстанавливается с исключенным средним значением по интервалу задания.

**Обратное преобразование Гильберта,** по существу, это второе преобразование в последовательном двойном преобразовании Гильберта с изменением знака результата:

$$x(t) = \text{TH}^{-1}[\tilde{x}(t)] = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\tilde{x}(\tau)}{t-\tau} d\tau = \tilde{x}(t) * (-1/\pi t). \quad (23.162)$$

Альтернативная форма вычисления  $x(t)$  из  $\tilde{x}(t)$ :

$$x(t) = \text{TF}^{-1}[j \operatorname{sgn}(f) \cdot \text{TF}[\tilde{x}(t)]]. \quad (23.163)$$

**Подобие** при изменении масштаба аргумента:

$$\text{TH}[x(at)] = \tilde{x}(at).$$

**Энергетическая эквивалентность:**

$$\int_{-\infty}^{\infty} x^2(t) dt = \int_{-\infty}^{\infty} \tilde{X}^2(t) dt. \quad (23.164)$$

Это следует из теоремы Парсеваля (энергия сигнала равна сумме энергии всех частотных составляющих сигнала) и равенства модулей спектров сигналов  $x(t)$  и  $\tilde{X}(t)$  (энергия сигнала не зависит от его фазовочастотной характеристики).

**Свойство ортогональности:**

$$\int_{-\infty}^{\infty} x(t) \cdot \tilde{X}(t) dt = 0. \quad (23.165)$$

Если все косинусные составляющие сигнала  $x(t)$  превращаются в ортогональные им синусные составляющие сигнала  $\tilde{X}(t)$ , а синусные – в ортогональные им косинусные, то и сигналы  $x(t)$  и  $\tilde{X}(t)$  должны быть ортогональны. Из теоремы Парсеваля следует:

$$\int_{-\infty}^{\infty} x(t) \cdot \tilde{X}(t) dt = \int_{-\infty}^{\infty} X^*(f) \cdot \tilde{X}(f) df.$$

Функция

$$X^*(f) \cdot \tilde{X}(f) = -X^* \cdot j \operatorname{sgn}(f) \cdot X(f) = -j \operatorname{sgn}(f) \cdot |X(f)|^2$$

является нечетной, а поэтому определенный интеграл от этой функции по симметричным относительно нуля пределам равен нулю. Ортогональность сигналов наглядно видна на рис. 23.59.

**Свойство свертки:**

$$\operatorname{TH}[x(t) * y(t)] = \tilde{X}(t) * y(t) = x(t) * \tilde{Y}(t). \quad (23.166)$$

Это вытекает из следующих соображений.

Примем  $z(t) = x(t) * y(t)$ , при этом:

$$Z(f) = X(f) \cdot Y(f), \quad \tilde{Z}(f) = -j \operatorname{sgn}(f) \cdot Z(f) = -j \operatorname{sgn}(f) X(f) \cdot Y(f).$$

$$\tilde{Z}(f) = [-j \operatorname{sgn}(f) X(f)] \cdot Y(f) = \tilde{X}(f) \cdot Y(f) \Leftrightarrow \tilde{X}(f) * y(t).$$

$$\tilde{Z}(f) = X(f) \cdot [-j \operatorname{sgn}(f) Y(f)] = X(f) \cdot \tilde{Y}(f) \Leftrightarrow x(t) * \tilde{Y}(t).$$

**Отсутствие коммутативности** с преобразованием Фурье:

$$\operatorname{TF}[\operatorname{TH}[x(t)]] \neq \operatorname{TH}[\operatorname{TF}[x(t)]]. \quad (23.167)$$

**Свойство модуляции:** Модулирующие сигналы  $u(t)$ , как правило, имеют ограниченный спектр, максимальные частоты которого  $\Omega$  много меньше значения несущей частоты  $\omega_0$ , при этом:

$$\operatorname{TH}[u(t) \cdot \cos(\omega_0 t)] = u(t) \cdot \sin(\omega_0 t). \quad (23.168)$$

Для четных функций  $u(t)$  это свойство очевидно. При переходе в частотную область:

$$\operatorname{TH}[u(t) \cdot \cos(\omega_0 t)] \Leftrightarrow -j \operatorname{sgn}(\omega) \cdot [U(\omega) * (\delta(\omega + \omega_0) + \delta(\omega - \omega_0))].$$

Множитель  $-j \operatorname{sgn}(\omega)$  является знаковой константой по  $\omega$  и

может быть внесен под интеграл свертки и умножен на  $(\delta(\omega+\omega_0)+\delta(\omega-\omega_0))$ , что, как уже рассматривалось ранее (см. 23.155 – 23.157), при обратном преобразовании Фурье дает  $u(t)\cdot\sin(\omega_0t)$ .

Аналогично можно показать, что

$$\text{TH}[u(t)\cdot\sin(\omega_0t)] = -u(t)\cdot\cos(\omega_0t). \quad (23.169)$$

## 23.23. Вычисление преобразования Гильберта

### Преобразование Гильберта аналоговых сигналов

целесообразно выполнять не по формулам линейной свертки с оператором  $1/\pi t$ , который стремится к  $\infty$  при  $t \Rightarrow 0$ , а через спектр аналитической функции:

$$z(t) = x(t) + j \cdot \tilde{X}(t) \Leftrightarrow X(f) + j \cdot \tilde{X}(f) = Z(f). \quad (23.170)$$

Заменяя в этом выражении функцию  $\tilde{X}(f) = -j \operatorname{sgn}(f) \cdot X(f)$ , получаем:

$$Z(f) = [1 + \operatorname{sgn}(f)] \cdot X(f), \quad (23.171)$$

где функция  $1 + \operatorname{sgn}(f)$  равна 0 при  $f < 0$ , 1 при  $f = 0$  и 2 при  $f > 0$ , при этом:

$$Z(f) = \begin{cases} 2 X(f), & f > 0, \\ X(0), & f = 0, \\ 0, & f < 0. \end{cases} \quad (23.172)$$

т.е. спектр функции  $z(t)$  является односторонним и устанавливается непосредственно по спектру функции  $x(t)$  при  $f \geq 0$ . Обратное преобразование Фурье функции  $Z(f)$  должно давать комплексную функцию  $z(t)$ , при этом из (23.172) следует:

$$x(t) = \operatorname{Re} \left[ 2 \int_0^{\infty} X(f) \exp(j2\pi ft) df \right], \quad (23.173)$$

$$\tilde{X}(t) = \operatorname{Im} \left[ 2 \int_0^{\infty} X(f) \exp(j2\pi ft) df \right]. \quad (23.174)$$

В дискретной форме, при общем числе  $N$  отсчетов функции  $x(t)$  с шагом  $\Delta t$ , с шагом по частоте  $\Delta f = 1/(N\Delta t)$ :

$$X(n\Delta f) = \Delta t \sum_{k=0}^{N-1} x(k\Delta t) \cdot \exp(-j2\pi kn/N), \quad n = 0, 1, \dots, N/2. \quad (23.175)$$

$$x(k\Delta t) = \Delta f \cdot \operatorname{Re} \left[ X_0 + 2 \sum_{n=1}^{N/2} X(n\Delta f) \cdot \exp(j2\pi kn/N) \right]. \quad (23.176)$$

$$\tilde{X}(k\Delta t) = 2\Delta f \cdot \text{Im} \left[ \sum_{n=1}^{N/2} X(n\Delta f) \cdot \exp(j2\pi kn/N) \right]. \quad (23.177)$$

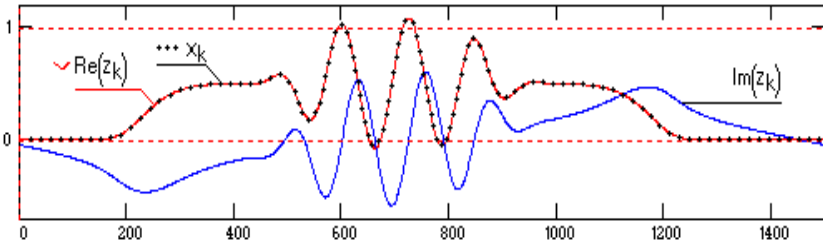


Рис. 23.64.

На рис. 23.64 приведен пример преобразования Гильберта, выполненный через частотную область. Естественно, что при реальном использовании преобразования Гильберта выполнять вычисления по (23.176) не требуется.

**Оператор дискретного преобразования Гильберта**

$hb(k\Delta t) \Leftarrow 1/\pi t$  на интервале от  $-T$  до  $T$  с шагом  $\Delta t$  можно получить обратным преобразованием Фурье частотной характеристики  $Hb(f)$  (выражение 23.151) в интервале от  $-f_N$  до  $f_N$  ( $f_N=1/2\Delta t$ ). При  $\Delta t=1$ :

$$hb(k\Delta t) = \int_{-f_N}^{f_N} Hb(f) \exp(j2\pi f k \Delta t) df = \int_{-1/2}^0 j \exp(j2\pi f k \Delta t) df - \int_0^{1/2} \exp(j2\pi f k \Delta t) df$$

$$df = [1/(2\pi k \Delta t)] \cdot [1 - \exp(-j\pi k \Delta t) - \exp(j\pi k \Delta t) + 1] =$$

$$= [1/(\pi k \Delta t)] \cdot [1 - (\exp(-j\pi k \Delta t) + \exp(j\pi k \Delta t))/2] =$$

$$= [1/(\pi k \Delta t)] \cdot (1 - \cos(\pi k \Delta t)) = [2/(\pi k \Delta t)] \sin^2(\pi k \Delta t/2). \quad (23.178)$$

$$hb(k\Delta t) = 2/(\pi k \Delta t), \quad k = \pm 1, \pm 3, \pm 5, \dots, \quad (23.179)$$

$$hb(k\Delta t) = 0, \quad k = \pm 0, \pm 2, \pm 4, \dots$$

Нетрудно убедиться, что коэффициент усиления постоянной составляющей оператора равен нулю, а коэффициент усиления дисперсии помех равен 1.

В частотной области при выполнении преобразования Гильберта спектральных функций оператор свертки  $hb(k\Delta f) \Leftarrow 1/\pi f$  не отличается от приведенного для временной области.

## **24. Вейвлетные преобразования сигналов**

### **24.1. Базисные функции вейвлет-преобразования**

Аналитика вейвлетных преобразований сигналов определяются математической базой разложения сигналов, которая аналогична преобразованиям Фурье. Основной отличительной особенностью вейвлет-преобразований является новый базис разложения сигналов - вейвлетные функции. Свойства вейвлетов принципиально важны как для самой возможности разложения сигналов по единичным вейвлетным функциям, так и для целенаправленных действий над вейвлетными спектрами сигналов, в том числе с последующей реконструкцией сигналов по обработанным вейвлетным спектрам.

Вейвлеты могут быть ортогональными, полуортогональными, биортогональными. Вейвлетные функции могут быть симметричными, асимметричными и несимметричными, с компактной областью определения и не имеющие таковой, а также иметь различную степень гладкости. Некоторые функции имеют аналитическое выражение, другие – быстрый алгоритм вычисления вейвлет-преобразования. Для практики желательно было бы иметь ортогональные симметричные и асимметричные вейвлеты, но таких идеальных вейвлетов не существует. Наибольшее применение находят биортогональные вейвлеты.

Базисными функциями вейвлет-преобразований могут быть самые различные функции с компактным носителем - модулированные импульсами синусоиды, функции со скачками уровня и т.п. Они обеспечивает хорошее отображение и анализ сигналов с локальными особенностями, в том числе со скачками, разрывами и перепадами значений с большой крутизной.

Было бы желательно иметь такое вейвлет-преобразование сигналов, которое обеспечивало полную информационную эквивалентность вейвлетного спектра сигналов временному представлению и однозначность декомпозиции - реконструкции сигналов. Однако это возможно только при использовании ортогональных и биортогональных вейвлетов. Для качественного анализа сигналов и локальных особенностей в сигналах может применяться более обширная номенклатура вейвлетных функций, которые хотя и не обеспечивают реконструкцию сигналов, но позволяют оценить информационное содержание сигналов и динамику

изменения этой информации.

**Определение вейвлета.** К вейвлетам относятся локализованные функции, которые конструируются из одного материнского вейвлета  $\psi(t)$  (или по любой другой независимой переменной) путем операций сдвига по аргументу ( $b$ ) и масштабного изменения ( $a$ ):

$$\psi_{ab}(t) = (1/\sqrt{|a|}) \psi((t-b)/a), \quad (a, b) \in \mathbb{R}, \quad \psi(t) \in L^2(\mathbb{R}).$$

где множитель  $(1/\sqrt{|a|})$  обеспечивает независимость нормы функций от масштабного числа 'a'.

Непрерывное вейвлет-преобразование сигнала  $s(t) \in L^2(\mathbb{R})$ , которое применяется для качественного частотно-временного анализа, по смыслу соответствует преобразованию Фурье с заменой гармонического базиса  $\exp(-j\omega t)$  на вейвлетный  $\psi((t-b)/a)$ :

$$C(a, b) = \langle s(t), \psi_{ab}(t) \rangle = (1/\sqrt{|a|}) \int_{-\infty}^{\infty} s(t) \psi((t-b)/a) dt, \quad (a, b) \in \mathbb{R}, \quad a \neq 0.$$

Вейвлетный масштабно-временной спектр  $C(a,b)$  в отличие от фурье-спектра является функцией двух аргументов: масштаба вейвлета 'a' (в единицах, обратных частоте), и временного смещения вейвлета по сигналу 'b' (в единицах времени), при этом параметры 'a' и 'b' могут принимать любые значения в пределах областей их определения.

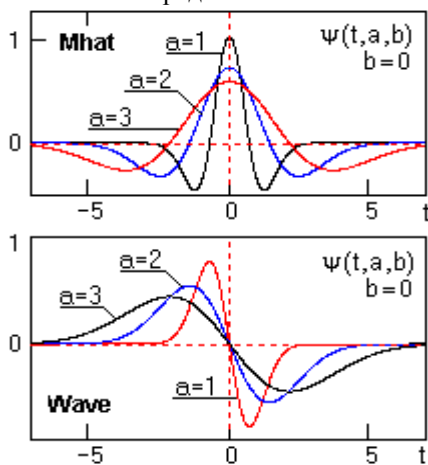


Рис. 24.1. Вейвлеты Mhat и Wave.

На рис. 24.1 приведены примеры простейших неортогональных вейвлетов четного (Mhat) и нечетного (Wave) типов.

Для количественных методов анализа в качестве вейвлетных базисов можно использовать любые локализованные функции  $\psi(t)$ ,

если для них существуют функции-двойники  $\psi^\#(t)$ , такие, что семейства  $\{\psi_{ab}(t)\}$  и  $\{\psi^\#_{ab}(t)\}$  могут образовывать парные базисы функционального пространства  $L^2(\mathbb{R})$ . Вейвлеты, определенные таким образом, позволяют представить любую произвольную функцию в пространстве  $L^2(\mathbb{R})$  в виде ряда:

$$s(t) = \sum_{a,b} C(a,b) \psi^\#_{ab}(t), \quad (a, b) \in I,$$

где коэффициенты  $C(a,b)$  – проекции сигнала на вейвлетный базис пространства:

$$C(a,b) = \langle s(t), \psi_{ab}(t) \rangle = \int_{-\infty}^{\infty} s(t) \psi_{ab}(t) dt.$$

Если вейвлет  $\psi(t)$  обладает свойством ортогональности, то  $\psi^\#(t) \equiv \psi(t)$  и вейвлетный базис ортогонален. Вейвлет может быть неортогональным, однако если он имеет двойника, и пара  $(\psi(t), \psi^\#(t))$  дает возможность сформировать семейства  $\{\psi_{mk}(t)\}$  и  $\{\psi^\#_{zp}(t)\}$ , удовлетворяющие условию биортогональности на целых числах  $I$ :

$$\langle \psi_{mk}(t), \psi^\#_{zp}(t) \rangle = \delta_{mz} \cdot \delta_{kp}, \quad m,k,z,p \in I,$$

то возможно разложение сигналов на вейвлетные ряды с обратной формулой реконструкции.

