

высшее

образование

В.Т. Першин

ОСНОВЫ СОВРЕМЕННОЙ РАДИОЭЛЕКТРОНИКИ



УДК 621.396
ББК 32.884я73
КТК 222
П27

Рецензент:

главный научный сотрудник
НИЧ Белорусского национального
технического университета доктор технических наук,
доцент *О. Г. Девойно.*

Першин В. Т.

П27 Основы современной радиоэлектроники : учебное пособие / В. Т. Першин. — Ростов н/Д : Феникс, 2009. — 541, [1] с. : ил. — (Высшее образование).

ISBN 978-5-222-14681-1

Учебное пособие отражает основные идеи курса «Основы современной радиоэлектроники». Рассматривает понятия энтропии и информации, сигналы и их анализ, источники информации и каналы связи, согласование источника информации с каналом связи, взаимодействие открытых систем, сигналы физического уровня, помехоустойчивое кодирование, системы связи с уплотнением и разделением каналов, основные виды импульсной модуляции и манипуляции, многопозиционные дискретные сигналы, основные протоколы взаимодействия радиоэлектронных устройств, цифровое радиовещание, принципы построения цифровых сетей сотовой подвижной связи, беспроводные технологии передачи информации. К наиболее сложным в методическом отношении вопросам даны примеры, снабженные подробными решениями. Ряд примеров выполнен с использованием пакета MATHCAD. Приведен перечень тем практических занятий, реализуемых средствами пакета MATLAB.

ISBN 978-5-222-14681-1

УДК 621.396
ББК 32.884я73

© Першин В. Т., 2009
© Оформление: ООО «Феникс», 2009

Начав свое развитие более 100 лет назад, радиоэлектроника в настоящее время охватывает широкую научно-техническую область знаний, в которой разрабатываются методы и устройства для получения, хранения, преобразования и передачи информации на расстояние как с помощью излучения электромагнитных волн, так и без него. Спектр проблем, в той или иной мере относящихся к радиоэлектронике, чрезвычайно широк. Тем не менее, в качестве составных частей радиоэлектроники можно выделить радиотехнику, электронику и микроэлектронику, а теперь еще и нанoeлектронику.

При этом важнейшими задачами радиоэлектроники в области радиотехники, электроники и микро(нано)электроники соответственно являются:

- разработка методов и устройств получения, преобразования и передачи информации на расстояние с помощью электромагнитного излучения, распространяющегося в пространстве, разделяющем передатчик и приемник, и без него;
- изучение вопросов взаимодействия электронов с электромагнитными полями и методов создания на их основе приборов и устройств;
- создание электронных функциональных устройств в микроминиатюрном исполнении на основе группового изготовления радиоэлементов и печатного монтажа.

Идеи современной радиоэлектроники проникают во многие области, которые раньше имели лишь отдаленное отношение к вопросам передачи и обработки информации. Однако в настоящее время разработаны новые концепции открытых систем и иерархий взаимодействия между их компонентами. Ввиду стремительного развития средств электронно-вычислительной техники и информационных технологий многие вопросы передачи информации между абонентами невозможно решить без хорошего знания и использования радиоэлектронных методов. То же можно сказать и о развитии интеллектуальных систем.

Классификация периодов развития радиоэлектроники

По мнению автора этого учебного пособия, весь период развития радиоэлектроники можно разделить на две части. Первая часть относится к периоду от изобретения радио до середины 40-х годов прошлого века, а вторая часть является продолжением первой вплоть до наших дней.

Основанием такой классификации является наполнение к середине 40-х годов прошлого столетия основных понятий энтропии и информации новым содержанием, связанным с осознанием того факта, что носитель информации — *сигнал* — имеет в общем случае случайную природу. При вероятностном подходе информация рассматривается как сообщение об исходе случайного события, и количество информации ставится в зависимость от априорных вероятностей этого события.

Понятие информации, которое невозможно определить, но можно описать конструкциями, имеющими различные общность и область применения, связывается с неопределенностью выбора среди возможных вариантов сообщений. Чтобы измерить количество информации исходя из понятия неопределенности, человек ставит себя на место приемника действительной или возможной информации. С точки зрения описания понятия неопределенности информация является абстракцией, исключающей связь со смысловым ее содержанием (семантикой).

Для современного состояния радиоэлектроники характерно объединение абсолютно всех средств и методов вычислительной техники с проблемами передачи информации, успешно решаемыми на основе развития возможностей микро- и нанoeлектроники, обеспечивающими в последнее время чрезвычайно высокую степень интеграции в производстве функциональных радиоэлектронных устройств.

Изложенное представляется автору главным тезисом предлагаемой им классификации периодов развития радиоэлектроники, в соответствии с которой автор и рассматривает эти два периода соответственно как *классический* и *современный*. Изложению идей современной радиоэлектроники и посвящено это учебное пособие.

Развитие радиоэлектроники послевоенного времени

Подлинный взрыв развития современной радиоэлектроники начался в послевоенные годы с появлением классических работ К. Шеннона и Н. Винера. Работы Н. Винера были тесно связаны с исследованиями военного времени по автоматическому управлению огнем, а труды К. Шеннона «Математическая теория связи» и «Связь при наличии шума» — с исследованиями по шифрованию сообщений и их передаче по секретным каналам связи. Математические модели К. Шеннона и Н. Винера сильно различаются: сигнал, по Н. Винеру, мог обрабатываться после воздействия на него шума, по К. Шеннону — сигнал мог обрабатываться как до, так и после передачи по каналу связи с шумами.

В 1946 г. В.А. Котельников защитил докторскую диссертацию «Теория потенциальной помехоустойчивости при флуктуационных помехах», в которой описал построенную им математическую модель идеального приемника, лучше которого по помехоустойчивости сделать невозможно. В.А. Котельников исследовал воздействие флуктуационных шумов, искажающих сигнал, и показал, что полностью исключить действие помех можно только в двух случаях: при бесконечной мощности сигнала и при мощности помех, стремящейся к нулю. Во всех остальных случаях избавиться от помех невозможно, но при передаче сигналов ДА-НЕТ вероятность ошибки определяется только отношением энергии сигнала к энергии помех при выбранном способе их передачи. В этой работе В.А. Котельников дал алгоритм работы идеального приемника: если сообщение ДА передается сигналом $A(t)$, а НЕТ — сигналом $B(t)$, то энергии этих сигналов должны быть равны, а по своей форме эти сигналы должны быть противоположны, т. е. $A(t) + B(t) = 0$. Эти разнополярные сигналы соответствуют сигналам, использующимся при фазовой манипуляции со сдвигом фаз на 180° .

В 1948 г. американский ученый К. Шеннон в монографии «Математическая теория связи», основываясь на фундаментальном понятии количественной меры неопределенности — *энтропии* — и связанного с ним количества информации, подошел к проблеме передачи информации с другой стороны.

В.А. Котельников изучал передачу сигналов без избыточности. Рассматриваемый им сигнал не содержал никаких дополнительных элементов, кроме необходимых для переноса информации. К. Шеннон построил математическую модель не только для идеального приемника В.А. Котельникова, но и для всей системы передачи информации. К.Шеннон предложил использовать сигналы, лучше которых нет. Это сигналы с избыточностью, в которых, кроме, например, посылок, несущих информацию, передаются посылки, не несущие информации. Эти дополнительные посылки вводятся по определенным правилам и позволяют во время приема снизить или даже исключить ошибки, вызванные действием помех. Этот метод защиты информации от разрушительного действия помех называется *избыточным кодированием*. Конечно, введение дополнительных элементов в структуру сигнала должно уменьшить скорость передачи информации по каналу связи. Более того, при высокой эффективности избыточного кодирования скорость передачи может уменьшиться в несколько раз.

Таким образом, результат В.А. Котельникова относится к высокоскоростным системам безыбыточного кодирования, а результат К. Шеннона — к системам с избыточным кодированием, передающим информацию с меньшей скоростью. К. Шеннон показал, что если система связи, в которой действует нормальный шум, занимает полосу частот Δf , а отношение мощностей сигнала и помех суть P_c/P_n , то максимальное количество C информации, которое можно передать в единицу времени со сколь угодно малой вероятностью ошибок равно

$$C = \Delta f \log(1 + P_c/P_n).$$

Основы помехоустойчивого кодирования

В 1948 г. К. Шеннон заложил основы теории кодирования, сформулировав два важных результата:

- для случая канала без помех можно осуществить кодирование дискретных сообщений так, чтобы среднее количество двоичных знаков на элемент исходного алфавита было как угодно близко, но не менее величины энтропии источника информации (при этом энтропия опре-

деляется статистическими свойствами источника сообщений, и такое кодирование называется *статистическим*, или *эффективным*);

- для канала связи с шумом существует такой способ кодирования конечного количества информации, при котором информация будет передана с какой угодно высокой достоверностью, если только скорость поступления ее не превышает *пропускную способность канала связи*.

Важно отметить, что эти две теоремы К. Шеннона унавливают только существование оптимальных или близких к ним кодов, но ничего не говорят о том, как эти коды можно построить. Исследования в этом направлении продолжаются и в настоящее время. А.Н. Колмогоров, немало сделавший для развития математического фундамента теории информации, справедливо отметил, что «строгое математическое обоснование своих идей в сколько-нибудь трудных случаях К. Шеннон предоставил своим продолжателям».

Решение поставленных в работах К. Шеннона задач было продолжено трудами А.Я. Хинчина, И.М. Гельфанда, А.Н. Колмогорова, А.М. Яглома, Л.М.Финка и других ученых советской школы. В частности, А.Я. Хинчин доказал ряд теорем теории информации для дискретных сообщений, А.М. Яглом рассмотрел более общий случай непрерывных систем, лишь слегка затронутый в работах К. Шеннона. Советский ученый Финк Л.М. предложил в 1955 г. сверточные коды, которые снова были открыты Хегельбергером в 1959 г. и теперь называются кодами Хегельбергера.

Выдвинутый А.Н. Колмогоровым тезис о том, что наряду с вероятностным подходом к определению количества информации правомерны, а часто и более естественны другие подходы. Так, комбинаторный и алгебраический методы описания информации привели к созданию алгоритмической теории информации.

Новые подходы к созданию современной теории информации

Алгоритмический подход дает возможность определить сложность описания различных объектов. По А.Н. Колмогорову, «относительной сложностью» объекта u при заданном x

считается минимальная длина $l(p)$ «программы» p получения y из x . Это определение зависит от «метода программирования», который представляет собой функцию $y = \varphi(p, x)$, описывающую соответствие объекта y программе p и объекту x .

До сих пор появляются частные подходы определения информации на семантическом и прагматическом уровнях ее рассмотрения, которые часто оказываются неконструктивными вследствие того, что смысл и практическая ценность передаваемого сообщения зависит от свойств потребителя — *человека*, его способности правильно воспринять и использовать полученную информацию.

Проблема сжатия информации

Следующей проблемой, непосредственно примыкающей к кодированию, является проблема сжатия информации, которая представляет собой кодирование источника с целью получения минимального объема генерируемых им сообщений. Сжатие информации приводит к уменьшению числа символов в сообщении до минимума, необходимого для представляемой информации, заключенной в сообщении.

Разработаны два класса методов сжатия информации:

- необратимого сжатия, называемого *энтропийным*, или *сжатием с потерями*;
- обратимого сжатия, т. е. полностью восстанавливаемого преобразования исходной информации.

Обратимое сжатие берет свое начало от работ Р. Хартли, К. Шеннона, Р. Файнштейна, Р. Фано. Наиболее интересные результаты в области обратимого сжатия случайных последовательностей с известным распределением вероятностей, получившими название *оптимального статистического кодирования*, получены в работах Д. Хаффмана, Р. Крафта, Л.Д. Девиссона.

Методы необратимого сжатия основаны на работах по способам восстановления непрерывного сигнала по дискретным отсчетам и работам по вычислению ε -энтропии класса функций, связанных с исследованиями В.А. Котельникова, А.Н. Колмогорова, А.А. Харкевича, Р.Л. Добрушина и других ученых.

Б.М. Фитингоф, Р.Е. Кричевский, Ю.М. Штарьков и другие ученые рассмотрели случайные последовательности с неизвестными заранее распределениями вероятностей и раз-

работали оптимальные для бесконечных последовательностей методы сжатия на основе наблюдения статистики первых символов последовательности. Это направление называют сейчас *оптимальным универсальным кодированием*.

Французский ученый А. Хоквингем (1959 г.) и американцы Р.К. Боуз и Д.К. Рой-Чоудхури (1960 г.) нашли большой класс кодов, обеспечивающий произвольное минимальное кодовое расстояние $d_{\min} \geq 5$. Они получили название БЧХ (Боуза-Чоудхури-Хоквингема).

Каскадный метод построения кодов впервые был предложен в 1966 г. Г.Д. Форни, затем, в начале 70-х годов, теория каскадных кодов была развита В.В. Зябловым, Э.Л. Блохом, В.А. Зиновьевым.

Современные стандарты сжатия информации

Хорошие результаты можно получить только для определенных источников на основе точного знания его статистических свойств. Примером является алгоритм сжатия неподвижных изображений, разработанный в 1990 г. Объединенной группой экспертов в области фотографии (США) — *Joint Photographic Expert Group*, JPEG. Основу алгоритма составляет известная математическая операция, называемая «дискретным косинусным преобразованием» с применением на последующих этапах кодирования по методу Хаффмана. Этот метод позволяет достичь очень высоких коэффициентов сжатия. Например, файл объемом в 14 Мбайт со сканера можно сжать примерно до 1,2 Мбайт. Этот алгоритм предусматривает потери, т.е. восстановленное после сжатия изображение будет отличаться от исходного.

Дальнейшее развитие идеи сжатия цифровых подвижных изображений получили в разработке стандартов MPEG (*Motion Picture Expert Group*).

Наиболее широко используются стандарты MPEG-2 и MPEG-4. Находится в стадии завершения разработка стандарта MPEG-7.

Кодирование речевых и аудиосигналов

Весьма важное место в технике связи занимает передача речи. Ученые и инженеры стремятся представить речь в

возможно более компактной форме. Появился термин «ком- пандирование речевых сигналов», т. е. КОМпрессия/ЭКСПАНДИРОВАНИЕ, что равносильно термину «устройство сжатия и расширения для изменения диапазона громкости».

В настоящее время разработаны алгоритмы сжатия информации без потерь, такие как Лемпеля-Зива, LZ77, RLE (*Run Length Encoding*), *Huffman Encoding*, а также с потерями JPEG, М-JPEG, MPEG.

Компандирование для оцифровки речевого сигнала описано в Рекомендациях МККТТ (Международный консультативный комитет по телефонии и телеграфии, или *Consultative Committee on International Telephony and Telegraphy* — CCITT) G.711 и G.712 для А- и Мю-законов, для которых предусмотрена частота дискретизации 8 кГц с погрешностью ± 50 отсчетов на 1 млн. и числом двоичных разрядов на отсчет, равным 8, что обеспечивает скорость передачи 64 Кбит/с. МККТТ входил до 28.02.93 г. в Международный союз электросвязи (*The International Telecommunication Union* — ITU), являвшийся органом ООН. Вместо него был создан *The Telecommunication Standardization Sector* (ITU-T) — постоянный орган, ответственный за изучение технических, рабочих и тарифных вопросов и разработку Рекомендаций для них с точки зрения стандартизации телекоммуникаций на международном уровне. *The World Telecommunication Standardization Conference* (WTSC), которая собирается каждые 4 года, устанавливает темы для изучения ITU-T *Study Groups*, которые, в свою очередь, разрабатывают Рекомендации по этим темам. Например, Рекомендации G.701 пересмотрены были ITU-T *Study Groups* XV в 1988–1993 г.г. и одобрены WTSC (Хельсинки, 11–12 марта, 1993 г.).

В конце 30-х годов XX века была осуществлена идея воссоздания речи путем генерирования основного тона и шума, которые возбуждали фильтр с передаточной функцией, соответствующей огибающей спектра и изменяющейся во времени со скоростью произнесения звуков речи. Впервые Дадли (*Dudley*) разработал устройство, которое преобразовывало речь в аналоговый сигнал с полосой частот шириной 300 Гц для передачи по каналу связи, т.е. сжатию речи в 10 раз. Это устройство было названо вокодером (VOice — голос, CODER — кодировщик) и давало синтезированную речь низкого каче-

ства. За прошедшие 60 лет вокодеры стали пригодными для коммерческих целей.

Современные вокодеры обеспечивают хорошее качество речи при скорости передачи порядка 800–2400 *бит/с* и качество речи, пригодное для ведения служебных переговоров при скорости передачи 1200 *бит/с*. Заметим, однако, что вокодеры используют в основном не для уплотнения каналов связи, а для их защиты от прослушивания, так как только преобразовав речь в цифровую форму и применив *шифрование*, можно гарантированно защититься от перехвата передаваемой информации.

Кодирование речевых и аудиосигналов имеют много общих черт, так как прежде всего основываются на свойствах их восприятия человеческой аудиторией. Речь можно кодировать очень эффективно, так как имеется много моделей создания речевых сигналов, в то время как ничего подобного не существует для аудиосигналов. Скрампное уменьшение скорости битов аудиосигналов получается путем мгновенного сжатия, т. е. преобразования однородной 14-битовой импульсно-кодовой модуляции (ИКМ) в неоднородную 11-битовую ИКМ или путем использования *Near Instantaneously Compounded Audio Multiplex* (NICAM) кодирования [ITU-R. Rec. 660]. Например, BBC использует NICAM-728 форматы кодирования для цифровой передачи звука в нескольких европейских вещательных телевизионных сетях. По сравнению с речевым кодированием аудиокodирование предполагает более высокие скорости дискретизации, лучшее разрешение по амплитуде, более широкий динамический диапазон, большие вариации в ширине плотности мощности спектра, представлении аудиосигнала стереофоническим или многоканальным и, наконец, более высокое ожидаемое качество для слушателей. Действительно, высокое качество CD с 16-битовой дискретизацией в формате ИКМ делают цифровые аудио популярными.

Построение реляционных баз данных

Реляционные системы берут свое начало в математической теории множеств. Доктор из фирмы IBM Э.Ф. Кодд предложил их в конце 1968 года, поскольку он первым осознал, что

можно использовать математику для придания надежной основы и строгости области управления базами данных.

Нечеткость многих терминов, используемых в сфере обработки данных, заставила Кодда отказаться от них и придумать новые или дать более точные определения существующим. Так, он не мог использовать широко распространенный термин «запись», который в различных ситуациях может означать экземпляр записи, либо тип записей, запись в стиле языка КОБОЛ (которая допускает повторяющиеся группы) или плоскую запись (которая их не допускает), логическую запись или физическую запись, хранимую запись или виртуальную запись и т. д. Вместо этого он использовал термин «кортеж длины n » или просто «кортеж», которому дал точное определение. В литературе можно подробно познакомиться с терминологией реляционных баз данных.

Разработка, в основном, шла в отделениях фирмы IBM (языки ISBL, SQL, QBE) и университетах США (PIQUE, QUEL). Последний создавался для СУБД INGRES (*Interactive Graphics and Retrieval System*), которая была разработана в начале 70-х годов в Университете шт. Калифорния и сегодня входит в пятерку лучших профессиональных СУБД. Сегодня из всех этих языков полностью сохранились и развиваются QBE (*Query-By-Example* — запрос по образцу) и SQL, а из остальных взяты в расширение внутренних языков СУБД только наиболее интересные конструкции. В начале 80-х годов SQL «победил» другие языки запросов и стал фактическим стандартом таких языков для профессиональных реляционных СУБД. В 1987 году он стал международным стандартом языка баз данных и начал внедряться во все распространенные СУБД персональных компьютеров. Почему же это произошло?

Непрерывный рост быстродействия, а также снижение энергопотребления, размеров и стоимости компьютеров привели к резкому расширению возможных рынков их сбыта, круга пользователей, разнообразия типов и цен. Естественно, что расширился спрос на разнообразное программное обеспечение.

Борясь за покупателя, фирмы, производящие программное обеспечение, стали выпускать на рынок все более и более интеллектуальные и, следовательно, объемные программные комплексы. Приобретая такие комплексы, многие организации и отдельные пользователи часто не могли разместить их

на собственных ЭВМ, однако не хотели и отказываться от нового сервиса. Для обмена информацией и ее обобществления были созданы сети ЭВМ, где обобществляемые программы и данные стали размещать на специальных обслуживающих устройствах — файловых серверах.

Завершение разработки теории преобразований Фурье

На 60-е годы XX века приходится завершение формирования семейства преобразований Фурье, начавшееся с опубликования статьи J.W. Cooley и J.W. Tukey о быстром преобразовании Фурье (БПФ). БПФ оказалось результатом переработки выражения для дискретного преобразования Фурье (ДПФ) и требовало только $(N/2)\log_2 N$ умножений комплексных чисел, где N представляло число рассчитываемых точек преобразования. Эффективность БПФ по сравнению с ДПФ становится весьма существенной, когда количество точек БПФ увеличивается до нескольких тысяч. Очевидно, что БПФ вычисляет все компоненты выходного спектра, точнее или все, или ни одного. Если необходимо рассчитать только несколько точек спектра ДПФ, то может оказаться, что в этом случае ДПФ будет эффективнее, чем БПФ. Вычисление, однако, выходного отсчета спектра с использованием ДПФ требует только N умножений комплексных чисел.

Реализация методов помехоустойчивого кодирования

Огромную роль в разработке и реализации CDMA сыграла американская компания *Qualcomm*, организованная в 1985 г. известными американскими учеными Э. Витерби и И. Джекобсом. Накопленный опыт позволил создать первые коммерческие спутниковые системы для сопровождения грузоперевозок — *Omnitracs* и *Euteltracs*. Следующий важный шаг *Qualcomm* (1991 г.) — разработка проекта стандарта CDMA для сотовых систем общего пользования. После двухлетней доработки базовую версию стандарта CDMA утвердила Ассоциация производителей оборудования связи TIA (США). Реальное воплощение технологии (согласно международной

классификации стандартов — IS-95) началось в сентябре 1995 г., когда в Гонконге была развернута система на базе CDMA — cdmaOne.20.

Наиболее горячими поклонниками CDMA стали азиатские страны — Япония, Китай, Малайзия и особенно Южная Корея, где построена крупнейшая в мире сеть сотовой связи на 2,6 млн абонентов. В России первая сеть фиксированной связи на базе оборудования компании *Qualcomm* была развернута АО «Связьинформ» Челябинской области в конце 1996 г., а ее коммерческая эксплуатация осуществляется с сентября 1997 г.

Цифровое радиовещание и телевидение

Наиболее совершенной из разработанных к настоящему времени систем цифрового звукового радиовещания (ЦЗРВ) (*Digital Audio Broadcasting*), которые относятся к первой категории, является «Эврика-147/ DAB». Она принята Европейским Радиовещательным Союзом (EBU) в качестве общеевропейской и рекомендована для внедрения во всем мире Межсоюзной технической комиссией всемирной конференции радиовещательных союзов (*Inter-Union Technical Committee of the World Conference of Broadcasting Unions*). Ее приняли не только государства Европы, но и Канада, Китай, Индия, Австралия и другие. Это открывает возможность беспрепятственного обмена радиопрограммами и информацией на международном уровне.

Ко второй категории можно отнести системы ЦЗРВ, разработанные в США. Эти системы предназначены для работы в УКВ диапазоне 87,5–108 МГц и СВ диапазоне 525–1608 кГц одновременно с существующими аналоговыми АМ- и ЧМ-радиостанциями.

Аналоговое телевизионное вещание появилось так давно, что смесь телевизионных служб и технических инфраструктур в каждой стране образует уникальное сочетание. Частотный диапазон, в котором ведется наземное вещание, близок к насыщению. Например, в странах Европы используется более 70 тысяч телевизионных передатчиков, многие из которых работают в метровом диапазоне. На частоте одного канала может работать до тысячи передатчиков.

Тем не менее, наземное вещание нельзя ликвидировать, заменив его кабельным и спутниковым.

Наземное телевидение — единственный способ доставки телевизионных программ к портативным переносным приемникам. Наземное цифровое телевидение может оказаться важным фактором ускорения замены аналоговых телевизоров на цифровые и сохранения переходного периода совместного существования аналогового и цифрового телевидения. Аналоговые телевизоры обладают высокой надежностью, и зрители могут продолжать смотреть привычные программы аналоговых каналов еще многие годы. Цифровое наземное телевидение может дать таким зрителям побудительный мотив для замены аналогового телевизора на цифровой.

Понимая эти обстоятельства, мировые лидеры цифрового телевидения (Европа, США, Япония) предлагают три стандарта цифрового наземного телевизионного вещания:

DVB (*Digital Video Broadcasting*) — Европа;

ATSC DTV (*Advanced Television Systems Committee Digital Television*) — США;

ISDB (*Integrated Services Digital Broadcasting*) — Япония.

В 1998 г. в США началось цифровое наземное телевизионное вещание по стандарту ATSC. Хотя с использованием этого формата могут передаваться в цифровой форме телевизионные изображения стандартного формата и разнообразные данные, главное его назначение — телевидение высокой четкости (ТВЧ).

Процесс, приведший к появлению стандарта ATSC DTV, зародился в 1987 году, когда Федеральная комиссия по связи (*Federal Communication Commission* — FCC) США образовала Консультативный комитет по перспективным телевизионным системам с целью разработки стандарта наземного вещания для ТВЧ. Первые 23 рассмотренные системы были аналоговыми, но в 1990 году была предложена первая цифровая система ТВЧ, за которой последовали и другие. В 1993 году было решено объединить лучшие стороны рассмотренных систем и создать цифровую систему ТВЧ. Был образован HDTV *Grand Alliance* (Большой Союз ТВЧ), объединивший представителей промышленности и науки. В декабре 1996 года FCC утвердила разработанную систему ТВЧ под названием ATSC DTV в качестве стандарта США.

В 1992 году состоялось первое заседание группы европейских специалистов, которые собрались для обсуждения проекта цифрового телевизионного вещания, а в 1993 году родился европейский проект DVB. Первым стандартом был DVB-S, который регламентировал цифровое спутниковое вещание. Первое практическое применение европейского стандарта состоялось в Южной Африке и Таиланде. К сегодняшнему дню этот стандарт приняли сотни вещательных организаций разных стран, включая страны Европы, Америки. Следующим был стандарт кабельного телевидения DVB-C, утвержденный через год. Затем пришла очередь наземного телевизионного вещания DVB-T, утвержденного первым в мире в 1996 году. Семейство стандартов DVB определяет единую среду, в которой телевизионные программы в закодированном виде могут переноситься разными носителями, пока не окажутся в приемнике-декодере домашнего телевизора.

Япония также завершила работу над стандартом цифрового телевидения, предложив систему цифрового вещания с предоставлением комплексных услуг. Однако японский стандарт появился позже своих конкурентов и пока рассматривается лишь странами, которые в силу ряда причин хотели бы получить цифровое телевидение в японском варианте.

Очевидно, принятие единого мирового стандарта наземного вещания было бы наилучшим вариантом развития цифрового телевидения. Но если ни один из существующих ныне стандартов не окажется безусловным победителем, то повторится ситуация, которая сложилась с системами цветного телевидения NTSC, PAL, SECAM.

Европейский стандарт EN 300 744 наземного цифрового телевидения

В этом стандарте передаваемые данные представляют собой информацию об изображении и звуковом сопровождении, а также о других данных. Эта информация в системе DVB-T должна быть закодирована в виде пакетов транспортного потока MPEG-2. Стандарт определяет структуру передаваемого потока данных, систему канального кодирования и модуляции для служб наземного телевидения, работающих в форматах стандартной, повышенной и высокой четкости.

Для обеспечения совместимости устройств различных производителей стандарт определяет параметры преобразования данных и сигналов в передающей части системы цифрового наземного телевизионного вещания.

Несмотря на то, что в этом учебном пособии содержатся результаты собственных исследований автора, работа посвящена главным образом изложению основ современных достижений радиоэлектроники, которые позволят учащимся полностью овладеть идеями современной радиоэлектроники и активно использовать их в своей будущей работе.

Выполнению этой задачи посвящено использование большого количества примеров, решение которых рассмотрено для облегчения понимания особенно трудных в методическом отношении вопросов.

При работе над пособием помощь советами и участием в подготовке рукописи оказали автору сотрудники кафедры электроники Белорусского государственного университета информатики и радиоэлектроники. Многочисленные и ценные замечания сделал при рецензировании настоящего учебного пособия главный научный сотрудник НИЧ Белорусского национального технического университета, доктор технических наук, доцент О.Г. Девойно.

Автор выражает своим коллегам по работе и рецензенту искреннюю благодарность.

ПОНЯТИЯ ЭНТРОПИИ И ИНФОРМАЦИИ

Государственные стандарты подготовки специалистов для работы в области радиоэлектроники прежде всего требуют овладения фундаментальными понятиями этой области науки и техники. К числу этих понятий относится целый ряд вопросов, основными из которых являются понятия *информации*, *энтропии* и ряда других, базирующихся на этих основных положениях. Радиоэлектроника как наука декларирует изучение всех аспектов передачи информации на расстояние и главным образом — с помощью электромагнитного излучения, распространяющегося в пространстве, разделяющем приемник и передатчик. При этом имеются в виду и информационные взаимодействия, при которых выполняются операции хранения, измерения, кодирования, шифрования и других видов обработки информации. Хотя существуют различные подходы к определению понятия информации, применяемые в технических системах, но энтропийный подход рассмотрения этой проблемы дает более интересную и значимую характеристику информационных потоков. Поэтому изложение курса «Основы современной радиоэлектроники» целесообразно начать именно с рассмотрения понятий информации и энтропии.

1.1. Понятие информации

Понятие информации является первичным понятием, точное определение которому дать *невозможно*. Поэтому ограничимся лишь некоторыми его толкованиями.

Информация трактует смысл полученного *сообщения*, представляет его *интерпретацию*.

Например, мы хотим помочь читателю сформировать понятие, обозначаемое словом *информация*. Мы располагаем

сведениями об этом понятии и стараемся описать предмет нашего разговора, тем самым передав читателю сведения, восприняв которые он получит представление о сути обсуждаемого предмета. В результате этого имеющиеся сведения о понятии *информация* переходят с помощью сообщений к читателю, но этот переход может с разной степенью адекватности соответствовать реальному положению вещей. Определение смысла передаваемых сведений — задача, выполняемая *человеком*, и именно *человек* определяет, как интерпретирует конкретное сообщение книга, компьютер или другое устройство. При этом книга или компьютер не работают с информацией, поскольку ее нельзя ни ввести в них, ни вывести, так как они всегда имеют дело с сообщениями, представленными данными, на основании которых только *человек* может получить информацию о сути обсуждаемого предмета.

Итак, можно сказать, что *информация* — это понимание смысла, представлений, интерпретаций, возникающее в аппарате мышления человека после получения им данных, связанное с предшествующими знаниями и понятиями о совокупности сведений, например изучаемых в данной книге.

Поэтому в дальнейшем под *информацией* будем понимать то, что содержится в совокупности сведений, находящихся в книгах, компьютерах, библиотеках и т. п.

Дополним описанное понятие информации, перечислив некоторые характерные особенности информации, которые состоят в том, что она:

- представляет собой очень важный ресурс современного производства, снижающий потребности в расходе сырья, энергии, капитала и т. п.;
- является товаром, причем продавец информации ее не теряет после продажи;
- придает дополнительную ценность другим ресурсам: чем интеллектуальнее труд, вложенный в разработку интегральной схемы, тем больше новых возможностей можно ожидать от ее использования;
- нематериальна, но проявляется в форме материальных носителей дискретных знаков или первичных сигналов;
- имеет объективную количественную сторону и не имеет смысловой, так как важность последней может оценить только *человек*, а не техническое устройство;

- отражает содержание сообщения, понижающего неопределенность некоторого опыта с неоднозначным исходом.

Информация играет чрезвычайно важную роль в современном мире. По мере накопления знаний о системных свойствах материи непрерывно растет понимание места информации от простого набора данных о состоянии физического тела до важнейших характеристик состояния всей материи.

Н. Винер подчеркивал, что общество простирается до тех пределов, до каких распространяется информация.

1.2. Информация и энтропия

Впервые на связь информации и энтропии в термодинамике при изучении тепловых процессов указал Л. Больцман, давший статистическое определение энтропии в 1877 году и заметивший, что *термодинамическая энтропия* характеризует *недостающую информацию*. Но эта фраза еще долго оставалась без внимания, пока спустя 70 лет К. Шеннон не сформулировал постулаты теории информации, а затем было замечено, что формула Больцмана инвариантна *информационной энтропии*, и была выявлена их системная связь, доказана системность этих фундаментальных понятий.

В каждом элементарном сообщении для получателя содержится определенная информация, представляющая совокупность сведений о состоянии источника сообщения. Определяя количественную меру этой информации, мы не учитываем ее смысловое содержание, так же как и ее важность для конкретного получателя. Очевидно, что при отсутствии сведений о состоянии источника имеется неопределенность относительно того, какое сообщение u_i из числа U возможных им выбрано, а при наличии этих сведений данная неопределенность полностью исчезает. Поэтому естественно количество информации, содержащейся в сообщении, измерять величиной исчезнувшей неопределенности. Введем меру этой неопределенности, которую можно рассматривать и как количественную меру информации. При этом эта мера должна возрастать с увеличением возможностей выбора, т. е. с увеличением объема алфавита источника N . Кроме того, нужно, чтобы вводимая мера обладала свойством аддитивности,

закключающемся в том, что если 2 независимых источника с объемами алфавита N и M рассматривать как один источник, одновременно реализующий пары состояний n_i и m_j , то в соответствии с принципом аддитивности неопределенность объединенного источника должна равняться сумме неопределенностей исходных источников. Поскольку объем алфавита объединенного источника $NМ$, то искомая функция при равной вероятности состояний источников должна удовлетворять условию $f(NМ) = f(N) + f(M)$. Можно математически строго показать, что единственной функцией, при перемножении аргументов которой значения функций складываются, является логарифмическая функция. Поэтому перечисленные требования выполняются, если в качестве меры $H(U)$ неопределенности источника с равновероятными состояниями и характеризующего его ансамбля U принять логарифм объема алфавита источника.

Впервые данная мера

$$H(U) = \log N \quad (1.1)$$

была предложена Р. Хартли в 1928 г. Основание логарифма в (1.1) не имеет принципиального значения и определяет только масштаб или единицу количества информации. Чаше всего в качестве основания используют число 2, при этом единица количества информации называется *двоичной единицей*, или *битом*, и представляет собой информацию, содержащуюся в одном дискретном сообщении источника равновероятных сообщений с объемом алфавита, равным двум. При выборе в (1.1) основания логарифма равным 10 получаем десятичную единицу, называемую *дитом*. Иногда используют натуральную единицу количества информации, называемую *натом*, при этом основание логарифма в (1.1) равно $e = 2,7182$. В дальнейшем мы будем использовать в качестве основания логарифма число 2.

Предложенная Р. Хартли мера количества информации имеет ограниченное применение, поскольку предполагает равную вероятность выбора источником любого из возможных его состояний. В общем случае, когда вероятности различных состояний источника не одинаковы, степень неопределенности конкретного состояния зависит не только от объема алфавита источника, но и от вероятности этого состояния.

Когда К. Шеннон искал способ количественного описания информации, содержащейся в некотором сообщении, то логические рассуждения вывели его на формулу фактически того же вида, что и у энтропии S , определяемой формулой Планка-Больцмана и характеризующейся *термодинамической вероятностью* W ,

$$S = k \cdot \ln W, \quad (1.2)$$

$k = 1,38 \cdot 10^{-16}$ эрг/град — постоянная Больцмана.

В противоположность обычной вероятности, которая всегда меньше единицы, термодинамическая вероятность — это большое целое положительное число, равное числу микросостояний (*комплексий*), соответствующих заданному макроскопическому состоянию. В трактовке Больцмана *упорядоченная система* имеет *меньшую* энтропию, так как порядок — *менее вероятное* состояние, а *неупорядоченная* — *высокую* энтропию, так как состояние беспорядка является *более* вероятным.

Энтропию производят любые сложные системы, между ними в процессе эволюции идет обмен информацией, который сопровождается изменением энтропии, например при испарении воды утрачивается информация о пространственном расположении молекул, энтропия возрастает; если вода замерзает, уменьшается неопределенность в расположении молекул, этот процесс сопровождается понижением энтропии. При этом каждый бит информации имеет *энтропийную* цену:

$$1 \text{ бит} = k \cdot \ln 2 = 0,97 \cdot 10^{-23} \text{ Дж/град.}$$

Приобретение информации происходит за счет изменения энтропии, т. е. *физическая* и *информационная* энтропии связаны друг с другом.

Численно количество информации равно той энтропии, которая имела место до получения информации. Количественное равенство информации и энтропии не означает тождественности этих понятий. Наоборот, они противоположны друг другу, так как информация равна не просто энтропии, а *устраненной* энтропии.

Ключом к новому пониманию сущности феномена информации и механизма информационных процессов послужила установленная Л. Бриллюэном взаимосвязь информации и физической энтропии.

Л. Бриллюэн писал, что информацию следует рассматривать как отрицательное слагаемое энтропии системы. Он считал, что информация есть *негэнтропия* и что информация и энтропия должны рассматриваться совместно и не могут рассматриваться порознь. Энтропия есть мера *недостатка* информации о системе. Следует отметить, что Л. Бриллюэн различал между собой *свободную* информацию и *связанную информацию*, и считал, что «только связанная информация» будет представлять негэнтропию.

Из этого следует, что Л. Бриллюэн не приписывал любой информации свойства негэнтропии, а это свойство приписывал только связанной информации. Важно понять, какое физическое или информационное наполнение содержит термин негэнтропия. Таким образом, для характеристики степени порядка материальных объектов часто используют *энтропию как неупорядоченность* и *негэнтропию*, т.е. связанную информацию, как *упорядоченность*.

Физическая энтропия является мерой энергетической упорядоченности объекта и представляет собой функцию от числа их возможных состояний. Любое *повышение упорядоченности* объектов ведет к снижению их совокупной энтропии, и наоборот.

Понимание физического смысла энтропии затруднено тем обстоятельством, что ее значение *не может быть измерено* никаким прибором, но зато *вычисляется*.

Физическая и термодинамическая энтропии и информационная энтропия системы принципиально различны. Если физическая и термодинамическая энтропии для замкнутых систем со временем не уменьшаются, а растут в соответствии со вторым началом термодинамики, то информационная энтропия со временем может не увеличиваться, а уменьшаться для любых систем.

Поскольку принципиально неустранимы броуновское движение, флуктуации и шумы как для микро-, так и для макрообъектов и макронаблюдений, то возникает как микро-, так и макронеупорядоченность, что позволяет использовать в качестве универсального параметра микро- и макросистем *обобщенную энтропию*. Понятие обобщенной энтропии представляет такие наиболее общие свойства действительности, как

неупорядоченность и упорядоченность, неопределенность и определенность, хаос и порядок.

Всякое явление *двойственно*, и оно одновременно содержит в себе некоторую хаотическую, броуновскую составляющую и упорядоченную составляющую, как составляющую хаоса, так и порядка.

Наряду с обычной положительной энтропией в природе существует *антиэнтропия*, или *отрицательная энтропия*, т. е. энтропия с *отрицательным знаком*, существенно отличающаяся от положительной энтропии по физическому содержанию.

К сожалению, алгоритмическое истолкование антиэнтропии (отрицательной энтропии) хотя не является чисто формальным, но еще не раскрывает ее физического содержания. Известно, что система обладает *антиэнтропией* даже тогда, когда для нее не существует *полной* процедуры воспроизведения. Полная процедура воспроизведения эквивалентна существованию полного алгоритма воспроизведения. Однако если простейшие системы алгоритмируемы, то природа в целом изобилует *неалгоритмируемыми* системами, к которым следует отнести сознание и, вероятно, все живые организмы. Крут математических задач, не имеющих общего алгоритма для своего решения, по-видимому, относится к системам, не имеющим полной алгоритмической процедуры для своего воспроизведения.

В такой ситуации К. Шеннон количество информации $I(u_i)$, содержащееся в дискретных сообщениях u_i , определил как функцию вероятности появлений этих сообщений $p(u_i) = p_i$ и охарактеризовал величиной

$$H(U) = - \sum_{i=1}^n p_i \log p_i, \quad (1.3)$$

названной им *энтропией*.

Позаимствовать термин из термодинамики посоветовал К. Шеннону знаменитый математик Джон фон Нейман (*John von Neumann*). Полушутя Д. Нейман обосновал свой совет тем, что в среде математиков и инженеров мало кто знает об энтропии, а поэтому Шеннон получит огромное преимущество в неизбежных спорах о новой теории.

Основание логарифма в (1.3) выбирается из тех же соображений, что и в (1.1). Знак минус в равенстве (1.3) необхо-

дим для того, чтобы количество информации $I(u_i)$ было неотрицательным числом, так как всегда $p_i = 1$. Очевидно что, так же, как и мера $H(U)$, определяемая (1.1), величина $I(u_i)$ обладает свойством аддитивности и в случае достоверного сообщения, когда $p(u_i) = 1$, $I(u_i) = 0$. Однако теперь количество информации, содержащееся в дискретном сообщении, зависит от степени неожиданности этого сообщения, характеризующей вероятностью его появления.

Количество информации, которое приходится на одно сообщение a_i , определяется выражением

$$I(a_i) = p(a_i) \log_2 \frac{1}{p(a_i)} = -p(a_i) \log_2 p(a_i), \quad (1.4)$$

т. е. количество информации в сообщении тем больше, чем меньше вероятность его наступления (чем более неожиданно его наблюдение).



Пример 1.1. Оценить информационную емкость двух сообщений о том, что некто 1) выиграл, 2) не выиграл единственный приз в лотерее, в которой было продано 1024 билета.

Решение. Вероятность, соответствующая первому опыту, равна $p_1 = 1/1024$, а второго $p_2 = 1023/1024 = 0,9990234375 \approx 1,000$. Информационная емкость первого сообщения составляет $I(p_1) = -p_1 \log_2 p_1$ 10 бит, а второго опыта приближается к нулю. Эти обстоятельства хорошо согласуются и с нашим практическим опытом. Действительно, сообщение о том, что кто-то ничего не выиграл в лотерее, нам собственно ничего нового и не говорит, потому что мы заранее знаем, что выигрыш в лотерее — это очень редкое событие.

Официальное рождение теории информации как научной дисциплины принято отсчитывать от основополагающей работы К. Шеннона «Математическая теория связи». Именно он ввел в обиход и двоичную единицу информации, впоследствии названную *бит* (*binary digit*), и широко используемую сегодня меру количества информации и энтропии.

Итак, энтропия в термодинамике, введенная в 1865 г. Р. Клаузиусом, образовавшим ее от корня греческого слова

«тропе», означающего «превращение», с добавлением заимствованной из слова «энергия» приставки «эн-», характеризует неопределенность теплового состояния системы, являясь мерой ее внутренней неупорядоченности. В информатике энтропия рассматривается как мера неопределенности системной ситуации.

Единицей измерения энтропии, как и информации, является бит. 1 *бит* — это та неопределенность, которую имеет источник с равновероятной выдачей двух возможных сообщений, обычно символов 0 и 1.

Можно сказать, что энтропия является мерой молекулярного хаоса; чем она больше, тем хаотичнее движение частиц (например, атомов и молекул). Второй закон термодинамики утверждает, что все самопроизвольные процессы в макром мире всегда протекают с увеличением хаоса, т. е. с увеличением энтропии. Это доказано громадным числом экспериментов и не подлежит сомнению. Но столь же надежно установлено, что атомы и молекулы, из которых состоят все окружающие нас тела, движутся по законам классической механики. А в механике доказана теорема Лиувилля, утверждающая, что при механическом движении частиц их энтропия не меняется. И термодинамика, и классическая механика относятся к фундаментальным физическим теориям. Поэтому их выводы не могут противоречить друг другу. А в данном случае противоречие налицо: в термодинамике энтропия возрастает, в механике — всегда постоянна. *Вот проблема для решения молодым и талантливым.*

1.3. Понятие сигнала

Материальным носителем информации является *сигнал как средство хранения и передачи информации*. Причем существенным является не наличие самих материальных объектов, а их состояние. Так, воздух может передавать посредством колебаний речь, музыку, шум самолета. Кроме полезной информации мы можем получать постороннюю, ненужную нам информацию, которую принято называть в теории информации *шумом*.

Кодированием будем называть *систему сигнального соответствия*. Примерами кода являются азбука Морзе, алфавит, жесты глухонемых и т. д. Важно согласование кодов для при-

нимающей и передающей стороны. Для человека, не понимающего китайский язык, речь китайца может являться лишь шумом. Иногда рассогласование кодов организуется специально, например для шифрованного сообщения. Но на самом деле речь идет лишь о смене кода.

Все сигналы делятся на два типа по отношению ко времени. Это статические и динамические сигналы. Статические более удобны для хранения информации. Это книга, фотография, магнитная пленка, компакт-диск. Примером динамического сигнала являются колебания воздуха при переносе речевых или музыкальных сигналов, колебания магнитного поля при переносе радиоволн, световые вспышки и т. д.

1.4. Свойства энтропии

Энтропия является мерой неопределенности опыта, в котором проявляются случайные события, и равна средней неопределенности всех возможных его исходов. Еще раз обсудим понятие о мере для ее измерения.

Рассмотрим опыт с n равновероятными исходами. Так как каждый исход является случайным, то он вносит свой вклад в неопределенность всего опыта. Поскольку все исходы равновероятны, то естественно допустить, что их неопределенности также одинаковы. Следовательно, мера суммарной неопределенности должна быть равна сумме мер неопределенности каждого из опытов, т. е. мера неопределенности аддитивна. Таким образом, за меру неопределенности опыта с n равновероятными исходами можно принять число $\log n$.

Легко распространить этот подход измерения величины неопределенности на случай, когда исходы опытов A_1 и A_2 неравновероятны и имеют, например, $p(A_1)$ и $p(A_2)$. Тогда

$$H_1 = -p(A_1)\log_2 p(A_1) \text{ и } H_2 = -p(A_2)\log_2 p(A_2), \quad (1.5)$$

$$H_3 = H_1 + H_2 = -p(A_1)\log_2 p(A_1) - p(A_2)\log_2 p(A_2), \quad (1.6)$$

Обобщая это выражение на ситуацию, когда опыт имеет n неравновероятных исходов $A_1, A_2, \dots, A_n$, получим выражение для $H(A)$, аналогичное (1.3):

$$H(A) = -\sum_{i=1}^n p(A_i) \log_2 p(A_i). \quad (1.7)$$

Введенная величина $H(A)$ называется энтропией опыта A .



Пример 1.2. Для передачи сообщений используется код, состоящий из четырех независимых друг от друга символов, вероятности появления которых равны 0,4, 0,3, 0,2 и 0,1. Оценить энтропию источника этих сообщений. Как изменится энтропия при равновероятном поступлении генерируемых сообщений?

Решение. В соответствии с формулой (1.7) составим выражение для энтропии источника сообщений. Имеем

$$H = -0,4 \log_2 0,4 - 0,3 \log_2 0,3 - 0,2 \log_2 0,2 - 0,1 \log_2 0,1 = \\ = 0,5288 + 0,5211 + 0,4644 + 0,3322 = 1,8465 \text{ бит.}$$

Для случая равновероятных сообщений

$$H = -4 \cdot 0,25 \cdot \log_2 0,25 = 4 \cdot 0,5000 = 2,0000 \text{ бит.}$$

На этом примере можно непосредственно убедиться в том, что неопределенность возрастает при равномерном распределении вероятностей наступления событий.

1.5. Условная энтропия

Введем понятие энтропии сложного опыта (A, B) , состоящего в том, что опыты не являются независимыми, т. е. на исход B оказывает влияние результат опыта A или исход B оказывает влияние на вероятность наблюдения опыта A . Например, если в урне находится всего два шара разного цвета и опыт A состоит в извлечении первого, а B — второго из них, то опыт A полностью снимает неопределенность сложного опыта (A, B) , т. е. оказывается $H(A, B) = H(A)$, а не сумме энтропий, как это следует из (1.6).

В условиях зависимых друг от друга опытов A и B исход опыта A может оказывать влияние на исход опыта B , т. е. оказывается, что некоторые пары событий A_i, B_j не являются независимыми. Тогда в (1.5) вероятность $p(A_i)$ надо заменить выражением

$$p(A_i, B_j) = p(A_i)p(B_j/A_i), \quad (1.8)$$

где $p(B_j/A_i)$ — вероятность наступления исхода B_j при условии, что в первом опыте имел место исход A_i . Тогда

$$H(A, B) = - \sum_{i=1}^n \sum_{j=1}^m p(A_i) p(B_j / A_i) \log_2 p(A_i) - \\ - \sum_{i=1}^n \sum_{j=1}^m p(A_i) p(B_j / A_i) \log_2 p(B_j / A_i) = H(A) + H(A / B), \quad (1.9)$$

где $H(A/B)$ — условная энтропия опыта B , т. е. энтропия опыта B при условии выполнения опыта A .

Выражение (1.9) представляет собой общее правило нахождения энтропии сложного опыта. Совершенно очевидно, что выражение (1.6) является частным случаем (1.9) при условии независимости опытов A и B .

Если случайная величина X принимает n значений, заданных своими вероятностями, то условная энтропия случайной величины X относительно значения y_j случайной величины Y , принимающих m значений, для дискретных X и Y определяется как

$$H(X / y_j) = - \sum_{i=1}^n p(X = x_i / Y = y_j) \log_2 p(X = x_i / Y = y_j). \quad (1.10)$$

Средней условной энтропией $H_y(X)$ называется математическое ожидание условной энтропии. Для дискретных случайных величин

$$H_y(X) = M[H(X / y)] = \\ = - \sum_{i=1}^n \sum_{j=1}^m p(Y = y_j) p(X = x_i / Y = y_j) \cdot \log_2 p(X = x_i / Y = y_j) \quad (1.11)$$

и для непрерывных случайных величин

$$H_y(X) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_y(y) f(x / y) \log_2 f(x / y) dx dy. \quad (1.12)$$



Пример 1.3. Из урны, в которой находятся 2 белых и 4 черных шара, извлекают последовательно два шара без возврата. Найти энтропии, характеризующие неопределенность первого, второго и обоих извлечений.

Решение. Извлечение первого шара имеет два возможных исхода: A_1 — вынут белый шар с вероятностью осуществления этого события $p(A_1) = 1/3$, A_2 — вынут

черный шар с вероятностью этого события $p(A_2) = 1 - p(A_1) = 2/3$. Эти данные дают возможность рассчитать величину неопределенности $H(A)$, а именно:

$$H(A) = -p(A_1) \log_2 p(A_1) - p(A_2) \log_2 p(A_2) = 0,918 \text{ бит.}$$

Опыт B , состоящий в извлечении второго шара, также имеет два исхода: B_1 — вынут белый шар, B_2 — вынут черный шар. Однако вероятности этих исходов зависят от того, каким был исход опыта с выниманием первого шара. Если имел место исход A_1 , тогда $p(B_1 / A_1) = 1/5$, $p(B_2 / A_1) = 4/5$. При A_2 имеем $p(B_1 / A_2) = 2/5$, $3/5$. Следовательно, величина неопределенности, характеризующая второй опыт, соответствует условной энтропии и, согласно (1.9), равна

$$H(B / A_1) = -(1/5) \log_2 (1/5) - (4/5) \log_2 (4/5) = 0,722 \text{ бит,}$$

$$H(B / A_2) = -(2/5) \log_2 (2/5) - (3/5) \log_2 (3/5) = 0,971 \text{ бит,}$$

$$\begin{aligned} H(B / A) &= P(A_1)H(B / A_1) + p(A_2)H(B / A_2) = \\ &= (1/3) \cdot 0,722 + (2/3) \cdot 0,971 = 0,888 \text{ бит.} \end{aligned}$$

В соответствии с (1.6) для неопределенности опыта обоих извлечений имеем:

$$H(A, B) = 0,918 + 0,888 = 1,806 \text{ бит.}$$

Относительно условной энтропии можно сформулировать такие ее особенности.

1. Условная энтропия представляет собой неотрицательную величину. Действительно, $H(B / A) = 0$ только в том случае, если любой исход A полностью определяет исход B (как в примере с двумя шарами разного цвета в урне), т. е. $H(B / A_1) = H(B / A_2) = \dots = H(B / A_n) = 0$. В этом случае $H(A, B) = H(A)$.

2. Если опыты A и B независимы, то $H(B / A) = H(B)$, причем это оказывается наибольшим значением условной энтропии. Иными словами, опыт A не может повысить неопределенность опыта B , он может либо не оказать никакого влияния (если опыты независимы), либо понизить энтропию B .

Эти особенности позволяют утверждать, что

$$0 \leq H(B / A) \leq H(B), \quad (1.13)$$

т. е. условная энтропия не превосходит безусловную. Отсюда следует, что

$$H(A, B) \leq H(A) + H(B), \quad (1.14)$$

причем равенство реализуется только в случае, когда опыты A и B независимы.

1.6. Энтропия объединения нескольких источников

Под объединением двух источников u и z понимают *обобщенный* источник сообщений (u, z) , характеризующийся совместными распределениями вероятностей $p(u_i, z_j)$ всех возможных комбинаций, где сообщения u_i получены от источника u , а состояния z_j — от источника z .

Энтропия объединения нескольких статистически независимых источников сообщений равна сумме энтропий исходных источников. Это положение характеризует соответствие энтропии обобщенного источника свойству *аддитивности* энтропии.

Действительно, не теряя общности, ограничимся рассмотрением объединения источников сообщений u и z с объемами алфавита соответственно N и M . Тогда в соответствии с определением (1.3) энтропия обобщенного источника будет равна

$$H(U, Z) = - \sum_{i=1}^N \sum_{j=1}^M p(u_i, z_j) \cdot \log_2 p(u_i, z_j). \quad (1.15)$$

В случае статистической независимости u и z , т. е. когда $p(u_i, z_j) = p(u_i) \cdot p(z_j)$, для энтропии обобщенного источника имеем:

$$\begin{aligned} H(U, Z) &= - \sum_{i=1}^N \sum_{j=1}^M p(u_i) \cdot p(z_j) \cdot \log_2 (p(u_i) \cdot p(z_j)) = \\ &= - \sum_{j=1}^M p(z_j) \cdot \sum_{i=1}^N p(u_i) \cdot \log_2 p(u_i) - \sum_{i=1}^N p(u_i) \cdot \sum_{j=1}^M p(z_j) \cdot \log_2 p(z_j) = \\ &= H(U) + H(Z), \end{aligned} \quad (1.16)$$

так как $\sum_{j=1}^M p(z_j) = 1, \sum_{i=1}^N p(u_i) = 1, - \sum_{i=1}^N p(u_i) \cdot \log_2 p(u_i) = H(U),$

$$- \sum_{j=1}^M p(z_j) \cdot \log_2 p(z_j) = H(Z).$$

1.7. Понятие дифференциальной энтропии

Энтропия непрерывного источника сообщений, характеризующегося законом распределения $p(u)$, бесконечна, так как неопределенность выбора из бесконечно большого числа возможных состояний бесконечно велика. Это означает, что собственная информация любой непрерывной случайной величины бесконечно велика. Производя предельный переход в формуле Шеннона (1.3), в которой суммирование заменено интегрированием, с учетом того, что приращение $\Delta u \rightarrow 0$, получим

$$\begin{aligned} H(U) &= - \int_{-\infty}^{\infty} p(u_i) \Delta u \log_2(p(u_i) \Delta u) du = \\ &= - \int_{-\infty}^{\infty} p(u) \log_2 p(u) du - \lim_{\Delta u \rightarrow 0} \log \Delta u. \end{aligned} \quad (1.17)$$

Полагая $p(u) \log_2 p(u) = 0$ для тех значений u , для которых $p(u) = 0$, первое слагаемое называют *дифференциальной энтропией*, так как ее можно трактовать как среднюю неопределенность выбора случайной величины U по сравнению со средней неопределенностью выбора случайной величины, изменяющейся в диапазоне, равном единице, и имеющей равномерное распределение.

Второе слагаемое стремится к бесконечности, и оно не зависит от плотности вероятности $p(u)$.

Обобщение понятия энтропии на случай непрерывных функций распределения путем интегрирования в области изменения U не приводит к нужному результату, ибо плотность $p(u)$ имеет размерность, а логарифм размерной величины не определен. Однако часто положение исправляется путем умножения $p(u)$ на некоторую величину e , имеющую такую же размерность, что и u .

Мы сравниваем неопределенность случайной величины с равномерно распределенной в единичном интервале. Поэтому дифференциальная энтропия может быть не только положительной.

Энтропия $H(U)$ является удобной мерой неопределенности законов распределения вероятностей, особенно в тех случаях, когда распределения являются асимметричными, многовер-

шинными и когда использование таких числовых характеристик, как среднее значение, дисперсия и моменты высших порядков, теряет всякую наглядность и удобство.

Дифференциальная энтропия зависит от масштаба (единицы измерения). Имеются разные подходы к определению понятия количества информации непрерывного источника, ввиду бесконечности его меры неопределенности. Наиболее простой из них — замена непрерывного сообщения дискретным, что позволяет полностью восстановить непрерывное сообщение, а следовательно, не потерять информацию.

Прямой же подход определяет количество информации как разность априорной и апостериорной дифференциальных энтропий. Последняя не равна нулю из-за некоторой неопределенности, остающейся после регистрации события вследствие помех, и конечной величины точности регистрации. При таком определении количества информации исчезает бесконечность после вычитания постоянного бесконечного члена.

Дифференциальная энтропия обладает следующими свойствами:

1. Дифференциальная энтропия в отличие от обычной энтропии дискретного источника не является мерой собственной информации, содержащейся в ансамбле значений случайной величины U . Она зависит от масштаба U и может принимать отрицательные значения. Информационный смысл имеет не сама дифференциальная энтропия, а разность двух дифференциальных энтропий, чем и объясняется ее название.

2. Дифференциальная энтропия не меняется при изменении всех возможных значений случайной величины U на постоянную величину C . Действительно, масштаб U при этом не меняется и справедливо равенство

$$\begin{aligned} H(U + C) &= - \int_{-\infty}^{\infty} p(u + C) \log p(u + C) d(u + C) = \\ &= - \int_{-\infty}^{\infty} p(u) \log p(u) du. \end{aligned} \quad (1.18)$$

Отсюда следует, что $H(U)$ не зависит от математического ожидания случайной величины, так как, изменяя все значения u на C , мы тем самым изменяем на C и ее среднее, то есть математическое ожидание.

3. Дифференциальная энтропия аддитивна, т. е. для объединения (X, Y) независимых случайных величин X и Y справедливо $H(X, Y) = H(X) + H(Y)$. Доказательство этого свойства аналогично доказательству (1.16) аддитивности обычной энтропии.



Пример 1.4. Оценить энтропию случайной величины X , распределенной по биномиальному закону: 1) в общем случае, 2) при $p = 0,5$ и $n = 5$.

Решение. Распределение случайной величины задано выражением

$$p(k) = \begin{cases} 0, & k < 0, \\ C_n^k p^k q^{n-k}, & 0 < k < n, \quad q = 1 - p, \\ 0, & k > n. \end{cases}$$

Рассмотрим первый случай.

$$\begin{aligned} H(X) &= - \sum_{k=0}^n C_n^k p^k q^{n-k} \log_2 (C_n^k p^k q^{n-k}) = \\ &= - \sum_{k=0}^n C_n^k p^k q^{n-k} \log_2 C_n^k - \sum_{k=0}^n C_n^k p^k q^{n-k} \log_2 p^k - \\ &\quad - \sum_{k=0}^n C_n^k p^k q^{n-k} \log_2 q^{n-k}. \end{aligned}$$

Так как при $k = 0$ и $k = n$

$$\log_2 C_n^k = 0, \quad \sum_{k=0}^n C_n^k p^k q^{n-k} = 1, \quad \sum_{k=0}^n k C_n^k p^k q^{n-k} = np,$$

то

$$H(X) = -n(p \log_2 p + q \log_2 q) - \sum_{k=1}^n C_n^k p^k q^{n-k} \log_2 C_n^k.$$

Рассмотрим второй случай. При $p = q = 0,5$ получаем

$$H(X) = -5(0,5 \log_2 0,5 + 0,5 \log_2 0,5) -$$

$$- \sum_{k=1}^4 C_5^k 0,5^k 0,5^{5-k} \log_2 C_5^k \approx 2,198 \text{ бит.}$$



Пример 1.5. Убедитесь, что дифференциальная энтропия сообщений, описываемых нормальным законом распределения, равна $\log \sqrt{2\pi e \sigma^2}$.

1.8. Взаимная информация между двумя непрерывными случайными величинами

Определим с помощью предельного перехода взаимную информацию между двумя непрерывными случайными величинами X и Y . Разбив области определения X и Y соответственно на небольшие интервалы Δx и Δy , заменим эти величины дискретными так же, как это делалось при выводе формулы (1.17).

Взаимную информацию между величинами X и Y можно определить

$$\begin{aligned} I(X, Y) &= \lim_{\substack{\Delta x \rightarrow 0 \\ \Delta y \rightarrow 0}} \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i, y_j) \Delta x \Delta y}{p(x_i) \Delta x p(y_j) \Delta y} \Delta x_i \Delta y_j = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y) dx dy}{p(x) p(y)} = M \left[\log \frac{p(x, y)}{p(x) p(y)} \right]. \quad (1.19) \end{aligned}$$

При этом предельном переходе никаких явных бесконечностей не появилось, т. е. взаимная информация оказывается величиной конечной, имеющей тот же смысл, что и для дискретных сообщений.

С учетом того, что $p(x, y) = p(y) p(x/y)$, равенство (1.19) можно представить в виде

$$\begin{aligned} I(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(y) p(x/y)}{p(x) p(y)} dx dy = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \left[\log \frac{1}{p(x)} - \log \frac{1}{p(x/y)} \right] dx dy = \quad (1.20) \\ &= \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)} dx - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x/y)} dx dy \\ &= H(X) - H(X/Y). \end{aligned}$$

Здесь $H(X)$ — определенная выражением (1.17) дифференциальная энтропия X , а $H(X/Y)$ — условная дифференциальная энтропия

$$H(X/Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x, y)} dx dy. \quad (1.21)$$

Можно показать, что во всех случаях $H(X/Y) = H(X)$. Формула (1.20) имеет ту же форму, что и (1.9), а отличается лишь заменой энтропии дифференциальной энтропией. Легко убедиться, что основные свойства 1 и 2 взаимной информации, описываемые равенствами (1.15)–(1.16), остаются справедливыми и в этом случае.

1.9. ϵ -энтропия и ϵ -производительность источника непрерывных сообщений

Непрерывные сообщения, характеризующиеся бесконечным количеством содержащейся в них информации (телефонные разговоры, телепередачи), успешно передаются по каналам связи. Это объясняется тем, что на практике никогда не требуется абсолютно точного воспроизведения переданного сообщения, а для передачи даже с очень высокой, но ограниченной точностью требуется конечное количество информации, так же как и при передаче дискретных сообщений. Данное обстоятельство и положено в основу определения количественной меры собственной информации, источников непрерывных сообщений. В качестве такой меры принимается минимальное количество информации, необходимое для воспроизведения непрерывного сообщения с заданной точностью. Очевидно, что при таком подходе собственная информация зависит не только от свойств источника сообщений, но и от выбора параметра ϵ , характеризующего точность воспроизведения. *Возможны различные подходы к определению ϵ в зависимости от вида и назначения передаваемой информации.* Наиболее часто в информационной технике в качестве ϵ используют среднеквадратическое отклонение между принятым y и переданным x сигналами, отражающими непрерывные сообщения, т. е.

$$\epsilon = \sqrt{(y - x)^2} = \sqrt{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - x)^2 p(x, y) dx dy}, \quad (1.22)$$

где X и Y — ансамбли сигналов, отражающих исходное и воспроизведенное сообщения.

Два варианта сообщения или сигнала, различающиеся не более чем на заданное значение ϵ , называются эквивалентными. Взаимная информация $I(X, Y)$ между двумя эквивалентными процессами $X(t)$ и $Y(t)$ может быть определена в соответствии с (1.21) как $I(X, Y) = H(X) - H(X/Y)$, где $H(X)$ и $H(X/Y)$ — соответственно дифференциальная и условная дифференциальная энтропии. Из приведенного выражения видно, что величина $I(X, Y)$ зависит не только от собственного распределения $p(x)$ ансамбля X , но и от условного распределения $p(x/y)$, которое определяется способом преобразования процесса X в Y . Для характеристики собственной информации, содержащейся в одном отсчете процесса X , нужно устранить ее зависимость от способа преобразования сообщения X в эквивалентное ему сообщение Y . Этого можно добиться, если под количеством собственной информации или ϵ -энтропией $H_\epsilon(X)$ процесса X понимать минимизированную по всем распределениям $H(X/Y)$ величину $I(X, Y)$, при которой сообщения X и Y еще эквивалентны.

Таким образом, ϵ -энтропия определяет минимальное количество информации, содержащейся в одном отсчете *непрерывного сообщения*, необходимое для воспроизведения его с заданной верностью. Если ансамбль сообщений X представляет собой процесс с *дискретным временем* с непрерывными отсчетами, то под ϵ -производительностью источника понимают величину

$$H'_\epsilon(X) = \nu_c H_\epsilon(X), \quad (1.23)$$

где ν_c — количество отсчетов сообщения, выдаваемых в единицу времени.

В том случае, когда X — непрерывный случайный процесс с ограниченным спектром, вся информация, содержащаяся в его значениях, эквивалентна информации, содержащейся в отсчетах процесса, следующих друг за другом с интервалом $\Delta t = 1/2f_m$, (f_m — граничная частота спектра), т. е. со скоростью

$$\nu_c = 2f_m. \quad (1.24)$$

При этом ϵ -производительность источника или процесса по-прежнему определяется выражением (1.23), где величина ν_c рассчитывается из условия (1.24).

В том случае, если следующие друг за другом отсчеты процесса коррелированы, величина $H_\epsilon(X)$ должна вычисляться с учетом вероятностных связей между отсчетами.

Таким образом, ϵ -производительность источника непрерывных сообщений представляет собой минимальное количество информации, которое нужно создать источнику в единицу времени, для воспроизведения его сообщений с заданной верностью.

Поэтому ϵ -производительность называют также скоростью создания информации при заданном критерии верности.

Максимально возможная ϵ -производительность $H'_{\epsilon \max}$ непрерывного источника X обеспечивается при гауссовском распределении X с дисперсией σ_x^2 (при этом условии $H(X)$ максимальна) (см. пример 1.5).



Пример 1.6. Оценить значение $H'_{\epsilon \max}$, когда непрерывное сообщение $X(t)$ представляет собой стационарный гауссовский процесс с равномерным энергетическим спектром, ограниченным частотой F_c , и с заданной мощностью (дисперсией) P_x , а критерий эквивалентности ϵ задан в виде (1.22).

Решение. Будем считать, что заданная верность воспроизведения обусловлена действием аддитивной статистически не связанной с сигналом помехи $\epsilon(t)$ с математическим ожиданием $M[\epsilon] = 0$ и дисперсией (мощностью) $\sigma_\epsilon^2 = P_\epsilon$. Исходный сигнал X рассматриваем как сумму воспроизводящего сообщения сигнала Y и помехи: $X = Y + \epsilon$.

При этом, поскольку $p(x/y) = p(y + \epsilon/y) = p(\epsilon/y) = p(\epsilon)$, то $H(X/Y)$ полностью определяется шумом воспроизведения $\epsilon(t)$. Поэтому $\max H(X/Y) = \max H(\epsilon)$. Так как шум воспроизведения имеет фиксированную дисперсию $\sigma_\epsilon^2 = \overline{\epsilon^2(t)}$, то дифференциальная энтропия имеет максимум при гауссовском распределении шума

$$\max H(\epsilon) = \log \sqrt{2\pi e \sigma_\epsilon^2}.$$

В свою очередь дифференциальная энтропия гауссовского источника с дисперсией $\sigma_x^2 = P_x$ (см. пример 1.5) $H(X) = \log \sqrt{2\pi e \sigma_x^2}$.

Следовательно, ε -энтропия на один отсчет сообщения

$$\begin{aligned} H_{\varepsilon \max}(X) &= \log \sqrt{2\pi e \sigma_x^2} - \log \sqrt{2\pi e \sigma_\varepsilon^2} = \\ &= \frac{1}{2} \log \frac{\sigma_x^2}{\sigma_\varepsilon^2} = \frac{1}{2} \log \frac{P_x}{P_\varepsilon} \end{aligned} \quad (1.25)$$

Величина $P_x/P_\varepsilon = c_0$ характеризует минимальное отношение сигнал/шум, при котором сообщения $X(t)$ и $Y(t)$ еще эквивалентны.

Согласно теореме Котельникова шаг дискретизации $\Delta t = 1/2F_c$, а $\nu_c = 2 F_c$. При этом равномерность спектра сообщения обеспечивает некоррелированность отстоящих на Δt друг от друга отсчетов, а гауссовский характер распределения $X(t)$ — их независимость. Следовательно, в соответствии с (1.23)

$$H'_{\varepsilon \max}(X) = 2F_c H_\varepsilon(X)$$

или с учетом (1.25)

$$H'_{\varepsilon \max}(X) = F_c \log \frac{P_x}{P_\varepsilon} = F_c \log \rho_0. \quad (1.26)$$

Количество информации, выданное таким источником за время T_c , равно

$$T_c H'_\varepsilon(X) = T_c F_c \log \rho_0. \quad (1.27)$$

Интересно отметить, что правая часть выражения (1.27) совпадает с наиболее общей характеристикой сигнала, называемой его объемом, если принять динамический диапазон сигнала $D = \log c_0$. Это означает, что объем сигнала равен максимальному количеству информации, которое может содержаться в сигнале длительностью T_c .

1.10. Связь между энтропией и информацией

Из рассмотрения свойств энтропии можно увидеть, как предшествующий опыт A уменьшает величину неопределенности последующего опыта B . Разность $H(B)$ и $H(B/A)$ показывает, какие новые сведения относительно B мы получаем, произведя опыт A . Эта величина называется *информацией*

относительно опыта B , содержащейся в опыте A , и обозначается равенством

$$I(A, B) = H(B) - H(B / A). \quad (1.28)$$

Данное выражение открывает возможность численного измерения количества информации на основании единиц, принятых нами для измерения энтропии, т. е. поскольку единицей измерения энтропии является бит, то в этих же единицах следует измерять и количество информации, получаемой в результате реализации соответствующего исхода опыта.

Пусть $A = B$, т. е. просто проведен опыт B . Так как этот опыт несет полную информацию о себе самом, неопределенность его исхода полностью снимается, т. е. $H(B / B) = 0$. Тогда $I(A, B) = H(B)$, т. е. можно считать, что энтропия опыта B равна информации, которая содержится в нем самом. Точнее говоря, энтропия опыта равна той информации, которую мы получаем в результате осуществления этого опыта.

Отметим ряд существенных свойств информации:

- из соотношений (1.9) и (1.28) приходим к выводу, что информация относительно опыта B , содержащаяся в опыте A $I(A, B) \geq 0$, причем $I(A, B) = 0$ тогда и только тогда, когда опыты A и B независимы;
- $I(A, B) = I(B, A)$, т. е. информация симметрична относительно последовательности опытов, т. е. информация, содержащаяся в опыте A относительно опыта B , равна информации, содержащейся в опыте B относительно опыта A ;
- из записанных положений и представления энтропии в виде (1.3) следует, что

$$I = - \sum_{i=1}^n p(A_i) \log_2 p(A_i), \quad (1.29)$$

т. е. информация опыта равна среднему значению количества информации, содержащейся в каком-либо одном его исходе.



Пример 1.7. Какое количество информации мы узнаем, когда фиксируем результат броска монеты?

Решение. В данном случае события равновероятны, т. е. $p_1 = p_2 = 0,5$. Согласно (1.30) имеем:

$$I = -0,5 \cdot \log_2 0,5 - 0,5 \cdot \log_2 0,5 = 1 \text{ бит.}$$



Пример 1.8. Какое количество информации мы получаем в результате броска игральной кости?

Решение. Вероятность интересующей нас грани равна $1/6$. Поэтому в соответствии с (1.30) получаем

$$I = -(1/6) \cdot \log_2(1/6) - (5/6) \cdot \log_2(5/6) \approx 0,65 \text{ бит.}$$

Из рассмотренных примеров можно сделать вывод о том, что когда все n исходов равновероятны, то в этом случае

$$I = \sum_{i=1}^n \frac{1}{n} \log_2 n = \log_2 n. \quad (1.30)$$

Эта формула связывает количество равновероятных состояний n в сообщении I , когда любое из этих сообщений реализовалось. Ее смысл в том, что если некоторое множество содержит n элементов и x принадлежит данному множеству, то для его выделения (однозначной идентификации) среди прочих требуется количество информации, равное $\log_2 n$.

Частным случаем применения формулы (1.21) является ситуация, когда $n = 2^k$, подставляя это значение в (1.21), получим

$$I = k \text{ бит.} \quad (1.31)$$

Именно эта ситуация была реализована в примере 1.4. Анализируя результаты решения, можно прийти к выводу, что k как раз равно количеству вопросов с бинарными равновероятными ответами, которые определяли количество информации в рассмотренном примере.



Пример 1.9. Источник информации генерирует множество кодовых комбинаций, одна из которых имеет априорную вероятность $p(x) = 1/8$, а апостериорные вероятности, соответствующие последовательному приему символов y , z , и u , равны $p(x/y) = 1/6$, $p(x/yz) = 1/2$, $p(x/yzu) = 1$. Оценить увеличение информации о сообщении x в процессе приема символов y , x и u .

Решение. Найдем величину априорной энтропии, связанную с приемом сообщения с вероятностью $p(x)$.