### **Свойства вейвлета.**

- **Локализация.** Вейвлет должен быть непрерывным, интегрируемым, иметь компактный носитель и быть локализованным как во времени (в пространстве), так и по частоте. Если вейвлет в пространстве сужается, то его "средняя" частота повышается, спектр вейвлета перемещается в область более высоких частот и расширяется. Этот процесс должен быть линейным – сужение вейвлета вдвое должно повышать его "среднюю" частоту и ширину спектра также вдвое.

- **Нулевое среднее значение,** т.е. выполнение условия для нулевого момента:

$$\int_{-\infty}^{\infty} \psi(t) dt = 0,$$

что обеспечивает нулевое усиление постоянной составляющей сигналов, нулевое значение частотного спектра вейвлета при  $\omega=0$ , и локализацию спектра вейвлета в виде полосового фильтра с центром на определенной (доминирующей) частоте  $\omega_0$ .

- **Ограниченность.** Необходимое и достаточное условие:

$$\|\psi(t)\|^2 = \int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty$$

- **Автомодельность базиса** или самоподобие. Форма всех

базисных вейвлетов  $\psi_{ab}(t)$  должна быть подобна материнскому вейвлету  $\psi(t)$ , т.е. должна оставаться одной и той же при сдвигах и масштабировании (растяжении/сжатии), иметь одно и то же число осцилляций.

**Отображение преобразования.** Результатом вейвлет-преобразования одномерного числового ряда (сигнала) является двумерный массив значений коэффициентов  $C(a,b)$ . Распределение этих значений в пространстве  $(a,b)$  - временной масштаб, временная локализация, дает информацию об изменении во времени относительного вклада в сигнале вейвлетных компонент разного масштаба и называется спектром коэффициентов вейвлет-преобразования, масштабно-временным (частотно-временным) спектром или просто вейвлет-спектром (wavelet spectrum).

Спектр  $C(a,b)$  одномерного сигнала представляет собой поверхность в трехмерном пространстве. Способы визуализации спектра могут быть самыми различными. Наиболее распространенный способ – проекция на плоскость  $ab$  с изолиниями (изоуровнями), что позволяет проследить изменения коэффициентов на разных масштабах во времени, а также выявить картину локальных экстремумов этих поверхностей ("холмов" и "впадин"), так называемый "скелет" (skeleton) структуры анализируемого процесса. При широком диапазоне масштабов применяются логарифмические координаты ( $\log a, b$ ). Пример вейвлетного спектра простейшего сигнала при его разложении вейвлетом Mhat приведен на рис. 24.2.

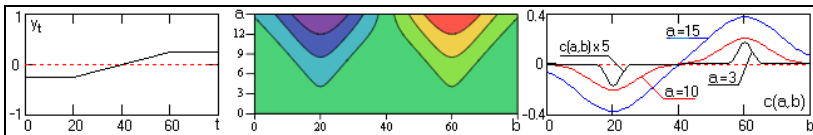


Рис. 24.2. Сигнал, вейвлетный Mhat - спектр и масштабные сечения спектра.

По вертикальным сечениям (сечениям сдвига  $b$ ) вейвлет-спектр отражает компонентный состав сигнала (из данного комплекта вейвлетов) в каждый текущий момент. По смыслу преобразования, как скалярного произведения сигнала с вейвлетом, ясно, что значения коэффициентов в каждой текущей временной точке по масштабным сечениям тем больше, чем сильнее корреляция между вейвлетом данного масштаба и поведением сигнала в окрестностях этой точки. Соответственно, сечения по параметру 'a' демонстрируют изменения в сигнале компоненты данного масштаба 'a' со временем.

Вейвлетные составляющие сигнала в сечениях его спектра не



имеют ничего общего с синусоидами, и представлены, как правило, сигналами достаточно сложной и не всегда понятной формы, что может затруднять их наглядное представление и понимание.

**Вейвлетные функции.** Выбор анализирующего вейвлета определяется тем, какую информацию необходимо извлечь из сигнала. С учетом характерных особенностей различных вейвлетов во временном и в частотном пространстве, можно выявлять в анализируемых сигналах те или иные свойства и особенности, которые незаметны на графиках сигналов, особенно в присутствии шумов. При этом задача реконструкции сигнала может и не ставится, что расширяет семейство используемых регулярных вейвлетных функций, в том числе неортогональных. Более того, вейвлет может конструироваться непосредственно под ту локальную особенность в сигнале, которая подлежит выделению или обнаружению, если ее форма априорно известна.

При анализе сигналов вейвлетами четного типа (симметричными или близкими к симметричным) гармоническим сигналам обычно соответствуют яркие горизонтальные полосы вейвлетных пиков и впадин на доминирующих частотах вейвлетов, совпадающих с частотой гармоник сигналов. Нарушения гладкости сигналов фиксируются вертикальными полосами, пики в сигналах выделяются максимумами, а впадины – минимумами вейвлетных коэффициентов. Напротив, вейвлеты нечетного типа более резко реагируют на скачки и быстрые изменения в сигналах, отмечая их максимумами или минимумами в зависимости от знака дифференциалов. Чем резче выражены особенности сигналов, тем сильнее они выделяются на спектрограммах.

Для конструирования таких вейвлетов часто используются производные функции Гаусса, которые имеют наилучшую локализацию как во временной, так и в частотной областях. В общей форме уравнение базового вейвлета:

$$\psi_n(x) = (-1)^{n+1} d^n[\exp(-x^2/2)]/dx^n, \quad n \geq 1, \quad (24.1)$$

**WAVE-вейвлет** вычисляется по первой производной (n=1) и приведен на рис. 24.3 во временной и частотной области для трех значений масштабных коэффициентов 'a'. Форма вейвлета относится к нечетным функциям и, соответственно, спектр вейвлета является мнимым. Уравнение вейвлета по (24.1) с единичной нормой:

$$\psi(t, a, b) = \frac{-1.786}{\sqrt{2}} \cdot \frac{t-b}{a} \cdot \exp \left[ -\left( \frac{t-b}{a} \right)^2 \right]. \quad (24.2)$$

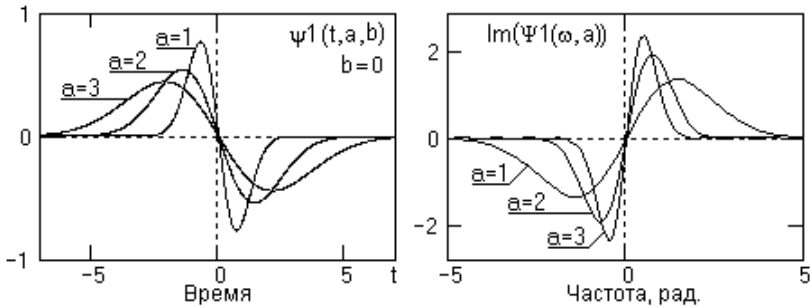


Рис. 24.3. Вейвлет Wave.

На рис. 24.4 приведен пример применения вейвлета для анализа двух однотипных сигналов, один из которых осложнен шумами с мощностью на уровне мощности самого сигнала. Как следует из рисунка, контурная масштабно-временная картина вейвлетных коэффициентов, а равно и ее сечения на больших значениях масштабных коэффициентов 'а' очень точно и уверенно фиксирует положение вершины информационного сигнала сменой знака коэффициентов  $C(a,b)$ .

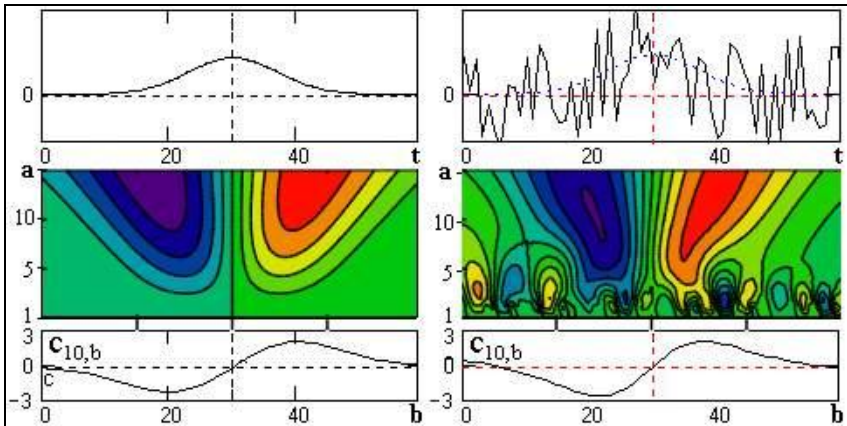


Рис. 24.4.

**МНАТ-вейвлет** (Mexican hat – мексиканская шляпа) вычисляется по второй производной ( $n=2$ ) и приведен на рис. 24.5. Вейвлет симметричен, спектр вейвлета представлен только действительной частью и хорошо локализован по частоте, нулевой и первый моменты вейвлета равны нулю. Применяется для анализа сложных сигналов. Уравнение вейвлета по (24.1):

$$\psi(t, a, b) = \frac{1.031}{\sqrt{2}} \cdot \left\{ \exp \left[ -\left( \frac{t-b}{a} \right)^2 \right] - 2 \left( \frac{t-b}{a} \right)^2 \cdot \exp \left[ -\left( \frac{t-b}{a} \right)^2 \right] \right\} \quad (24.3)$$

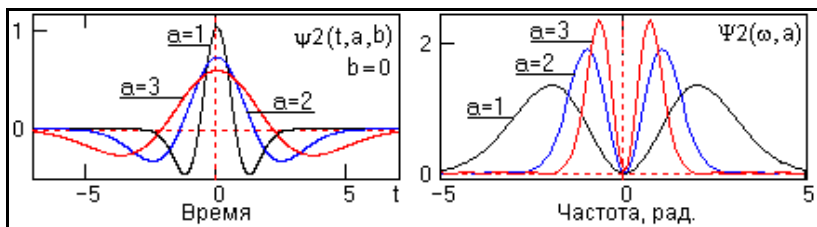


Рис. 24.5. Вейвлет МНАТ.

На рис. 24.6 приведен пример использования вейвлета для анализа сложного сигнала  $y(t)$ . Модель сигнала образована суммой сигналов разной структуры. Сигналы  $y_1$ - $y_2$  представляют собой функции Гаусса разного масштабного уровня, сигнал  $y_3$  - прямоугольный импульс, сигнал  $y_4$  задан в виде тренда с постоянным значением дифференциала. На контурном графике вейвлет-коэффициентов можно видеть выделение всех трех основных структур сигнала при полном исключении тренда. Особенно четко выделяются границы скачков прямоугольной структуры. Справа на рисунке приведена полная трехмерная картина вейвлет-преобразования.

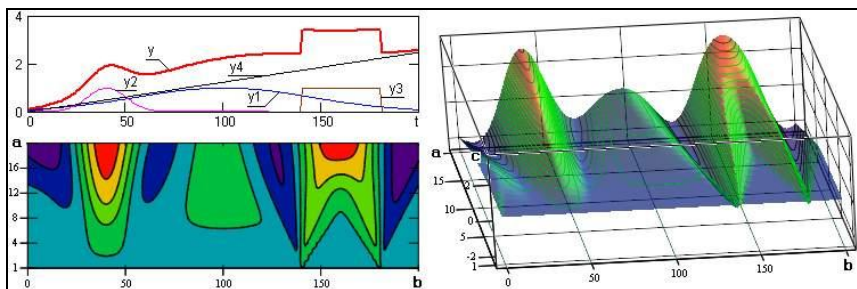


Рис. 24.6.

Вейвлет широко используется в двумерном варианте для анализа изотропных полей. На его основе возможно также построение двумерного неизотропного базиса с хорошей угловой избирательностью при добавлении к сдвигам и масштабированию вейвлета его вращения.

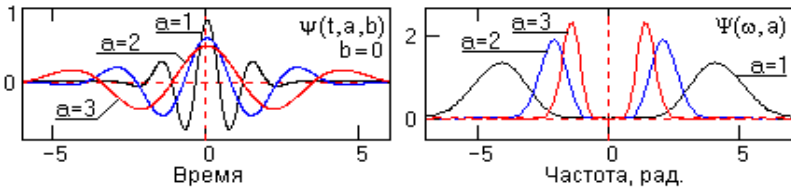


Рис. 24.7.

При повышении номера производной функции (24.1) временная область определения вейвлета несколько увеличивается при существенном повышении доминирующей частоты вейвлета и степени его локализации в частотной области. Вейвлеты  $n$ -го порядка позволяют анализировать более тонкие высокочастотные структуры сигналов, подавляя низкочастотные компоненты. Пример вейвлета по восьмой производной приведен на рис. 24.7.

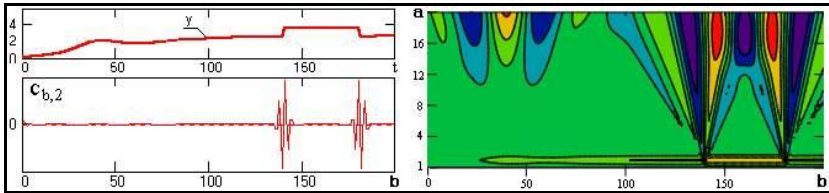


Рис. 24.8.

Практическое следствие повышения степени локализации вейвлетов в частотной области наглядно видно на рис. 24.8 на примере преобразования той же функции, что и на рис. 24.6. Сравнение рисунков показывает существенное повышение чувствительности вейвлета к высокочастотным составляющим сигнала на малых масштабных коэффициентах.

## 24.2. Свойства вейвлет-преобразования

Результаты вейвлет-преобразования, как скалярного произведения вейвлета и сигнальной функции, содержат комбинированную информацию об анализируемом сигнале и самом вейвлете. Получение объективной информации о сигнале базируется на свойствах вейвлет-преобразования, общих для вейвлетов всех типов. Рассмотрим основные из этих свойств. Для обозначения операции вейвлет-преобразования произвольных функций  $s(t)$  будем применять индекс  $TW[s(t)]$ .

### *Линейность.*

$$TW[\alpha \cdot s_1(t) + \beta \cdot s_2(t)] = \alpha \cdot TW[s_1(t)] + \beta \cdot TW[s_2(t)]. \quad (24.4)$$

*Инвариантность относительно сдвига.* Сдвиг сигнала во

времени на  $t_0$  приводит к сдвигу вейвлет-спектра также на  $t_0$ :

$$TW[s(t-t_0)] = C(a, b-t_0). \quad (24.5)$$

**Инвариантность относительно масштабирования.**

Растяжение (сжатие) сигнала приводит к сжатию (растяжению) вейвлет-спектра сигнала:

$$TW[s(t/a_0)] = (1/a_0) \cdot C(a/a_0, b/a_0). \quad (24.6)$$

**Дифференцирование.**

$$d^n \{TW[s(t)]\} / dt^n = TW[d^n(s(t)) / dt^n]. \quad (24.7)$$

$$TW[d^n(s(t)) / dt^n] = (-1)^n \int_{-\infty}^{\infty} s(t) [d^n(\psi(t)) / dt^n] dt. \quad (24.8)$$

Если анализирующий вейвлет задан формулой, то это может быть очень полезным для анализа сигналов. Проанализировать особенности высокого порядка или мелкомасштабные вариации сигнала  $s(t)$  можно дифференцированием нужного числа раз либо вейвлета, либо самого сигнала.

**Аналог теоремы Парсеваля** для ортогональных и биортогональных вейвлетов.

$$\int_{\mathbb{R}} s_1(t) \cdot s_2^*(t) = C_{\psi}^{-1} \int_{\mathbb{R}} a^{-2} C(a,b) C^*(a,b) da db. \quad (24.9)$$

Отсюда следует, что энергия сигнала может вычисляться через коэффициенты вейвлет-преобразования.

Определения и свойства одномерного непрерывного вейвлет-преобразования обобщаются на многомерный и на дискретный случаи.

### 24.3. Вейвлет-преобразование простых сигналов

Вейвлет-преобразование, выполняемое при анализе сигналов для выявления в них каких-либо особенностей и места их локализации без обратной реконструкции, допускает применение любых типов вейвлетов, как ортогональных, так и неортогональных. Чаще всего для этих целей используются симметричные вейвлеты. Ниже приводятся результаты применения вейвлета  $Mhat$  для анализа сигналов простых форм. Вычисления выполнены с вейвлетом (24.3) по формуле:

$$c(a,b) = \sum_{t=b-3a}^{b+3a} s(t) \psi(t,a,b), \quad (24.10)$$

где суммирование выполняется в растворе угла влияния (по области достоверности) с шагом  $\Delta t = \Delta b = \Delta a = 1$ . Так как при непрерывном разложении скейлинг-функция не используется, отсчет значений 'a' начинается с 1, а ряд коэффициентов  $c(0,b)$  оставляется нулевым и определяет нулевой фон контурных графиков спектра.

Импульсы Кронекера (положительный и отрицательный), вейвлет-спектр импульсов и сечения спектра на трех значениях

параметра 'a' приведены на рис. 24.9. Цветовая гамма спектра здесь и в дальнейшем соответствует естественному цветоряду от красного (большие значения) к фиолетовому (малые значения коэффициентов).

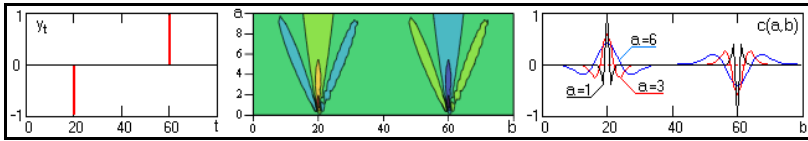


Рис. 24.9. Преобразование импульсов Кронекера.

На сечениях спектра видно, что свертка единичных импульсов с разномасштабными вейвлетами повторяет форму вейвлетов, как это и положено при операции свертки. Соответственно, линии максимальных экстремумов на сечениях ("хребты" и "долины", в зависимости от полярности) определяют временное положение импульсов, а боковые экстремумы противоположной полярности образуют характерные лепестки в конусе угла влияния, который хорошо выражен.

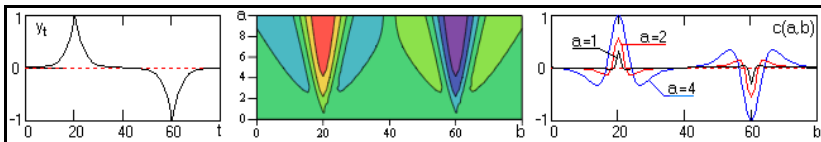


Рис. 24.10. Преобразование функций Лапласа.

Аналогичный характер спектра сохраняется и для любых локальных неоднородностей на сигналах в форме пиков (рис. 24.10) со смещением максимумов (минимумов) коэффициентов  $c(a,b)$  со значений  $a=1$  в область больших значений 'a' (в зависимости от эффективной ширины пиков).

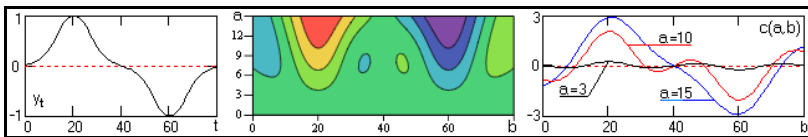


Рис. 24.11. Преобразование функций Гаусса.

На рис. 24.11 приведен спектр функций Гаусса. При сглаживании вершин пиковых неоднородностей форма цветовых конусов также сглаживается, но "хребтовые" ("долинные") линии достаточно точно фиксируют на временной оси положение центров локальных неоднородностей.

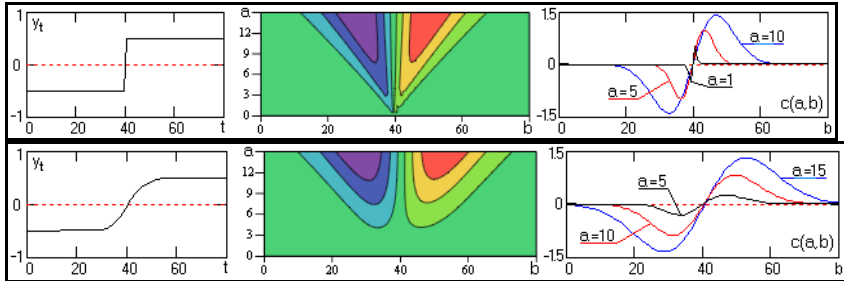


Рис. 24.12. Преобразование перепада постоянного значения функций.

На рис. 24.12 приведены спектры двух разных по крутизне перепадов постоянных значений функции. Центры перепадов фиксируются по переходу через нуль значений коэффициентов  $c(a,b)$ , а крутизна перепадов отражается, в основном, на значениях функции  $c(a,b)$  при малых значениях параметра 'a'.

При изломах функций спектрограммы уверенно фиксируют место изломов максимумами (минимумами) значений коэффициентов  $c(a,b)$ , как это показано на рис. 24.13. При наложении на такие функции шумов точное определение места изломов по масштабным сечениям на малых значениях параметра 'a' становится невозможным, однако на больших значениях параметра 'a' такая возможность сохраняется, естественно, с уменьшением точности локализации.

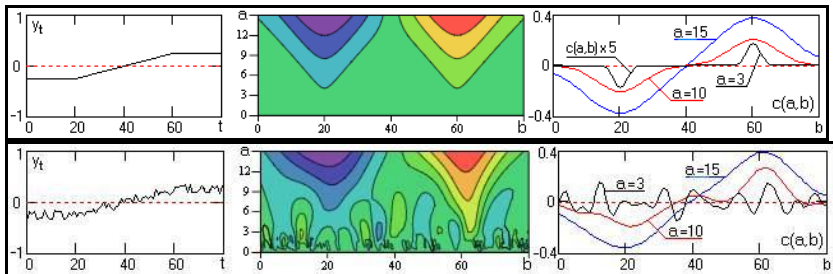


Рис. 24.13. Преобразование изломов функций.

Аналогичный характер имеет влияние шумов и на другие локальные сигналы (рис. 24.9-24.12). Если спектральные особенности сигналов распространяются на диапазон значений параметра 'a', то имеется возможность идентификации этих сигналов и их места на временной оси.

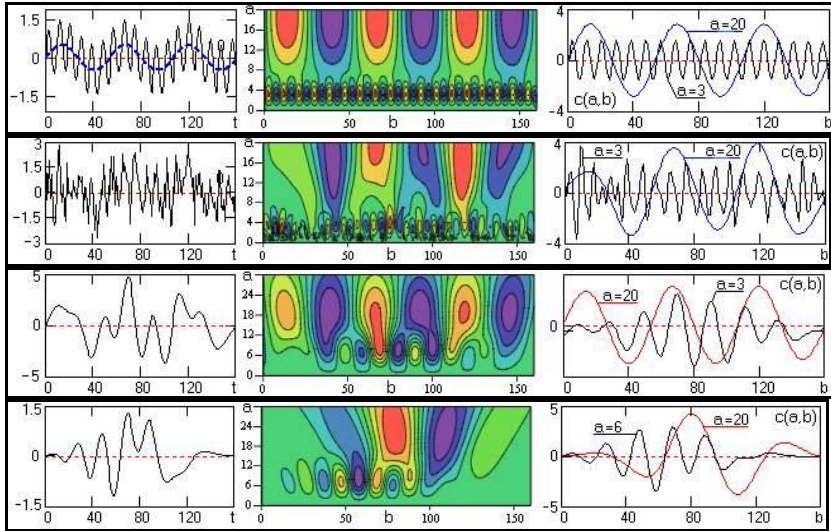


Рис. 24.14. Преобразование гармонических функций.

Разделение гармонических функций на масштабной оси спектров, в том числе при наложении сильных шумовых процессов, приведено в примерах на рис. 24.14. Приведенный пример имеет чисто иллюстративный характер, так как для выделения гармонических процессов с постоянной частотой во времени целесообразно использовать спектральный анализ и частотные полосовые фильтры. Тем не менее, для локальных сигналов, типа модулированных гармоник, вейвлет-спектры достаточно хорошо показывают место их локализации на временной оси.

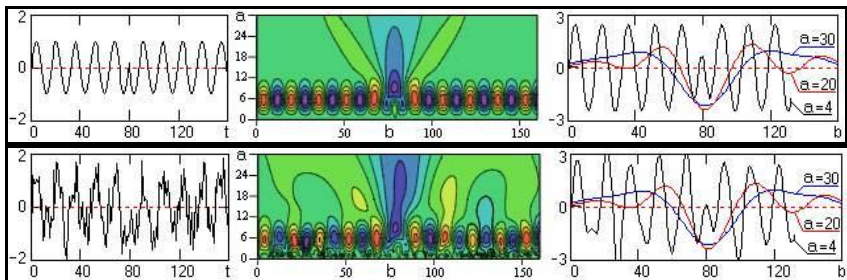


Рис. 24.15. Изменение фазы гармонического сигнала.

На рис. 24.15 приведен пример еще одной характерной особенности гармонического сигнала – изменение его фазы на  $180^\circ$ , которое хорошо фиксируется на всех масштабах вейвлета, а,



следовательно, достаточно легко определяется даже в присутствии сильных шумовых сигналов. При наложении синусоидальных сигналов на тренд вейвлет-преобразование на больших масштабах позволяет достаточно уверенно выделять характерные особенности тренда. Пример выделения изломов тренда приведен на рис. 24.16.

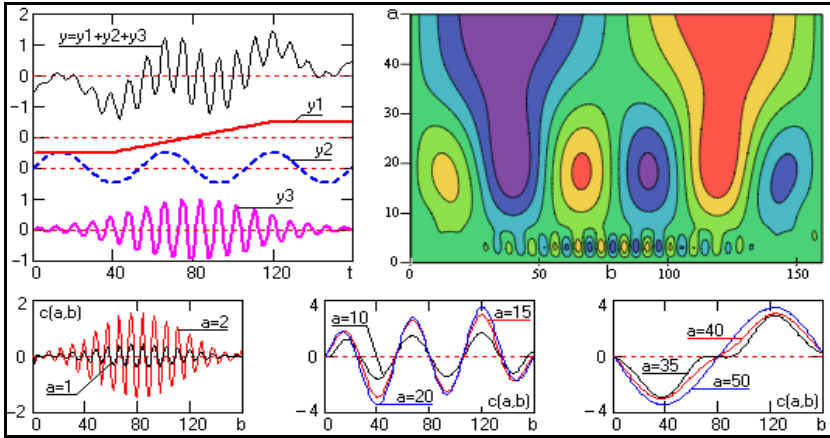


Рис. 24.16. Преобразование суммы трех сигналов.

Форма вейвлета (четность или нечетность), доминирующая частота и степень ее локализации существенно влияют на вейвлет-спектры анализируемых сигналов и на возможности выделения его локальных особенностей. На нижеследующих рисунках приведены сравнительные спектры простых сигналов при использовании вейвлетов Wave (нечетный, рис. 24.3), Mhat (четный, рис. 24.5) и вейвлета по 8-й производной Гаусса (рис. 24.17-24.24), который также является четным, и имеет в 4 раза более высокую доминирующую частоту, чем вейвлет Mhat.

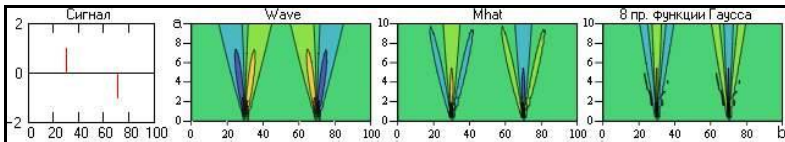


Рис. 24.17. Импульсы Кронекера.

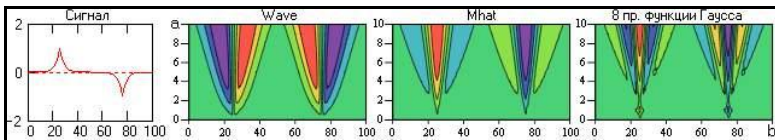


Рис. 24.18. Пики Лапласа.

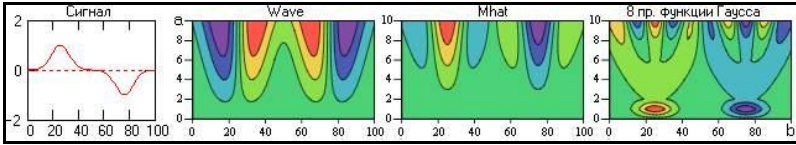


Рис. 24.19. Функции Гаусса.

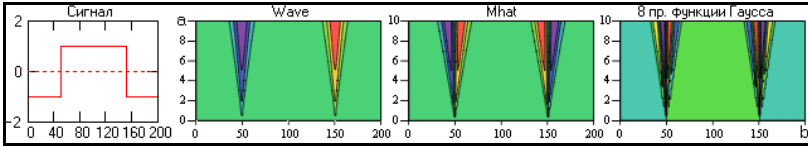


Рис. 24.20. Крутые скачки.

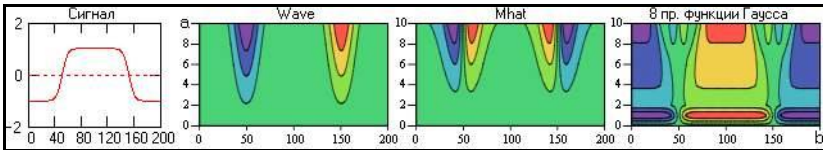


Рис. 24.21. Сглаженные скачки.

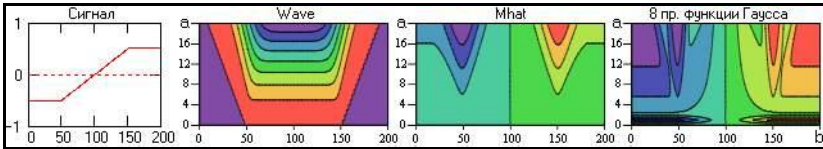


Рис. 24.22. Изломы функций

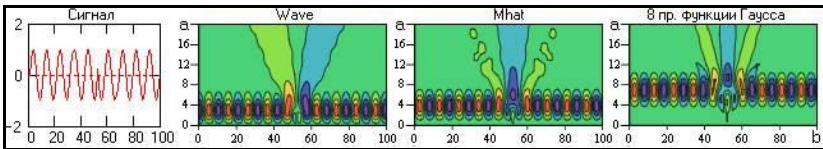


Рис. 24.23. Фазовые скачки гармоник.