В соответствии с (1.1) количество информации, доставляемое самим символом x или любым другим однозначно с ним связанным, в нашем случае определяется

формулой $I = \log_2 p(x)$, и поэтому априорная энтропия равна $H = \log_2 p(x) = \log_2 (1/8) = 3$ бита.

После приема символа y энтропия опыта снижается на величину, равную содержащейся в этом символе взаимной информации, содержащейся в символе y и относящейся к символу x

$$I(x; y) = \log_2(p(x/y)/p(x)) = \log_2(4/3) \approx 0,415 \text{ бит.}$$

Второй принятый символ z доставит дополнительную информацию

$$I(x; z/y) = \log_2(p(x/yz)/p(x/y)) = \log_2 3 \approx 1,585 \text{ бит.}$$

Третий символ u доставит информацию о сообщении x

$$I(x; u/yz) = \log_2(p(x/yzu)/p(x/yz)) = \log_2 2 = 1 \text{ бит.}$$

Полная информация

$$I(x; yzu) = I(x; y) + I(x; z/y) + I(x; u/yz) = 3 \text{ бита,}$$

что и было правильно рассчитано в начале задачи.

1.11. Взаимная энтропия и информация

Определенная равенством (1.3) энтропия характеризует информационные свойства одного дискретного источника или ансамбля. Однако в технике связи очень часто представляет интерес выявление количества информации содержащегося в одном ансамбле сообщений U с объемом алфавита N относительно другого в общем случае зависящего от него ансамбля Z , с объемом алфавита M . В качестве U и Z можно рассматривать, например, ансамбль сообщений U и сигналов Z , с помощью которых передают сообщения U . Для определения такой информационной характеристики мы ввели понятие *условной энтропии*, которое обозначали выражением $H(U/Z)$, определяющим среднее количество информации, даваемое сообщением ансамбля U при условии, что сообщение ансамбля Z уже известно. Если оба ансамбля имеют независимые элементы, то мера неопределенности $H(U/Z)$ находится усреднением по всем значениям z_j средней неопределенности элементов ансамбля $H(U/z_j)$ при данном z_j . Последнее находится аналогично энтропии $H(U)$ заменой безусловных вероятностей $p(u_k)$ (появление сообщения u_k) на условные (при условии z_j) появления сообщений u_k с вероятностями $p(u_k/z_j)$

$$H(U / Z) = - \sum_{j=1}^M p(z_j) \sum_{k=1}^N p(u_k / z_j) \cdot \log_2 p(u_k / z_j), \quad (1.32)$$

где $\sum_{k=1}^N p(u_k / z_j) \cdot \log_2 p(u_k / z_j) = H(U / z_j)$ — усредненное значение неопределенности элементов ансамбля $H(U/z_j)$ при данном z_j .

По теореме умножения вероятностей имеем:

$$p(z_j) \cdot p(u_k / z_j) = p(u_k, z_j), \quad (1.33)$$

где $p(u_k, z_j)$ — вероятность совместного появления сообщений u_k и z_j . С учетом выражения (1.33) выражение (1.32) можно переписать в виде

$$H(U/Z) = \sum_{j=1}^M \sum_{k=1}^N p(u_k, z_j) \cdot \log_2 (p(u_k, z_j) / p(z_j)). \quad (1.34)$$

Возможно также другое представление (1.32)

$$H(U / Z) = M[-\log_2 p(u_k / z_j)] = - \sum_{j=1}^M \sum_{i=1}^N p(u_k, z_j) \log_2 p(u_k, z_j), \quad (1.35)$$

где $M[\cdot]$ — символ математического ожидания.

Условная энтропия удовлетворяет неравенству (1.13).

Когда $H(U / Z) = 0$, то по реализации ансамбля Z можно точно установить реализацию ансамбля U (канал без помех).

Когда $H(U / Z) = H(U)$, ансамбли U и Z независимы и знание реализации Z ничего не говорит о реализации U . В общем случае $H(U / Z) \leq H(U)$ и знание реализаций Z снижает первоначальную неопределенность U . На основании этого можно ввести информационную характеристику двух ансамблей U и Z называемую *взаимной информацией* между U и Z , или *количеством информации*, содержащейся в Z относительно U , которая определяется, как

$$I(U; Z) = H(U) - H(U / Z). \quad (1.36)$$

Взаимная информация измеряется в тех же единицах, что и энтропия, например в битах.

Величина $I(U; Z)$ показывает, сколько в среднем бит информации о реализации ансамбля U дает наблюдение реализации ансамбля Z .

Подставляя (1.3) и (1.11) в (1.36), имеем:

$$I(U; Z) = - \sum_{k=1}^N p(u_k) \cdot \log_2 p(u_k) + \\ + \sum_{k=1}^N \sum_{j=1}^M p(u_k, z_j) \cdot \log_2 (p(u_k, z_j) / p(z_j)).$$

Учитывая, что $p(u_k) = \sum_{j=1}^M p(u_k, z_j)$, выражение для $I(U; Z)$

можно записать в виде

$$I(U; Z) = - \sum_{k=1}^N \sum_{j=1}^M p(u_k, z_j) \cdot \log_2 p(u_k) + \\ + \sum_{k=1}^N \sum_{j=1}^M p(u_k, z_j) \cdot \log_2 (p(u_k, z_j) / p(z_j)) = \quad (1.37) \\ = \sum_{k=1}^N \sum_{j=1}^M p(u_k, z_j) \left[\frac{1}{\log_2 p(u_k)} + \log_2 \frac{p(u_k, z_j)}{p(z_j)} \right] = \\ = \sum_{k=1}^N \sum_{j=1}^M p(u_k, z_j) \cdot \log_2 \frac{p(u_k, z_j)}{p(u_k) p(z_j)} = I(Z; U),$$

обсудим полученные результаты. Мы видим из (1.37), что взаимная информация обладает следующими свойствами.

$$1. I(U; Z) = 0, \quad (1.38)$$

причем равенство имеет место только в том случае, когда U и Z независимы между собой. Это следует из определения (1.12) и неравенства (1.13).

$$2. I(U; Z) = I(Z; U), \quad (1.39)$$

т. е. Z содержит столько же информации относительно U , сколько U содержит информации относительно Z , это свойство вытекает из симметрии (1.14). Поэтому можно также записать

$$I(U; Z) = H(Z) - H(Z / U). \quad (1.40)$$

3. Можно записать

$$I(U; Z) \leq H(U) \quad (1.41)$$

$$I(U; Z) \leq H(Z), \quad (1.42)$$

причем равенство в этих неравенствах имеет место, когда по реализации z можно точно восстановить реализацию u или наоборот.

4. Полагая в (1.13) $Z = U$ и учитывая, что $H(U/U) = 0$, получаем $I(U; U) = H(U)$. Это позволяет интерпретировать энтропию источника как его собственную информацию ансамбля U о самом себе.

Пусть U ансамбль дискретных сообщений, а Z ансамбль дискретных сигналов, в которые преобразуется сообщение U , тогда $I(U; Z) = H(U)$ только в том случае, когда преобразование U в Z обратимо, т. е. однозначно. При необратимом преобразовании $I(U; Z) < H(U)$ и разность

$$H(U) - I(U; Z) = H(U/Z) \quad (1.43)$$

называют потерей информации, или *ненадежностью преобразования* U в Z .

Таким образом, информация не теряется только при обратимых преобразованиях, величина

$$H(Z/U) = H(Z) - I(U; Z) \quad (1.44)$$

называется *энтропией шума преобразования*, или *ложной информацией*, создаваемой при ее образовании.

1.12. Условная взаимная информация

В дискретных линиях передачи информации как реализации сообщений, так и реализации соответствующих им сигналов представляют собой последовательности символов алфавита, содержащего конечное число элементарных символов.

Если x_i , $i = 1, 2, \dots, n$ представляют собой совокупность возможных различных символов алфавита на входе системы, а y_j , $j = 1, 2, \dots, m$ — совокупность символов на ее выходе, то тогда

- $p(x_i)$ — априорная вероятность появления символа x_i на входе системы;
- $p(y_j)$ — априорная вероятность появления символа y_j на выходе системы;

- $p(x_i, y_j)$ — совместная вероятность появления на выходе системы символа y_j , если на входе был символ x_i ;
- $p(x_i/y_j)$ — апостериорная вероятность появления на входе системы символа x_i , когда на ее выходе был зафиксирован символ y_j ;
- $p(x_i; y_j)$ — взаимная вероятность появления двух символов x_i и y_j , характеризующая количество информации, получаемой об ансамбле y_j в результате наблюдения символов x_i или получаемой об ансамбле x_i в результате наблюдения символов y_j ;
- $p(y_j/x_i)$ — условная вероятность появления на выходе системы символа y_j при условии, что на входе был символ x_i .

Для этих вероятностей выполняются соотношения

$$p(x_i) = \sum_{j=1}^m p(x_i, y_j); \quad p(y_j) = \sum_{i=1}^n p(x_i, y_j);$$

$$\sum_{i=1}^n p(x_i) = \sum_{j=1}^m p(y_j) = 1. \quad (1.45)$$

Тогда количество информации, доставляемое самим символом x_i или другим однозначно с ним связанным процессом, можно назвать *собственной информацией символа x_i* и определить формулой

$$I(x_i) = -\log p(x_i). \quad (1.46)$$

Другие понятия информации определим следующими соотношениями:

• *условная собственная информация $I(x_i/y_j)$ символа x_i при известном y_j*

$$I(x_i / y_j) = -\log p(x_i / y_j); \quad (1.47)$$

• *количество информации относительно символа x_i , доставляемое символом y_j (количество информации относительно символа y_j , доставляемое символом x_i), которое представляет собой взаимную информацию $I(x_i; y_j)$ двух случайных символов x_i и y_j относительно друг друга*

$$I(x_i; y_j) = \log_2 \frac{p(x_i, y_j)}{p(x_i) \cdot p(y_j)}; \quad (1.48)$$

- собственная информация $I(x_i, y_j)$ совместного события x_i, y_j

$$I(x_i, y_j) = -\log p(x_i, y_j); \quad (1.49)$$

- среднее количество информации $I(X, y_j)$, доставляемое принятым символом y_j относительно множества всех передаваемых символов $X = \{x_i\}, i = 1, 2, \dots, n$

$$I(X; y_j) = \sum_{i=1}^n I(x_i; y_j) p(x_i / y_j) = \sum_{i=1}^n p(x_i / y_j) \cdot \log_2 \frac{p(x_i / y_j)}{p(x_i)}; \quad (1.50)$$

- среднее количество взаимной информации $I(x_i; Y)$ по множеству символов $Y = \{y_j\}, j = 1, 2, \dots, m$ при фиксированном x_i

$$I(x_i; Y) = \sum_{j=1}^m I(x_i; y_j) \cdot p(y_j / x_i) = \sum_{j=1}^m p(y_j / x_i) \cdot \log_2 \frac{p(y_j / x_i)}{p(y_j)}; \quad (1.51)$$

- полное среднее количество взаимной информации $I(X; Y)$ в множестве символов Y относительно множества символов X

$$I(X; Y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 \frac{p(x_i / y_j)}{p(x_i)} = \sum_{j=1}^m p(y_j) \cdot I(X; y_j). \quad (1.52)$$

Последнее соотношение является наиболее часто используемым соотношением при решении большинства задач, связанных с построением систем передачи и преобразования информации.

Заметим, что если символ x_i , отражающий, например, информационное содержание сообщения, поступающего на вход линии передачи информации, статистически связан не только с символом передачи y_j , соответствующим выходному сигналу линии передачи, но и с третьим символом $z_k, k = 1, 2, \dots, l$, то при известных вероятностях $p(x_i, y_j, z_k)$ условная взаимная информация равна

$$I(x_i; y_j / z_k) = \log_2 \frac{p(x_i / y_j z_k)}{p(x_i)} = \log_2 \frac{p(x_i; y_j / z_k)}{p(x_i / z_k) p(y_j / z_k)}. \quad (1.53)$$

Соотношения (1.21)–(1.29) позволяют записать целый ряд других соотношений, справедливых при обработке информации

и часто оказывающихся весьма полезными при проведении практических расчетов.

Для любых символов x_i, y_j и z_k имеют место соотношения

$$I(x_i; y_j) = I(y_j; x_i), I(x_i; y_j) = I(x_i), I(x_i; y_j) = I(y_j); \quad (1.54)$$

$$I(x_i; y_j z_k) = I(x_i; y_j) + I(x_i; z_k / y_j) = I(x_i; z_k) + I(x_i; y_j / z_k). \quad (1.55)$$

Полное среднее количество взаимной информации $I(x_i; y_j)$ в множестве символов Y показывает, какое количество информации о случайной величине X может быть получено в результате наблюдения случайной величины Y , и измеряется разностью энтропий случайной величины X и ее средней условной энтропии относительно случайной величины Y :

$$I(x_i; y_j) = I(x_i) - I(x_i / y_j) = I(y_j) - I(y_j / x_i) = i(x_i) + I(y_j) - I(x_i y_j), \quad (1.56)$$

$$I(x_i y_j) = I(x_i) + I(y_j) - I(x_i; y_j). \quad (1.57)$$

С учетом средних количеств информации, доставляемых принятым символом y_j относительно множества всех передаваемых символов $X = \{x_i\} i = 1, 2, \dots, n$, а также отражающих информационное содержание сообщений x_i относительно принятых сигналов $Y = \{y_j\}, j = 1, 2, \dots, m$, можем записать

$$I(X; y_j) = 0, I(x_i; Y) = 0, I(X; Y) = 0, I(Y; X) = 0, \quad (1.58)$$

$$I(X; YZ) = I(X; Y) + I(X; Z/Y), \quad (1.59)$$

$$I(YZ; X) = I(Y; X) + I(Z; X/Y). \quad (1.60)$$

Используя определение среднего количества взаимной информации, запишем для средней собственной информации, доставляемой случайной величиной X ,

$$I(X) = \sum_{i=1}^n p(x_i) \cdot I(x_i) = - \sum_{i=1}^n p(x_i) \cdot \log_2 p(x_i) = H(X), \quad (1.61)$$

где $H(X)$ — энтропия дискретной случайной величины X , определяющая количественную меру неопределенности о сообщении до его приема. Для энтропии справедливы следующие выражения:

$$\begin{aligned} H(Y/X) &= - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(x_i / y_j) = \\ &= - \sum_{i=1}^n p(x_i) \sum_{j=1}^m p(y_j / x_i) \cdot \log_2 p(y_j / x_i) = I(Y / X). \end{aligned} \quad (1.62)$$

$$\begin{aligned}
 H(X/Y) &= - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log_2 p(x_i / y_j) = \\
 &= - \sum_{j=1}^m p(y_j) \sum_{i=1}^n p(x_i / y_j) \cdot \log_2 p(x_i / y_j)
 \end{aligned} \tag{1.63}$$

$$H(XY) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \cdot \log p(x_i, y_j) \tag{1.64}$$

$$H(XY) = H(X) + H(Y/X) = H(Y) + H(X/Y), \tag{1.65}$$

где $H(Y/X)$ — условная энтропия множества событий Y при данном множестве событий X , $H(X/Y)$ — условная энтропия множества событий X при данном множестве Y ; $H(XY)$ — энтропия множества совместных событий XY .

Записанные соотношения будут использованы в вычислениях количества переданной информации по каналам связи, характеризующимися некоторыми дополнительными особенностями.

2.1. Классификация сигналов

Поскольку реальные физические процессы протекают во времени, то в качестве математической модели сигнала, представляющего эти процессы, используют функции времени, отражающие изменения физических процессов.

Все сигналы можно разделить на детерминированные (известные) и случайные. К детерминированным относятся сигналы, значения которых известны в любой момент времени. Если же значения сигнала невозможно предсказать с вероятностью, близкой к единице, то такой сигнал мы будем называть случайным. По форме все сигналы можно разделить на три группы: аналоговые, дискретные и цифровые. *Аналоговый* сигнал описывается непрерывной (или кусочно-непрерывной) функцией $x_a(t)$, причем сама функция, как и ее аргумент, может принимать любые значения из выбранных пределов. *Дискретный* сигнал получается в результате *дискретизации непрерывной функции*, представляющей замену непрерывной функции ее дискретными значениями, и описывается решетчатой функцией (последовательным временным рядом) $x(nT)$, который может принимать любые значения в некотором интервале, в то время как независимая переменная n принимает дискретные значения $n = 0, \pm 1, \pm 2, \dots$, а T представляет собой интервал дискретизации. Выполнение операции дискретизации основано на теореме Котельникова, согласно которой функция с верхней частотой спектра f_g полностью определяется последовательностью значений в точках отсчета (отсчетов), отстоящих одна от другой на временной интервал, равный $1/2f_g$. *Цифровой* сигнал описывается *квантованной* решетчатой функцией (квантованной последовательностью, квантованным временным рядом) $x_q(nT)$, т. е. решетчатой функцией, принимающей лишь ряд диск-

ретных значений, называемых *уровнями квантования*, в то время как независимая переменная n принимает значения $0, \pm 1, \pm 2, \dots$

Операции квантования и восстановления сигнала не являются в общем случае взаимно обратными, так как квантование в общем случае выполняется с неустранимой погрешностью. Переход от дискретного сигнала к цифровому в общем случае также осуществляется неточно.

Кроме того, все сигналы можно также разделить на две категории: периодические и непериодические. К периодическим сигналам отнесем те сигналы, которые можно описать некоторой временной функцией $s(t)$, такой, что для нее можно указать некое число T , такое, что для всякого $n = 0, \pm 1, \pm 2, \dots$ будет выполняться условие $s(t) = s(t + nT)$. Если такого числа T для функции $s(t)$ указать невозможно, то тогда сигнал, описываемый функцией $s(t)$, называется непериодическим. Число T называют периодом функции, а значит, и сигнала $s(t)$ и для определенности полагают $T > 0$. Период сигнала связан с его частотой F известным соотношением $T = 1/F$. Частота колебания, имеющего период $T = 1$ с, называется 1 Гц.

В радиотехнике широко используется так называемая циклическая частота $\Omega = 2\pi F = 2\pi / T$, размерность которой представляет отношение *радиан/с*.

При решении конкретных практических задач обязательно нужно иметь в виду различие между размерностями частоты $F = 1/T$ колебания и циклической частоты $\Omega = 2\pi/T$ этого же колебания. С физической точки зрения это различие заключается в том, что частота F показывает, сколько оборотов в единицу времени совершает, например, радиус-вектор вращающейся точки, а частота Ω — сколько радиан проходит в единицу времени тот же радиус-вектор вращающейся точки.

И в завершение классификации отметим, что все сигналы можно разделить на две категории по ширине их спектра: с бесконечно широким спектром и ограниченным (*финитным*) спектром. Конечно, сигналы с бесконечно широким спектром имеют чисто теоретический интерес, так как все физически реализуемые системы имеют ограниченную полосу рабочих частот. Поэтому на практике используются сигналы с

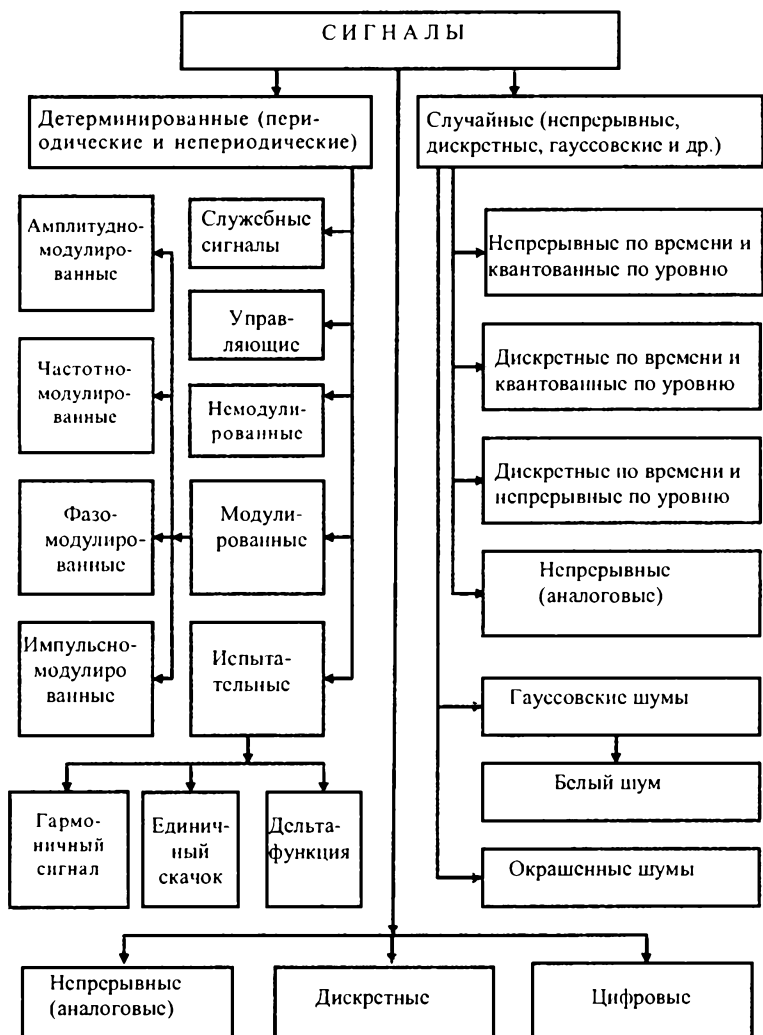


Рис. 2.1. Общая классификация сигналов

ограниченным спектром, которые представляют в большинстве случаев физические модели сигналов с неограниченным спектром. Общая классификация сигналов иллюстрируется рис. 2.1.

Классификация сигналов нужна для правильного выбора их в случае системного использования. Например, если мы выбираем для передачи информации сигнал с амплитудной модуляцией, то тем самым мы выбираем и ряд требований ко всему каналу связи: ширину канала связи, энергетические показатели системы связи, эффективность использования мощности несущей частоты при передаче информационного сигнала и т. д.

2.2. Детерминированные сигналы

Среди детерминированных сигналов выделим пять классов сигналов:

- служебные сигналы;
- управляющие (модулирующие);
- высокочастотные немодулированные;
- модулированные (радиосигналы);
- испытательные.

2.2.1. Служебные сигналы

Служебные сигналы — это сообщения, передаваемые от одного узла системы связи другому с целью управления вызовом. Служебные сигналы обеспечивают взаимодействие службы с клиентами, подключение и отключение абонентов, а также учет всех корректных вызовов.

Не для всех соединений нужны служебные сигналы. В традиционных линиях передачи данных, построенных на принципе мультиплексирования с разделением во времени (TDM — *Time Division Multiplex*), используется передача служебных сигналов по индивидуальному каналу (CAS — *Channel Associated Signaling*), а в сетях с комплексным обслуживанием (ISDN — *Integrated Services Digital Network*) — передача служебных сигналов по общему каналу (CCS — *Common Channel Signaling*).

Независимо от того, используется стандартная или какая-то другая система обмена служебными сигналами, ее целью всегда является управление функциями канала.

2.2.2. Управляющие сигналы

Управляющие сигналы представляют собой передаваемые сообщения, преобразованные в электрическую форму. Процесс получения электрического сигнала, например соответствующего звучанию камертона, состоит в замене механических колебаний электрическими, изменяющимися по тому же закону, что и механические.

2.2.3. Немодулированные (высокочастотные) колебания

Высокочастотные колебания используются для переноса информации, содержащейся в управляющих колебаниях, поскольку они могут создавать электромагнитные волны, распространяющиеся с малыми потерями в пространстве, разделяющем приемник и передатчик.

Аналитическое выражение для высокочастотного колебания имеет вид: $U(t) = U_m \cos(\omega t + \varphi)$. Мы видим, что в зависимости от того, какой параметр высокочастотного колебания изменяется в соответствии с законом передаваемого сообщения, возможны три основных вида модуляции: амплитудная, частотная и фазовая. При изменении амплитуды (частоты, начальной фазы) высокочастотного колебания по закону передаваемого сообщения информация об этом сообщении запечатлевается в изменении амплитуды (частоты, начальной фазы) высокочастотного колебания. Частота высокочастотного сигнала, на которой осуществляется перенос электрического сигнала на расстояние по каналу связи, называется *несущей*.

2.2.4. Модулированные сигналы

Модулированные сигналы формируются на выходе модулятора и представляют собой высокочастотные радиосигналы, несущие в себе информацию, содержащуюся в передаваемом сообщении. В зависимости от того, какой параметр высокочастотного колебания изменяется в соответствии с законом передаваемого сообщения, различают амплитудно-модулированные колебания и колебания с угловой модуляцией, которые в свою очередь можно разделить на колебания с частотной

и фазовой модуляцией. Например, при передаче напряжения треугольной формы мы получаем три различных модулированных колебания, форма которых показана на рис. 2.2.

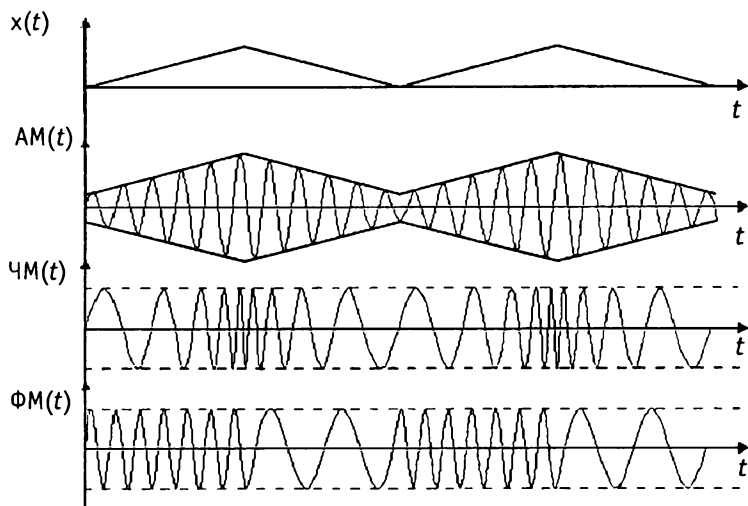


Рис. 2.2. Электрический сигнал треугольной формы $x(t)$ и модулированные этим сигналом по амплитуде $AM(t)$, частоте $ЧМ(t)$ и начальной фазе $ФМ(t)$ колебания

2.2.5. Испытательные сигналы

Рассмотрим несколько испытательных сигналов, необходимость в существовании которых обусловлена потребностями испытания характеристик разрабатываемых электронных устройств.

Гармоническое колебание. Самым распространенным испытательным сигналом является гармоническое колебание, которое используется в измерительной практике для оценки частотных свойств устройств различного назначения. При этом, например, на входы разных экземпляров одного и того же устройства подается один и тот же гармонический сигнал и регистрируются отклики каждого устройства на этот сигнал.

Единичный скачок. Другим испытательным сигналом является единичный скачок (функция Хевисайда)

$$l(t) = \begin{cases} 1 & \text{при } t \geq 0, \\ 0 & \text{при } t < 0. \end{cases} \quad (2.1)$$

Единичный скачок представляет собой безразмерную величину, поэтому умножение сигнала $s(t)$ на функцию единичного скачка равносильно включению этого сигнала в момент $t = 0$

$$s(t)l(t) = \begin{cases} s(t) & \text{при } t \geq 0, \\ 0 & \text{при } t < 0. \end{cases}$$

Это свойство единичной функции часто используется при описании как ограниченных, так и неограниченных во времени сигналов.

Дельта-функция. По определению δ -функция удовлетворяет следующим условиям:

$$\delta(t) = \begin{cases} \infty & \text{при } t = 0, \\ 0 & \text{при } t \neq 0 \end{cases} \quad (2.2)$$

и

$$\int_{-\infty}^{\infty} \delta(t) dt = 1. \quad (2.3)$$

Таким образом, δ -функция имеет размерность $[c^{-1}]$ и равна нулю при всех отличных от нуля значениях аргумента, принимая в точке $t = 0$ бесконечно большое значение. Площадь под кривой, ограниченной δ -функцией, равна единице.

Графики рассмотренных испытательных сигналов показаны на рис. 2.3.

П р и м е ч а н и е. Функция единичного скачка $l(t)$ и δ -функция не являются функциями в обычном математическом смысле. Они определяются в теории обобщенных функций последовательностями регулярных функций и все операции над ними рассматриваются как операции над этими последовательностями.

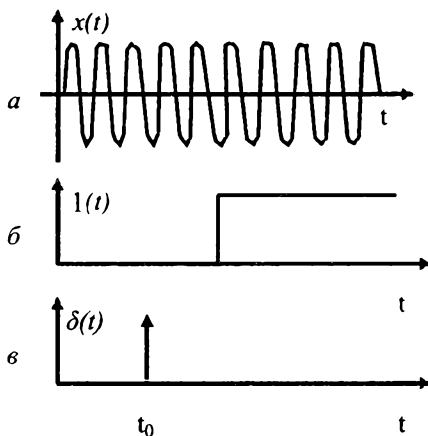


Рис. 2.3. Графическое представление испытательных сигналов типа гармонического колебания (а), единичного скачка $l(t)$ (б) и дельта-функции $\delta(t)$ (в)



Пример 2.1. Вычислить интеграл

$$I = \int_a^b s(t) \delta(t - t_0) dt, \quad a < t_0 < b.$$

Решение. Представим δ -функцию как предел прямоугольного импульса $g(t)$ длительностью Δ и амплитудой

$1/\Delta$ при $\Delta \rightarrow 0$: $\delta(t) = \lim_{\Delta \rightarrow 0} g(t)$. Тогда

$$I = \int_a^b s(t) \lim_{\Delta \rightarrow 0} g(t - t_0) dt = \lim_{\Delta \rightarrow 0} \int_a^b s(t) g(t - t_0) dt.$$

Так как $a < t_0 < b$, то

$$I = \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} \int_{c-\Delta/2}^{c+\Delta/2} s(t) dt.$$

Если функция $s(t)$ непрерывна в точке $t = t_0$, то по теореме о среднем

$$\int_{c-\Delta/2}^{c+\Delta/2} s(t) dt = \Delta s(t_0 + \xi \frac{\Delta}{2}),$$

где $-1 < \xi < 1$. Поэтому $I = \lim_{\Delta \rightarrow 0} s(t_0 + \xi \frac{\Delta}{2}) = s(t_0)$.

Итак, при $a < t_0 < b$ и непрерывности функции $s(t)$ в точке $t = t_0$

$$\int_a^b s(t) \delta(t - t_0) dt = s(t_0). \quad (2.4)$$

Это равенство описывает фильтрующее свойство δ -функции, т. е. δ -функция действует на функции $s(t)$ как временной фильтр, выбирающий из всех ее значений лишь одно в точке $t = t_0$ соответствующей обращению в нуль аргумента δ -функции. Иными словами, интеграл от произведения произвольной функции $s(t)$, ограниченной в интервале времени (a, b) , на δ -функцию $\delta(t - t_0)$ равен значению функции $s(t)$ в точке $t = t_0$.

Итак, мы показали, что формула (2.4) следует из определения δ -функции с помощью выражений (2.2) и (2.3). При строгом подходе δ -функцию следует определять выражением (2.4), так как формулы (2.2) и (2.3) не определяют однозначную функцию. Формула (2.4) определяет δ -функцию с помощью ее интегральных свойств, а не как функцию времени для каждого значения t .

Очевидно, что

$$\int_{-\infty}^t \delta(t) dt = \begin{cases} 1 & \text{при } t \geq 0, \\ 0 & \text{при } t < 0, \end{cases}$$

откуда, учитывая (2.1), приходим к выводу

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

или

$$\delta(t) = \frac{d1(t)}{dt}.$$

Следовательно, с помощью понятия δ -функции можно выразить производную от разрывной функции в точке ее разрыва.

Помимо рассмотренных примеров испытательных сигналов в различных областях радиоэлектроники используется большое число других сигналов, предназначенных для оценки характеристик электронных устройств. Например, в телевидении используется испытательный сигнал, формирующий цветные полосы на экране кинескопа, сигналы испытательной таблицы, сетчатого поля, шахматного поля и т. д.

2.2.6. Понятие цифровой последовательности

Формально последовательность записывается следующим образом:

$$x = \{x(n)\}, \quad -\infty < n < \infty.$$

Более удобно говорить о «последовательности $x(n)$ ».

Примеры последовательностей, играющих роль испытательных сигналов в системах обработки цифровой информации:

$$\delta(n) = \begin{cases} 1, & n = 0; \\ 0, & n \neq 0; \end{cases}$$

а) единичный импульс:

$$u(n) = \begin{cases} 1, & n \geq 0; \\ 0, & n < 0; \end{cases}$$

б) единичная ступенчатая последовательность:

$$u(n) = \begin{cases} 1, & n \geq 0; \\ 0, & n < 0; \end{cases}$$

в) экспоненциальная последовательность (действительная):

$$x(n) = a^n, \quad a = e^{-aT}, \quad e^{at} \Big|_{t=nT} = e^{-anT} = a^n,$$

где a — действительное число;

г) косинусоидальная (синусоидальная) последовательность:

$$x(n) = A \cos(\omega_0 n + \varphi);$$

$$x(n) = A \cos(\Omega_0 t + \varphi) \Big|_{t=nT} = A \cos(\Omega_0 nT + \varphi) = A \cos(\omega_0 n + \varphi),$$

где $\omega_0 = \Omega_0 T$ — цифровая круговая частота, которая имеет ограниченный диапазон изменения: $0 \leq \omega_0 \leq 2\pi$ или $-\pi \leq \omega_0 \leq \pi$;

д) комплексная экспоненциальная последовательность:

$$e^{(\alpha + i\omega)n} = e^{\alpha n} (\cos \omega n + i \sin \omega n).$$

Видим, что сигналы могут быть как действительными, так и комплексными, т. е. состоящими из действительной и мнимой частей. Комплексное представление сигнала используется не только для упрощения математического аппарата анализа сигналов, но и является вполне естественным при представлении сигнала с синфазной (совпадающей по фазе с некоторым иным сигналом) и квадратурной (отличающейся по

фазе от этого сигнала на 90°) составляющими сигнала. Использование комплексных чисел позволяет совместить в одном числе два разных, но так, чтобы можно было оперировать с ними одновременно. Комплексные числа удобно представлять точками на плоскости. Тогда их сложение представляет собой сложение векторов, а умножение соответствует повороту вектора на плоскости. Последовательное возведение комплексного числа в степень соответствует вращению вектора вокруг начала координат.

Когда на вход системы поступает сигнал в виде экспоненциальной последовательности z^n , (z — произвольное комплексное число), это соответствует испытательному сигналу, содержащему две квадратурные составляющие.

Последовательность $x(n)$ по определению называется *периодической* с периодом N , если $x(n) = x(n+N)$ для всех n . Например,

$$e^{i\omega_0 n} = e^{i\omega_0 (n+N)} = e^{i\omega_0 n} e^{i\omega_0 N}.$$

Очевидно, если $e^{i\omega_0 N} = 1$, то $\omega_0 N = 2\pi k$, k — целое число.

Если отношение $2\pi/\omega_0 = N$ представляет собой целое число, то N является периодом функции $e^{i\omega_0 n}$. Если $2\pi/\omega_0$ является рациональным числом, то период этой функции будет больше, чем $2\pi/\omega_0$. Если же $2\pi/\omega_0$ — иррациональное число, то функция $e^{i\omega_0 n}$ представляет собой непериодическую функцию n .

Графическое представление последовательности

$x(n) = a_4 \delta(n+4) + a_3 \delta(n+3) + a_1 \delta(n-1) + a_3 \delta(n-3)$ показано на рис. 2.4.

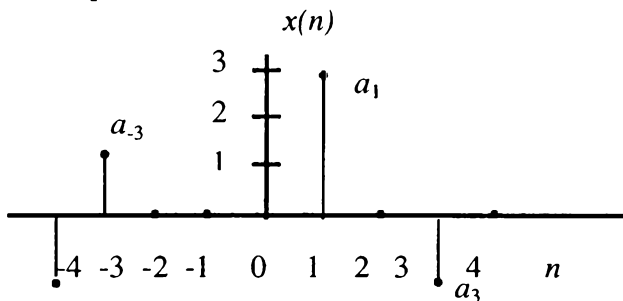


Рис. 2.4. Пример последовательности, представляющей собой сумму взвешенных и задержанных δ -импульсов



Пример 2.2. Вычислить цифровую круговую частоту последовательности с параметрами $T_0 = 1/12$ с и $T = 1/116$ с, где T — период следования временных отсчетов, а T_0 — период повторения последовательности.

Решение. Цифровая круговая частота получается равной

$$\omega_0 = \Omega_0 \cdot T = 2\pi \cdot 12 / 116 = 6\pi / 29 \text{ рад} = 3 \cdot 2\pi / 29.$$

Это значит, что в трех периодах косинусной функции $\cos(2\pi t / T_0)$ помещается ровно 29 интервалов дискретизации T .

Соответствующая косинусная последовательность

$$\begin{aligned} \cos(\Omega n) &= \cos[\Omega(n + i \cdot N)], \\ i &= 0, \pm 1, \pm 2, \dots \end{aligned} \quad (2.5)$$

является периодической в n с периодом $N = 29$.

Последовательность

$$x(n) = \cos(n) \quad (2.6)$$

является непериодической, так как нормированная круговая частота $\Omega = 1$ не является рациональной, кратной 2π (рис. 2.5).

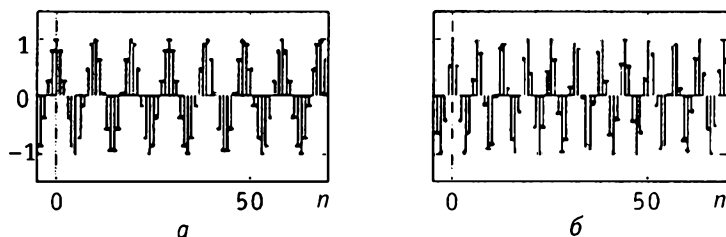


Рис. 2.5. Периодическая косинусная последовательность по (2.5) (а) и непериодическая косинусная последовательность по (2.6) (б)

Энергия последовательности $x(n)$ рассчитывается по формуле $E = \sum_{n=-\infty}^{\infty} |x(n)|^2$.

2.3. Случайные сигналы

Все реальные процессы наиболее точно описываются непрерывными функциями. Однако при практическом анализе случайных колебаний и сигналов используют математические модели, опирающиеся на дискретное представление как аргументов, так и функций. Применение той или иной модели определяется необходимой степенью детализации различных сторон явления. Наиболее характерные примеры представления случайных процессов дает следующая классификация.

Все случайные сигналы можно разделить на два типа: стационарные (в узком и широком смысле) и нестационарные. Применительно к задачам радиоэлектроники мы рассмотрим следующие четыре подхода к описанию случайных сигналов.

Непрерывный случайный процесс описывается непрерывной функцией непрерывного аргумента. Примером такого процесса может служить флуктуационный шум на выходе любого радиоприемного устройства даже в отсутствие принимаемого полезного сигнала. Уровень непрерывного случайного сигнала может принимать любое значение в заданных пределах.

Случайный процесс, непрерывный по времени и квантованный по уровню, отличается от непрерывного тем, что его мгновенные значения не образуют непрерывного ряда. Примером может служить ток через телеграфный ключ, с помощью которого формируется последовательность посылок и пауз. Если содержание телеграммы заранее неизвестно, то мгновенное значение тока в любой из моментов можно рассматривать как случайное, и для его описания необходимо использовать вероятностные характеристики этого тока.

Случайный процесс, дискретный по времени и непрерывный по уровню, т. е. дискретный поток непрерывных случайных величин. Примером такого процесса может служить последовательность видеоимпульсов, амплитуда которых изменяется от импульса к импульсу по случайному модулирующему закону, причем модулирующая кривая представляет непрерывный случайный процесс.

Случайный процесс, дискретный по времени и квантованный по уровню, т. е. дискретный поток дискретных случайных величин. Пример такого процесса — передача округленных квантованных показаний измерительных приборов в системах телеизмерений.

2.3.1. Гауссовские случайные процессы

Гауссовский процесс — это наиболее распространенный тип случайного процесса. Он представляет собой предельный вид случайного процесса, к которому приближаются другие виды случайных процессов при некоторых условиях, часто выполняющихся на практике. Особенность гауссовских процессов — их n -мерная функция распределения (плотность распределения) вероятности любой мерности определяется математическим ожиданием и корреляционной функцией. Для гауссовских процессов аппарат анализа является наиболее простым и хорошо разработанным.

2.3.2. Белый шум

Случайный процесс, обладающий равномерным энергетическим спектром, называют *белым шумом* по аналогии с белым светом, имеющим в видимой части равномерный сплошной спектр.

Несмотря на то что белый шум — неосуществимая на практике абстракция (его дисперсия, т. е. средняя мощность, бесконечна), ценность этого понятия и для теории, и для практики велика. Дело в том, что объектом внимания в прикладных дисциплинах являются не столько сами по себе случайные процессы, сколько реакция на них тех или иных физических систем. Любая же реальная система обладает конечной полосой пропускания. Поэтому произвольный случайный процесс (шум со спектром, равномерным в пределах этой полосы), действующий в этой системе, можно заменить белым шумом, не видя погрешности, поскольку добавленные внеполосные составляющие никакого воздействия на систему не окажут.

Наибольший интерес при анализе различных радиоэлектронных систем имеет аддитивный белый гауссовский шум (АБГШ). Аналитическое выражение для нормальной плотности в общем случае имеет вид:

$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

где μ — среднее значение, σ — среднее квадратическое отклонение. Идеальный шум представляет собой последовательность коротких импульсов, имеющих случайную высоту и следующих

друг за другом случайно. Спектральная плотность идеального шума, т. е. мощность шума, приходящаяся на конечную полосу частот, бесконечно мала. Для анализа процессов в области реальных положительных частот используют одностороннюю спектральную плотность N_0 , Вт/Гц. При теоретическом анализе поведение радиоэлектронных систем рассматривают в области положительных и отрицательных частот. В этом случае вводят двустороннюю спектральную плотность $N_0/2$, Вт/Гц. В обоих случаях мощность шума остается одной и той же. Постоянство спектральной плотности означает, что в бесконечно широкой полосе частот его мощность бесконечно велика, т. е. такое свойство шума является математической идеализацией. Однако практически полоса пропускания системы всегда ограничена, что автоматически ограничивает мощность шума. Поэтому *реальный* белый шум соответствует *идеальному* белому шуму, прошедшему через фильтр.

Обычно при расчетах мощность реального белого шума N в полосе частот W вычисляется по формуле $N = N_0 \cdot W$.

2.3.3. Окрашенный шум

Вводят понятие *окрашенного* шума. Например, белый шум со значением спектральной плотности на нулевой частоте, равной нулю, называется *зеленым* шумом, а шум с частотным спектром, являющимся плоским в логарифмическом представлении, имеет одинаковую мощность в полосе, пропорциональной ее ширине, и называется *розовым*. Это значит, что розовый шум, занимающий полосу частот от 40 до 60 Гц, имеет ту же мощность, что и шум, занимающий полосу частот от 4000 до 6000 Гц. Плотность мощности розового шума по сравнению с плотностью мощности белого шума уменьшается на 3 дБ на октаву, т. е. плотность мощности пропорциональна $1/f$.

Коричневый (или *красный*) шум очень похож на розовый шум и отличается от него только тем, что его плотность мощности уменьшается на 6 дБ на октаву с увеличением частоты (плотность шума пропорциональна $1/f^2$). При этом частота $f=0$ не входит в состав этого шума. Этот шум может генерироваться алгоритмом, который имитирует броуновское движение и потому не намекает на коричневый цвет, а представ-

ляет собой искажение от *Brownian motion*. Известен также как *random walk* или *drunkard's walk* шум.

Плотность мощности *голубого* шума увеличивается на 3 дБ на октаву с ростом частоты в ограниченной области частот, а плотность мощности *фиолетового* шума возрастает на 6 дБ на октаву, т. е. пропорционально f^2 , в ограниченной частотной области. Этот шум представляет собой результат дифференцирования белого шума.

Серый шум представляет собой шум, вид плотности мощности которого подчиняется закону, совпадающему с кривой психоакустической одинаковой громкости в пределах данной частотной области, т. е. звучит такой шум одинаково громко на всех частотах. *Черный* шум имеет несколько определений, одно из которых соответствует тишине. Существуют также «менее официальные» определения окрашенных шумов, однако телекоммуникационный словарь Федерального стандарта США 1037C определяет только белый, розовый, голубой и черный шумы.

2.4. Квантование сигнала по уровню

Квантование сигнала по уровню — это замена сигнала $s(t)$, изменяющегося от $s_{\min}$ до $s_{\max}$, счетным множеством значений его уровней $s_0, s_1, s_2, \dots, s_N$, отстоящих друг от друга на величину $\Delta s = (s_{\max} - s_{\min})/N$, называемую *шагом квантования*. Эти уровни можно перенумеровать, например, двоичными числами, количество разрядов которых

$$n \geq \log_2 N. \quad (2.7)$$

В результате квантования сигнала $s(t)$ получается новый сигнал $s'(t)$, имеющий ступенчатую форму (рис. 2.6). При таком преобразовании сигнала появляется *ошибка квантования*, характеризующаяся величиной шага квантования, так как мгновенное значение квантуемой величины может оказаться между соседними уровнями квантования. Часто точность выполнения операции квантования характеризуют относительной ошибкой квантования

$$\delta = \Delta s / s_{\max} = (s_{\max} - s_{\min}) / N s_{\max}. \quad (2.8)$$

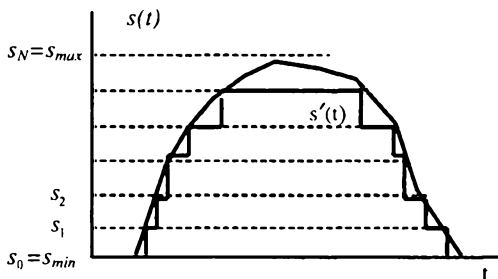


Рис. 2.6. Квантование сигнала по уровню

Из (2.7) видно, что чем больше величина N , тем меньше сигнал $s'(t)$ отличается от сигнала $s(t)$. С увеличением N возрастает число разрядов n представления двоичных чисел, что ведет к усложнению аппаратуры для обработки таких сигналов. Тем не менее, квантование сигнала по уровню широко используется в микроэлектронных устройствах цифровой техники.



Пример 2.3. Определить число разрядов двоичных чисел для измерения изменяющегося от 0 до 100 В напряжения U с погрешностью δ , не превышающей 0,5 В.

Решение. Шаг квантования $\Delta U = 2 \delta = 2 \cdot 0,5 = 1$ В. Количество уровней квантования $N = U_{max} / \Delta U = 100$.

По формуле (2.7) разрядность двоичных чисел

$$n \geq \log_2 N = 7.$$

2.5. Спектральный (гармонический) анализ сигналов

2.5.1. Понятие гармонического анализа

Гармонический анализ — это раздел математики, который изучает возможности представления функций в виде тригонометрических рядов и интегралов. Основным понятием в гармоническом анализе является гармоническое колебание, которое математически можно записать следующим образом:

$$s(t) = A \cos(\Omega t + \varphi),$$

где A , Ω , φ — соответственно амплитуда, циклическая частота, начальная фаза колебания.

В гармоническом анализе вводится понятие n -й гармоники гармонического колебания частоты Ω , под которой понимают опять же гармоническое колебание с частотой, в n раз превышающей частоту основного гармонического колебания. Математически выражение для n -й гармоники $s_n(t)$ основного тона Ω можно записать следующим образом:

$$s_n(t) = A_n \cos(n\Omega t + \varphi_n),$$

где A_n , $n\Omega$, φ_n — амплитуда, циклическая частота, начальная фаза n -й гармоники основного тона Ω соответственно.

Следующим важным понятием является спектр сигнала. Под спектром сигнала понимают совокупность его гармонических составляющих. Введение понятия спектра сигнала обусловило использование в технических приложениях название спектрального анализа для гармонического анализа сигналов.

2.5.2. Спектральный анализ периодических сигналов

Тригонометрическая форма ряда Фурье

Основа спектрального анализа периодических сигналов — ряд Фурье для периодической функции $s(t)$: практически всякую периодическую функцию $s(t)$ можно представить в виде ряда Фурье:

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\Omega t + b_n \sin n\Omega t), \quad (2.9)$$

где $\frac{a_0}{2} = \frac{1}{T} \int_{-T/2}^{T/2} s(t) dt$ соответствует постоянной составляющей сигнала, а $a_n = \frac{2}{T} \int_{-T/2}^{T/2} s(t) \cos n\Omega t dt$ и $b_n = \frac{2}{T} \int_{-T/2}^{T/2} s(t) \sin n\Omega t dt$ представляют амплитуды косинусоидальных и синусоидальных гармоник сигнала соответственно.

На практике удобнее пользоваться представлением функции $s(t)$ в виде суммы только косинусоидальных гармоник. Если в формуле (2.9) выражение, стоящее под суммой, умножить и разделить на $\sqrt{a_n^2 + b_n^2}$, то получим

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(\frac{a_n \cos n\Omega t}{\sqrt{a_n^2 + b_n^2}} + \frac{b_n \sin n\Omega t}{\sqrt{a_n^2 + b_n^2}} \right) \sqrt{a_n^2 + b_n^2}.$$

Так как

$$\frac{a_n}{\sqrt{a_n^2 + b_n^2}} \leq 1, \quad \frac{b_n}{\sqrt{a_n^2 + b_n^2}} \leq 1, \\ \left(\frac{a_n}{\sqrt{a_n^2 + b_n^2}} \right)^2 + \left(\frac{b_n}{\sqrt{a_n^2 + b_n^2}} \right)^2 = 1,$$

то можно обозначить

$$\frac{a_n}{\sqrt{a_n^2 + b_n^2}} = \cos \varphi_n \quad \text{и} \quad \frac{b_n}{\sqrt{a_n^2 + b_n^2}} = \sin \varphi_n.$$

И тогда можно записать

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \sqrt{a_n^2 + b_n^2} \cos(n\Omega t - \varphi_n),$$

$$\varphi_n = \operatorname{arctg}(b_n / a_n).$$

Введя обозначения $c_n = \sqrt{a_n^2 + b_n^2}$, $c_0 = a_0$, окончательно получим

$$s(t) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos(n\Omega t - \varphi_n). \quad (2.10)$$

Из формул (2.9) и (2.10) следует, что спектр периодического сигнала дискретен. Спектр периодического сигнала состоит из отдельных частот, соответствующих дискретным частотам $0, \Omega, 2\Omega, 3\Omega, \dots$ Важно отметить, что если периодический сигнал следует с частотой, например, 1 кГц , то в его спектре могут быть только частоты, кратные частоте 1 кГц , т. е. $0, 1 \text{ кГц}, 2 \text{ кГц}$ и т. д. Это означает, что в спектре такого периодического сигнала не может быть частот, отличных от вышеперечисленных, например частоты $1,5 \text{ кГц}$ или $1,2 \text{ кГц}$.

Графически спектр сигнала можно представить следующим образом. По оси абсцисс откладывают частоты гармонических составляющих сигнала, а по оси ординат — величины амплитуд этих гармоник. Для исчерпывающей характеристики спектра это графическое изображение дополняют заданием начальной фазы каждой гармонической составляющей спектра. Это обстоятельство и определяет название спектра периодического сигнала: *линейчатый*, или *дискретный*, спектр.

Рис. 2.7 дает общее представление об амплитудно-частотной (АЧХ) и фазочастотной (ФЧХ) характеристиках периодического сигнала, из которого видно, что частоты спектральных составляющих кратны частоте следования периодического сигнала.

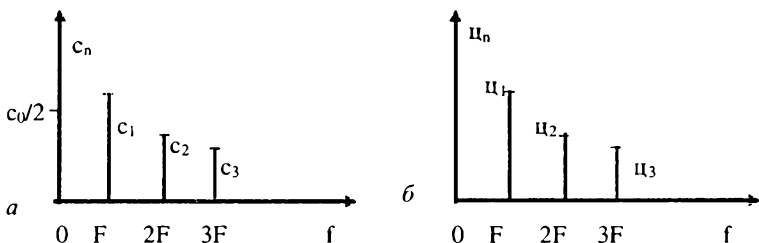


Рис. 2.7. АЧХ (а) и ФЧХ (б) периодического сигнала

Таким образом, структура спектра периодического сигнала полностью определяется его АЧХ и ФЧХ, т. е. амплитудами и начальными фазами спектральных составляющих. Графическое изображение спектра дает наглядное представление об относительной величине отдельных составляющих, позволяя легко оценить, например, техническую ширину спектра сигнала.

Комплексная форма ряда Фурье

В радиотехнических устройствах сигнал подвергается различным формам обработки. Для того чтобы рассчитать результат этих преобразований, используется комплексная форма ряда Фурье. Эти преобразования имеют формальный характер и связаны с применением комплексной формы для представления функции вещественной переменной. Поэтому при проведении расчетов реальный сигнал можно представить в виде функции комплексной переменной, преобразовав его соответствующим образом, а затем снова перейти к вещественному представлению.

Для вывода ряда Фурье в комплексной форме за основу возьмем уравнение (2.10). Учитывая, что

$$\cos(n\Omega t - \varphi_n) = \frac{e^{i(n\Omega t - \varphi_n)} + e^{-i(n\Omega t - \varphi_n)}}{2},$$

подставим это выражение в (2.10). Получим

$$s(t) = \frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \frac{e^{i(n\Omega t - \varphi_n)} + e^{-i(n\Omega t - \varphi_n)}}{2} =$$

$$= \frac{c_0}{2} + \frac{1}{2} \sum_{n=1}^{\infty} c_n e^{i(n\Omega t - \varphi_n)} + \frac{1}{2} \sum_{n=1}^{\infty} c_n e^{-i(n\Omega t - \varphi_n)}.$$

Преобразуем третье слагаемое следующим образом:

$$\frac{1}{2} \sum_{n=1}^{\infty} c_n e^{-i(n\Omega t - \varphi_n)} = \frac{1}{2} \sum_{n=-1}^{-\infty} c_{-n} e^{i(n\Omega t - \varphi_{-n})}.$$

Теперь все три слагаемых можно записать в виде одной формулы

$$s(t) = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} c_n e^{i(n\Omega t - \varphi_n)} = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} c_n e^{-i\varphi_n} e^{in\Omega t} = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} C_n e^{in\Omega t}, \quad (2.11)$$

где

$$C_n = a_n - ib_n = \frac{2}{T} \int_{-T/2}^{T/2} s(t) \cos n\Omega t dt - i \frac{2}{T} \int_{-T/2}^{T/2} s(t) \sin n\Omega t dt =$$

$$= \frac{2}{T} \int_{-T/2}^{T/2} s(t) e^{-in\Omega t} dt.$$

Входящая в (2.11) комплексная амплитуда C_n связана с a_n и b_n очевидными соотношениями

$$C_n = c_n e^{-i\varphi_n} = a_n - i b_n,$$

$$C_{-n} = c_n e^{i\varphi_n} = a_n + i b_n.$$

Отсюда видно, что комплексные амплитуды C_n и C_{-n} являются взаимно сопряженными комплексными величинами.

Нетрудно убедиться, что функция φ_n нечетная относительно n , а следовательно, и частоты, т. е. $\varphi_n = -\varphi_{-n}$. Четность модуля c_n вытекает непосредственно из его определения.

Отрицательные частоты, которые вводятся для представления реального сигнала $s(t)$ комплексным рядом Фурье, носят формальный характер и не имеют никакого физического смысла.

Сравним величины коэффициентов в выражениях (2.10) и (2.11) разложения функции $s(t)$ в ряд Фурье при тригонометрической и комплексной формах представления ряда. Срав-

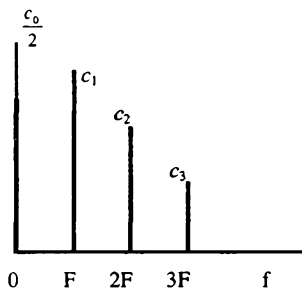


Рис. 2.8. АЧХ спектра, полученная с помощью ряда Фурье в тригонометрической форме

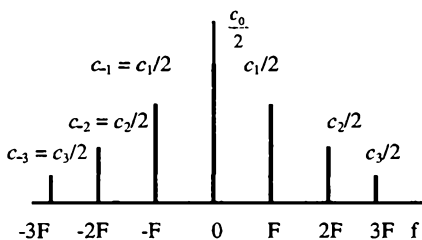


Рис. 2.9. АЧХ спектра, полученная с помощью ряда Фурье в экспоненциальной форме

нивая выражения (2.10) и (2.11), обнаруживаем, что при $n = 0$ имеет место равенство коэффициентов, рассчитанных по каждому из этих выражений. Сравнение же коэффициентов, рассчитываемых по каждому из этих рядов при других значениях n , приводит к следующему выводу. Коэффициент c_n , определяемый из ряда (2.11), составляет только половину такого же коэффициента, определяемого из ряда (2.10). Рис. 2.7 и 2.8 показывают, что при переходе от тригонометрической к экспоненциальной форме ряда Фурье (с искусственным введением в рассмотрение отрицательных частот) уменьшаются в два раза амплитудные значения коэффициентов разложения. Для полного представления о спектре сигнала в виде ряда (2.11) его АЧХ должна быть дополнена заданием ФЧХ, которая получается точно такой же, как и в представлении сигнала тригонометрическим рядом (2.10), но с расширением ее в область отрицательных частот.



Пример 2.4. Найти спектр последовательности прямоугольных импульсов, имеющих амплитуду E , длительность τ и следующих с периодом T .

Решение. Запишем анализируемый импульс в виде (рис. 2.10)

$$s(t) = \begin{cases} E, & -\tau/2 \leq t \leq \tau/2, \\ 0, & \tau/2 < t < T/2, \quad -T/2 < t < -\tau/2. \end{cases} \quad (2.12)$$

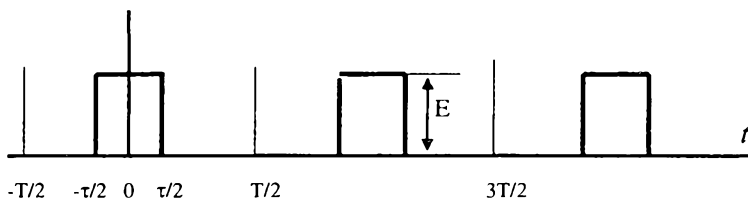


Рис. 2.10. Последовательность прямоугольных импульсов, следующих с периодом T и описываемых четной функцией $s(t)$

Так как функция $s(t)$ четная, то

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} s(t) \sin n\Omega t dt = 0.$$

Тогда функцию $s(t)$ можно представить в виде ряда

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos n\Omega t.$$

Найдем величину постоянной составляющей сигнала:

$$\frac{a_0}{2} = \frac{1}{T} \int_{-\tau/2}^{\tau/2} E dt = \frac{E\tau}{T}.$$

Эта формула показывает, что постоянная составляющая исследуемого сигнала зависит от отношения длительности импульса τ к его периоду следования T . Чем короче длительность импульса по отношению к периоду следования, тем меньше уровень постоянной составляющей сигнала. Полученная формула правильно оценивает также постоянную составляющую импульса с длительностью, равной периоду их следования ($\tau = T$). Величина $q = T/\tau$ называется *скважностью* последовательности импульсов.

Теперь найдем амплитуды косинусоидальных составляющих сигнала

$$\begin{aligned} a_n &= \frac{2}{T} \int_{-\tau/2}^{\tau/2} E \cos n\Omega t dt = \frac{2E}{T} \frac{1}{n\Omega} \int_{-\tau/2}^{\tau/2} \cos n\Omega t d(n\Omega t) = \\ &= \frac{E}{\pi n} \sin n\Omega t \Big|_{-\tau/2}^{\tau/2} = \frac{2E}{\pi n} \sin \frac{n\Omega\tau}{2} = \frac{2E\tau}{T} \frac{\sin \frac{\pi n\tau}{T}}{\frac{\pi n\tau}{T}}. \end{aligned} \quad (2.13)$$

Здесь было учтено, что $\Omega = 2\pi / T$. Вторым сомножителем в выражении для a_n представляет собой функцию $\sin x / x$, график которой показан на рис. 2.11 и является огибающей амплитуд спектральных составляющих исследуемого сигнала. *График этой функции совпадает с огибающей амплитуд всех гармоник спектра, за исключением его нулевой гармоники (постоянной составляющей).*

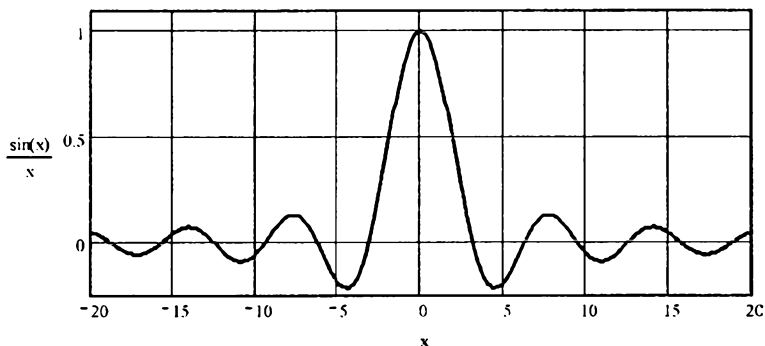


Рис. 2.11. График функции $\sin x/x$

Рассмотрим, как зависит распределение спектральных составляющих сигнала от отношения τ/T . Амплитуды спектральных

составляющих обращаются в нуль при $\sin \frac{\pi n \tau}{T} = 0$, т. е.

при значениях аргумента синуса, равных $\pi, 2\pi, 3\pi, \dots$ Определим номер гармоники, амплитуда которой первой обращается в нуль. Это произойдет при фазе синуса, равной π ,

т. е. при выполнении условия $\frac{\pi n \tau}{T} = \pi$, откуда $n = T / \tau$. На-

пример, если мы имеем дело с последовательностью импульсов длительностью 1 мкс, следующих с периодом 4 мкс, то первой спектральной составляющей, амплитуда которой обратится в нуль, будет четвертая гармоника. Чтобы определить номер следующей спектральной составляющей, амплитуда которой обратится в нуль, надо составить уравнение, соответствующее фазе синуса, равной 2π , т. е. записать $\frac{\pi n \tau}{T} = 2\pi$. Отсюда видно, что такой будет восьмая гармоника спектра сигнала (рис. 2.12).

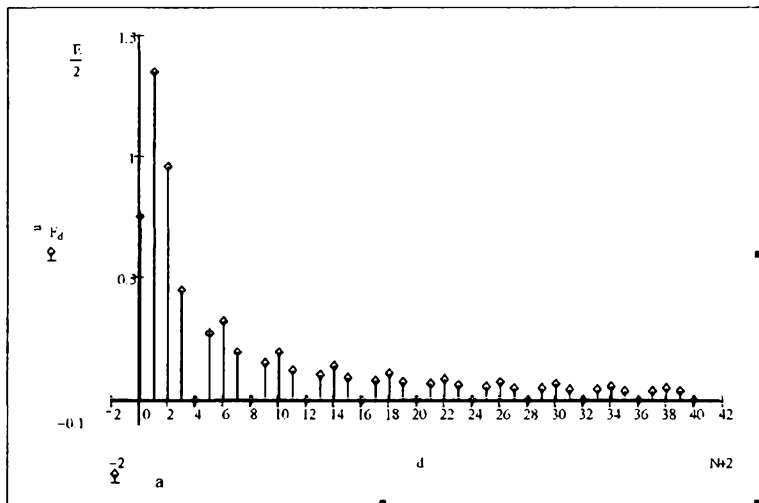


Рис. 2.12. Спектр последовательности прямоугольных импульсов с отношением периода следования импульсов к их длительности, равным 4

Если уменьшить длительность импульса вдвое, сохранив при этом период следования, то первой составляющей спектра, обратившейся в нуль, станет восьмая гармоника, а следующей — шестнадцатая (рис. 2.13).

Итак, получается наглядное представление об АЧХ последовательности прямоугольных импульсов, заданной тригонометрическим рядом Фурье. Чтобы получить исчерпывающую информацию о спектре этой последовательности, нужно конкретизировать ее ФЧХ. Поскольку все коэффициенты $b_n = 0$, то фазы гармоник спектральных составляющих сигнала $s(t)$ будут определяться начальной фазой косинуса соответствующей гармоники спектра. Мы видим, что фазы всех гармоник от первой до первой, обращаемой в нуль (восьмой), равны нулю. Фазы гармоник, начиная с девятой и заканчивая шестнадцатой, равны π . Далее фазы гармоники становятся равными 2π , 3π и т. д. На рис. 2.13, б показана ФЧХ этой последовательности.

Отметим, что когда мы будем получать информацию о ФЧХ с помощью пакета Mathcad, то получим ее в виде, представленном на рис. 2.13, в.

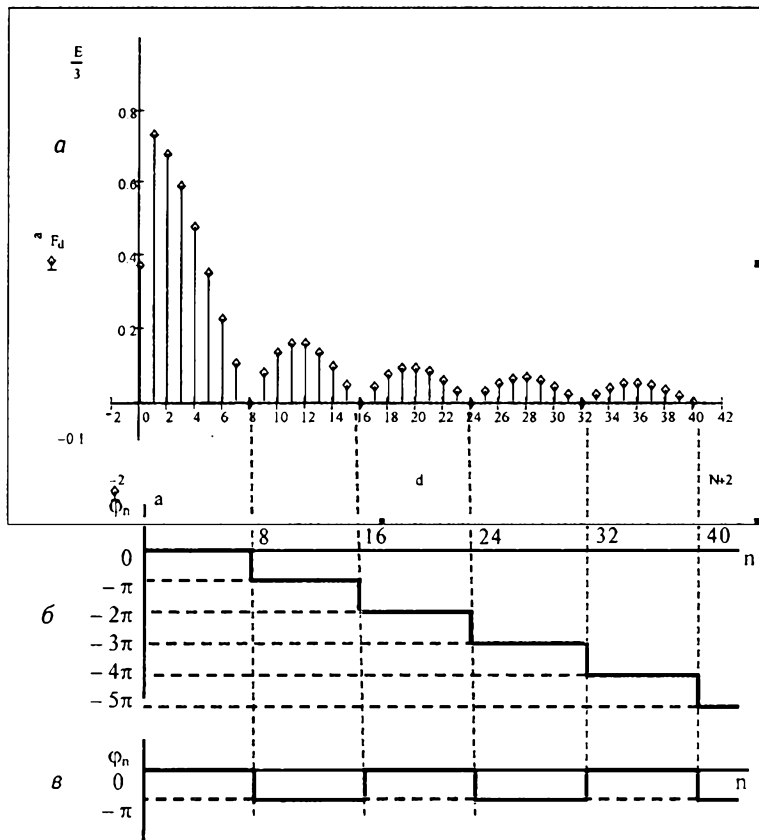


Рис. 2.13. АЧХ (а), ФЧХ (б) последовательности прямоугольных импульсов с отношением периода следования импульсов к их длительности, равным 8, и ФЧХ, которая получается в результате ее расчета в среде Mathcad (в)

Дело в том, что ЭВМ при вычислениях не различает величин фаз $0, 2\pi, 4\pi$, и будет выдавать, если не принять специальных мер, одно и то же значение начальной фазы, равное нулю. Это же относится и к изображению начальной фазы, принимающей значения $\pi, 3\pi, \dots$ Программа *Mathcad* будет выдавать для всех перечисленных значений начальной фазы гармоник единственное значение, равное π . Поэтому при

анализе спектров в системе *Mathcad* всегда надо контролировать выдаваемые машиной результаты и правильно их интерпретировать.



Пример 2.5. Дать наглядное представление о спектре рассмотренной в примере 2.4 последовательности, представив ее экспоненциальным рядом Фурье.

Решение. При переходе к экспоненциальной форме ряда Фурье представления реального сигнала мы ввели отрицательные частоты, которых в природе нет. В результате этого мы изображаем амплитуду гармоники $|C_n|$ двумя амплитудами $|C_n|/2$ и $|C_{-n}|/2$, расположенными симметрично относительно нулевой частоты. При этом устраняется нарушение стройности представления огибающей спектра анализируемой последовательности, так как теперь функция $\sin x/x$ описывает огибающую всех гармоник, включая и амплитуду постоянной составляющей. Таким образом, представление спектра последовательности прямоугольных импульсов экспоненциальным рядом Фурье становится изящнее (рис. 2.14, а). Что касается ФЧХ этого сигнала, то она сохраняет вид, который мы получили при представлении сигнала тригонометрическим рядом Фурье, и показана на рис. 2.14, б.

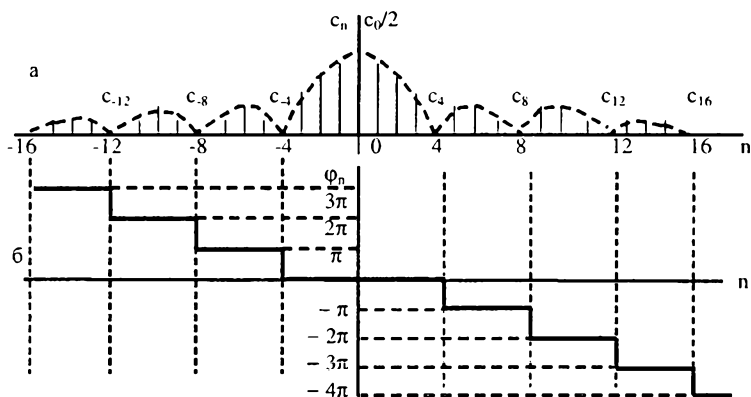


Рис. 2.14. АЧХ (а) и ФЧХ (б) последовательности прямоугольных импульсов при представлении сигнала экспоненциальным рядом Фурье

Мы видим, что с уменьшением длительности импульса возрастает номер его спектральной составляющей, амплитуда которой первой обращается в нуль. Это обстоятельство хорошо иллюстрирует соотношение для простых сигналов $\Delta f_{cn} = 1/\tau$, связывающее техническую ширину спектра Δf_{cn} с длительностью сигнала τ . Кроме того, с уменьшением длительности сигнала относительно периода следования спектр сигнала становится более насыщенным, что, несомненно, должно учитываться, например при разработке хороших электронных музыкальных инструментов.

2.5.3. Спектральный анализ непериодических сигналов

Для проведения гармонического анализа непериодического сигнала $s(t)$, действующего в конечном интервале $t_1 < t < t_2$, поступим следующим образом. Превратим наш непериодический сигнал $s(t)$ в периодический путем повторения его с периодом $T > t_2 - t_1$. Теперь для этого сигнала можно записать ряд Фурье. Устремив период T к бесконечности, в пределе получим бесконечно малые амплитуды гармонических составляющих, сумма которых изображает исходную непериодическую функцию $s(t)$, заданную в интервале $t_1 < t < t_2$.

Число гармонических составляющих, входящих в ряд Фурье для функции $s(t)$, будет при этом бесконечно большим, так как при $T \rightarrow \infty$ частота первой гармоники функции $\Omega = 2\pi/T \rightarrow 0$. Это значит, что расстояние между спектральными линиями, равное первой гармонике сигнала, становится бесконечно малым и спектр становится сплошным.

Отсюда следует, что при гармоническом анализе непериодического сигнала получается сплошной спектр, состоящий из бесконечно большого числа гармоник с бесконечно малыми амплитудами.

Математически это можно выразить следующим образом (рис. 2.15).

Полагая, что наш непериодический сигнал является периодическим, запишем для него ряд Фурье в комплексной форме:

$$s(t) = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} C_n e^{in\Omega t}.$$

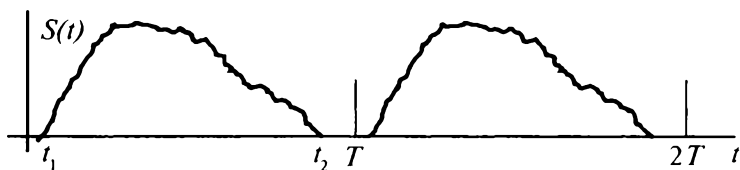


Рис. 2.15. Одиночный (непериодический) сигнал длительностью $t_2 - t_1$, представленный как периодический с периодом $T \geq t_2 - t_1$

Подставим в этот ряд значение $\hat{c}_n$:

$$s(t) = \frac{1}{2} \sum_{n=-\infty}^{\infty} \left(\frac{2}{T} \int_{-T/2}^{T/2} s(t) e^{-in\Omega t} dt \right) e^{in\Omega t} =$$

$$\frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \left(\int_{-T/2}^{T/2} s(t) e^{-n\Omega t} dt \right) e^{in\Omega t} \Omega.$$

Теперь устремим период T следования сигнала к бесконечности. При этом сигнал из периодического превращается в сигнал с бесконечно большим периодом, т. е. сигнал непериодический. Суммирование можно заменить интегрированием в бесконечных пределах. Частоту $n\Omega$ заменим текущей частотой ω , а частоту Ω — приращением текущей частоты $d\omega$. Тогда получим

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{i\omega t} d\omega, \quad (2.14)$$

где внутренний интеграл

$$S(\omega) = \int_{-\infty}^{\infty} s(t) e^{-i\omega t} dt \quad (2.15)$$

называется *спектральной плотностью*, или *спектральной характеристикой* сигнала $s(t)$.

Выражение (2.14) представляет непериодический сигнал в виде суммы (интеграла) гармонических колебаний с бесконечно малыми амплитудами. Сравнивая выражение (2.14) с выражением (2.13) ряда Фурье для периодического сигнала, мы видим, что амплитуды составляющих непериодического сигнала равны $\frac{1}{\pi} S(\omega) d\omega$.

Выясним смысл названия спектральной плотности. Для этого выделим какую-либо дискретную частоту $\Omega_n = n\Omega$, соответствующую в случае периодической функции $s(t)$ n -й гармонике, и запишем для амплитуды этой гармоники выражение

$$C_n = \frac{2}{T} \int_{t_1}^{t_2} s(t) e^{-i\Omega_n t} dt.$$

В случае же непериодической функции, совпадающей с $s(t)$ в интервале $t_1 < t < t_2$, получим для спектральной плотности, соответствующей той же частоте Ω_n , следующее выражение:

$$S(\Omega_n) = \int_{t_1}^{t_2} s(t) e^{-i\Omega_n t} dt.$$

Отсюда видно, что

$$S(\Omega_n) = T \frac{C_n}{2} = \frac{C_n}{F}.$$

Таким образом, значение спектральной плотности $S(\Omega_n)$ получается путем деления амплитуды C_n (поскольку одной действительной амплитуде спектральной составляющей сигнала соответствуют две спектральные составляющие с амплитудами $C_n/2$ и $C_{-n}/2$) гармоники на полосу частот F , отделяющую соседние линии дискретного спектра, т. е. $S(\Omega)$ имеет смысл *плотности амплитуд* и размерность [амплитуда/герц].

Из полученного выражения вытекает также следующее важное положение: *огibaющая сплошного спектра (модуль спектральной плотности) непериодической функции и огibaющая линейчатого спектра периодической функции, представленной экспоненциальным рядом Фурье (полученной из непериодической путем продолжения ее с периодом T), совпадают по форме и отличаются только масштабом и, конечно, размерностью.*

Однако это положение не распространяется на постоянную составляющую периодического сигнала: спектральное представление непериодического прямоугольного импульса описывается функцией $\sin x/x$ без особенностей, которые имеют место при расчете постоянной составляющей спектра последовательности этих импульсов.



Пример 2.6. Оценить влияние грозových разрядов на прием сигналов радиовещательных станций, работающих в диапазонах ДВ, СВ, КВ, УКВ.

Решение. В качестве модели тока, формирующего грозовой разряд, возьмем уравнение $i(t) = I_0 e^{-\alpha t}$, где $\alpha = 10^4 \dots 10^5 \text{ c}^{-1}$ характеризует скорость убывания тока грозового разряда (рис. 2.16). Чтобы оценить влияние грозового разряда на прием вещательных станций, оценим спектр тока, формирующего этот разряд, и сделаем соответствующие выводы.

Итак, имеем

$$S(\omega) = \int_0^{\infty} i(t) e^{-i\omega t} dt = \int_0^{\infty} I_0 e^{-\alpha t} e^{-i\omega t} dt =$$

$$= \frac{I_0}{-(\alpha + i\omega)} \int_0^{\infty} e^{-(\alpha + i\omega)t} d[-(\alpha + i\omega)t] = \frac{I_0}{\alpha + i\omega}.$$

Поскольку для оценки влияния грозových разрядов на прием сигналов станций, работающих в диапазонах ДВ, СВ, КВ, УКВ, основное значение имеет амплитудный спектр излучения грозового разряда, то запишем

$$S(\omega) = \frac{I_0}{\sqrt{\alpha^2 + \omega^2}}.$$

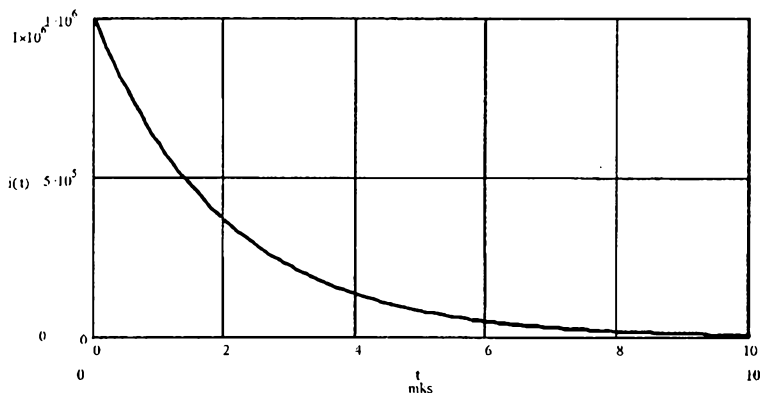


Рис. 2.16

Оценим частоту, на которой мощность излучения грозового разряда уменьшается вдвое. Это легко сделать, составив уравнение

$$\frac{I_0}{\sqrt{\alpha^2 + \omega_{zp}^2}} = \frac{I_0}{\alpha \sqrt{2}}.$$

Из этого уравнения легко находим $\omega_{zp} = \alpha = 10^4 \dots 10^5 \text{ с}^{-1}$. Это значит, что на частоте

$$f_{zp} = \omega_{zp} / 2\pi = (10^4 \dots 10^5) / 2\pi \approx 10^3 \dots 10^4 \text{ Гц}$$

мощность излучения грозового разряда составит половину от максимальной в начальный момент и с увеличением частоты будет уменьшаться. Этот результат дает возможность заключить, что наиболее сильно будет влиять излучение грозовых разрядов на прием сигналов станций, работающих в диапазоне ДВ, менее значительно оно будет сказываться на приеме сигналов радиостанций, работающих в диапазоне СВ и КВ, и почти никак не будет проявляться в диапазоне УКВ.

2.6. Дискретизация сигналов

Операция замены непрерывного сигнала $s(t)$ последовательностью его значений (выборок) в соответствии с теоремой Котельникова называется *дискретизацией*.

Теорема Котельникова. Если функция $s(t)$ не содержит частот выше f_g , то она полностью определяется последовательностью своих значений в моменты, отстоящие друг от друга на $\Delta t = \pi / \omega_g = 1 / 2f_g$.

Дополним функцию $S(\omega)$ значениями, симметричными относительно нуля значениям на участке $0 \dots \omega_g$, и повторим ее с периодом $2\omega_g$. Тогда функцию $S(\omega)$ можно будет представить рядом Фурье

$$S(\omega) = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} c_n e^{in \frac{2\pi}{2\omega_g} \omega} = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} c_n e^{in \frac{\pi}{\omega_g} \omega}, \quad (2.16)$$

где

$$c_n = \frac{2}{2\omega_g} \int_{-\omega_g}^{\omega_g} S(\omega) e^{-in \frac{2\pi}{2\omega_g} \omega} d\omega = \frac{1}{\omega_g} \int_{-\omega_g}^{\omega_g} S(\omega) e^{-in \frac{\pi}{\omega_g} \omega} d\omega. \quad (2.17)$$

С другой стороны,

$$s(t) = \frac{1}{2\pi} \int_{-\omega_g}^{\omega_g} S(\omega) e^{i\omega t} d\omega = \frac{1}{2\pi} \int_{-\omega_g}^{\omega_g} S(\omega) e^{-in\frac{\pi}{\omega_g}\omega} d\omega = s(-n\Delta t). \quad (2.18)$$

Отсюда получаем:

$$c_n = \frac{2\pi}{\omega_g} s(-n\Delta t). \quad (2.19)$$

Таким образом, коэффициенты ряда Фурье функции $S(\omega)$ определяются через значения функции $s(t)$ в точках отсчета.

Запишем ряд Фурье, определяющий спектральную плотность $S(\omega)$, по коэффициентам $\hat{c}_n$

$$\begin{aligned} S(\omega) &= \frac{1}{2} \sum_{n=-\infty}^{\infty} \frac{2\pi}{\omega_g} s(-n\Delta t) e^{in\frac{\pi}{\omega_g}\omega} = \frac{\pi}{\omega_g} \sum_{n=-\infty}^{\infty} s(-n\Delta t) e^{in\frac{\pi}{\omega_g}\omega} = \\ &= \frac{\pi}{\omega_g} \sum_{n=-\infty}^{\infty} s(n\Delta t) e^{-in\frac{\pi}{\omega_g}\omega}. \end{aligned} \quad (2.20)$$

Полученное выражение означает, что функция $s(t) = s(n\Delta t)$, которой соответствует спектральная плотность $S(\omega)$, однозначна и проходит через заданные значения в точках отсчета $n\Delta t$, лежащих на расстоянии $\Delta t = 1/2f_g$ одна от другой.

Запишем выражение для функции $s(t)$, подставив в (2.18) значение $S(\omega)$:

$$\begin{aligned} s(t) &= \int_{-\omega_g}^{\omega_g} S(\omega) e^{i\omega t} d\omega = \frac{1}{2\pi} \int_{-\omega_g}^{\omega_g} \frac{\pi}{\omega_g} \left(\sum_{n=-\infty}^{\infty} s(n\Delta t) e^{-in\frac{\pi}{\omega_g}\omega} \right) e^{i\omega t} d\omega = \\ &= \frac{1}{2\omega_g} \int_{-\omega_g}^{\omega_g} \left(\sum_{n=-\infty}^{\infty} s(n\Delta t) \right) e^{i\omega(t - n\frac{\pi}{\omega_g})} d\omega = \sum_{n=-\infty}^{\infty} s(n\Delta t) \frac{\sin \omega_g(t - n\Delta t)}{\omega_g(t - n\Delta t)}. \end{aligned} \quad (2.21)$$

Если функция $s(t)$ конечна в интервале T , а спектр ее ограничен полосой частот вплоть до ω_g , то число дискретных значений (отсчетов) функции $s(n\Delta t)$, которое необходимо для полного задания ее в интервале T , будет равно

$$n = T / \Delta t + 1 = 2f_g T + 1 \approx 2f_g T, \quad (2.22)$$

так как $T/\Delta t \gg 1$.

Выражение для $s(t)$ при этом принимает вид

$$s(t) = \sum_{n=-f_s T}^{n=f_s T} s(n\Delta t) \frac{\sin \omega_s (t - n\Delta t)}{\omega_s (t - n\Delta t)}. \quad (2.23)$$

Принципиальное отличие дискретизации с использованием ряда Котельникова заключается в том, что отсчеты мгновенных значений сигнала производятся непосредственно в процессе его поступления на вход преобразующего устройства, а для представления в виде ряда Фурье необходимо предварительно полностью зарегистрировать сигнал. Большим преимуществом ряда Котельникова является простота определения его коэффициентов. Полученные в процессе дискретизации мгновенные значения сигнала могут быть равны любой величине в диапазоне изменения сигнала $s(t)$ от s_{min} до s_{max} .



Пример 2.7. Найти минимальное число отсчетов, с помощью которых можно передать информацию, содержащуюся в прямоугольном видеоимпульсе длительностью $\tau = 1$ мкс.

Решение. Поскольку спектр видеоимпульса является бесконечно широким, то ограничимся его технической шириной, определяемой частотой, соответствующей первому нулю. Эта частота равна $f_g = 1/\tau$. Поэтому будем считать, что анализируемый видеоимпульс обладает финитным спектром с наивысшей частотой $f_g = 10^6$ Гц = 1 МГц. В соответствии с требованиями теоремы Котельникова расстояние между отсчетами нужно выбирать равным $1/2 f_g = 0,5$ мкс. Это значит, что достаточно трех выборок, чтобы передать информацию, содержащуюся в прямоугольном видеоимпульсе. Однако лучше выбрать пять отсчетов. При этом возрастет точность передачи передаваемой информации за счет ее количественного увеличения вследствие появления избыточности.



Пример 2.8. Изобразить спектральную диаграмму последовательности $n = 5$ отсчетов, соответствующих дискретизированному видеоимпульсу длительностью 1 с.

Решение. Прежде всего отметим, что в отличие от спектра исходного видеоимпульса, который считается финитным и ограниченным верхней частотой f_g , спектр последовательности отсчетов является бесконечным, содержит гармоники,

кратные частоте $1/\Delta t$, и имеет период $n\Delta t$, а огибающая описывается функцией, напоминающей $\sin x/x$, (рис. 2.17). Из графика следует, что первый нуль соответствует частоте $f = 1/n\Delta t$, где $n\Delta t$ соответствует длительности дискретизируемого импульса. Расстояние между нулевыми точками на оси частот соответствует частоте f . Таким образом, спектр дискретизованного сигнала по главному лепестку совпадает со спектром исходного континуального сигнала, ограниченного по частоте интервалом до первого нуля. Максимальные величины пиков составляют nU_0 , т. е. в n раз превышают амплитуду исходного сигнала. Расстояние по оси частот между основными всплесками спектра дискретизованного сигнала соответствуют величине $1/\Delta t$, т. е. определяется периодом следования дискретизирующих импульсов. Кроме того, размерность спектра дискретизованного сигнала $[B \cdot c]$ отличается от размерности дискретизованного сигнала $[B]$. Это соответствует ситуации, когда отсчеты формируются с помощью δ -импульсов. Однако на практике моделью δ -импульса часто служит прямоугольный видеоимпульс достаточно малой длительности. Если допустить, что длительность отсчетного импульса составляет 0,04 длительности исходного импульса, т. е. 40 мс, то его огибающая будет накладываться на бесконечный периодический спектр. При этом частота, соответствующая первому нулю огибающей спектра импульса длительностью 40 мс, равна 25 Гц (рис. 2.18).

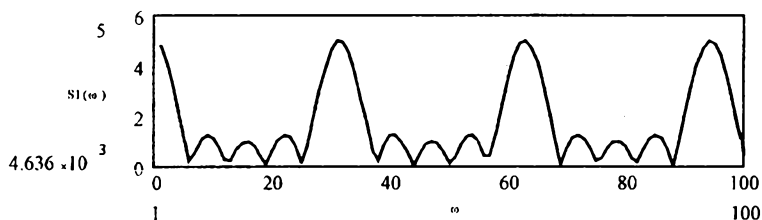


Рис. 2.17. Огибающая спектра дискретизованного сигнала, полученного из видеосигнала длительностью 1 мкс, с помощью пяти выборок, представляющих собой δ -импульсы

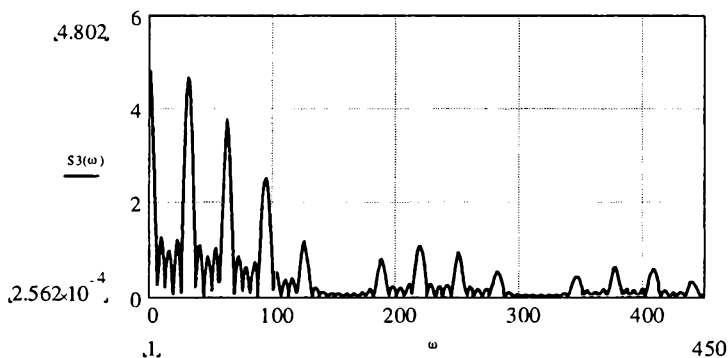


Рис. 2.18. Огибающая спектра дискретизированного сигнала, полученного из видеосигнала длительностью 1 с помощью пяти выборок, представляющих собой импульсы длительностью 40 мс

2.7. Скорость передачи информации по каналу связи и скорость модуляции

Если сигнал изменяется так, что равновероятно можно различать только 2 составляющие его информационного параметра, то, в соответствии с оценкой Р. Хартли, любое изменение сигнала соответствует наименьшей единице информации — биту.

Если же сигнал может иметь более двух различных изменений, то любое его изменение несет несколько бит информации.

Количество изменений информационного параметра сигнала в секунду измеряется в *Бодах (Baud)*. Промежуток времени между соседними изменениями информационного параметра сигнала называется тактом работы передатчика, или *бодовым интервалом*.

При использовании кодированных сигналов нужно различать скорость передачи информации, измеряемой количеством бит, передаваемых по каналу связи в одну секунду, и скорость модуляции, представляющей собой скорость генерации сигнальных посылок и измеряемой в *Бод/с*.

Связь между скоростью модуляции D и скоростью передачи битов по каналу связи $R = 1/t_b$, где t_b — длительность передачи одного бита, характеризуется соотношением: $D = R/b$, где b — число бит в посылке.

Если сигнал имеет более 2 различных состояний, то скорость передачи в *бит/с* будет выше, чем число *Бод*. Например, если информационными параметрами являются фаза и амплитуда синусоиды, причем различаются 4 состояния фазы в 0° , 90° , 180° , 270° и два значения амплитуды сигнала, то информационный сигнал в этом случае может иметь 8 различных состояний. В этом случае модем, работающий со скоростью модуляции, равной 2400 *Бод/с* (с тактовой частотой 2400 *Гц*), передает информацию со скоростью 7200 *бит/с*, так как при одном изменении сигнала передается 3 бита информации.

При использовании сигнала с двумя различными состояниями может наблюдаться обратная картина. Это часто происходит потому, что для надежного распознавания приемником передаваемой информации каждый бит в последовательности кодируется с помощью нескольких изменений информационного параметра несущего колебания. Например, при кодировании единичного значения бита импульсом положительной полярности, а нулевого значения — импульсом отрицательной полярности, физически сигнал дважды изменяет свое состояние при передаче каждого бита. При таком кодировании скорость передачи в 2 раза ниже, чем число *бод*, передаваемых по линии.

В 1924 г. *Harold Nyquist* открыл функциональную зависимость между полосой пропускания $f_{\text{ш}}$ канала связи и максимальной скоростью модуляции, измеряемой в *Бод/с*, при передаче информации по этому каналу связи. *Nyquist* установил, что скорость модуляции в *Бод/с* не может превышать более чем в два раза полосу пропускания канала.

Без учета шума в канале связи зависимость скорости передачи информации, измеряемой в *бит/с*, определяется выражением:

$$C = 2f_{\text{ш}} \log_2 m, \quad (2.24)$$

где m — число различных состояний информативного параметра сигнала.

Таким образом, для телефонной линии с полосой пропускания 3000 Гц скорость модуляции не может быть выше чем 6000 Бод/с.

В схеме «1 бит за 1 Бод» скорость передачи информации в бит/с равна скорости модуляции в Бод/с. Многие схемы кодирования сигналов способны передавать более одного бита в течение одного сигнализирующего события. Поэтому в таких системах «несколько бит за 1 Бод» скорость передачи информации превышает скорость модуляции в Бод/с.

Хотя формула Найквиста явно не учитывает наличие шума, косвенно его влияние отражается в выборе числа состояний информационного параметра сигнала. Для повышения пропускной способности канала хотелось бы увеличивать это число до значительных величин, но на практике этого делать нельзя из-за шума на линии. Поэтому число возможных состояний сигнала фактически ограничивается отношением мощности сигнала к мощности шума, а формула Найквиста определяет предельную скорость передачи данных в том случае, когда количество состояний уже выбрано с учетом возможностей устойчивого распознавания приемником.

В 1948 г. *Cloude Shannon* указал, что теорема Найквиста применима только к каналам связи с отсутствием шума. В реальном мире таких каналов нет. Шум устанавливает ограничения, но не на скорость модуляции в Бод/с, а на количество бит, которые могут быть закодированы одним сигнализирующим событием. Из теоремы Шеннона следует, что для обычной телефонной линии (с отношением сигнал/шум, равным приблизительно 30 дБ) максимальное количество бит, которое может быть закодировано одним сигнализирующим событием, равно 5. Следовательно, максимальная скорость передачи информации, достигаемая при таком ограничении, приблизительно составляет 30 Кбит/с (передается 6000 Бод по 5 бит, передаваемых за 1 Бод).

Выбор способа представления дискретной информации в виде сигналов, подаваемых на вход линии связи, называется *линейным* или *физическим кодированием*. *Логическое кодирование* выполняется до физического и подразумевает замену бит исходной информации новой последовательностью бит, несущей ту же информацию, но обладающей дополнительными свойствами, например возможностью для приемной стороны

обнаруживать ошибки в принятых данных. Другой пример логического кодирования — шифрование данных. Логическое кодирование уменьшает скорость передачи информации по линии связи.

2.8. Семейство преобразований Фурье

Рассмотренные выше сигналы представляли собой две их разновидности: периодические и непериодические. При этом каждой разновидности сигнала соответствует свое преобразование Фурье. Если сигнал $s(t)$ представляет собой периодический процесс с периодом T , то его можно представить в виде ряда Фурье, имеющим разные формы представления (2.9)–(2.11), коэффициенты которого рассчитываются по соответствующим формулам.

Если сигнал является непериодическим, его можно представить в виде преобразований Фурье (2.14) и (2.15).

Строго говоря, понятия «ряд Фурье» и «преобразование Фурье» различаются.

Преобразование Фурье предполагает, что частоты заданы непрерывно.

Ряд Фурье записывается на дискретном наборе частот.

Кроме того, мы можем говорить о непрерывном или дискретном времени, т. е. сигнал может представляться как непрерывной функцией времени, так и в виде своих временных отсчетов. Это дает 4 варианта преобразований:

- преобразование Фурье с непрерывным временем;
- преобразование Фурье с дискретным временем;
- ряд Фурье с непрерывным временем;
- ряд Фурье с дискретным временем.

С точки зрения цифровой обработки сигналов наиболее практична дискретизация и во временной, и в частотной области, но она является только аппроксимацией непрерывного преобразования.

Непрерывное преобразование Фурье точно представляет любые, не обязательно периодические явления. Если же приходится ограничиваться дискретным набором частот, то представимый рядом Фурье сигнал может быть только периодическим, и сигналы, не обладающие свойством периодичности, можно представить в этом случае только приближенно.

Для этого сигнал мысленно надо продолжить, как бы превращая в периодический. На стыках отрезков сигнала при этом могут возникать разрывы и изломы сигнала, что приводит к необходимости ввода окон, сглаживающих эти явления.

Дискретное преобразование Фурье (ДПФ) оперирует с дискретной во времени выборкой периодического сигнала во временной области. Чтобы его можно было представить в виде суммы синусоид, сигнал должен быть периодическим. Но в качестве набора входных данных для ДПФ доступно только конечное число отсчетов. Эту ситуацию можно разрешить, если мысленно поместить бесконечное число одинаковых групп отсчетов до и после обрабатываемой группы, образуя математическую, но не реальную периодичность, как это показано на рис. 2.19.

Существует два основных типа ДПФ: вещественное и комплексное. Так как входные отсчеты во временной области являются вещественными и не имеют мнимой части, мнимая часть входных отсчетов всегда принимается равной нулю.

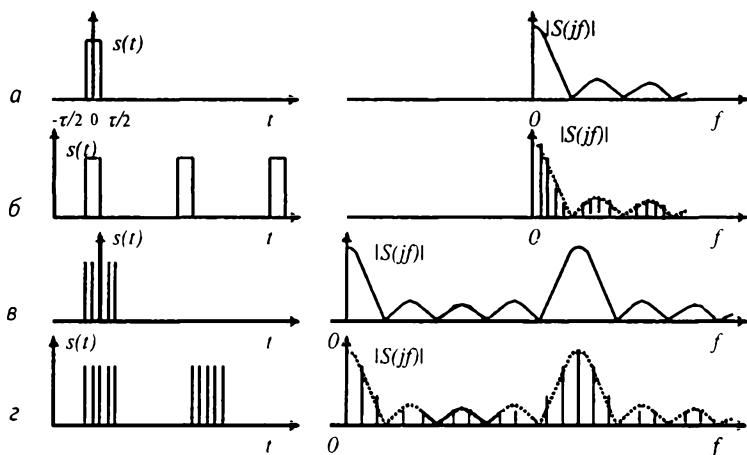


Рис. 2.19. Семейство преобразований Фурье.

Сигналы непрерывный и аperiodический (преобразование Фурье) (а), непрерывный и периодический (ряды Фурье) (б), дискретный и аperiodический (дискретные ряды Фурье) (в), дискретный и периодический (дискретные ряды Фурье) (з) и модули их спектральных характеристик

Выход ДПФ содержит вещественную и мнимую компоненты, которые могут быть преобразованы в амплитуду и фазу соответствующего гармонического колебания. Вещественное ДПФ несколько проще комплексного. Если известны выходные частоты вещественного ДПФ и требуется использовать обратное комплексное ДПФ для вычисления отсчетов во временной области, то надо знать, как разместить выходные точки вещественного ДПФ в формате комплексного ДПФ перед выполнением обратного комплексного ДПФ. Комплексное ДПФ имеет вещественные и мнимые значения и на входе, и на выходе. Мнимые части отсчетов во временной области устанавливаются в ноль. При рассмотрении спектра, получаемого в результате вычисления комплексного ДПФ, полезно знать, как связать его с результатом вычисления вещественного комплексного ДПФ и наоборот.

2.8.1. Непрерывно-временной ряд Фурье

Применительно к нуждам цифровой обработки для сигнала $s(t)$, ограниченного по длительности интервалом $0 \leq t \leq NT$, спектральное представление иногда связывают с непрерывно-временным рядом Фурье

$$S(k) = \int_0^{NT} s(t) e^{-i2\pi kt / NT} dt, \quad (2.25)$$

$$s(t) = \frac{1}{NT} \sum_{k=-\infty}^{\infty} S(k) e^{i2\pi kt / NT},$$

который по существу на всей временной оси представляет периодически (с периодом NT) протяженную функцию $s(t)$, т. е.

$$s(t) = \sum_{k=-\infty}^{\infty} x(k) \delta(t - kNT) = \sum_{k=-\infty}^{\infty} x(t - kNT). \quad (2.26)$$

Таким образом, коэффициенты ряда Фурье представляют собой отсчеты непрерывного по частоте спектра дискретной последовательности точек.

2.8.2. Дискретно-временной ряд Фурье (ДВРФ)

Для последовательности $s(n)$ конечной длины в N отсчетов спектральные представления связываются с дискретно-временным рядом Фурье:

$$S(k) = T \sum_{n=0}^{N-1} s(n) e^{-i2\pi kn/N}, \quad 0 \leq k \leq N-1,$$

$$s(n) = \frac{1}{NT} \sum_{k=0}^{N-1} S(k) e^{i2\pi kn/N}, \quad 0 \leq n \leq N-1.$$

Обе последовательности $S(k)$ и $s(n)$ периодичны с периодом N и на интервале в один период представляют соответствующую совокупность отсчетов спектра $S(f)$ в точках $f_k = k/NT$ и отсчетов сигнала $s(t)$, в точках $t = nT$.

Эти соотношения называют еще парой дискретных преобразований Фурье (ДПФ). Их отличие от традиционно используемых определяется наличием множителей T и $1/T$, связанных с желанием обеспечить корректность масштабов при вычислении энергии и мощности анализируемых сигналов.

2.8.3. Дискретно-временное преобразование Фурье (ДВПФ)

В этом случае сигнал $s(t)$ представляется своими отсчетами непрерывной функции времени, взятыми с периодом T , $s(t) = s(t)|_{t=nT}$.

При этом спектр последовательности определяет периодическую (с периодом $1/T$) функцию частоты, которая является периодическим продолжением спектральной плотности $S(f)$ непрерывного сигнала $s(t)$.

Если $S(f)$ ограничена по частоте на оси частот, например в интервале $-f_g \leq f \leq f_g$, а период отсчетов удовлетворяет соотношению $T \leq (1/2) / f_g$, то спектры дискретной функции не перекрываются.

Следовательно, спектр дискретной последовательности $s(n)$ повторяет форму спектра функции $s(t)$, которая может быть восстановлена по ее отсчетам. В противном случае имеет место эффект *наложения* спектров, и восстановление функции $s(t)$ по ее отсчетам $s(n)$ окажется невозможным.

Когда представляемая *последовательность имеет конечную длительность*, т. е. имеет конечное число ненулевых значений, возможно еще одно представление в частотной области, называемое *дискретным преобразованием Фурье (ДПФ)*, которое также является дискретной последовательностью и *соответствует равноудаленным по частоте отсчетам ДВПФ*. Кроме своей теоретической важности, ДПФ играет центральную роль при реализации ряда алгоритмов обработки сигналов, поскольку разработаны эффективные алгоритмы вычисления ДПФ, получившие название алгоритмов *быстрого преобразования Фурье (БПФ)*.

Имеются различные подходы к выводу и интерпретации представления последовательности конечной длительности с помощью ДПФ.

Представление по Фурье периодических последовательностей важно само по себе и используется также для представления последовательностей конечной длины. Это делается путем создания периодической последовательности, каждый период которой совпадает с последовательностью конечной длины. Как будет показано далее, ряд Фурье этой периодической последовательности соответствует ДПФ последовательности конечной длины.

2.8.4. Представление периодических последовательностей дискретным рядом Фурье

Периодические последовательности $x(n)$ с периодом в N отсчетов, т. е. $x(n) = x(n + kN)$ для любого целого k , можно представить *рядом Фурье*:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i \frac{2\pi}{N} kn}, \quad (2.27)$$

т. е. суммой комплексных экспонент с частотами, кратными основной частоте $2\pi/N$ периодической последовательности:

$$e_k = e^{i \frac{2\pi}{N} kn}, \quad k = 0, 1, 2, \dots, N-1.$$

В отличие от ряда непрерывной периодической функции в этом представлении периодической последовательности есть только N различных комплексных экспонент $e_k(n)$ с периодом,

равным целой части основного периода N . Так получается вследствие того, что комплексная экспонента $e_k(n)$ периодична по k с периодом N .

Следовательно, множество N комплексных экспонент определяет все различные комплексные экспоненты с частотами, кратными основной частоте $2\pi/N$. Поэтому представление $x(n)$ рядом Фурье содержит только N комплексных экспонент.

Выразим коэффициенты $X(k)$ через значения последовательности $x(n)$, умножив (2.27) на $\exp(i 2\pi kn/N)$ и просуммировав результаты по n :

$$\begin{aligned} \sum_{n=0}^{N-1} x(n) e^{-i \frac{2\pi}{N} m} &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} X(k) e^{i \frac{2\pi}{N} (k-r)n} = \\ &= \sum_{k=0}^{N-1} X(k) \frac{1}{N} \sum_{n=0}^{N-1} e^{i \frac{2\pi}{N} (k-r)n} = X(r). \end{aligned} \quad (2.28)$$

Здесь было использовано соотношение

$$\frac{1}{N} \sum_{n=0}^{N-1} e^{i \frac{2\pi}{N} m} = \frac{1}{N} \begin{pmatrix} 1 - e^{i \frac{2\pi}{N} rN} \\ 1 - e^{i \frac{2\pi}{N} r} \end{pmatrix} = \begin{cases} 1, & r = mN, \quad m — \text{целое число;} \\ 0, & r \neq mN. \end{cases} \quad (2.29)$$

Таким образом, коэффициенты $X(k)$ в (2.27) определяются соотношением:

$$X(k) = \sum_{n=-\infty}^{\infty} x(n) e^{-i \frac{2\pi}{N} kn}. \quad (2.30)$$

Соотношения (2.27) и (2.30) являются парой преобразований, которые называются *дискретным рядом Фурье (ДРФ)* периодической последовательности. Эту последовательность, заданную во временной области своими эквидистантными значениями, представляют на множестве не непрерывных, а дискретных частот, т. е. последовательность коэффициентов ДРФ $X(k)$ представляет собою периодическую последовательность с периодом N : $X(k) = X(k + lN)$, l — целое число.

2.8.5. Представление по Фурье последовательности конечной длительности — дискретное преобразование Фурье

Пусть имеем последовательность $s(k)$ конечной длины N , для которой $s(k) = 0$ всюду, за исключением интервала $0 \leq k \leq N - 1$.

Иными словами, $s(k)$ представляет собою один период периодической последовательности, т.е. коэффициенты ДРФ образуют последовательность эквидистантных отсчетов ДВПФ последовательности $s(k)$ в точках $\omega_k = 2\pi k/N$.



Пример 2.9. Найти ДПФ прямоугольного видеоимпульса напряжения $s(t)$ амплитуды U_0 , представленного своими $N = 5$ отсчетами (рис. 2.20, а).

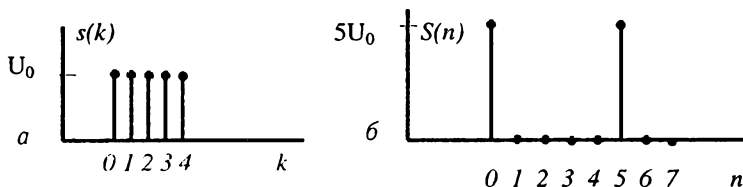


Рис. 2.20. Отсчеты прямоугольного видеоимпульса (а) и их ДПФ (б)

Для ДПФ сигнала $s(t)$ имеем

$$S(n) = U_0 \sum_{k=0}^{N-1} e^{-j\frac{2\pi}{N}nk} = U_0 \frac{1 - e^{-j2\pi n}}{1 - e^{-j2\pi n/N}} =$$

$$= \begin{cases} U_0 N, & n = 0, \pm N, \pm 2N, \dots, \\ 0 & \text{при других } n. \end{cases}$$

Полученный результат, показанный на рис. 2.20,б, можно истолковать как значения сплошного спектра дискретизованного сигнала на частотах $\omega_n = n2\pi/NT$, и в данном примере при n , кратных N , эти значения равны нулю. Таким образом, в пределах периода N лишь одна спектральная составляющая имеет амплитуду NU_0 .

Поэтому коэффициенты ДРФ однозначно представляют последовательность конечной длительности $s(k)$, так как по ним можно восстановить исходную последовательность.

Таким образом, ДРФ и ДПФ практически представляют собой одни и те же преобразования, которые можно использовать для представления как последовательности конечной длительности, соответствующей N отсчетам, так и ее периодического продолжения с периодом N на всю числовую ось.

2.8.6. ДПФ последовательности, дополненной нулями

Представление последовательностей конечной длительности в частотной области путем вычисления ДПФ можно использовать для получения значений преобразования Фурье в L точках ($L > N$), равномерно распределенных по единичной окружности.



Пример 2.10. Сигнал из примера 2.9 дополнен N нулевыми отсчетами (рис. 2.21, а). Найти ДПФ этого сигнала и сравнить его с результатом, полученным в примере 2.9.

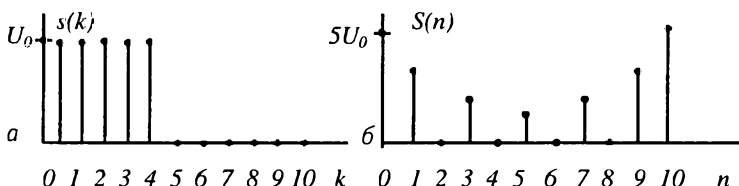


Рис. 2.21. Отсчеты, дополненные нулями (а), и их ДПФ (б)

Решение. При увеличении числа временных отсчетов с N до $2N$ добавлением N нулевых отсчетов интервал между соседними дискретными частотами сокращается вдвое. ДПФ сигнала с удвоенным периодом определяет и большее число спектральных коэффициентов, модули которых определяются выражением

$$|S(n)| = |U_0(1 - e^{-i\pi n})/(1 - e^{-i\pi n/N})| = |U_0 \cdot \sin(n\pi/2)/\sin(n\pi/2N)|,$$

представленным на рис. 2.21, б.

Таким образом, простое дополнение последовательности $s(k)$ нулевыми отсчетами до длины $L > N$ с последующим вычисле-

нием L -точечного ДПФ позволяет получать отсчеты ДВПФ для совокупности точек, равномерно распределенных по единичной окружности с частотным интервалом $2\pi/L$, т. е. при $L \gg 1$, сколь угодно малым.

Операция, состоящая в увеличении периода последовательности за счет добавления нулевых отсчетов, широко используется в практике спектрального анализа конечных последовательностей.

2.9.1. Уменьшение частоты дискретизации в целое число M раз

Эта процедура называется также *прореживанием*, или *децимацией* дискретного сигнала $x(n)$. При этом новый дискретный сигнал $x_d(n)$ получается прореживанием исходного, т. е. сохранением лишь каждого M -го отсчета сигнала $x(n)$: $x_d(n) = x(Mn)$. Такой сигнал может быть получен в результате периодической (с периодом $T_1 = MT$) дискретизации сигнала $x(t)$, т. е. с частотой дискретизации $F_{T_1} = F_T/M$. На рис. 2.22 представлено условное графическое отображение прореживателя в M раз.

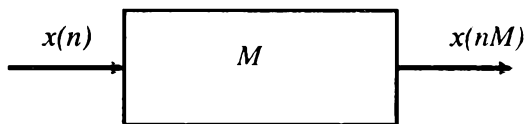


Рис. 2.22. Прореживатель в M раз

Запишем соотношение, связывающее спектры дискретного сигнала $x(n)$ и сигнала непрерывного времени $x(t)$, в виде:

$$X(e^{j\omega T}) = (1/T) \sum_{n=-\infty}^{\infty} X[j\omega(T - 2\pi n)/T]. \quad (2.31)$$

На рис. 2.23 на основе (2.31) представлены графически спектры дискретных сигналов $x(n)$ и $x_d(n)$ для $M = 2$. Чтобы гарантировать отсутствие эффекта наложения в связи с расширением спектра в M раз при прореживании дискретного сигнала, необходимо перед прореживанием пропустить его через дискретный фильтр нижних частот (ДФНЧ) с полосой пропускания $|\omega| \leq \pi/M$. Таким образом, совокупность каскад-

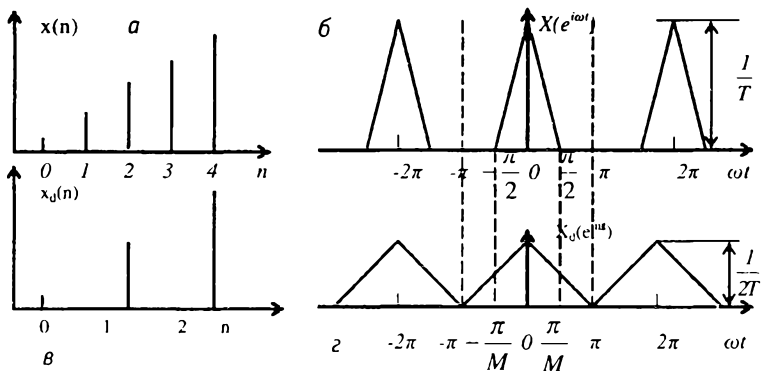


Рис. 2.23. Отсчеты сигнала $x(t)$, следующие с периодом T (а), спектр $X(e^{j\omega t})$ сигнала $x(n)$ (б), отсчеты прореженного сигнала $x_d(n)$ (в) и спектр $X_d(e^{j\omega t})$ сигнала $x_d(n)$ (г)

но включенных ДФНЧ с полосой пропускания $|\omega| \leq \pi/M$ и прореживателя в M раз выполнит операцию уменьшения частоты дискретизации в M раз без эффекта наложения. Такую систему (рис. 2.24) называют *компрессором частоты дискретизации*.

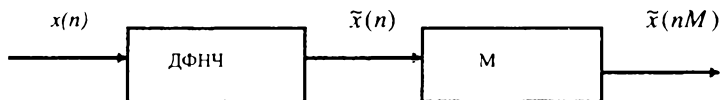


Рис. 2.24. Компрессор частоты дискретизации

2.9.2. Увеличение частоты дискретизации в целое число L раз

Эту процедуру называют *интерполяцией*, или восстановлением отсутствующих отсчетов дискретного сигнала. Для интерполяции дискретного сигнала $x(n)$ его сначала удлиняют в L раз путем вставки между каждыми двумя соседними отсчетами $L - 1$ нулевых отсчетов и получают сигнал $x_{ydl}(n)$.

Спектр сигнала $x_{ydl}(n)$ примет форму

$$\begin{aligned}
 X_{ydl}(e^{j\omega}) &= \sum_{n=-\infty}^{\infty} x_{ydl}(n) e^{-j\omega n} = \sum_{n=kL}^{\infty} x(n/L) e^{-j\omega n} = \\
 &= \sum_{k=-\infty}^{\infty} x(k) e^{-j\omega Lk} = X(e^{j\omega L}).
 \end{aligned}
 \quad (2.32)$$

Из (2.32) следует вывод о том, что при описанной выше операции удлинения дискретного сигнала спектр сигнала $x_{удл}(n)$ сжимается по оси частот в L раз в сравнении со спектром исходного сигнала, а период его повторения будет равен $2\pi/L$. Если затем пропустить $x_{удл}(n)$ через идеальный дискретный ФНЧ с полосой пропускания $|\omega| \leq \pi/M$ и усилением L , то лишние спектральные полосы будут удалены, а выходной сигнал фильтра будет точно соответствовать дискретному сигналу $x_{ин}$ с периодом дискретизации T/L , т. е. частота дискретизации его увеличится в L раз, а нужные отсчеты будут восстановлены. На рис. 2.25 представлены графически: отсчеты исходного сигнала (а) и спектр дискретизованного исходного сигнала (б), сигнал, удлинённый добавлением нулевых отсчетов (в) и его спектр (г), интерполированный сигнал на выходе ДФНЧ (д) и его спектр (е). На рис. 2.26 представлена структура интерполятора (*экспандера* частоты дискретизации), увеличивающего частоту дискретизации в L раз.

Теорема отсчетов для последовательностей описывает правила, согласно которым мы можем без потери информации «выбрасывать» некоторые ее части, т. е. прореживать цифровой сигнал. Операция прореживания называется еще *децимацией*, ее выполняет устройство, называемое *дециматором*. Обратная ей операция восстановления сигнала называется *интерполяцией*, ее реализует *интерполятор*. Децимация обозначается стрелкой, направленной вниз: $\downarrow$, а интерполяция — стрелкой, направленной вверх: $\uparrow$. Рядом со стрелкой ставят коэффициент децимации (интерполяции). Например, запись $2\downarrow$ означает, что из последовательности выбрасывается каждый второй отсчет, запись $2\uparrow$ означает вставку нуля между соседними отсчетами.

Если необходимо изменить частоту дискретизации исходного дискретного сигнала в L/M (рациональное число) раз, то такая операция может быть выполнена каскадным соединением интерполятора в L раз и дециматора в M раз, как это представлено на рис. 2.27. Два последовательно включенных дискретных фильтра нижних частот ДФНЧ1 и ДФНЧ2 могут быть заменены одним с меньшей частотой среза.

Рассмотрим примеры применения дециматоров и интерполяторов в цифровой аудиотехнике. При преобразовании в

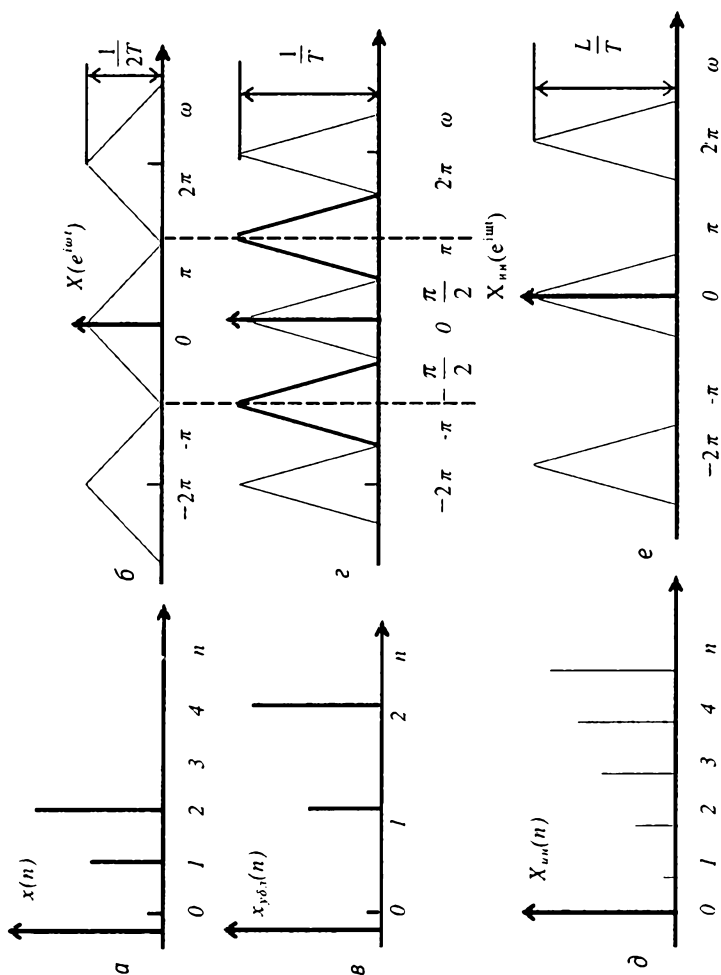
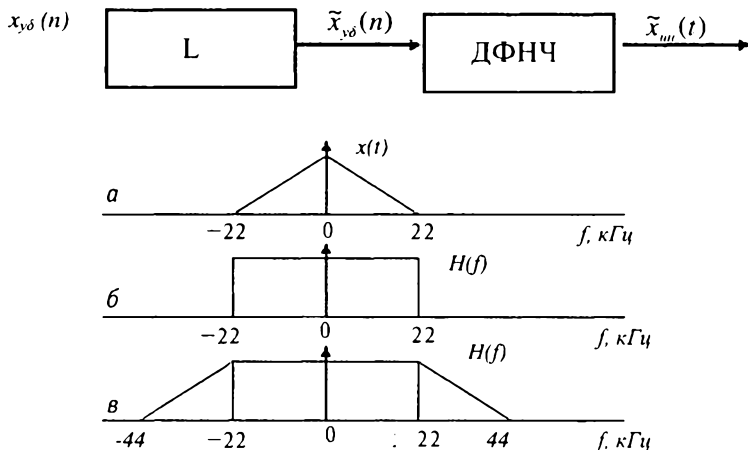
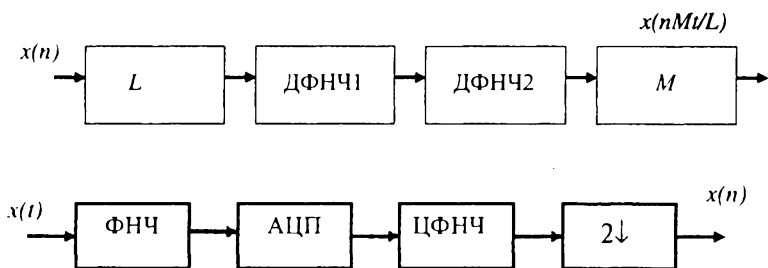


Рис. 2.25. Отсчеты $x(n)$ сигнала $x(t)$, следующие с периодом T (а), спектр $X(e^{j\omega t})$ дискретизованного сигнала $x(t)$ (б), отсчеты удлинённого во времени сигнала $x_{y\delta n}(n)$ (в), спектр $X_{y\delta n}(e^{j\omega t})$ сигнала $x_{y\delta n}(n)$ (г), отсчеты интерполированного сигнала на выходе ДФНЧ (д) и его спектр (е)



Спектр аналогового сигнала (а), требуемая амплитудно-частотная характеристика ФНЧ при частоте дискретизации 44,1 кГц (б), требуемая амплитудно-частотная характеристика ФНЧ при частоте дискретизации 88,2 кГц (в)

Рис. 2.26. Экспандер частоты дискретизации



Обработка сигнала цифровым ФНЧ с требуемой крутизной переходной полосы и прореживателем частоты дискретизации в 2 раза для получения сигнала с частотой дискретизации 44,1 кГц

Рис. 2.27. Каскадное соединение интерполятора в L раз и дециматора в M раз

цифровую форму аналогового музыкального сигнала полагают, что полоса его соответствует интервалу частот 22 кГц и минимальная частота дискретизации равна $44,1 \text{ кГц}$. Перед дискретизацией необходима аналоговая низкочастотная фильтрация (ФНЧ) для исключения эффекта наложения и внеполосного шума. При этом ФНЧ должен иметь высокую равномерность частотной характеристики в полосе пропускания и узкую переходную полосу от полосы пропускания к полосе задерживания.

Такие фильтры, как правило, имеют очень нелинейную фазо-частотную характеристику (ФЧХ) у края полосы пропускания (у частоты 22 кГц), что считается недопустимым для высококачественного воспроизведения музыки. Распространенный способ решения этой проблемы состоит в повышении вдвое (иногда вчетверо) частоты дискретизации. При этом аналоговый ФНЧ может иметь более широкую переходную полосу в интервале частот $22 \text{ кГц} — 44,1 \text{ кГц}$, и нелинейность ФЧХ оказывается приемлемо малой. На рис. 2.25 графически отображены: а) спектр аналогового сигнала; б) требуемая амплитудно-частотная характеристика ФНЧ при частоте дискретизации $44,1 \text{ кГц}$; в) требуемая амплитудно-частотная характеристика ФНЧ при частоте дискретизации $88,2 \text{ кГц}$.

Полученный на выходе аналого-цифрового преобразователя (АЦП) цифровой сигнал с частотой дискретизации $88,2 \text{ кГц}$ далее обрабатывается цифровым ФНЧ (ЦФНЧ) с требуемой крутизной переходной полосы и прореживается вдвое для получения нужного сигнала с частотой дискретизации $44,1 \text{ кГц}$, как это представлено на рис. 2.27. Заметим здесь, что реализация ЦФНЧ с линейной ФЧХ не вызывает принципиальных затруднений. Для этой цели используются цифровые фильтры с симметричной импульсной характеристикой.

Чтобы сохранить максимально возможную полосу частот в сигнале, являющемся промежуточным результатом, интерполяция должна быть осуществлена перед децимацией. В противном случае часть полосы исходного сигнала была бы отфильтрована дециматором. Характерным примером является преобразование частоты дискретизации проигрывателя компакт-дисков, которая равна $44,1 \text{ кГц}$, в частоту дискретизации,

используемую при цифровой звукозаписи в формате *DAT*, которая равна $48,0 \text{ кГц}$. Коэффициент интерполяции при этом равен 160, а коэффициент децимации — 147. На практике интерполяционный фильтр $h_1(k)$ и прореживающий фильтр $h_2(k)$ объединяются в один фильтр $h(k)$.

Полная функция преобразования частоты дискретизации интегрирована в микросхемах семейства *AD1890*, *AD1892*, *AD1893*, которые работают на частотах от 8 кГц до 56 кГц (48 кГц для *AD1892*). Новая модель *AD1896* работает на частотах до 196 кГц .

Аналогичная проблема возникает в ступени цифро-аналогового преобразования, когда цифровой музыкальный сигнал должен быть преобразован в аналоговый путем обработки в ФНЧ. Для этой цели нужен аналоговый ФНЧ с частотой среза 22 кГц и узкой переходной полосой и, следовательно, он будет иметь сильно нелинейную ФЧХ, что недопустимо. Для решения этой проблемы используют интерpolator, повышающий частоту дискретизации цифрового сигнала вдвое, после чего такой сигнал преобразуется в аналоговый в ЦАП с аналоговым ФНЧ, переходная полоса которого может быть в интервале частот $22 \text{ кГц} - 44,1 \text{ кГц}$, а нелинейность ФЧХ будет приемлемо малой. Структурная схема такой обработки представлена на рис. 2.28.

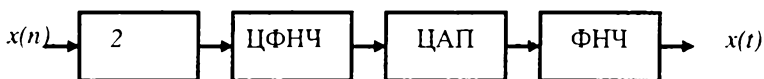


Рис. 2.28. Структурная схема цифро-аналогового преобразования сигнала с помощью ФНЧ с частотой среза 22 кГц и узкой переходной полосой

2.10. Дискретизация сигналов с непрерывным временем

В предыдущих разделах мы старались не сопоставлять дискретные сигналы и системы с аналоговыми сигналами и системами, за исключением указаний на сходство некоторых основных теоретических понятий. Однако часто дискретные сигналы получаются из аналоговых сигналов с помощью пе-

риодической дискретизации, поэтому важно понять, как последовательности, полученные таким образом, связаны с исходным сигналом.

Рассмотрим аналоговый сигнал $x(t)$, имеющий спектр — преобразование Фурье (НВПФ) $X(i\Omega)$:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(j\Omega) e^{i\Omega t} d\Omega \quad (2.33)$$

$$X(j\Omega) = \int_{-\infty}^{\infty} x(t) e^{-i\Omega t} dt. \quad (2.34)$$

Говорят, что последовательность $x(n)$ со значениями $x(n) = x(nT) = x(t)|_{t=nT}$ получена из $x(t)$ периодической дискретизацией, а T называется *периодом дискретизации*. Величина, обратная T , называется *частотой дискретизации* $F = 1/T$, или *скоростью дискретизации*. Чтобы определить, в каком смысле $x(n)$ представляет исходный сигнал $x(t)$, удобно связать $X(j\Omega)$ — преобразование Фурье аналогового сигнала $x(t)$ с $X(e^{i\Omega T})$ — преобразованием Фурье последовательности $x(n)$. Для последовательности $x(n)$ пару ДВПФ (2.32) и (2.33) запишем в виде:

$$x(n) = \frac{T}{2\pi} \int X(e^{i\Omega T}) e^{i\Omega nT} d\Omega \quad (2.35)$$

$$X(e^{i\Omega T}) = \sum_{n=-\infty}^{\infty} x(n) e^{-i\Omega nT}. \quad (2.36)$$

Идеализированную процедуру периодической дискретизации аналогового сигнала $x(t)$ представим в виде импульсного сигнала непрерывного времени $x_d(t)$:

$$x_d(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} x(nT) \delta(t - nT),$$

который содержит только отсчеты $x(n) = x(nT)$. Преобразование Фурье $X_d(i\Omega)$ сигнала $x_d(t)$ равно

$$\begin{aligned} X_d(i\Omega) &= \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT) \delta(t - nT) e^{-i\Omega t} dt = \\ &= \sum_{n=-\infty}^{\infty} x(nT) \int_{-\infty}^{\infty} \delta(t - nT) e^{-i\Omega t} dt = \sum_{n=-\infty}^{\infty} x(nT) e^{-i\Omega nT} = X(e^{i\Omega T}). \end{aligned}$$

И, очевидно, совпадает с ДВПФ $X(e^{j\Omega T})$ последовательности $x(n)$. Поэтому $X(e^{j\Omega T})$ можно выразить в виде свертки преобразований Фурье сомножителей $x(t)$ и $\sum_{n=-\infty}^{\infty} \delta(t - nT)$, образующих $x_d(t)$. Учитывая, что преобразование Фурье сигнала

$\sum_{n=-\infty}^{\infty} \delta(t - nT)$ есть периодическая с периодом $\Omega_0 = 2\pi/T$ последовательность дельта-функций частоты, равная $\Omega_0 = \sum_{n=-\infty}^{\infty} \delta(\Omega - n\Omega_0)$, свертка примет вид:

$$\begin{aligned} \frac{\Omega_0}{2\pi} \int_{-\infty}^{\infty} X[i(\Omega - \nu)] \sum_{n=-\infty}^{\infty} \delta(\nu - n\Omega_0) d\nu = \\ = \frac{\Omega_0}{2\pi} \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} X[i(\Omega - \nu)] \delta(\nu - n\Omega_0) d\nu = \frac{1}{T} \sum_{n=-\infty}^{\infty} X[i(\Omega - n\Omega_0)]. \end{aligned} \quad (2.37)$$

Из соотношения (2.37) следует, что ДВПФ последовательности $x(n)$ представляет собой периодически с периодом Ω_0 размноженное по оси частот преобразование Фурье (НВПФ) сигнала непрерывного времени $x(t)$. Графическое отображение спектров сигналов представлено на рис. 2.29. Если спектр $X(i\Omega)$ аналогового сигнала имеет ограниченную локализацию по оси частот, например в диапазоне частот $-\Omega_g \leq \Omega \leq \Omega_g$ ($\Omega_g = 2\pi F_g$ — верхняя частота спектра), как это представлено на рис. 2.29, а, то спектр $X(e^{j\Omega T})$ дискретизованного сигнала будет иметь вид, представленный на рис. 2.29, в, когда соседние спектральные полосы не перекрываются. Если период дискретизации выбран слишком большим, таким, что $-\pi/T < \Omega_g < \pi/T$ ($F < F_g$), то в спектре $X(e^{j\Omega T})$ дискретизованного сигнала будет иметь место перекрытие соседних спектральных полос (*эффект наложения спектров*), как это представлено на рис. 2.29, б. Из рис. 2.29, в и соотношения (2.27) видно, что при частоте дискретизации, по крайней мере вдвое большей верхней частоты спектра $X(i\Omega)$ ($F_{\text{дв}} = 2F_g$), справедливо соотношение:

$$X(e^{j\Omega T}) = \frac{1}{T} X(i\Omega), \quad -\pi/T \leq \Omega \leq \pi/T, \quad (2.38)$$

т. е. спектр $X(e^{j\Omega T})$ совпадает со спектром $X(i\Omega)$ и, следовательно, по дискретной последовательности $x(n)$ можно вос-

становить исходный аналоговый сигнал $x(t)$. Это утверждение составляет содержание так называемой *теоремы отсчетов* (в отечественной литературе часто именуемой теоремой Котельникова).

Сделаем замечание относительно выбора шкалы частот при спектральных представлениях дискретных последовательностей. В соотношениях (2.34) и (2.35), соответствующих паре ДВПФ, мы использовали аналоговые частоты Ω , измеряемые в радиан/с. Аналоговые частоты F , измеряемые в герцах ($1/c$), определяются как $F = \Omega/2\pi$. В соотношениях (2.30) и (2.31) для ДВПФ использовались *цифровые частоты* $\omega = \Omega T = \Omega/F_g$, измеряемые в радианах, которые можно трактовать как аналоговые, нормированные к частоте дискретизации. В большинстве случаев используются именно цифровые частоты Ω вместе с цифровыми безразмерными частотами $f = \omega/2\pi = F/F_g$. Переход от цифровых частот к аналоговым осуществляется домножением на частоту дискретизации. На рис. 2.29, б и 2.29, в использованы шкалы цифровых частот ω .

Найдем интерполяционную формулу для восстановления $x(t)$ по $x(n)$. Учитывая соотношения (2.35), (2.36) и (2.38), получим

$$\begin{aligned} x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(i\Omega) e^{i\Omega t} d\Omega = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} X(e^{i\Omega T}) e^{i\Omega t} d\Omega = \\ &= \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} \sum_{n=-\infty}^{\infty} x(n) e^{-i\Omega n T} e^{i\Omega t} d\Omega = \sum_{n=-\infty}^{\infty} x(n) \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} e^{i\Omega(t-nT)} d\Omega = \\ &= \sum_{n=-\infty}^{\infty} x(n) \frac{\sin[(\pi/T)(t-nT)]}{(\pi/T)(t-nT)}. \end{aligned} \quad (2.39)$$

Представление аналогового сигнала в виде (2.39) справедливо только для функций с ограниченным спектром при достаточно малом T , т. е. отсутствии эффекта наложения. Если $-\pi/T > \Omega_g > \pi/T$ ($F_g < 2F_g$), то из-за наложения спектров восстановление аналогового сигнала без искажений невозможно. Заметим, что техническая реализация операции восстановления аналогового сигнала по его отсчетам может быть выполнена пропусканием импульсного сигнала через идеальный

фильтр нижних частот, пропускающий без искажений полосу частот $-\pi/T < \Omega < \pi/T$.

При дискретизации аналоговых видеосигналов с неограниченным спектром можно уменьшить влияние эффекта наложения спектров, если аналоговый сигнал перед подачей на устройство дискретизации (АЦП) пропустить через фильтр нижних частот (ФНЧ) с целью ограничения его спектра диапазоном частот $|\Omega| < \pi/T$. При этом останутся лишь искажения, связанные с отбрасыванием хвостов спектра. Эта процедура низкочастотной фильтрации целесообразна и в случае, когда дискретизируемый сигнал смешан с широкополосной помехой, чтобы после дискретизации не ухудшалось отношение сигнал/помеха вследствие эффекта наложения.

Рассмотрим особо дискретизацию во времени аналоговых радиосигналов, которые могут быть представлены в виде $x(t) = a(t)\cos[\Omega_0 t + \varphi(t)]$ и спектр которых локализован в полосе частот в окрестности средней частоты Ω_0 . Таким образом представляются, например, сигналы, получаемые при модуляции гармонического колебания с частотой Ω_0 по амплитуде и (или) углу (частоте, фазе) некоторым низкочастотным сигналом. Таким же образом могут быть представлены и любые другие узкополосные сигналы, для которых $\Delta\Omega/\Omega_0 \ll 1$. Понятно, что такие сигналы могут быть дискретизованы как видеосигналы с верхней частотой спектра, равной $\Omega_0 + \Delta\Omega/2$, однако такая дискретизация оказывается для них чрезвычайно избыточной. Покажем, что узкополосные радиосигналы можно дискретизовать, образуя отсчеты с периодом $T \leq 2\pi/\Delta\Omega$, где $\Delta\Omega$ — ширина спектра радиосигнала. Для этого запишем $x(t)$ в виде

$$x(t) = a(t)\cos[\Omega_0 t + \varphi(t)] = a(t)\cos\varphi(t)\cos\Omega_0 t - a(t)\sin\varphi(t)\sin\Omega_0 t = a_c \cos\Omega_0 t - a_s \sin\Omega_0 t = \operatorname{Re}\{z(t)\exp(i\Omega_0 t)\}$$

где $a_c(t) = a(t)\cos\varphi(t)$, $a_s(t) = a(t)\sin\varphi(t)$, $z(t) = a_c(t) + ia_s(t) = a(t)e^{i\varphi(t)}$ — комплексная огибающая радиосигнала. При известной частоте Ω_0 узкополосный радиосигнал полностью описывается его комплексной огибающей $z(t)$, которая является низкочастотным процессом с шириной спектра $\Delta\Omega$. Найдем связь между спектром $X(i\Omega)$ сигнала $x(t)$ и спектром $Z(i\Omega)$ комплексной огибающей $z(t)$.

Поскольку

$$x(t) = (1/2)[z(t)\exp(i\Omega_0 t) + z^*(t)\exp(-i\Omega_0 t)],$$

следовательно,

$$X(j\Omega) = (1/2)\{Z[i(\Omega - \Omega_0) + Z^*[i(-\Omega - \Omega_0)]\}.$$

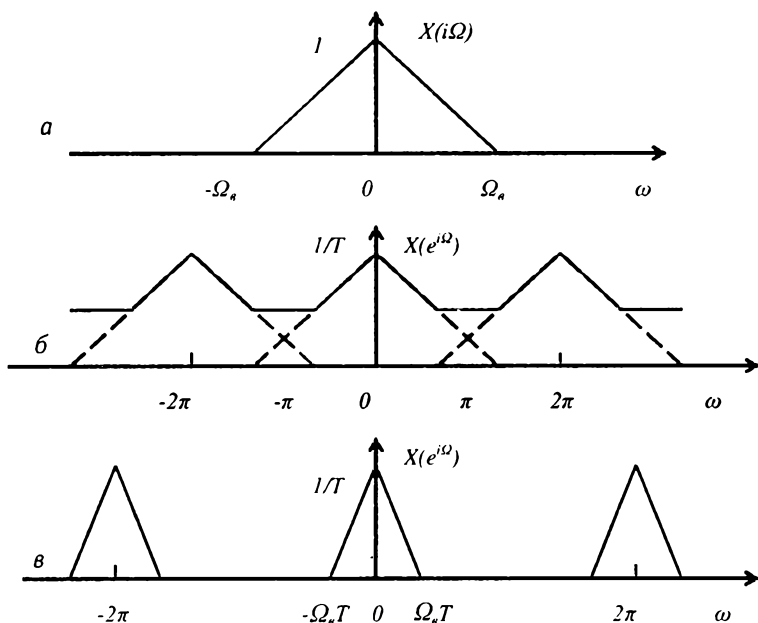


Рис. 2.29. Преобразование Фурье аналогового сигнала (а), преобразование Фурье дискретного сигнала, полученного при периодической дискретизации; период дискретизации велик, и поэтому периодически повторяющиеся преобразования Фурье аналогового сигнала перекрываются (б); период дискретизации мал настолько, что периодически повторяющиеся преобразования Фурье аналогового сигнала не перекрываются (в)

На рис. 2.30 графически отображены спектр $|X(i\Omega)|$ сигнала $x(t)$ и спектр $|Z(i\Omega)|$ комплексной огибающей $z(t)$.

Итак, при известной частоте Ω_0 дискретизация узкополосного сигнала эквивалентна процессу образования выборок его комплексной огибающей с периодом $T \leq 2\pi/\Delta\Omega$, по которым этот сигнал может быть восстановлен без искажений. Заметим, что выборки комплексной огибающей, т. е. комплексная последовательность $z(n) = z(t)|_{t=nT} = z(nT) = a_c(nT) + ia_s(nT)$ есть совокупность действительных последовательностей $a_c(n) = a_c(t)|_{t=nT} = a_c(nT) \cdot a_s(n) = a_s(nT)$.

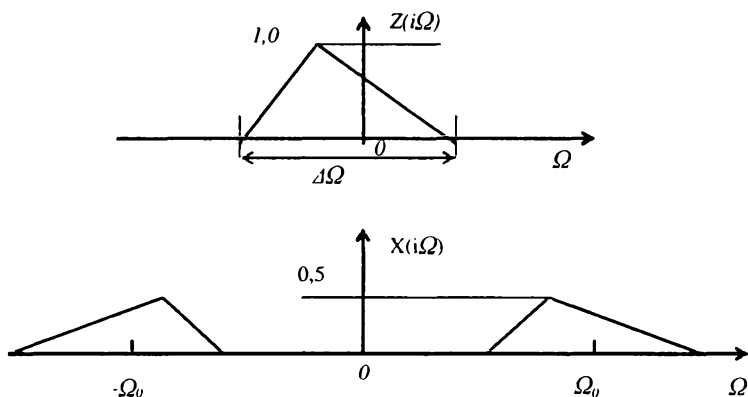


Рис. 2.30. Сравнение спектральных характеристик $Z(i\Omega)$ и $X(i\Omega)$

Поэтому при дискретизации радиосигнала $x(t)$ можно предварительно выделить из него низкочастотные сигналы $a_c(t)$ и $a_s(t)$, например с помощью двухканальной схемы, представленной на рис. 2.31 и называемой блоком квадратурного разложения, с последующей их периодической дискретизацией. Каждый из каналов схемы состоит из перемножителя входного сигнала $x(t)$ с опорным — гармоническим колебанием с частотой Ω_0 и фильтра нижних частот Φ с полосой пропускания $0 - \Delta\Omega/2$, причем опорные колебания каналов сдвинуты по фазе на $\pi/2$. Возможны также и другие варианты получения выборок сигналов $a_c(t)$ и $a_s(t)$, например непосредственной дискретизацией $x(t)$.

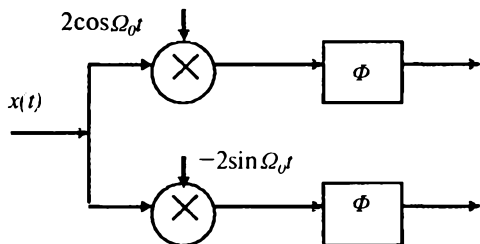


Рис. 2.31. Блок квадратурного разложения аналогового радиосигнала

2.11. ДПФ и дискретное косинусное преобразование

Чтобы иметь дело только с действительными числами, обычно используют разложение с помощью дискретного косинусного преобразования (ДКП). Это разложение имеет вид

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} C(k) \cos\left(\frac{2n+1}{2N} k\pi\right), \quad (2.40)$$

где коэффициенты ДКП находят по формулам:

$$C(0) = \frac{1}{\sqrt{2N}} \sum_{n=0}^{N-1} x(n),$$

$$C(k) = \frac{1}{\sqrt{2N}} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2n+1}{2N} k\pi\right),$$

$$k = 0, 1, \dots, N-1. \quad (2.41)$$

Как и в случае ДПФ, нахождение коэффициентов $C(k)$ по (2.40) называется прямым ДКП, а представление сигнала в виде суммы гармоник (2.41) называется обратным ДКП.

Существуют также быстрые алгоритмы вычисления ДКП. В цифровом телевидении на сегодняшний день главную роль играет ДКП блоков 8×8 пикселей, при выполнении которого используется алгоритм быстрого вычисления одномерного ДКП отрезка цифрового сигнала, содержащего 8 элементов. При этом сначала вычисляются ДКП для каждого столбца блока элементов изображения, а затем в полученной матрице 8×8 чисел вычисляются ДКП для каждой строки.

В современной аппаратуре, в том числе и для цифрового телевидения, ДПФ и ДКП, как правило, выполняются в реальном времени с применением цифровых процессоров обработки сигналов параллельных вычислительных устройств и других средств. Такие средства могут входить в состав БИС обработки телевизионных сигналов, например декодеров MPEG-2.

В заключение кратко поясним основной принцип использования ДПФ и ДКП в современной радиотехнике вообще и в цифровом телевидении в частности. Входной дискретный сигнал (одномерный или двумерный) подвергается прямому ДПФ или ДКП. Полученные коэффициенты обрабатываются. Выполняя операции над коэффициентами ДПФ или ДКП, мы, тем самым, выполняем операции над отдельными частотными составляющими сигнала. Например, можно усилить одни частотные составляющие и ослабить другие. Без выполнения ДПФ или ДКП такое выборочное воздействие на частотные составляющие осуществить гораздо труднее. После обработки выполняется обратное ДПФ или ДКП, в результате чего получается выходной дискретный сигнал.

ДКП лежит в основе наиболее широко используемых в настоящее время методов кодирования JPEG, MPEG-2 и некоторых других.

3.1. Источники сообщений без памяти и с памятью

Из определения энтропии $H(U)$ дискретного источника, создающего ансамбль сообщений U , а также из определения дифференциальной энтропии следует, что величина $H(U)$ зависит от распределения вероятности появления сообщений на выходе источника. Энтропия максимальна только в том случае, когда все сообщения источника равновероятны.

При генерировании источником последовательности сообщений существует еще один фактор, оказывающий влияние на величину энтропии, а именно наличие или отсутствие у источника памяти. Источник дискретных сообщений называется *источником с памятью*, если вероятность выдачи им очередного элементарного сообщения u_k зависит от того, какое (или какие) элементарные сообщения были выданы ранее. Иначе говоря, сообщения источника с памятью являются зависимыми. Стационарный источник независимых сообщений называется *источником без памяти*.

В качестве примера источника дискретных сообщений с памятью можно привести источник связного русского текста, в качестве элементарных сообщений которого рассматриваются отдельные буквы русского алфавита. Наряду с тем, что различными являются вероятности появления разных букв в тексте (*а* чаще, чем *з*), имеет место зависимость вероятности появления каждой буквы от того, какая буква ей предшествовала. Так, сочетание букв «*ар*» может встретиться чаще, чем «*аз*». В то же время если передаче подлежат расчеты ЭВМ и в качестве элементарных сообщений выступают отдельные цифры, то есть основания в общем случае такой источник информации считать источником без памяти. Количество информации,

содержащееся в одном элементарном сообщении источника с памятью, определяется с помощью условных вероятностей. Следовательно, и энтропия такого источника, определяемая на основе (1.3), также будет равна условной энтропии $H(u/z)$ сообщения ансамбля U при условии, что ему предшествовало сообщение (или несколько сообщений) ансамбля Z . В соответствии со свойством условной энтропии (1.6) для случая зависимых ансамблей U и Z всегда $H(U/Z) < H(U)$, т. е. энтропия источника с памятью всегда меньше энтропии источника независимых сообщений. Таким образом, энтропия дискретного источника максимальна в том случае, когда выполняются 2 условия:

- 1) все сообщения источника независимы (источник без памяти);
- 2) все сообщения источника равновероятны.

Невыполнение любого из этих требований уменьшает энтропию источника и является причиной избыточности.

Для характеристики величины, на которую увеличивается длина соответствующего символа по сравнению с минимальной длиной, необходимой для передачи той же информации, вводят специальный параметр R , называемый *избыточностью*. Избыточностью источника дискретных сообщений с энтропией H и объемом алфавита N называется величина

$$R = \frac{H_{\max} - H}{H_{\max}} = 1 - \frac{H}{H_{\max}} = 1 - \mu,$$

$$H_{\max} = \log_2 N, \quad 0 \leq R \leq 1,$$

где N — число различных букв используемого алфавита, H — энтропия, приходящаяся на одну букву смыслового текста при учете всех многобуквенных сочетаний, $H_{\max} = \log_2 N$ — максимальная энтропия, приходящаяся на букву, когда буквы независимы и равновероятны, μ — коэффициент сжатия текста.

Наличие избыточности позволяет ставить вопрос о сжатии информации без ее потери в передаваемых сообщениях.

Избыточность показывает, какая доля максимально возможной при заданном объеме алфавита неопределенности (энтропии) не используется источником. В частности, избыточность современного английского текста составляет около 50%, избыточность русского текста находится вблизи 70%.

3.2. Производительность источника дискретных сообщений. Скорость передачи информации

Обычно источники передают сообщения с некоторой скоростью, затрачивая в среднем время τ на передачу одного сообщения.

Производительностью источника

$$H'(U) = \frac{1}{\tau} \cdot H(U) \quad (3.1)$$

называется суммарная энтропия сообщений, переданных за единицу времени. Производительность измеряется в *бит/с*. Если сообщение может быть представлено в виде последовательности элементарных дискретных сообщений u_k источника с энтропией $H(U)$, следующих со скоростью $v_k = 1/\tau$ элементов в секунду, то

$$H'(U) = v_k \cdot H(U). \quad (3.2)$$

Аналогичным образом получим соответствующие равенства для условной энтропии и количества информации, рассчитанные на одно сообщение в единицу времени:

$$I'(U; Z) = (1/\tau) \cdot I(U; Z). \quad (3.3)$$

Величина $I'(U; Z)$ называется *скоростью передачи информации* от U к Z или, наоборот, от Z к U . Если, например, U -ансамбль сигналов на входе дискретного канала, а Z -ансамбль сигналов на его выходе, то скорость передачи информации по каналу

$$\begin{aligned} I'(U; Z) &= H'(U) - H'(U/Z) = \\ &= H'(Z) - H'(Z/U) = H'(U; Z). \end{aligned} \quad (3.4)$$

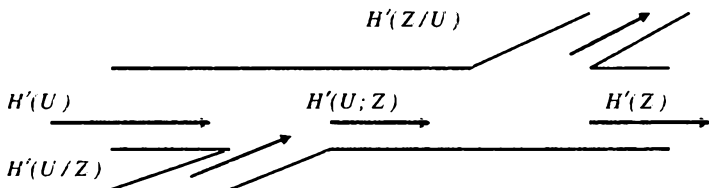


Рис. 3.1. Скорость потери информации (ненадежность канала)

$H'(U/Z)$, скорость создания ложной информации в канале

$H'(Z/U)$ и скорость передачи информации по каналу $H'(U; Z)$

Это соотношение наглядно иллюстрируется на рис. 3.1. Здесь $H'(Z)$ производительность источника передаваемого сигнала U , а $H'(Z)$ «производительность» канала, т. е. полная собственная информация в принятом сигнале за единицу времени. Величина $H'(U / Z)$ представляет собой *потерю информации или ненадежность канала* в единицу времени, а $H'(Z / U)$ — скорость создания ложной, посторонней информации в канале, не имеющей отношения к U и обусловленная присутствующими в канале помехами. По определению Шеннона, ненадежность канала $H'(U / Z)$ является энтропией входа, когда выход известен, т. е. ее можно считать мерой средней неопределенности принятого сигнала. Величина же $H'(Z / U)$ есть энтропия выхода, когда вход известен и служит мерой средней неопределенности передаваемого сигнала. Соотношение между $H'(U / Z)$ и $H'(Z / U)$ зависит от свойств канала. Так, например, при передаче звукового сигнала по каналу с узкой полосой пропускания, недостаточной для качественного воспроизведения сигнала, и с низким уровнем помех теряется часть полезной информации, но почти не получается бесполезной, так как в этом случае $H'(U / Z) \gg H'(Z / U)$. Если же сигнал воспроизводится качественно, но при этом прослушиваются наводки от соседнего радиоканала, то это означает, что, почти не теряя полезной информации, мы получили много лишней, мешающей информации и $H'(U / Z) \ll H'(Z / U)$.