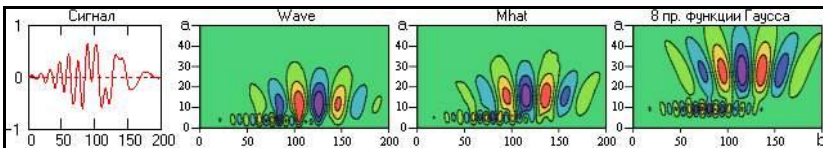


Рис. 24.24. Сумма двух модулированных синусоид.

Заметим, что при анализе произвольных сигналов использование разнотипных вейвлетов позволяет повысить достоверность выделения локальных особенностей сигналов.

## **24.4. Введение в вейвлетный кратномасштабный анализ**

В практике передачи информации часто требуется представить сигнал в виде совокупности его последовательных приближений. Например, при просмотре и передаче изображений с выборкой из некоторой базы данных можно сначала передать грубую его версию, а затем (при необходимости) последовательно ее уточнять. При сжатии изображений часто без потери качества можно убирать из изображения незначимые мелкомасштабные детали.

Произвольный информационный сигнал обычно рассматривается в виде суммы разнотипных составляющих: региональной функции тренда, циклических компонент с определенным периодом повторения, локальных особенностей (аномалий) разного порядка и флуктуаций (шумов) вокруг всех вышеперечисленных составляющих сигнала. Инструментом разделения (декомпозиции) сигналов на такие составляющие, анализа их порядка и реконструкции сигналов из определенных составляющих (или с исключением определенных составляющих, например шумов или малозначимых деталей) является кратномасштабный (многомасштабный) анализ (КМА).

КМА позволяет получить хорошее разрешение по времени (плохое по частоте) на высоких частотах и хорошее разрешение по частоте (плохое по времени) на низких частотах. Этот подход становится эффективным, если сигнал имеет короткие высокочастотные компоненты и протяженные низкочастотные компоненты. Именно такие сигналы и встречаются чаще всего.

Идея кратномасштабного анализа заключается в том, что разложение сигнала производится по ортогональному базису, образованному сдвигами и кратномасштабными копиями вейвлетной функции. Свертка сигнала с вейвлетами позволяет выделить характерные особенности сигнала в области локализации этих вейвлетов, причем, чем больший масштаб имеет вейвлет, тем более широкая область сигнала будет оказывать влияние на результат свертки. Понятие кратномасштабного анализа (Multiresolution analyses) является фундаментальным в теории вейвлетов. Для кратномасштабного анализа разработан быстрый каскадный алгоритм вычислений, подобный быстрому преобразованию Фурье.

## **24.5. Принципы кратномасштабного анализа**

### **Дискретные ортогональные преобразования.**

Непрерывное вейвлет-преобразование, равно как и его дискретный аналог с произвольным шагом по масштабу и сдвигу, обладает сильной избыточностью. Интуитивно понятно, что если какая-либо информация заключена в  $N$  отсчетах сигнала, то при любых преобразованиях сигнала для отображения этой информации без потерь в новом базисном пространстве должно быть необходимо и достаточно то же самое количество отсчетов  $N$ . С учетом принципа неопределенности Гейзенберга это означает, что для точного восстановления сигнала достаточно знать его вейвлет-преобразование на некоторой решетке частотно-временной области, густой в области высоких частот сигнала, и редкой в области низких частот. Идея КМА заключается в том, чтобы масштабировать вейвлет в некоторое постоянное число раз (например, 2), и при скольжении по сигналу сдвигать его во времени с шагом, равным интервалу носителя масштабированного вейвлета. Если обозначить количество масштабных строк индексом  $m$ , и принять  $N=2^m$ , то при  $N=32$  решетка вейвлетного спектра будет иметь всего  $m=5$  масштабных строк с количеством отсчетов в первой строке 16, во второй 8, в третьей 4, в четвертой 2, и в пятой 1, с общим количеством отсчетов 32, как и в исходном сигнале. При этом все сдвиги одного масштаба будут попарно ортогональны (нет перекрытия сдвигов), равно как и вейвлеты разных масштабов в силу их нулевого первого момента.

**Вейвлет Хаара.** Простейшие методы КМА, без всякой теоретической базы, использовались при обработке числовых данных уже достаточно давно. Рассмотрим один из таких методов на практическом примере анализа гистограмм, который обычно выполняется функцией Хаара (Haar), в дальнейшем получившей название вейвлета Хаара (рис. 24.25).

$$\psi_0(x) := \text{if}(x < 0 \vee x \geq 1, 0, \text{if}(x < 0.5, 1, -1))$$

$$\varphi_0(x) := \text{if}[(x < 0 \vee x \geq 1), 0, 1]$$

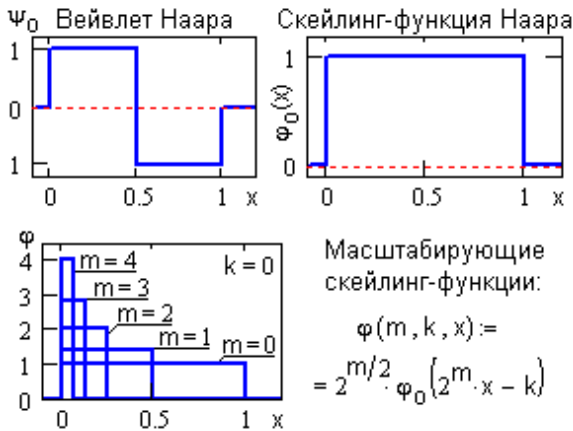


Рис. 24.25.

Допустим, что мы анализируем определенную зависимость  $s(x)$  на интервале  $0 \leq x \leq 1$ , показанную на рис. 24.26.

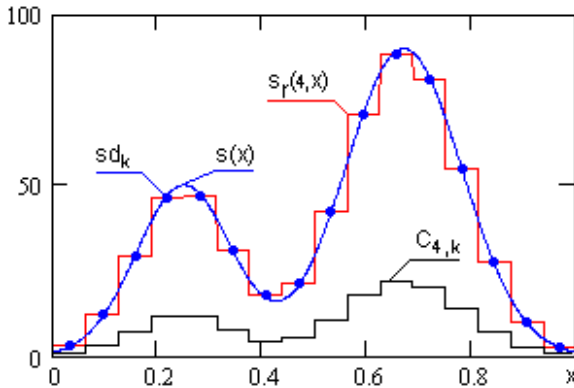


Рис. 24.26.

Функция нецентрированная, и для использования вейвлет-преобразования с последующим восстановлением исходного сигнала требует применения как вейвлета, так и его скейлинг-функции. На основе базовых функций вейвлета и скейлинг-функции Хаара, приведенных на рис. 24.25, записываем масштабированные функции:

$$\varphi_{m,k}(x) = 2^{m/2} \varphi(2^m x - k), \quad (24.11)$$

$$\psi_{m,k}(x) = 2^{m/2} \psi(2^m x - k). \quad (24.12)$$

Эти функции образуют нормированные взаимно ортогональные базисы пространства вейвлетных коэффициентов, на которые может быть разложен анализируемый сигнал. Ортогональность базисных функций является обязательным условием КМА, обеспечивающим возможность обратной реконструкции сигнала.

Для коротких и достаточно гладких кривых нет смысла устанавливать много уровней декомпозиции сигнала. Примем максимальное значение  $m$ , равным 4, при этом  $N=1/2^m=16$  с интервалом дискретизации данных, соответственно,  $\Delta x=1/N$ . В принципе, можно применять и задание исходного значения  $\Delta x$  с последующим определением количества уровней разложения.

При сдвиговой ортогональности прямоугольных базисных функций прямое преобразование (проекции сигнала на базис (24.11)) для непрерывных сигналов выполняется по формуле:

$$C_{m,k} = \int_{k\Delta x}^{(k+1)\Delta x} s(x) \varphi(2^m x - k) dx. \quad (24.13)$$

Значения коэффициентов при  $m=4$ :

Восстановление сигнала с четвертого уровня декомпозиции соответственно выполняется по формуле реконструкции:

$$s_r(m,x) = \sum_{k=0}^{N-1} C_{m,k}$$

$C_{m,k}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	0	0.774	2.996	7.384	11.596	11.65	7.733	4.471	5.355	10.615	17.721	22.048	20.246	13.711	6.848	2.522	0.684

$s_r(k\Delta x + \frac{\Delta x}{2})$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	0	3.095	11.884	29.536	46.395	46.599	30.933	17.884	21.418	42.462	70.983	88.193	80.983	54.846	27.391	10.086	2.738

$$\varphi_{m,k}(x), \quad m=4, N=16. \quad (24.14)$$

Восстановление исходной непрерывной функции сигнала  $s(x)$  скейлинг-функцией Хаара невозможно в силу того, что значение скейлинг-функции – константа шириной  $\Delta x$ , на которую умножается соответствующее значение  $C_{4,k}$  и распространяется на весь интервал  $k\Delta x - (k+1)\Delta x$  (кривая  $s_r(x)$  на рис. 24.26). Если выполнить перевод сигнала  $s(t)$  во временной ряд  $sd_k, k=0 \dots N-1$ , с осреднением по интервалам  $\Delta x$ , или с использованием (в общем случае произвольного вейвлета) его скейлинг-функции:

$$sd_k = 2^{m/2} \int_{k\Delta x}^{(k+1)\Delta x} s(x) \varphi_{m,k}(x) dx, \quad (24.15)$$

sd =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
	0	3.095	11.984	29.536	46.385	46.599	30.933	17.884	21.418	42.462	70.883	88.193	80.983	54.846	27.391	10.086	2.738

то нетрудно убедиться, что  $sd_k = s_i(k\Delta x + \Delta x/2)$  (числовые отсчеты  $sd_k$  на рис. 24.26 отнесены к середине интервалов  $\Delta x$ ).

В принципе, гистограмма  $sd_k$  может представлять собой непосредственные исходные дискретные данные (результаты измерений и т.п.). Сравнением выражений (24.15) и (24.13) нетрудно убедиться, что нулевой уровень разложения ( $m=m_{max}$ ) может быть получен непосредственно из дискретных данных:

$$C_{m,k} = sd_k / 2^{m/2} . \tag{24.16}$$

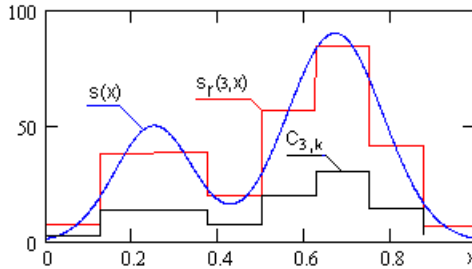


Рис. 24.27.

На следующем уровне разложения функции, при  $m=3$ , скейлинг-функция (24.11) расширяется по  $x$  в 2 раза (в нашем примере до  $1/8$ ), т.е. производится усреднение отсчетов по двум соседним интервалам исходной гистограммы. Количество коэффициентов соответственно в 2 раза уменьшается. Расчет коэффициентов  $C_{3,k}$  может выполняться непосредственно по (24.13), реконструкция  $s_{r3}(x)$  – по (24.14), при  $m=3$ ,  $N=8$ . Тем самым аппроксимация исходного сигнала выполняется на более "грубом" уровне декомпозиции, на основании чего скейлинг-функции вейвлетов называют аппроксимирующими или масштабными функциями, а сами коэффициенты, выделенные скейлинг-функциями – аппроксимирующими.

Но при известных значениях коэффициентов  $C_{4,k}$  предшествующего уровня следующий уровень может выполняться непосредственно по ним с учетом изменения нормировочного множителя в формуле скейлинг-функции (24.11). В общей форме:

$$C_{m-1,k} = (1/\sqrt{2}) (C_{m,2k} + C_{m,2k+1}). \tag{24.17}$$

$$C_{3,k} = \{2.665, 13.421, 13.706, 6.948, 20.037, 29.906, 14.538, 2.267\}.$$

Кроме аппроксимирующих коэффициентов  $C_{m-1,k}$  из предшествующей гистограммы аппроксимации  $C_{m,k}$  могут быть

выделены также коэффициенты изменения сигнала в пределах нового интервала усреднения, т.е. коэффициенты разности значений первой и второй половины интервала:

$$D_{m-1,k} = (1/\sqrt{2})(C_{m,2k} - C_{m,2k+1}), \quad (24.17)$$

которые называют детализирующими коэффициентами.

$$D_{3,k} = \{-1.571, -2.979, 2.769, -0.625, -5.024, 1.275, 4.853, 1.299\}.$$

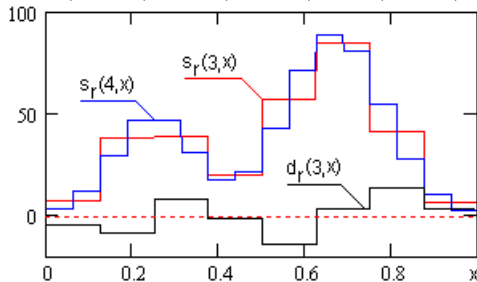


Рис. 24.28

На рис. 24.28 показан график  $d_r(3,x)$  детализирующих коэффициентов ( $m=3$ ), приведенный к масштабу исходного сигнала по формуле (24.14) при  $m=3$  и  $N=2^m=8$ , по которому нетрудно понять их физическую сущность. Так как значения сигнала в интервале разложения  $2\Delta t$  по  $m=3$  представляют собой среднее значение сигналов в двух интервалах  $\Delta t$  разложения по  $m=4$ , которые они перекрывают, а детализирующий коэффициент (с учетом приведения к масштабу исходного сигнала) равен половине разности сигналов этих двух интервалов, то его значение есть не что иное, как флюктуация сигнала по  $m=4$  относительно его аппроксимации по  $m=3$ . Если детализирующий коэффициент отрицателен, то эта флюктуация отрицательна относительно аппроксимированного значения в первой половине его интервала и положительна во второй, и наоборот. Т.е. соответствующие коэффициенты аппроксимации  $C_{m-1,k}$  и детализации  $D_{m-1,k}$  разделяют коэффициенты  $C_{m,k}$  предшествующего уровня декомпозиции сигнала на аппроксимированную (низкочастотную) и флюктуационную (высокочастотную) части.