3.3. Коэффициент сжатия

Наличие в сообщениях избыточности позволяет ставить вопрос о сжатии данных, т. е. о передаче того же количества информации с помощью последовательностей символов меньшей длины. Для этого используются специальные алгоритмы сжатия, уменьшающие избыточность. Эффект сжатия оценивают *коэффициентом сжатия* $= n/q$, где n — число минимально необходимых символов для передачи сообщения (практически это число символов на выходе эталонного алгоритма сжатия); q — число символов в сообщении, сжатом данным алгоритмом. Так, при двоичном кодировании n равно энтропии источника информации.



Пример 3.1. Источник сообщений выдает символы из ансамбля $A=\{a_i\}$, $i = 1, 2, 3, 4$ с вероятностями $p(a_1) = 0,2$; $p(a_2) = 0,3$; $p(a_3) = 0,4$; $p(a_4) = 0,1$. Найти количество информации, содержащейся в каждом из символов источника при их независимом выборе, т. е. вычислить величины частных энтропий (источник без памяти). Вычислить энтропию и избыточность источника.

Решение. Частные энтропии:

$$H(a_1) = -\log_2 0,2 = 1,22 \text{ бит},$$

$$H(a_2) = -\log_2 0,3 = 1,75 \text{ бит},$$

$$H(a_3) = -\log_2 0,4 = 1,33 \text{ бит},$$

$$H(a_4) = -\log_2 0,1 = 3,33 \text{ бит}.$$

Энтропия источника сообщений:

$$\begin{aligned} H(A) &= 0,2\log_2 0,2 + 0,3\log_2 0,3 + 0,4\log_2 0,4 + 0,1\log_2 0,1 = \\ &= 1,8465 \text{ бит}. \end{aligned}$$

Максимальная энтропия, соответствующая равновероятному распределению вероятностей получения сообщений:

$$H_{\max} = \log_2 N = \log_2 4 = 2 \text{ бита}.$$

Избыточность источника сообщений:

$$R = \frac{H_{\max} - H}{H_{\max}} = 1 - \frac{1,8465}{2} = 0,07675.$$

3.4. Понятие двоичного симметричного канала связи

Рассмотрим случай, когда входной алфавит A совпадает с выходным алфавитом B и равен $\{0, 1\}$. Пусть при посылке 0 он принимается как 0, а 1 как 1, но иногда 0 может быть принят как 1, а 1 — как 0. Пусть в среднем один из, например, 1000 символов будет ошибочным. Это означает, что для каждого символа имеется вероятность $p = 0,001$ того, что в канале связи произойдет ошибка, т. е. для переходных (условных) вероятностей $p(0/0)$, $p(0/1)$, $p(1/1)$ и $p(1/0)$ имеем $p(0/0) = p(1/1) = p$, $p(1/0) = p(0/1) = 1 - p$.

В данном случае переходные вероятности образуют симметричную матрицу, и поэтому такая модель канала связи называется *двоичным симметричным каналом связи*.

Существуют каналы и с другими типами ошибок. Если для условных вероятностей есть запреты, то канал связи получается с односторонними ошибками, если в алфавит (входной и/или выходной) включен пустой символ, то получаем канал со вставками или выпадениями.

Если сигнал, представляющий собой 0 или 1, искажается в канале так, что его нельзя интерпретировать ни как 0, ни как 1, то говорят, что в этом случае имеет место ошибка типа стирания и принятый символ заменяется некоторым третьим символом.



Пример 3.2. По двоичному симметричному каналу связи с помехами передаются сигналы x_1 и x_2 с априорными вероятностями $p(x_1) = 0,75$ и $p(x_2) = 0,25$. Из-за наличия помех вероятность правильного приема каждого из сигналов x_1 и x_2 снижается до $7/8$. Дать графическую интерпретацию работы канала связи, рассчитав величины:

- средней собственной информации $I(X)$ и $I(Y)$, т. е. величины информации, совпадающие с энтропией дискретных случайных величин X и Y , количественно определяющие априорные меры неопределенности этих сообщений;
- средних значений информации $I(X/Y)$ и $I(Y/X)$, совпадающих с условными энтропиями ансамбля событий X при данном ансамбле событий Y и ансамбля событий Y при данном ансамбле событий X соответственно;
- полного среднего количества взаимной информации $I(X; Y)$ в ансамбле сигналов Y относительно ансамбля символов X ;
- величину переданной по каналу информации.

Дать графическую интерпретацию рассчитанных потоков информации.

Решение. По условию $p(x_1) = 0,75$, $p(x_2) = 0,25$, $p(y_1/x_1) = p(y_2/x_2) = 7/8$. Кроме того, легко установить, что $p(y_1/x_2) = p(y_2/x_1) = 1/8$, т. е. при передаче символа x_1 вероятность приема сигнала y_2 равна вероятности приема сигнала y_1 при передаче символа x_2 .

Построим модель анализируемого канала связи (рис. 3.2).

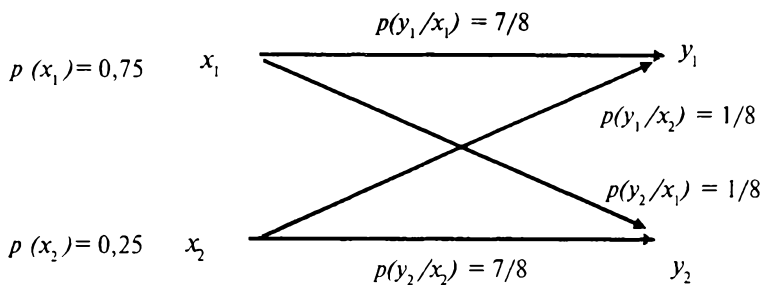


Рис. 3.2. Модель анализируемого двоичного симметричного канала

Вычислим вероятности

$p(y_j)$, $p(x_i, y_j)$ и $p(x_i/y_j)$, $i = 1, 2, j = 1, 2$.

$$p(y_1) = p(x_1) p(y_1/x_1) + p(x_2) p(y_1/x_2) = 11/16;$$

$$p(y_2) = p(x_2) p(y_2/x_2) + p(x_1) p(y_2/x_1) = 5/16;$$

$$p(x_1, y_1) = p(x_1) p(y_1/x_1) = 21/32;$$

$$p(x_1, y_2) = p(x_1) p(y_2/x_1) = 3/32;$$

$$p(x_2, y_1) = p(x_2) p(y_1/x_2) = 1/32;$$

$$p(x_2, y_2) = p(x_2) p(y_2/x_2) = 7/32;$$

$$p(x_1/y_1) = p(x_1) p(y_1/x_1)/p(y_1) = 21/22;$$

$$p(x_1/y_2) = p(x_1) p(y_2/x_1)/p(y_2) = 3/10;$$

$$p(x_2/y_1) = p(x_2) p(y_1/x_2)/p(y_1) = 1/22;$$

$$p(x_2/y_2) = p(x_2) p(y_2/x_2)/p(y_2) = 7/10.$$

$$I(X) = H(X) = - \sum_{i=1}^2 p(x_i) \log p(x_i) = 0,811 \text{ бит.}$$

$$I(Y) = H(Y) = - \sum_{j=1}^2 p(y_j) \log p(y_j) = 0,90 \text{ бит.}$$

$$I(X/Y) = - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log p(x_i / y_j) = 0,4588 \text{ бит.}$$

$$I(Y/X) = - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log p(y_j / x_i) = 0,55 \text{ бум.}$$

$$I(X, Y) = - \sum_{i=1}^2 \sum_{j=1}^2 p(x_i, y_j) \log [p(x_i / y_j) / p(x_i)] = 0,3525 \text{ бум.}$$

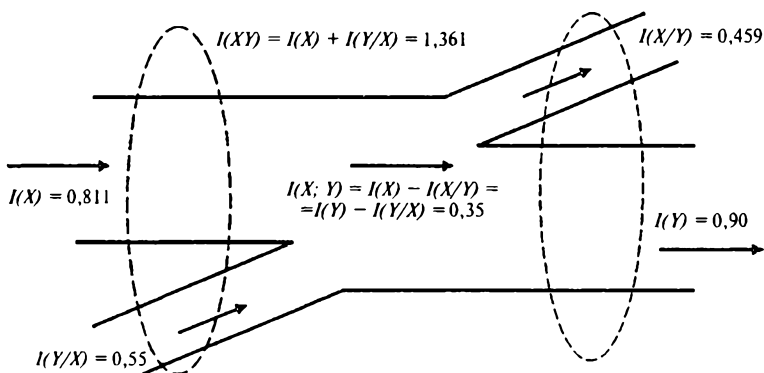


Рис. 3.3. Графическая интерпретация потоков информации

3.5. Пропускная способность дискретного канала

В любой системе связи по каналу передается информация, скорость ее передачи определяется выражением (3.3). Как видно из него, эта скорость зависит не только от свойств самого канала, но и от подаваемого на его вход сигнала и поэтому не может характеризовать канал как средство передачи информации. Попытаемся найти объективную характеристику способности канала передавать информацию. Рассмотрим дискретный канал, через который в единицу времени передается v_k символов источника с объемом алфавита M . При передаче каждого символа в среднем по каналу проходит количество информации

$$I(U; Z) = H(U) - H(U / Z) = H(Z) - H(Z / U), \quad (3.5)$$

где U и Z — ансамбли сообщений на входе и выходе канала. Из четырех фигурирующих здесь энтропий лишь $H(U)$ — собственная информация источника передаваемых символов, определяется источником входного сигнала и не зависит от свойств

канала. Остальные три энтропии в общем случае зависят как от свойств источника, так и от канала. Представим себе, что на вход канала можно подавать символы от различных источников, характеризующихся различными распределениями вероятностей $p(U)$ при одних и тех же значениях v_k и M . Для каждого такого источника количество информации, переданной по каналу, принимает свое значение. Очевидно, существует какой-то источник входного сигнала с некоторым распределением $p(U)$, для которого величина $I(U; Z)$ максимальна. Максимальное количество переданной информации C , взятое по всевозможным сообщениям входного сигнала, определяемым распределением вероятностей $p(U)$, характеризует сам канал, и называется пропускной способностью канала в расчете на один символ.

$$C = \max[I(U; Z)_{p(U)}], \text{ бит/символ.} \quad (3.6)$$

В формуле (3.6) максимизация производится по всем возможным многомерным (т. е. учитывающим статистическую взаимозависимость последовательно выдаваемых элементарных сообщений) распределениям вероятностей $p(U)$. Обычно определяют пропускную способность в расчете на единицу времени, которую и называют просто *пропускной способностью канала*. Пропускная способность канала удовлетворяет неравенству

$$0 \leq C \leq v_k \cdot \log_2 M, \quad (3.7)$$

причем $C = 0$ при независимых входе и выходе канала, т.е. $H(U/Z) = H(U)$ (обрыв канала или сильные помехи). Ограниченное значение

$$C = v_k \cdot \log_2 M \quad (3.8)$$

наблюдается в том случае, когда помех в канале нет, $H(U/Z) = H(Z/U) = 0$, при этом

$$H(U) = H(Z) = I(U; Z), \quad (3.9)$$

если учесть, что при заданном M

$$H_{\max}(U) = \log_2 M. \quad (3.10)$$

Таким образом, пропускная способность дискретного канала без шума определяется равенством (3.8), при наличии шума $C < v_k \cdot \log_2 M$.

В качестве примера вычислим пропускную способность дискретного симметричного канала без памяти. В таком канале каждый переданный кодовый символ может быть принят ошибочно с фиксированной вероятностью p и правильно с вероятностью $1 - p$, причем в случае ошибки вместо переданного символа u_k может быть с равной вероятностью принят любой другой символ. Таким образом, вероятность того, что принят символ z_j при условии, что передавался символ u_k , равна

$$p(z_j / u_k) = \begin{cases} p / (N - 1) & \text{при } j \neq k, \\ 1 - p & \text{при } j = k, \end{cases} \quad (3.11)$$

где N — объем алфавита источника. Термин без памяти означает, что вероятность ошибки в канале не зависит от того, какие символы передавались ранее и как они были приняты.



Пример 3.3. Вычислить пропускную способность C двоичного симметричного канала (рис. 3.4) при условии, что все символы x_i и соответствующие им элементарные сигналы y_j имеют одинаковую длительность τ при $F = 1/\tau$ — верхняя частота спектра передаваемого импульса. Например, при передаче сигналов с помощью прямоугольных импульсов спектр их ограничивается по первому нулю. Частота F определяет верхнюю рабочую частоту канала связи. Построить график зависимости $C/C_{\max} = f(p_e)$, где $C_{\max}$ — максимальная пропускная способность при отсутствии помех, а p_e — вероятность ошибочного приема.

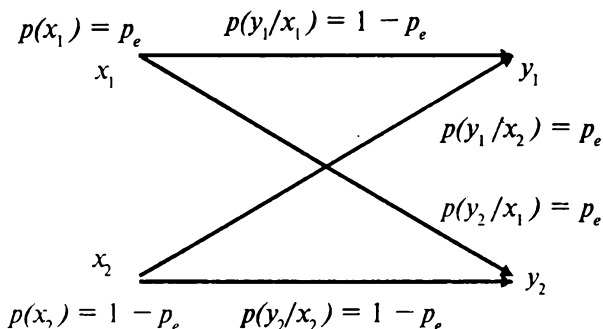


Рис. 3.4. Двоичный симметричный канал связи

Решение. Пропускная способность канала связи C представляет собой максимальную скорость передачи информации, которая может быть достигнута выбором оптимального распределения вероятностей передачи $p(x_j)$ символов сообщения, составляющих ансамбль U :

$$C = \max[F \cdot I(U; Z)_{p(U)}] = \max[F \cdot (H(U) - H(U/Z)_{p(U)})] = \\ = \max[F \cdot (H(Z) - H(Z/U)_{p(U)})].$$

В случае симметричного двоичного канала, когда передаются только $n = 2$ символа, соответствующие, например, передаче логических 0 или 1 :

$$H(Z) = \sum_{j=1}^2 p(z_j) \log_2 p(z_j), \\ H(Z/U) = \sum_{j=1}^2 p(u_i) \sum_{i=1}^2 p(z_j / u_i) \log_2 p(z_j / u_i).$$

Так как

$$p(z_1) = p(u_1)p(z_1 / u_1) + p(u_2)p(z_1 / u_2) = p + p_e - 2pp_e, \\ p(z_2) = p(u_1)p(z_2 / u_1) + p(u_2)p(z_2 / u_2) = 1 - p - p_e + 2pp_e,$$

то

$$H(Z) = -(p + p_e - 2pp_e) \cdot \log_2(p + p_e - 2pp_e) - \\ - (1 - p - p_e + 2pp_e) \log_2(1 - p - p_e + 2pp_e), \\ H(Z/U) = -p(u_1)[p(z_1 / u_1) \log_2 p(z_1 / u_1) + \\ + p(z_2 / u_1) \log_2 p(z_2 / u_1)] - \\ - p(u_2)[p(z_1 / u_2) \log_2 p(z_1 / u_2) + \\ + p(z_2 / u_2) \log_2 p(z_2 / u_2)] = \\ = -(1 - p_e) \log_2(1 - p_e) - p_e \log_2 p_e.$$

Из выражения для $H(Z/U)$ видно, что ввиду симметрии канала связи условная энтропия $H(Z/U)$ не зависит от вероятности передачи p . Поэтому максимальное значение $I(U; Z)$ достигается просто максимизацией $H(Z)$.

Максимум $H(Z)$ достигается тогда, когда сигналы z_1 и z_2 независимы и равновероятны, что, в свою очередь, имеет место при равновероятных передаваемых символах. Следовательно,

$$C = F [1 + (1 - p_e) \log_2(1 - p_e) + p_e \log_2 p_e].$$

При отсутствии помех, т. е. когда $p_e = 0$, $C = C_{\max} = F$. Поэтому

$$\frac{C}{C_{\max}} = f(p_e) = 1 + (1 - p_e) \log_2(1 - p_e) + p_e \log_2 p_e.$$

График относительной зависимости пропускной способности $C/C_{\max}$ от вероятности ошибочного приема p_e изображен на рис. 3.5.

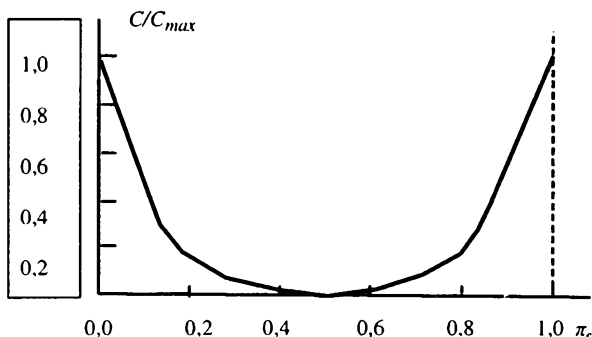


Рис. 3.5. Зависимость пропускной способности двоичного симметричного канала от вероятности ошибочного приема

Обсудим полученный результат. При $p_e = 0,5$ пропускная способность двоичного канала $C = 0$, так как при такой вероятности ошибки последовательность выходных двоичных символов можно получить, совсем не передавая сигнал по каналу, а выбирая их значения наугад. При $p_e = 0$ последовательности на входе и выходе совпадают. То, что пропускная способность при $p_e = 1$ в двоичном канале такая же, что и при $p_e = 0$ (канал без шумов), объясняется тем, что при $p_e = 1$ достаточно все выходные символы инвертировать, чтобы правильно восстановить исходный сигнал.



Пример 3.4. Симметричный двоичный канал с шириной рабочих частот $f_0 = 1000$ Гц характеризуется вероятностями $p(y_1/x_1) = p(y_2/x_2) = 0,9$, $p(y_1/x_2) = p(y_2/x_1) = 0,1$. Найти количество передаваемой по каналу информации.

Решение. Переданная по каналу информация определяется соотношением

$$K(Z; U) = K(Z) - K(Z/U), \text{ где}$$

$$K(Z) = -p(y_1) \log_2 p(y_1) - p(y_2) \log_2 p(y_2).$$

Передача информации достигает максимума, когда входные и выходные сигналы равновероятны: $p(x_1) = p(x_2) = 0,5$ и $p(y_1) = p(y_2) = 0,5$. В этом случае

$$K(Z) = -0,5 \log_2 0,5 - 0,5 \log_2 0,5 = 1.$$

$$K(Z/U) = -(1 - p_e) \log_2 (1 - p_e) - p_e \log_2 p_e =$$

$$= -0,9 \log_2 0,9 - 0,1 \log_2 0,1 = -0,1368 - 0,3322 = 0,4690 \text{ бит.}$$

Из $K(Z; U) = K(Z) - K(Z/U) = 1 - 0,469 = 0,531$, бит, поэтому

$$C = f_0 K(Z/U) = 531 \text{ бит/с.}$$

Таким образом, мы видим, что если вероятность ошибочного приема составляет всего 10%, то пропускная способность канала снижается почти на половину. Легко получить, что при $p_e = 0,5$, пропускная способность канала вообще оказывается равной нулю, поэтому в этом случае пропадает всякая необходимость пользоваться таким каналом связи.



Пример 3.5. Оценить снижение пропускной способности двоичного симметричного канала вследствие того, что на каждые 100 переданных сигналов появляется один ошибочный.

Решение. Вероятность появления ошибки $p_e = 0,01$. Следовательно,

$$\frac{C}{C_{\max}} = 1 + (1 - p_e) \log_2 (1 - p_e) + p_e \log_2 p_e = 0,9192,$$

т. е. пропускная способность канала уменьшилась почти на 8%.

3.6. Однородный симметричный канал со стиранием

Дискретный канал называется однородным, если для любой пары i и j условная вероятность $p(b_j / a_i)$ с течением времени не изменяется, т. е. влияние помех все время одинаково.

Рассмотрим двоичный симметричный канал, на входе которого действуют сигналы a_1 и a_2 с вероятностями появления $p_1(a_1)$ и $p_2(a_2)$ соответственно, таким образом, что принятый сигнал не идентифицируется ни с одним из поступающих на вход, т. е. сигнал так искажается, что становится неузнаваемым. Поэтому можно считать, что принимается с вероятностью q новый сигнал a_3 , появление которого можно интерпретировать как пропажу (стирание) входного сигнала. Тогда

$$\begin{aligned} p(a_1 / a_1) &= p(a_2 / a_2) = 1 - p - q, \\ p(a_2 / a_1) &= p(a_1 / a_2) = p, \\ p(a_3 / a_1) &= p(a_3 / a_2) = q. \end{aligned} \quad (3.12)$$

Таким образом, используя энтропийный подход при анализе процесса передачи информации по дискретному каналу с помехами, будем считать, что опыт b состоит в выяснении того, какой сигнал из b_j , общее число которых изменяется в соответствии с $j = 1, \dots, m$, был принят на приемном конце канала. Опыт a заключается в выяснении того, какой сигнал, в общем случае из a_i , $i = 1, \dots, n$, был послан на вход канала. Опыт b несет в себе информацию об опыте a , значение которой определяется выражением

$$I(a; b) = H(a) - H(a/b). \quad (3.13)$$

Применительно к рассматриваемой ситуации смысл этого выражения заключается в том, что распознанный на приемном конце сигнал содержит информацию о сигнале, который был отправлен, но в общем случае — не полную. Из-за влияния помех в канале часть начальной информации теряется, и исход опыта b не несет полной информации относительно значения предшествующего исхода опыта a .

Таким образом, $H(a)$ — это энтропия, определяющая меру неопределенности того, какой сигнал передан:

$$H(a) = - \sum_{i=1}^n p(a_i) \log_2 p(a_i), \quad (3.14)$$

а $H(a/b)$ — условная энтропия, т. е. энтропия опыта a при условии, что ему предшествовал опыт b :

$$H_b(a) = - \sum_{i=1}^n p_{b_i}(a_i) \log_2 p_{b_i}(a_i), \quad (3.15)$$

$$H(b/a) = \sum_{j=1}^m p(b_j) H(a/b_j) = \\ - \sum_{j=1}^m p(b_j) \sum_{i=1}^n p(a_i/b_j) \log_2 p(a_i/b_j). \quad (3.16)$$

Для средней информации на один элементарный сигнал имеем

$$I(b; a) = - \sum_{i=1}^n p(a_i) \log_2 p(a_i) - \sum_{j=1}^m p(b_j) \sum_{i=1}^n p(a_i/b_j) \log_2 p(a_i/b_j). \quad (3.17)$$

Поскольку $I(a; b) = I(b; a)$, то

$$I(a; b) = - \sum_{j=1}^m p(b_j) \log_2 p(b_j) - \sum_{i=1}^n p(a_i) \sum_{j=1}^m p(b_j/a_i) \log_2 p(b_j/a_i). \quad (3.18)$$

Отсюда следует, что для получения информации, содержащейся в сигнале, принятом на приемном конце канала, нужно располагать сведениями об априорных и апостериорных вероятностях. При этом априорные вероятности определяются особенностями источника (точнее, кодера), апостериорные — свойствами канала связи. Используя (3.17) или (3.18) и изменяя значения $p(a_i)$ в допустимых по условию задачи пределах, можно найти наибольшее значение информации, которую несет в себе опыт b об опыте a , т. е. информацию $\max[I(b; a)]$. Тогда пропускная способность канала

$$C = C_{\max} \cdot \max[I(b; a)]. \quad (3.19)$$

Выражение (3.19) позволяет сформулировать порядок решения задачи о нахождении пропускной способности конкретного канала, сводящийся к вычислению априорных и апостериорных вероятностей, нахождению по (3.17) максимального информационного содержания элементарного сигнала, расчету по (3.19) пропускной способности канала.

Например, если в канале отсутствуют помехи, тогда $m = n$ и сигналы на приемном конце канала совпадают с теми, что действуют на входе этого канала, т. е. апостериорные вероятности $p(a_i/b_j) = 1$ при $i = j$ и $p(a_i/b_j) = 0$ при $i \neq j$. В этом

случае $H(a/b) = 0$ и $\max[I(b; a)] = \max[H(a)] = \log_2 n$. Максимальное значение неопределенности получается при равновероятном появлении сигналов на входе канала. Из (3.19) следует, что $C = C_{\max} \cdot \log_2 n$, что совпадает с определением пропускной способности канала без помех и является частным случаем общего определения (3.19).



Пример 3.6. Рассчитать пропускную способность C двоичного симметричного канала со стиранием (рис. 3.6), если символы a_1 и a_2 имеют одинаковую длительность τ и $F = 1/\tau$ представляет верхнюю частоту спектра прямоугольных посылок.

Решение. Рассчитаем условную энтропию приема сигналов a_1 или a_2 , распространяющихся по каналу со стиранием. Имеем $H(b/a_1) = H(b/a_2) = -(1-p-q) \log_2(1-p-q) -$

$$-p \log_2 p - q \log_2 q,$$

$$H(b/a) = p(a_1) \cdot H(b/a_1) + p(a_2) \cdot H(b/a_2) =$$

$$= [p(a_1) + p(a_2)] [-(1-p-q) \log_2(1-p-q) - p \log_2 p - q \log_2 q] =$$

$$= -(1-p-q) \log_2(1-p-q) - p \log_2 p - q \log_2 q.$$

Так как $H(b/a)$ не зависит от значений априорных вероятностей, то $I(a; b)$ достигает максимума при таких $p(a_1)$ и $p(a_2)$, когда наибольшее значение приобретает энтропия $H(b)$. Для расчета величины $H(b)$ надо рассчитать величины вероятностей появления всех сигналов на выходе канала $q_j, j = 1, 2, 3$.

Рассчитаем величины q_j для a_1, a_2 и a_3 . Для a_1 вероятность $q_1 = p(a_1) \cdot (1-p-q) + p(a_2) \cdot p$; для a_2 имеем $q_2 = p(a_2) \cdot p + p(a_1) \cdot (1-p-q)$; для a_3 вероятность $q_3 = q$. Тогда

$$H(b) = -q_1 \cdot \log_2 q_1 - q_2 \cdot \log_2 q_2 - q \cdot \log_2 q.$$

Так как q определяется особенностями канала и не зависит от априорных вероятностей сигналов на входе, наибольшим значением для $H(b)$ будет то, которое определится максимальным значением выражения $-q_1 \cdot \log_2 q_1 - q_2 \cdot \log_2 q_2$, причем при любых $p(a_1)$ и $p(a_2)$ справедливо $q_1 + q_2 = 1 - q$. Нетрудно показать, что это выражение достигает максимума при условии $q_1 = q_2 = 0,5(1 - q)$. Тогда

$$\max[H(b)] = -(1-q) \cdot \log_2 \frac{1-q}{2} - q \cdot \log_2 q,$$

$$\begin{aligned} \max[I(a; b) &= -(1-p) \cdot \log_2 \frac{1-q}{2} + \\ &+ (1-p-q) \cdot \log_2(1-p-q) + p \cdot \log_2 p = \\ &= (1-p)[1 - \log_2(1-p)] + (1-p-q) \cdot \log_2(1-p-q) + p \cdot \log_2 p. \end{aligned}$$

Окончательно для пропускной способности канала со стиранием получаем

$$C = F\{(1-q)[1 - \log_2(1-q)] + (1-p-q) \cdot \log_2(1-p-q) + p \cdot \log_2 p\}.$$

Из полученного результата следует, что величина C уменьшается как с увеличением p , так и с увеличением q . Если $p \neq 0$ и $q \neq 0$, то $C < C_{\max} = F$. В реальных двоичных каналах со стиранием $p < q$, т. е. вероятность такого искажения входного сигнала, при котором его невозможно распознать, выше вероятности такого искажения, при котором сигнал становится похожим на другой из используемых в системе передачи сигналов. Если вероятность p пренебрежимо мала и единственным искажением оказывается стирание сигнала, пропускная способность оказывается равной $C = C_{\max} \cdot (1-q)$. График этой функции приведен на рис. 3.7. Из этого рисунка видно, что при $p = 0$ из $C_{\max}$ двоичных сигналов, передаваемых по каналу в единицу времени, в среднем $C_{\max} \cdot q$ будет стираться, но при остальных значениях $C_{\max} \cdot (1-q)$ сигналы будут на приемном конце расшифровываться без потерь, и с каждым из них связан ровно 1 бит информации.

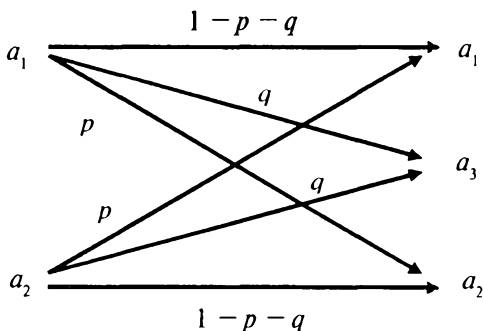


Рис. 3.6. Двоичный симметричный канал связи со стиранием

Таким образом, помехи, существующие в реальном канале связи, приводят к снижению его пропускной способности, если сравнивать его с таким же каналом, но без помех, и оценить ее снижение можно по известным заранее априорным и апостериорным вероятностям, характеризующим работу реального канала.

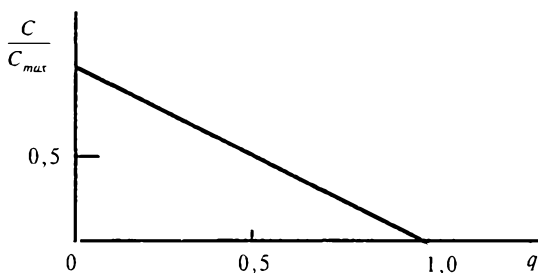


Рис. 3.7. Зависимость относительной пропускной способности C/C_{max} однородного двоичного симметричного канала от вероятности стирания q

СОГЛАСОВАНИЕ ИСТОЧНИКА ИНФОРМАЦИИ С КАНАЛОМ СВЯЗИ

4.1. Согласование дискретного источника с дискретным каналом без шума. Кодирование информации

Рассмотрим дискретный канал, вероятность возникновения ошибки в котором близка к нулю. Такой канал называют идеальным каналом, или каналом без шума. В соответствии с (3.8) пропускная способность канала определяется $C = \nu_k \log_2 M$. Естественно поставить вопрос о возможности передачи по нему без потерь информации от произвольного дискретного источника U , характеризуемого производительностью $H'(U)$ со скоростью, равной пропускной способности канала. Схема построения такой системы передачи информации должна выглядеть так, как показано на рис. 4.1.

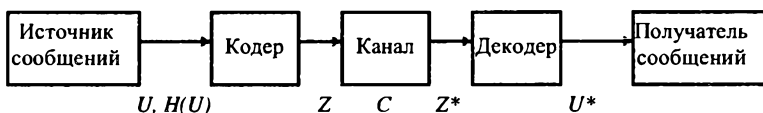


Рис. 4.1. Структурная схема дискретного идеального канала связи

Покажем необходимость включения *кодера* и *декодера*, которые выполняют операции кодирования и декодирования информации, в составе этой схемы. Чтобы скорость передачи информации в канале была равна его пропускной способности, на входе канала должен действовать дискретный источник с определенными статистическими свойствами, максимизирующими величину передаваемой по каналу информации.

В случае идеального канала без помех такой источник должен просто обладать максимальной энтропией, или нулевой избыточностью. Иначе говоря, такой источник должен выдавать независимые равновероятные сообщения.

Если же передаются сообщения от произвольного источника с любыми статистическими свойствами, то в этом случае источник характеризуется ненулевой избыточностью. Таким образом, функция кодера заключается в согласовании в статистическом смысле сообщений источника со входом канала. Задача этого согласования в конечном итоге сводится к устранению избыточности сообщений. Таким образом, кодер осуществляет кодирование сообщений, т. е. каждому дискретному сообщению по определенному правилу ставят в соответствие последовательность символов из алфавита объемом M . При этом по отношению ко входу канала выдаваемые кодером символы сами являются дискретными элементами сообщений, статистические свойства которых должны отличаться от статистических свойств сообщений исходного источника. Решение поставленной задачи определяется возможностью построения кодера, полностью устраняющего избыточность произвольного источника исходных сообщений. При этом скорость передачи информации по каналу связи равна пропускной способности канала

$$H'(U) = v_c \cdot H(U) = v_k \cdot \log_2 M = C,$$

$$v_k / v_c = H(U) / \log_2 M.$$

При полном ее решении оказывается справедливым равенство

$$H'(U) = v_c \cdot H(U) = v_k \log_2 M = C, \quad (4.1)$$

откуда имеем

$$\eta = v_k / v_c = H(U) / \log_2 M, \quad (4.2)$$

где $H(U)$ — энтропия источника передаваемых сообщений, v_k и v_c — средние количества символов соответственно сообщения и кода, передаваемых в единицу времени. $\eta = v_k / v_c$ — среднее количество символов кода, приходящихся на одно сообщение.

4.1.1. Эффективное кодирование

Степень приближения к точному выполнению равенств (4.1) и (4.2) зависит от степени уменьшения избыточности источника сообщений. Кодирование, позволяющее устранять избыточность источника сообщений, называется *эффективным*, или *статистическим*. Коды, получаемые в результате такого кодирования, называются *эффективными*, или *статистическими*. Рассмотрим основные идеи, которые могут быть положены в основу эффективного кодирования.

Избыточность дискретных источников обуславливается двумя причинами: 1) *памятью источника* и 2) *неравновероятностью сообщений*.

Универсальным способом уменьшения избыточности, обусловленной памятью источника, является укрупнение элементарных сообщений. При этом кодирование осуществляется длинными блоками. Вероятностные связи между блоками меньше, чем между отдельными элементами сообщений, и чем длиннее блоки, тем меньше связь между ними. Например, если вероятностные связи между буквами в любом языке относительно сильны, то между словами они значительно меньше, еще слабее между фразами, еще меньше между абзацами. Поэтому, применяя кодирование слов, фраз, абзацев, мы можем достаточно полно устранить избыточность, обусловленную вероятностными связями. Однако при этом возрастает задержка передачи сообщений, так как сначала нужно дождаться формирования всего длинного блока сообщений и затем его закодировать и передать. Уменьшение избыточности, обусловленной неравномерностью сообщений, может быть достигнуто применением неравномерных кодов. Основная идея построения таких кодов состоит в том, что наиболее вероятным сообщениям ставятся в соответствие наиболее короткие блоки кодовых символов (кодовые комбинации), а наименее вероятным — более длинные. В силу неравномерности таких кодов и случайного характера сообщения U передача информации осуществляется без потерь с постоянной скоростью следования кодовых символов v_k и может быть обеспечена лишь при наличии буферного накопителя с большой памятью и, следовательно, при допустимости больших задержек.

4.1.2. Асинхронное и синхронное кодирование

Для правильного распознавания позиций символов в передаваемом сообщении получатель должен знать границы передаваемых элементов сообщения. Для этого необходима синхронизация передатчика и приемника. Использование специального дополнительного провода для сигналов синхронизации (в этом случае имеем *битовую* синхронизацию) слишком дорого, поэтому используют другие способы синхронизации.

В *асинхронном режиме* применяют коды, в которых явно выделены границы каждого символа (байта) специальными стартовым и стоповым символами. Подобные побайтно выделенные коды называют *байт-ориентированными*, а способ передачи — *байтовой синхронизацией*. Однако это увеличивает число битов, не относящихся собственно к сообщению.

В *синхронном режиме* синхронизм поддерживается во время передачи всего информационного блока без обрамления каждого байта. Такие коды называют *бит-ориентированными*. Для входа в синхронизм нужно обозначать границы лишь всего передаваемого блока информации с помощью специальных начальной и конечной комбинаций байтов (обычно это двубайтовые комбинации). В этом случае синхронизация называется *блочной (фреймовой)*.

Для обрамления текстового блока (текст состоит только из печатаемых символов) можно использовать символы, отличающиеся от печатаемых. Для обрамления двоичных блоков применяют специальный символ (обозначаемый DLE — *Data Link Escape*), который благодаря *стаффингу (stuffing)* становится уникальным. Уникальность заключается в том, что если DLE встречается внутри блока, то сразу вслед за ним вставляется еще один DLE. Приемник будет игнорировать каждый второй идущий подряд символ DLE. Если же DLE встречается без добавления, то это граница блока.

З а м е ч а н и е. Существуют понятия бит- и байт-стаффинга. В случае последовательного канала передачи информации для правильного считывания посылки нужно найти ее начало. Во избежание ложных срабатываний эта последовательность исключается из совокупности передаваемых сообщений. Выполняется это вставкой дополнительных битов при угрозе появления ложной исключительной ситуации. В этом

суть бит-стаффинга. Бит-стаффинг экономичнее байт-стаффинга, так как вместо лишнего байта при его выполнении вставляется только один бит. Следовательно, скорость передачи данных в этом случае замедляется в меньшей степени.

4.2. Первая и вторая теоремы Шеннона

4.2.1. Передача информации по линии связи

Информация, идущая от источника, кодируется кодером и превращается преобразователем «коды-сигналы» в последовательность сигналов, поступающих в канал связи. Преобразователь «сигналы-коды» и декодер осуществляют обратное преобразование. На канал связи действуют *шумы* — помехи, искажающие передаваемый сигнал, что компенсируется защитой от шумов (рис. 4.2).

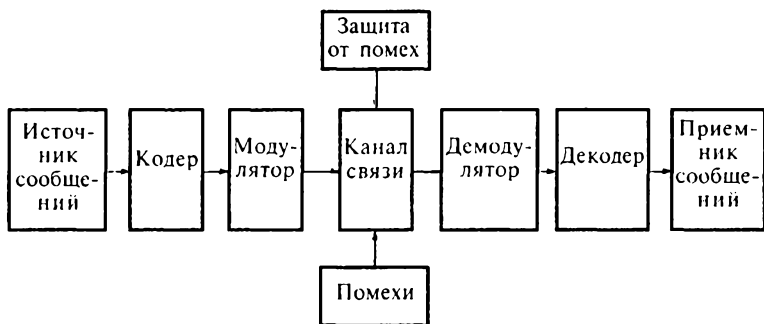


Рис. 4.2. Структурная схема линии связи с помехами и с защитой от них

Первичным называется алфавит, используемый в сообщении, которое выдает источник. Кодер преобразует сообщение, используя *вторичный* алфавит. Пусть первичный алфавит A содержит N знаков, а вторичный B — M знаков. На каждый знак первичного и вторичного алфавита приходится средняя информация I_A и I_B .

Передаваемое сообщение в первичном алфавите содержит n знаков, а закодированное — m знаков, они несут количество информации $I(A) = nI_A$ и $I(B) = mI_B$.

Кодирование будет обратимым при условии неисчезновения информации, когда

$$I(A) = I(B),$$

т. е. количество информации при обратимом кодировании не уменьшается:

$$I_A = (m/n)I_B = K(A,B)I_B,$$

где $K(A,B) = m/n$ — длина кода, т. е. среднее число знаков вторичного алфавита B , используемых для кодирования одного знака первичного алфавита A . Итак, длина кода

$$K(A, B) = I_A / I_B,$$

а ее минимальное значение

$$K^{\min}(A, B) = I_A / I_B.$$

Если $I_A > I_B$, то $K(A,B) > 1$, т. е. одному знаку первичного алфавита соответствует несколько знаков вторичного. *Относительная избыточность кода* равна

$$Q(A,B) = (K(A,B) - K^{\min}(A,B)) / K^{\min}(A,B).$$

Проблема оптимизации кода состоит в нахождении метода кодирования, при котором

$$K(A,B) \rightarrow K^{\min}(A,B), \quad Q(A,B) \rightarrow 0,$$

т. е. средняя длина кода стремится к своему пределу, и операция кодирования становится более эффективной. Ее решение позволит затратить на передачу сообщения меньше энергии и времени, а при его хранении использовать меньше поверхности носителя. Все это без учета других показателей, важность которых, вообще говоря, может быть весьма существенной.

4.2.2. Понятие о теоремах Шеннона

При передаче сообщений по каналам связи могут возникать помехи, способные привести к искажению принимаемых знаков. Вообще, передача сообщений при наличии помех является серьезной теоретической и практической задачей. Ее значимость возрастает в связи с повсеместным внедрением компьютерных телекоммуникаций, в которых помехи

неизбежны. При работе с кодированной информацией, искажаемой помехами, можно выделить следующие основные проблемы:

- установление самого факта того, что произошло искажение информации (в этом случае в канале связи создается сигнал о повторении передачи);
- выяснение того, в каком конкретно месте передаваемого текста это произошло, и автоматическое исправление ошибки хотя бы с некоторой степенью достоверности.

Помехи в передаче информации — вполне обычное дело во всех сферах профессиональной деятельности и в быту. Один из примеров — разговор по телефону, в трубке которого «трещит», вождение автомобиля в тумане и т. д. Чаще всего человек вполне справляется с каждой из указанных выше задач, хотя и не всегда отдает себе отчет, как он это делает (т. е. неалгоритмически, а исходя из каких-то ассоциативных связей). Известно, что естественный язык обладает большой избыточностью (в европейских языках — до 50%), чем объясняется большая помехоустойчивость сообщений, составленных из знаков алфавитов таких языков. Таким образом, в этом случае избыточность — полезное свойство.

Избыточность могла бы быть использована и при передаче кодированных сообщений в технических системах. Например, каждый фрагмент текста («предложение») передается трижды, и верным считается та пара фрагментов, которая полностью совпала. Однако большая избыточность приводит к большим временным затратам при передаче информации и требует большого объема памяти при ее хранении. Впервые теоретическое исследование эффективного кодирования предпринял К. Шеннон.

4.2.3. Первая теорема Шеннона

Эта теорема декларирует *возможность создания системы эффективного кодирования дискретных сообщений, у которой среднее число двоичных символов на один символ сообщения асимптотически стремится к энтропии источника сообщений (в отсутствии помех).*

4.2.4. Вторая теорема Шеннона

Вторая теорема Шеннона гласит, что *при наличии помех в канале всегда можно найти такую систему кодирования, при которой сообщения будут переданы с заданной достоверностью*. При наличии ограничения пропускная способность канала должна превышать производительность источника сообщений. Таким образом, *вторая теорема Шеннона устанавливает принципы помехоустойчивого кодирования*.

Для дискретного канала с помехами *теорема утверждает, что если скорость создания сообщений меньше или равна пропускной способности канала, то существует код, обеспечивающий передачу со сколь угодно малой частотой ошибок*.

Доказательство теоремы основывается на следующих рассуждениях. Первоначально последовательность $X = \{x\}$ кодируется символами из B так, что достигается максимальная пропускная способность (канал не имеет помех). Затем в последовательность из B длины n вводится r символов, и по каналу передается новая последовательность из $n + r$ символов. Число возможных последовательностей длины $n + r$ больше числа возможных последовательностей длины n . Множество всех последовательностей длины $n + r$ может быть разбито на n подмножеств, каждому из которых сопоставлена одна из последовательностей длины n . При наличии помехи последовательность из $n + r$ символов извлекает ее из соответствующего подмножества с вероятностью сколь угодно малой.

Это позволяет определить на приемной стороне канала, какому подмножеству принадлежит искаженная помехами принятая последовательность длины $n + r$, и тем самым восстановить исходную последовательность длины n .

Эта теорема не дает конкретного метода построения кода, но указывает на пределы достижимого в создании помехоустойчивых кодов, стимулирует поиск новых путей решения этой проблемы.

4.2.5. Выводы из теорем Шеннона

Итак, первая теорема Шеннона (основная теорема о кодировании при отсутствии помех) гласит: *при отсутствии помех возможно кодирование, при котором избыточность кода будет сколь угодно близка к нулю*.

Если вторичный алфавит имеет 2 знака (двоичный код), то в случае оптимального кодирования средняя длина двоичного кода должна быть равна среднему информационному содержанию знака первичного алфавита.

На реальный канал связи действуют помехи, в результате чего возникают ошибки. *Уровень достоверности* передаваемого сообщения может быть определен как отношение $D = n_{\text{error}}/n$ числа ошибок к общему количеству знаков n .

Вторая теорема Шеннона: *если скорость передачи не превышает пропускной способности канала связи с шумом, то всегда найдется способ кодирования, при котором сообщение будет передаваться с требуемой достоверностью.*

Возможны два способа решения проблемы:

- 1) способ кодирования только устанавливает факт искажения сообщения, что позволяет потребовать повторную передачу;
- 2) используемый код находит и автоматически исправляет ошибку передачи, не обращаясь к передатчику за дополнительными сведениями.

4.3. Кодирование в дискретном канале связи без помех

С точки зрения осуществления кодирования такой канал связи является самым простым, так как в нем можно использовать неизбыточные коды, например передавая наиболее вероятные сообщения более короткими кодами.

Рассмотрим ансамбль M сообщений, заданный своими вероятностями. Выберем для кодирования этих сообщений код с основанием K и закодируем каждое сообщение n_j некоторым словом, тем меньшей длины, чем больше вероятность появления этой кодовой комбинации. При этом, конечно, количество информации в кодовом слове I_j должно быть не меньше информации, содержащейся в кодируемом сообщении, а именно:

$$I_j = -\log p_j = n_j \log K.$$

Отсюда следует, что при $K = 2$ $n_j = -\log p_j$, т. е. n_j тем больше, чем меньше p_j . Рассчитаем среднюю длину кодового слова:

$$n_{cp} = M[n_j] = -\sum_{j=1}^M p_j \log p_j = H.$$

Таким образом, средняя длина кодового слова определяет предельно короткую длину кодовой комбинации, которую можно получить при наилучшем выборе кода. Учитывая это, коды, средняя длина слов которых удовлетворяет условию

$$H = n_{cp} = H + 1,$$

называются *оптимальными*. При этом n_{cp} можно не округлять до натурального числа.

Разработаны методы построения оптимальных кодов для каналов связи без помех. Из них наибольшее распространение получили коды Шеннона-Фано и коды Хаффмена.

Какая же связь между значением оцифрованного звука и генерируемым звуковой картой напряжением? Наиболее очевидный подход заключается в использовании линейной связи (*linear relation*), при которой, например, увеличение значения цифрового представления звука вдвое будет приводить к увеличению напряжения также в два раза. Однако этот подход не эффективен. Человеческое ухо воспринимает звук нелинейно: разница между малыми цифровыми представлениями звукового сигнала может быть слишком велика для слабых звуков, в то время как разница между большими представлениями будет слишком мала, чтобы ухо ее различило.

Принимая во внимание указанную природу человеческого слуха, вводят логарифмическую шкалу. Соотношения *-Law* и *A-Law* соответствуют этой шкале. Соотношение *-Law* используется, прежде всего, в Северной Америке и в Японии. Для преобразования значения линейной дискретизации в дискретизацию y используется следующее уравнение:

$$y_{\mu} = \frac{\text{sign} m}{\ln(1 + \mu)} \ln \left(1 + \mu \left| \frac{m}{m_p} \right| \right), \quad (4.13)$$

где m_p — максимальное входное значение оцифрованного звука, а μ — константа, обычно 100 или 255.

A-Law используется в Европе. Оно также используется для преобразования значения линейной дискретизации в дискретизацию y_A . A — это константа 87.6:

$$y_A = \begin{cases} \frac{A}{1 + \ln A} \left(\frac{m}{m_p} \right) & \left| \frac{m}{m_p} \right| \leq \frac{1}{A}, \\ \frac{\text{sign} m}{1 + \ln A} \left(1 + \ln A \left| \frac{m}{m_p} \right| \right) & \frac{1}{A} \leq \left| \frac{m}{m_p} \right| \leq 1. \end{cases} \quad (4.14)$$

Соотношения *-Law* и *A-Law* позволяют восьмиразрядные измерения представлять в том же диапазоне, что и линейные 12-разрядные. Таким образом, можно получить более чем 30% сжатия.

4.4. Оптимальные коды Шеннона-Фано

Принцип построения оптимального кода Шеннона-Фано сводится к следующему алгоритму:

1. Сообщения, составляющие ансамбль, располагаются по мере убывания вероятностей их появления.

2. Выбирается основание кода K .

3. Все сообщения разбиваются на K групп с равными суммарными вероятностями внутри каждой группы. Всем сообщениям первой группы в качестве первого символа присваивается 0, сообщениям второй группы — 1, сообщениям K -й группы — символ $K - 1$. Этим обеспечивается равная вероятность появления всех символов 0, 1, ..., K на первой позиции в кодовых словах.

4. Каждая из групп делится на K подгрупп с равной суммарной вероятностью в каждой подгруппе. Всем сообщениям первых подгрупп в качестве второго символа присваивается 0, всем сообщениям вторых подгрупп — 1, а сообщениям $K - x$ подгрупп — символ $K - 1$.

5. Процесс продолжается до тех пор, пока в каждой подгруппе не окажется по одной комбинации.



Пример 4.1. Построить оптимальный двоичный код для ансамбля сообщений, заданного вероятностями 1/4, 1/4, 1/8, 1/8, 1/16, 1/16, 1/16, 1/16. Рассчитайте среднюю длину кодового слова. Считая, что ансамбль представляет собой 8 букв русского алфавита А, З, Л, Н, О, П, Р, У, закодировать полученным кодом фразу А РОЗА УПАЛА НА ЛАПУ АЗОРА.

Решение. В соответствии с требованиями алгоритма располагаем сообщения в столбец по убыванию вероятностей

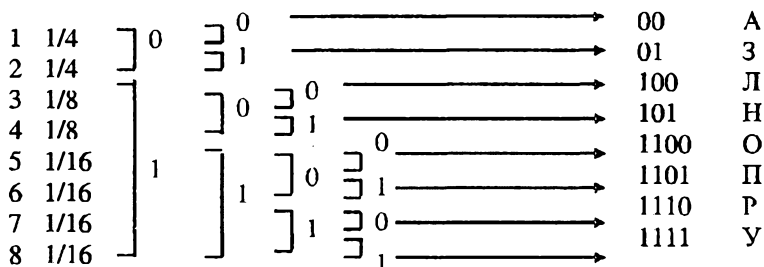


Рис. 4.2. Построение оптимального кода Шеннона-Фано

Фразу можно представить последовательностью двоичных цифр

001110110001001111110100100001010010000110111110001111000.

Полученный оптимальный код является *неравномерным*, так как наиболее вероятные сообщения кодируются коротким сигналом. В неравномерных кодах возникает проблема разделения кодовых комбинаций. Решение данной проблемы обеспечивается применением префиксных кодов. В этом случае для упрощения процедуры декодирования нужно выполнить условие однозначной различимости кодовых комбинаций. Одним из способов выполнения этого условия является построение кодовых слов таким образом, чтобы *никакая комбинация не являлась началом другой*. Альтернативой этому требованию служит введение специальных разделительных символов, которые должны выдаваться в конце каждого кодового слова. Их введение приведет к увеличению основания кода. Префиксные коды всегда однозначно декодируемы. Как нетрудно непосредственно убедиться, различимость кодов Шеннона-Фано обеспечивается первым способом. Таким образом, при кодировании методом Шеннона-Фано любой последовательности символов можно поставить в соответствие непрерывный ряд сообщений.

Далее мы предлагаем читателю решить самостоятельно пару примеров, надеясь, что это поможет существенно дополнить понимание рассматриваемых здесь вопросов.



Пример 4.2. Решить этот же пример при условии равновероятного появления сообщений.

Убедиться в том, что в этом случае на каждое сообщение будет приходиться ровно три двоичных цифры и длина закодированной фразы возрастет с 57 до $3 \cdot 21 = 63$ двоичных цифр. Этот результат можно объяснить тем, что при равновероятном распределении сообщений возрастает энтропия, что ведет к большей неопределенности опыта, т. е. количество получаемой информации увеличивается.



Пример 4.3. Решить пример 4.2, построив оптимальный троичный код для ансамбля сообщений, заданный вероятностями $1/3, 1/3, 1/9, 1/9, 1/27, 1/27, 1/54, 1/54$. Как изменится длина закодированной троичным кодом фразы?

Как показывают рассмотренные примеры, оптимальность кода достигается за счет пропорциональности вероятностей сообщений основанию кода, т. е.

$$p_j = K^m, m \in \mathbb{N}.$$

Только в этом случае осуществляется точное разделение всех групп сообщений на подгруппы с равной суммарной вероятностью сообщений в каждой подгруппе. Если это условие не выполняется, т. е. когда сообщения имеют произвольное распределение вероятностей, разбиение на группы возможно только с приближенным равенством суммарных вероятностей. В результате все символы кода будут иметь лишь приближенно равные вероятности появления, и средняя длина кодового слова

$$n_{cp} \neq H / \log K = n_{opt}, \text{ т. е. } n_{cp} \approx n_{opt}.$$

Такие коды называются *квазиоптимальными*. Это означает, что существует возможность построения более оптимального кода. Таким кодом является код Хаффмена.

4.4. Код Хаффмена

Одним из часто используемых методов эффективного кодирования является так называемый код Хаффмена.

Пусть сообщения входного алфавита $A = \{a_1, a_2, \dots, a_k\}$ имеют соответственно вероятности их появления $p_1, p_2, \dots, p_k$. Тогда алгоритм кодирования Хаффмена состоит в следующем:

1. Сообщения располагаются в столбец в порядке убывания вероятности их появления.

2. Два самых маловероятных сообщения объединяем в одно сообщение b , которое имеет вероятность, равную сумме вероятностей сообщений a_{k-1} и a_k , т. е. $p_{k-1} + p_k$. В результате получим сообщения $a_1, a_2, \dots, a_{k-2}, b$, вероятности которых $p_1, p_2, \dots, p_{k-2}, p_{k-1} + p_k$.

3. Повторяем шаги 1 и 2 до тех пор, пока не получим единственное сообщение, вероятность которого равна 1.

4. Проводя линии, объединяющие сообщения и образующие последовательные подмножества, получаем дерево, в котором отдельные сообщения являются концевыми узлами. Соответствующие им кодовые слова можно определить, приписывая правым ветвям объединения символ «1», а левым — «0». Впрочем, понятия «правые» и «левые» ветви в данном случае относительны.

Введем понятие *кодowego* дерева для множества кодовых слов. Наглядное графическое изображение множества кодовых слов можно получить, установив соответствие между сообщениями и концевыми узлами двоичного дерева. Пример двоичного кодowego дерева изображен на рис. 4.3.

Две ветви, идущие от корня дерева к узлам первого порядка, соответствуют выбору между «0» и «1» в качестве первого символа кодowego слова: левая ветвь соответствует «0», а правая — «1». Две ветви, идущие из узлов первого порядка, соответствуют второму символу кодовых слов, левая означает «0», а правая — «1» и т. д. Ясно, что последовательность символов каждого кодowego слова определяет необходимые правила продвижения от корня дерева до концевого узла, соответствующего рассматриваемому сообщению.

Формально кодовые слова могут быть приписаны также промежуточным узлам. Например, промежуточному узлу второго порядка на рис. 4.3 можно приписать кодовое слово 11, которое соответствует первым двум символам кодовых слов, соответствующих концевым узлам, порождаемым этим узлом. Однако кодовые слова, соответствующие промежуточным узлам, не могут быть использованы для представления сообще-

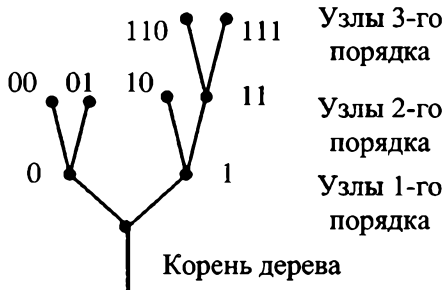


Рис. 4.3. Пример двоичного кодового дерева

ний, так как в этом случае нарушается требование *префиксности* кода. Требование, чтобы только концевые узлы сопоставлялись сообщениям, эквивалентно условию, чтобы ни одно из кодовых слов не совпало с началом (префиксом) более длинного кодового слова.

Префиксный код — это код, обладающий тем свойством, что никакое кодовое слово не является префиксом любого другого кодового слова. Двоичное дерево имеет свойство прямого родства, если каждая вершина, за исключением корня, имеет прямого потомка и если вершины дерева могут быть упорядочены не по возрастанию вероятностей, с которыми каждая вершина оказывается соседней со своим прямым потомком. Двоичный префиксный код является кодом Хаффмана тогда и только тогда, когда кодовое дерево обладает свойством прямого родства.

Любой код, кодовые слова которого соответствуют различным концевым вершинам некоторого двоичного кодового дерева, является префиксным, т. е. однозначно декодируемым.



Пример 4.4. Построить оптимальный двоичный код для ансамбля сообщений, заданного вероятностями 0,25, 0,22, 0,13, 0,11, 0,10, 0,09, 0,07, 0,03. Считая, что ансамбль представляет собой 8 букв русского алфавита А, З, Л, Н, О, П, Р, У, закодировать полученным кодом фразу А РОЗА УПАЛА НА ЛАПУ АЗОРА.

Решение. Фразу можно представить последовательностью двоичных цифр:

100001111101100000111010110100011011010111000001001000110.

Так как в процессе кодирования сообщениям сопоставляются только концевые узлы, полученный код является префиксным и всегда однозначно декодируем.

При равномерных кодах одиночная ошибка в кодовой комбинации приводит к неправильному декодированию только этой комбинации. Одним из серьезных недостатков префиксных кодов является появление трека ошибок, т. е. одиночная ошибка в кодовой комбинации, при определенных обстоятельствах, способна привести к неправильному декодированию не только данной, но и нескольких последующих кодовых комбинаций.

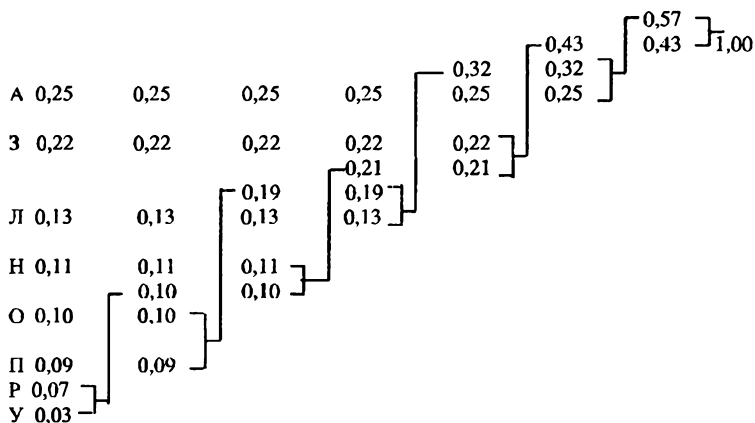


Рис. 4.4. Расчет таблицы для построения кодового дерева

Рассмотрим особенности декодирования оптимального кода Хаффмена, когда в канале связи возможно появление ошибок в приеме двоичных символов.



Пример 4.5. Декодировать последовательность, представленную кодом Хаффмена в примере 4.5, при наличии ошибки в первом разряде первого символа.

Решение. Если вместо 10000111110110 на приемное устройство поступает информация 00000111110110, то это означает, что вместо сообщения А РОЗА на приемном конце ка-

нала связи будет зарегистрировано сообщение УЗОЗА. Таким образом, ошибка в одном разряде кодового слова привела к неправильному декодированию двух символов.

Это явление называется *треком ошибок* при декодировании оптимальных кодов.

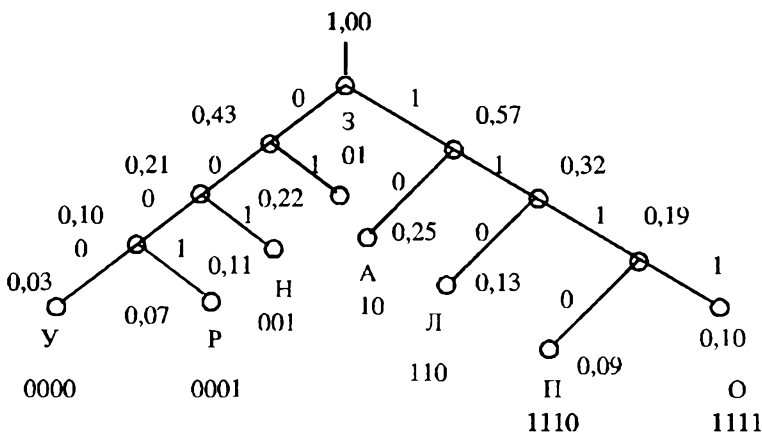


Рис. 4.5. Кодовое дерево префиксного кода Хаффмена

4.6. Канальное кодирование

Для защиты от ошибок в каналах радиосвязи используется помехоустойчивое канальное кодирование сигнала, которое осуществляется путем введения в состав передаваемого сигнала достаточно большого объема дополнительной (избыточной) информации. Как правило, канальное кодирование реализуется в виде четырех процедур:

- блочного кодирования (*block-encoding*);
- сверточного кодирования (*convolutional encoding*);
- перемежения (*interleaving*);
- скремблирования (*scrambling*).

При *блочном кодировании* входная информация разбивается на блоки по k символов, которые преобразуются по определенному закону в n -символьные блоки, где $n > k$. Блочное кодирование предназначено, в основном, для обнаружения одиночных и групповых ошибок в канале связи и в определенных случаях для их исправления.

При *сверточном кодировании* каждый символ входной информационной последовательности, состоящий из k бит, преобразуется в n -битовый символ выходной последовательности, причем $n > k$. Сверточное кодирование является мощным средством борьбы с одиночными ошибками, хотя и не обеспечивает их обнаружения.

При *перемежении* производится изменение порядка следования символов информационной последовательности таким образом, что стоящие рядом символы оказываются разделенными несколькими другими. Перемежение обеспечивает преобразование групповых ошибок в канале связи в одиночные.

Основная причина использования перемежения — резкое снижение корректирующих способностей кодов при появлении в канале *пакетов ошибок*, вызванных сильными помехами и замираниями. Попытки создания кодов, способных исправлять такие серии ошибок, предпринимались неоднократно (коды Хагельбаргера, Файера и др.). Однако из-за вычислительных сложностей, экспоненциально возрастающих с увеличением длины пакета ошибок, они не нашли практического применения. Суть процесса перемежения сводится к перестановке символов кодированной последовательности после демодуляции. Перестановка позволяет так разнести стоящие рядом символы, чтобы они оказались разделены группой других символов, передаваемых в том же блоке данных. Эта операция не вносит избыточности, а только изменяет порядок следования символов или бит.

Рассмотрим пример многошагового перемежителя (*Multi-Stage Interliving*). Исходная кодовая последовательность из L символов разбивается на M блоков по N символов в каждом и преобразуется в матрицу размером $L = M \cdot N$, где N — число символов в строке, в M — число столбцов. Операция блочно-го перемежения заключается в последовательной построчной записи входных данных и считывании этой информации по столбцам. В результате порядок следования символов в выходной последовательности изменяется. Максимальное расстояние, т. е. *глубина перемежения*, на которое разносятся соседние символы, определяется как $d = T \cdot M$, где T — длительность символа. Если время, в течение которого происходит сбой сигнала, меньше глубины перемежения, любой пакет ошибок будет преобразован в группу из M одиночных оши-

бок, которые легко устраняются соответствующим помехоустойчивым кодом.

В общем случае выбор глубины перемежения зависит от двух факторов. С одной стороны, чем больше расстояние между соседними символами, тем большей длины пакет ошибок может быть исправлен. С другой стороны, чем больше глубина перемежения, тем сложнее аппаратно-программная реализация оборудования и больше задержка сигнала.

Скремблирование состоит в преобразовании входной информационной последовательности в выходную путем ее побитного сложения по модулю 2 со специально формируемой шифровой последовательностью. Скремблирование используется для определенной защиты передаваемой информации, а также для аутентификации абонентов.

Структура канального кодирования является общей для всех типов логических каналов, хотя параметры каждой из процедур канального кодирования для различных логических каналов, как правило, отличаются. Поэтому в дальнейшем будут приводиться общие алгоритмы преобразований по каждой из процедур, а затем будут указаны параметры этих преобразований для каждого типа логического канала.

4.7. Достоинства и недостатки эффективных кодов

Достоинства:

1. На передачу сообщений затрачивается минимальное количество символов: $n_c = H(X)_c$.
2. Преобразование сообщений в сигнал производится с минимальной избыточностью.
3. Решается успешно задача согласования источника сигнала с каналом связи.
4. Не требуется введения специальных символов между кодовыми комбинациями (маркеров): свойство неприводимости (префиксные коды), коды без запятой.

Недостатки:

1. Неравномерность кодов.
2. Отсутствие помехозащищенности кодов. Это главный недостаток эффективных кодов (см. пример 4.5).

ВЗАИМОДЕЙСТВИЕ ОТКРЫТЫХ СИСТЕМ

В начале 1980 года Международная Организация по Стандартизации (*International Standards Organization, ISO*) признала необходимость создания общей модели сети. Среди предпосылок к ее появлению можно назвать:

- необходимость эталонной системы, обеспечивающей взаимодействие сетевых средств, предлагаемых различными разработчиками;
- необходимость теоретически обоснованной сетевой модели, в достаточной степени решающей задачу перемещения информации между компьютерами различных систем;
- разбиение общей задачи перемещения информации на более мелкие подзадачи, что позволило бы разработчикам сетевых приложений сконцентрироваться на решении конкретных прикладных задач.

Эталонная модель под названием «Взаимодействие Открытых Систем» (*Open Systems Interconnection, OSI*) была выпущена в 1984 году. OSI описана в стандарте ISO-7498.

Общение двух приложений посредством сети является довольно сложной задачей, которая включает в себя:

- поиск приложения, с которым будем обмениваться информацией;
- установление и поддержание связи;
- обработку сигналов и помех при обмене.

Если бы реализация всех операций, необходимых для обмена информацией, лежала только на приложениях, то создание последних было бы крайне головоломной задачей. Помимо этого, возникла бы проблема согласования транспортных средств для приложений, выпущенных различными

разработчиками. В OSI взаимодействие двух систем представляется в виде двух моделей — горизонтальной и вертикальной:

- в *горизонтальной модели* рассматривается прямое взаимодействие (обмен данными) одинаковых уровней в двух конечных точках; для организации такого взаимодействия в каждой из конечных точек должны поддерживаться одинаковые протоколы для данного уровня;
- в *вертикальной модели* рассматривается взаимодействие (обмен информацией) между соседними уровнями одной системы, при этом каждый уровень может предоставлять свои услуги вышележащему уровню и пользоваться услугами нижележащего уровня.

Модель OSI разделяет задачу сетевого обмена на семь более мелких задач, что упрощает решение. Каждая из подзадач сформулирована таким образом, чтобы для ее решения требовался минимум внешней информации. Каждый уровень модели OSI соответствует своей подзадаче. Из этого следует, что каждый уровень модели в достаточной степени автономен. *Поэтому реальные реализации сетей могут использовать не все уровни, а только часть из них.*

Основные принципы разработки сетевых уровней, соответствующих модели взаимодействия открытых систем:

- каждый уровень соответствует уровню абстракции программного обеспечения. Следовательно, новый сетевой уровень вводится, если ПО требует дополнительный уровень абстракции;
- каждый уровень должен выполнять строго определенную функцию;
- набор функций, выполняемых сетевым уровнем, приводится в соответствие с общепринятыми международными стандартами;
- границы уровня выбираются таким образом, чтобы минимизировать проходящий через них поток данных;
- количество сетевых уровней должно быть достаточно большим, чтобы не заставлять разработчиков приложений размещать различные функции на одном и том же уровне. С другой стороны, избыточное количество уровней чересчур усложняет модель, делая ее необъятной.

Однако только лишь самые низкие, физические, уровни общаются между собой напрямую. Все остальные уровни обращаются к более низким через точку доступа к услугам уровня (*Service Access Point, SAP*). Каждый уровень, принимая информацию от верхнего, добавляет к ней свои данные, которые необходимы для функционирования этого уровня. При прохождении очередного уровня сверху вниз данные получают новое внешнее оформление. На принимающей стороне данные проходят путь, направленный снизу вверх и на каждом уровне соответствующее оформление снимается. Поэтому каждый уровень на принимающей стороне получает данные точно в том виде, в котором они были отправлены соответствующим уровнем на противоположной стороне.

В системах передачи информации идеологической основой стандартизации является *многоуровневый подход* к разработке средств сетевого взаимодействия. Именно на основе этого подхода была создана стандартная семиуровневая *модель взаимодействия открытых систем*, ставшая своего рода универсальным языком сетевых специалистов.

Организация взаимодействия между устройствами сети является сложной задачей. Как известно, для решения сложных задач используется универсальный прием — *декомпозиция*, т. е. разбиение одной задачи на несколько задач-модулей. *Декомпозиция* состоит в четком определении функций каждого модуля, а также порядка их взаимодействия (*интерфейсов*). В результате достигается логическое упрощение задачи, а кроме того, появляется возможность модификации отдельных модулей без изменения остальной части системы.

При декомпозиции часто используют *многоуровневый подход*. Он заключается в следующем:

- все множество модулей, решающих частные задачи, разбивают на группы и упорядочивают по уровням, образующим *иерархию*;
- в соответствии с принципом *иерархии* для каждого промежуточного уровня можно указать непосредственно примыкающие к нему соседние вышележащий и нижележащий уровни;
- группа модулей, составляющих каждый уровень, должна быть сформирована таким образом, чтобы все модули этой группы для выполнения своих задач обращались

с запросами только к модулям соседнего нижележащего уровня;

- с другой стороны, результаты работы всех модулей, отнесенных к некоторому уровню, могут быть переданы только модулям соседнего вышележащего уровня.

Такая иерархическая декомпозиция задачи предполагает четкое определение функции каждого уровня и интерфейсов между уровнями. Интерфейс определяет набор функций, которые нижележащий уровень предоставляет вышележащему. В результате иерархической декомпозиции достигается относительная независимость уровней, а значит, возможность их автономной разработки и модификации. Средства решения задачи организации сетевого взаимодействия, конечно, тоже могут быть представлены в виде иерархически организованного множества модулей. Например, модулям нижнего уровня можно поручить вопросы, связанные с надежной передачей информации между двумя соседними узлами, а модулям следующего, более высокого уровня — транспортировку *сообщений* в пределах всей сети. Очевидно, что последняя задача — организация связи двух любых, не обязательно соседних, узлов — является более общей и поэтому ее можно решить посредством многократных обращений к нижележащему уровню.

5.1. Понятие открытой системы

Модель OSI, как это следует из ее названия, описывает взаимосвязи *открытых систем*. Такой моделью может быть любая система передачи информации, построенная в соответствии с открытыми спецификациями.

Под термином «спецификация» понимают формализованное описание аппаратных или программных компонентов, способов их функционирования, взаимодействия с другими компонентами, условий эксплуатации, ограничений и особых характеристик. Понятно, что не всякая спецификация является стандартом. В свою очередь, под открытыми спецификациями понимаются опубликованные, общедоступные спецификации, соответствующие стандартам и принятые в результате достижения согласия после всестороннего обсуждения заинтересованными сторонами.

Использование при разработке систем открытых спецификаций позволяет третьим сторонам разрабатывать для этих систем различные аппаратные или программные средства расширения и модификации, а также создавать программно-аппаратные комплексы из продуктов разных производителей.

Для реальных систем полная открытость является недостижимым идеалом. Как правило, даже в системах, называемых открытыми, этому определению соответствуют лишь некоторые их части, поддерживающие внешние интерфейсы. Чем больше открытых спецификаций использовано при разработке системы, тем более открытой она является.

Модель OSI касается только одного аспекта открытости, а именно — открытости средств взаимодействия устройств, связанных в систему передачи информации. Здесь под открытой системой понимается сетевое устройство, готовое взаимодействовать с другими сетевыми устройствами с использованием стандартных правил, определяющих формат, содержание и значение принимаемых и отправляемых сообщений. Если две сети построены с соблюдением принципов открытости, то это дает следующие преимущества:

- возможность построения сети из аппаратных и программных средств различных производителей, придерживающихся одного и того же стандарта;
- возможность замены отдельных компонентов сети другими, более совершенными, что позволяет сети развиваться с минимальными затратами;
- возможность легкого сопряжения одной сети с другой;
- простота освоения и обслуживания сети.

Ярким примером открытой системы является международная сеть Internet. Эта сеть развивалась в полном соответствии с требованиями, предъявляемыми к открытым системам. В разработке ее стандартов принимали участие тысячи специалистов-пользователей этой сети из различных университетов, научных организаций и фирм-производителей вычислительной аппаратуры и программного обеспечения, работающих в разных странах. Само название стандартов, определяющих работу сети *Internet*, — *Request For Comments (RFC)*, что можно перевести как «запрос на комментарии», — показывает гласный и открытый характер принимаемых стандартов. В результате сеть *Internet* сумела объединить в себе самое разнообраз-

разное оборудование и программное обеспечение огромного числа сетей, разбросанных по всему миру.

Открытая система — система передачи информации, состоящая из аппаратных и программных продуктов и технологий, разработанных в соответствии с общедоступными и общепринятыми (международными) стандартами.

Обязательными свойствами открытых систем являются:

- переносимость;
- интероперабельность;
- масштабируемость;
- доступность программного и аппаратного обеспечения для развития и реструктуризации.

5.2. Взаимодействие открытых систем

Взаимодействие открытых систем — правила сопряжения систем с открытой архитектурой от различных производителей.

Стандартизация функции взаимосвязи компьютерных сетей на международном уровне осуществляется с середины 70-х годов XX века. В настоящее время сформирована обширная система стандартов. Стандартизация взаимосвязи систем охватывает три уровня описания средств информационного обмена.

На первом уровне, который можно назвать концептуальным, специфицируется эталонная модель взаимосвязи открытых систем OSI RM (*Open Systems Interconnection Reference Model*), в рамках которой определяются основные понятия и общая структура взаимосвязи, описываются принципы построения системы базовых стандартов. Таким образом, эталонная модель формирует методологические основы построения и язык описания стандартов взаимосвязи открытых систем.

На втором уровне определяются спецификации сервиса (услуг), предоставляемого отдельными компонентами архитектуры OSI RM, т. е. на данном уровне стандартизируются спецификации функциональных возможностей отдельных компонент модели.

Наиболее детальным уровнем описания взаимосвязи открытых систем является спецификация протоколов информационного обмена между функциональными элементами эталонной

модели, определяющими правила и форматы взаимодействия элементов. Изучение системы стандартов сетевых технологий, разработанных на основе эталонной модели OSI RM, начнем с самой эталонной модели. Следует отметить, что стандарты сетевых технологий, разработанные в соответствии с эталонной моделью OSI, не ограничиваются стеком сетевых протоколов модели OSI. К ним можно отнести обширное множество стандартов технологий локальных сетей, цифровых сетей с интегральным сервисом ISDN, системы сигнализации № 7 (SS7), сетевых технологий X.25, ATM и Frame Relay, радиосетей стандарта GSM и пр.

В начале 80-х годов ISO признала необходимость создания модели сети, на основе которой поставщики оборудования телекоммуникаций могли создавать взаимодействующие друг с другом сети. В 1984 году такой стандарт был выпущен под названием «Эталонная модель взаимодействия открытых систем» (*Open System Interconnect, OSI*), или OSI/ISO. Эталонная модель OSI стала основной архитектурной моделью для систем передачи сообщений. При рассмотрении конкретных прикладных телекоммуникационных систем производится сравнение их архитектуры с моделью OSI/ISO. Эта модель является наилучшим средством для изучения современной технологии связи. Эталонная модель *OSI* делит проблему передачи информации между абонентами на семь менее крупных и, следовательно, более легко разрешимых задач.

Конкретизация каждой задачи производилась по принципу относительной автономности. Очевидно, автономная задача решается легче. Каждой из семи областей проблемы передачи информации ставится в соответствие один из уровней эталонной модели. Два самых низших уровня эталонной модели *OSI* реализуются аппаратным и программным обеспечением, остальные пять высших уровней, как правило, реализуются программным обеспечением. Эталонная модель OSI описывает, каким образом информация проходит через среду передачи (например, металлические провода) от прикладного процесса-источника (например, по передаче речи) до процесса-получателя.

В качестве примера связи типа OSI предположим, что система А на рис. 5.1 имеет информацию для отправки в систему В. Прикладной процесс системы А общается с уровнем 7

системы *A* (верхний уровень), который сообщается с уровнем 6 системы *A*, который в свою очередь сообщается с уровнем 5 системы *A*, и так далее до уровня 1 системы *A*. Задача уровня 1 — отдавать (а также забирать) информацию в физическую среду. После того, как информация проходит через физическую среду и принимается системой *B*, она поднимается через слои системы *B* в обратном порядке (сначала уровень 1, затем уровень 2 и т. д.), пока она, наконец, не достигнет прикладного процесса системы *B*.

Каждый из уровней сообщается с выше- и нижестоящими уровнями данной системы. Однако для выполнения присущих уровню задач необходимо сообщение с соответствующим уровнем другой системы, т. е. главной задачей уровня 1 системы *A* является связь с уровнем 1 системы *B*; уровень 2 системы *A* сообщается с уровнем 2 системы *B* и т. д.

Уровневая модель OSI исключает прямую связь между соответствующими уровнями разных систем. Следовательно, каждый уровень системы *A* использует услуги, предоставляемые ему смежными уровнями, чтобы осуществить связь с соответствующим ему уровнем системы *B*. Нижестоящий уровень называется источником услуг, а вышестоящий — пользователем услуг. Взаимодействие уровней происходит в так называемой точке предоставления услуг. Взаимоотношения между смежными уровнями отдельной системы показаны на рис. 5.1. Обмен управляющей информацией между соответствующими уровнями разных систем производится в виде обмена специальными «заголовками», добавляемыми к полезной информационной нагрузке. Обычно заголовок предшествует фактической прикладной информации. Каждый нижележащий уровень передающей системы добавляет к поступившему от вышележащего уровня информационному блоку свой заголовок с необходимой управляющей информацией для соответствующего уровня другой системы (рис. 5.1).

В принимающей системе производится анализ данной управляющей информации и удаление соответствующего заголовка перед передачей информационного блока вышележащему уровню. Таким образом, размер информационного блока увеличивается при движении сверху вниз по уровням в передающей системе и уменьшается при движении снизу вверх по уровням в принимающей системе. Эталонная модель OSI

не является реализацией сети. Она только определяет функции протокола каждого уровня. Модель OSI RM разработана международной организацией по стандартизации ISO. Ее описание приведено в документах, имеющих индекс ISO 7498, а также в рекомендации X.200 организации ITU-T (ранее, до 1994 г., называвшейся CCITT). Оба документа являются эквивалентными с технической точки зрения и имеют статус формального международного стандарта. Формирование стандарта OSI RM осуществлялось на протяжении почти десятилетия. Текст первой редакции ISO 7498 публиковался и принимался в качестве международного стандарта по частям с 1984 по 1989 гг. В 1994 г. стандарт пересматривался и претерпел некоторую редакцию и технические исправления. Представленная в документах ISO 7498 и X.200 модель взаимосвязи открытых систем определяет базовые понятия, структуру, семантику, механизмы исполнения телекоммуникационной функции (т. е. функции взаимосвязи удаленных систем посредством обмена данными), нотации для спецификации сервисов сетевых протоколов.

OSI RM предназначена для определения общей основы процесса стандартизации в области взаимосвязи систем, обеспечивающей целостность и взаимную согласованность стандартов. Разработанные на этой основе стандарты позволяют реализовывать унифицированные средства обмена данными между системами в соответствии с согласованными на международном уровне требованиями, определенными в модели OSI RM. Системы, взаимодействующие посредством такого рода стандартных процедур обмена данными, называются «открытыми системами», а реализуемая ими взаимосвязь — «взаимосвязью открытых систем».

5.3. Протокол. Интерфейс. Стек протоколов

Многоуровневое представление средств сетевого взаимодействия имеет свою специфику, связанную с тем, что в процессе обмена *сообщениями* участвуют две стороны, то есть в данном случае необходимо организовать согласованную работу двух «*иерархий*», работающих на разных компьютерах. Оба участника сетевого обмена должны принять множество соглашений. Например, они должны согласовать уровни и

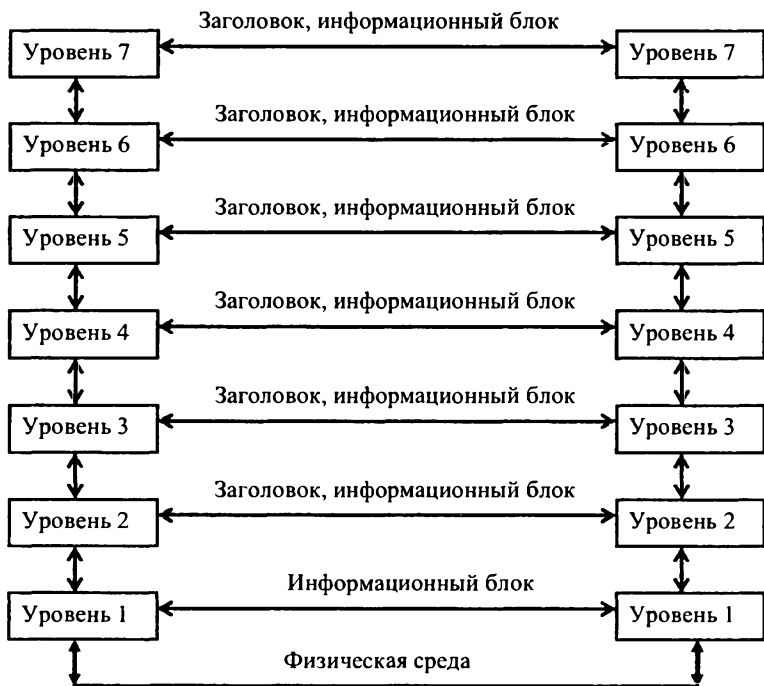


Рис. 5.1. Взаимодействие между уровнями одной системы и формирование информационных блоков для связи между соответствующими уровнями приемной и передающей систем

форму *электрических сигналов*, способ определения длины *сообщений*, договориться о методах контроля достоверности и т. п. Другими словами, соглашения должны быть приняты для всех уровней, начиная от самого низкого — уровня передачи *битов* — до самого высокого, реализующего сервис для пользователей сети.

Сетевые компоненты, лежащие на одном уровне, но в разных узлах, называются *протоколом*. Модули, реализующие *протоколы* соседних уровней и находящиеся в одном узле, также взаимодействуют друг с другом в соответствии с четко определенными правилами с помощью стандартизированных форматов *сообщений*. Эти правила принято называть *интерфейсом*.

Интерфейс — определяет последовательность и формат сообщений, которыми обмениваются сетевые компоненты, лежащие на соседних уровнях в одном узле, а также набор услуг, предоставляемый данным уровнем соседнему уровню. В сущности, *протокол* и *интерфейс* выражают одно и то же понятие, но традиционно в сетях за ними закреплены разные области действия: протоколы определяют правила взаимодействия модулей одного уровня в разных узлах, а *интерфейсы* — модулей соседних уровней в одном узле. Средства каждого уровня должны отрабатывать, во-первых, собственный *протокол*, а во-вторых, *интерфейсы* с соседними уровнями.

Иерархически организованный набор *протоколов*, достаточный для организации взаимодействия узлов в сети, называется *стеком коммуникационных протоколов*. *Коммуникационные протоколы* могут быть реализованы как программно, так и аппаратно. *Протоколы* нижних уровней часто реализуются комбинацией программных и аппаратных средств, а *протоколы* верхних уровней — как правило, чисто программными средствами.

Программный модуль, реализующий некоторый протокол, часто для краткости также называют протоколом. При этом соотношение между протоколом как формально определенной процедурой и протоколом — программным модулем, реализующим эту процедуру, — аналогично соотношению между алгоритмом решения некоторой задачи и программой, решающей эту задачу.

Понятно, что один и тот же алгоритм может быть запрограммирован с разной степенью эффективности. Точно так же и протокол может иметь несколько программных реализаций. Именно поэтому при сравнении протоколов следует учитывать не только логику их работы, но и качество программных решений. Более того, на эффективность взаимодействия устройств в сети влияет качество всей совокупности протоколов, составляющих *стек*, в частности то, насколько рационально распределены функции между протоколами разных уровней и насколько хорошо определены *интерфейсы* между ними.

Протоколы реализуются не только компьютерами, но и другими сетевыми устройствами — *концентраторами*, *мостами*, *коммутаторами*, *маршрутизаторами* и т. д. Действи-

тельно, в общем случае связь компьютеров в сети осуществляется не напрямую, а через различные коммуникационные устройства. В зависимости от типа устройства в нем должны быть встроенные средства, реализующие тот или иной набор протоколов.

Чтобы еще раз пояснить понятия «протокол» и «интерфейс», рассмотрим пример, не имеющий отношения к вычислительным сетям, а именно, обсудим взаимодействие двух предприятий, *А* и *В*. Между этими предприятиями существуют многочисленные договоренности и соглашения, например о регулярных поставках продукции. В соответствии с договоренностью, начальник отдела продаж предприятия *А* регулярно в начале каждого месяца посылает официальное сообщение начальнику отдела закупок предприятия *В* о том, сколько какого товара может быть поставлено в этом месяце. В ответ на это сообщение начальник отдела закупок предприятия *В* посылает заявку установленного образца на нужное количество продукции. Возможно, подобная процедура включает дополнительные согласования; в любом случае, существует установленный порядок взаимодействия, который можно считать «протоколом уровня начальников». Начальники посылают свои сообщения и заявки через секретарей. Порядок взаимодействия начальника и секретаря соответствует понятию межуровневого *интерфейса* «начальник — секретарь». На предприятии *А* обмен документами между начальником и секретарем идет через специальную папку, а на предприятии *В* начальник общается с секретарем по факсу. Таким образом, *интерфейсы* «начальник — секретарь» на этих двух предприятиях отличаются.

После того как сообщения переданы секретарям, начальников не волнует, каким образом эти сообщения будут перемещаться дальше — по обычной почте или электронной, факсом или нарочным. Выбор способа передачи — это уровень компетенции секретарей, они могут решать этот вопрос, не уведомляя о том своих начальников, так как их протокол взаимодействия связан только с передачей поступающих сверху сообщений и не касается содержания этих сообщений. Из рис. 5.1 видно, что в качестве протокола взаимодействия «секретарь — секретарь» используется обмен письмами. При решении иных вопросов начальники могут взаимодействовать по другим правилам-протоколам, но это не повлияет на работу

секретарей, для которых не важно, какие сообщения отправлять, а важно, чтобы они дошли до адресата. Итак, в данном случае мы имеем дело с двумя уровнями — уровнем начальников и уровнем секретарей, и каждый из них имеет собственный протокол, который может быть изменен независимо от протокола другого уровня. В этой независимости протоколов друг от друга и состоит преимущество *многоуровневого подхода*.

5.4. Модель OSI

В начале 80-х годов ряд международных организаций по стандартизации — ISO, ITU-T и некоторые другие — разработали модель, которая сыграла значительную роль в развитии сетей. Эта модель называется моделью ISO/OSI.

Эталонная модель OSI, иногда называемая стеком OSI, представляет собой 7-уровневую сетевую иерархию (рис. 5.1) разработанную Международной организацией по стандартам (*International Standardization Organization* — ISO).

В горизонтальной модели двум программам требуется общий протокол для обмена данными. В вертикальной — соседние уровни обмениваются данными с использованием интерфейсов API.

В результате систематического проектирования архитектуры для среды взаимосвязи открытых систем была определена семиуровневая модель архитектуры OSI RM, включающая следующие уровни:

- Прикладной (*Application* — *A*) — уровень 7;
- Представительный (*Presentation* — *P*) — уровень 6;
- Сеансовый (*Session* — *S*) — уровень 5;
- Транспортный (*Transport* — *T*) — уровень 4;
- Сетевой (*Network* — *N*) — уровень 3;
- Канальный или звена данных (*Data Link* — *DL*) — уровень 2;
- Физический (*Physical* — *Ph*) — уровень 1.

5.5. Физический уровень (уровень 1)

Этот уровень определяет электротехнические, механические, процедурные и функциональные характеристики установления, поддержания и разъединения физического канала между конечными системами. Спецификации физического

уровня определяют такие характеристики, как величины напряжений, параметры синхронизации, скорость передачи физической информации, максимальные расстояния передачи информации, физические соединители и другие аналогичные характеристики. Физической средой в различных телекоммуникационных системах могут быть самые разнообразные средства от простейшей пары проводов до сложной системы передачи синхронной цифровой иерархии. Данный курс лекций посвящен рассмотрению именно физических сред и физического уровня эталонной модели взаимодействия открытых систем.

Физический уровень (Physical layer) имеет дело с передачей битов по *физическим каналам* связи, таким как коаксиальный кабель, витая пара, оптоволоконный кабель, или цифровой территориальный канал. К этому уровню имеют отношение характеристики физических сред передачи данных, такие как полоса пропускания, помехозащищенность, волновое сопротивление и другие. На этом же уровне определяются характеристики *электрических сигналов*, передающих дискретную информацию, такую как крутизна фронтов импульсов, уровни напряжения или тока передаваемого сигнала, тип *кодирования*, скорость передачи сигналов. Кроме того, здесь стандартизируются типы разъемов и назначение каждого контакта.

Физический уровень:

- передача битов по *физическим каналам*;
- формирование *электрических сигналов*;
- кодирование *информации*;
- синхронизация;
- модуляция.

Реализуется аппаратно.

Функции *физического уровня* реализуются во всех устройствах, подключенных к сети. Со стороны компьютера функции *физического уровня* выполняются *сетевым адаптером* или последовательным портом. Примером протокола *физического уровня* может служить спецификация 10Base-T технологии Ethernet, которая определяет в качестве используемого кабеля неэкранированную витую пару категории 3 с волновым сопротивлением 100 Ом, разъем RJ-45, максимальную длину *физического сегмента* 100 метров, манчестерский код для

представления данных в кабеле, а также некоторые другие характеристики среды и *электрических сигналов*.

5.6. Канальный уровень (уровень 2)

Этот уровень, формально называемый информационно-канальным уровнем, обеспечивает надежный транзит данных через физический канал. Выполняя эту задачу, канальный уровень решает вопросы физической адресации (в противоположность сетевой или логической адресации), топологии сети, линейной дисциплины (каким образом конечной системе использовать сетевой канал), уведомления об ошибках, упорядоченной доставки блоков данных и управления потоком информации.

На *физическом уровне* просто пересылаются *биты*. При этом не учитывается, что в тех сетях, в которых линии связи используются (разделяются) попеременно несколькими парами взаимодействующих компьютеров, физическая среда передачи может быть занята. Поэтому одной из задач *канального уровня* (*Data Link layer*) является проверка *доступности среды передачи*.

Другая задача *канального уровня* — реализация механизмов *обнаружения и коррекции ошибок*. Для этого на *канальном уровне биты* группируются в наборы, называемые *кадрами* (*frames*). *Канальный уровень* обеспечивает корректность передачи каждого *кадра*, помещая специальную последовательность *бит* в начало и конец каждого *кадра* для его выделения, а также вычисляет *контрольную сумму*, обрабатывая все байты *кадра* определенным способом, и добавляет *контрольную сумму* к *кадру*. Когда *кадр* проходит по сети, получатель снова вычисляет *контрольную сумму* полученных данных и сравнивает результат с *контрольной суммой* из *кадра*. Если они совпадают, *кадр* считается правильным и принимается. Если же *контрольные суммы* не совпадают, то фиксируется ошибка. *Канальный уровень* может не только обнаруживать ошибки, но и исправлять их за счет повторной передачи поврежденных *кадров*. Необходимо отметить, что функция исправления ошибок для *канального уровня* не является обязательной, поэтому в некоторых протоколах этого уровня она отсутствует, например в *Ethernet* и *frame relay*.

Функции канального уровня

Надежная доставка пакета:

- между двумя соседними станциями в сети с произвольной топологией;
- между любыми станциями в сети с типовой топологией;
- проверка доступности разделяемой среды;
- выделение *кадров* из потока данных, поступающих по сети; формирование *кадров* при отправке данных;
- подсчет и проверка *контрольной суммы*.

Реализуются программно-аппаратно.

В протоколах *канального уровня*, используемых в локальных сетях, заложена определенная структура связей между компьютерами и способы их *адресации*. Хотя *канальный уровень* и обеспечивает доставку *кадра* между любыми двумя узлами локальной сети, он это делает только в сети с определенной топологией связей, именно той топологией, для которой он был разработан. К таким типовым топологиям, поддерживаемым протоколами *канального уровня* локальных сетей, относятся «общая шина», «кольцо» и «звезда», а также структуры, полученные из них с помощью *мостов* и *коммутаторов*. Примерами протоколов *канального уровня* являются протоколы *Ethernet*, *Token Ring*, *FDDI*, *100VG-AnyLAN*.

В локальных сетях протоколы *канального уровня* используются компьютерами, *мостами*, *коммутаторами* и *маршрутизаторами*. В компьютерах функции *канального уровня* реализуются совместными усилиями *сетевых адаптеров* и их *драйверов*.

В глобальных сетях, которые редко обладают регулярной топологией, *канальный уровень* часто обеспечивает обмен сообщениями только между двумя соседними компьютерами, соединенными индивидуальной линией связи. Примерами протоколов «точка-точка» (как часто называют такие протоколы) могут служить широко распространенные протоколы *PPP* и *LAP-B*. В таких случаях для доставки сообщений между конечными узлами через всю сеть используются средства *сетевого уровня*. Именно так организованы сети *X.25*. Иногда в глобальных сетях функции *канального уровня* в чистом виде выделить трудно, так как в одном и том же протоколе они объединяются с функциями *сетевого уровня*. Примерами такого подхода могут служить протоколы технологий *ATM* и *frame relay*.

В целом, *канальный уровень* представляет собой весьма мощный набор функций по пересылке сообщений между узлами сети. В некоторых случаях протоколы *канального уровня* оказываются самодостаточными транспортными средствами, и тогда поверх них могут работать непосредственно протоколы прикладного уровня или приложения, без привлечения средств *сетевого* и транспортного уровней. Например, существует реализация протокола управления сетью SNMP непосредственно поверх *Ethernet*, хотя стандартно этот протокол работает поверх *сетевого протокола* IP и транспортного протокола UDP. Естественно, что применение такой реализации будет ограниченным — она не подходит для *составных сетей* разных технологий, например *Ethernet* и X.25, и даже для такой сети, в которой во всех *сегментах* применяется *Ethernet*, но между *сегментами* существуют петлевидные связи. А вот в двухсегментной сети Ethernet, объединенной *мостом*, реализация SNMP над *канальным уровнем* будет вполне работоспособна.

Тем не менее, для обеспечения качественной транспортировки сообщений в сетях любых топологий и технологий функций *канального уровня* оказывается недостаточно, поэтому в модели OSI решение этой задачи возлагается на два следующих уровня — *сетевой* и транспортный.

Канальный уровень обеспечивает передачу пакетов данных, поступающих от протоколов верхних уровней, узлу назначения, адрес которого также указывает протокол верхнего уровня. Протоколы *канального уровня* оформляют переданные им пакеты в кадры собственного формата, помещая указанный адрес назначения в одно из полей такого *кадра*, а также сопровождая *кадр* *контрольной суммой*. Протокол *канального уровня* имеет локальный смысл, он предназначен для доставки кадров данных, как правило, в пределах сетей с простой топологией связей и однотипной или близкой технологией, например в односегментных сетях *Ethernet* или же в многосегментных сетях *Ethernet* и *Token Ring* иерархической топологии, разделенных только *мостами* и *коммутаторами*. Во всех этих конфигурациях адрес назначения имеет локальный смысл для данной сети и не изменяется при прохождении *кадра* от узла-источника к узлу назначения. Возможность передавать данные между локальными сетями разных технологий связана с тем, что в этих технологиях используются адреса одинакового фор-

мата, к тому же производители сетевых адаптеров обеспечивают уникальность адресов независимо от технологии.

Другой областью действия протоколов *канального уровня* являются связи типа «точка-точка» глобальных сетей, когда протокол *канального уровня* ответственен за доставку *кадра* непосредственному соседу. Адрес в этом случае не имеет принципиального значения, а на первый план выходит способность протокола восстанавливать искаженные и утерянные *кадры*, так как плохое качество территориальных каналов, особенно коммутируемых телефонных, часто требует выполнения подобных действий. Если же перечисленные выше условия не соблюдаются, например связи между *сегментами Ethernet* имеют петлевидную структуру, либо объединяемые сети используют различные способы *адресации*, как в сетях *Ethernet* и X.25, то протокол *канального уровня* не может в одиночку справиться с задачей передачи *кадра* между узлами и требует помощи протокола *сетевого уровня*.

Поскольку сетевые адаптеры работают на физическом и канальном уровнях семиуровневой модели OSI, то канальный уровень принято разделять на два подуровня: MAC (*Media Access Control*) и LLC (*Logical Link Control*). Подуровень MAC — это подуровень управления доступом к среде передачи данных, обеспечивающий корректное совместное использование общей разделяемой среды передачи данных, предоставляя ее в соответствии с определенным алгоритмом в распоряжение той или иной станции. Подуровень LLC отвечает за передачу кадров между узлами с различной степенью надежности, а также реализует функции интерфейса с прилегающим к нему третьим (сетевым) уровнем.

5.7. Сетевой уровень (уровень 3)

Сетевой уровень представляет собой комплексный уровень, который обеспечивает возможность соединения и выбор маршрута между двумя конечными системами. Поскольку две конечные системы, желающие организовать связь, может разделять значительное географическое расстояние и множество подсетей, сетевой уровень является доменом маршрутизации. Протоколы маршрутизации выбирают оптимальные маршруты через последовательность соединенных между собой подсетей.

Традиционные протоколы сетевого уровня передают информацию вдоль этих маршрутов.

Сетевой уровень (*Network layer*) служит для образования единой транспортной системы, объединяющей несколько сетей, причем эти сети могут использовать различные принципы передачи сообщений между конечными узлами и обладать произвольной структурой связей. Функции сетевого уровня достаточно разнообразны. Рассмотрим их на примере объединения локальных сетей.

Протоколы канального уровня локальных сетей обеспечивают доставку данных между любыми узлами только в сети с соответствующей типовой топологией, например топологией иерархической звезды. Это жесткое ограничение, которое не позволяет строить сети с развитой структурой, например сети, объединяющие несколько сетей предприятия в единую сеть, или высоконадежные сети, в которых существуют избыточные связи между узлами. Можно было бы усложнять протоколы канального уровня для поддержания петлевидных избыточных связей, но принцип разделения обязанностей между уровнями приводит к другому решению. Чтобы, с одной стороны, сохранить простоту процедур передачи данных для типовых топологий, а с другой — допустить использование произвольных топологий, вводится дополнительный сетевой уровень.

На сетевом уровне сам термин «сеть» наделяют специфическим значением. В данном случае под сетью понимается совокупность компьютеров, соединенных между собой в соответствии с одной из стандартных типовых топологий и использующих для передачи данных один из протоколов канального уровня, определенный для этой топологии:

Внутри сети доставка данных обеспечивается соответствующим канальным уровнем, а вот доставкой данных между сетями занимается сетевой уровень, который и поддерживает возможность правильного выбора маршрута передачи сообщения даже в том случае, когда структура связей между составляющими сетями имеет характер, отличный от принятого в протоколах канального уровня.

Сети соединяются между собой специальными устройствами, называемыми маршрутизаторами. Маршрутизатор — это устройство, которое собирает информацию о топологии

межсетевых соединений и пересылает пакеты сетевого уровня в сеть назначения. Чтобы передать сообщение от отправителя, находящегося в одной сети, получателю, находящемуся в другой сети, нужно совершить некоторое количество транзитных передач между сетями, или хопов (от слова *hop* — прыжок), каждый раз выбирая подходящий маршрут. Таким образом, маршрут представляет собой последовательность маршрутизаторов, через которые проходит пакет.

Сетевой уровень — доставка пакета:

- между любыми двумя узлами сети с произвольной топологией;
- между любыми двумя сетями в составной сети;
- сеть — совокупность компьютеров, использующих для обмена данными единую сетевую технологию;
- маршрут — последовательность прохождения пакетом маршрутизаторов в составной сети.

Проблема выбора наилучшего пути называется маршрутизацией, и ее решение является одной из главных задач сетевого уровня. Эта проблема осложняется тем, что самый короткий путь — не всегда самый лучший. Часто критерием при выборе маршрута является время передачи данных; оно зависит от пропускной способности каналов связи и интенсивности трафика, которая может с течением времени изменяться. Некоторые алгоритмы маршрутизации пытаются приспособиться к изменению нагрузки, в то время как другие принимают решения на основе средних показателей за длительное время. Выбор маршрута может осуществляться и по другим критериям, таким как надежность передачи.

В общем случае функции сетевого уровня шире, чем функции передачи сообщений по связям с нестандартной структурой, которые мы рассмотрели на примере объединения нескольких локальных сетей. Сетевой уровень также решает задачи согласования разных технологий, упрощения адресации в крупных сетях и создания надежных и гибких барьеров на пути нежелательного трафика между сетями.

Сообщения сетевого уровня принято называть пакетами (*packet*). При организации доставки пакетов на сетевом уровне используется понятие «номер сети». В этом случае адрес получателя состоит из старшей части — номера сети и младшей — номера узла в этой сети. Все узлы одной сети должны

иметь одну и ту же старшую часть адреса, поэтому термину «сеть» на сетевом уровне можно дать и другое, более формальное определение: сеть — это совокупность узлов, сетевой адрес которых содержит один и тот же номер сети.

На сетевом уровне определяется два вида протоколов. Первый вид — сетевые протоколы (*routed protocols*) — реализуют продвижение пакетов через сеть. Именно эти протоколы обычно имеют в виду, когда говорят о протоколах сетевого уровня. Однако часто к сетевому уровню относят и другой вид протоколов, называемых протоколами обмена маршрутной информацией или просто протоколами маршрутизации (*routing protocols*). С помощью этих протоколов маршрутизаторы собирают информацию о топологии межсетевых соединений. Протоколы сетевого уровня реализуются программными модулями операционной системы, а также программными и аппаратными средствами маршрутизаторов.

На сетевом уровне работают протоколы еще одного типа, которые отвечают за отображение адреса узла, используемого на сетевом уровне, в локальный адрес сети. Такие протоколы часто называют протоколами разрешения адресов — *Address Resolution Protocol*, ARP. Иногда их относят не к сетевому уровню, а к канальному, хотя тонкости классификации не изменяют сути.

Примерами протоколов сетевого уровня являются протокол межсетевого взаимодействия IP стека TCP/IP и протокол межсетевого обмена пакетами IPX стека *Novell*.

5.8. Транспортный уровень (уровень 4)

Граница между сеансовым и транспортным уровнями может быть представлена как граница между протоколами высших (прикладных) уровней и протоколами низших уровней. В то время как прикладной, представительный и сеансовый уровни заняты прикладными вопросами, четыре низших уровня решают проблемы транспортировки данных. Транспортный уровень обеспечивает услуги по транспортировке данных, что избавляет высшие слои от необходимости вникать в ее детали. Функцией транспортного уровня является надежная транспортировка данных через сеть. Предоставляя надежные услуги, транспортный уровень обеспечивает механизмы для

установки, поддержания и упорядоченного завершения действия каналов, систем обнаружения и устранения неисправностей транспортировки и управления информационным потоком (с целью предотвращения переполнения системы данными из другой системы).

На пути от отправителя к получателю пакеты могут быть искажены или утеряны. Хотя некоторые приложения имеют собственные средства обработки ошибок, существуют и такие, которые предпочитают сразу иметь дело с надежным соединением. Транспортный уровень (*Transport layer*) обеспечивает приложениям или верхним уровням стека — прикладному и сеансовому — передачу данных с той степенью надежности, которая им требуется. Модель OSI определяет пять классов сервиса, предоставляемых транспортным уровнем. Эти виды сервиса отличаются качеством предоставляемых услуг: срочностью, возможностью восстановления прерванной связи, наличием средств мультиплексирования нескольких соединений между различными прикладными протоколами через общий транспортный протокол, а главное — способностью к обнаружению и исправлению ошибок передачи, таких как искажение, потеря и дублирование пакетов.

Выбор класса сервиса транспортного уровня определяется, с одной стороны, тем, в какой степени задача обеспечения надежности решается самими приложениями и протоколами более высоких, чем транспортный, уровней, а с другой стороны, зависит от того, насколько надежной является система транспортировки данных в сети, обеспечиваемая уровнями, расположенными ниже транспортного, — сетевым, канальным и физическим. Так, например, если качество каналов передачи связи очень высокое и вероятность наличия ошибок, не обнаруженных протоколами более низких уровней, невелика, стоит воспользоваться одним из облегченных сервисов транспортного уровня, не обремененных многочисленными проверками, квитированием и другими приемами повышения надежности. Если же транспортные средства нижних уровней изначально очень ненадежны, то целесообразно обратиться к наиболее развитому сервису транспортного уровня, который работает, используя максимум средств для обнаружения и устранения ошибок, — с помощью предварительного установления логического соединения, отслеживания доставки

сообщений по контрольным суммам и циклической нумерации пакетов, установления тайм-аутов доставки и т. п.

Транспортный уровень — обеспечение доставки информации с требуемым качеством между любыми узлами сети:

- разбивка сообщения сеансового уровня на пакеты, их нумерация;
- буферизация принимаемых пакетов;
- упорядочивание прибывающих пакетов;
- адресация прикладных процессов;
- управление потоком.

Как правило, все протоколы, начиная с транспортного уровня и выше, реализуются программными средствами конечных узлов сети — компонентами их сетевых операционных систем. В качестве примера транспортных протоколов можно привести протоколы TCP и UDP стека TCP/IP и протокол SPX стека *Novell*. Протоколы четырех нижних уровней обобщенно называют сетевым транспортом, или транспортной подсистемой, так как они полностью решают задачу транспортировки сообщений с заданным уровнем качества в составных сетях с произвольной топологией и различными технологиями. Остальные три верхних уровня решают задачи предоставления прикладных сервисов на основании имеющейся транспортной подсистемы.

5.9. Сеансовый уровень (уровень 5)

Этот уровень устанавливает, управляет и завершает сеансы взаимодействия между прикладными задачами. Сеансы состоят из диалога между двумя или более объектами представления. Сеансовый уровень синхронизирует диалог между объектами представительного уровня и управляет обменом информацией между ними. Кроме того, сеансовый уровень предоставляет средства для отправки информации, класса услуг и уведомления в исключительных ситуациях о проблемах сеансового, представительного и прикладного уровней.

Сеансовый уровень (*Session layer*) обеспечивает управление диалогом: фиксирует, какая из сторон является активной в настоящий момент, предоставляет средства синхронизации. Последние позволяют вставлять контрольные точки в длинные передачи, чтобы в случае отказа можно было вернуться

назад к последней контрольной точке, а не начинать все сначала. На практике немногие приложения используют сеансовый уровень, и он редко реализуется в виде отдельных протоколов, хотя функции этого уровня часто объединяют с функциями прикладного уровня и реализуют в одном протоколе.

Сеансовый уровень — управление диалогом объектов прикладного уровня:

- установление способа обмена сообщениями (дуплексный или полудуплексный);
- синхронизация обмена сообщениями;
- организация «контрольных точек» диалога.

5.10. Представительный уровень (уровень 6)

Представительный уровень отвечает за то, чтобы информация, посылаемая из прикладного уровня одной системы, была читаемой для прикладного уровня другой системы. При необходимости представительный уровень осуществляет трансляцию между множеством форматов представления информации путем использования общего формата представления информации. Представительный уровень занят не только форматом и представлением фактических данных пользователя, но также структурами данных, которые используют программы. Поэтому, кроме трансформации формата фактических данных (если она необходима), представительный уровень согласует синтаксис передачи данных для прикладного уровня.

Представительный уровень (*Presentation layer*) имеет дело с формой представления передаваемой по сети информации, не меняя при этом ее содержания. За счет уровня представления информация, передаваемая прикладным уровнем одной системы, всегда понятна прикладному уровню другой системы. С помощью средств данного уровня протоколы прикладных уровней могут преодолеть синтаксические различия в представлении данных или же различия в кодах символов, например в кодах ASCII и EBCDIC. На этом уровне может выполняться шифрование и дешифрование данных, благодаря которым секретность обмена данными обеспечивается сразу для всех прикладных служб. Примером такого протокола является протокол *Secure Socket Layer* (SSL), который

обеспечивает секретный обмен сообщениями для протоколов прикладного уровня стека TCP/IP.

Уровень представления согласовывает представление (синтаксис) данных при взаимодействии двух прикладных процессов:

- преобразование данных из внешнего формата во внутренний;
- шифрование и расшифровка данных.

5.11. Прикладной уровень (уровень 7)

Прикладной уровень — это самый близкий к пользователю уровень OSI. Он отличается от других уровней тем, что не обеспечивает услуг ни одному из других уровней OSI. Он обеспечивает услугами прикладные процессы, лежащие за пределами масштаба модели OSI. Примерами таких прикладных процессов могут служить процессы передачи речевых сигналов, базы данных, текстовые процессоры и т. д. Прикладной уровень идентифицирует и устанавливает наличие предполагаемых партнеров для связи, синхронизирует совместно работающие прикладные процессы, а также устанавливает и согласовывает процедуры устранения ошибок и управления целостностью информации. Прикладной уровень также определяет, имеется ли в наличии достаточно ресурсов для предполагаемой связи.

Прикладной уровень (*Application layer*) — это в действительности просто набор разнообразных протоколов, с помощью которых пользователи сети получают доступ к разделяемым ресурсам, таким как файлы, принтеры или гипертекстовые Web-страницы, а также организуют совместную работу, например с помощью протокола электронной почты. Единица данных, которой оперирует прикладной уровень, обычно называется сообщением (*message*).

Прикладной уровень — набор всех сетевых сервисов, которые предоставляет система конечному пользователю:

- идентификация, проверка прав доступа;
- принт- и файл-сервис, почта, удаленный доступ...

Существует очень много различных служб прикладного уровня. Приведем в качестве примера хотя бы несколько наиболее распространенных реализаций файловых служб: NCP в опера-

ционной системе Novell NetWare, SMB в *Microsoft Windows* NT, NFS, FTP и TFTP, входящие в стек TCP/IP.

5.12. Сетезависимые и сетезависимые уровни

Функции всех уровней модели OSI могут быть отнесены к одной из двух групп: либо к функциям, зависящим от конкретной технической реализации сети, либо к функциям, ориентированным на работу с приложениями.

Три нижних уровня — физический, канальный и сетевой — являются сетезависимыми, то есть протоколы этих уровней тесно связаны с технической реализацией сети и используемым коммуникационным оборудованием. Например, переход на оборудование FDDI означает полную смену протоколов физического и канального уровней во всех узлах сети.

Три верхних уровня — прикладной, представительный и сеансовый — ориентированы на приложения и мало зависят от технических особенностей построения сети. На протоколы этих уровней не влияют какие бы то ни было изменения в топологии сети, замена оборудования или переход на другую сетевую технологию. Так, переход от *Ethernet* к высокоскоростной технологии 100VG-AnyLAN не потребует никаких изменений в программных средствах, реализующих функции прикладного, представительного и сеансового уровней.

Транспортный уровень является промежуточным, он скрывает все детали функционирования нижних уровней от верхних. Это позволяет разрабатывать приложения, не зависящие от технических средств непосредственной транспортировки сообщений.

5.13. Инкапсуляция и обработка пакетов

При продвижении пакета данных по уровням сверху вниз каждый новый уровень добавляет к пакету свою служебную информацию в виде заголовка и, возможно, трейлера (информации, помещаемой в конец сообщения). Эта операция называется инкапсуляцией данных верхнего уровня в пакете нижнего уровня. Служебная информация предназначена для объекта того же уровня на удаленном компьютере, ее формат и интерпретация определяются протоколом данного уровня.

Разумеется, данные, приходящие с верхнего уровня, могут на самом деле представлять собой пакеты с уже инкапсулированными данными еще более верхнего уровня.

С другой стороны, при получении пакета от нижнего уровня он разделяется на заголовок (трейлер) и данные. Служебная информация из заголовка (трейлера) анализируется и в соответствии с ней данные, возможно, направляются одному из объектов верхнего уровня. Тот в свою очередь рассматривает эти данные как пакет со своей служебной информацией и данными для еще более верхнего уровня, и процедура повторяется, пока пользовательские данные, очищенные от всей служебной информации, не достигнут прикладного процесса.

Возможно, что пакет данных не будет доведен до самого верхнего уровня, например в случае, если данный компьютер представляет собой промежуточную станцию на пути между отправителем и получателем. В этом случае объект соответствующего уровня при анализе служебной информации заметит, что пакет на этом уровне адресован не ему (хотя с точки зрения нижележащих уровней он был адресован именно этому компьютеру). Тогда объект выполнит необходимые действия для перенаправления пакета к месту назначения или возврата отправителю с сообщением об ошибке, но в любом случае не будет продвигать данные на верхний уровень.

Модель OSI предложена достаточно давно, однако протоколы, на ней основанные, используются редко, во-первых, в силу своей не всегда оправданной сложности, во-вторых, из-за существования хотя и не соответствующих строго модели OSI, но уже хорошо зарекомендовавших себя стеков протоколов (например, TCP/IP).

Поэтому модель OSI стоит рассматривать, в основном, как опорную базу для классификации и сопоставления протокольных стеков.

Эталонная модель взаимодействия открытых систем OSI (*Open Systems Interconnection*), формализующая стандарты связи и передачи данных в любых линиях передачи информации, разбивает все функции взаимодействия систем на семь уровней. Нижний, или физический, уровень устанавливает тактовую частоту, метод кодирования и обеспечивает преобразование данных из цифровой формы в электромагнитные сигналы, предназначенные для определенной среды передачи, и наоборот. Сигналы, поступающие с физического уровня на второй, канальный уровень, не зависят от типа среды передачи. Качество сигналов, отправляемых по физическим каналам, напротив, непосредственно обусловлено возможностями волокна или витой пары, типом интерфейса (двух-, трех-, четырехпарная передача), наличием внешних и внутренних помех.

6.1. Требования к методам цифрового кодирования

При использовании прямоугольных импульсов для передачи дискретной информации необходимо выбрать такой способ кодирования, который одновременно достигал бы нескольких целей:

- имел при одной и той же битовой скорости наименьшую ширину спектра результирующего сигнала;
- обеспечивал синхронизацию между передатчиком и приемником;
- обладал способностью распознавать ошибки;
- обладал низкой стоимостью реализации.

Более узкий спектр сигналов позволяет на одной и той же линии (с одной и той полосой пропускания) добиваться более

высокой скорости передачи данных. Кроме того, часто к спектру сигнала предъявляется требование отсутствия постоянной составляющей, то есть отсутствия постоянного тока между передатчиком и приемником. В частности, применение различных трансформаторных схем гальванической развязки препятствует прохождению постоянного тока.

Синхронизация передатчика и приемника нужна для того, чтобы приемник точно знал, в какой момент времени необходимо считывать новую информацию с линии связи. Эта проблема в сетях решается сложнее, чем при обмене данными между близко расположенными устройствами, например между блоками внутри компьютера или же между компьютером и принтером. На небольших расстояниях хорошо работает схема, основанная на отдельной тактирующей линии связи, так что информация снимается с линии данных только в момент прихода тактового импульса. В сетях использование этой схемы вызывает трудности из-за неоднородности характеристик проводников в кабелях. На больших расстояниях неравномерность скорости распространения сигнала может привести к тому, что тактовый импульс придет несколько позже или раньше соответствующего сигнала данных, и бит данных будет пропущен или считан повторно. Другой причиной, по которой в сетях отказываются от использования тактирующих импульсов, является экономия проводников в дорогостоящих кабелях.

Поэтому в сетях применяются так называемые самосинхронизирующиеся коды, сигналы которых несут для передатчика указания о том, в какой момент времени нужно осуществлять распознавание очередного бита (или нескольких битов, если код ориентирован более чем на два состояния сигнала). Любой резкий перепад сигнала — так называемый фронт — может служить хорошим указанием для синхронизации приемника с передатчиком. При использовании синусоид в качестве несущего сигнала результирующий код обладает свойством самосинхронизации, так как изменение амплитуды несущей частоты дает возможность приемнику определить момент появления входного кода. Распознавание и коррекцию искаженных данных сложно осуществить средствами физического уровня, поэтому чаще всего эту работу берут на себя протоколы, лежащие выше: канальный, сетевой, транспортный или при-

кладной. С другой стороны, распознавание ошибок на физическом уровне экономит время, так как приемник не ждет полного помещения кадра в буфер, а отбраковывает его сразу при распознавании ошибочных битов внутри кадра. Требования, предъявляемые к методам кодирования, являются взаимно противоречивыми, поэтому каждый из рассматриваемых ниже популярных методов цифрового кодирования обладает своими преимуществами и своими недостатками по сравнению с другими.

Скорость следования импульсов, определяемая тактовой частотой, влияет на скорость передачи данных, но не равна ей. Если обратиться к автомобильной аналогии и представить себе, что тактовая частота — это обороты двигателя автомобиля, то скорость передачи данных можно сопоставить со скоростью его движения. Преобразование одного вида движения в другой обеспечивается кодированием, или — по той же аналогии — работой коробки передач.

Для передачи данных по электропроводным кабелям или радиоканалам требуется выполнить следующие операции:

- синхронизировать тактовые частоты передатчика и приемника;
- преобразовать цифровые сигналы в аналоговые;
- сузить спектр электрических сигналов с помощью фильтров;
- обеспечить передачу отфильтрованных сигналов по линии связи;
- усилить принятые сигналы и восстановить их первоначальную форму;
- и, наконец, преобразовать аналоговые сигналы в цифровые.

Рассмотрим, как связаны между собой тактовая частота и последовательность битов. Тактовая частота, измеряемая в герцах, — это число изменений сигнала в единицу времени. Скорость же передачи данных равна числу передаваемых битов в секунду. Кажущаяся аналогия этих параметров породила множество ошибочных представлений. На самом деле численное равенство скорости передачи данных и тактовой частоты является лишь частным случаем — когда применяется простейшее кодирование сигнала. Для увеличения скорости передачи используются более сложные способы кодирования — элект-

рический сигнал может иметь два, три, пять и более уровней и передаваться по нескольким линиям параллельно, например по всем четырем витым парам кабеля. Каждому протоколу соответствует определенная ширина спектра сигнала и требуется некая пропускная способность информационной магистрали. Однозначного соответствия между мегагерцами и мегабитами в секунду не существует. Скорость передачи данных может равняться тактовой частоте либо в два, четыре и более раз превосходить ее, что зависит от метода кодирования.

Рассмотрим методы кодирования в порядке их усложнения. Цифровая передача данных требует выполнения нескольких обязательных операций:

- синхронизации тактовой частоты передатчика и приемника;
- преобразования последовательности битов в электрический сигнал;
- уменьшения частоты спектра электрического сигнала с помощью фильтров;
- передачи урезанного спектра по каналу связи;
- усиления сигнала и восстановление его формы приемником;
- преобразования аналогового сигнала в цифровой.

6.2. Методы линейного кодирования

Линейное кодирование представляет собой совокупность методов полярного, униполярного, биполярного и манчестерского кодирования.

6.2.1. Униполярное кодирование

Униполярное кодирование имеет два уровня напряжения, соответствующих уровням логических нуля и единицы. Так как униполярное линейное кодирование соответствует уровням TTL-логики, то он находит широкое применение в компьютерах и цифровой технике. Этот вид кодирования используется во внутренних устройствах ЭВМ, где путь сигнала представляет собой небольшую величину, но на длинных расстояниях передачи сигналов он сталкивается с проблемами, связанными с наличием паразитных емкостей в передающей среде.

Рассмотрим взаимосвязь тактовой частоты и битовой последовательности. Битовый поток передается со скоростью, определяемой числом бит в единицу времени. Другими словами, биты в секунду — это число дискретных изменений сигнала в единицу времени. Тактовая частота, измеряемая в герцах, — это число синусоидальных изменений сигнала в единицу времени.

6.2.2. Полярное кодирование (Код NRZ)

Когда цифровое кодирование симметрично около нулевого уровня напряжения, оно называется *Polar Code*. Интерфейс RS-232D использует полярное линейное кодирование. Сигнал не возвращается к нулю, а остается либо положительным, либо отрицательным. Полярное кодирование называют также *None Return to Zero* (NRZ).

Код NRZ — без возврата к нулю — это простейший двухуровневый код. Нулю соответствует нижний уровень, единице — верхний. Информационные переходы происходят на границе битов. Вариант кода NRZI (*Non Return to Zero Inverted*) — соответствует обратной полярности используемой последовательности (рис. 6.1).

Рисунок 6.1 иллюстрирует уже упомянутый ранее метод потенциального кодирования, называемый также кодированием без возвращения к нулю (*Non Return to Zero, NRZ*). Последнее

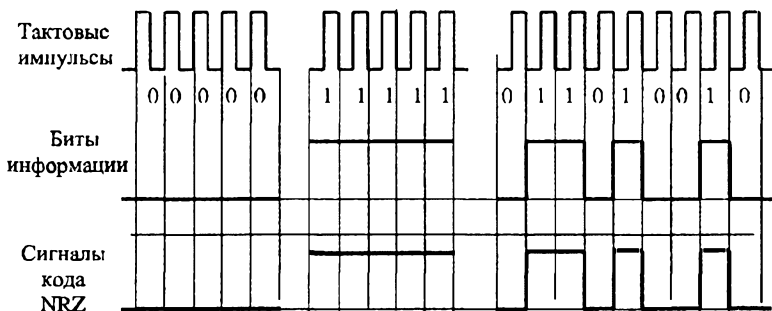


Рис. 6.1. Иллюстрация соотношения между тактовой частотой и частотой следования битов информации и двухуровневых сигналов кода NRZ

название отражает то обстоятельство, что при передаче последовательности единиц сигнал не возвращается к нулю в течение такта (как мы увидим ниже, в других методах кодирования возврат к нулю в этом случае происходит). Метод NRZ прост в реализации, обладает хорошей распознаваемостью ошибок (из-за двух резко отличающихся потенциалов), но не обладает свойством самосинхронизации. При передаче длинной последовательности единиц или нулей сигнал на линии не изменяется, поэтому приемник лишен возможности определять по входному сигналу моменты времени, когда нужно в очередной раз считывать данные. Даже при наличии высокоточного тактового генератора приемник может ошибиться с моментом съема данных, так как частоты двух генераторов никогда не бывают полностью идентичными. Поэтому при высоких скоростях обмена данными и длинных последовательностях единиц или нулей небольшое рассогласование тактовых частот может привести к ошибке в целый такт и, соответственно, считыванию некорректного значения бита.

Другим серьезным недостатком метода NRZ является наличие низкочастотной составляющей, которая приближается к нулю при передаче длинных последовательностей единиц или нулей. Из-за этого каналы связи, не обеспечивающие прямого гальванического соединения между приемником и источником, этот вид кодирования не поддерживают. В результате в чистом виде код NRZ в сетях не используется. Тем не менее используются его различные модификации, в которых устраняют как плохую самосинхронизацию кода NRZ, так и проблемы постоянной составляющей.

Привлекательность кода NRZ, из-за которой имеет смысл заняться его улучшением, состоит в достаточно низкой частоте основной гармоники f_0 . У других методов кодирования, например манчестерского, основная гармоника имеет более высокую частоту.

Несомненное достоинство кода — простота. Сигнал не надо кодировать и декодировать.

Кроме того, скорость передачи данных вдвое превышает тактовую частоту (при тактовой частоте 1 Гц обеспечивается передача двух битов). Наибольшая частота будет фиксироваться при чередовании единиц и нулей. Для других комбинаций

частота будет меньше. При передаче последовательности одинаковых битов частота изменения сигнала равна нулю.

Код NRZ (NRZI) не имеет синхронизации. Это является самым большим его недостатком. Если тактовая частота приемника отличается от частоты передатчика, теряется синхронизация, биты преобразуются, что приводит к большим потерям передаваемой информации.

Для синхронизации начала приема пакета используется стартовый служебный бит, например единица. Наиболее известное применение кода NRZI — стандарт ATM155. Самый распространенный протокол RS-232. Этот стандарт соединения оборудования был разработан в 1969 г. рядом крупных промышленных корпораций и опубликован Ассоциацией электронной промышленности США (*Electronic Industries Association, EIA*) как вариант С рекомендуемого стандарта *Recommended Standard* № 232, применяемого для соединений через последовательный порт персонального компьютера. Он использует код NRZ. Международный союз электросвязи ITU-T подобные рекомендации обозначает как V.24 и V.28. Передача информации ведется байтами по 8 бит, сопровождаемыми стартовыми и стоповыми битами.

Код NRZI называется потенциальным кодом с инверсией при единице (*Non Return to Zero with ones Inverted, NRZI*). Он очень похож на код AMI, однако обладает только двумя уровнями сигнала. При передаче нуля он передает потенциал, который был установлен в предыдущем такте, т. е. не меняет его, а при передаче единицы потенциал инвертируется на противоположный. Он удобен в тех случаях, когда наличие третьего уровня сигнала весьма нежелательно, например в оптических кабелях, где устойчиво распознаются два состояния сигнала — свет и темнота.

Для улучшения потенциальных кодов, подобных AMI и NRZI, используются два метода. Первый метод основан на добавлении в исходный код избыточных битов, содержащих логические единицы. Очевидно, что в этом случае длинные последовательности нулей прерываются, и код становится самосинхронизирующимся для любых передаваемых данных. Исчезает также постоянная составляющая, а значит, еще более сужается спектр сигнала. Но этот метод снижает полезную пропускную способность линии, так как избыточные

единицы пользовательской информации не несут. Другой метод основан на предварительном «перемешивании» исходной информации таким образом, чтобы вероятность появления единиц и нулей на линии становилась близкой. Устройства, или блоки, выполняющие такую операцию, называются скремблерами (*scramble* — свалка, беспорядочная куча). При скремблировании используется известный алгоритм, поэтому приемник, получив двоичные данные, передает их на дескремблер, который восстанавливает исходную последовательность битов. Избыточные биты при этом по линии не передаются. Оба метода относятся к логическому, а не физическому кодированию, так как форму сигналов на линии они не определяют.

6.2.3. Биполярное кодирование (Код RZ)

RZ — это трехуровневый код, обеспечивающий возврат к нулевому уровню после передачи каждого бита информации. Его так и называют кодирование с возвратом к нулю (Return to Zero). Логическому нулю соответствует положительный импульс, логической единице — отрицательный (рис. 6.2).

Информационный переход осуществляется в начале бита, возврат к нулевому уровню — в середине бита. Особенностью кода RZ является то, что в центре бита всегда есть переход

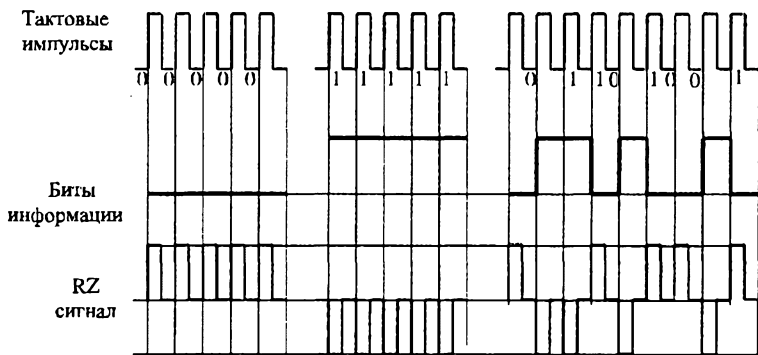


Рис. 6.2. Иллюстрация соотношения между тактовой частотой и частотой следования битов информации и трехуровневых RZ сигналов

(положительный или отрицательный). Следовательно, каждый бит обозначен. Приемник может выделить синхроимпульс (строб), имеющий частоту следования импульсов, из самого сигнала. Привязка производится к каждому биту, что обеспечивает синхронизацию приемника с передатчиком. Такие коды, несущие в себе строб, называются самосинхронизирующимися.

Недостаток кода RZ состоит в том, что он не дает выигрыша в скорости передачи данных. Для передачи со скоростью 10 Мбит/с требуется частота несущей 10 МГц. Кроме того, для различения трех уровней необходимо лучшее соотношение сигнал/шум на входе приемника, чем для двухуровневых кодов.

Наиболее часто код RZ используется в оптоволоконных сетях. При передаче света не существует положительных и отрицательных сигналов, поэтому используют три уровня мощности световых импульсов.

6.2.4. Код AMI

Кроме потенциальных кодов в сетях используются и импульсные коды, в которых данные представлены полным импульсом или же его частью — фронтом. Наиболее простым случаем такого подхода является биполярный импульсный код, в котором единица представлена импульсом одной полярности, а ноль — другой (рис. 6.3). Каждый импульс длится половину такта.

Одной из модификаций метода NRZ является метод биполярного кодирования с альтернативной инверсией (*Bipolar Alternate Mark Inversion*, AMI). В этом методе (рис. 6.3) используются три уровня потенциала — отрицательный, нулевой и положительный. Для кодирования логического нуля используется нулевой потенциал, а логическая единица кодируется либо положительным потенциалом, либо отрицательным, при этом потенциал каждой новой единицы противоположен потенциалу предыдущей.

Код AMI обладает отличными самосинхронизирующими свойствами, но постоянная составляющая может присутствовать, например, при передаче длинной последовательности единиц или нулей. Кроме того, спектр у него шире, чем у потенциальных кодов. Так, при передаче всех нулей

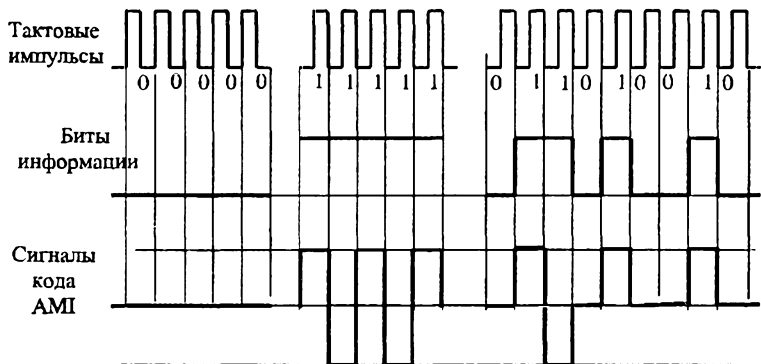


Рис. 6.3. Иллюстрация соотношения между тактовой частотой и частотой следования битов информации и трехуровневых сигналов кода АМІ

или единиц частота основной гармоники кода будет равна N Гц, что в два раза выше основной гармоники кода NRZ и в четыре раза выше основной гармоники кода АМІ при передаче чередующихся единиц и нулей. Из-за слишком широкого спектра биполярный импульсный код используется редко.

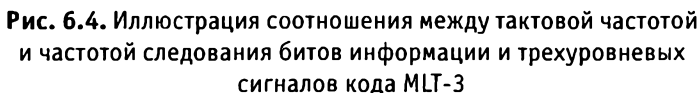
Код АМІ частично ликвидирует проблемы постоянной составляющей и отсутствия самосинхронизации, присущие коду NRZ. Это происходит при передаче длинных последовательностей единиц. В этих случаях сигнал на линии представляет собой последовательность разнополярных импульсов с тем же спектром, что и у кода NRZ, передающего чередующиеся нули и единицы, то есть без постоянной составляющей и с основной гармоникой $N/2$ Гц (где N — битовая скорость передачи данных). Длинные же последовательности нулей так же опасны для кода АМІ, как и для кода NRZ — сигнал вырождается в постоянный потенциал нулевой амплитуды. Поэтому код АМІ требует дальнейшего улучшения, хотя задача упрощается — осталось справиться только с последовательностями нулей.

В целом, для различных комбинаций битов на линии использование кода АМІ приводит к более узкому спектру сигнала, чем для кода NRZ, а значит, и к более высокой пропускной способности линии. Например, при передаче чере-

В коде АМІ используются не два, а три уровня сигнала на линии. Дополнительный уровень требует увеличения мощности передатчика примерно на 3 дБ для обеспечения той же достоверности приема битов на линии передачи, что является общим недостатком кодов с несколькими состояниями сигнала по сравнению с кодами, которые различают только два состояния.

Код трехуровневой передачи MLT-3 (*Multi Level Transmis- sion* — 3) имеет много общего с кодом NRZ. Важнейшее отличие — три уровня сигнала.

Единице соответствует переход с одного уровня сигнала на другой. Изменение уровня сигнала происходит последовательно с учетом предыдущего перехода. Максимальной частоте сигнала соответствует передача последовательности единиц.



При передаче нулей сигнал не меняется. Информационные переходы фиксируются на границе битов. Один цикл сигнала вмещает четыре бита (рис. 6.4).

Недостаток кода MLT-3, как и кода NRZ, — отсутствие синхронизации. Эту проблему решают с помощью преобразования данных, которое исключает длинные последовательности нулей и устраняет возможность рассинхронизации.

6.2.6. Код Манчестер-II

Код Манчестер-II или манчестерский код получил наибольшее распространение в локальных сетях. Он также относится к самосинхронизирующимся кодам, но в отличие от кода RZ имеет не три, а только два уровня, что обеспечивает лучшую помехозащищенность (рис. 6.5).

В локальных сетях до недавнего времени самым распространенным методом кодирования был манчестерский код. Он применяется в технологиях *Ethernet* и *Token Ring*. В манчестерском коде для кодирования единиц и нулей используется перепад потенциала, т. е. фронт импульса. При манчестерском кодировании каждый такт делится на две части. Информация кодируется перепадами потенциала, происходящими в середине каждого такта. Единица кодируется перепадом от низкого уровня сигнала к высокому, а ноль — обратным перепадом. В начале каждого такта может происходить служебный перепад сигнала, если нужно представить несколько

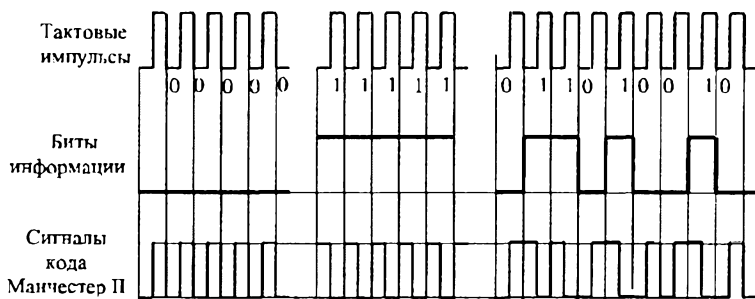


Рис. 6.5. Иллюстрация соотношения между тактовой частотой и частотой следования битов информации и двухуровневых сигналов кода Манчестер II

единиц или нулей подряд. Так как сигнал изменяется по крайней мере один раз за такт передачи одного бита данных, то манчестерский код обладает хорошими самосинхронизирующими свойствами. Полоса пропускания манчестерского кода уже, чем у биполярного импульсного. У него также нет постоянной составляющей, а основная гармоника в худшем случае (при передаче последовательности единиц или нулей) имеет частоту $N \text{ Гц}$, а в лучшем (при передаче чередующихся единиц и нулей) она равна $N/2 \text{ Гц}$, как и у кодов АМІ или NRZ. В среднем ширина полосы манчестерского кода в полтора раза уже, чем у биполярного импульсного кода, а основная гармоника колеблется вблизи значения $3N/4$. Манчестерский код имеет еще одно преимущество перед биполярным импульсным кодом. В последнем для передачи данных используются три уровня сигнала, а в манчестерском — два.

Логическому нулю соответствует переход на верхний уровень в центре битового интервала, логической единице — переход на нижний уровень. Логика кодирования хорошо видна на примере передачи последовательности единиц или нулей. При передаче чередующихся битов частота следования импульсов уменьшается в два раза.

Информационные переходы в середине бита остаются, а граничные (на границе битовых интервалов) — при чередовании единиц и нулей отсутствуют. Это выполняется с помощью последовательности запрещающих импульсов. Эти импульсы синхронизируются с информационными и обеспечивают запрет нежелательных граничных переходов.

Изменение сигнала в центре каждого бита позволяет легко выделить синхросигнал. Самосинхронизация дает возможность передачи больших пакетов информации без потерь из-за различий тактовой частоты передатчика и приемника.

Большое достоинство манчестерского кода — отсутствие постоянной составляющей при передаче длинной последовательности единиц или нулей. Благодаря этому гальваническая развязка сигналов выполняется простейшими способами, например, с помощью импульсных трансформаторов.

Частотный спектр сигнала при манчестерском кодировании включает только две несущие частоты. Для десятимегабитного протокола — это 10 МГц при передаче сигнала, состоящего из одних нулей или одних единиц, и 5 МГц — для сигнала с

чередованием нулей и единиц. Поэтому с помощью полосовых фильтров можно легко отфильтровать все другие частоты.

Код Манчестер-II нашел применение в оптоволоконных и электропроводных сетях. Самый распространенный протокол локальных сетей *Ethernet 10 Мбит/с* использует именно этот код.

6.2.7. Кодирование данных 4В5В

Протоколы, использующие код NRZ, чаще всего дополняют кодированием данных 4В5В (от четырехбитовой последовательности к пятибитовой). В отличие от кодирования сигналов, которое использует тактовую частоту и обеспечивает переход от импульсов к битам и наоборот, кодирование данных преобразует одну последовательность битов в другую.

В коде 4В5В используется пятибитовая основа для передачи четырехбитовых информационных сигналов. Пятибитовая схема дает 32 (два в пятой степени) двухразрядных буквенно-цифровых символа, имеющих значение в десятичном коде от 00 до 31. Для данных отводится четыре бита или 16 символов.

Четырехбитовый информационный сигнал перекодируется в пятибитовый сигнал в кодере передатчика. Преобразованный сигнал имеет 16 значений для передачи информации и 16 избыточных значений. В декодере приемника пять битов расшифровываются как информационные и служебные сигналы. Для служебных сигналов отведены девять символов, семь символов — исключены.

Исключены комбинации, имеющие более трех нулей (01 — 00001, 02 — 00010, 03 — 00011, 08 — 01000, 16 — 10000). Такие сигналы интерпретируются символом V и командой приемника VIOLATION — сбой. Команда означает наличие ошибки из-за высокого уровня помех или сбоя передатчика. Единственная комбинация из пяти нулей (00 — 00000) относится к служебным сигналам, означает символ Q и имеет статус QUIET — отсутствие сигнала в линии.

Кодирование данных решает две задачи — синхронизации и улучшения помехоустойчивости. Синхронизация происходит за счет исключения последовательности более трех нулей. Высокая помехоустойчивость достигается контролем принятых данных на пятибитовом интервале.

**Соответствие исходных
и результирующих кодов 4B5B**

Исходный код	Результирующий код	Исходный код	Результирующий код
0000	11110	1000	10010
0001	01001	1001	10011
0010	10100	1010	10110
0011	10101	1011	10111
0100	01010	1100	11010
0101	01011	1101	11011
0110	01110	1110	11100
0111	01111	1111	11101

Цена кодирования данных — снижение скорости передачи полезной информации. В результате добавления одного избыточного бита на четыре информационных, эффективность использования полосы частот в протоколах с кодом MLT-3 и кодированием данных 4B5B уменьшается соответственно на 25 %.

При совместном использовании кодирования сигналов MLT-3 и данных 4B5B скорость передачи снижается, так как оказывается равной передаче 3 битов информации на 1 герц несущей частоты сигнала. Такая схема используется в протоколе TP-PMD.

6.2.8. Потенциальный код 2B1Q

В локальных сетях до недавнего времени самым распространенным методом кодирования был так называемый манчестерский код (рис. 6.5). Он применяется в технологиях *Ethernet* и *Token Ring*.

На рис. 6.6 показан потенциальный код с четырьмя уровнями сигнала для кодирования данных. Это код 2B1Q, название которого отражает его суть — каждые два бита (2B) передаются за один такт сигналом, имеющим четыре состояния (1Q). Паре бит 00 соответствует потенциал $-2,5B$; паре бит 01 — потенциал $-0,833 B$; паре 11 — потенциал $+0,833 B$; а паре 10 — потенциал $+2,5 B$. При этом способе кодирования

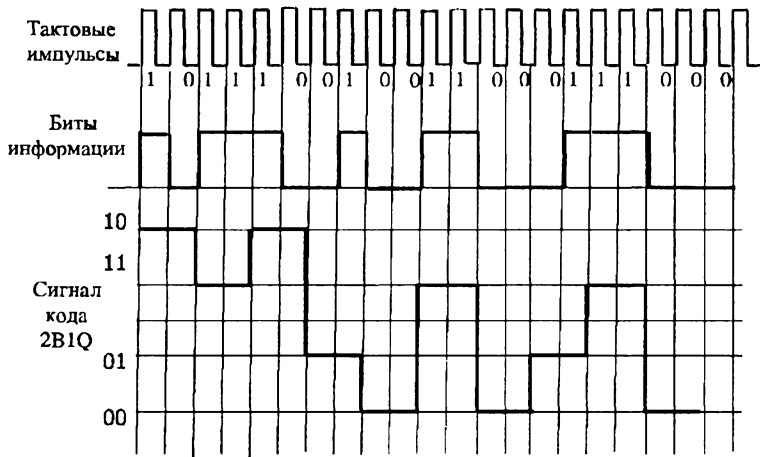


Рис. 6.6. Иллюстрация соотношения между тактовой частотой и частотой следования битов информации и пятиуровневых сигналов кода PAM-5

требуется дополнительные меры по борьбе с длинными последовательностями одинаковых пар битов, так как при этом сигнал превращается в постоянную составляющую. При случайном чередовании битов спектр сигнала в два раза уже, чем у кода NRZ, так как при той же битовой скорости длительность такта увеличивается в два раза. Таким образом, с помощью кода 2B1Q можно по одной и той же линии передавать данные в два раза быстрее, чем с помощью кода AMI или NRZI. Однако для его реализации мощность передатчика должна быть выше, чтобы четыре уровня четко различались приемником на фоне помех.

6.2.9. Код PAM-5

Рассмотренные выше схемы кодирования сигналов были битовыми. При битовом кодировании каждому биту соответствует значение сигнала, определяемое логикой протокола.

При байтовом кодировании уровень сигнала задают два бита и более.

В пятиуровневом коде PAM-5 (рис. 6.7) используется 5 уровней амплитуды и двухбитовое кодирование. Для каж-

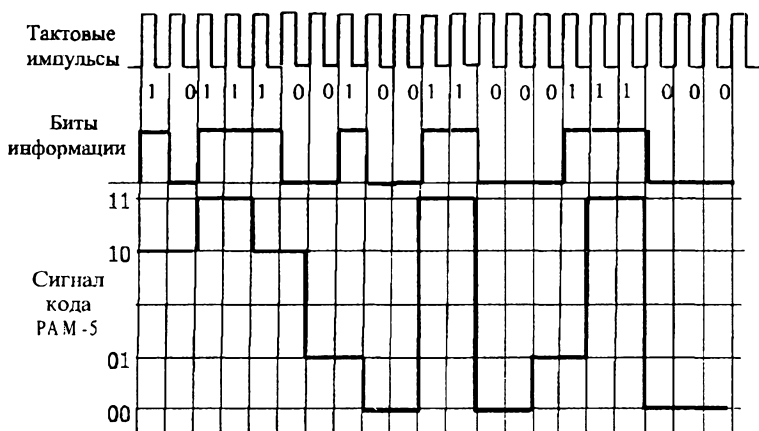


Рис. 6.7. Иллюстрация соотношения между тактовой частотой и частотой следования битов информации и пятиуровневых сигналов кода

дой комбинации задается уровень напряжения. При двухбитовом кодировании для передачи информации необходимо четыре уровня (два во второй степени — 00, 01, 10, 11). Передача двух битов одновременно обеспечивает уменьшение в два раза частоты изменения сигнала.

Пятый уровень добавлен для создания избыточности кода, используемого для исправления ошибок. Это дает дополнительный резерв соотношения сигнал/шум 6 дБ.

Код РАМ-5 используется в протоколе 1000 *Base T Gigabit Ethernet*. Данный протокол обеспечивает передачу данных со скоростью 1000 Мбит/с при ширине спектра сигнала всего 125 МГц.

Как это достигается? Данные передаются по всем четырем парам одновременно. Следовательно, каждая пара должна обеспечить скорость 250 Мбит/с. Максимальная частота спектра несущей при передаче двухбитовых символов кода РАМ-5 составляет 62,5 МГц. С учетом передачи первой гармоники протоколу 1000 *Base T* требуется полоса частот до 125 МГц. Но о несущей, гармониках и полосе частот следует говорить отдельно.

6.3. Ширина магистральной — требуемая полоса частот

Скорость движения зависит не только от возможностей автомобиля, но и от качества магистральной. То же самое справедливо и для передачи данных.

Рассмотрим возможности информационных магистралей.

Кодирование сигналов — это способ преобразования тактовой частоты в скорость передачи данных. С какой целью выполняют преобразование? Для того чтобы увеличить скорость без изменения частотного диапазона канала связи. Кодирование требует использования более сложной приемопередающей аппаратуры. Это минус. Зато при переходе к более скоростным протоколам можно использовать те же кабели. А это уже большой плюс.

Например, протокол *Fast Ethernet 100 Base T4* обеспечивает работу линии передачи информации со скоростью 100 Мбит/с на кабелях категории 3 (16 МГц). *Gigabit Ethernet 1000 Base T* реализован таким образом, чтобы на базе каналов категории 5 (100 МГц), имеющий некоторый резерв, передавать 1000 Мбит/с.

6.4. Ширина спектра сигнала

Сигнал, имеющий синусоидальную форму, называется гармоническим. Его параметры определяются частотой и амплитудой. Чем больше форма сигнала отличается от синусоиды, тем больше гармонических составляющих он несет. Частоты гармоник кратны частоте несущей. Стандарты электропитания, например, требуют оценки качества напряжения сигнала вплоть до тридцатой гармоники.

Диапазон частот сложного сигнала называется спектральной шириной сигнала. Он включает основную составляющую, которая определяет несущую, и гармонические составляющие, которые задают форму импульсов.

Восстановление формы импульсов производится на аппаратном уровне, поэтому гармонические составляющие убирают с помощью фильтров.

Спектральная ширина сигнала зависит от тактовой частоты, метода кодирования и характеристик фильтра передатчика.