Отсюда следует, что ряды коэффициенты  $C_{m-1,k}$  и  $D_{m-1,k}$  (количество точек  $2^{m-1}$  в каждом ряде) содержат полную информацию, адекватную информации в  $C_{m,k}$  предшествующего уровня (количество точек  $2^m=2^{m-1}+2^{m-1}$ ), что позволяют полностью восстановить значения коэффициентов более высокого уровня  $m$ :

$$C_{m,2k} = (1/\sqrt{2})(C_{m-1,k} + D_{m-1,k}), \quad C_{m,2k+1} = (1/\sqrt{2})(C_{m-1,k} - D_{m-1,k}), \quad (24.19)$$

а, следовательно, и восстановить исходный дискретный сигнал. Для восстановления значений в исходных интервалах при  $m=4$ , значение



аппроксимирующего коэффициента на первой половине интервала при  $m=3$  складывается с детализирующим коэффициентом, а на второй – вычитается. Для математического отображения этой операции введем функцию  $\psi$  (24.12), форма которой приведена на рис. 24.25, и обеспечим ее сдвиг по координате синхронно со скейлинг-функцией. Эта функция является ортонормированным базисом разложения детализирующих коэффициентов. Именно она и получила название вейвлета (вейвлетной или детализирующей функции). С ее использованием уравнение (24.14) с входящими в него уравнениями (24.19) приводятся к следующей форме (с уровня  $m=3$ ,  $2^{m-1}=8$ ):

$$s_r(3, x) = \sum_{k=0}^7 C_{3,k} \varphi_{3,k}(x) + \sum_{k=0}^7 D_{3,k} \psi_{3,k}(x). \quad (24.20)$$

Как и значения коэффициентов  $C_{m,k}$ , значения детализирующих коэффициентов могут вычисляться непосредственно по формуле (24.13) с заменой скейлинг-функции на вейвлет-функцию.

Аналогичным образом операция разделения на аппроксимирующие и детализирующие коэффициенты может быть продолжена над значениями коэффициентов  $C_{3,k}$  по уровню  $m=2$ , с выделением коэффициентов аппроксимации  $C_{2,k}$  и детализации  $D_{2,k}$ , и далее по уровням  $m=1$  и  $m=0$ . На последнем уровне  $m=0$  получаем только 1 коэффициент аппроксимации  $C_0$  и детализации  $D_0$  по всему интервалу задания сигнала  $0 \leq x \leq 1$ . Применяя последовательно, начиная с  $m=0$ , функцию "сборки" сигнала (24.20), получаем общую формулу реконструкции сигнала:

$$s_r(x) = C_0 \cdot \varphi_0(x) + D_0 \cdot \psi_0(x) + \sum_{k=0}^1 D_{1,k} \cdot \psi_{1,k}(x) + \sum_{k=0}^3 D_{2,k} \cdot \psi_{2,k}(x) + \sum_{k=0}^7 D_{3,k} \cdot \psi_{3,k}(x). \quad (24.21)$$

**Свойства преобразования.** Отметим на этом примере характерные и очевидные особенности нового представления сигнала:

- Общее количество коэффициентов разложения равно количеству отсчетов исходного сигнала (условие необходимости и достаточности сохранения в новом математическом представлении исходного объема информации).
- Вейвлет и его скейлинг-функция должны иметь однозначную связь. Это определяется тем, что разложение сигнала может быть выполнено с использованием только скейлинг-функции, а детализирующие коэффициенты определяться по разности  $m$  и  $m+1$  уровней аппроксимации, и наоборот.
- Значение  $C_0$  представляет собой среднее значение исходного сигнала по интервалу его задания. Для центрированных сигналов это

значение равно нулю. При выполнении разложения без скейлинг-функции (с вейвлетом в (24.12)) картина детальных особенностей нецентрированных сигналов остается без изменений, но полная реконструкция сигнала невозможна. Без значения  $C_0 \cdot \varphi_0(x)$  при полном разложении сигнал центрируется, при реконструкции с других масштабов декомпозиции искажается за счет отсутствия соответствующих коэффициентов  $C_{m,k}$ .

- Увеличение масштабного значения  $m$  разложения соответствует возрастанию временного разрешения сигнала ( $1/2^m$ ). Коэффициенты вейвлет-преобразования вскрывают флюктуационную структуру сигнала на разных масштабах и в разных временных точках. В областях "гладких" значений сигнала коэффициенты детализации близки к нулевым и ими можно пренебречь, что позволяет осуществлять сжатие информации для хранения.

- Реконструкция сигнала возможна с любого масштабного уровня декомпозиции, причем все особенности сигнала сохраняются без искажений с временным разрешением первого вейвлета (с минимальной шириной окна).

## **24.6. Математические основы кратномасштабного анализа**

Разложение сигнала на сумму аппроксимирующих и детализирующих составляющих производится с использованием ортогональных и биортогональных вейвлетов. На таких вейвлетах выполняется быстрое вейвлет-преобразование. При выполнении КМА пространство сигналов  $L^2(\mathbb{R})$  представляется в виде системы вложенных подпространств  $V_m$ , отличающихся друг от друга перемасштабированием независимой переменной.

**Исходные условия** ортогонального кратномасштабного анализа.

В качестве пространства сигналов будем рассматривать  $L^2(\mathbb{R})$  – пространство функций  $s(t)$  с конечной энергией. В этом пространстве определено скалярное произведение и норма функций:

$$\langle s(t), g(t) \rangle = \int_{-\infty}^{\infty} s(t) g^*(t), \quad \|s(t)\| = \sqrt{\langle s(t), s(t) \rangle}.$$

Базисом в пространстве  $V \subset L^2(\mathbb{R})$  называется такая система функций  $\{v_n(t)\}$ , что любая функция  $v(t) \in V$  единственным образом записывается в виде  $v(t) = \sum_n c_n v_n(t)$ . Базис называется ортонормированным, если  $\langle v_i(t), v_j(t) \rangle = \delta_{ij}$ . В этом случае  $c_n = \langle v(t), v_n(t) \rangle$ .

Под кратномасштабным анализом понимается описание

пространства  $L^2(\mathbb{R})$  через иерархические вложенные подпространства  $V_m \subset L^2(\mathbb{R})$ ,  $m = 0, \pm 1, \pm 2, \dots$ , которые не пересекаются и объединение которых в пределе дает  $L^2(\mathbb{R})$ . Система подпространств должна удовлетворять следующим условиям.

1. Условие вложенности:  $V_m \subset V_{m+1}$ .

Все пространство сигналов  $L^2(\mathbb{R})$  в целом может быть представлено в виде последовательности вложенных друг в друга замкнутых подпространств соответствующих уровней  $m$  декомпозиции сигнала:

$$\dots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \dots \subset V_m \subset V_{m+1} \dots$$

"Размеры" подпространств непрерывно расширяются по мере роста значения  $m$ , а объединение всех подпространств в пределе дает пространство  $L^2(\mathbb{R})$ .

2. Условие полноты и плотности разбиения:

$$\bigcup_{m \in I} V_m = L^2(\mathbb{R}). \quad (24.22)$$

3. Условие ортогональности подпространств:

$$\bigcap_{m \in I} V_m = \{0\}. \quad (24.23)$$

4. Условие сохранения в подпространстве при сдвигах функций:

$$v(t) \in V_m \Leftrightarrow v(t+1) \in V_m.$$

5. Для любой функции  $v(t) \in V_m$  ее масштабное преобразование по аргументу в 2 раза перемещает функцию в соседнее подпространство:

$$v(t) \in V_m \Leftrightarrow v(2t) \in V_{m+1}, \quad v(t) \in V_m \Leftrightarrow v(t/2) \in V_{m-1} \quad (24.24)$$

6. Для пространства  $V_0$  существует  $\phi$ -функция  $\phi(t) \in V_0$ , целочисленные сдвиги которой по аргументу образуют ортонормированный базис пространства  $V_0$ :

$$\phi_{0,k} = \phi(t-k), \quad k \in I \quad (k=0, \pm 1, \pm 2, \dots). \quad (24.25)$$

Функция  $\phi(t)$  называется скейлинг-функцией (scaling function).

Условие нормирования скейлинг-функции:

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

Из этих условий следует, что если подпространство  $V_0$  имеет ортонормированный базис  $\phi_{0,k}$ , то и все остальные подпространства также имеют ортонормированные базисы, которые образуются масштабным преобразованием базиса  $\phi_{0,k}$ :

$$\phi_{m,k}(t) = a^{m/2} \phi(a^m t - k), \quad m, k \in I. \quad (24.26)$$

Стандартное значение параметра 'a' в кратномасштабном анализе равно 2. Так, например, если  $\phi_{0,k}(t)=1$  на интервале  $[0,1)$  и

$\varphi_{0,k}(t)=0$  вне этого интервала, то целочисленные сдвиги этой функции попарно ортогональны, и пространство  $V_0$  состоит из функций, имеющих постоянные значения на интервалах вида  $[k, k+1)$ ,  $V_1$  – из функций, постоянных на интервалах  $[k/2, (k+1)/2)$ ,  $V_{-1}$  – из функций, постоянных на интервалах  $[2k, 2k+1)$ , и т.д.

Все условия в совокупности позволяют разложить произвольный сигнал  $s(t) \in L^2(\mathbb{R})$  по подпространствам  $V_m$ , т.е. на множество последовательных разномасштабных и ортогональных друг другу функций  $v_m(t) \in V_m$ , объединение которых дает исходный сигнал  $s(t)$ , или аппроксимирует сигнал с определенной точностью в зависимости от ограничения количества значений масштабирующего коэффициента  $m$  (и, соответственно, количества подпространств  $V_m$ ). Функции  $v_m(t)$  являются ортогональными проекциями сигнала  $s(t)$  на подпространства  $V_m$ . Отсюда появляется возможность анализа функции или сигнала на различных уровнях разрешения, или масштаба. Переменная  $m$  называется масштабным коэффициентом, или уровнем анализа. Если значение  $m$  мало, то функция  $v_m(t)$  есть грубая аппроксимация  $s(t)$ , в которой отсутствуют детали. При увеличении значений  $m$  точность аппроксимации повышается.

Кратность КМА, равную 2, в принципе, можно заменить любым целым числом, большим 1, но использование двоичной кратности оптимально и позволяет использовать быстрое вейвлет-преобразование.

**Масштабирующая функция.** Для того чтобы задать КМА, достаточно знать только одно из подпространств  $V_m$ , остальные определяются уравнением (24.26). Допустим, что это подпространство  $V_0$ , состоящее из сигналов, заданных "с разрешением 1". Тогда в пространстве  $V_m$  задаются сигналы с разрешением  $2^m$ , и оно отличается от  $V_0$  только перемасштабированием базисной функции в соответствии с (24.26). Так, если пространство  $V_0$  имеет скейлинг-функцию  $\varphi_0(t)$ , то соответствующее уравнение для скейлинг-функции  $\varphi_1(t)$  пространства  $V_1$  определяется выражением  $\sqrt{2} \varphi(2t-k)$ .

Поскольку  $V_0 \subset V_1$ , то функцию  $\varphi_0(t)$  можно представить линейной комбинацией сдвигов функции  $\varphi_1(t)$  (с учетом ее более компактного носителя) с определенными весовыми коэффициентами перемасштабирования  $h_k$ . Так, для скейлинг-функции Хаара, имеющей прямоугольное окно, каждая функция  $\varphi_{m-1}(t)$  образуется суммой двух последовательных функций  $\varphi_m(t)$  с соответствующим коэффициентом для сохранения единичной нормы. В общем случае, носитель функции может иметь произвольный размер с числом отсчетов  $2M$  (в единицах

k), при этом уравнение линейной связи базисных функций пространств, которое обычно называют функциональным уравнением масштабирования (уравнение рескейлинга), записывается в следующем виде:

$$\varphi(t/2) = \sqrt{2} \sum_{k=0}^{2M-1} h_k \varphi(t-k), \quad (24.27)$$

или, в эквивалентной форме:

$$\varphi(t) = \sqrt{2} \sum_{k=0}^{2M-1} h_k \varphi(2t-k). \quad (24.28)$$

Это уравнение называется масштабирующим. Решение этого уравнения и дает скейлинг-функцию, которую иногда называют "отцовским" вейвлетом.

Значения  $h_k$  определяются из условия для ортонормальных базисов:

$$h_k = \sqrt{2} \int_t \varphi(t) \varphi^*(2t-k) dt. \quad (24.29)$$

При дискретных значениях параметров сдвига масштабирующий вейвлет также дискретен и при задании функции  $\varphi(t)$  на конечном интервале имеет конечное число коэффициентов  $h_k$ , отличных от нуля. Условие нормировки масштабирующих коэффициентов:

$$\int_{-\infty}^{\infty} \varphi(t) dt = 1, \quad (24.30)$$

откуда следует:

$$\sum_k (h_k)^2 = 1. \quad (24.31)$$

Простейший и самый короткий вейвлет соответствует  $M=1$ . Так, для скейлинг-функции Наара при определении коэффициентов  $h_k$  уравнение рескейлинга (24.27) отображает растяжение вдвое исходного интервала более высокого уровня, число ненулевых коэффициентов  $h_k$  равно  $2M = 2$ , при этом, с учетом коэффициента нормировки  $\sqrt{2}$ , значения коэффициентов  $h_0 = h_1 = 1/\sqrt{2}$ . Подставляя значения коэффициентов в (24.27), получаем:

$$\varphi(t/2) = \varphi(t) + \varphi(t-1).$$

Решение этого функционального уравнения:

$$\varphi(t) = \theta(t) + \theta(1-t),$$

где  $\theta(t)$  – функция Хевисайда:  $\theta(t) = 1$  при  $t \geq 0$ ,  $\theta(t) = 0$  при  $t < 0$ .

Из совокупности исходных условий кратномасштабного

анализа и уравнения (24.26) следует, что перевод сигнала из пространства  $V_{m+1}$  с более высоким разрешением в пространство  $V_m$ , по существу, представляет собой нормированную децимацию сигнала - двукратное прореживание, с соответствующим уменьшением в 2 раза числа отсчетов сигнала. Это эквивалентно низкочастотной фильтрации сигналов  $v_{m+1}(k) \in V_{m+1}$  оператором  $h_k$  с частотой среза, равной половине частоты Найквиста сигналов  $v_{m+1}(k)$ , с автоматическим сокращением (за счет прореживания) главного частотного диапазона децимированного сигнала  $v_m(k)$  в 2 раза.

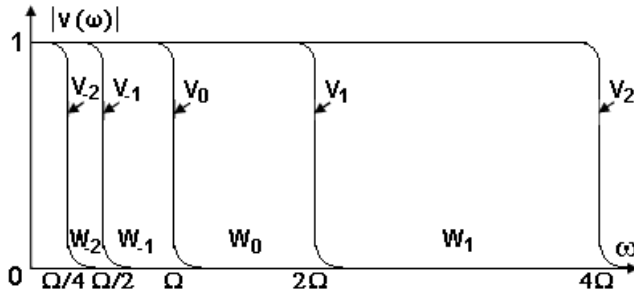


Рис. 24.29.

На рис. 24.29 приведено отображение такой операции в частотной области представления пространств  $V_m(\omega)$ , которое имеет вполне конкретный физический смысл разделения спектров сигналов (и самих сигналов при их восстановлении из спектров) на низкочастотную  $V_{m-1}(\omega)$  и высокочастотную  $W_{m-1}$  части.

Для исключения потерь высокочастотной информации, которая может потребоваться при восстановлении сигнала, она должна "переводиться" и сохраняться в новые подпространства  $W_m$ , ортогональные подпространствам  $V_m$ , такие, что

$$V_{m+1} = V_m \oplus W_m. \tag{24.32}$$

Подпространства  $W_m$  называются *детализирующими* в том смысле, что именно они содержат ту дополнительную информацию (не пересекающуюся с пространством  $V_m$ ), необходимую для повышения уровня разрешения сигнала с  $V_m$  на  $V_{m+1}$  при его восстановлении.

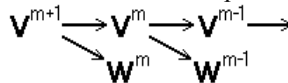


Рис. 24.30.

Этот процесс нагляднее рассматривать в обратном порядке, как это показано на рис. 24.30. Так как наиболее детальный уровень разрешения сигнала соответствует максимальным значениям  $m$  (при  $m \Rightarrow \infty$  сигнал становится непрерывным), то разложение сигнала по

подпространствам начинается с максимальных значений  $m$ . На каждом цикле разделения, при переходе из пространства  $V_{m+1}$  в пространство  $V_m$ , от пространства  $V_{m+1}$  отделяется подпространство  $W_m$ , в которое отфильтровывается высокочастотная информация пространства  $V_{m+1}$ , а остающаяся информация более "грубого" разрешения перемещается в пространство  $V_m$  и представляет собой аппроксимацию данных пространства  $V_{m+1}$ . В пределе, с учетом свойства ортогональности пространств:

$$V_{m+1} = \bigoplus_{m=-\infty}^m W_m, \quad \bigoplus_{m=-\infty}^{\infty} W_m = L^2(\mathbb{R}). \quad (24.32)$$

Таким образом, физический смысл процесса разложения пространств достаточно прост. Исходное пространство ( $V_{m+1}$  на рисунке) является пространством сигналов и функций с определенным частотным диапазоном. При разложении сигнала в пространство  $W_m$  отделяются высокочастотные составляющие пространства  $V_{m+1}$ , а в пространстве  $V_m$  остаются его низкочастотные составляющие.

**Базисный вейвлет.** Детализирующие подпространства  $W_m$  в совокупности также образуют взаимно ортогональный набор и имеют свой ортонормальный базис при любом заданном значении  $m$ . Если скейлинг-функция установлена, то базисная функция детализирующего пространства, которую называют вейвлетом ("материнским"), должна иметь определенную связь со скейлинг-функцией.

Уравнение (24.27), по существу, представляет собой нормированное уравнение свертки, где  $h_k$  представляет собой оператор низкочастотного фильтра:

$$\varphi(t/2) = \sqrt{2} h_k \otimes \varphi(t-k). \quad (24.33)$$

При переходе в частотную область:

$$\varphi(2\omega) = H(\omega) \varphi(\omega). \quad (24.34)$$

Уравнение (24.34) является масштабирующим уравнением в частотной области и полностью определяется  $2\pi$  периодической функцией  $H(\omega)$ . Отсюда следует, что в пространстве  $W_m$  должна сохраняться высокочастотная часть информации сигнала, что может выполняться квадратурным обращением низкочастотного фильтра:

$$G(\omega) = -\exp(-j\omega) H^*(\omega+\pi). \quad (24.35)$$

Фурье-образ искомого вейвлета:

$$\Psi(2\omega) = G(\omega) \varphi(\omega). \quad (24.36)$$

При переходе во временную область:

$$\psi(t/2) = \sqrt{2} g_k \otimes \varphi(t-k) = \sqrt{2} \sum_{k=0}^{2M-1} g_k \varphi(t-k). \quad (24.37)$$

При этом значения  $g_k$  могут быть вычислены и

непосредственно во временной области соответствующим квадратурным обращением оператора  $h_k$ :

$$g_k = (-1)^k h_{2^m-1-k} \equiv (-1)^k (h_k)_{\text{rev}}. \quad (24.38)$$

где  $(h_k)_{\text{rev}}$  – реверсированный массив оператора  $h_k$ , записанный в обратном порядке. Соответственно, для вейвлета Хаара эти коэффициенты равны:  $g_0 = 1/\sqrt{2}$ ,  $g_1 = -1/\sqrt{2}$ . Именно этот вейвлет и известен, как вейвлет Хаара (рис. 24.25). В функциональном анализе он применяется с 1910 года. Масштабированные и смещенные версии скейлинг-функции и вейвлета:

$$\begin{aligned} \varphi_{m,k} &= 2^{m/2} \varphi(2^m t - k). \\ \psi_{m,k} &= 2^{m/2} \psi(2^m t - k). \end{aligned}$$

Вейвлет Хаара знакопеременен, при этом

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

Условие знакопеременности является общим условием для всех "материнских" вейвлетов, которое обеспечивает безусловную устойчивость базиса при восстановлении исходного сигнала.

**Разложение функций на вейвлетные ряды** на заданном уровне разрешения  $m$  выполняется по формуле:

$$s(t) = \sum_k C_{m,k} \varphi_{m,k} + \sum_{m,k} D_{m,k} \psi_{m,k}. \quad (24.39)$$

Значения коэффициентов (которые обычно называют суммами и разностями):

$$C_{m,k} = \int_t s(t) \varphi_{m,k}(t) dt. \quad (24.40)$$

$$D_{m,k} = \int_t s(t) \psi_{m,k}(t) dt. \quad (24.41)$$

На практике значения коэффициентов определяются с помощью быстрого вейвлет-преобразования, которое будет рассмотрено ниже.

Первая сумма в (24.39) содержит усредненные (с весовыми функциями  $\varphi_{m,k}$ ) значения функции  $s(t)$  по диадным интервалам  $[k \cdot 2^{-m}, (k+1) \cdot 2^{-m}]$ , вторая – значения флюктуаций на данных интервалах. По мере возрастания значения  $m$  длина интервалов уменьшается и уровень детализации (разрешения) функции  $s(t)$  увеличивается. На самом детальном уровне  $m = m_{\text{max}} = M$  ряд представлен только скейлинг-функцией и, в пределах точности разложения, практически совпадает с исходной функцией:

$$s(t) = \sum_k C_{M,k} \varphi_{M,k}.$$



На низшем уровне разрешения (на наиболее широких интервалах) первая сумма ряда (24.39) содержит всего одно усредненное взвешенное значение сигнала, а вторая сумма показывает флуктуации на всех без исключения уровнях. Числовой ряд на каждом из уровней является "истинным" представлением сигнала с тем же объемом информации, но только в другом (вейвлетном) математическом представлении. Вейвлет-преобразованием сигнала очень часто называют полную комбинацию рядов только второй суммы (24.39), дающей представление о локальных особенностях и флуктуациях сигнала на всех уровнях разрешения, которые обычно и являются предметом изучения.

Таким образом, выражение (24.39) показывает возможность аппроксимации любой произвольной функции  $s(t)$  набором простых локальных функций  $\varphi_{m,k}(t)$  и  $\psi_{m,k}(t)$ , ортогональных на разных уровнях значений  $m$  и полностью покрывающих пространство  $L^2(\mathbb{R})$  за счет смещений  $k$ . Первая сумма выражения (24.39) дает "сглаженные средние" значения функции  $s(t)$  на разных масштабных уровнях, вторая сумма вейвлетных функций добавляет к "грубой" аппроксимации сигнала все более подробные детали на все меньших масштабных интервалах.

**Вычисление вейвлетных рядов.** Допустим, что сигнал  $s(t)$  известен с разрешением 1, т.е. задан числовым рядом  $s_n \equiv C_{m,n}$  в пространстве  $V_0$  с масштабом  $m=0$ , который можно рассматривать, как результат его разложения по сдвигам скейлинг-функции:

$$s(t) = \sum_n s_n \varphi(t-n).$$

Версия масштаба  $m-1$  ортогональной проекции  $s(t)$  на подпространство  $V_{-1}$  будет задаваться набором скалярных произведений  $s(t)$  с функциями из базиса  $V_{-1}$ :

$$C_{m-1,n} = \langle s(t), (1/\sqrt{2})\varphi(t/2-n) \rangle.$$

Из уравнения (24.27) и условий ортогональности следует:

$$C_{m-1,n} = \sum_s h_s C_{m,2n-s}.$$

Другими словами, вычисление вейвлетных коэффициентов аппроксимации масштаба  $m-1$  может осуществляться путем свертки коэффициентов предыдущего масштаба ( $m$ ) с низкочастотным фильтром  $h_s$  и прореживания вдвое, которое «встроено» в эту формулу через индекс  $2n-s$ .

В качестве деталей сигнала  $s(t)$ , исчезающих при переходе к новому масштабу  $m-1$ , следует взять компоненту  $s(t)$ , ортогональную к сигналам масштаба  $m-1$  пространства  $V_{-1}$ , и спроецировать ее на базис пространства детализирующих коэффициентов  $W_{-1}$ :

$$D_{m-1,n} = \langle s(t), (1/\sqrt{2})\psi(t/2-n) \rangle.$$

Или, с использованием уравнения (24.37):

$$D_{m-1,n} = \sum_s g_s C_{m,2n-s}.$$

Т.е. действует та же схема свертки аппроксимационных коэффициентов предыдущего масштаба с оператором высокочастотного фильтра и прореживания вдвое.

Схемы последовательного масштабного разложения сигналов действуют на любых масштабах. Эту процедуру вычисления вейвлетных рядов называют быстрым вейвлет-преобразованием (БВП) или алгоритмом Малла по фамилии его автора.

### **24.7. Быстрое вейвлет-преобразование**

**Принцип преобразования.** Любую функцию  $s(t)$  можно рассматривать на любом  $m'$  - уровне разрешения. Для разделения функции на этом уровне между ее усредненными значениями и флюктуациями вокруг средних значений преобразуем формулу 24.421 к следующему виду:

$$s(t) = \sum_{k=-\infty}^{\infty} C_{m',k} \varphi_{m',k}(t) + \sum_{m=m'}^{\infty} \sum_{k=-\infty}^{\infty} D_{m,k} \psi_{m,k}(t). \quad (24.42)$$

При бесконечных пределах первая сумма в этом выражении стремится к нулю и может быть опущена, давая "чистое" вейвлет-преобразование. В общем случае коэффициенты  $C_{m,k}$  и  $D_{m,k}$  можно вычислять непосредственно по формулам (24.40) и (24.41). На практике мы обычно имеем дело с цифровыми данными в виде конечного набора отсчетов, а, соответственно, наилучший уровень разрешения определен интервалом, содержащим один отсчет, и суммирование выполняется в конечных пределах. Значение  $m=0$  обычно принимается для этого наилучшего уровня разрешения. Для принятой нами формы вейвлетов  $\varphi_{m,k} = 2^{m/2} \varphi(2^m t - k)$  усреднение отсчетов (расширение размеров вейвлетов) происходит при уменьшении значений  $m$ , т.е. при  $m = 0, -1, -2, \dots$  Для исключения использования отрицательных индексов масштабирования знак "минус" обычно вводится непосредственно в функции вейвлетов, т.е.  $\varphi_{m,k} = 2^{-m/2} \varphi(2^{-m} t - k)$ , при этом вейвлет-коэффициенты вычисляются для  $m > 0$ .

**Алгоритм Малла.** Кратномасштабный анализ при последовательном увеличении значений  $m$  приводит к естественной форме быстрых итерационных вычислений:

$$C_{m+1,k} = \sum_n h_n C_{m,2k+n}, \quad (24.43)$$

$$D_{m+1,k} = \sum_n g_n C_{m,2k+n}, \quad (24.44)$$

$$C_{0,k} = \int_{k\Delta t}^{(k+1)\Delta t} s(t) \varphi(t-k) dt. \quad (24.45)$$

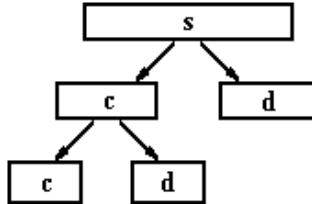


Рис. 24.31.

Уравнения обеспечивают пирамидальный алгоритм вычисления вейвлет-коэффициентов (алгоритм Малла), приведенный на рис. 24.31. Явный вид вейвлета требуется только для расчета коэффициентов  $h_n$  и  $g_n$ , при самом преобразовании используются значения коэффициентов  $h_n$  и  $g_n$ . Уравнение (24.45) применяется при известной аналитической форме функции  $s(t)$ . Для цифровых данных в качестве значений  $C_{0,k}$  принимаются исходные значения данных, т.е.  $C_{0,k} = s(k)$ .

Сущность операций, выполняемых формулами (24.43) и (24.44), заключается в следующем. На первом этапе преобразования цифровой фильтр  $h_n$  из сигнала  $s_k = C_{0,k}$  выделяет низкие частоты  $|\omega| \leq \pi/2$ , а октавный фильтр  $g_n$  выделяет верхние частоты  $\pi/2 \leq |\omega| \leq \pi$ . Поскольку на выходе фильтра  $h_n$  отсутствует верхняя половина частот, то частота дискретизации выходного сигнала может быть уменьшена в 2 раза, т.е. выполнена децимация выходного сигнала, что производится в формуле (24.43) сдвигами  $(2k+n)$  через 2 отсчета по входному сигналу. На выходе фильтра  $g_n$  освобождается место в области низких частот, и аналогичное прореживание выходного сигнала приводит к транспонированию верхних частот на освободившееся место. Таким образом, каждый из выходных сигналов несет информацию о своей половине частот, при этом выходная информация представлена таким же количеством отсчетов, что и входная.

**Реконструкция сигналов.** Поскольку в формулах (24.43, 24.44) вместо базисных функций используются фильтры, то обратные преобразования, т.е. последовательную сборку сигнала от больших  $m$  к малым и реконструкцию сигналов по значениям его вейвлет-коэффициентов с любого уровня разрешения, имеет смысл также выразить через фильтры реконструкции:

$$C_{m-1} = \sum_{n \in I} C_{m,n} h_{k-2n}^r + \sum_{n \in I} D_{m,n} g_{k-2n}^r, \quad (24.46)$$

Алгоритм вычислений по (24.46) обратен алгоритму декомпозиции, т.е. представляет собой аппроксимацию коэффициентов  $C_m$  и  $D_m$  на новый, в 2 раза меньший, шаг дискретизации с двукратным увеличением частоты Найквиста и восстановлением спектра коэффициентов  $C_m$  в низкочастотную часть нового главного диапазона спектра  $C_{m-1}$ , а спектра коэффициентов  $D_m$  в высокочастотную часть спектра  $C_{m-1}$ . Это выполняется расстановкой нулевых значений между коэффициентами  $C_m$  и  $D_m$  (увеличение в 2 раза числа отсчетов), фильтрацией полученных массивов низкочастотным  $h^r(k)$  и высокочастотным  $g^r(k)$  фильтрами реконструкции, и сложением результатов фильтрации. Модули частотных характеристик фильтров  $h^r(k)$  и  $g^r(k)$  должны повторять модули частотных характеристик фильтров  $h(k)$  и  $g(k)$ . Но фильтры декомпозиции  $h(k)$  и  $g(k)$  являются односторонними и фазосдвигающими, и при реконструкции коэффициентов  $C_{m-1}$  этот сдвиг фазы должен ликвидироваться. Последнее достигается реверсом значений коэффициентов фильтров декомпозиции, т.е.:

$$h^r(k) = \text{reverse}(h(k)), \quad g^r(k) = \text{reverse}(g(k)). \quad (24.47)$$

Точность реконструкции сигналов зависит от потерь информации при выполнении прореживания спектров, причем эти потери наблюдаются на срезах полос пропускания фильтров низких и высоких частот, крутизна которых зависит от порядка фильтров, их согласованности, и типа вейвлетных функций.

Обязательным условием преобразования сигнала является его задание количеством точек (отсчетов), равном  $N=2^m$ , где значение  $m \geq 1$  определяет максимально возможное число уровней декомпозиции сигнала при целочисленных значениях кратности сдвигов операторов фильтров количеству отсчетов вейвлетных коэффициентов на каждом уровне декомпозиции. Для выполнения этого условия количество отсчетов сигнала, как правило, дополняется до ближайшего большего значения  $N$  методами, известными из практики задания начальных/конечных условий свертки (нулями, конечными значениями сигналов, четными или нечетными значениями относительно конечных отсчетов, периодическим продолжением и т.п.). Может применяться также передискретизация исходного сигнала до необходимого количества отсчетов.