Из рис. 6.8 видно, как правильный выбор метода кодирования позволяет уменьшить частоту несущей. Для трех мето-

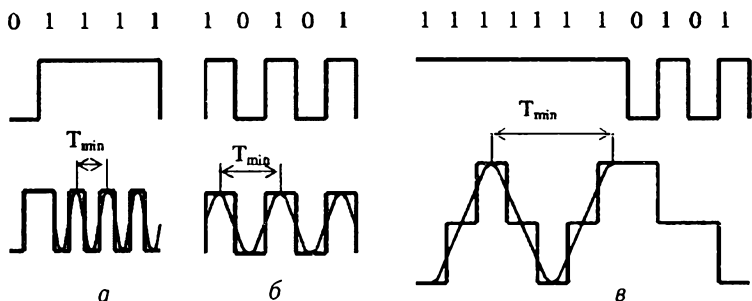


Рис. 6.8. Изменение длительности минимального периода тактовой частоты в зависимости от метода кодирования последовательности битов информации: код Манчестер II (а), коды NRZ и NRZI (б), код MLT-3 (в)

дов кодирования приведены ситуации, требующие максимальной частоты несущей, т. е. $f_{max} = 1/T_{min}$. Один герц несущей передает один бит (рис. 6.8, а) при манчестерском кодировании, два бита (01) кода NRZ или NRZI (рис. 6.8, б) и четыре бита (1111) кода MLT-3 (рис. 6.8, в). *Фактор кодирования* (передача) составляет соответственно один, два и четыре.

Другие комбинации битов требуют меньшей частоты. Например, при чередовании нулей и единиц частота несущей кода MLT-3 уменьшается еще в два раза, последовательность большой продолжительности нулей может уменьшить частоту несущей до нуля.

Спектральную ширину сигнала не следует путать с тактовой частотой. Тактовая частота — это метроном, задающий ритм работы всей системы передачи информации. На рис. 6.8 тактовой частоте соответствует скорость чередования битов. Ширина спектра сигнала в данной аналогии — это диапазон частот сигнала, позволяющий восстановить исходный импульсный сигнал.

Если при аналоговой передаче широкополосного сигнала попытаться передать его по телефону, придется пожертвовать значительной частью спектра, так как линия связи, имеющая узкую полосу пропускания, не пропустит верхние гармоники. При этом качество воспроизведения этого сигнала на выходе узкополосного канала связи ухудшится.

При цифровой передаче для восстановления исходного сигнала требуется меньше гармоник, чем для аналогового. Технология передачи и приема цифровых сигналов позволяет восстановить исходный сигнал по несущей спектра. Однако для уменьшения коэффициента ошибок необходимо присутствие первой гармоники, что удваивает ширину спектра или частотный диапазон передаваемых сигналов.

6.5. Логическое кодирование

Логическое кодирование используется для улучшения потенциальных кодов типа AMI, NRZI или 2Q1B. Логическое кодирование должно заменять длинные последовательности битов, приводящие к постоянному потенциалу, вкраплениями единиц. Как уже отмечалось выше, для логического кодирования характерны два метода — избыточные коды и скремблирование.

6.5.1. Избыточные коды

Избыточные коды основаны на разбиении исходной последовательности битов на порции, которые часто называют символами. Затем каждый исходный символ заменяется новым, который имеет большее количество битов, чем исходный. Например, логический код 4B5B, используемый в технологиях FDDI и *Fast Ethernet*, заменяет исходные символы длиной в 4 *бита* символами длиной в 5 *бит*. Так как результирующие символы содержат избыточные биты, то общее количество битовых комбинаций в них больше, чем в исходных. Так, в коде 4B5B результирующие символы могут содержать 32 битовых комбинации, в то время как исходные символы — только 16 (табл. 6.1). Поэтому в результирующем коде можно отобрать 16 таких комбинаций, которые не содержат большого количества нулей, а остальные считать запрещенными кодами (*code violation*).

Кроме устранения постоянной составляющей и придания коду свойства самосинхронизации, избыточные коды позволяют приемнику распознавать искаженные биты. Если приемник принимает запрещенный код, значит, на линии произошло искажение сигнала. После разбиения получившийся код 4B/5B передается по линии путем физического кодирования по одному из методов потенциального кодирования,

чувствительному только к длинным последовательностям нулей. Символы кода 4B5B длиной 5 *бит* гарантируют, что при любом их сочетании на линии не может встретиться более трех нулей подряд.

Буква В в названии кода означает, что элементарный сигнал имеет два состояния (от английского *binary* — двоичный). Имеются также коды и с тремя состояниями сигнала, например, в коде 8B6T для кодирования 8 *бит* исходной информации используется код из 6 сигналов, каждый из которых имеет три состояния. Избыточность кода 8B6T выше, чем кода 4B5B, так как на 256 исходных кодов приходится $3^6 = 729$ результирующих символов.

Использование таблицы перекодировки является очень простой операцией, поэтому этот подход не усложняет сетевые адаптеры и интерфейсные блоки коммутаторов и маршрутизаторов.

Для обеспечения заданной пропускной способности линии передатчик, использующий избыточный код, должен работать с повышенной тактовой частотой. Так, для передачи кодов 4B5B со скоростью 100 *Мбит/с* передатчик должен работать с тактовой частотой 125 *МГц*. При этом спектр сигнала на линии расширяется по сравнению со случаем, когда по линии передается чистый, не избыточный код. Тем не менее, спектр избыточного потенциального кода оказывается уже спектра манчестерского кода, что оправдывает дополнительный этап логического кодирования, а также работу приемника и передатчика на повышенной тактовой частоте.

6.5.2. Скремблирование

Скремблирование наряду с избыточным кодированием является способом логического кодирования информации. Методы скремблирования заключаются в побитном вычислении результирующего кода на основании битов исходного кода и полученных в предыдущих тактах битов результирующего кода.

Например, скремблер может реализовывать следующее соотношение:

$$B_i = A_i \oplus B_{i-3} \oplus B_{i-5}.$$

Здесь B_i — двоичная цифра результирующего кода, полученная на i -м такте работы скремблера, A_i — двоичная цифра

исходного кода, поступающая на i -м такте на вход скремблера, B_{i-3} и B_{i-5} — двоичные цифры результирующего кода, полученные на предыдущих тактах работы скремблера, соответственно на 3 и на 5 тактов ранее текущего такта, и объединенные операцией исключающего ИЛИ (сложение по модулю 2).

Например, для исходной последовательности 110110000001 скремблер даст следующий результирующий код (первые три цифры результирующего кода будут совпадать с исходным кодом, так как еще нет нужных предыдущих цифр): Таким образом, на выходе скремблера появится последовательность 110001101111, в которой нет последовательности из шести нулей, присутствовавшей в исходном коде. После получения результирующей последовательности приемник передает ее дескремблеру, который восстанавливает исходную последовательность на основании обратного соотношения:

$$C_i = B_i \oplus B_{i-3} \oplus B_{i-5} = (A_i \oplus B_{i-3} \oplus B_{i-5}) \oplus B_{i-3} \oplus B_{i-5} = A_i$$



Пример 6.1. На вход скремблера, работающего по соотношению

$$B_i = A_i \oplus B_{i-3} \oplus B_{i-5},$$

где B_i — двоичная цифра результирующего кода, полученная на i -м такте скремблера, A_i — двоичная цифра исходного кода, поступающая на i -м такте на вход скремблера, B_{i-3} и B_{i-5} — двоичные цифры результирующего кода, полученные на предыдущих тактах работы скремблера, соответственно на 3 и 5 тактов ранее текущего такта, $\oplus$ — операция исключающего ИЛИ (сложение по модулю 2), поступает последовательность 110110000001. Получить результирующий код.

Решение. Первые 3 цифры результирующего кода будут совпадать с первыми 3 цифрами исходного кода, так как еще нет нужных предыдущих цифр. Поэтому

$B_1 = A_1 = 1$, $B_2 = A_2 = 1$, $B_3 = A_3$. Далее вычислим значения остальных цифр. Имеем

$$B_4 = A_4 \oplus B_1 = 1 \oplus 1 = 0,$$

$$B_5 = A_5 \oplus B_2 = 1 \oplus 1 = 0,$$

$$B_6 = A_6 \oplus B_3 \oplus B_1 = 0 \oplus 0 \oplus 1 = 1,$$

$$B_7 = A_7 \oplus B_4 \oplus B_2 = 0 \oplus 0 \oplus 1 = 1,$$

$$B_8 = A_8 \oplus B_5 \oplus B_3 = 0 \oplus 0 \oplus 0 = 0,$$

$$B_9 = A_9 \oplus B_6 \oplus B_4 = 0 \oplus 1 \oplus 0 = 1,$$

$$B_{10} = A_{10} (B_7 (B_5 = 0) (0 = 1,$$

$$B_{11} = A_{11} (B_8 (B_6 = 0) (0 (1 = 1,$$

$$B_{12} = A_{12} (B_9 (B_7 = 1) (1 (1 = 1.$$

Различные алгоритмы скремблирования отличаются количеством слагаемых, дающих цифру результирующего кода, и сдвигом между слагаемыми. Так, в сетях ISDN при передаче данных от сети к абоненту используется преобразование со сдвигами в 5 и 23 позиций, а при передаче данных от абонента в сеть — со сдвигами 18 и 23 позиций.

Существуют и более простые методы борьбы с последовательностями единиц, также относимые к классу скремблирования. Для улучшения кода *Bipolar AMI* используются два метода, основанные на искусственном искажении последовательности нулей запрещенными символами. Рис. 6.9 иллюстрирует использование метода *B8ZS* (*Bi polar with 8-Zeros Substitution*) и метода *HDB3* (*High-Density Bipolar 3-Zeros*) для корректировки кода *AMI*. Исходный код состоит из двух длинных последовательностей нулей: в первом случае — из 8, а во втором — из 5. Код *B8ZS* исправляет только последовательности, состоящие из 8 нулей. Для этого он после первых трех нулей вместо оставшихся пяти нулей вставляет пять цифр: $V-1^*-0-V-1^*$. Здесь V обозначает сигнал единицы, запрещенной для данного такта полярности, т. е. сигнал, не изменяющий полярность предыдущей единицы, 1^* — сигнал единицы корректной полярности (знак звездочки отмечает тот факт, что в исходном коде в этом такте была не единица, а ноль). В результате на 8 тактах приемник наблюдает два искажения — очень маловероятно, что это случилось из-за шума на линии или других сбоев передачи. Поэтому приемник считает такие нарушения кодировкой 8 последовательных нулей и после приема заменяет их исходными 8 нулями. Код *B8ZS* построен так, что его постоянная составляющая равна нулю при любых последовательностях двоичных цифр.

Код *HDB3* исправляет любые четыре подряд идущих нуля в исходной последовательности. Правила формирования кода *HDB3* более сложные, чем кода *B8ZS*. Каждые четыре нуля

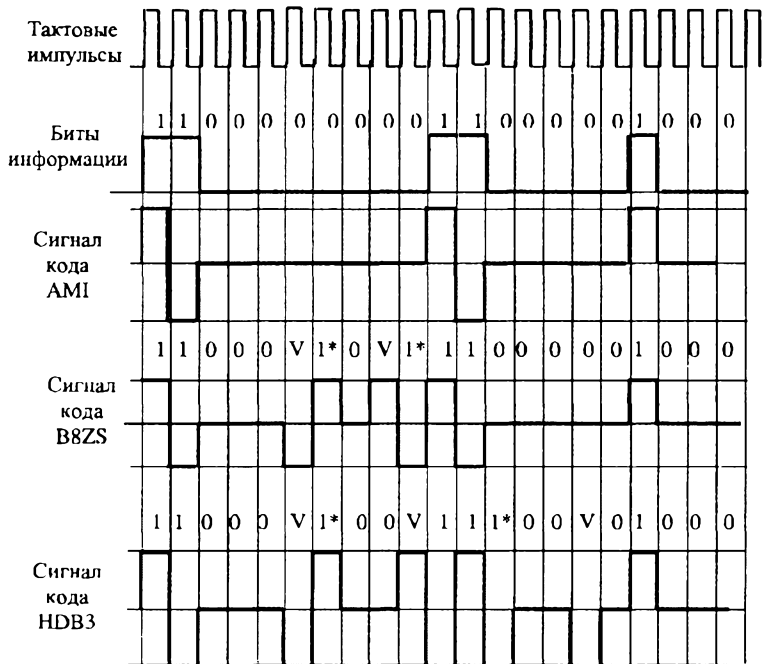


Рис. 6.9. Соотношения между тактовой частотой и частотой следования битов информации и сигналов кодов AMI, B8ZS и HDB3

заменяются четырьмя сигналами, в которых имеется один сигнал *V*. Для подавления постоянной составляющей полярность сигнала *V* чередуется при последовательных заменах. Кроме того, для замены используются два образца четырех-тактных кодов. Если перед заменой исходный код содержал нечетное число единиц, задействуется последовательность 000*V*, а если число единиц было четным — последовательность 1*00*V*. Коды, полученные из потенциального путем логического кодирования, обладают более узким спектром, чем манчестерский, даже при повышенной тактовой частоте (на рисунке спектр кода 4B5B должен был бы примерно совпадать с кодом B8ZS, но он сдвинут в область более высоких частот, так как его тактовая частота повышена на 1/4 по сравнению с другими кодами). Этим объясняется применение

ние потенциальных избыточных и скремблированных кодов в современных технологиях, подобных *FDDI*, *Fast Ethernet*, *Gigabit Ethernet*, *ISDN* и т. п. вместо манчестерского и биполярного импульсного кодирования.

Улучшенные потенциальные коды обладают достаточно узкой полосой пропускания для любых последовательностей единиц и нулей, которые встречаются в передаваемых данных. На рис. 6.10 приведены спектры сигналов разных кодов, полученные при передаче произвольных данных, в которых различные сочетания нулей и единиц в исходном коде равновероятны.

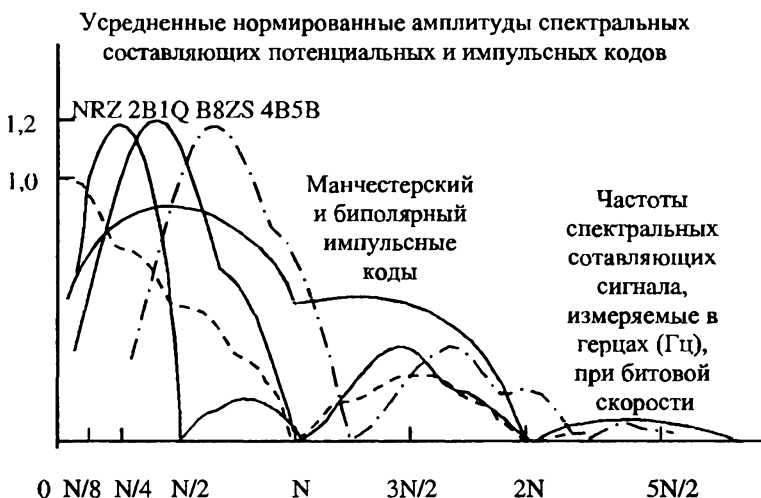


Рис. 6.10. Спектры потенциальных и импульсных кодов

6.6. Выводы

На широкополосных каналах связи применяются потенциальные и импульсные методы кодирования, в которых данные представлены различными уровнями постоянного потенциала сигнала либо полярностями импульса или его фронта.

При использовании потенциальных кодов особое значение приобретает задача синхронизации приемника с передатчиком, так как при передаче длинных последовательностей нулей или

единиц сигнал на входе приемника не изменяется и приемнику сложно определить момент съема очередного бита данных.

Наиболее простым потенциальным кодом является код без возвращения к нулю (NRZ), однако он не является самосинхронизирующимся и создает постоянную составляющую. Наиболее популярным импульсным кодом является манчестерский код, в котором информацию несет направление перепада сигнала в середине каждого такта. Манчестерский код применяется в технологиях *Ethernet* и *Token Ring*.

Для улучшения свойств потенциального кода NRZ используются методы логического кодирования, исключаящие длинные последовательности нулей. Эти методы основаны:

- на введении избыточных битов в исходные данные (коды типа 4B5B);
- скремблировании исходных данных (коды типа 2B1Q).

Улучшенные потенциальные коды обладают более узким спектром, чем импульсные, поэтому они находят применение в высокоскоростных технологиях, таких как FDDI, *Fast Ethernet*, *Gigabit Ethernet*.

7.1. Помехоустойчивые коды и их основные параметры

Проблема повышения соответствия принятой информации той информации, которая передавалась по каналу связи, обусловлена тем, что реальные каналы связи не соответствуют требованиям, предъявляемым при передаче информации.

7.1.1. Принцип построения помехоустойчивых кодов

Простые коды, называемые *примитивными*, характеризуются тем, что для передачи информации используются все *кодированные слова (комбинации)*, количество которых равно $N = q^n$ (q — основание кода, а n — длина кода). Поэтому даже один ошибочно принятый символ приводит к замене одного кодированного слова другим и, следовательно, к неправильному приему сообщения в целом.

Помехоустойчивыми называются коды, позволяющие обнаруживать и/или исправлять ошибки в кодированных словах, которые возникают при передаче по каналам связи. Эти коды строятся таким образом, что для передачи сообщения используется лишь часть кодированных слов, которые отличаются друг от друга более чем в одном символе. Эти кодированные слова называются *разрешенными*. Все остальные кодированные слова не используются и относятся к числу *запрещенных*.

Применение помехоустойчивых кодов для повышения верности передачи данных связано с решением задач кодирования и декодирования. Задача кодирования заключается в получении при передаче для каждой k -элементной комбинации из множества q^k соответствующего ей кодированного слова длиной n из множества q^n . Задача декодирования состоит в получении

k -элементной комбинации из принятого n -разрядного кодового слова при одновременном обнаружении или исправлении ошибок.

7.1.2. Классификация помехоустойчивых кодов

Помехоустойчивые коды можно разделить на 4 основные группы:

- избыточные коды с обнаружением ошибок при передаче информации;
- избыточные коды с исправлением ошибок;
- блочные коды;
- непрерывные коды.

Избыточные коды с исправлением ошибок могут использоваться как для исправления независимых ошибок, так и для исправления ошибок, объединенных в пакеты.

Блочные коды бывают равномерными и неравномерными, делимыми и неделимыми. Когда используются только *равномерные коды*, то в этом случае недополучение одного из битов кодового слова свидетельствует об ошибочности передачи. Таким образом, постоянство длины кодовой цепочки оказывается дополнительным средством контроля правильности передачи. Равномерные коды используются во внутренних компьютерных линиях связи для увеличения скорости передачи за счет параллельной передачи нескольких бит по шинам фиксированной ширины.

При *блочном кодировании* передаваемые двоичные сообщения представляют собой последовательности отдельных блоков — *кодовых комбинаций*, которыми кодируются знаки (или группы знаков) первичного алфавита. Если все кодовые комбинации имеют одинаковую длину, код называется *равномерным*; если нет — *неравномерным*.

Следующий классификационный признак, относящийся как к блочным, так и к непрерывным кодам, подразделяет коды на *разделимые* и *неразделимые*. *Разделимыми* называются коды, в которых информационные и проверочные биты располагаются в строго определенных позициях. В *неразделимых* кодах такой определенности нет, что затрудняет их кодирование и декодирование. Поэтому практический интерес представляют в основном разделимые коды, а из неразделимых — только *коды с постоянным весом*.

Третий классификационный признак относится только к блочным разделимым кодам — они подразделяются на *систематические* (линейные) и *несистематические*.

В *несистематических* (нелинейных) кодах информационные и проверочные биты либо вообще не имеют связи, либо эта связь нелинейна — такие коды применяются крайне редко.

Разделимые коды делятся на *линейные* и *нелинейные*. Линейные коды называют еще *систематическими*, в которых информационные и проверочные биты связаны линейными соотношениями, описываемыми линейными уравнениями.

В помехоустойчивых кодах k информационных бит дополняются r проверочными, в результате чего коды становятся *избыточными*, количественной мерой которых является *относительная избыточность помехоустойчивого кода*:

$$F(n, k) = (n - k)/k = r/k.$$

Функция $F(n, k)$ показывает, какая часть переданной кодовой комбинации не содержит первичной информации.

При блочном кодировании последовательность элементарных сообщений разбивается на отрезки и каждому отрезку ставится в соответствие определенная последовательность (*блок*) кодовых символов, называемых кодовой комбинацией. Множество кодовых комбинаций, возможных при данном способе кодирования, и есть блочный код. Длина кода может быть как постоянной, так и переменной.

При декодировании удобнее (проще) иметь дело с равномерным кодом, поэтому именно он используется в помехоустойчивом кодировании. *Непрерывные* (цепные) коды представляют собой непрерывную последовательность бит, не разделяемую на блоки (информационные и проверочные биты в них чередуются по определенному правилу). Наиболее часто в линиях связи используются систематические коды.

В разделимых кодах, обозначаемых (n, k) , символы делятся на информационные k и проверочные $n - k$. Число кодовых комбинаций равно 2^k .

К неразделимым относятся коды, символы которых нельзя разделить на информационные и проверочные.

Линейные коды — это такие коды, в которых поразрядная сумма по модулю 2 любых кодовых слов также является кодовым словом. Линейный код называется систематическим, если

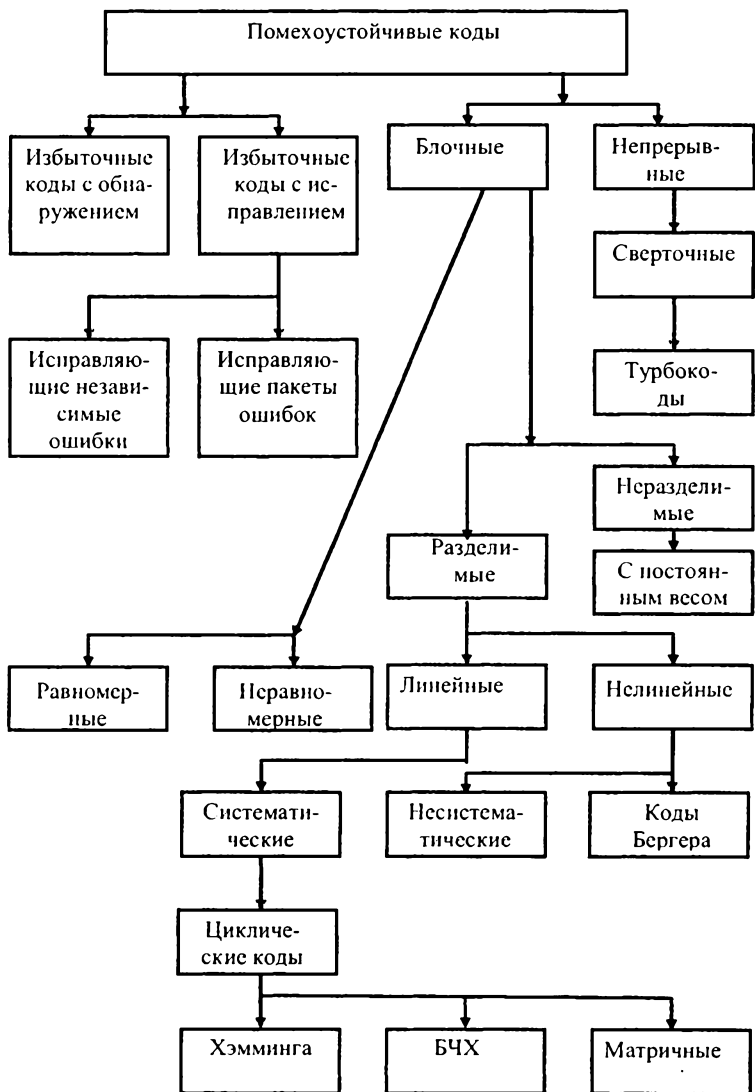


Рис. 7.1. Классификация помехоустойчивых кодов

первые его k символов являются информационными, а остальные $n - k$ — проверочными.

Среди линейных систематических кодов наиболее простой код $(n, n - k)$, содержащий 1 проверочный символ, который равен сумме по модулю 2 всех информационных символов. Этот код называется *кодом с проверкой на четность* и позволяет обнаружить все сочетания ошибок нечетной кратности.



Пример 7.1. Создать код обнаружения ошибок при передаче трехразрядных информационных слов, введя четвертый разряд для индикации бита положительной четности. Какие ошибки можно обнаружить с помощью данного кода? Вычислить вероятность необнаруженной ошибки сообщения, предполагая, что все символьные ошибки являются независимыми событиями и вероятность ошибки в канальном символе равна $p = 10^{-3}$.

Решение. Код представляет собой таблицу, приведенную на рис. 7.2. Код может выявить все комбинации с одной или тремя ошибками. Вероятность необнаруженной ошибки равна вероятности появления где-либо в кодовом слове двух или четырех ошибок:

$$P = C_4^2 p^2 (1 - p^2) + C_4^4 p^4 = 6 \times 10^{-6}.$$

Можно сделать вывод, что теперь по каналу нужно передавать не трехразрядные слова, а четырехразрядные. К тому же сохраняется значительная вероятность появления необнаруженной ошибки, так как двойные ошибки в каждом слове этот код обнаружить не в состоянии.

Циклические коды являются подклассом линейных кодов и представляют собой наборы, образованные циклической перестановкой любой кодовой последовательности, также являющейся кодовой комбинацией. Это свойство позволяет упростить кодирующее и декодирующее устройства, особенно при обнаружении ошибок и при исправлении одиночной ошибки.

Информационные слова	Биты четности
000	0
001	1
010	1
011	0
100	1
101	0
110	0
111	1

Рис. 7.2. Код с проверкой на четность

Код, в котором кодовая комбинация, полученная путем циклического сдвига разрешенной кодовой комбинации, является также разрешенной кодовой комбинацией, называется циклическим (*полиномиальным, кодом с циклическими избыточными проверками — ЦИП*). Сдвиг осуществляется справа налево, при этом крайний левый символ переносится в конец комбинации.

Циклический код относится к линейным, блочным, корректирующим, равномерным кодам.

В циклических кодах кодовые комбинации представляются в виде многочленов, что позволяет свести действия над кодовыми комбинациями к действиям над многочленами (используя аппарат полиномиальной алгебры).

Циклические коды являются разновидностью систематических кодов и поэтому обладают всеми их свойствами. Первоначально они были созданы для упрощения схем кодирования и декодирования. Их эффективность при обнаружении и исправлении ошибок обеспечила им широкое применение на практике. Циклические коды используются в ЭВМ при последовательной передаче данных.

Примерами циклических кодов являются коды Хэмминга, Боуза-Чоудхури-Хоквингема и другие.

Примером нелинейного кода является код Бергера, у которого проверочные символы представляют двоичную запись числа двоичных единиц в последовательности информационных символов. Пример такого кода: 00000, 00101, 01001, 10001, 10110, 11111, 11010. Коды Бергера применяются в ассиметричных каналах связи. В симметричных каналах они обнаруживают все одиночные ошибки и некоторую часть многократных.

Непрерывные коды характеризуются тем, что операции кодирования и декодирования производятся над непрерывной последовательностью символов без разбиения ее на блоки. Среди таких кодов наиболее распространены сверточные коды.

В некоторых каналах возможно образование *пакетов ошибок*, т. е. нескольких следующих подряд и связанных друг с другом ошибок — для защиты от них требуются иные, нежели для независимых ошибок методы кодирования.

Различают каналы с независимыми и группирующимися ошибками. Соответственно этому помехоустойчивые коды

можно разбить на исправляющие независимые ошибки и исправляющие пакеты ошибок. Для исправления пакетов разработано много весьма эффективных кодов. На практике целесообразно использовать коды, исправляющие независимые ошибки вместе с устройством перемежения символов или декорреляции ошибок. При этом символы кодовой комбинации не передаются друг за другом, а перемешиваются с символами других кодовых комбинаций.

Если интервал между символами, принадлежащими одной кодовой комбинации, сделать больше, чем «память» канала, то ошибки в пределах одной кодовой комбинации можно считать независимыми, что и позволяет использовать коды, исправляющие независимые ошибки.

7.2. Понятие о конечных полях

Полем называют множество элементов, на котором определены две операции. Одна из них называется сложением и обозначается $a + b$, а другая — умножением и обозначается $a \cdot b$, даже если эти операции не являются обычными операциями сложения и умножения чисел. Для того чтобы множество элементов, на котором заданы операции сложения и умножения, можно было назвать полем, необходимо, чтобы по каждой из этих операций выполнялись все групповые аксиомы, а также выполнялся дистрибутивный закон, т. е. для трех любых элементов поля a , b , c были справедливы равенства $a \cdot (b + c) = a \cdot b + a \cdot c$ и $(b + c) \cdot a = b \cdot a + c \cdot a$. Кроме того, по каждой операции группа должна быть коммутативной, т. е. должно выполняться $a + b = b + a$ и $a \cdot b = b \cdot a$. Следует заметить, что групповые свойства по операции умножения справедливы для всех ненулевых элементов поля.

Поля с конечным числом элементов q называют полями Галуа по имени их первого исследователя Эвариста Галуа и обозначают $GF(q)$.

Число элементов поля q называют порядком поля. Конечные поля используются для построения большинства известных кодов и их декодирования. В зависимости от значения q различают простые или расширенные поля. Поле называют простым, если q — простое число. Для обозначения простых чисел будем использовать символ p . Простое поле представ-

ляет собою набор чисел по модулю p : $0, 1, 2, \dots, p - 1$, а операции сложения и умножения выполняются по модулю p .

Порядком поля называется число его элементов. Поле F называется конечным, если оно имеет конечный порядок. В противном случае поле называется бесконечным.

Примеры бесконечных полей $\langle Q; +, * \rangle$, $\langle R; +, * \rangle$, $\langle C; +, * \rangle$, где Q , R , C обозначают множества рациональных, вещественных и комплексных чисел соответственно, а операции $+$ и $*$ являются обычными операциями сложения и умножения по модулю простого числа p . Такие простые поля называются простыми и обозначаются F_p .

Характеристикой поля F называется наименьшее положительное целое число p такое, что в поле F справедливо равенство

$$1 + 1 + \dots + 1 = 0,$$

т. е. единица складывается p раз и в сумме получается нуль.

Поскольку поле не содержит делителей нуля, характеристика p всегда является простым числом. Если такого числа не существует, то поле F имеет характеристику, равную нулю. Любое конечное поле имеет характеристику, отличную от нуля.

Наименьшее число элементов, образующих поле, равно 2. Такое поле должно содержать 2 единичных элемента: 0 относительно операции сложения и 1 относительно операции умножения. Это поле обозначается $GF(2)$ и называется *двоичным*. Правила сложения и умножения для элементов $GF(2)$ приводятся рис. 7.3.

Поле, составленное из элементов 0, 1, 2, обозначается $GF(3)$ и называется *троичным* полем. Правила сложения и умножения для поля $GF(3)$ приведены на рис. 7.4.

Формирование таблиц, по которым выполняются правила сложения и умножения, производится приведением результата сложения или умножения чисел, записанных в строках или столбцах, по модулю p , т. е. в качестве результата операции принимается остаток от деления полученного числа на p .

Анализируя состав таблиц, приведенных на рис. 7.3 и 7.4, легко убедиться, что 0 и 1 как единичные элементы по операции сложения и умножения соответственно не изменяют значения других элементов поля по соответствующей операции. Кроме того, видно, что для каждого элемента по операции сложения и для ненулевых элементов по операции умножения имеются обратные.

+	0	1
0	0	1
1	1	0

a

*	0	1
0	0	0
1	0	1

б

Рис. 7.3. Правила сложения (*a*) и умножения (*б*) для элементов $GF(2)$

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

a

*	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

б

Рис. 7.4. Правила сложения (*a*) и умножения (*б*) для элементов троичного поля $GF(3)$

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

a

*	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

б

Рис. 7.5. Попытка построения правил сложения (*a*) и умножения (*б*) для $p = 4$, которое не является простым

Из рис. 7.5, *б* видно, что для элемента 2 по операции умножения отсутствует обратный, т. е. набор чисел 0, 1, 2, 3 не является полем при введении операции по модулю 4. Такой результат объясняется тем, что 4 не является простым числом. Для поля $GF(5)$ с элементами 0, 1, 2, 3, 4 правила сложения и умножения приведены на рис. 7.6.

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

а

*	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

б

Рис. 7.6. Правила сложения (а) и умножения (б) для поля $GF(5)$

Изучим возможность построения полей с элементами в виде последовательностей чисел. Определим условия, при которых последовательности длины m с элементами из поля $GF(p)$ образуют поле.

Порядком произвольного элемента α некоторого конечно-го поля называется наименьшее целое положительное число k такое, что $\alpha^k = 1$. В конечном поле $GF(p^m)$ для элемента β порядка k все элементы $1, \beta, \beta^2, \dots, \beta^{k-1}$ различны. Поэтому порядок каждого элемента поля $GF(p^m)$ конечен и не превышает числа $p^m - 1$. Элемент β поля $GF(p^m)$ называется *примитивным*, если его порядок равен $p^m - 1$. Многочлен, корнем которого является примитивный элемент, называется *примитивным* многочленом. Заметим, что не всякий неприводимый многочлен является примитивным.



Пример 7.2. Показать, что многочлен $f(x) = x^4 + x^3 + 1$ неприводим над полем $GF(2)$, составленным из элементов 0 и 1.

Решение. Действительно, так как элементы 0 и 1 не являются корнями, то $f(x)$ не имеет линейных многочленов x и $x + 1$ в качестве делителей. Легко проверить, что единственный неприводимый над полем $GF(2)$ многочлен второй степени $x^2 + x + 1$ также не делит $f(x)$. Следовательно, многочлен $f(x)$ неприводим над полем $GF(2)$.



Пример 7.3. Построить конечное поле $GF(2^4)$, используя неприводимый над $GF(2)$ полином $f(x) = x^4 + x + 1$.

Решение. Поле $GF(2^4)$ состоит из полиномов степени, меньшей 4:

0	x	x^2	x^3
1	$x + 1$	$x^2 + 1$	$x^3 + 1$
		$x^2 + x$	$x^3 + x$
		$x^2 + x + 1$	$x^3 + x + 1$
			$x^3 + x^2$
			$x^3 + x^2 + 1$
			$x^3 + x^2 + x$
			$x^3 + x^2 + x^2 + 1.$

Определим на множестве элементов операции умножения и взятия обратного элемента. Особенно удобно производить эти операции с помощью представления всех ненулевых элементов поля $GF(2)$ в виде степеней некоторого примитивного элемента α . Выберем его. Легко проверить, что в качестве α можно взять x . Действительно, все его степени по модулю $f(x)$ различны между собой:

$x^1 = x$	$x^2 = x^2$	$x^3 = x^3$
$x^4 = x + 1$	$x^5 = x^2 + x$	$x^6 = x^3 + x^2$
$x^7 = x^3 + x + 1$	$x^8 = x^2 + 1$	$x^9 = x^3 + x$
$x^{10} = x^2 + x + 1$	$x^{11} = x^3 + x^2 + x$	$x^{12} = x^2 + x^2 + x + 1$
$x^{13} = x^3 + x^2 + 1$	$x^{14} = x^{13} + 1$	$x^{15} = 1.$

Следовательно, порядок x равен 15.

Поясним, как получилась эта таблица. Построение первых трех строк тривиально. Чтобы построить значение, находящееся в четвертой строке, надо разделить x^4 на неприводимый полином $f(x) = x^4 + x + 1$. Получившийся остаток от деления x^4 на неприводимый полином $f(x)$ и является искомым значением для x^4 . Аналогично получены остальные строки построенной таблицы.

$$\begin{array}{r|l}
 x^4 & x^4 + x + 1 \\
 \oplus & \hline
 x^4 + x + 1 & 1 \\
 \hline
 & x + 1
 \end{array}$$

Представим полученное поле с помощью таблицы, в которой число i для элемента $\gamma = a^i$ называется логарифмом по основанию выбранного примитивного элемента a . Логарифм элемента 0 полагают обычно равным $-\infty$.

Логарифм	Степень примитивного элемента	Многочлен	Вектор
$-\infty$	0	0	0000
0	1	1	1000
1	6	x	0100
2	6^2	x^2	0010
3	6^3	x^3	0001
4	6^4	$x + 1$	1100
5	6^5	$x^2 + x$	0110
6	6^6	$x^3 + x^2$	0011
7	6^7	$x^3 + x + 1$	1101
8	6^8	$x^2 + 1$	1010
9	6^9	$x^3 + x$	0101
10	6^{10}	$x^2 + x + 1$	1110
11	6^{11}	$x^3 + x^2 + x$	0111
12	6^{12}	$x^3 + x^2 + x + 1$	1111
13	6^{13}	$x^3 + x^2 + 1$	1011
14	6^{14}	$x^3 + 1$	1001

Сложение элементов поля является обычным сложением по модулю 2 и не зависит от выбора примитивного элемента в поле. Например,

$$\alpha^7 + \alpha^{11} = (x^3 + x + 1) + (x^3 + x^2 + x) = x^2 + 1 = \alpha^8.$$

Умножение ненулевых элементов поля, представленных в виде степеней примитивного элемента, проводится путем сложения показателей степеней по модулю 15. Например,

$$(x^3 + x^2) \cdot (x^{13} + x^2 + 1) = \alpha^6 \cdot \alpha^{13} = \alpha^{19(mod 15)} = \alpha^4 = x + 1.$$

Операция умножения, в отличие от сложения зависит от выбора многочлена $f(x)$. Тем не менее, какие бы неприводимые многочлены одинаковой степени не использовались для построения поля, все построенные поля будут изоморфны между собой.

Нахождение обратного элемента рассмотрим на конкретном примере. Найдем обратный элемент для многочлена $x^3 + x^2 + 1 = \alpha^{13}$, т. е. такой ненулевой элемент α^k , что

$$\alpha^{13} \cdot \alpha^k = 1.$$

Для этого запишем

$$(x^3 + x^2 + 1)^{-1} = \alpha^{13} = \alpha^{15-13} = \alpha^2 = x^2.$$

Легко проверить, что решение найдено верно:

$$\begin{aligned} (x^3 + x^2 + 1) \cdot (x^2) &= x^5 + x^4 + x^2 = \\ &= ((x^2 + x) + (x + 1) + 1^2) \bmod(f(x)) = 1 \bmod(f(x)). \end{aligned}$$

Возможность находить для заданного ненулевого многочлена обратный обеспечивается неприводимостью многочлена $f(x)$ (аналогичная и в общем случае для любого поля $GF(q)$).

Нетрудно убедиться, что в построенном поле элементы α^2 , α^4 также примитивные, а элементы α^3 , α^5 примитивными не являются. Например, степени элемента α^3 порождают не все поле, а только элементы α^2 , α^6 , α^9 , α^{12} , $\alpha^{15} = 1$.

Заметим, что многочлен $f(x)$ в нашем случае является примитивным, так как x , его корень по построению, является примитивным элементом α .



Пример 7.4. Построить конечное поле $GF(2^4)$ с помощью неприводимого над полем $GF(2)$ многочлена $f(x) = x^4 + x^3 + x^2 + x + 1$.

Найдем некоторый примитивный элемент поля $GF(2^4)$. Например, элемент x не является примитивным элементом, так как его порядок равен 5, что меньше $2^4 - 1 = 15$. Проведем расчеты аналогично тому, как это мы сделали в предыдущем примере, и представим их результаты в виде таблицы.

Логарифм	Степень примитивного элемента	Многочлен	Вектор
$-\infty$	0	0	0000
0	1	1	1000
1	6	$x+1$	1100
2	6^2	x^2+1	1010
3	6^3	x^3+x^2+x+1	1111
4	6^4	x^3+x^2+x	0111
5	6^5	x^3+x^2+1	1011
6	6^6	x^3	0001
7	6^7	x^2+x+1	1110
8	6^8	x^3+1	1001
9	6^9	x^2	0010
10	6^{10}	x^3+x^2	0011
11	6^{11}	x^3+x+1	1101
12	6^{12}	x	0100
13	6^{13}	x^2+1	0110
14	6^{14}	x^3+x	0101

Таким образом, мы убедились, что с помощью различных многочленов можно найти различные, но эквивалентные поля Галуа $GF(2^4)$.

Рассмотрим последовательности длины 4 с элементами из $GF(2)$. Такие последовательности можно складывать как векторы, и нулевым элементом по операции сложения является 0000. Для задания операции умножения сопоставим каждой последовательности многочлен от α :

Последовательность	Многочлен
0000	0
1000	1
0100	α
1100	$\alpha + 1$
0010	α^2
1010	$\alpha^2 + 1$
0001	α^3
...	...
1111	$\alpha^3 + \alpha^2 + \alpha + 1$

Умножение таких многочленов может дать степень, большую чем 3, т. е. последовательность, не принадлежащую рассматриваемому множеству. Например, $(1101) \cdot (1001) = (\alpha^3 + \alpha + 1) \cdot (1 + \alpha^3) = \alpha^6 + \alpha^4 + \alpha + 1$. Для того чтобы свести ответ к многочлену степени не более 3, положим, что α удовлетворяет уравнению степени 4, например $f(\alpha) = \alpha^4 + \alpha + 1 = 0$, или $\alpha^4 = \alpha + 1$.

Тогда

$$\alpha^5 = \alpha^2 + \alpha, \quad \alpha^6 = \alpha^3 + \alpha^2;$$

$$\alpha^6 + \alpha^4 + \alpha + 1 = \alpha^3 + \alpha^2 + \alpha + 1 \quad \alpha + 1 = \alpha^3 + \alpha^2.$$

Это эквивалентно делению на многочлен $\alpha^4 + \alpha + 1$ и нахождению остатка от деления:

$$\begin{array}{r|l}
 \oplus \quad \alpha^6 + \alpha^4 + \alpha + 1 & \alpha^4 + \alpha + 1 \\
 \alpha^6 + \alpha^3 + \alpha^2 & \hline
 \hline
 \alpha^4 + \alpha^3 + \alpha^2 + \alpha + 1 & \\
 \oplus \quad \alpha^4 + & \\
 & \alpha + 1 \\
 \hline
 & \alpha^3 + \alpha^2 - \text{остаток}
 \end{array}$$

Таким образом, имеет место аналогия при формировании поля из чисел и последовательностей чисел (многочленов). Эта аналогия распространяется и на то, что для обратимости введенной операции умножения (чтобы система элементов в виде последовательностей длины m или многочленов степени, меньшей m , образовывала поле) многочлен $f(\alpha)$ должен быть неприводим над полем своих коэффициентов.

Поле, образованное многочленами над полем $GF(p)$ по модулю неприводимого многочлена $p(x)$ степени m , называется расширением поля степени m над $GF(p)$ или расширенным полем. Оно содержит p^m элементов и обозначается $GF(p^m)$.

Поле, образованное шестнадцатью двоичными последовательностями длины 4, или многочленами степени 3 и менее с коэффициентами из $GF(2)$ по модулю многочлена $x^4 + x + 1$, неприводимого над $GF(2)$, является примером расширенного поля $GF(2^4)$, которое может быть обозначено также $GF(16)$.

Важнейшим свойством конечных полей является следующее. Множество всех ненулевых элементов конечного поля образует группу по операции умножения, т. е. мультипликативную группу порядка $q - 1$.

Рассмотрим совокупность элементов мультипликативной группы, образованную некоторым элементом α и всеми его степенями α^2, α^3 и т. д. Так как группа конечна, должно появиться повторение, т. е. $\alpha^i = \alpha^j$. Умножая это равенство на $(\alpha^i)^{-1} = (\alpha^{-1})^i$, получим $\alpha^{j-i} = 1$. Следовательно, некоторая степень α равна 1.

Наименьшее положительное число e , такое, что $\alpha^e = 1$, называется порядком элемента α . Совокупность элементов $1, \alpha, \alpha^2, \dots, \alpha^{e-1}$ образует подгруппу, поскольку произведение любых двух элементов принадлежит этой совокупности, а элемент, обратный α^j , равен α^{e-j} и тоже входит в эту совокупность.

Группа, которая состоит из всех степеней одного из ее элементов, называется *циклической* группой.

Из рассмотренного свойства конечных полей вытекают два важных следствия.

Первое из них утверждает, что многочлен $x^{q-1} - 1$ имеет своими корнями все $q - 1$ ненулевых элементов поля $GF(q)$, т. е.

$$x^{q-1} - 1 = \prod_{\alpha \in GF(q), \alpha \neq 0} (x - \alpha).$$

В поле $GF(q)$ элемент b , имеющий порядок $e = q - 1$, называется примитивным. Отсюда следует, что любой ненулевой элемент $GF(q)$ является степенью примитивного элемента.

Второе следствие из рассмотренного свойства утверждает, что любое конечное поле $GF(q)$ содержит примитивный элемент, т. е. мультипликативная группа $GF(q)$ является циклической.

По результатам решения примеров 2 и 3 представлены изоморфные таблицы для полей $GF(2^4)$, построенные для различных неприводимых многочленов.

Из рассмотрения поля $GF(2^4)$, построенного по модулю $x^4 + x + 1$, мы видим, что примитивный элемент поля b является корнем этого многочлена. Многочлен, корнем которого является примитивный элемент, называется примитивным многочленом. Если в качестве $f(x)$ выбрать примитивный неприводимый многочлен степени m над полем $GF(2)$, то получим поле $GF(2^m)$ из всех 2^m двоичных последовательностей длины m .

Выше было показано, что $GF(4)$ нельзя представить в виде совокупности чисел 0, 1, 2, 3. Построим его как расширенное поле по модулю многочлена $f(x) = x^2 + x + 1$.



Пример 7.5. Построить поле $GF(4)$ по неприводимому многочлену $f(x) = x^2 + x + 1$.

Решение. Элементы поля $GF(4)$ можно представить различными способами. Здесь принято, что примитивный элемент b является корнем $f(b)$, т. е. $b^2 + b + 1 = 0$. Тогда получим для элементов поля $GF(4)$ следующий набор элементов.

Последовательность длины 2	Многочлен	Степень	Логарифм
00	0	0	$-\infty$
10	1	1	0
01	b	b	1
11	$b + 1$	b^2	2

Правила сложения и умножения в этом поле приведены на рис. 7.7.

+	0	1	α	α^2
0	0	1	α	α^2
1	1	0	α^2	α
α	α	α^2	0	1
α^2	α^2	α	1	0

a

$\times$	0	1	α	α^2
0	0	0	0	0
1	0	1	α	α^2
α	0	α	α^2	1
α^2	0	α^2	1	α

b

Рис. 7.7. Правила сложения (a) и умножения (b) для элементов поля $GF(4)$

Формирование первой строки, первого столбца и диагональных элементов таблицы сложения, а также двух первых строк и двух первых столбцов таблицы умножения не вызывает затруднения. Поясним формирование других элементов:

$$\alpha + 1 = \alpha^2, \alpha^2 + 1 = \alpha, \alpha^2 + \alpha = 1;$$

$$\alpha \cdot \alpha^2 = \alpha^3 = \alpha \cdot (1 + \alpha) = \alpha + \alpha^2 = 1,$$

на основе соотношения получаем для примитивного элемента $\alpha^2 + \alpha + 1 = 0$.

7.3. Способы задания линейных кодов

1. Линейные коды можно задать, перечислив все их кодовые слова, т. е. составив список всех кодовых слов кода.



Пример 7.6. Перечислить все кодовые слова $(5, 3)$ -кода (a_i — информационные, а b_i — проверочные символы).

В табл. 7.1 показаны все кодовые слова линейного $(5, 3)$ -кода.

Таблица 7.1

Кодовые слова $(5, 3)$ -кода

a_1	a_2	a_3	b_1	b_2
0	0	1	1	0
0	1	0	1	1
0	1	1	0	1
1	0	0	0	1
1	0	1	1	1
1	1	0	1	0
1	1	1	0	0

2. Составив систему проверочных уравнений, определяющих правила формирования проверочных символов по известным информационным:

$$b_j = \sum_{i=1}^k a_i \cdot h_{ij},$$

где j — номер проверочного символа; i — номер информационного символа; h_{ij} — коэффициенты, принимающие значения 0 или 1 в соответствии с правилами формирования конкретных групповых кодов.



Пример 7.7. Составить систему проверочных уравнений для (5, 3)-кода.

Решение. Проверочные уравнения для (5, 3)-кода имеют вид:

$$b_1 = a_2 + a_3;$$

$$b_2 = a_1 + a_2.$$

3. Матричный способ, основанный на построении *порождающей и проверочной* матриц. Векторное пространство V_n над $GF(2)$ включает в себя 2^n векторов (n -последовательностей), а подпространством его является множество из 2^k кодовых слов длины n , которое однозначно определяется его базисом, состоящим из k линейно независимых векторов. Поэтому линейный (n, k) -код полностью определяется набором из k кодовых слов, принадлежащих этому коду. Набор из k кодовых слов, соответствующих базису, обычно представляется в виде матрицы, которая называется порождающей.

7.4. Принцип построения циклических кодов

Идея построения циклических кодов базируется на использовании *неприводимых многочленов*. *Неприводимым* называется многочлен, который не может быть представлен в виде произведения многочленов низших степеней, т. е. такой многочлен делится только на самого себя или на единицу и не делится ни на какой другой многочлен. На такой многочлен делится без остатка двучлен $x^n + 1$. Неприводимые многочлены в теории циклических кодов играют роль *образующих* полиномов.

Чтобы понять принцип построения циклического кода, умножаем комбинацию простого k -значного кода $Q(x)$ на одночлен x^r , а затем делим на образующий полином $P(x)$, степень которого равна r . В результате умножения $Q(x)$ на x^r степень каждого одночлена, входящего в $Q(x)$, повышается на r . При делении произведения $x^r Q(x)$ на образующий полином получается частное $C(x)$ такой же степени, как и $Q(x)$.

Частное $C(x)$ имеет такую же степень, как и кодовая комбинация $Q(x)$ простого кода, поэтому $C(x)$ является кодовой комбинацией этого же простого k -значного кода. Следует заметить, что степень остатка не может быть больше степени образующего полинома, т. е. его наивысшая степень может быть равна $(r - 1)$. Следовательно, наибольшее число разрядов остатка $R(x)$ не превышает числа r .

Итак, представим сказанное в виде математических зависимостей:

$$Q(x) \cdot x^r / P(x) = C(x) + R(x) / P(x),$$

$$Q(x) \cdot x^r = C(x) P(x) + R(x),$$

$$Q(x) \cdot x^r + R(x) = C(x) P(x) = F(x).$$

Таким образом, кодовая комбинация циклического n -значного кода может быть получена двумя способами:

1) умножением кодовой комбинации $Q(x)$ простого кода на одночлен x^r и добавлением к этому произведению остатка $R(x)$, полученного в результате деления произведения $Q(x) x^r$ на образующий полином $P(x)$;

2) умножения кодовой комбинации $C(x)$ простого k -значного на образующий полином $P(x)$.

При построении циклических кодов первым способом расположение информационных символов во всех комбинациях строго упорядочено — они занимают k старших разрядов комбинации, а остальные $(n - k)$ разрядов отводятся под контрольные.

При втором способе образования циклических кодов информационные и контрольные символы в комбинациях циклического кода не отделены друг от друга, что затрудняет процесс декодирования.

7.5. Получение кодовой комбинации добавлением остатка $R(x)$



Пример 7.8. Построить циклический код для передачи 31 разрядной кодовой комбинации с исправлением однократной ошибки ($n = 31, s = 1$).

Приведем поэтапное решение поставленной задачи.

1. Определим число контрольных разрядов m :

$$m = \log_2 (n + 1) = \log_2 (31 + 1) = 5.$$

2. Определим количество информационных разрядов k :

$$k = n - m = 26, \text{ т. е. получили } (31, 26)\text{-код.}$$

3. Строим информационный полином, соответствующий информационному слову длиной k бит:

$$G(x) = 000000000000000000000000101 = x^2 + 1.$$

4. Осуществляем сдвиг кода влево на $m = n - k = 5$ разрядов, т. е. полином $G(x)$ умножается на x^m :

$$\begin{aligned} x^m G(x) &= (x^2 + 1) x^5 = x^7 + x^5 = \\ &= 00000000000000000000000010100000. \end{aligned}$$

5. Выбираем образующий многочлен $P(x)$ по таблице неприводимых многочленов. Для исправления одиночной ошибки ($d_0 = 3$) образующий полином $P(x)$ должен быть степени $m = n - k = 5$ и количеством ненулевых членов не меньше минимального кодового расстояния $d_0 = 3$. Исходя из этого образующий полином $P(x)$ равен: $P(x) = x^5 + x^4 + x^3 + x^2 + 1 = 111101$.

6. Определим остаток $R(x)$ от деления $G(x) \cdot x^m$ на образующий полином $P(x)$

$\begin{array}{r} x^7 + x^5 \\ x^7 + x^6 + x^5 + x^4 + x^2 \\ \hline x^6 + x^4 + x^2 \\ x^6 + x^5 + x^4 + x^3 + x \\ \hline x^5 + x^3 + x^2 + x \\ x^5 + x^4 + x^3 + x^2 + 1 \\ \hline x^4 + x + 1 \end{array}$	$\begin{array}{r} 10100000 \\ \underline{111101} \\ 101010 \\ \underline{111101} \\ 101110 \\ \underline{111101} \\ 10011 \end{array}$	$\begin{array}{r} 111101 \\ \underline{111} \end{array}$

$$\text{Остаток } R(x) = x^4 + x + 1 = 10011.$$

7. Строим передаваемый кодовый полином $F(x)$:

$$\begin{aligned} F(x) &= x^m G(x) \oplus R(x) = x^7 + x^5 + x^4 + x + 1 = \\ &= 00000000000000000000000010110011. \end{aligned}$$

8. Пусть в принятом сообщении произошла ошибка в тридцать первом разряде, при этом принятое кодовое сообщение имеет вид:

$$F'(x) = F(x) \oplus E(x) = 100000000000000000000000010110011.$$

9. Разделим многочлен $F(x)$, соответствующий полученной кодовой комбинации, на образующий полином, при этом вес остатка (количество единиц в коде остатка) должен быть меньше или равен количеству ошибок $w \leq s$:

```

100000000000000000000000010110011 | 111101
111101
  111010
   111101
     111000
      111101
        101000
         111101
           101010
            111101
              101110
               111101
                 100110
                  111101
                    110110
                     111101
                       101100
                        111101
                          100010
                           111101
                             111110
                              111101
                                110010
                                 111101
                                  111111
                                   111101
                                     100011
                                      111101
                                        11110

```

Сравниваем вес полученного остатка w с числом исправляемых ошибок $w > s$.

10. Производим циклический сдвиг принятой кодовой комбинации на один разряд влево и повторяем п. 9 пока $w \leq s$.

[illegible]

Складываем по модулю 2 последнее делимое с последним остатком:

$$\begin{array}{r} \text{6) } \quad 000000000000000000000000101100111 \\ \oplus \quad \underline{\hspace{28em}} \hspace{0.5em} 1 \\ \quad 000000000000000000000000101100110 \end{array}$$

Осуществляем обратный сдвиг на 1 разряд полученной комбинации

[illegible]

Отбросив контрольные разряды, получаем переданное информационное слово.

7.6. Построение кодовой комбинации путем умножения на образующий полином



Пример 7.9. Построить циклический код для передачи 31-разрядной кодовой комбинации с исправлением однократной ошибки ($n = 31, s = 1$) путем умножения образующего многочлена на многочлен полного 31-разрядного кода.

Решение может быть представлено следующим образом.

1. Строим информационный полином, соответствующий информационному слову длиной k бит:

$$G(x) = 0000000000000000000000000000101 = x^2 + 2.$$

2. Строим передаваемый кодовый полином

$$\begin{array}{r}
 000000000000000000000000101 \\
 111101 \\
 \hline
 000000000000000000000000101 \\
 000000000000000000000000101 \\
 000000000000000000000000101 \\
 000000000000000000000000101 \\
 000000000000000000000000101 \\
 \hline
 00000000000000000000000011001001
 \end{array}$$

3. Процесс исправления однократной ошибки аналогичен описанному в 7.4.

7.7. Операции над циклическими кодами

1. Сдвиг справа налево осуществляется путем умножения полинома $G(x)$ на x :

$$G(x) = x^4 + x^2 + 1 \text{ соответствует } 0010101;$$

$$G(x) \cdot x = x^5 + x^3 + x \text{ соответствует } 0101010.$$

2. Операции сложения и вычитания выполняются по модулю 2. Они являются эквивалентными и ассоциативными:

$$G_1(x) + G_2(x) \Rightarrow G_3(x);$$

$$G_1(x) - G_2(x) \Rightarrow G_3(x);$$

$$G_2(x) + G_1(x) \Rightarrow G_3(x).$$

Например:

$$G_1(x) = x^5 + x^3 + x;$$

$$G_2(x) = x^4 + x^3 + 1;$$

$$G_3(x) = G_1(x) \oplus G_2(x) = x^5 + x^4 + x + 1.$$

3. Операция деления является обычным делением многочленов, только вместо вычитания используется сложение по модулю 2:

$$G_1(x) = x^6 + x^4 + x^3;$$

$$G_2(x) = x^3 + x^2 + 1.$$

7.8. Способы кодирования циклическим кодом

Задача кодирования заключается в формировании по информационным словам $Q(x)$ кодовых слов $F(x)$ циклического

(n, k) -кода, который по своей структуре может быть несистематическим и систематическим.

Формирование кодовых слов несистематического кода заключается в умножении многочлена $Q(x)$, отображающего информационную последовательность длины k , на порождающий многочлен, т. е. $F(x) = Q(x)P(x)$. Формирование кодовых слов систематического кода заключается в преобразовании информационной последовательности $Q(x)$ в соответствии с выражением $F(x) = Q(x)x^r + R(x)$. Проверочная последовательность $R(x)$ определяется двумя способами: при использовании «классического» способа кодирования $r(x) = R_{g(x)}[a(x) \cdot x^r]$; при использовании способа кодирования, рекомендованного МККТТ, $r(x) = \overline{R}_{g(x)}[a(x) \cdot x^r + x(1)^{r-1} \cdot x^k]$, где $x(1)^{r-1}$ — единичный многочлен степени $(r-1)$, а оператор $\lfloor \rfloor$ означает округление к минус бесконечности.

Указанные выше математические операции выполняют кодеры несистематического и систематического кодов.

7.9. Матричное задание циклических кодов

Циклический код может быть задан порождающей и проверочной матрицами. Для их построения достаточно знать порождающий $g(x)$ и проверочный $h(x)$ многочлены. Для несистематического циклического кода матрицы строятся циклическим сдвигом порождающего и проверочного многочленов, т. е. путем их умножения на x :

$$G_{(n,k)} = \begin{vmatrix} g(x) \\ x \cdot g(x) \\ x^2 \cdot g(x) \\ \dots \\ x^{k-1} \cdot g(x) \end{vmatrix} \quad \text{и} \quad H_{(n,k)} = \begin{vmatrix} h(x) \\ x \cdot h(x) \\ x^2 \cdot h(x) \\ \dots \\ x^{r-1} \cdot h(x) \end{vmatrix}.$$

При построении матрицы $H_{(n,k)}$ старший коэффициент многочлена $h(x)$ располагается справа.



Пример 7.10. Для циклического $(7,4)$ -кода с порождающим многочленом $g(x) = x^3 + x + 1$ матрицы $G_{(n,k)}$ и $H_{(n,k)}$ имеют вид:

$$G_{(7,4)} = \begin{vmatrix} x^3 + x + 1 \\ x(x^3 + x + 1) \\ x^2(x^3 + x + 1) \\ x^3(x^3 + x + 1) \end{vmatrix} = \begin{vmatrix} x^3 + x + 1 \\ x^4 + x^2 + x \\ x^5 + x^3 + x^2 \\ x^6 + x^4 + x^3 \end{vmatrix} = \begin{vmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \end{vmatrix} =$$

$$H_{(7,4)} = \begin{vmatrix} 1 + x + x^2 + x^4 \\ x(1 + x + x^2 + x^4) \\ x^2(1 + x + x^2 + x^4) \end{vmatrix} = \begin{vmatrix} 1 + x + x^2 + x^4 \\ x + x^2 + x^3 + x^5 \\ x^2 + x^3 + x^4 + x^6 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{vmatrix}$$

где $h(x) = x^7 + 1 / g(x) = x^7 + 1 / x^3 + x + 1 = x^4 + x^2 + x + 1$.

Для систематического циклического кода матрица $G_{(n,k)}$ определяется из выражения

$$G_{(n,k)} = [I_k, R_{k,r}],$$

где I_k — единичная матрица; $R_{k,r}$ — прямоугольная матрица. Строки матрицы $R_{k,r}$ определяются из выражений

$r_i(x) = R_{g(x)} [a_i(x) \cdot x^r]$ или $r_i(x) = R_{g(x)} [x^{n-i}]$, где $a_i(x)$ — значение i -той строки матрицы I_k ; i — номер строки матрицы $R_{k,r}$,

$\lfloor \rfloor$ — означает округление к минус бесконечности.

7.10. Коды Хэмминга

Код Хэмминга представляет собой блочный код, который позволяет выявить и исправить ошибочно переданный бит в пределах переданного блока. Обычно код Хэмминга характеризуется двумя целыми числами, например $(11,7)$, используемый при передаче 7-битных ASCII-кодов. Такая запись говорит, что при передаче 7-битного кода используется 4 контрольных битов ($7 + 4 = 11$). При этом предполагается, что имела место ошибка в одном бите и что ошибка в двух или более битах существенно менее вероятна. С учетом этого исправление ошибки осуществляется с определенной вероятностью.



Пример 7.11. Рассмотреть режимы передачи кода буквы $s = 0 \times 073 = 1110011$ с использованием кода Хэмминга (11,7) при различных ошибках в передаваемых битах.

Решение. Представим передаваемый символ в виде таблицы

Позиция бита	11	10	9	8	7	6	5	4	3	2	1
Значение бита	1	1	1	*	0	0	1	*	1	*	*

Символами * помечены четыре позиции, где должны размещаться контрольные биты. Эти позиции определяются целой степенью 2 (1, 2, 4, 8). Контрольная сумма формируется путем выполнения операции XOR (исключающее ИЛИ) над кодами позиций ненулевых битов. В данном случае это позиции 11, 10, 9, 5 и 3. Вычислим контрольную сумму:

11 =	1011
10 =	1010
09 =	1001
05 =	0101
03 =	0011
S =	1110

Таким образом, приемник получит код:

Позиция бита	11	10	9	8	7	6	5	4	3	2	1
Значение бита	1	1	1	1	0	0	1	1	1	1	0

Просуммируем снова коды позиций ненулевых битов и получим нуль.

11 =	1011
10 =	1010
09 =	1001
08 =	1000
05 =	0101
04 =	0100
03 =	0011
02 =	0010
S =	0000

Теперь рассмотрим два случая ошибок в одном из битов посылки, например в бите 7 (1 вместо 0) и в бите 5 (0 вместо 1). Просуммируем коды позиций ненулевых бит еще 2 раза соответственно допущенным ошибкам в 7 и 5 битах.

11 =	1011
10 =	1010
09 =	1001
08 =	1000
07 =	0111
05 =	0101
04 =	0100
03 =	0011
02 =	0010
S =	0111

11 =	1011
10 =	1010
09 =	1001
08 =	1000
04 =	0100
03 =	0011
02 =	0010
S =	0101

В обоих случаях контрольная сумма равна позиции бита, переданного с ошибкой. Теперь для исправления ошибки достаточно инвертировать бит, номер которого указан в контрольной сумме. Понятно, что если ошибка произойдет при передаче более чем одного бита, код Хэмминга при данной избыточности окажется бесполезен.

Число возможных кодовых комбинаций M помехоустойчивого кода делится на n классов, где N — число разрешенных кодов. Разделение на классы осуществляется так, чтобы в каждый класс вошел один разрешенный код и ближайшие к нему (по расстоянию Хэмминга) запрещенные коды. В процессе приема данных определяется, к какому классу принадлежит пришедший код. Если код принят с ошибкой, он заменяется ближайшим разрешенным кодом. При этом предполагается, что кратность ошибки не более q_m .

7.11. Коды БЧХ

Обобщением кодов Хэмминга являются циклические коды *BCH* (*Bose-Chadhuri-Nocquenghem*). Это коды с широким выбором длины и возможностей исправления ошибок. Циклические коды характеризуются полиномом $g(x)$ степени $n-k$, $g(x) = 1 + g_1x + g_2x^2 + \dots + x^{n-k}$. $g(x)$ называется порождающим многочленом циклического кода. Если многочлен $g(x)$ $n-k$ и является делителем многочлена $x^n + 1$, то код $C(g(x))$ является линейным циклическим (n,k) -кодом. Число цикли-

ческих n -разрядных кодов равно числу делителей многочлена $x^n + 1$. При кодировании слова все кодовые слова кратны $g(x)$. Многочлен $g(x)$ определяется на основе сомножителей полинома $x^n + 1$ как:

$$x^n + 1 = g(x)h(x).$$

Например, если $n = 7$ ($x^7 + 1$), его сомножители $(1 + x + x^3) \times (1 + x + x^2 + x^4)$, а $g(x) = 1 + x + x^3$.

Чтобы представить сообщение $h(x)$ в виде циклического кода, в котором можно указать постоянные места проверочных и информационных символов, нужно разделить многочлен $x^{n-k}h(x)$ на $g(x)$ и прибавить остаток от деления к многочлену $x^{n-k}h(x)$. Привлекательность циклических кодов заключается в простоте аппаратной реализации с использованием сдвиговых регистров.

Пусть общее число бит в блоке равно N , из них полезную информацию несут в себе K бит, тогда в случае ошибки, имеется возможность исправить t бит. Таблица содержит зависимость t от N и K для кодов BCH.

Общее число битов, N	Число полезных битов, M	Число исправляемых битов, t
31	26	1
31	21	2
31	16	3
63	57	1
63	51	2
53	45	3
127	120	1
127	113	2
127	106	3

Увеличивая разность $N-M$, можно не только нарастить число исправляемых битов t , но открыть возможность обнаружить множественные ошибки. В следующей таблице приведен процент обнаруживаемых множественных ошибок в зависимости от M и $N - M$.

Число полезных битов, M	Число избыточных битов ($n - m$)		
	6	7	8
32	48%	74%	89%
40	36%	68%	84%
48	23%	62%	81%

Другой блочный метод предполагает «продольное и поперечное» контрольное суммирование передаваемого блока. Блок при этом представляется в виде N строк и M столбцов. Вычисляются биты четности для всех строк и всех столбцов, в результате получается два кода, соответственно длиной N и M бит. На принимающей стороне биты четности для строк и столбцов вычисляются повторно и сравниваются с присланными. При выявлении отличия в бите i -кода битов четности строк и бите j -кода столбцов, позиция неверного бита оказывается определенной (i, j) . Понятно, что если выявится два и более неверных битов в контрольных кодах строк и столбцов, задача коррекции становится неразрешимой. Уязвим этот метод и для двойных ошибок, когда сбой был, а контрольные коды остались корректными.

Применение кодов свертки позволяют уменьшить вероятность ошибок при обмене, даже если число ошибок при передаче блока данных больше 1.

Число проверочных элементов кода БЧХ удовлетворяет соотношению $r = n - k = mt_u$.

7.12. Коды Рида-Соломона (РС)

Коды РС являются недвоичными циклическими кодами, символы кодовых слов которых берутся из конечного поля $GF(q)$. Здесь q степень некоторого простого числа, например $q = 2^m$.

Допустим, что РС-код построен над $GF(8)$, которое является расширением поля $GF(2)$ по модулю примитивного многочлена $f(z) = z^3 + z + 1$. В этом случае символы кодовых слов кода будут иметь значения, представленные в таблице:

000	0	0	011	$z + 1$	6^3
001	1	6^0	110	$z^2 + z$	6^4
010	z	6^1	111	$z^2 + z + 1$	6^5
100	z^2	6^2	101	$z^2 + 1$	6^6

Кодовые слова РС-кода отображаются в виде многочленов

$$V(x) = V_{N-1}x^{N-1} + V_{N-2}x^{N-2} + \dots + V_1x^1 + V_0,$$

где N — длина кода; V_i — q -ичные коэффициенты (символы кодовых слов), которые могут принимать любое значение из

$GF(q)$. Эти коэффициенты, как это следует из таблицы, также отображаются многочленами с двоичными коэффициентами $V_i = a_{m-1}z^{m-1} + a_{m-2}z^{m-2} + \dots + a_1z^1 + a_0$. Коды PC являются максимальными, так как при длине кода N и информационной последовательности k они обладают наибольшим кодовым расстоянием $d = N - k + 1$.

Порождающим многочленом $g(x)$ PC -кода является делитель двучлена $x^N + 1$, степени меньшей N , с коэффициентами из $GF(q)$ при условии, что элементы $\alpha^1, \alpha^2, \dots, \alpha^{d-1}$ этого поля являются корнями $g(x)$. Здесь α — примитивный элемент $GF(q)$. На основе этого определения, а также теоремы Безу, выражение для порождающего многочлена PC -кода будет иметь вид:

$$g(x) = \prod_{i=1}^{d-1} (x - \alpha^i).$$

Степень $g(x)$ равна $d - 1 = N - k = R$.

В PC -кодах принадлежность кодовых слов данному коду определяется выполнением $d - 1$ уравнений в соответствии с выражением

$$\sum_{i=0}^{N-1} V_i \cdot z_i^m = 0, \quad (m = m_0, m_0 + 1, \dots, m_0 + d - 2), \quad (7.1)$$

где V_i — символы-коэффициенты из $GF(q)$; $z_0, z_1, \dots, z_{N-1}$ — ненулевые элементы $GF(q)$. Элементы $z_0, z_1, \dots, z_{N-1}$ называются локаторами, т. е. элементами, указывающими на номер позиции символа кодового слова. Например, указателем i -позиции является локатор z^i или элемент $GF(q)$. Так как все локаторы должны быть различны и причем ненулевыми, то их число в $GF(q)$ равно $q - 1$. Следовательно, такое количество символов должно быть в кодовых словах кода. Поэтому обычно длина PC -кода определяется из выражения $N = q - 1$.

Допустим, что длина PC -кода равна N , кодовое расстояние $d = 3$, то в соответствии с (7.1) проверочными уравнениями будут

$$V_{N-1}(\alpha^{N-1})^1 + V_{N-2}(\alpha^{N-2})^1 + \dots + V_1(\alpha^1)^1 + V_0(\alpha^0)^1 = 0;$$

$$V_{N-1}(\alpha^{N-1})^2 + V_{N-2}(\alpha^{N-2})^2 + \dots + V_1(\alpha^1)^2 + V_0(\alpha^0)^2 = 0.$$

Свойства РС-кодов

1. Циклический сдвиг кодовых слов, символы которых принимают значение из $GF(q)$, порождает новые кодовые слова этого же кода.

2. Сумма по $\text{mod } 2$ двух и более кодовых слов дает кодовое слово, принадлежащее этому же коду.

3. Кодовое расстояние РС-кода определяется не по двоичным элементам, а по q -ичным символам.

4. В РС-коде, исправляющем t_u ошибок, порождающий многочлен определяется из выражения

$$g(x) = (x - \alpha^{m_0}) \cdot (x - \alpha^{m_0+1}) \dots (x - \alpha^{m_0+2t_u-1}).$$

Обычно m_0 принимают равным 1. Однако, с помощью разумного выбора значения m_0 иногда можно упростить схему кодера.

5. Корректирующие способности РС-кода определяются его кодовым расстоянием:

$$t_0 = d - 1, T_0 = (t_0 - 1)m + 1, t_u = d - 0,5,$$

$$T_u = (t_u - 1)m + 1,$$

где T_0 и T_u — длины пакетов, в которых обнаруживаются и исправляются ошибки. Обнаружение ошибок в кодовых словах состоит в проверке этих условий, т. е. определении синдрома $S = S_{d-2} S_1 S_0$, элементы которого определяются из выражения

$$S_j = \sum_{i=0}^{N-1} V_i z_i^{m_0+j} (j = 0, \dots, d-2).$$



Пример 7.12. Требуется сформировать кодовое слово РС-кода над $GF(2^3)$, соответствующее двоичной информационной последовательности:

$$a(1,0)=000000011100101.$$

Решение. Так как $m = 3$, то каждый q -ичный символ кода состоит из трех двоичных элементов. Поэтому с учетом таблицы 6 $a(x) = \alpha^3 x^2 + \alpha^2 x + \alpha^6$.

Определяем параметры кода. $N = q - 1 = 7$; $k = 5$; $R = 2$; $d = N - k + 1 = 3$;

$$g(x) = \prod_{i=1}^{d-1} (x - \alpha^i) = (x - \alpha^1)(x - \alpha^2) = \\ = x^2 + (\alpha^1 + \alpha^2)x + \alpha^3 = x^2 + \alpha^4x + \alpha^3.$$

Кодовое слово формируется в соответствии с выражением

$$V(x) = \alpha(x) \cdot x^R + R(x), \quad R(x) = R_{g(x)}[\alpha(x) \cdot x^R], \text{ где}$$

$$\alpha(x) \cdot x^R = (\alpha^3 x^2 + \alpha^2 x + \alpha^6) \cdot x^2 = \alpha^3 x^4 + \alpha^2 x^3 + \alpha^6 x^2$$

$$R(x) = R_{g(x)}[\alpha^3 x^4 + \alpha^2 x^3 + \alpha^6 x^2] = \alpha^6 x + \alpha^6.$$

В результате $V(x) = \alpha^3 x^4 + \alpha^2 x^3 + \alpha^6 x^2 + \alpha^6 x + \alpha^6$, или в двоичной форме $V(1,0) = 000.000.011.100.101.101.101$.

7.13. Методы описания сверточных кодов

Кодер СК содержит регистр памяти для хранения определенного числа информационных символов и преобразователь информационной последовательности в кодовую последовательность. Процесс кодирования производится непрерывно. Скорость кода $R = k/n$, где k — число информационных символов, одновременно поступающих на вход кодера, n — число соответствующих им символов на выходе кодера. Схема простого кодера показана на рис. 7.8, а.