**Пакетные вейвлеты.** Основная информация обычно заключена в низкочастотной части сигнала, разложение которой может быть продолжено вплоть до нулевого уровня. Но аналогичная операция может применяться и к любой высокочастотной части разложения. Это

соответствует замене вейвлета  $\psi(t)$  на два новых вейвлета

$$\psi_1(t) = \sum_n h_n \psi(t-n), \quad \psi_2(t) = \sum_n g_n \psi(t-n),$$

которые тоже локализованы в пространстве, но на вдвое более широком интервале, чем исходный вейвлет. Бинарное дерево разложения (рис. 24.31) "расщепляется" и для коэффициентов 'D' любого уровня. Такое расщепление является адаптивным и легко приспосабливается к индивидуальным особенностям сигналов. Функции адаптивного преобразования называют вейвлет-пакетом.

### 24.8. Фильтры дуальной декомпозиции и реконструкции сигналов

Рассмотренные математические основы дуального вейвлет-преобразования показывают, что основную роль в реализации вейвлетных преобразований играют низкочастотные и высокочастотные фильтры декомпозиции и реконструкции сигналов.

**Идеальные фильтры.** Преобразование Фурье произвольной числовой последовательности  $\{s_k\}$  является  $2\pi$ -периодической функцией и определяется числовыми значениями на главном частотном диапазоне  $[-\pi, \pi]$ . При этом полагается, что шаг дискретизации данных  $\Delta t=1$ , а частота Найквиста сигнала  $s_k$  равна  $\omega_N = \pi/\Delta t = \pi$ . Передаточная функция  $H(\omega)$  низкочастотного фильтра  $h_n$ , удовлетворяющего условию (24.29), концентрируется в интервале  $[-\pi/2, \pi/2]$ . При разделении сигнала на два частотных поддиапазона с полным сохранением исходной информации должно выполняться условие (рис. 24.32):

$$H(\omega) + G(\omega) = 1. \tag{24.28}$$

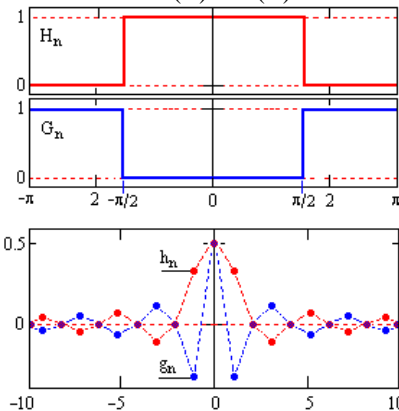


Рис. 24.32.

Отсюда следует, что передаточная функция  $G(\omega)$  высокочастотного фильтра  $g_n$ , сосредоточена на  $[-\pi, -\pi/2]$  и  $[\pi/2, \pi]$ . Соответственно, идеальные фильтры  $H(\omega)$  и  $G(\omega)$  в пределах главного частотного диапазона задаются выражениями:

$$H(\omega) = \begin{cases} 1, & \omega \in [-\pi/2, \pi/2] \\ 0, & \omega \notin [-\pi/2, \pi/2] \end{cases}, \quad G(\omega) = \begin{cases} 0, & \omega \in [-\pi/2, \pi/2] \\ 1, & \omega \notin [-\pi/2, \pi/2] \end{cases}. \quad (24.49)$$

Коэффициенты фильтров (обратное преобразование Фурье, рис. 24.32):

$$h_0 = 0.5; \quad h_{2k} = 0, k \neq 0; \quad h_{2k+1} = (-1)^k / (\pi(2k+1)); \quad k=0, \pm 1, \pm 2, \pm 3, \dots \quad (24.50)$$

$$g_0 = 0.5; \quad g_{2k} = 0, k \neq 0; \quad g_{2k+1} = (-1)^{k+1} / (\pi(2k+1)); \quad k=0, \pm 1, \pm 2, \pm 3, \dots \quad (24.51)$$

Связь значений коэффициентов:

$$g_n = (-1)^n h_n. \quad (24.52)$$

Разложение сигнала  $s(k)$  на низкочастотную и высокочастотную части в спектральной и временной области:

$$S(\omega) = H(\omega)S(\omega) + G(\omega)S(\omega) = S_h(\omega) + S_g(\omega). \quad (24.53)$$

$$s(k) = h(n) \otimes s(k-n) + g_g(n) \otimes s(k-n) = s_h(k) + s_g(k). \quad (24.54)$$

Так как носитель функции  $S_h(\omega)$  находится на интервале  $[-\pi/2, \pi/2]$ , то  $S_h(\omega)$  можно разложить в ряд Фурье, как  $\pi$ -периодическую функцию с частотой Найквиста  $\pi/2$ :

$$S_{h2\downarrow}(\omega) \leftrightarrow 2 \sum_k s_h(2k) \exp(-j 2k\omega), \quad \omega \in [-\pi/2, \pi/2], \quad (24.55)$$

т.е. функция  $s_h(k)$  избыточна по количеству отсчетов и может быть децимирована. Соответственно, не требуется вычислять свертку по всем нечетным значениям  $s(k)$ . Двукратная децимация обычно обозначается индексом  $2\downarrow$ :

$$s_{h2\downarrow}(m) = s_h(2k) = h(n) \otimes s(2k-n) = C(m),$$

где  $m$  – последовательная нумерация четных отсчетов  $s_h(2k)$  ( $m = \text{int}(k/2)$ ).

Учитывая периодичность частотных функций, спектр  $S_g(\omega)$  можно рассматривать на интервале  $[0, 2\pi]$ , где ненулевые отсчеты  $S_g(\omega)$  находятся на интервале  $[\pi/2, 3\pi/2]$ . При аналогичном разложении  $S_g(\omega)$  в этом интервале, как  $\pi$ -периодической функции:

$$S_{g2\downarrow}(\omega) \leftrightarrow 2 \sum_k s_g(2k) \exp(-j 2k\omega), \quad \omega \in [\pi/2, 3\pi/2], \quad (24.56)$$

$$s_{g2\downarrow}(m) = s_g(2k) = g(n) \otimes s(2k-n) = D(m).$$

Для обратного преобразования спектров  $S_{h2\downarrow}(\omega)$  и  $S_{g2\downarrow}(\omega)$  в спектры  $S_h(\omega)$  и  $S_g(\omega)$  главный диапазон спектров нужно увеличить в 2 раза дополнением нулями. Во временной области эта операция может быть выполнена передискретизацией значений  $s_h(2k)$  и  $s_g(2k)$  с шага

$2\Delta k$  на шаг  $\Delta k$  рядом Котельникова-Шеннона. Альтернативная более быстрая операция - обратная децимация массивов  $C(m) \rightarrow C_{2\uparrow}(k)$  и  $D(m) \rightarrow D_{2\uparrow}(k)$ , которая выполняется дополнением массивов нулями между всеми отсчетами (обозначается индексом  $2\uparrow$ ), с последующей фильтрацией фильтрами  $2h(n)$  и  $2g(n)$ :

$$s_h(k) = 2h(n) \otimes C_{2\uparrow}(k-n), \quad s_g(k) = 2h(n) \otimes D_{2\uparrow}(k-n), \quad (24.57)$$

что обеспечивает восстановление исходного сигнала:

$$s(k) = s_h(k) + s_g(k). \quad (24.58)$$

**Реальные фильтры.** Операторы идеальных фильтров, заданные в частотной области прямоугольными импульсами (24.49), имеют бесконечные импульсные характеристики (24.50, 24.51) и убывают достаточно медленно. При усечении таких операторов на частотной характеристике проявляется явление Гиббса, что увеличивает погрешности декомпозиции и реконструкции сигналов. С практической точки зрения целесообразно использовать фильтры с плавным переходом от полосы пропускания в полосу подавления, которые имеют конечное число ненулевых коэффициентов. При задании таких низкочастотных  $H(\omega)$  и высокочастотных  $G(\omega)$  фильтров, удовлетворяющих условию (24.48), разложение сигнала с децимацией остается без изменений. Добавляя к операторам фильтров декомпозиции индекс  $d$ , получаем:

$$\begin{aligned} s(k) &\rightarrow s_h(k) = h_d(n) \otimes s(k-n) \rightarrow 2\downarrow \rightarrow C(m) = s_h(2k), \\ s(k) &\rightarrow s_g(k) = g_d(n) \otimes s(k-n) \rightarrow 2\downarrow \rightarrow D(m) = s_g(2k). \end{aligned} \quad (24.59)$$

Однако точное восстановление сигналов по формулам (24.57, 24.58) возможно только для взаимно ортогональных фильтров. Для ограниченных перекрывающихся по спектру фильтров для постановления сигналов необходимы парные к ним фильтры реконструкции, компенсирующие возможные искажения восстановления. Для упрощения выражений для числовых рядов перейдем в  $z$ -область представления сигналов.

$$H_d(z) = \sum_n h_d(n) z^n, \quad G_d(z) = \sum_n g_d(n) z^n, \quad S(z) = \sum_n s_d(n) z^n,$$

где  $z = \exp(-j\omega)$  – комплексная переменная.

Отфильтрованные низкочастотный и высокочастотный сигналы:

$$C_d(z) = H_d(z) S(z), \quad D_d(z) = G_d(z) S(z). \quad (24.60)$$

Децимация сигналов в  $z$ -области выполняется простыми выражениями:

$$C(z^2) = 0.5 (C_d(z) + C_d(-z)), \quad D(z^2) = 0.5 (D_d(z) + D_d(-z)). \quad (24.61)$$

Обозначим фильтры реконструкции сигнала индексами  $r$ . Уравнение реконструкции:

$$S(z) = 2[H_r(z) C(z^2) + G_r(z) D(z^2)]. \quad (24.62)$$

Подставляя в это выражение функции (24.61) и (24.60), получаем:

$$S(z) = [H_r(z)H_d(z)+G_r(z)G_d(z)]S(z) + [H_r(z)H_d(-z)+G_r(z)G_d(-z)]S(-z). \quad (24.63)$$

Отсюда следует, что искомые фильтры должны удовлетворять системе уравнений:

$$\begin{aligned} H_r(z)H_d(z)+G_r(z)G_d(z) &= 1, \\ H_r(z)H_d(-z)+G_r(z)G_d(-z) &= 0. \end{aligned} \quad (24.64)$$

Решение системы существует, если определитель отличен от нуля всюду на единичной окружности  $z=\exp(-j\omega)$ :

$$H_d(z)C_d(-z) - H_d(-z)G_d(z) \neq 0.$$

## 24.9. Ортогональные и биортогональные вейвлеты

**Коэффициенты вейвлета.** Значения коэффициентов  $h_k$  и  $g_k$  в рамках КМА определяются на основании общих свойств скейлинг-функций и вейвлетов. Уравнения функций:

$$\varphi(t) = \sqrt{2} \sum_{k \in I} h_k \varphi(2t-k), \quad \psi(t) = \sqrt{2} \sum_{k \in I} g_k \varphi(2t-k).$$

Из свойства ортогональности масштабных функций следует первое уравнение на значения коэффициентов  $h_k$ :

$$\int_t \varphi(t) \varphi(t-x) dt = \delta_x, \quad \sum_k h_k h_{k+2x} = \delta_x. \quad (24.65)$$

Из условий нормировки скейлинг-функции следует второе уравнение:

$$\int_t \varphi(t) dt = 1. \quad \sum_k h_k = \sqrt{2}. \quad (24.66)$$

Из ортогональности вейвлетных и масштабных функций следует уравнение, решением которого являются значения коэффициентов  $g_k$ :

$$\int_t \psi(t) \varphi(t-n) dt = 0. \quad \sum_k h_k g_{k+2x} = 0. \quad (24.67)$$

$$g_k = (-1)^k h_{2M-1-k}. \quad (24.68)$$

Точность и масштабная разрешающая способность аппроксимации анализируемых функций вейвлетами зависит от их гладкости, т.е. от порядка дифференцируемости. При использовании вейвлет-преобразования для сжатия информации (отбрасыванием малозначимых коэффициентов разложения) вейвлеты с высокой гладкостью обеспечивают более точную реконструкцию сигналов. Для обеспечения знакопеременности и заданной гладкости до степени  $M-1$  вейвлеты должны быть ортогональны полиномам соответствующих



степеней:

$$\int_t t^m \psi(t) dt = 0, \quad m = 0, 1, \dots, M-1, \quad \sum_k k^m g_k = 0. \quad (24.69)$$

$$\sum_k (-1)^k k^m h_k = 0. \quad (24.70)$$

**Пример расчета** коэффициентов выполним при  $M=2$ .

Запишем уравнения (24.65, 66 и 70) в явном виде:

$$\begin{aligned} h_0 h_2 + h_1 h_3 &= 0, & h_0 + h_1 + h_2 + h_3 &= \sqrt{2}, \\ h_0 - h_1 + h_2 - h_3 &= 0, & -h_1 + 2h_2 - 3h_3 &= 0. \end{aligned}$$

Решение этой системы уравнений:

$$h_0 = 2^{-3/2} - h_3, \quad h_1 = 2^{-1/2} - h_3, \quad (24.71)$$

$$h_2 = 2^{-3/2} + h_3, \quad h_3 = 2^{-5/2} (1 \pm \sqrt{3}).$$

Примем для коэффициента  $h_3$  знак минус в скобках, при этом:

$$h_0 = 2^{-5/2} (1 + \sqrt{3}) = 0.483, \quad h_1 = 2^{-5/2} (3 + \sqrt{3}) = 0.837,$$

$$h_2 = 2^{-5/2} (3 - \sqrt{3}) = 0.224, \quad h_3 = 2^{-5/2} (1 - \sqrt{3}) = -0.129.$$

Соответственно, значения коэффициентов  $g_k$ , вычисленные по (24.68):

$$g_0 = -0.129, \quad g_1 = -0.224, \quad g_2 = 0.837, \quad g_3 = -0.483.$$

Спектры коэффициентов  $h_k$  и  $g_k$  приведены на рис. 24.33.

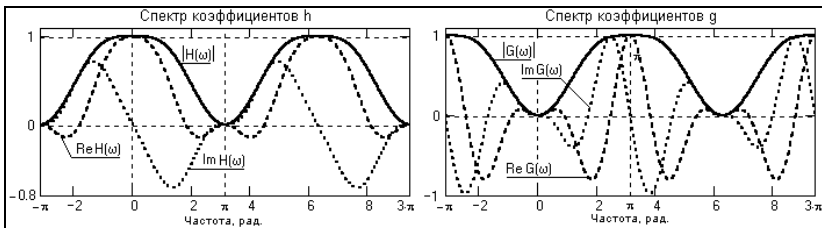


Рис. 24.33.

Из рисунка 24.33 можно видеть, что спектры коэффициенты  $h_k$  и  $g_k$  представляют собой передаточные функции односторонних согласованных цифровых фильтров, низкочастотного и высокочастотного соответственно.

Выбор знака в (24.71) для  $h_3$  несколько изменяет форму вейвлета, особенно для вейвлетов более высокого порядка. При повышении порядка гладкость вейвлета повышается и, как правило, несколько увеличивается область его определения  $(2M-1)$ .

**Вейвлет Добеши.** Существуют семейства ортогональных вейвлетов, которые вообще не имеют аналитического выражения и определяются только фильтрами. К ним относятся вейвлеты Добеши. Скейлинг-функции и вейвлеты Добеши – это непрерывные функции, не тождественные нулю и недифференцируемые на конечном отрезке.

Вычисленные выше при  $M=2$  коэффициенты  $h_k$  и  $g_k$  определяют простейший вейвлет второго порядка  $D_2$  из семейства ортонормальных вейвлетов Добеши, приведенный на рис. 24.34.

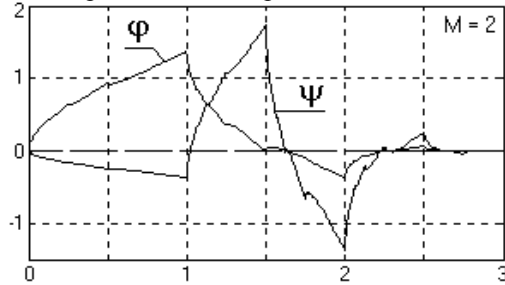


Рис. 24.34.

Область задания вейвлетов Добеши шире, чем вейвлетов Хаара, но при этом они обеспечивают при вейвлет-преобразовании большее количество малозначимых коэффициентов разложения и, при отбрасывании последних, более сильное сжатие данных.

Вид скейлинг-функции и вейвлета Добеши вычисляются методом итераций (можно посмотреть в Matlab), а в практических приложениях используются только вейвлет-коэффициенты  $h_k$  и  $g_k$  без вычисления конкретной формы вейвлетов.

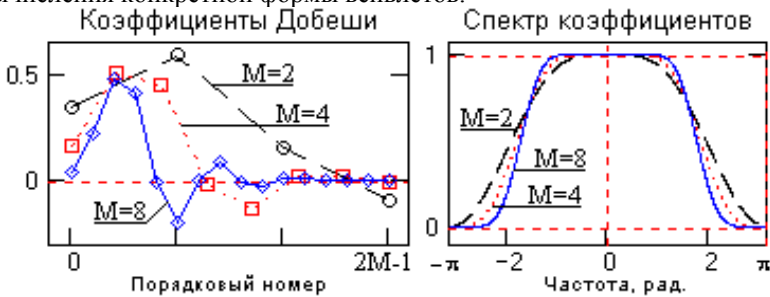


Рис. 24.35.

На рис. 24.35 приведены три вектора коэффициентов Добеши из системы Matlab (db2, db4 и db8) и спектры коэффициентов. Как видно из рисунка, по мере увеличения порядка векторов крутизна среза их частотных характеристик увеличивается, а, соответственно, качество разложения сигналов и их реконструкции также будут улучшаться.

За исключением вейвлетов Хаара, все вещественные ортогональные вейвлеты асимметричны. Наиболее близкую к симметричной форму имеет вейвлет Коифлетса.

**Биортогональные вейвлеты.** Требование

ортогональности вейвлетного базиса является сильным ограничением. Например, невозможно построить ортонормированный базис для вейвлетов с гладким симметричным носителем. Для обеспечения симметрии при точной реконструкции сигналов применяются биортогональные вейвлеты. Для вейвлетов  $\psi_{m,k}(t)$  с базисом Рисса можно определить дуальный вейвлет  $\psi^{\#}_{m,k}(t)$ , который образует биортогональную пару с вейвлетом  $\psi_{m,k}(t)$ , удовлетворяющую требованию ортогональности их скалярного произведения:

$$\langle \psi_{m,k}(t), \psi^{\#}_{m,k}(t) \rangle = \delta_{m,k}.$$

При использовании биортогональной пары декомпозиция сигналов может производиться вейвлетом  $\psi_{m,k}(t)$ , а реконструкция парным вейвлетом  $\psi^{\#}_{m,k}(t)$ , или наоборот. Разложение функций с биортогональными вейвлетами может производиться в двух эквивалентных формах:

$$s(t) = \sum_{m,k} \langle s(t), \psi^{\#}_{m,k}(t) \rangle \psi_{m,k}. \quad (24.72)$$

$$s(t) = \sum_{m,k} \langle s(t), \psi_{m,k}(t) \rangle \psi^{\#}_{m,k}. \quad (24.73)$$

Свойства регулярности биортогональных вейвлетов могут заметно отличаться. Если один из них обладает гладкостью порядка  $n$ , то дуальный ему вейвлет может иметь, по крайней мере,  $n$  нулевых моментов. Большое число нулевых моментов дает хорошие результаты при сжатии информации, а большая степень гладкости вейвлета обеспечивает более точную реконструкцию сигналов. При этом оба вейвлета можно выполнить симметричными.

Ортогональное преобразование можно рассматривать, как частный случай биортогонального при  $\psi_{m,k}(t) = \psi^{\#}_{m,k}(t)$ . Для биортогонального преобразования действительны все свойства ортогонального.

ЛИТЕРАТУРА

ОСНОВНАЯ

1. Баскаков С.И. Радиотехнические цепи и сигналы: Учебник для вузов. - М.: Высшая школа, 1988.
2. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. - М.: Мир, 1989. - 540 с.
3. Васильев Д.В. Радиотехнические цепи и сигналы: Учебное пособие для вузов. - М.: Радио и связь, 1982. - 528 с.
4. Гольденберг Л.М. и др. Цифровая обработка сигналов: Справочник. - М.: Радио и связь, 1985.
5. Гольденберг Л.М. и др. Цифровая обработка сигналов: Учебное пособие для вузов.- М.: Радио и связь, 1990.- 256 с.
6. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. - М.: Мир, 1988. - 488 с.
7. Дмитриев В.И. Прикладная теория информации: Учебник для вузов. - М.: Высшая школа, 1989.
8. Зиновьев А.Л., Филиппов Л.И. Введение в теорию сигналов и цепей: Учебное пособие для вузов. - М.: Высшая школа, 1975. - 264 с.
9. Игнатов В.А. Теория информации и передачи сигналов. - М.: Советское радио, 1979.
10. Купер Дж., Макгиллем А. Вероятностные методы анализа сигналов и систем. - М.: Мир, 1989.
11. Лосев А.К. Линейные радиотехнические цепи: Учебник для вузов. - М.: Высшая школа, 1971.
12. Макс Ж. Методы и техника обработки сигналов при физических измерениях: В 2-х томах.- М.: Мир, 1983.
13. Никитин А.А. Теоретические основы обработки геофизической информации: Учебник для вузов. - М.: Недра, 1986.- 342 с.
14. Оппенгейм А.В., Шафер Р.В. Цифровая обработка сигналов. - М.: Связь, 1979. - 416 с.
15. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. - М.: Мир, 1982. - 428 с.
16. Рапопорт М.Б. Вычислительная техника в полевой геофизике: Учебник для вузов. - М.: Недра, 1993.- 350 с.
17. Сиберт У.М. Цепи, сигналы, системы. - М.: Мир, 1988. - 336 с.
18. Хемминг Р.В. Цифровые фильтры. - М.: Недра, 1987. 221с.
19. Сергиенко А.Б. Цифровая обработка сигналов. / Учебник для вузов. - СПб.: Питер, 203. - 608 с.
20. Вероятностные методы в вычислительной технике:

Учебное пособие для вузов./ А.В.Крайников и др. - М.: Высшая школа, 1986. - 312 с.

21. Гурский Е.И. Теория вероятностей с элементами математической статистики: Учебное пособие для вузов. - М.: Высшая школа, 1971.- 328 с.

22. Игнатов В.А. Теория информации и передачи сигналов. - М.: Советское радио, 1979.

23. Колесник В.Д., Полтырев Г.Ш. Курс теории информации. - М.: Наука, 1982. - 416 с.

24. Сато Ю. Обработка сигналов. Первое знакомство. - Изд.: ДОДЭКА, 2002.

25. Харкевич А.А. Борьба с помехами. - М.: Наука, 1965.

26. Корн Г., Корн Е. Справочник по математике для научных работников и инженеров. - М.: Наука, 1984.

27. Айфичер Э., Джервис Б. Цифровая обработка сигналов. Практический подход. / М., "Вильямс", 2004, 992 с.

28. Дремин И.Л. и др. Вейвлеты и их использование. / Успехи физических наук, 2001, т.171, № 5, стр. 465-501.

29. Дьяконов В., Абраменкова И. MATLAB. Обработка сигналов и изображений. Специальный справочник. - СПб.: Питер, 2002, 608 с.

30. Петухов А.П. Введение в теорию базисов всплесков. - СПб.: Изд. СПбГТУ, 1999, 132 с.

31. Смоленцев Н.К. Основы теории вейвлетов. Вейвлеты в Matlab. М.: LVR Пресс, 2005. - 304 с.

32. Переберин А.В. О систематизации вейвлет-преобразований. - Вычислительные методы и программирование, 2001, т.2.

33. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. - СПб, ВУС, 1999. 204 с.

### **Дополнительная**

34. *Андронов И. С, Финк Л. М.* Передача дискретных сообщений по параллельным каналам.— М. : Сов. радио, 1971.— 360 с.

35. *Аппаратура* передачи дискретной информации «МС-5»/ Под ред. Ю. Б. ОКВ-нева.— М. : Связь, 1970.— 240 с.

36. *Блох Э. Л., Попов О. В., Турин В. Я.* Модели источника ошибок в каналах передачи цифровой информации.— М. : Связь, 1971.— 312 с.

37. *Бомштейн Б. Д., Киселев Л. К., Моргачев Е. Т.* Методы борьбы с помехами в каналах проводной связи.— М. : Связь, 1975.— 284 с.

38. *Бонгард М. М.* О понятии «полезная информация».— В сб.: Проблемы кибернетики. Вып. 9. М., Физматгиз, 1963.
39. *Ван Трис Г. Л.* Теория обнаружения, оценок и модуляции.— М. : Сов. радио, 1972,— Т. 1. 744 с.
40. *Варакин Л. Е.* Теория систем сигналов.— М. : Сов. радио, 1978.— 304 с.
41. *Возенкрафт Дж., Джекобе И.* Теоретические основы техники связи. М. : Мир, 1969.— 640 с.
42. *Грекова И.* К вопросу об информации.— Наука и жизнь, 1967, № 3.
43. *Гренандер У.* Случайные процессы и статистические выводы.— М. ; Изд-во иностр. лит., 1961.— 168 с.
44. *Гуров В. С, Емельянов Г. А., Етрухин Н. И.* Передача дискретной информации и телеграфия.— М. : Связь, 1974.— 560 с.
45. *Зюко А. Г.* Элементы передачи информации.— Киев : Техніка, 1969.— 282 с.
46. *Каневский З. М.* Передача сообщений с обратной связью.— М. : Знание, 1969.— 120 с.
47. *Карнап Р.* Значение необходимости.— М. : Изд-во иностр. лит, 1959.— 280 с.
48. *Катков Ф. А., Дидык Б. С, Стулов В. А.* Телемеханика.— Киев : Вища школа, 1974.— 248 с.
49. *Кеннеди Р.* Каналы связи с замиранием и рассеянием.— М.: Сов. радио, 1973.— 320 с.
50. *Клюев Н. И.* Информационные основы передачи сообщений.— М. : Сов. радио, 1966.— 328 с.
51. *Колмогоров А. И.* Три подхода к определению понятия «количество информации».— В кн.: Проблемы передачи информации, 1965, т. 1, вып. 1.
52. *Котельников В. А.* Теория потенциальной помехоустойчивости.— М. : Гос-энергоиздат, 1956.— 152 с.
53. *Левенштейн В. И.* О границах вероятностей необнаружения ошибки.— Проблемы передачи информации, 1977, т. 13.
54. *Леман Е.* Проверка статистических гипотез.— М. : Наука, 1966.— 410 с.
55. *Окунев Ю. Б.* Система связи с инвариантными характеристиками помехоустойчивости.— М. : Связь, 1973.— 80 с.
56. *Передача информации с обратной связью / З. М. Каневский и др.*— М. Связь, 1976.— 352 с.
57. *Пирс Дж.* Символы, сигналы, шумы.— М. : Мир, 1967.— 320 с.

58. *Пугачев В. С.* Введение в теорию вероятностей.— М. : Наука, 1968.— 248 с.
59. *Сикарев А. А., Фалько А. И.* Оптимальный прием дискретных сообщений.— М. : Связь, 1978.— 328 с.
60. *Стратонович Р. Л.* Условные и марковские процессы и их применение в теории оптимального управления.— М. : Изд-во МГУ, 1966.— 168 с.
61. *Стратонович Р. Л.* Принципы адаптивного приема.— М. : Сов. радио, 1973 — 144 с.
62. *Сосулин Ю. Г.* Теория обнаружения и оценивания стохастических сигналов.— М. : Сов. радио, 1978.— 320 с.
63. *Тихонов В. И. Кульман Н. К.* Нелинейная фильтрация и квазигогерентный прием сигналов.— М.; Сов. радио, 1975.— 704 с.
64. *Тимошук Л. Н.* Информационные носители, их характеристики в области применения.— М.; Знание, 1967, № 1.
65. *Харкевич А. А.* О ценности информации.— В кн.: Проблемы кибернетики. М. : Физматгиз, 1960, вып. 4.
66. *Хармут Х.* Теория секвентного анализа. Основы и применения.— М. : Мир, 1980.—576 с.
67. *Хелстром К.* Статистическая теория обнаружения сигналов.— М. : Изд-во иностр. лит. 1963.—432 с.
68. *Цымбал В. П.* Устройства дискретной техники на логических тиратронах.— Киев : Техшка, 1969.— 128 с.
69. *Шляпоберсий В. И.* Основы техники передачи дискретных сообщений.— М. : Связь. 1973.—480 с.
70. *Элементы теории передачи дискретной информации* / Под ред. Л. П. Пурто-ва.—М. : Связь, 1977.—328 с.
71. *Экономическая информация.*— М.: Статистика, 1974.—348с.
72. *Элементы теории передачи дискретной информации.*—М. : Связь, 1972,—232 с.
73. *Энциклопедия кибернетики*, т. 1, 2.— Киев, 1974.— 1226 с.
74. *Яглом А. М., Яглом И. М.* Вероятность и информация.— М. : Наука, 1973.—512 с.
75. *Ясин Е. Г.* К проблеме измерения количества, содержательности и ценности информации.— В кн.: Экономическая семистика. М. : Наука, 1970.

Научно-практическое издание

**Кононюк Анатолий Ефимович**

# **Информациология**

## **Общая теория информации**

*Книга 4*

Авторская редакция

Подписано в печать 31.03.2011 г.

Формат 60x84/16.

Усл. печ. л. 16,5. Тираж 300 экз.

**Издатель и изготовитель:**

Издательство «Освита Украины»

04214, г. Киев, ул. Героев Днепра, 63, к. 40

Свидетельство о внесении в Государственный реестр  
издателей ДК №1957 от 23.04.2009 г.

Тел./факс (044) 411-4397; 237-5992

E-mail: osvita2005@ukr.net, [www.rambook.ru](http://www.rambook.ru)

**Издательство «Освита Украины» приглашает**  
авторов к сотрудничеству по выпуску изданий,  
касающихся вопросов управления, модернизации,  
инновационных процессов, технологий, методических  
и методологических аспектов образования  
и учебного процесса в высших учебных заведениях.

Предоставляем все виды издательских  
и полиграфических услуг.