Информационные двоичные символы u поступают на вход регистра с K разрядами. На выходах сумматоров по модулю 2 образуются кодовые символы $a(1)$ и $a(2)$. Входы сумматоров соединены с определенными разрядами регистра. За время

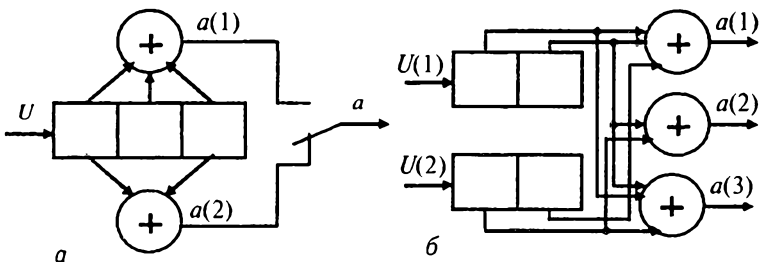


Рис. 7.8. Схема простого декодера (а) и формирование за время одного информационного символа двух кодовых символов (б)

одного информационного символа на выходе образуются два кодовых символа ($R = 1/2$). Возможно кодирование и с другими скоростями. При скорости $2/3$ на вход кодера одновременно поступает $k = 2$ информационных символа, на выходе при этом образуется $n = 3$ кодовых символа. Схема такого кодера показана на рис. 7.8, б.

Рассматриваемый код называется сверточным, постольку последовательность a кодовых символов может быть определена как свертка информационных символов u с импульсным откликом кодера. На рис. 7.9 показано прохождение последовательности $u = 100$ через кодер.

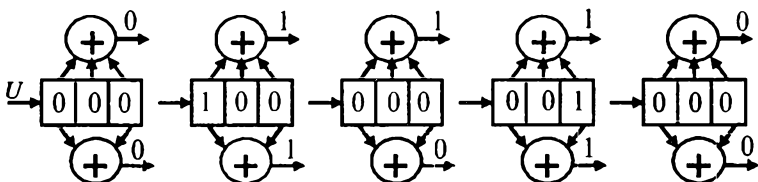


Рис. 7.9. Прохождение единичной последовательности через кодер

Символы $a(1)$ и $a(2)$ на его выходе образуют импульсный отклик $h = 00111011\ 00\dots$ Таким образом, если на входе кодера действует произвольная информационная последовательность u , то последовательность на его выходе есть сумма по модулю 2 всех импульсных откликов, обусловленных действием смещенных во времени символов 1. Сверточный кодер, как автомат с конечным числом состояний, может быть описан диаграммой состояний. Диаграмма представляет собой направленный граф и описывает все возможные переходы кодера из одного состояния в другое, а также содержит символы выходов кодера, которые сопровождают эти переходы.

Первоначально кодер находится в состоянии 00, и поступление на его вход информационного символа $u = 0$ перевозят его также в состояние 00. При этом на выходе кодера будут символы $a(1)a(2) = 00$. На диаграмме этот переход обозначается петлей 00, выходящей из состояния 00 и вновь возвращающейся в это состояние. Далее, при поступлении символа

$u = 1$ кодер переходит в состояние 10, при этом на выходе будут символы $a(1)a(2) = 11$. Этот переход из состояния 00 в состояние 10 обозначается пунктирной линией. Далее возможно поступление на вход кодера информационных символов 0 либо 1. При этом кодер переходит в состояние 01 либо 11, а символы на выходе будут 10 либо 01 соответственно. Процесс построения диаграммы заканчивается, когда будут просмотрены все возможные переходы из одного состояния во все остальные.

Решетчатая диаграмма является разверткой диаграммы состояний во времени. На решетке состояния показаны узлами, а переходы соединяющими их линиями. После каждого перехода из одного состояния в другое происходит смещение на один шаг вправо. Решетчатая диаграмма дает наглядное представление всех разрешенных путей, по которым может продвигаться кодер при кодировании. Каждой информационной последовательности на входе кодера соответствует единственный путь по решетке. Построение решетки производится на основе диаграммы состояний. Исходное состояние $S(1)S(2) = 0$. С поступлением очередного символа $u = 0$ либо 1 возможны переходы в состояния 00 либо 10, обозначаемые ветвями 00 и 11. Процесс следует продолжить, причем через три шага очередной фрагмент решетки будет повторяться. Пунктиром показан путь 11100001..., соответствующий поступлению на вход кодера информационной последовательности 1011....

Для описания кодера последовательности символов на его входе и выходе представляют с использованием оператора задержки:

$$U_{(j)}(D) = U_{(j)0}D^0 + U_{(j)1}D^1 + U_{(j)2}D^2 + ,$$

$$U_{(i)}(D) = U_{(i)0}D^0 + U_{(i)1}D^1 + U_{(i)2}D^2 + ,$$

Здесь индексы в скобках обозначают: i – номер входа кодера, $1 = j = n$, j – номер выхода кодера, $1 = i = k$. Индексы без скобок (0, 1, 2, ...) обозначают дискретные моменты времени.

Процесс кодирования может быть представлен как умножение многочлена входной информационной последовательности $u(D)$ на порождающие многочлены кода $G^{(j)}(D)$, которые описывают связи ячеек регистра кодера с его выходами:

$$a^{(j)}(D) = u(D)G^{(j)}(D), \quad 1 = j = n.$$

Порождающий многочлен представим в виде ряда

$$G^{(j)}(D) = g_0^{(j)} + g_1^{(j)}D + g_2^{(j)}D^2 + \dots + g_v^{(j)}D^v,$$

СК можно также задавать порождающей матрицей

$$G = \begin{vmatrix} G_0 & G_1 & G_2 & \dots & G_{k-1} & \dots \\ \dots & G_0 & G_1 & G_2 & \dots & G_{k-1} & \dots \\ \dots & \dots & G_0 & G_1 & G_2 & \dots & G_{k-1} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{vmatrix}$$

Порождающая матрица состоит из сдвигов базисной порождающей матрицы (верхняя строка матрицы G), которая, в свою очередь, состоит из элементарных матриц G_i , $0 = i = k - 1$, содержащих k строк и n столбцов. Элементами этих матриц двоичных кодов являются символы 0 и 1.

Как и при использовании блочных кодов, процесс кодирования может быть представлен в матричной форме:

$$A = U \cdot G,$$

где U — полубесконечная матрица входных информационных символов, A — полубесконечная матрица символов на выходе кодера.

7.14. Декодирование сверточных кодов

Алгебраические методы декодирования основаны на использовании алгебраических свойств кодовых последовательностей. В ряде случаев эти методы приводят к простым реализациям кодека. Такие алгоритмы являются неоптимальными, так как используемые алгебраические процедуры декодирования предназначены для исправления конкретных (и не всех) конфигураций ошибок в канале. Алгебраические методы отождествляют с поэлементным приемом последовательностей, который для кодов с избыточностью, как известно, дает худшие результаты, чем прием в целом.

Вероятностные методы декодирования значительно ближе к оптимальному приему в целом, так как в этом случае декодер оперирует с величинами, пропорциональными вероятностям, оценивает и сравнивает вероятности различных гипотез и на этой основе выносит решения о передаваемых символах.

В 1967 году Витерби разработал и проанализировал алгоритм, в котором, по сути, реализуется декодирование, основанное на принципе максимального правдоподобия; однако в нем уменьшается вычислительная нагрузка за счет использования особенностей структуры конкретной решетки кода. Преимущество декодирования Витерби заключается в том, что сложность декодера Витерби не является функцией количества символов в последовательности кодовых слов. Алгоритм включает в себя вычисление *меры подобия* (или *расстояния*) между сигналом, полученным в момент времени t_1 , и всеми путями решетки, входящими в каждое состояние в момент времени t_1 . В алгоритме Витерби не рассматриваются те пути решетки, которые, согласно принципу максимального правдоподобия, заведомо не могут быть оптимальными. Если в одно и то же состояние входят два пути, выбирается тот, который имеет лучшую метрику; такой путь называется *выживающим*. Отбор выживающих путей выполняется для каждого состояния. Таким образом, декодер углубляется в решетку, принимая решения путем исключения менее вероятных путей. Предварительный отказ от маловероятных путей упрощает процесс декодирования. Отметим, что задачу отбора оптимальных путей можно выразить как выбор кодового слова с *максимальной метрикой правдоподобия* или *минимальной метрикой расстояния*.

7.15. Пороговое декодирование

Вероятностные методы декодирования достаточно сложны в реализации, хотя и обеспечивают высокую помехоустойчивость. Наряду с ними широко применяют более простые алгоритмы. Для этой цели используют класс СК, допускающих пороговое декодирование.

Рассмотрим систематический код со скоростью $1/2$ и многочленами:

$$G^{(1)}(D) = 1,$$

$$G^{(2)}(D) = 1 + D^{(2)} + D^{(5)} + D^{(6)}$$

Схема кодека приведена на рис. 7.9. Моделью двоичного канала являются сумматоры по модулю 2, на входы которых, кроме кодовых последовательностей $a(1)$ и $a(2)$, поступают

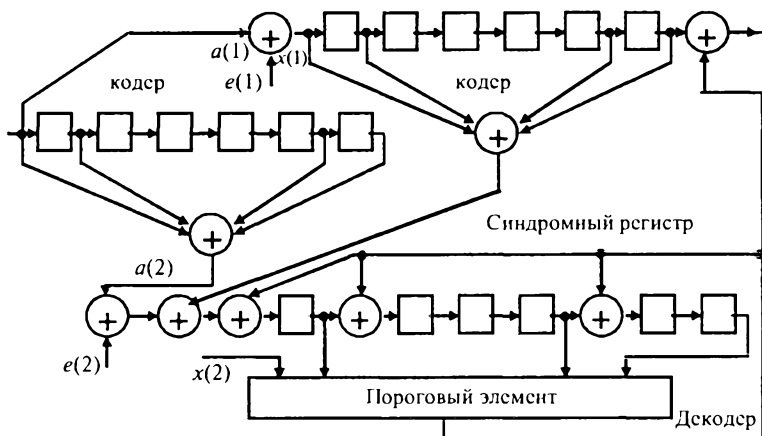


Рис. 7.9. Схема сверточного кодера порогового декодирования

ошибки $e(1)$ и $e(2)$. Декодер содержит аналог кодера, в котором принятым символам формируется копия проверочной последовательности. В формирователе синдрома (сумматоре по модулю 2) образуется последовательность синдромов, которая поступает на вход синдромного регистра. Наборам ошибок соответствуют определенные конфигурации синдромов последовательности S . Если количество ненулевых синдромов превышает определенный порог, на выходе порогового элемента появляется символ коррекции, который в корректоре используется для исправления ошибки в информационном символе.

7.16. Турбокоды

Последнее десятилетие характеризуется бурным ростом передачи цифровых данных и возрастанием требований к их достоверности. Этим обусловлено появление и широкое практическое внедрение новых эффективных методов модуляции, помехоустойчивого кодирования и сжатия данных.

В 1993 году группой французских ученых во главе с Берроу (*C. Berrou, A. Glaviux, P. Thitimajshima*) был предложен новый класс помехоустойчивых кодов, который получил название

турбокоды. Турбокоды сразу привлекли к себе широкое внимание, так как обеспечивают корректирующую способность, близкую к границе Шеннона, при приемлемой сложности декодирования. По структуре построения кодера турбокод можно отнести к классу компонентных кодов (*constituent code*) — он образуется при параллельном каскадировании двух и более компонентных кодов.

Компонентные коды в общем случае могут быть разными и иметь различные относительные информационные скорости. Информационная последовательность подается в первый кодер непосредственно, а в остальные кодеры через устройства псевдослучайного перемежения. Использование перемежителей позволяет предотвратить появление последовательностей коррелированных ошибок при декодировании. С целью повышения относительной информационной скорости турбокода используется *выкалывание*, или удаление, некоторых символов слова (*puncturing*).

В качестве компонентных кодов могут использоваться рекурсивные сверточные коды (сверточные коды с обратной связью), блочные коды, сверточные и блочные коды одновременно.

Такое усложнение структуры кодера существенно повышает скрытность канала связи. Для несанкционированного доступа к информации, передаваемой по системам связи с использованием турбокодирования, необходимо знать число и структуру компонентных кодов, параметры перемежения, наличие и параметры перфорации.

Сложность указанных задач обусловлена многообразием вариантов построения кодеров турбокода и входящих в них элементов, возможностью динамического изменения параметров кодера в зависимости от соотношения сигнал/шум в канале связи. Следовательно, при отсутствии информации о структуре турбокода задача доступа к закодированной информации является весьма трудоемкой. Следующая задача при попытке несанкционированного доступа — разработка и реализация декодера.

В общем виде, турбокоду с N компонентными кодами соответствует декодер, который содержит N компонентных декодеров.

Особенностью построения декодеров турбокода является использование «мягких» решений (величин, пропорциональных достоверности бита) и обмен между компонентными декодерами информацией (использование итеративного подхода). Из-за этого декодеры турбокода требуют больших аппаратных и вычислительных затрат, что делает задачу их реализации достаточно сложной.

Высокие исправляющие способности турбокодов позволяют работать при более низких отношениях сигнал/шум, чем используемые блочные или сверточные коды. Поэтому в настоящее время турбокоды нашли широкое применение в различных цифровых системах связи. В частности, они приняты в качестве стандарта для систем связи третьего поколения IMT-2000, стандарта сотовой связи CDMA-2000, цифрового телевидения DVB и в ряде других систем.

Сложность структуры построения кодера турбокода, наличие псевдослучайных перемежителей, сложность построения декодеров и работа систем связи при малых отношениях сигнал/шум повышают структурную скрытность каналов связи, в которых используется турбокодирование.

7.17. Граничные соотношения между параметрами помехоустойчивых кодов

Одной из важнейших задач построения помехоустойчивых кодов с заданными характеристиками является установление соотношения между его способностью обнаруживать или исправлять ошибки и избыточностью. Существуют граничные оценки, связывающие d_0 , n и k .

Граница Хэмминга, которая близка к оптимальной для высокоскоростных кодов, определяется соотношениями:

для q -ного кода

$$n - k \geq \log_2 \sum_{i=0}^{t_n} C_n^i \cdot (q - 1)^i;$$

для двоичного кода

$$n - k \geq \log_2 \sum_{i=0}^{t_n} C_n^i .$$

Граница Плоткина, которую целесообразно использовать для низкоскоростных кодов, определяется соотношениями:

для q -ного кода

$$d_0 \leq n \cdot (q - 1) \cdot q^{k-1} / q^k - 1;$$

для двоичного кода

$$d_0 \leq n \cdot 2^{k-1} / 2^k - 1.$$

Границы Хэмминга и Плоткина являются верхними границами для кодового расстояния при заданных n и k , задающими минимальную избыточность, при которой существует помехоустойчивый код, имеющий минимальное кодовое расстояние и гарантированно исправляющий t_u — кратные ошибки.

Граница Варшавова-Гильберта (нижняя граница) определяется соотношениями:

для q -ного кода

$$q^{n-k} > \sum_{i=0}^{d_0-2} C_{n-1}^i \cdot (q-1)^i$$

и для двоичного кода

$$2^{n-k} > \sum_{i=0}^{d_0-2} C_{n-1}^i.$$

Эти соотношения показывают, при каком значении $n - k$ определенно существует код, гарантированно исправляющий ошибки кратности t_u .

8.1. Уплотнение

В процессе линейного квантования, при котором все интервалы между точками на вертикальной оси равны, импульсы с меньшей амплитудой могут иметь большую процентную величину погрешности, чем импульсы с большей амплитудой. Такой эффект влияет на точность воспроизведения информации, передаваемой при помощи импульсов небольшой амплитуды. Поэтому уделяют особое внимание импульсам с малыми амплитудами. При этом размещают точки на вертикальной оси ближе к низким значениям амплитуды. Таким образом можно достичь прекрасного качества для основной части трафика и хорошего качества для остальной его части.

Для снижения ошибок на этапе квантования в настоящее время используются два алгоритма:

- алгоритм μ -метода, известного также под названием μ -255 (используется в Северной Америке и Японии);
- алгоритм A -метода (используется в Европе и других странах).

Оба метода называют алгоритмами уплотнения, так как они эффективно используются для сжатия сигналов с целью уменьшения ошибок на передающей стороне и обратном их развертывании к исходному состоянию на принимающей.

Оба алгоритма используют приблизительно логарифмические соотношения. На практике для обоих алгоритмов используется схема со 128 положительными десятичными квантовыми значениями (от 0 до 127) и 128 отрицательными (от 0 до 127), отложенными на шкале квантования. Тем не менее, эти значения кодируются по-разному. При кодировании с использованием алгоритма A -метода применяются некоторые математические хитрости, которые помогают преодолеть

несовместимость между алгоритмами уплотнения, а также между сигналами до квантования и сигналами, кодированными битовыми потоками.

8.1.1. Коэффициент сжатия

Наличие в сообщениях избыточности позволяет ставить вопрос о сжатии данных, т. е. о передаче того же количества информации с помощью последовательностей символов меньшей длины. Для этого используются специальные алгоритмы сжатия, уменьшающие избыточность. Эффект сжатия оценивают *коэффициентом сжатия*

$$K = n/q,$$

где n — число минимально необходимых символов для передачи сообщения (практически это число символов на выходе эталонного алгоритма сжатия); q — число символов в сообщении, сжатом данным алгоритмом. Так, при двоичном кодировании n равно энтропии источника информации.

Наряду с методами сжатия, не уменьшающими количество информации в сообщении, применяются методы сжатия, основанные на потере малосущественной информации.

8.1.2. Алгоритмы сжатия

Сжатие данных осуществляется либо на прикладном уровне с помощью программ сжатия, таких как ARJ, либо с помощью устройств защиты от ошибок (УЗО) непосредственно в составе модемов по протоколам типа V.42bis.

Очевидный способ сжатия числовой информации, представленной в коде ASCII, заключается в использовании сокращенного кода с четырьмя битами на символ вместо восьми, так как передается набор, включающий только 10 цифр, символы «точка», «запятая» и «пробел».

Среди простых алгоритмов сжатия наиболее известны *алгоритмы RLE (Run Length Encoding)*. В них вместо передачи цепочки из одинаковых символов передаются символ и значение длины цепочки. Метод эффективен при передаче растровых изображений, но малополезен при передаче текста.

К методам сжатия относят также *методы разностного кодирования*, поскольку разности амплитуд отсчетов представля-

ются меньшим числом разрядов, чем сами амплитуды. Разностное кодирование реализовано в методах дельта-модуляции и ее разновидностях.

Предсказывающие (предиктивные) методы основаны на экстраполяции значений амплитуд отсчетов, и если выполнено условие $A_r - A_p > d$, то отсчет должен быть передан, иначе он является избыточным. Здесь A_r и A_p — амплитуды реального и предсказанного отсчетов, d — допуск (допустимая погрешность представления амплитуд). Здесь точками показаны предсказываемые значения сигнала. Если точка выходит за пределы «коридора» (допуска d), показанного пунктирными линиями, то происходит передача отсчета. На рисунке передаваемые отсчеты отмечены темными кружками в моменты времени t_1, t_2, t_4, t_7 . Если передачи отсчета нет, то на приемном конце принимается экстраполированное значение.

Методы MPEG (Moving Pictures Experts Group) используют предсказывающее кодирование изображений. Так, если передавать только изменившиеся во времени пиксели изображения, то достигается сжатие в несколько десятков раз. Этот алгоритм сжатия используется также в стандарте H.261 ITU. Методы MPEG становятся мировыми стандартами для цифрового телевидения.

Для сжатия данных об изображениях можно использовать также *методы типа JPEG (Joint Photographic Expert Group)*, основанные на потере малосущественной информации (не различимые для глаза оттенки кодируются одинаково, коды могут стать короче). В этих методах передаваемая последовательность пикселей делится на блоки, в каждом блоке производится преобразование Фурье, устраняются высокие частоты, передаются коэффициенты разложения для оставшихся частот, по ним в приемнике изображение восстанавливается.

Другой принцип воплощен во *фрактальном кодировании*, при котором изображение, представленное совокупностью линий, описывается уравнениями этих линий.

8.1.3. Сжатие способом кодирования серий

Наиболее известный простой подход и алгоритм сжатия информации обратимым путем — это кодирование серий последовательностей (*Run Length Encoding*). Суть методов дан-

ного подхода состоит в замене цепочек или серий повторяющихся байтов или их последовательностей на один кодирующий байт и счетчик числа их повторений. Проблема всех аналогичных методов заключается лишь в определении способа, при помощи которого распаковывающий алгоритм мог бы отличить в результирующем потоке байтов закодированную серию от других — не закодированных последовательностей байтов. Решение проблемы достигается обычно простановкой меток в начале закодированных цепочек. Такими метками могут быть, например, характерные значения битов в первом байте закодированной серии, значения первого байта закодированной серии и т. п. Данные методы, как правило, достаточно эффективны для сжатия растровых графических изображений (BMP, PCX, TIF, GIF:), так как последние содержат достаточно много длинных серий повторяющихся последовательностей байтов. Недостатком метода RLE является достаточно низкая степень сжатия или стоимость кодирования файлов с малым числом серий и, что еще хуже — с малым числом повторяющихся байтов в сериях.

8.1.4. Процесс кодирования и его методы

Под кодированием обычно понимают обработку потока символов (в нашем случае байтов или полубайтов) в некотором алфавите, причем частоты появления символов в потоке различны. Целью кодирования является преобразование этого потока в поток бит минимальной длины, что достигается уменьшением энтропии входного потока путем учета частот символов. Длина кода, представляющего символы из алфавита потока, должна быть пропорциональна объему информации входного потока, а длина символов потока в битах может быть не кратна 8 и даже переменной. Если распределение вероятностей частот появления символов из алфавита входного потока известно, то можно построить модель оптимального кодирования. Однако ввиду существования огромного числа различных форматов файлов задача значительно усложняется, так как распределение частот символов данных заранее неизвестно. В таком случае используются два под-
хода.

Первый заключается в просмотре входного потока и построении кодирования на основании собранной статистики (при этом требуется два прохода по файлу — один для просмотра и сбора статистической информации, второй — для кодирования, что несколько ограничивает сферу применения таких алгоритмов, так как таким образом исключается возможность однократного кодирования «на лету», применяемого в телекоммуникационных системах, где и объем данных подчас не известен, а их повторная передача или разбор может занять неоправданно много времени). В таком случае в выходной поток записывается статистическая схема использованного кодирования. Данный метод известен как статическое кодирование Хаффмена.

Второй метод — метод адаптивного кодирования (*adaptive coder method*). Его общий принцип состоит в том, чтобы менять схему кодирования в зависимости от характера изменений входного потока. Такой подход имеет однократный алгоритм и не требует сохранения информации об использованном кодировании в явном виде. Адаптивное кодирование может дать большую степень сжатия по сравнению со статическим, поскольку более полно учитываются изменения частот входного потока. Данный метод известен как динамическое кодирование Хаффмена.

В статическом кодировании Хаффмена входным символам (цепочкам битов различной длины) ставятся в соответствие цепочки битов также переменной длины — их коды. Длина кода каждого символа берется пропорциональной двоичному логарифму его частоты, взятому с обратным знаком. А общий набор всех встретившихся различных символов составляет алфавит потока. Это кодирование является префиксным, что позволяет легко декодировать результирующий поток, так как при префиксном кодировании код любого символа не является префиксом кода никакого другого символа — алфавит уникален.

8.1.5. Арифметическое кодирование

Совершенно иное решение предлагает так называемое арифметическое кодирование. Арифметическое кодирование является методом, позволяющим упаковывать символы входного

алфавита без потерь при условии, что известно распределение частот этих символов, и является наиболее оптимальным, так как достигается теоретическая граница степени сжатия.

Предполагаемая требуемая последовательность символов при сжатии методом арифметического кодирования рассматривается как некоторая двоичная дробь из интервала $(0, 1)$. Результат сжатия представляется как последовательность двоичных цифр из записи этой дроби. Идея метода состоит в следующем: исходный текст рассматривается как записи этой дроби, где каждый входной символ является «цифрой» с весом, пропорциональным вероятности его появления. Этим объясняется интервал, соответствующий минимальной и максимальной вероятностям появления символа в потоке. Поясним работу метода на примере. Пусть алфавит состоит из двух символов: a и b с вероятностями соответственно $3/4$ и $2/4$. Как уже говорилось выше, кодирование Хаффмена не может упаковывать слова в данном алфавите, так как не справляется без сегментации с двухсимвольным алфавитом.

Рассмотрим наш интервал вероятностей $(0, 1)$. Разобьем его на части, длина которых пропорциональна вероятностям символов. В нашем случае это $(0, 3/4)$ и $(3/4, 1)$. Суть алгоритма в следующем: каждому слову во входном алфавите соответствует некоторый подынтервал из интервала $(0, 1)$, а поэтому слову соответствует весь интервал $(0, 1)$. После получения каждого следующего символа интервал уменьшается с выбором той его части, которая соответствует новому символу. Кодом цепочки является интервал, выделенный после обработки всех ее символов, точнее, двоичная запись любой точки из этого интервала, а длина полученного интервала пропорциональна вероятности появления кодируемой цепочки.

Алгоритм декодирования работает синхронно с кодирующим: начав с интервала $(0, 1)$, он последовательно определяет символы входной цепочки. В частности, в нашем случае он вначале разделит (пропорционально частотам символов) интервал $(0, 1)$ на $(0, 0.11)$ и $(0.11, 1)$. Поскольку число 0.0111 (код цепочки «*aaba*») находится в первом из них, можно получить первый символ: «*a*». Затем делим первый подынтервал $(0, 0.11)$ на $(0, 0.1001)$ и $(0.1001, 0.1100)$ (пропорционально частотам символов). Опять выбираем первый, так как $0 < 0.0111 < 0.1001$. Продолжая этот процесс, мы од-

нозначно декодируем все четыре символа. Для того, чтобы декодирующий алгоритм мог определить конец цепочки, мы можем либо передавать ее длину отдельно, либо добавить к алфавиту дополнительный уникальный символ «конец цепочки».

При разработке этого метода возникают две проблемы: во-первых, необходима арифметика с плавающей точкой, теоретически, неограниченной точности, и во-вторых — результат кодирования становится известен лишь при окончании входного потока. Однако дальнейшие исследования показывают, что можно практически без потерь обойтись целочисленной арифметикой небольшой точности (16–32 разряда), а также добиться инкрементальной работы алгоритма: цифры кода могут выдаваться последовательно по мере чтения входного потока при ограничении числа символов входной цепочки каким-либо разумным числом.

8.1.6. Некоторые стандарты сжатия звука для систем связи

Рассмотрим кратко некоторые стандарты кодирования звуковых сигналов, используемые в системах компьютерной видеосвязи и других цифровых системах связи. Следует отметить, что если стандарты MPEG-1,2 наряду с методами сжатия изображений содержат и методы сжатия звукового сопровождения, то рекомендации H.261-H.263 касаются только изображений, а для сжатия звука в системах видеосвязи должны применяться методы, описанные в особых стандартах.

Рекомендация G.711 определяет метод передачи речи с помощью ИКМ. Верхняя граничная частота передаваемого сигнала $f_b = 3,4$ кГц, частота дискретизации $f_d = 8$ кГц, количество разрядов квантования $m = 12$. После квантования выполняется нелинейное преобразование цифрового сигнала (командирование или компрессия) по так называемым m -закону или A -закону, в результате которого шаг квантования при больших уровнях сигнала увеличивается. Передаточная характеристика этого преобразования по форме похожа на характеристику гамма-коррекции. В результате компрессии количество разрядов квантования уменьшается до 8.

Более грубое квантование на больших уровнях громкости не создает заметных искажений звука.

Таким образом, скорость передачи двоичных символов для стандарта G.711 составляет $8 \text{ (бит)} \cdot 8 \text{ (кГц)} = 64 \text{ кбит/с}$. В декодере выполняется обратное нелинейное преобразование с восстановлением 12 бит на каждый отсчет, после чего с помощью ЦАП получается аналоговый звуковой сигнал.

Рекомендации G.721 и G.726 определяют методы сжатия речи с помощью адаптивной дифференциальной импульсно-кодовой модуляции (АДИКМ). Частота дискретизации и число разрядов квантования звукового сигнала такие же, как в рекомендации G.711. Число разрядов квантования разностного сигнала (ошибки предсказания) равно 4. Параметры предсказателя и шаг квантования разностного сигнала (ошибки предсказания) автоматически регулируются в зависимости от текущего спектра и амплитуды входного сигнала. Скорость передачи двоичных символов равна 32 кбит/с .

Рекомендация G.722 направлена на повышение качества передачи речи. Верхняя граничная частота сигнала $4 = 7 \text{ кГц}$, частота дискретизации $4 = 16 \text{ кГц}$, число разрядов квантования $b = 14$. Кодирование включает выделение двух частотных поддиапазонов $50 \text{ Гц} \dots 4 \text{ кГц}$ и $4 \text{ кГц} \dots 7 \text{ кГц}$ и применение АДИКМ с числом разрядов квантования ошибки предсказания 6 в низкочастотном поддиапазоне и 2 в высокочастотном поддиапазоне. Результирующая скорость передачи двоичных символов равна 64 кбит/с .

Рекомендации G.728 и G.729 описывают методы сжатия сигналов речи с использованием кодирования с линейным предсказанием (*Linear Predictive Coding*, LPC). Эти методы основаны на модели голосового аппарата человека в виде линейного фильтра, на вход которого подается или квазипериодическая последовательность импульсов (при формировании гласных и некоторых согласных звуков) или шум (при формировании шипящих, свистящих и некоторых других согласных звуков). При использовании этих методов разборчивая речь на выходе декодера может быть получена при скоростях передачи двоичных символов $4,8 \text{ кбит/с}$ и даже $2,4 \text{ кбит/с}$.

8.1.7. Усовершенствованная система кодирования звука AAC

В 1997 году стандарт MPEG-2 был дополнен усовершенствованной системой кодирования звука AAC (*Advanced Audio Coding*), имеющей следующие основные характеристики:

- число звуковых программ в потоке данных до 16, полное число основных звуковых каналов до 48, низкочастотных каналов до 15, каналов многоязычного сопровождения до 15 и каналов данных до 15;
- частота дискретизации звуковых сигналов от 8 до 96 кГц;
- число обрабатываемых частотных составляющих увеличено до 1024 (на *Layer III* было $32 \times 18 = 576$ составляющих);
- предусмотрены три профиля кодирования: основной (*Main*), упрощенный (*Low Complexity, LC*) и масштабируемый (*Scalable Sampling Rate, SSR*);
- используется адаптивное предсказание кодируемого звукового кадра по предыдущему;
- улучшен алгоритм распределения битов.

Кроме того, AAC содержит ряд других дополнений и усовершенствований по сравнению с *Layer III*.

В результате хорошее качество звучания достигается при сжатии стереозвука до 96 кбит/с, а при 128 кбит/с звук практически неотличим от исходного несжатого звука.

8.2. Системы связи с частотным разделением каналов

Структурная схема многоканальной системы связи с частотным разделением каналов (*Frequency Division Multiplexing, FDM*) приведена на рис. 8.1. Метод разделения состоит в следующем.

Сообщение от источника с номером i ($i = 1n$) поступает на соответствующий канальный модулятор (KM_i), где модулирует по одному из параметров вспомогательное синусоидальное колебание (поднесущую f_i), вырабатываемое генератором поднесущей. Частоты поднесущих отдельных каналов выбирают так, чтобы спектры модулированных поднесущих практически не перекрывались. Модулированные поднесущие назы-

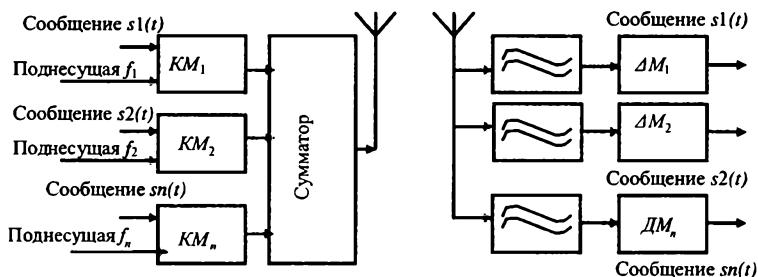


Рис. 8.1. Структурная схема системы связи с частотным разделением каналов

ваются канальными сигналами. В зависимости от вида модуляции канальные сигналы могут быть АМ, ЧМ, ФМ, ИМ и т. д. Канальные сигналы поступают на полосовые фильтры, чтобы уменьшить перекрытие спектров. После суммирования канальных сигналов образуется групповой сигнал. Так как по каналам многоканального передатчика могут передаваться сигналы от различных источников: телефонных аппаратов, фототелеграфных аппаратов и других источников сигнала, сигнал в групповом тракте возбудителя и линейного усилителя является сложным многоисточниковым сигналом. Как видно, групповой сигнал можно рассматривать как случайный нестационарный узкополосный сигнал с известным спектром мощности.

Групповой сигнал поступает на общий модулятор (ОМ) для модуляции несущего колебания, создаваемого в передатчике. При этом также может использоваться АМ, ЧМ, ФМ или ИМ. Таким образом, модуляция получается двухступенчатой.

Высокочастотная несущая, модулированная групповым сигналом, поступает в антенну и далее — в эфир.

В селективных частях приемного устройства происходит выделение и частотное разделение сигналов разных каналов. После усиления принятых сигналов они подвергаются демодуляции в схемах соответствующих демодуляторов (ДМ₁).

Примерами таких систем являются системы радиовещания и телевидения. Например, весь диапазон СВ содержит 120 каналов. Первому каналу соответствует частота 531 кГц, второму — 540 кГц и т. д. до 120-го, которому соответствует

частота 1602 кГц. Как видим, каналы отстоят друг от друга на 9 кГц. На приемной стороне спектры различных каналов выделяются соответствующими полосовыми фильтрами. После фильтров сигналы детектируются для получения переданных сообщений.

8.3. Системы связи с временным разделением

В соответствии с теоремой Котельникова информацию, содержащуюся в непрерывном сигнале с ограниченным частотой f_m спектром, можно передавать в виде отсчетов этого сигнала, взятых через промежутки времени $\Delta t = 1/2 f_m$.

Для передачи канал используется не все время, и поэтому можно использовать временное разделение. При этом образуются выборки всех подлежащих передаче сигналов, которые затем передаются в интервалах между импульсами других сигналов. В приемнике отсчеты, принадлежащие каждому сигналу, выделяются с помощью соответствующих устройств. Структурная схема канала с временным разделением каналов (TDM — *Time Division Multiplexing*) показана на рис. 8.2.

В передатчике коммутатор, управляемый генератором импульсов $G1$, подключен ко всем источникам сообщений. Этот же генератор импульсов вырабатывает отсчетные импульсы. Таким образом, коммутатор последовательно подключает источник сообщений к устройству выборки, которое последовательно берет отсчеты с помощью импульсов, вырабатываемых задающим генератором. Коммутатор и устройство выборки действуют синхронно. Таким образом, выходом устройства

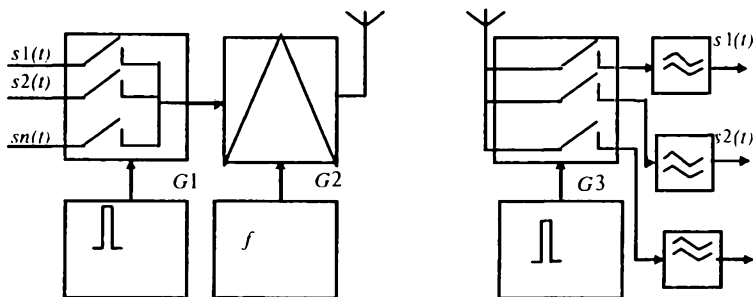


Рис. 8.2. Система связи с временным разделением каналов

выборки являются сигналы, которые состоят из последовательных отсчетов всех сообщений.

В приемнике для распределения импульсов по отдельным каналам используется свой задающий генератор импульсов и коммутатор, которые работают синхронно с коммутатором передатчика.

Отсчеты сообщений оказываются разделенными. Непрерывные сообщения в каждом канале восстанавливаются с помощью фильтров низких частот.

Принципы временного разделения каналов базируются на теореме Котельникова. Возможность передачи не всего непрерывного сообщения, а только отдельных значений лежит в основе всех систем связи, использующих временное разделение каналов. При этом отсчеты сообщений N различных каналов нужно передавать не одновременно, а поочередно, один за другим, так, чтобы каждому каналу на некотором временном интервале $\Delta t = T_n$ предоставлялся свой промежуток ΔT_k , называемый канальным интервалом.

Практически передача мгновенных значений сообщений осуществляется с помощью последовательности коротких импульсов длительностью $\tau \ll T_n$, которая модулируется по одному из параметров: амплитуде, длительности или временно-положению.

Таким образом, в канальных трактах возможен один из трех видов импульсной модуляции: АИМ, ШИМ или ВИМ.

В общем тракте импульсная последовательность модулируется высокочастотным гармоническим колебанием. Здесь возможен любой вид модуляции: АМ, ЧМ или ФМ.

Системы с временным разделением каналов принято классифицировать по видам модуляции, используемым в канальном и общем трактах: АИМ-АМ, АИМ-ЧМ, ВИМ-АМ и т. д.

8.4. Сравнение систем с частотным и временным разделением

В системе связи с частотным разделением все сигналы непрерывны и смешиваются во временной области. Однако спектры различных сигналов занимают разные полосы в частотной области и могут быть разделены с помощью фильтров. При временном разделении отсчеты каждого сигнала различ-

ны и могут быть разделены во временной области. Тем не менее, частотные спектры этих различных выборочных сигналов занимают одну и ту же полосу частот. Итак, при частотном разделении сохраняется частотный признак сигнала, тогда как признак формы колебания сохраняется при временном разделении. Поскольку сигнал полностью определяется либо своим частотным представлением, либо временным, принятые сигналы можно разделить в приемнике соответствующими способами.

При частотном разделении каждый сигнал передается по каналу все время, и все сигналы смешиваются, но каждый из них занимает свой конечный частотный диапазон, не занятый другими сигналами.

При временном разделении каждый сигнал занимает свой временной интервал, не занятый другими сигналами. Но спектры всех сигналов лежат в одной и той же частотной области.

Полосы частот, необходимые для передачи данного числа сигналов, одинаковы при частотном и временном разделении. Отсюда следует, что в данной полосе частот можно одновременно передавать одинаковое число сигналов, используя частотное или временное разделение.

С практической точки зрения, системы с временным разделением имеют преимущества перед системами с частотным разделением. Система с временным разделением проще, так как в системе с частотным разделением надо генерировать свои несущие частоты, при этом каждый канал занимает свою полосу частот, и требуются различные полосовые фильтры. А в системе с временным разделением в каждом канале используется одинаковое оборудование, состоящее из простых коммутаторов. Фильтры, необходимые для детектирования, — это ФНЧ одинаковые для всех каналов, они проще модуляторов, демодуляторов, генераторов и полосовых фильтров в системе с частотным разделением.

Система с временным разделением нечувствительна к искажениям в канале, которые возникают в системе с частотным разделением вследствие нелинейности вольт-амперных характеристик усилителей.

В системе с временным разделением сигналы передаются не одновременно, поэтому требования к нелинейности тракта

многоканальной системы такие же, как и в одноканальной системе. По этим причинам в телефонной связи на большие расстояния используются системы с временным разделением каналов.

8.5. Методы множественного доступа

Понятие множественного доступа (*Multiple Access*) связано с организацией совместного использования ограниченного участка спектра многими пользователями. Можно выделить 5 вариантов множественного доступа:

- множественный доступ с временным разделением каналов связи (*Time Division Multiplex Access*, TDMA);
- множественный доступ с частотным разделением каналов связи (*Frequency Division Multiplex Access*, FDMA);
- множественный доступ с кодовым разделением каналов связи (*Code Division Multiplex Access*, CDMA);
- множественный доступ с пространственным разделением каналов связи (*Space Division Multiplex Access*, SDMA);
- множественный доступ с поляризационным разделением каналов связи (*Polarization Division Multiplex Access*, PDMA).

Практический интерес представляют первые три.

8.5.1. TDMA

Технология TDMA представляет собой мультиплексирование с временным разделением сигналов в пределах одного канала. При создании системы TDMA канал связи должен иметь достаточно высокую пропускную способность, чтобы обеспечить множественным абонентам возможность пользоваться этим каналом одновременно. Система TDMA, как правило, сочетается с FDMA. Различные спецификации сотовых сетей предусматривают несколько вариантов ширины канала связи.

В целях поддержания совместимости с аналоговыми системами североамериканский стандарт TDMA предусматривает использование каналов связи с полосой рабочих частот шириной 30 кГц. Европейский стандарт оперирует с каналами шириной 200 кГц. Число временных интервалов у этих каналов также разное.

Дуплексная связь в системе с TDMA осуществляется за счет подключения технологии FDMA, т. е. представляет собой дуплексную связь с частотным разделением каналов.

8.5.2. FDMA

Технология FDMA представляет собой систему связи, использующей частотное разделение каналов. Доступный спектр частот разбит на каналы нескольких типов, которые предназначены для управления, сигнализации или передачи собственно полезной информации, т. е. человеческой речи. Данная технология применяется с традиционными аналоговыми системами. Ширина каждого канала при этом составляет 30 кГц.

FDMA решает вопрос распределения каналов связи, однако не может справиться с проблемой дуплексных операций. При осуществлении дуплексной связи с частотным разделением каналов решение этой проблемы состоит в использовании двух симплексных каналов связи, которые изолированы друг от друга для снижения вероятности интерференции в конечных точках электрических цепей.

Сам по себе множественный доступ с частотным разделением каналов способен функционировать только тогда, когда каналы для приема и передачи данных разделены друг от друга. В том случае, когда доступен только один канал из всего частотного диапазона, применяется дуплексная связь с временным разделением каналов. При сочетании FDMA с дуплексной связью и временным разделением каналов типичное разделение каналов по частоте сменяется их разделением по времени.

В отличие от «чистых» FDMA применяют дуплексные связи с временным разделением каналов в рамках системы. FDMA требует, чтобы все соты были синхронизированы. Более того, временное разделение каналов в данном случае, как правило, не обеспечивает увеличения эффективности спектра с точки зрения объема передаваемой информации или емкости канала связи. Это связано с тем, что защитный промежуток между временными интервалами получается больше, чем в «чистых» системах с временным разделением каналов (TDMA), а также больше времени прохождения полного цикла в системе с частотным разделением каналов (FDMA).

8.5.3. CDMA

Множественный доступ с кодовым разделением каналов относится к области военной и спутниковой связи, где главными достоинствами связи являются скрытность и высокая помехоустойчивость. Важная особенность CDMA — это работа в полосе частот, значительно превышающей полосу частот речевого сигнала, в сочетании ее с таким кодированием информации, которое позволяет выделять ее из общей широкой полосы, используемой одновременно всеми физическими каналами. Системы связи, реализующие CDMA, являются системами с расширенным спектром (*spread spectrum*), спектр информационного сообщения в них искусственно расширяется посредством модуляции (кодирования) периодической псевдослучайной последовательностью импульсов с достаточно малым дискретом.

Множественный доступ с кодовым разделением каналов связи основан на принципе, который подразумевает, что если известно, где нужно услышать и что нужно услышать, то можно выделить отдельный диалог из множества диалогов, ведущихся вполголоса (т. е. на уровне фонового шума) в комнате, полной людей. В случае применения технологии CDMA множество абонентов использует один и тот же диапазон частот, не рискуя при этом помешать «чужим» диалогам. Предоставление каждому абоненту возможности однозначно идентифицировать нужный ему диалог среди множества других позволяет намного более интеллектуально использовать спектр частот.

8.5.4. Множественный доступ с пространственным и поляризационным разделением каналов

Спутник INTELSAT IVA использует метод множественного доступа к пространственным разделениям каналов, называемый многолучевым многократным использованием частоты. Полосы частот одинаковы, но сигналы разнесены в пространстве. В таких случаях полосу частот называют многократно используемой.

Спутник COMSTAR 1 использует множественный доступ с поляризационным разделением каналов, который еще назы-

вают методом двойного поляризационного многократного использования частоты. В данном случае используется антенна с разными поляризациями, каждая из которых соотносится со своим приемником. Это позволяет реализовать одновременный доступ к спутнику пользователям, находящимся на небольших расстояниях друг от друга. Каждая из передающих антенн на Земле должна быть поляризована в соответствии с антенной спутника.

8.6. Растянутый спектр

Разновидность связи с растянутым спектром не подпадает ни под одну из разновидностей, рассмотренных выше, так как она позволяет передавать аналоговые или цифровые сигналы с помощью аналоговых сигналов.

Технология растянутого спектра изначально разрабатывалась для удовлетворения требований военных и разведки. Суть метода сводится к растягиванию спектра информационного сигнала на более широкую полосу с целью затруднить его подавление и перехват.

Метод перескока частоты — первый метод связи с растянутым спектром. Более современный способ — метод прямой последовательности. Оба эти метода используются в беспроводной связи.

На рис. 8.3 отмечены ключевые характеристики системы с расширением спектра.

Входной сигнал подается на кодер, генерирующий сигнал с относительно узкой полосой. Затем сигнал модулируется с использованием последовательности псевдослучайных чисел. Задача модулирования — расширить полосу частот, занимаемых спектром сигнала. На приемном конце канала связи производится демодуляция сигнала с расширенным спектром, для чего используется та же последовательность псевдослучайных чисел. После этого сигнал подается на каналный декодер для восстановления сигнала.

Последовательность псевдослучайных чисел генерируется по алгоритму, в котором используется некоторое исходное число. Алгоритм является детерминированным, поэтому генерируемые числа не являются случайными в статистическом смысле слова. В то же время последовательность успешно

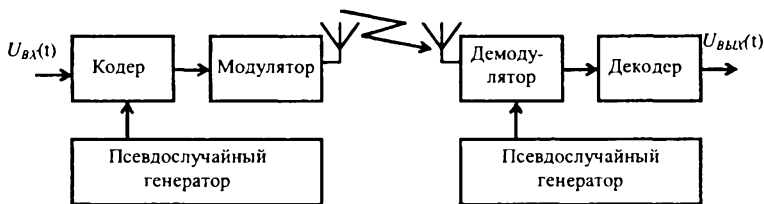


Рис. 8.3. Канал связи с расширением спектра

проходит испытания через тесты на их случайность. Поэтому числа называются псевдослучайными.

Предсказать последовательность без знания алгоритма и начального числа невозможно. Следовательно, декодировать сигнал сможет только приемник, совместно с передатчиком владеющий нужной информацией.

8.6.1. Метод перескока частоты

При использовании этого метода для передачи сигнала используются якобы случайные последовательности радиочастот, причем переход с одной частоты на другую происходит за доли секунды. Приемник, синхронно с передатчиком «перескакивающий» с частоты на частоту, может принимать сообщения. Те же, кто будет подслушивать передачу, смогут услышать лишь невнятный шум. Глушить сигнал бесполезно: это сделает неразборчивыми лишь несколько первых битов сигнала.

8.6.2. Метод простой последовательности

При использовании этой схемы каждый бит исходного сигнала представляется несколькими битами передаваемого сигнала, называемого *раздробленным кодом*. Данный код растягивает сигнал на более широкую полосу частот, причем степень растяжения полосы прямо пропорциональна количеству используемых битов. Следовательно, десятибитовый растягивающий код растянет сигнал на полосу частот в 10 раз большую, чем однобитовый код.

Один из приемов — объединение цифрового информационного потока с потоком последовательности случайных чисел, для чего используется схема исключающего ИЛИ. Сказанное иллюстрируется рис. 8.4.

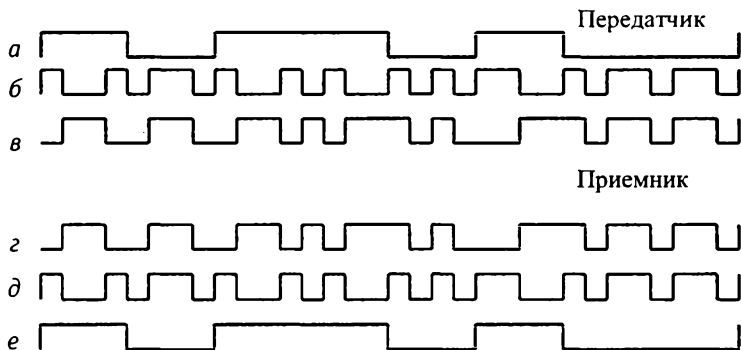


Рис. 8.4. Растягивание спектра методом прямой последовательности:

a — входные данные; *b* — локально генерируемый псевдослучайный поток битов; *v* — переданный сигнал ($v = a \oplus b$);
z — принятый сигнал *v*; *d* — локально генерируемый псевдослучайный поток битов, поток битов, идентичный потоку *b*;
e — выходные данные ($e = z \oplus d$)

Отметим, что единичный бит информации инвертирует псевдослучайный бит, тогда как нулевой бит информации псевдослучайный бит не изменяет. Сигнал, получающийся после сложения информационного и псевдослучайного сигналов, передается со скоростью, равной скорости передачи исходной псевдослучайной последовательности. Поэтому полоса полученного сигнала шире полосы информационного сигнала. В примере, приведенном на рис. 8.4, скорость псевдослучайного потока битов в 4 раза выше скорости информационного потока.

Рисунок 8.4. раскрывает идею реализации метода прямой последовательности. В этом случае информационный поток и поток случайных чисел пропускаются через схему исключающего ИЛИ, после чего осуществляется процесс модуляции. Расширение спектра, происходящее за счет использования метода прямой последовательности, определяется довольно просто. Предположим, например, что ширина бита информационного сигнала равна t_b . Тогда ширина спектра сигнала равна $2/T_c$, где T_c — ширина спектра бита псевдослучайной последовательности. Ширина спектра суммарного сигнала

приблизительно равна сумме всех битовых спектров. Очевидно, что величина полученного расширения прямо пропорциональна скорости передачи псевдослучайного потока: чем выше скорость передачи этого потока, тем больше расширение спектра.

8.7. Технологии уширения спектра беспроводных протоколов

В основе всех беспроводных протоколов семейства 802.11 лежит технология уширения спектра (*Spread Spectrum*, SS). Данная технология подразумевает, что первоначально узкополосный (в смысле ширины спектра) полезный информационный сигнал при передаче преобразуется таким образом, что его спектр оказывается значительно шире спектра первоначального сигнала, т. е. спектр сигнала как бы «размазывается» по частотному диапазону. Одновременно с уширением спектра сигнала происходит и перераспределение спектральной энергетической плотности сигнала — энергия сигнала также «размазывается» по спектру. В результате максимальная мощность преобразованного сигнала оказывается значительно ниже мощности исходного сигнала. При этом уровень полезного информационного сигнала может в буквальном смысле сравниваться с уровнем естественного шума. В результате сигнал становится в каком-то смысле «невидимым» — он просто теряется на уровне естественного шума.

Собственно, именно в изменении спектральной энергетической плотности сигнала и заключается идея уширения спектра. Дело в том, что если подходить к проблеме передачи данных традиционным способом, т. е. так, как это делается в радиоэфире, где каждой радиостанции отводится свой диапазон вещания, то мы неизбежно столкнемся с проблемой, что в ограниченном радиодиапазоне, предназначенном для совместного использования, невозможно «уместить» всех желающих. Поэтому необходимо найти такой способ передачи информации, при котором пользователи могли бы сосуществовать в одном частотном диапазоне и при этом не мешать друг другу. Именно эту задачу и решает технология уширения спектра.

Существует несколько различных технологий уширения спектра, однако для дальнейшего понимания протоколов беспроводной передачи информации нам необходимо детально познакомиться лишь с технологией уширения спектра методом прямой последовательности (*Direct Sequence Spread Spectrum*, DSSS), кодов Баркера и применения ССК-последовательностей.

8.7.1. Технология DSSS

При потенциальном кодировании информационные биты — логические нули и единицы — передаются прямоугольными импульсами напряжений. Прямоугольный импульс длительности T имеет спектр, ширина которого обратно пропорциональна длительности импульса. Поэтому чем меньше длительность информационного бита, тем больший спектр занимает такой сигнал.

Для преднамеренного уширения спектра первоначально узкополосного сигнала в технологии DSSS в каждый передаваемый информационный бит (логический 0 или 1) в буквальном смысле встраивается последовательность так называемых чипов. Если информационные биты — логические нули или единицы — при потенциальном кодировании информации можно представить в виде последовательности прямоугольных импульсов, то каждый отдельный чип — это тоже прямоугольный импульс, но его длительность в несколько раз меньше длительности информационного бита. Последовательность чипов представляет собой последовательность прямоугольных импульсов, т. е. нулей и единиц, однако эти нули и единицы не являются информационными. Поскольку длительность одного чипа в n раз меньше длительности информационного бита, то и ширина спектра преобразованного сигнала будет в n раз больше ширины спектра первоначального сигнала. При этом и амплитуда передаваемого сигнала уменьшится в n раз.

Чиповые последовательности, встраиваемые в информационные биты, называют шумоподобными кодами (PN -последовательности), что подчеркивает то обстоятельство, что результирующий сигнал становится шумоподобным и его трудно отличить от естественного шума.

Как уширить спектр сигнала и сделать его неотличимым от естественного шума, понятно. Для этого, в принципе, можно воспользоваться произвольной (случайной) чиповой последовательностью. Однако возникает вопрос: а как такой сигнал принимать? Ведь если он становится шумоподобным, то выделить из него полезный информационный сигнал не так просто, если вообще возможно. Оказывается, возможно, но для этого нужно соответствующим образом подобрать чиповую последовательность. Используемые для уширения спектра сигнала чиповые последовательности должны удовлетворять определенным требованиям автокорреляции. Под термином автокорреляции в математике подразумевают степень подобия функции самой себе в различные моменты времени. Если подобрать такую чиповую последовательность, для которой функция автокорреляции будет иметь резко выраженный пик лишь для одного момента времени, то такой информационный сигнал возможно будет выделить на уровне шума. Для этого в приемнике полученный сигнал умножается на ту же чиповую последовательность, т. е. вычисляется автокорреляционная функция сигнала. В результате сигнал становится опять узкополосным, поэтому его фильтруют в узкой полосе частот и любая помеха, попадающая в полосу исходного широкополосного сигнала, после умножения на чиповую последовательность, наоборот, становится широкополосной и обрезается фильтрами, а в узкую информационную полосу попадает лишь часть помехи, по мощности значительно меньшая, чем помеха, действующая на входе приемника.

8.7.2. Коды Баркера

Чиповых последовательностей, отвечающих указанным требованиям автокорреляции, существует достаточно много, но для нас особый интерес представляют так называемые коды Баркера, поскольку именно они используются в протоколах стандарта 802.11.

Коды Баркера обладают наилучшими среди известных псевдослучайных последовательностей свойствами шумоподобности, что и обусловило их широкое применение. Например, в протоколах стандарта 802.11 используется код Баркера длиной в 11 чипов (11100010010).

Для того чтобы передать сигнал, логическая единица передается прямой последовательностью Баркера, а логический ноль — инверсной последовательностью.

Скорость 1 Мбит/с.

В стандарте 802.11 предусмотрено два скоростных режима: 1 и 2 Мбит/с. Для кодирования данных на физическом уровне используется метод DSSS с 11-чиповыми кодами Баркера. При информационной скорости 1 Мбит/с скорость следования отдельных чипов последовательности Баркера составляет $11 \cdot 10^6$ чип/с, а ширина спектра такого сигнала составляет 22 МГц. Учитывая, что ширина частотного диапазона составляет 83,5 МГц, получаем, что всего в данном частотном диапазоне можно уместить 3 неперекрывающихся частотных канала. Весь частотный диапазон, однако, принято делить на 11 частотных перекрывающихся каналов по 22 МГц, отстоящих друг от друга на 5 МГц. К примеру, первый канал занимает частотный диапазон от 2400 до 2423 МГц и центрирован относительно частоты 2412 МГц. Второй канал центрирован относительно частоты 2417 МГц, а последний, 11 канал центрирован относительно частоты 2462 МГц. При таком рассмотрении первый, шестой и 11 каналы не перекрываются друг с другом и имеют 3-мегагерцовый зазор друг относительно друга. Именно эти три канала могут использоваться независимо друг от друга.

Для модуляции синусоидального несущего сигнала (процесс, необходимый для информационного наполнения несущего сигнала) используется относительная двоичная фазовая модуляция (*Differential Binary Phase Shift Key*, DBPSK). При этом кодирование информации происходит за счет сдвига фазы синусоидального сигнала по отношению к предыдущему состоянию сигнала. Двоичная фазовая модуляция предусматривает два возможных значения сдвига фазы: 0 и π . Тогда логический ноль может передаваться синфазным сигналом (сдвиг по фазе равен 0), а единица — сигналом, который сдвинут по фазе на π .

Скорость 2 Мбит/с.

Информационная скорость 1 Мбит/с является обязательной в стандарте IEEE 802.11 (*Basic Access Rate*), но опционально возможна и скорость в 2 Мбит/с (*Enhanced Access Rate*). Для передачи данных на такой скорости используется та же

технология DSSS с 11-чиповыми кодами Баркера, но для модуляции несущего колебания применяется относительная квадратурная фазовая модуляция (*Differential Quadrature Phase Shift Keying*). При относительной квадратурной фазовой модуляции сдвиг фаз может принимать четыре различных значения: 0, $\pi/2$, π и $3\pi/2$. Используя четыре различных состояния сигнала, можно в одном дискретном состоянии закодировать последовательность двух информационных бит (дибит) и тем самым в два раза повысить информационную скорость передачи. К примеру, дибиту 00 может соответствовать сдвиг фазы, равный 0; дибиту 01 — сдвиг фазы, равный $\pi/2$; дибиту 11 — сдвиг фазы, равный π ; дибиту 10 — сдвиг фазы, равный $3\pi/2$.

В заключение рассмотрения физического уровня протокола 802.11 отметим, что при информационной скорости 2 Мбит/с скорость следования отдельных чипов последовательности Баркера остается прежней, т. е. $11 \cdot 10^6$ чип/с, а следовательно, не меняется и ширина спектра передаваемого сигнала.

Физический уровень протокола 802.11b/b+.

Протокол IEEE 802.11b, принятый в июле 1999 года, является своего рода расширением базового протокола 802.11 и кроме скоростей 1 и 2 Мбит/с предусматривает скорости 5,5 и 11 Мбит/с. Для работы на скоростях 1 и 2 Мбит/с используются технология уширения спектра с использованием кодов Баркера, а для скоростей 5,5 и 11 Мбит/с используются так называемые комплементарные коды (*Complementary Code Keying*, ССК).

8.7.3. ССК-последовательности

Комплементарные коды или ССК-последовательности обладают тем свойством, что сумма их автокорреляционных функций для любого циклического сдвига, отличного от нуля, всегда равна нулю. Последнее обстоятельство позволяет легко выделять эти последовательности на уровне шума, что в значительной степени увеличивает помехоустойчивость при передаче данных.

Элементы 8-чиповой ССК-последовательности образуются по формулам:

$$c_0 = e^{i(\varphi_1 + \varphi_2 + \varphi_3 + \varphi_4)}, \quad c_1 = e^{i(\varphi_1 + \varphi_3 + \varphi_4)},$$

$$c_2 = e^{i(\varphi_1 + \varphi_2 + \varphi_4)}, \quad c_3 = e^{i(\varphi_1 + \varphi_4)},$$

$$c_4 = e^{i(\varphi_1 + \varphi_2 + \varphi_3)}, \quad c_5 = e^{i(\varphi_1 + \varphi_3)},$$

$$c_6 = -e^{i(\varphi_1 + \varphi_2)}, \quad c_7 = e^{i\varphi_1}.$$

Значения фазы φ_1 выбирается по первому дибиту, φ_2 — по второму, φ_3 — по третьему и φ_4 — по четвертому. Таким образом, для однозначного определения ССК-последовательности требуется 8 *бит* входных данных. Обратите внимание, что фаза $i\varphi$, присутствует во всех членах последовательности. Практически это означает сдвиг по фазе всех членов последовательности на одно и то же значение. По этой причине первый дибит данных задает сдвиг целого символа по фазе по отношению к фазе предыдущего переданного символа.

Для скорости 5,5 *Мбит/с* в одном символе кодируется 4 *бита* (два дибита): $\{d_0, d_1, d_2, d_3\}$. Для того чтобы закодировать 4 *бита* в одном символе, необходимо иметь 16 различных дискретных состояний сигнала, каждое из которых определяется той или иной комплементарной ССК-последовательностью. Поэтому для реализации скорости 5,5 *Мбит/с* требуется иметь набор из 16 различных комплементарных последовательностей. Такой набор сформировать нетрудно, поскольку сами последовательности являются 8-чиповыми и определяются на множестве комплексных элементов $\{+1, -1, +j, -j\}$, т. е. всего можно сформировать 65 536 различных последовательностей.

Выбор требуемой последовательности для кодирования символа происходит следующим образом. Поступающий поток битов группируется по два дибита, то есть каждая группа формирует один символ. При этом символы делятся на четные и нечетные, а первый дибит $\{d_0, d_1\}$ задает фазовый сдвиг четных и нечетных символов. Поскольку каждый дибит может принимать четыре различных значения, то и для четных, и нечетных символов существуют по четыре возможных сдвига фаз (табл. 8.1).

Следующий дибит $\{d_2, d_3\}$ определяет остальные фазы ССК-последовательности по формулам: $\varphi_2 = d_2\pi + \pi/2$, $\varphi_3 = 0$, $\varphi_3 = d_3\pi$.

**Фазовые сдвиги символов,
определяемые первым дибитом $\{d_0, d_1\}$**

$\{d_0, d_1\}$	Фазовый сдвиг четных символов	Фазовый сдвиг нечетных символов
00	0	π
01	$\pi/2$	$-\pi/2$
11	π	0
10	$-\pi/2$	$\pi/2$

Таким образом, для задания СКК-последовательности при кодировании 4 *бит/символ* используют только 2 *бита* данных (второй дибит). Этот дибит, принимающий одно из четырех возможных значений, позволяет задать одну из четырех комплементарных СКК-последовательностей. Первый дибит определяет сдвиг по фазе всего символа (причем возможен сдвиг на четыре различных значения) и применяется в относительной фазовой модуляции. Следовательно, набор из четырех возможных комплементарных последовательностей с последующей относительной квадратурной фазовой модуляцией позволяет сформировать 16 дискретных состояний сигнала и закодировать 4 *бита* в одном символе. Как нетрудно рассчитать, при информационной скорости 5,5 *Мбит/с* символьная скорость составит $5,5/4 = 1,375$ мегасимволов в секунду. Учитывая, что сами по себе СКК-последовательности являются 8-чиповыми, частота следования отдельных чипов составляет 11 *МГц*, а ширина спектра сигнала — соответственно 22 *МГц*.

При скорости 11 *Мбит/с* в одном символе кодируется одновременно 8 *бит* данных. При этом первый дибит последовательности данных, как и прежде, задает сдвиг фазы при относительной фазовой модуляции целого символа в зависимости от того, четный он или нечетный, а остальные три дибита 8-битовой последовательности данных определяют оставшиеся фазы, причем значение φ_2 выбирается по второму дибиту, φ_3 — по третьему и φ_4 — по четвертому. Значение сдвига фаз определяется по табл. 8.2.

Итак, для задания СКК-последовательности используют только 6 бит данных (второй, третий и четвертый дибиты). А поскольку 6 бит данных могут иметь 64 различные комбина-

Фазовые сдвиги символа, определяемые вторым, третьим и четвертым дибитами

$\{d_p, d_{i+1}\}$	Фазовый сдвиг символа
00	0
01	$p/2$
10	p
11	$-p/2$

ции, то при кодировании каждого символа применяется одна из 64 возможных восьмиразрядных ССК-последовательностей.

Набор из 64 возможных комплементарных последовательностей с последующей относительной квадратурной фазовой модуляцией позволяет сформировать 256 дискретных состояний сигнала и закодировать 8 бит в одном символе. При информационной скорости 11 Мбит/с символьная скорость составит $11/8 = 1,375 \cdot 10^6$ символов в секунду. Учитывая, что сами по себе ССК-последовательности являются 8-чиповыми, частота следования отдельных чипов составляет 11 МГц ($1,375 \cdot 8$ МГц), а ширина спектра сигнала — 22 МГц.

В протоколе 802.11g используются сверточные кодеры, состоящие из шести запоминающих ячеек (рис. 8.6). Поэтому всего существует семь различных состояний входного сигнала: текущее и шесть значений, хранящихся в запоминающих ячейках. Такие кодеры называются сверточными кодерами на семь состояний ($K = 7$). Кроме того, каждому входному биту X_0 ставятся в соответствие два выходных бита — Y_0 и

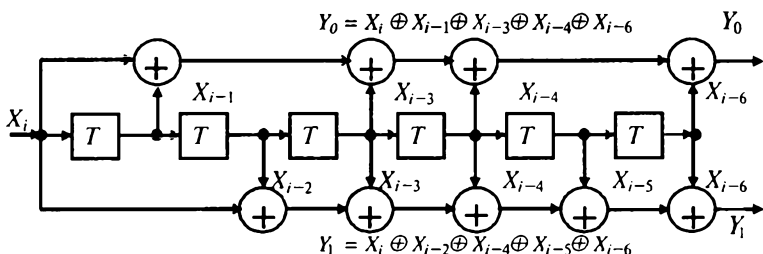


Рис. 8.6. Схема сверточного кодера ($K = 7$),
скорость кодирования $r = 1/2$

Y_j . Если скорость входной последовательности составляет k бит/с, то скорость выходной последовательности — k ди-бит/с или $2k$ бит/с. При этом говорят, что скорость кодирования равна $1/2$. Значение каждого формируемого дибита $\{Y_0, Y_j\}$ зависит не только от входящего информационного бита X_0 , но и от шести предыдущих битов, значения которых хранятся в 6 запоминающих ячейках.

Главным достоинством сверточных кодеров является помехоустойчивость формируемой ими последовательности. Дело в том, что при избыточности кодирования (вспомним, что каждому информационному биту ставится в соответствие дибит) даже в случае возникновения ошибок приема (к примеру, вместо дибита 11 ошибочно принят дибит 10) исходная последовательность битов может быть безошибочно восстановлена. Для этого на стороне приемника применяется декодер Витерби.

В стандарте IEEE 802.11b речь идет о комплексных элементарных 8-чиповых последовательностях, определенных на множестве комплексных элементов. Используя множество комплексных элементов $\{1, -1, j, -j\}$ можно сформировать восемь одинаковых по модулю, но отличающихся по фазе комплексных чисел. Иными словами, элементы 8-чиповой ССК-последовательности могут принимать одно из следующих 8 значений: $1, -1, j, -j, 1 + j, 1 - j, -1 + j, -1 - j$. Основное отличие ССК-последовательностей от рассмотренных ранее кодов Баркера заключается в том, что существует не строго заданная последовательность, посредством которой можно было кодировать либо логический нуль, либо единицу, а целый набор последовательностей. Учитывая, что каждый элемент 8-чиповой последовательности может принимать одно из восьми значений в зависимости от значения фазы, ясно, что можно скомбинировать $8^8 = 16777216$ вариантов последовательностей, однако не все они будут комплементарными. Но даже с учетом требования комплементарности можно сформировать достаточно большое число разных ССК-последовательностей. Это обстоятельство позволяет кодировать в одном передаваемом символе несколько информационных бит и тем самым повысить информационную скорость передачи.

Вообще говоря, использование ССК-кодов позволяет кодировать 8 бит на один символ при скорости 11 Мбит/с и

4 *бит* на символ при скорости 5,5 *Мбит/с*. При этом в обоих случаях символьная скорость передачи составляет $1,375 \cdot 10^6$ символов в секунду ($11/8 = 5,5/4 = 1,375$), а учитывая, что каждый символ задается 8-чиповой последовательностью, получаем, что в обоих случаях скорость следования отдельных чипов составляет $11 \cdot 10^6$ чипов в секунду. Соответственно, и ширина спектра сигнала как скорости 11 *Мбит/с* и 5,5 *Мбит/с* составляет 22 *МГц*.

Рассматривая возможные скорости передачи 5,5 и 11 *Мбит/с* в протоколе 802.11b, мы до сих пор оставляли без внимания вопрос, зачем нужна скорость 5,5 *Мбит/с*, если использование ССК-последовательностей позволяет передавать данные на скорости 11 *Мбит/с*. Теоретически это действительно так, но только если не учитывать при этом помеховой обстановки. В реальных условиях зашумленность каналов передачи и соответственно соотношение уровней шума и сигнала может оказаться таковым, что передача на высокой информационной скорости, т. е. когда в одном символе кодируется множество информационных бит, может оказаться невозможной по причине их ошибочного распознавания. Отметим, что чем выше зашумленность каналов связи, тем меньше информационная скорость передачи. При этом важно, что приемник и передатчик должны правильно оценивать помеховую обстановку и выбирать приемлемую скорость передачи.

Пунктурное кодирование

Опционально в протоколе 802.11g технология двоичного пакетного сверточного кодирования PBCC может применяться при скоростях передачи 22 и 33 *Мбит/с*. Отметим, что скорость 22 *Мбит/с* при использовании технологии PBCC предусмотрена и в стандарте 802.11b+. При скорости 22 *Мбит/с*, по сравнению с уже рассмотренной нами схемой PBCC, имеются два отличия. Прежде всего, применяется не квадратичная, а 8-позиционная фазовая модуляция 8-PSK, т. е. фаза сигнала может принимать восемь различных значений. Это позволяет в одном символе кодировать уже не два, а три бита и, следовательно, увеличить информационную скорость передачи.

Кроме сверточного кодера в схему добавлен пунктурный кодер (*Puncture*). Смысл такого решения достаточно прост:

избыточность сверточного кодера, равная 2 (на каждый входной бит приходится два выходных), достаточно высока и при определенных условиях помеховой обстановки является излишней, поэтому можно уменьшить избыточность, чтобы, к примеру, каждым двум входным битам соответствовали три выходных.

Для этого можно, конечно, разработать соответствующий сверточный кодер, но лучше добавить в схему блок, который будет просто уничтожать лишние биты. Допустим, что пунктурный кодер удаляет один бит из каждых четырех входных битов. Тогда каждым четырем входящим битам будут соответствовать три выходящих. Скорость такого кодера составляет 4:3.

Если же такой кодер используется в паре со сверточным кодером со скоростью $1/2$, то общая скорость кодирования составляет уже $2/3$, то есть каждым двум входным битам будут соответствовать 3 выходных.

Ортогональное частотное разделение каналов

Передача радиосигналов в открытом пространстве неизбежно сопровождается многолучевой интерференцией. Этот эффект заключается в том, что в результате многократных отражений один и тот же сигнал может попадать в приемник разными путями. Но различные пути распространения имеют и разные длины, а потому ослабление сигнала для них будет неодинаковым. Следовательно, в точке приема результирующий сигнал представляет собой суперпозицию (интерференцию) многих сигналов с различными амплитудами и начальными фазами, что эквивалентно сложению сигналов с разными фазами.

Следствием многолучевой интерференции является искажение принимаемого сигнала. Многолучевая интерференция присуща любому типу сигналов, но особенно негативно она сказывается на широкополосных сигналах. Дело в том, что при использовании широкополосного сигнала в результате интерференции одни частоты складываются синфазно, что приводит к увеличению сигнала, а другие, наоборот, противофазно, вызывая ослабление сигнала на данной частоте.

Говоря о многолучевой интерференции, возникающей при передаче сигналов, различают два крайних случая. В первом

случае максимальная задержка между различными сигналами не превышает длительности одного символа, и интерференция возникает в пределах одного передаваемого символа. Во втором случае максимальная задержка между различными сигналами больше длительности одного символа, а в результате интерференции складываются сигналы, представляющие разные символы, — так возникает межсимвольная интерференция (*Inter Symbol Interference*, ISI).

Наиболее сильно на искажении сигнала сказывается межсимвольная интерференция. Поскольку символ — это дискретное состояние сигнала, характеризующееся значениями частоты несущей, амплитуды и фазы, то для различных символов меняются амплитуда и фаза сигнала, поэтому восстановить исходный сигнал крайне сложно.

Чтобы избежать, а точнее, частично компенсировать эффект многолучевого распространения, используются частотные эквалайзеры, однако по мере роста скорости передачи данных, либо за счет увеличения символьной скорости, либо за счет усложнения схемы кодирования эффективность их применения падает.

При скорости передачи 11 или 22 Мбит/с в случае использования ССК-кодов или пакетного сверточного кодирования схемы компенсации межсимвольной интерференции вполне успешно справляются с возложенной на них задачей, но при более высоких скоростях такой подход становится неприемлемым. Поэтому для реализации более высоких скоростей передачи в стандарте 802.11g применяется принципиально иной метод кодирования данных, который состоит в том, что поток передаваемых данных распределяется по множеству частотных подканалов и передача ведется параллельно на всех этих подканалах. При этом высокая скорость передачи достигается именно за счет одновременной передачи данных по всем каналам, причем скорость передачи в отдельном подканале может быть и невысокой.

Поскольку в каждом из частотных подканалов скорость передачи данных можно сделать не слишком высокой, это создает предпосылки для эффективного подавления межсимвольной интерференции.

При частотном разделении каналов необходимо, чтобы ширина каждого канала была, с одной стороны, достаточно

узкой для минимизации искажения сигнала в его пределах, а с другой — достаточно широкой для обеспечения требуемой скорости передачи. Кроме того, для экономного использования всей полосы канала, разделяемого на подканалы, желательно как можно плотнее расположить частотные подканалы, но при этом избежать межканальной интерференции, чтобы обеспечить полную независимость каналов друг от друга. Частотные каналы, удовлетворяющие перечисленным требованиям, называются ортогональными. Несущие сигналы всех частотных подканалов (а точнее, функции, описывающие эти сигналы) ортогональны друг другу. Ортогональность несущих сигналов гарантирует частотную независимость каналов друг от друга, а следовательно, отсутствие межканальной интерференции.

Рассмотренный способ деления широкополосного канала на ортогональные частотные подканалы называется ортогональным частотным разделением с мультиплексированием (*Orthogonal Frequency Division Multiplexing*, OFDM). Эта технология была заимствована из стандарта 802.11a. Для ее реализации в передающих устройствах используется обратное быстрое преобразование Фурье (IFFT), переводящее предварительно мультиплексированный на N -каналов сигнал из временного представления в частотное.

Как уже отмечалось, одним из ключевых преимуществ метода OFDM является сочетание высокой скорости передачи с эффективным противостоянием многолучевому распространению. Если точнее, то сама по себе технология OFDM не предотвращает многолучевого распространения, но создает предпосылки для устранения эффекта межсимвольной интерференции. Дело в том, что неотъемлемой частью технологии OFDM является понятие «защитный интервал» (*Guard Interval*, GI) — это циклическое повторение окончания символа, пристраиваемое в начале символа. Охранный интервал является избыточной информацией и в этом смысле снижает полезную (информационную) скорость передачи. GI добавляется к отсылаемому символу в передатчике и отбрасывается при приеме символа в приемнике, но именно он защищает от возникновения межсимвольной интерференции.

Защитный интервал задает паузы между отдельными символами, и если его длительность превышает максимальное

время задержки сигнала в результате многолучевого распространения, то межсимвольной интерференции не возникает.

В стандарте IEEE 802.11g технология ортогонального частотного разделения каналов OFDM является обязательной при скоростях передачи 6, 12 и 24 Мбит/с и опциональной при скоростях передачи 9, 18, 36, 48 и 54 Мбит/с.

Технология гибридного кодирования ССК-OFDM

Кроме технологии ортогонального частотного разделения каналов OFDM, для опциональных скоростей (6, 9, 12, 18, 24, 36, 48 и 54 Мбит/с) может использоваться также технология гибридного кодирования ССК-OFDM.

Для того чтобы понять сущность этого термина, вспомним, что любой передаваемый пакет данных содержит заголовок/преамбулу со служебной информацией и поле данных. Когда речь идет о пакете в формате ССК, подразумевается, что заголовок и данные кадра передаются в формате ССК. Аналогично при применении технологии OFDM заголовок кадра и данные передаются посредством OFDM-кодирования. При использовании гибридной технологии ССК-OFDM заголовок кадра кодируется с помощью ССК-кодов, но сами данные кадра передаются с помощью многочастотного OFDM-кодирования. Таким образом, технология ССК-OFDM является своеобразным гибридом ССК и OFDM. Технология ССК-OFDM — не единственная гибридная технология: при применении пакетного кодирования РВСС заголовок кадра также передается посредством ССК-кодов и только данные кадра кодируются с помощью РВСС.

8.7.4. Двоичное пакетное сверточное кодирование РВСС

Кроме комплементарных ССК-последовательностей, в стандарте IEEE 802.11g для скоростей 5,5 и 11 Мбит/с опционально предусмотрено использование технологии двоичного пакетного сверточного кодирования РВСС.

В основе метода РВСС лежит так называемое сверточное кодирование со скоростью $1/2$. Сверточное кодирование подразумевает, что входной последовательности битов $\{X\}$ ставится в соответствие по определенному алгоритму выходная последовательность битов $\{Y\}$, причем значение каждого бита

выходной последовательности зависит от значения нескольких битов входной последовательности, то есть для расчета одного бита выходной последовательности учитывается некоторая предыстория входной последовательности.

Для того чтобы значение выходного бита зависело от значений нескольких битов входной последовательности, в сверточном кодере применяются запоминающие ячейки и логические элементы XOR. Кроме того, любой сверточный кодер подразумевает определенную степень избыточности, что повышает помехоустойчивость закодированной информации. Дибит $\{Y_0, Y_1\}$, формируемый в сверточном кодере, используется в дальнейшем в качестве передаваемого символа, но предварительно он подвергается фазовой модуляции (рис. 8.7).

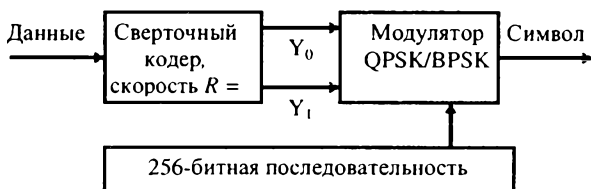


Рис. 8.7. Схема PBCC-модулятора

Если скорость передачи составляет 11 Мбит/с , то применяется квадратичная фазовая модуляция QPSK. В данном случае каждому из четырех возможных состояний дибита соответствует одна из четырех возможных фаз. При этом в каждом символе кодируется по одному входному биту, а скорость передачи битов соответствует скорости передачи символов. Если же скорость передачи составляет $5,5 \text{ Мбит/с}$, то используется двоичная фазовая модуляция BPSK. Каждый бит Y_0 или Y_1 , формируемый сверточным кодером, последовательно подвергается фазовой модуляции. Поскольку каждому входному биту в данном случае соответствуют два выходных символа, скорость передачи битов равна половине скорости передачи символов. Поэтому для скорости и $5,5$ и 11 Мбит/с символьная скорость составляет $11 \cdot 10^6$ символов в секунду.

При применении технологии двоичного пакетного сверточного кодирования PBCC необходимо, чтобы ширина спектра

передаваемого сигнала оставалась такой же, как и в случае ССК-последовательностей или кодов Баркера, т. е. была бы равна 22 МГц. В отличие от технологии DSSS, в этом случае используется не технология уширения спектра, а шумоподобные последовательности, однако и здесь предусмотрено уширение спектра до стандартных 22 МГц. Для этого применяют вариации возможных сигнальных созвездий QPSK и BPSK.

Напомним, что сигнальные созвездия представляют собой геометрическое отображение возможных выходных состояний сигнала. Для QPSK-модуляции имеются четыре дискретных состояния сигнала: 00, 01, 10 и 11. Каждому из этих дибитов соответствует одна из четырех возможных фаз несущего сигнала. Понятно, что расположение точек на сигнальном созвездии может быть различным. Именно этот принцип реализован в методе PBCC для уширения спектра выходного сигнала: используются по два сигнальных созвездия QPSK и BPSK (рис. 8.8).

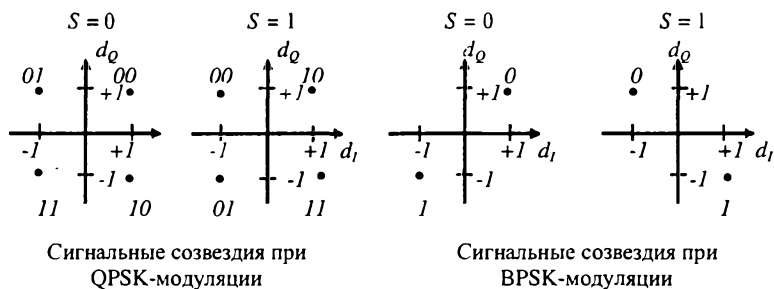


Рис. 8.8. Возможные типы сигнальных созвездий

Выбор между конкретным типом применяемого созвездия задается управляющим сигналом S , принимающим значение 0 или 1. Этот сигнал задается псевдослучайной последовательностью с периодом повторения 256 бит, которая формируется из 16-битной базовой последовательности 0011001110001011. Для того чтобы из данной базовой 16-битной последовательности получить 256-битную, используют циклический сдвиг трех первых символов одновременно. Так получают еще пятнадцать 16-битовых последовательностей, что в сумме дает одну 256-битную.

Как уже отмечалось, рассмотренный метод PBCC-кодирования опционально используется как в протоколе 802.11g, так и в протоколе 802.11b/b+ на скоростях 5,5 и 11 Мбит/с.

Для дальнейшего рассмотрения протокола 802.11b/b+ нам предстоит ознакомиться с еще одним типом кодирования — так называемым двоичным пакетным сверточным кодированием (*Packet Binary Convolutional Coding*, PBCC).

Идея сверточного кодирования заключается в следующем. Входящая последовательность информационных бит преобразуется в специальном сверточном кодере таким образом, чтобы каждому входному биту соответствовало более одного выходного, т. е. сверточный кодер добавляет определенную избыточную информацию к исходной последовательности. Если, к примеру, каждому входному биту соответствует два выходных, то говорят о сверточном кодировании со скоростью $R = 1/2$. Если же каждым двум входным битами соответствует три выходных, то скорость сверточного кодирования будет составлять уже $2/3$.

Любой сверточный кодер строится на основе нескольких последовательно связанных запоминающих ячеек и логических элементов, связывающих эти ячейки между собой. Количество запоминающих ячеек определяет количество возможных состояний кодера. Если, к примеру, в сверточном кодере используется шесть запоминающих ячеек, то в кодере хранится информация о шести предыдущих состояниях сигнала, а с учетом значения входящего бита получим, что в таком кодере используется семь бит входной последовательности. Такой сверточный кодер называется кодером на семь состояний ($K = 7$).

Выходные биты, формируемые в сверточном кодере, определяются значениями входного бита и битами, хранимыми в запоминающих ячейках, т.е. значение каждого формируемого выходного бита зависит не только от входящего информационного бита, но и от нескольких предыдущих битов.

Дибит, формируемый в сверточном кодере, используется в дальнейшем в качестве передаваемого символа, но предварительно этот дибит подвергается фазовой модуляции. Причем в зависимости от скорости передачи возможна двоичная, квадратурная или даже восьмипозиционная фазовая модуляция.

Метод пакетного сверточного кодирования опционально предусмотрен как альтернативный метод кодирования в протоколе 802.11b на скоростях передачи 5,5 и 11 Мбит/с. Кроме того, именно данный режим кодирования лег в основу протокола 802.11b+ — расширения протокола 802.11b. Собственно, протокола 802.11b+ как такового официально не существует, однако данное расширение поддержано многими производителями беспроводных устройств. В протоколе 802.11b+ предусматривается еще одна скорость передачи данных — 22 Мбит/с с использованием технологии РВСС.

При скорости передачи 5,5 Мбит/с для модуляции дибита, формируемого сверточным кодером, используется двоичная фазовая модуляция, а при скорости 11 Мбит/с — квадратурная фазовая модуляция. При этом для скорости 11 Мбит/с в каждом символе кодируется по одному входному биту и скорость передачи бит соответствует скорости передачи символов, а при скорости 5,5 Мбит/с скорость передачи битов равна половине скорости передачи символов (поскольку каждому входному биту в данном случае соответствует два выходных символа). Поэтому и для скорости 5,5 Мбит/с, и для скорости 11 Мбит/с символьная скорость составляет $11 \cdot 10^6$ символов в секунду.

Для скорости 22 Мбит/с по сравнению с уже рассмотренной нами схемой РВСС передача данных имеет две особенности. Прежде всего, используется фазовая 8-позиционная фазовая модуляция (8-PSK), то есть фаза сигнала может принимать восемь различных значений, что позволяет в одном символе кодировать уже 3 бита. Кроме того, в схему кроме сверточного кодера добавлен пунктурный кодер (*Puncture*). Смысл такого решения довольно прост: избыточность сверточного кодера, равная 2 (на каждый входной бит приходится два выходных), достаточно высока и при определенных условиях помеховой обстановки является излишней, поэтому можно уменьшить избыточность, чтобы, к примеру, каждым двум входным битам соответствовало три выходных.

Для этого можно, конечно, разработать соответствующий сверточный кодер, но лучше добавить в схему специальный пунктурный кодер, который будет просто уничтожать лишние биты.

Допустим, что пунктурный кодер удаляет один бит из каждых четырех входных битов. Тогда каждым четверем входящим битам будет соответствовать три выходящих. Скорость такого кодера составляет 4:3.

Если же такой кодер используется в паре со сверточным кодером со скоростью $1/2$, то общая скорость кодирования составит уже $2/3$, то есть каждым двум входным битам будет соответствовать три выходных.

Разобравшись с принципом работы пунктурного кодера, вернемся к рассмотрению кодирования RVCC на скорости 22 Мбит/с в протоколе 802.11b+.

В сверточный кодер ($K = 7$, $R = 1/2$) данные поступают со скоростью 22 Мбит/с . После добавления избыточности в сверточном кодере биты со скоростью потока 44 Мбит/с поступают в пунктурный кодер 4:3, в котором избыточность уменьшается так, чтобы на каждые четыре входных бита приходилось три выходных. Следовательно, после пунктурного кодера скорость потока составит уже 33 Мбит/с (не информационная, а общая скорость с учетом добавленных избыточных битов). Полученная в результате последовательность направляется в фазовый модулятор 8-PSK, где каждые три бита упаковываются в один символ. При этом скорость передачи составит $11 \cdot 10^6$ символов в секунду, а информационная скорость — 22 Мбит/с (рис. 8.9).

В заключение обсуждения протокола 802.11b/b+ приведем таблицу соответствия между скоростями передачи и типом кодирования (табл. 8.3).

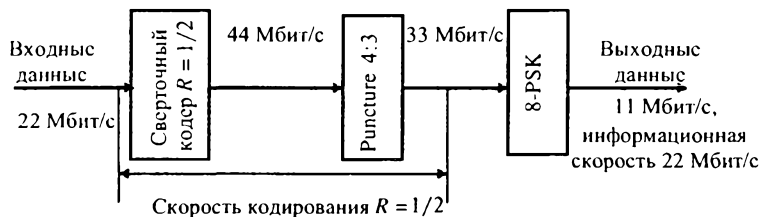


Рис. 8.9. Реализация скорости 22 Мбит/с в протоколе 802.11g.

**Соотношение между скоростями передачи
и типом кодирования в стандарте 802.11b/b+**

Скорость передачи, Мбит/с		Метод кодирования	Модуляция	Скорость сверточного кодирования	Символьная скорость, 10^6 сим-вол/с	Количество бит в одном символе
1	обязательно	Код Баркера	DBPSK		1	1
2	обязательно	Код Баркера	DQPSK		1	2
5,5	обязательно	ССК			1,375	2
	опционально	PBCC	DBPSK	1/2	11	0,5
11	обязательно	ССК	DQPSK		1,375	8
	опционально	PBCC	DQPSK	1/2	11	1
22*	обязательно	PBCC	DQPSK	3/4	11	2

* *Примечание.* Скорость 22 Мбит/с относится только к протоколу 802.11b+.

9.1. Основные виды импульсной модуляции

Успехи в освоении СВЧ и развитие импульсной техники способствовали созданию новых видов управления колебаниями, а именно импульсной модуляции. Термин «импульсная модуляция» означает, по существу, выполнение двойной модуляции: первичной модуляции некоторой вспомогательной последовательности передаваемых сообщений и вторичной модуляции радиочастотного колебания (несущей) напряжением, полученным от первичной модуляции.

Структурная схема передающей части системы импульсной радиосвязи приведена на рис. 9.1.

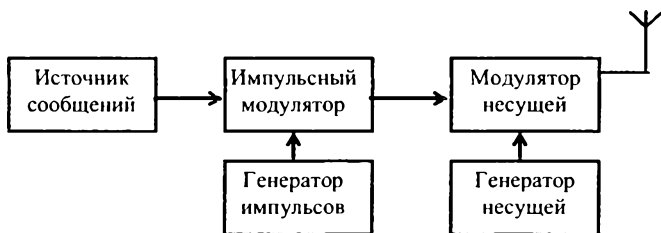


Рис. 9.1. Структурная схема передающей части системы импульсной радиосвязи

Вторичная модуляция, представляющая собой обычную амплитудную модуляцию несущего колебания импульсами, не требует дополнительного рассмотрения. Как и при простой амплитудной модуляции, спектр колебания в этом случае будет состоять из несущей частоты и двух симметричных полос боковых частот, каждая из которых определяется спектром импульсной последовательности.

В зависимости от выбора меняющегося параметра импульсная модуляция подразделяется на следующие виды:

- *амплитудно-импульсная модуляция (АИМ)*, когда по закону управляющего сигнала меняется приращение амплитуды импульсов;
- *модуляция по длительности импульсов (ДИМ)*, называемая еще модуляцией по ширине импульса (ШИМ), когда по закону управляющего сигнала меняется длительность импульсов;
- *временная импульсная модуляция (ВИМ)*, когда по закону управляющего сигнала происходит смещение импульсов по временной оси. При этом ВИМ можно разделить на частотно-импульсную (ЧИМ) и фазоимпульсную (ФИМ), в зависимости от того, изменяется ли положение импульса на оси времени пропорционально частоте или фазе передаваемого сигнала, как при ЧИМ, так и при ФИМ. Наиболее широко распространены высокочастотная импульсная модуляция по частоте (В-ЧИМ) и по фазе (В-ФИМ).

На рис. 9.2 показаны реализации модулирующего линейно изменяющегося сигнала $s(t)$ и результаты его запечатления в изменяющихся параметрах высокочастотного колебания. При этом возможно осуществить различные виды импульсной модуляции. Наиболее широко распространены высокочастотная модуляция по частоте и по фазе. В этих случаях ни амплитуда, ни длительность, ни тактовая частота следования импульсов не меняются. Меняются в процессе модуляции только частота или начальная фаза высокочастотного заполнения по закону изменяющегося сообщения.

Теоретической базой всех перечисленных видов импульсной модуляции является теорема В.А. Котельникова, на основании которой выбирается тактовая частота следования импульсов.

Амплитудный спектр последовательности прямоугольных импульсов мы уже не раз строили. Положение спектральных компонент на оси частот определяется частотой повторения импульсов, а амплитуды этих компонент зависят от длительности импульсов τ . Интервал частот от 0 до $1/\tau$ часто называют главным лепестком. Число гармоник частоты повторения, расположенных на этом интервале, равно скважности

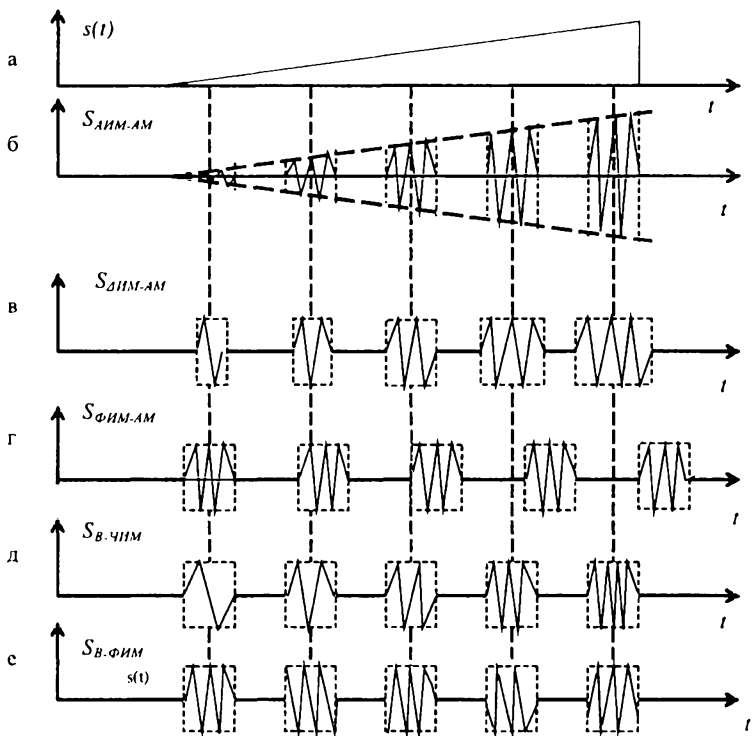


Рис. 9.2. Основные виды импульсной модуляции

импульсов $q = T/\tau$. В главном лепестке сосредоточена большая часть мощности периодической последовательности импульсов, поэтому эквивалентной шириной спектра периодической импульсной последовательности считают ширину главного лепестка $\Delta f = 1/\tau$.

Спектр сигнала после первичной импульсной модуляции лежит в низкочастотной области. Поэтому для радиопередачи производится повторная модуляция: модулированными видеоимпульсами модулируется высокочастотное гармоническое колебание. В результате антенна передатчика излучает в окружающее пространство последовательность модулированных радиоимпульсов.

9.1.1. Амплитудно-импульсная модуляция

АИМ — простейший вид импульсной модуляции, поскольку обладает наибольшей простотой, как при получении модулированных колебаний, так и при их детектировании. Однако этому виду модуляции присущ недостаток — он не допускает амплитудного ограничения. В связи с этим АИМ мало пригодна в тех случаях, когда одним из главных требований является повышение помехоустойчивости радиоканала. АИМ находит применение в качестве промежуточного преобразования при осуществлении, а также при детектировании более сложных видов модуляции (ДИМ и ВИМ).

Определим спектр АИМ-колебания при модуляции прямоугольных видеоимпульсов напряжением $s(t)$. На рис. 9.3, а изображена исходная периодическая последовательность $s_u(t)$ прямоугольных импульсов, имеющих длительность τ и период следования T , а на рис. 9.3, б — та же последовательность, промодулированная сигналом $s(t)$, близким к гармоническому колебанию.

Тонально модулированную последовательность видеоимпульсов в соответствии с законом модуляции $s(t) = U_m \cos \omega t$ можно записать (при $U_m = 1$):

$$s(t) = (1 + m \cos \Omega t) \left[\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos(n\omega_u t - \varphi_n) \right] =$$

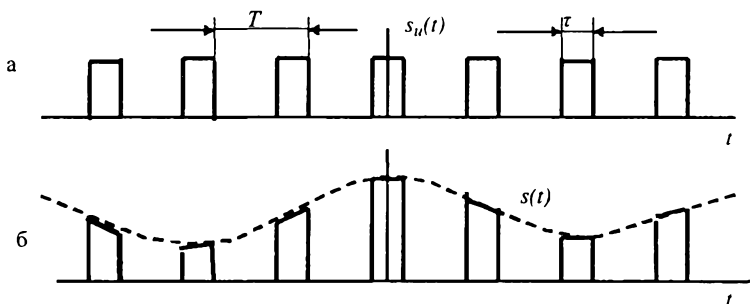


Рис. 9.3. Преобразование электрического управляющего сигнала $s_u(t)$ (а) в управляющий сигнал (б) для получения АИМ-сигнала

$$\begin{aligned}
&= \left[\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos(n\omega_u t - \varphi_n) \right] + \left[\frac{m c_0}{2} \cos \Omega t + \right. \\
&+ m \cos \Omega t \sum_{n=1}^{\infty} c_n \cos(n\omega_u t - \varphi_n) \left. \right] = \left[\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \cos(n\omega_u t - \varphi_n) \right] + \\
&+ \left[\frac{m c_0}{2} \cos \Omega t + \frac{m}{2} \sum_{n=1}^{\infty} c_n \cos((n\omega_u + \Omega)t - \varphi_n) + \right. \\
&+ \left. \frac{m}{2} \sum_{n=1}^{\infty} c_n \cos((n\omega_u - \Omega)t - \varphi_n) \right].
\end{aligned}$$

На основании этого выражения построим спектр модулированной импульсной последовательности (рис. 9.4).

Можно заметить, что при $m = 0$ в правой части анализируемого соотношения остаются только первые два слагаемых, которые образуют спектр исходной периодической последовательности прямоугольных импульсов. Благодаря модуляции около каждой компоненты (включая нулевую составляющую $c_0/2$), играющей роль несущей, появляются боковые составляющие с амплитудами $m c_n/2$ и частотами $n\omega_u \pm \Omega$, в том числе низкочастотная составляющая с амплитудой $m c_0/2$ и частотой модуляции Ω .

В случае модуляции спектром частот при АИМ около каждой гармонической составляющей спектра $n\omega_u$ появляются боковые составляющие с частотами $n\omega_u + \Omega$ и $n\omega_u - \Omega$. Если АИМ сигнал получен в результате модуляции спектром частот от Ω_{min} до Ω_{max} , то с обеих сторон каждой из составляющих спектра $n\omega_u$ появятся нижняя и верхняя боковые полосы частот шириной от Ω_{min} до Ω_{max} . При этом, чтобы боковые полосы при каждой следующей и предыдущей частотах $n\omega_u$ не перекрывались, вследствие чего появились бы искажения при передаче сигнала, необходимо, чтобы частота следования импульсов $f_u = \omega_u/2\pi = 1/T$ была не менее чем в два раза больше максимальной частоты спектра управляющего сигнала $F_{max} = 2\pi/\Omega_{max}$.

Наконец, при модуляции непериодическим сигналом, обладающим спектром $S(\omega)$, вместо отдельных пар боковых составляющих наблюдаются сплошные боковые полосы, смещенные на частоты $n\omega_u$, т. е. формируется спектр вида,

полученный для дискретизированного по времени непрерывного сигнала.

Чтобы уменьшить ошибки дискретизации исходного колебания, возникающие при использовании АИМ, тактовую частоту $\omega_u = 2\pi/T$ импульсов выбирают из условия $\omega_u > (25)\omega_g$, где $\omega_g = 2\pi f_g$ — максимальная частота в спектре $S(\omega)$ модулирующего сигнала. Этому условию соответствует интервал дискретизации $T = 2\pi/\omega_u$, меньший, чем интервал дискретизации, который определяется теоремой Котельникова: $\Delta t < 1/(25)f_g$. Поскольку передаваемое сообщение заложено в низкочастотную часть спектра АИМ, прилегающую к нулевой частоте, то для восстановления сообщения на приемном конце канала связи после амплитудного детектора достаточно использовать ФНЧ.

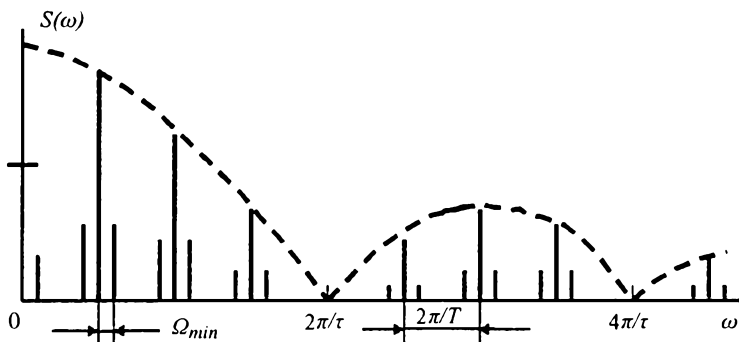


Рис. 9.4. Спектр модулированной импульсной последовательности

Приходим к выводу, что ширина спектра АИМ-сигнала не зависит от частоты модулирующего сигнала, поскольку определяется только длительностью импульса. Для прямоугольных импульсов несущей последовательности ширину спектра АИМ-сигнала можно принять равной $\Delta\omega_{\text{сн АИМ}} = 2\pi/\tau$.

9.1.2. Широтно-импульсная модуляция (ШИМ)

При широтно-импульсной модуляции модулируется ширина (длительность) импульсов, а остальные параметры остаются

неизменными. Различают одностороннюю и двустороннюю ШИМ. При односторонней ШИМ положение одного из фронтов импульса (например, переднего) остается неизменным относительно тактовой точки, а положение другого фронта (заднего) меняется в соответствии с модулирующим сообщением. Максимальный временной сдвиг модулируемого фронта $\Delta\tau_m$ относительно его начального значения называется *девиацией фронта импульса*. При ШИМ всегда должно выполняться условие $\Delta\tau_m/\tau < 1$.

Благодаря модуляции около каждой компоненты (включая нулевую составляющую $c_0/2$), играющей роль несущей, появляются боковые составляющие с амплитудами $mc_n/2$ и частотами $n\omega_u \pm \Omega$, в том числе низкочастотная составляющая с амплитудой $mc_0/2$ и частотой модуляции Ω .

Спектр ШИМ отличается от спектра АИМ тем, что при ШИМ гармоники частоты повторения оказываются модулированными по амплитуде и по фазе. При малых величинах индекса модуляции, что практически всегда имеет место, различие между спектрами АИМ и ШИМ незначительно.

9.1.3. Временная импульсная модуляция (ВИМ)

При ВИМ модулируется положение импульсов t_k относительно тактовых точек t_{0k} немодулированной последовательности. Длительность импульсов и их амплитуды при модуляции остаются неизменными. Имеем

$$t_k = t_{0k} + \Delta\tau_m x(t_k), \quad \tau = const, \quad U_0 = const.$$

Максимальный временной сдвиг импульсов $\Delta\tau_m$ относительно тактовой точки называется девиацией импульса. При ВИМ $\Delta\tau/\tau > 1$ каждому импульсу длительности τ нужно отвести временной интервал не менее $\tau + 2\Delta\tau_m$.

Анализ показывает, что низкочастотная компонента ВИМ описывается выражением

$$u_{ВИМ}(t) = \frac{U_0\tau}{T_p} + \frac{U_0\tau}{T_p} \Delta\tau_m x'(t) + \dots$$

Гармоники частоты повторения имеют амплитудную и фазовую модуляцию. Если сообщение $x(t) = \sin\Omega_M t$, то $x'(t) = \Omega_M \cos\Omega_M t$ и

$$u_{ВИМ}(t) = \frac{U_0 \tau}{T_p} + \frac{U_0 \tau}{T_p} \Delta \tau_m \Omega_M \cos \Omega_M t + \dots$$

Наиболее важное различие между спектрами ВИМ и ШИМ заключается в величине амплитуд составляющих на частоте F_M . При ШИМ имеем

$$U_{F_M} = \frac{U_0 \tau_m}{T_p},$$

а при ВИМ

$$U_{F_M} = \frac{U_0 \Delta \tau_m}{T_p} \tau \Omega_M.$$

Если число каналов велико ($N \gg 1$), то можно показать, что $\tau \Omega_M \ll 1$. Отсюда следует, что амплитуда составляющей на частоте F_M при ВИМ много меньше, чем при ШИМ. Кроме того, при ВИМ эта амплитуда зависит от частоты модулирующего напряжения F_M .

Анализируя спектры различных импульсно-модулированных сигналов, приходим к выводу, что ширина спектра этих сигналов определяется прежде всего длительностью последовательности импульсов, которая подвергается амплитудной, частотной или фазовой модуляции.

9.2. Основные виды манипуляции сигналов

В цифровой радиосвязи в качестве переносчика первичного сигнала можно использовать периодическую последовательность видео- или радиоимпульсов, характеризующуюся амплитудой, длительностью, частотой следования, положением импульсов во времени относительно тактовых импульсов. Изменением одного из перечисленных параметров в соответствии с изменением модулирующей функции можно получать различные виды импульсной модуляции: АИМ, ДИМ (или ШИМ), ЧИМ, ФИМ и т. д.

Спектр сигнала после первичной импульсной модуляции лежит в низкочастотной области, поэтому для радиопередачи производится повторная модуляция: модулированными видеоимпульсами модулируется гармоническое колебание высокой частоты. В результате передающее устройство излучает в

окружающее пространство последовательность модулированных радиопульсов.

В цифровой связи передается не аналоговый сигнал, а последовательность целых чисел $n_0, n_1, n_2, \dots$, которые могут принимать значения из некоторого конечного множества. Эти числа, называемые символами, поступают от источника информации с периодом T , а частота, соответствующая этому периоду, называется символьной скоростью. На практике часто используется двоичная последовательность из символов, когда каждое из n_i может принимать одно из двух значений — 0 или 1. Последовательность передаваемых импульсов является дискретным сигналом. Так как символы принимают значения из конечного множества, то этот сигнал является и квантованным, т. е. его можно назвать цифровым сигналом.

Способ модуляции, когда параметры несущего колебания меняются скачкообразно, называется манипуляцией. В зависимости от того, какие именно параметры несущего колебания изменяются в соответствии с модулирующим сигналом, различают АМн-, ФМн-, ЧМн-, квадратурную (КАМн) манипуляцию. Кроме того, при передаче цифровой информации можно использовать несущее колебание, отличное по форме от гармонического. Так, при использовании в качестве несущего колебания последовательности прямоугольных импульсов возможны АИМ, ШИМ и ВИМ и другие виды модуляции.

9.2.1. Частотная манипуляция

При частотной манипуляции каждому возможному значению передаваемого символа соответствует своя частота. В течение каждого символьного интервала передается гармоническое колебание с частотой, соответствующей текущему символу.

Сформируем двухпозиционный (бинарный) ЧМн-сигнал, в котором возможным значениям символов 0 и 1 соответствуют частоты f_1 и f_2 . Так как каждая посылка формируется независимо, сформированный сигнал содержит скачки фазы при переходе от одной посылки к другой. Чтобы спектр сигнала был компактнее, используется частотная модуляция с непрерывной фазовой функцией (*continuous phase frequency shift*

keying, CPFSK). При этом формируется линейно меняющаяся полная фаза колебания, а передаваемые символы управляют скоростью ее изменения. При этом передаваемые символы переключают значения мгновенной частоты; эта частота интегрируется, в результате чего получается непрерывная функция, описывающая поведение фазового угла колебания, косинус которого также получается непрерывной функцией.

Прием ЧМн-сигнала, как правило, осуществляется корреляционным методом. Сущность его состоит в вычислении функции взаимной корреляции между принятым сигналом и колебанием, играющим роль образцового, которое может генерироваться в приемнике. В качестве выходного символа выбирается тот, частота которого оказывается максимально коррелирована со входным сигналом.

Корреляционный прием может быть когерентным или некогерентным. Когерентный алгоритм может использоваться, если известна начальная фаза колебания. Опорными сигналами в этом случае служат вещественные гармонические сигналы, обладающие нужными частотами и начальными фазами. Для принятия решения о принятом символе сравниваются вещественные результаты вычисления корреляционных сумм.

На практике начальные фазы, как правило, неизвестны. Тогда используют некогерентный корреляционный прием, при котором опорные сигналы представляют гармонические колебания нужной частоты. Для принятия решения о принятом символе сравниваются модули результатов вычисления корреляционных сумм. Модуль комплексной взаимной функции корреляции не зависит от начальной фазы сигналов, но помехоустойчивость такого алгоритма хуже, чем в случае когерентного приема.

Для повышения помехоустойчивости ЧМн желательно, чтобы посылки, соответствующие разным символам, были некоррелированы. Считая начальные фазы посылок нулевыми, ЧМн-сигналы для символов 0 и 1 можно записать:

$$S_0(t) = A \cos \omega_0 t, \quad 0 \leq t \leq T,$$

$$S_1(t) = A \cos \omega_1 t, \quad 0 \leq t \leq T.$$

Из выражения для взаимной корреляционной функции при нулевом временном сдвиге получаем:

$$\begin{aligned}
 R_{01}(0) &= \int_0^T S_0(t) S_1(t) dt = A^2 \int_0^T \cos \omega_0 t \cos \omega_1 t dt = \\
 &= \frac{A^2}{2} \int_0^T \cos(\omega_1 + \omega_0) t dt + \frac{A^2}{2} \int_0^T \cos(\omega_1 - \omega_0) t dt = \\
 &= \frac{A^2 \sin(\omega_1 + \omega_0) T}{2(\omega_1 + \omega_0)} + \frac{A^2 \sin(\omega_1 - \omega_0) T}{2(\omega_1 - \omega_0)}.
 \end{aligned}$$

Если $(\omega_1 + \omega_0)T \gg 1$, то первое слагаемое значительно меньше второго и им можно пренебречь. Тогда

$$R_{01}(0) = \frac{A^2 \sin(\omega_1 - \omega_0) T}{2(\omega_1 - \omega_0)}.$$

Это значение функции взаимной корреляции равно нулю при $(\omega_1 - \omega_0)T = \pi k$, где k — целое число, не равное нулю.

Таким образом, минимальное расстояние между частотами манипуляции, при котором посылки, соответствующие разным символам, оказываются некоррелированными, составляет $\Delta\omega_{\min} = \pi/T$ или $\Delta f_{\min} = 1/2T = f_T/2$, где f_T — символьная скорость.

Двухпозиционная (бинарная) ЧМн, частота которой выбрана согласно этому условию, называется минимальной частотной манипуляцией (*minimum shift keying, MSK*).

В бинарной системе передачи термин «символ» синонимичен биту, поэтому символьная скорость f_s эквивалентна битовой скорости f_b , т. е. $f_s = f_b$ только для двоичной передачи. В других случаях символ может содержать несколько битов. В системе QPSK $f_s = f_b/2$, $T_s = 2T_b$, т. е. один символ содержит 2 бита информации.

В системе GSM для модуляции радиосигнала используется спектрально эффективная гауссовская частотная манипуляция с минимальным частотным сдвигом (GMSK). Последовательность информационных битов до модулятора проходит через фильтр НЧ с гауссовской АЧХ, что дает значительное уменьшение ширины полосы частот излучаемого сигнала.

GMSK-сигнал формируется следующим образом. На интервале, соответствующем 1 биту, фаза несущей изменяется на 90° . Это наименьшее изменение фазы, которое может быть обнаружено при данном типе манипуляции. Выходной сиг-

нал с непрерывным изменением фазы аналогичен сигналу, полученному в результате частотной модуляции с дискретным изменением частоты.

В системе GSM используется модуляция с величиной нормированной полосы $BT = 0,3$, где B — ширина полосы фильтра на уровне минус 3 дБ, T — длительность передачи 1 бита.

Основой формирования GMSK-сигнала является квадратурный (I/Q) модулятор с синфазной (*Inphase*) и квадратурной (*Quadrature*) компонентами, который состоит из двух умножителей и одного сумматора. Применение модуляции GMSK позволяет:

- использовать передающие устройства с усилителем мощности, работающем в классе C;
- получить низкий уровень внеполосного излучения;
- обладает хорошей помехозащищенностью.

В системе GSM для максимальной защиты от несанкционированного подключения применяется специальный модуль подлинности абонента — SIM-карта (*Subscriber Identification Module*) — модуль идентификации абонента. Это карточка со встроенной микросхемой, в которой хранится специальная информация о конкретном абоненте. SIM-карта выдается оператором сотовой связи при подключении телефона и может быть использована с любой моделью мобильного аппарата стандарта GSM.

9.2.2. Фазовая манипуляция

При ФМн скачкообразно изменяется фаза несущего колебания. На практике ФМн используется при небольшом числе возможных значений начальной фазы — как правило, 2, 4 или 8. Поскольку при приеме трудно измерить абсолютное значение начальной фазы, то значительно проще определить относительный фазовый сдвиг между двумя соседними символами. Поэтому обычно используется фазоразностная манипуляция (*Differential Phase Shift Keying* — DPSK).

9.2.3. Квадратурная манипуляция

При квадратурной манипуляции (*Quadrature Amplitude Shift Keying*, QASK) каждому из возможных значений дискретного символа C_i ставится в соответствие пара величин: амплитуды

синфазной и квадратурной составляющих либо амплитуда и начальная фаза несущего колебания. Параметры аналогового колебания, соответствующего дискретному символу, удобно представлять в алгебраической или комплексной форме. Совокупность этих комплексных чисел для всех возможных значений дискретного символа называется *созвездием*. На практике используются созвездия, содержащие от 4 до нескольких тысяч точек. Модемы, предназначенные для передачи данных по телефонным линиям связи, используют созвездия:

- А. 16 точек (протокол v.32, скорость 9600 бит/с),
- Б. 128 точек (протокол v.32bis, скорость 14400 бит/с),
- В. 640 точек (протокол v.34, скорость 28800 бит/с).

График сигнала с квадратурной манипуляцией не очень нагляден из-за смешанного характера модуляции (амплитудной и фазовой). Изменения амплитуды и начальной фазы при переходе от символа к символу могут быть небольшими и потому плохо заметными на графике.

9.2.4. Квадратурная амплитудная модуляция

Ранее мы отметили, что существует три основных вида модуляции: амплитудная, частотная и фазовая, в зависимости от того, в каком параметре высокочастотного колебания запечатлевается информация о передаваемом сообщении. Однако если изменять амплитуду в соответствии с сигналом $s_1(t)$, а фазу — в соответствии с законом $s_2(t)$, то можно запечатлеть информацию об этих двух сигналах и передавать одновременно информацию о двух сообщениях $s_1(t)$ и $s_2(t)$ на одной несущей частоте ω_0 с помощью сигнала

$$S_{1,2}(t) = U_m(t)\cos(\omega_0 t + \varphi_2(t)), \quad (9.1)$$

где в изменении амплитуды $U_m(t)$ содержится информация о сообщении $s_1(t)$, а в изменении начальной фазы $\varphi_2(t)$ несущего колебания — информация о сообщении $s_2(t)$. Такую модуляцию можно назвать амплитудно-фазовой. Сигналы $s_1(t)$ и $s_2(t)$ модулируют разные параметры высокочастотного колебания, поэтому нужно принимать специальные меры, чтобы уравновесить их воздействие на несущее колебание. Но если (9.1) переписать иначе, а именно как

$$S_{1,2}(t) = U_m(t)\cos\varphi_2(t) \cdot \cos\omega_0 t - U_m(t)\sin\varphi_2(t) \cdot \sin\omega_0 t = \\ = a(t)\cos\omega_0 t + b(t)\sin\omega_0 t,$$

где $a(t) = U_m(t)\cos\varphi_2(t)$, $b(t) = -U_m(t)\sin\varphi_2(t)$, то теперь сигнал $S_{1,2}(t)$ можно рассматривать как сумму двух АМ-колебаний с различающимися на 90° начальными фазами одной несущей частоты ω_0 .

Процесс модуляции сигналами $s_1(t)$ и $s_2(t)$ двух колебаний одной несущей частоты с различающимися на 90° начальными фазами называется квадратурной амплитудной модуляцией (КАМ) (QAM — *Quadrature Amplitude Modulation*). Модулирующие сигналы $s_1(t)$ и $s_2(t)$ воздействуют при этом только на амплитуду несущих колебаний $\cos\omega_0 t$ и $\sin\omega_0 t$ в соответствии с величиной своей интенсивности.

Итак, КАМ представляет собой сумму двух АМ-колебаний, полученных в результате модуляции сигналами $s_1(t)$ и $s_2(t)$ двух несущих одной частоты, но сдвинутых относительно друг друга на 90° . Спектр сигнала с КАМ можно представить состоящим из спектра одного АМ-колебания с несущей частотой ω_0 (несущее колебание частоты ω_0 и две боковые полосы, сформированные под действием сигнала $s_1(t)$) и спектра другого АМ-колебания с этой же несущей частотой ω_0 , но сдвинутой на 90° по фазе относительно несущей частоты первого АМ-колебания (несущее колебание частоты ω_0 и две боковые полосы, сформированные под действием сигнала $s_2(t)$). Если спектры модулированных сигналов занимают примерно одну и ту же полосу частот, то их амплитудные спектры будут перекрываться. При этом фазовый спектр синусной несущей сдвинут на 90° относительно косинусного фазового спектра. Это обстоятельство позволяет при демодуляции сигнала с КАМ выделить из него сигналы $s_1(t)$ и $s_2(t)$.

Спектр амплитуд такого сигнала состоит из несущей ω_0 и двух пар боковых: $\omega_0 \pm \Omega_1$ и $\omega_0 \pm \Omega_2$. Поскольку частота Ω_2 в два раза превышает частоту Ω_1 , то ширина спектра такого сигнала с КАМ равна $2\Omega_2$.

Удобным представлением квадратурно-модулированного колебания является векторная диаграмма (рис. 9.5).

КАМ модуляция устраняет бесполезную с точки зрения передачи информации передачу несущей частоты. Улучшить используемый сигнал для передачи информации можно в

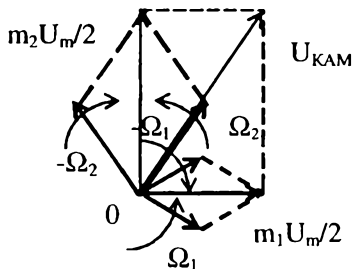


Рис. 9.5. Векторная диаграмма колебания с КАМ с подавленной несущей частотой

результате подавления несущего колебания путем использования балансных модуляторов. Тогда получается результат, представленный на рис. 9.5, из которого видно, что в этом случае с помощью КАМ передаются два сигнала с эффективностью, равной 100%.

Сигнал sI (рис. 9.6, а) модулирует косинусоидальное колебание, а сигнал sQ (рис. 9.6, б) — синусоидальное. Затем два модулированных колебания (рис. 9.6, в и г) складываются по правилу сложения векторов, образуя квадратурно-модулированное колебание (рис. 9.6, д). Мы видим, что моду-

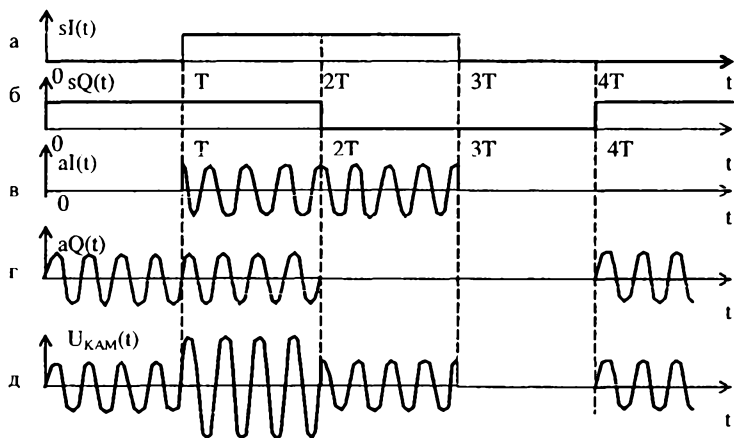


Рис. 9.6. Временные диаграммы сигнала с КАМ при модуляции спектром частот

ления является сложной: в моменты T и $2T$ наблюдаются как скачки амплитуды, так и скачки фазы модулированного колебания.

Отметим, что квадратурная модуляция используется в системах аналогового цветного телевидения NTSC и PAL при модуляции цветовой поднесущей двумя цветоразностными сигналами.

Обозначая вещественную и мнимую оси как I и Q соответственно, можно перейти к диаграмме, на которой колебания вместо векторов представлены точками. Такой способ является очень удобным при представлении более сложных модулированных колебаний. Например, колебание с КАМ, представленное в векторной форме на рис. 9.5, можно изобразить также и в форме, показанной на рис. 9.7. Как видно из рис. 9.7, пространство комплексной

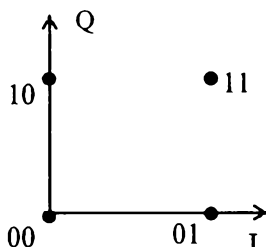


Рис. 9.7. Представление сигнала с КАМ в виде созвездия на комплексной плоскости

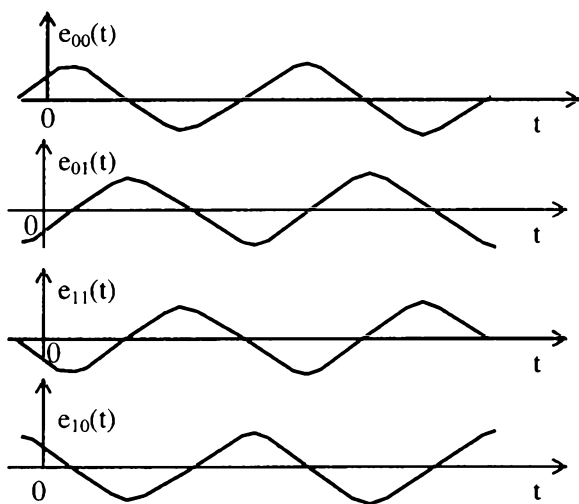


Рис. 9.8. Временные диаграммы сигнала с четырехпозиционной фазовой манипуляцией

плоскости используется не очень эффективно. На рис. 9.8 показаны временные диаграммы квадратурно-модулированного колебания, вектор которого при модуляции двух квадратурных компонент занимает также 4 точки, но уже в 4 квадрантах, что свидетельствует о повышении помехоустойчивости такого колебания.

Вид сигнального созвездия этого колебания показан на рис. 9.9. Поскольку амплитуды всех сигналов одинаковы, то такую модуляцию можно назвать и четырехпозиционной фазовой модуляцией или, точнее, манипуляцией (*Quadrature Phase Shift Keying*, QPSK).

Развивая этот подход, можно интерпретировать и сигналы других систем многопозиционной квадратурной модуляции. Например, на рис. 9.10 приведено векторное представление 16-позиционного квадратурно-модулированного сигнала. Увеличение числа позиций, или уровней, позволяет увеличить удельную скорость модуляции, но за счет увеличения мощности излучаемого колебания.

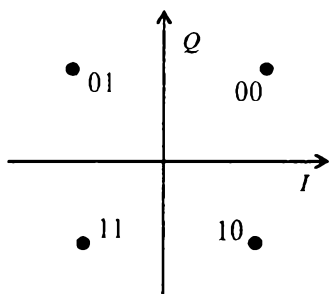


Рис. 9.9. Сигнальное созвездие сигнала с четырехпозиционной фазовой манипуляцией

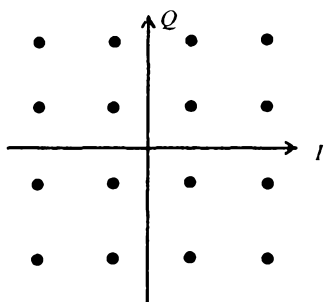


Рис. 9.10. Сигнальное созвездие 16-позиционного КАМ-сигнала

Так, при 16 позициях за время передачи одного символа может быть передано 4-разрядное двоичное число, что означает четырехкратное увеличение скорости передачи.

9.3. Ортогональная модуляция с частотным разделением

Новым способом модуляции, предложенным сравнительно недавно, является OFDM (*Orthogonal Frequency Division Multiplexing*). Подобно квадратурной модуляции, этот способ использует ортогональные несущие, но в отличие от квадратурной модуляции частоты этих несущих не являются одинаковыми, они расположены в некотором диапазоне, отведенном для передачи данных путем модуляции. Частоты несущих соответствуют уравнению

$$e_n(t) = \cos[2\pi (f_0 + n / T_s)t],$$

где f_0 — начало интервала, в котором производится частотное разделение, n — номер несущей, находящейся в диапазоне частот от 0 до N , т. е. всего несущих N , T_s — длительность интервала передачи одного символа.

Поскольку среднее произведение этих колебаний равно нулю, то это означает, как и в случае квадратурной модуляции, возможность их разделения на приеме даже при частичном перекрытии их боковых полос. Принцип ортогональной модуляции с частотным разделением иллюстрируется рис. 9.11.

Сигнал исходных данных разделяется на N потоков, формируясь в параллельную форму. Каждый из параллельных сигналов поступает на свой модулятор, в котором одна из ортогональных несущих подвергается модуляции. После сложения

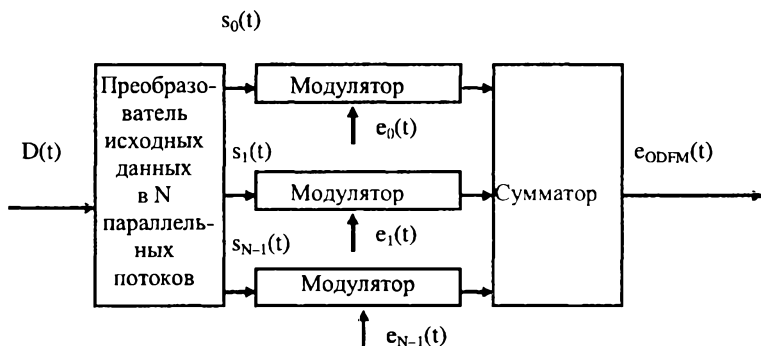


Рис. 9.11. Структурная схема получения сигнала с OFDM

модулированных ортогональных колебаний формируется сигнал OFDM.

Преимущества сигнала с OFDM проявляются при большом числе несущих. Например, если несущих 8000, то скорость данных, модулирующих каждую несущую, уменьшается в 8000 раз.

Следовательно, в 8000 раз увеличивается длительность передачи одного символа (структура импульса, конечно, при этом усложняется). При этом вставляются еще защитные интервалы между соседними символами, что позволяет бороться с межсимвольными искажениями.

Другим преимуществом этого метода модуляции является возможность борьбы с многолучевым распространением. Но реализовать схему, показанную на рис. 9.11, при таком числе несущих ранее было практически невозможно. Однако разработки интегральных схем, реализующих алгоритмы преобразования Фурье, позволили решить эту проблему (рис. 9.12). Перемножение некоторых коэффициентов на гармонические колебания разных частот, удовлетворяющих приведенным выше условиям, и суммирование полученных произведений представляет собой не что иное, как обратное преобразование Фурье, коэффициентами которого являются распараллеленные потоки данных. Модуляция реализуется как обратное преобразование Фурье, а демодуляция — как прямое преобразование Фурье. Структурная схема, иллюстрирующая процесс обработки сигнала с OFDM, приведена на рис. 9.12. По-

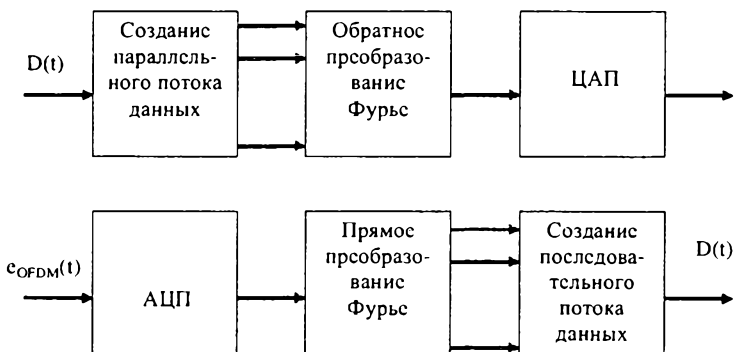


Рис. 9.12. Структурная схема обработки сигнала с OFDM

сколько все вычисления производятся в цифровой форме, то на входе схемы появляется аналого-цифровой преобразователь, а на выходе схемы — цифроаналоговый преобразователь.

Модуляция OFDM нашла применение в системах цифрового радиовещания и в системе цифрового телевизионного вещания DVB (*Digital Video Broadcasting*), разработанного группой представителей ряда европейских стран.

Сигналы ортогональны, если они взаимно независимы друг от друга. Ортогональность — это свойство, которое позволяет сложному информационному сигналу проходить без искажений по каналу связи и детектироваться без интерференции. Многие схемы с мультиплексированием являются схемами глубоко ортогональными. Например, TDM обеспечивает передачу сложных информационных сигналов, назначая уникальные временные слоты для каждого информационного сигнала. В течение каждого временного слота только сигнал от одного источника передается, предотвращая какую-либо интерференцию между другими источниками. Поэтому технологию передачи информации TDM принято считать по природе своей ортогональной.

В частотной области FDM системы также ортогональны, так как каждый сигнал размещается в своей полосе частот, чтобы предотвратить какое-либо взаимодействие с сигналами, находящимися в своих полосах частот, разделенных значительными защитными промежутками. Как видим, технологии TDM и FDM обеспечивают ортогональность передаваемых сигналов, но термин «ортогональность» зарезервирован для специальной формы FDM, называемой OFDM, у которой поднесущие размещены так тесно, как это теоретически возможно, одновременно обеспечивая их ортогональность даже в условиях взаимного перекрытия спектров поднесущих частот. Полоса частот каждой поднесущей выбирается так, чтобы равнялась целому числу, умноженному на величину, обратную длительности символа. При этом каждая поднесущая имеет целое число циклов, укладывающихся на длительности символа.

Модуляция поднесущих осуществляется битами, предназначенными для передачи, которые, например, изменяют амплитуды и фазы поднесущего колебания, формируют отображение, показанное на рис. 9.13 в виде комплексных чисел

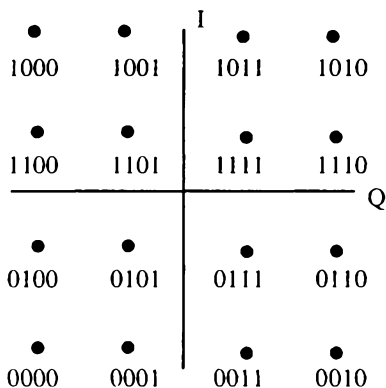


Рис. 9.13. Сигнальное созвездие 16-QAM модуляции с кодированием Грея

с составляющими I и Q . На рис. 9.13 показано отображение, соответствующее сигналу 16-QAM, у которого 4 бита информации запечатлеваются в каждом символе. Каждой комбинации из 4 битов, закодированных по коду Грея, соответствует уникальный вектор (I, Q) .

9.4. Генерирование OFDM-сигналов

Сначала выбирается требуемый спектр, основанный на входных данных и схеме используемой модуляции DBPSK, QPSK или QAM. Подлежащие передаче данные связываются с каждой поднесущей, амплитуды и фазы которых устанавливаются по используемому виду модуляции. После этого спектр превращается во временную область с помощью алгоритма обратного преобразования Фурье. Следующий шаг — добавление защитного интервала к передаваемому символу, назначение которого — предотвратить влияние многолучевого распространения на результат приема переданных символов. Эта операция сводится к формированию циклического префикса, располагающегося в начальной части символа и представляющего его конечную часть. В результате этого импульс не теряет своих периодичности и ортогональности. Это играет решающую роль в устранении межсимвольной интерференции и

интерференции между несущими. Циклический префикс вносит дополнительные потери в отношении сигнал/шум, но они пренебрежимо малы по сравнению с эффектом смягчения явлений интерференции. Так как первые OFDM-системы не использовали цифровой модуляции и демодуляции, то непрерывная во времени модель OFDM-системы может рассматриваться как идеальная OFDM-система.

Если допустить, что в OFDM-системе с N несущими, полосой частот W и длительностью символа T и длительностью префикса T_{cp} передатчик формирует сигнал

$$\varphi(t) = \begin{cases} \frac{1}{\sqrt{T - T_{cp}}} e^{2\pi i \frac{W}{N} k(t - T_{cp})}, & t \in [0, T], \\ 0, & t \notin [0, T] \end{cases}$$

где $T = (N/W) + T_{cp}$. Заметим, что $\varphi(t) = \varphi(t + N/W)$, когда t находится внутри циклического префикса. Так как $\varphi(t)$ представляет собой прямоугольный импульс с модулированной несущей частотой kW/N , то OFDM-сигнал можно интерпретировать как сигнал, составленный из N несущих, и каждая несущая переносится с низкой битовой скоростью. Поэтому форма сигнала $\varphi(t)$, используемого при модуляции и передаче в полосе своих l частот, можно представить следующим образом:

$$s_k(t) = \sum_{k=0}^{N-1} x_{k,l} \varphi_k(t - lT),$$

где $x_{0,l}, x_{1,l}, \dots, x_{N-1,l}$ — комплексные числа, полученные из точек сигнального созвездия. Когда передается бесконечная последовательность OFDM-символов, выход передатчика формируется наложением индивидуальных OFDM-символов

$$s(t) = \sum_{l=-\infty}^{\infty} s_l(t) = \sum_{l=-\infty}^{\infty} \sum_{k=0}^{N-1} x_{k,l} \varphi_k(t - lT).$$

Важное допущение, что влияет на импульсный отклик физического канала, который может быть или не быть инвариантным во времени, ограничивается временным периодом $\tau_h [0, T_{cp}]$, т. е. длительностью циклического префикса. Принятый сигнал тогда получается равным

$$r(t) = (g * s)(t) = \int_0^{T_{cp}} g(\tau; t) s(t - \tau) d\tau + n(t),$$

где $n(t)$ — аддитивный, белый и комплексный гауссов шум.

9.5. Прием OFDM-сигнала

Набор фильтров, согласованных с последней частью $[T_{cp}, T]$ переданного колебания $\Psi_k(t)$, т. е.

$$\Psi_k(t) = \begin{cases} \varphi_k(T - t), & t \in [0, T - T_{cp}] \\ 0, & t \notin [0, T - T_{cp}] \end{cases}.$$

Эта операция эффективно устраняет циклический префикс в приемной ступени системы. Вся интерференция между символами содержится в циклическом префиксе и не проявляет себя в выборках на выходе системы, полученных с помощью набора приемных фильтров. Мы можем теперь устранить временной индекс l , когда рассчитаем выходные отсчеты у k -го согласованного фильтра

$$y_k = (r * \phi_k)(t) \Big|_{t=T} = \int_{-\infty}^{\infty} r(t) \phi_k(T - t) dt.$$

Рассматривая канал, который является фиксированным на длительности OFDM-символа, и обозначая его $g(\phi)$, после упрощений получим следующий результат:

$$y_k = G\left(\frac{kW}{N}\right)x_k + n_k,$$

где $G(f)$ представляет собой преобразование Фурье от $g(\tau)$ и n_k — аддитивный белый гауссов шум.

Модуляция и демодуляция (с $\Psi_k(t)$ & $\psi(t)$) в непрерывной во времени модели заменяется IDFT и DFT соответственно, в то время как канал представляет собой дискретно-временную свертку. Циклический префикс работает таким же образом в этой системе, и расчеты выполняются, по существу, тем же способом. Так как во всех других случаях интегралы заменяются суммами, как в дискретно-временном домене.

Применение циклического префикса с длительностью большей, чем она формируется в канале, превращает линейную свертку в циклическую. Можем записать следующее уравнение для всей OFDM-системы

$$Y_l = DFT(IDFT(x_l) * g_l + n_l) = DFT(IDFT(x_l) * g_l) - n_l,$$

где y_l содержит N принятых точек данных, x_l — N переданных точек созвездия, g — импульсный отклик канала (с добавленными нулями, чтобы получить длину N), n_l — шум канала. Так как допускаем, что шум канала является белым и гауссовским, член $n_l = DFT(n_l)$ представляет собой некоррелированный гауссовский шум. Используя результат, что DFT циклически свернутых сигналов эквивалентно произведению их индивидуальных DFT, получим

$$y_l = x_l \cdot DFT(g_l) - n_l = x_l \cdot h_l + n_l.$$

Следующие четыре главных критерия используются для оценки изготовления и работы OFDM-системы:

- устойчивость к многолучевому распространению;
- клиппирование пиковой мощности;
- шум канала;
- ошибки временной синхронизации.

9.6. Кодированное разделение с ортогональными несущими

Многолучевое распространение радиосигнала с OFDM приводит к ослаблению или даже полному подавлению некоторых несущих вследствие интерференции прямого и задержанного сигналов. Решению этой проблемы помогает использование кодирования с целью обнаружения и исправления ошибок в канале передачи данных. Кодирование превращает OFDM в COFDM — *Coded Orthogonal Frequency Division Multiplex*. Эффективность использования COFDM в условиях многолучевого приема объясняется следующими обстоятельствами. Если по каналу связи с резко выраженной неравномерностью частотной характеристики передается одна модулированная несущая, то ослабление отдельных частотных составляющих можно компенсировать с помощью частотного корректора, правда, при этом снижается отношение сигнал/шум. Но если какая-нибудь составляющая подавлена полностью, то корректирующий фильтр помочь не может в принципе, и сигнал претерпевает необратимые искажения. Однако если данные передаются с помощью частотного разделения, то даже

полное исчезновение сигналов отдельных несущих не является катастрофой, поскольку данные, переносимые этими несущими могут быть восстановлены за счет канального кодирования. Система с COFDM хорошо работает в условиях передачи данных в наземном телевидении благодаря возможности раздельной обработки сигналов большого числа несущих.

При COFDM последовательный цифровой поток преобразуется в большое число параллельных потоков (субпотоков), каждый из которых передается на отдельной несущей. Группа несущих частот, которая в данный момент времени переносит биты параллельных цифровых потоков, называется «символом COFDM». Благодаря тому, что используется большое число параллельных потоков (обычно 1705 или 6817 субпотоков), длительность символа в параллельных потоках получается существенно больше, чем в последовательном потоке данных (соответственно 280 или 1120 мкс — в зависимости от числа используемых субпотоков). Это позволяет в декодере задержать оценку значений принятых символов на время, в течение которого изменения параметров радиоканала из-за действия отраженных сигналов прекратятся и канал стабилизируется.

Система COFDM превратилась в эталонный способ модуляции для наземной передачи данных в условиях многолучевого приема. Важное достоинство системы COFDM — гибкость, обеспечиваемая возможностью оперативного изменения скорости потока данных и параметров передачи с целью приспособления к условиям распространения радиоволн. Принятие COFDM позволит свободно устанавливать параметры передачи в соответствии с условиями распространения радиоволн и требованиями к скорости потока данных и надежности приема. Иерархическая передача, возможная в рамках COFDM, дает вещателям дополнительную свободу. 8-VSB пока не может обеспечить надежный прием на простые антенны в городских условиях в зоне устойчивого приема сигналов NTSC, а COFDM может.

9.7. Сложные виды модуляции и манипуляции

9.7.1. Импульсно-кодовая модуляция

В этой системе каждый отсчет сигнала $s(t)$ передается кодовой комбинацией, составленной из нескольких импульсов.

Преобразование аналоговых сигналов в цифровые, которое соответствующим образом отображает форму исходного сигнала, осуществляется при помощи четырех последовательных этапов:

- фильтрация;
- получение выборок сигнала;
- квантование сигнала;
- кодирование сигнала.

Совместное выполнение этих этапов и составляет процесс, который называется импульсно-кодовой модуляцией (ИКМ) — (*Pulse Code Modulation*, PCM). Этот процесс задокументирован в стандарте G.711, разработанном Международным Союзом телекоммуникаций (ITU).

После фильтрации сигнала берется выборка, затем каждый отсчет округляется до значения ближайшего разрешенного уровня. Каждый уровень представляется определенной комбинацией импульсов. Таким образом, вместо отдельных отсчетов передается кодовая комбинация.

Чаще всего для образования кодовых комбинаций используются двоичные импульсы, т. е. импульсы, которые могут принимать только два значения. В общем случае могут использоваться s , тем меньше требуется разрядов в кодовой комбинации. Таким образом, при ИКМ вместо одного импульса на отсчет (как в случае АИМ) нужно передавать группу импульсов. Поэтому ИКМ требует большего времени для передачи одинаковой информации, но этот недостаток перекрывается ее большей помехоустойчивостью. Для приема сигнала требуется только установить, присутствует импульс или отсутствует независимо от его амплитуды и формы. Внешняя помеха, которая вносит искажение высоты или ширины импульсов, не влияет на прием сигналов ИКМ.

Несколько слов о кодировании, являющемся последним этапом цифро-аналогового преобразования. Результатом кодирования является цифровой поток двоичных цифр, называемый еще цифровым трафиком. Он предназначен для передачи по цифровому каналу. Кодирование в данном контексте — это процесс, при котором каждое значение амплитуды в дискретном сигнале преобразуется в последовательность, например, из 8 битов. Подобно схеме квантования, метод кодирования основан на алгоритмах уплотнения, а 8-битовые

коды, формируемые для одних и тех же квантовых значений, совершенно могут отличаться друг от друга.

9.7.2. Дифференциальная ИКМ

Дифференциальная ИКМ (*Differential Pulse Code Modulation, DPCM*) имеет место, когда квантованию подвергаются не сами импульсы, а разница их амплитуд.

Схема кодирования может изменять диапазон возможных значений, если разница между значениями амплитуд выходит за границы текущего диапазона возможных значений. В этом случае говорят о адаптивной дифференциальной ИКМ (*Adaptive Differential Pulse Code Modulation, ADPCM*). Такой подход позволяет учесть все возможные разницы амплитуд, но может привести к увеличению погрешности квантования.

Для речевых сигналов со стандартной для передачи речи полосой 0,3–3,5 кГц обычно используют частоту дискретизации $F_{\text{дискр}} \geq 2F_{\text{max}} = 8 \text{ кГц}$. Экспериментально показано, что при равномерном квантовании для получения практически идеального качества речи нужно квантовать сигнал не менее чем на ± 2000 уровней, иными словами, для представления каждого отсчета понадобится 12 бит, а результирующая скорость кода будет составлять:

$$R = 8000 \text{ отсчетов/с} * 12 \text{ бит/отсчет} = 96000 \text{ бит/с}.$$

Используя *неравномерное квантование* (более точное для малых уровней сигнала и более грубое для больших его уровней) таким образом, чтобы относительная ошибка квантования была постоянной для всех уровней сигнала, можно достичь того же самого субъективного качества восстановления речевого сигнала, но при гораздо меньшем числе уровней квантования — порядка ± 128 . В этом случае для двоичного представления отсчетов сигнала понадобится уже 8 бит и результирующая скорость кода составит 64 кбит/с.

С учетом статистических свойств речевого сигнала (вида распределения вероятностей мгновенных значений), а также нелинейных свойств слуха, гораздо лучше различающего слабые звуки, оптимальной является логарифмическая шкала квантования, которая и была принята в качестве стандарта еще в середине 60-х годов и сегодня повсеместно используется. Правда, в США и Европе стандарты нелинейного кван-

тования несколько различаются (*m-law companding* и *A-law compression*), что приводит к необходимости перекодирования сигналов.

Таким образом, исходной для любого сравнения эффективности и качества кодирования речевых сигналов может служить скорость кода, равная 64 кбит/с.

Следующим приемом, позволяющим уменьшить результирующую скорость кода, может быть попытка предсказания значения текущего отсчета сигнала по нескольким предыдущим его значениям и далее кодирование уже не самого отсчета, а ошибки его предсказания — *разницы между истинным значением текущего отсчета и его предсказанным значением*. Если точность предсказания достаточно высока, то ошибка предсказания очередного отсчета будет значительно меньше величины самого отсчета, и для ее кодирования понадобится гораздо меньшее число бит. Таким образом, *чем более предсказуемым будет поведение кодируемого сигнала, тем более эффективным будет его сжатие*.

Описанная идея лежит в основе так называемой *дифференциальной импульсно-кодовой модуляции* (ДИКМ) (DPCM) — способа кодирования, при котором кодируются не сами значения сигнала, а их *отличия от некоторым образом предсказанных значений*. Простейшим способом предсказания является *использование предыдущего отсчета x_{i-1} сигнала в качестве предсказания его текущего значения: $x_i^* = x_{i-1}$* .

Это так называемое *предсказание нулевого порядка* $e_i = x_i - x_i^*$, самое простое, но и наименее точное. Более точным, очевидно, будет предсказание текущего отсчета на основе линейной комбинации двух предшествующих и т. д.

К сожалению, точность предсказания не всегда растет с ростом порядка предсказания, поскольку свойства сигнала между отсчетами изменяются. Поэтому обычно ограничиваются предсказанием не выше 2–3-го порядка.

На рис. 9.14 и 9.15 приведены схемы ДИКМ кодера и декодера.

При кодировании речевых сигналов с учетом степени их кратковременной (на несколько очередных отсчетов) предсказуемости результирующая скорость кода для ДИКМ обычно составляет 5–6 бит на отсчет, или 40–48 кбит/с.



Рис. 9.14. Схема кодера ДИКМ Рис. 9.15. Схема декодера ДИКМ

Эффективность ДИКМ может быть несколько повышена, если предсказание и квантование сигнала будут выполняться не на основе некоторых усредненных его характеристик, а с учетом их текущего значения и изменения во времени, т. е. адаптивно. Так, если скорость изменения сигнала стала большей, можно увеличить шаг квантования, и, наоборот, если сигнал стал изменяться гораздо медленнее, величину шага квантования можно уменьшить. При этом ошибка предсказания уменьшится и, следовательно, будет кодироваться меньшим числом бит на отсчет. Такой способ кодирования называется *адаптивной* ДИКМ, или АДИКМ (ADPCM). Сегодня такой способ кодирования стандартизован и широко используется при сжатии речи в междугородных цифровых системах связи, в системе микросотовой связи DECT, в цифровых бесшнуровых телефонах и т. д. Использование АДИКМ со скоростью кода 4 бита/отсчет, или 32 *кбит/с*, обеспечивает такое же субъективное качество речи, что и 64 *кбит/с* с ИКМ, но при вдвое меньшей скорости кода.

На сегодня стандартизованы также АДИКМ — кодеки для скоростей 40, 24 и 16 *кбит/с* (в последнем случае с несколько худшим, чем для 32 *кбит/с* — АДИКМ — качеством сигнала). Таким образом, видно, что сжатие речевых сигналов на основе кодирования их формы обеспечивает в лучшем случае двух - трехкратное уменьшение скорости кода. Дальнейшее снижение скорости ведет к резкому ухудшению качества кодируемого сигнала.

Описанные выше кодеры формы сигнала использовали чисто временной подход к описанию этого сигнала. Однако возможны и другие подходы. Примером может служить так называемое *кодирование поддиапазонов* (*Sub-Band Coding, SBC*), при котором входной сигнал разбивается (или расфилъ-

тровывается) на несколько частотных поддиапазонов (поддиапазонов — *sub-bands*) и сигнал в каждом из этих поддиапазонов кодируется по отдельности, например с использованием техники АДИКМ.

Поскольку каждый из частотных поддиапазонов имеет более узкую полосу (все поддиапазоны в сумме дают полосу исходного сигнала), то и частота дискретизации в каждом поддиапазоне также будет меньше. В результате суммарная скорость всех кодов будет по крайней мере не больше, чем скорость кода для исходного сигнала. Однако у такой техники есть определенные преимущества. Дело в том, что субъективная чувствительность слуха к сигналам и их искажениям различна на разных частотах. Она максимальна на частотах $1\text{--}1,5\text{ кГц}$ и уменьшается на более низких и более высоких частотах. Таким образом, если в диапазоне более высокой чувствительности слуха квантовать сигнал более точно, а в диапазонах низкой чувствительности более грубо, то можно получить выигрыш в результирующей скорости кода. Действительно, при использовании технологии *кодирования поддиапазонов* получено хорошее качество кодируемой речи при скорости кода $16\text{--}32\text{ кбит/с}$. Кодер получается несколько более сложным, чем при простой АДИКМ, однако гораздо проще, нежели для других эффективных способов сжатия речи.

Упрощенная схема подобного кодера (с разбиением на 2 поддиапазона) приведена на рис. 9.16.

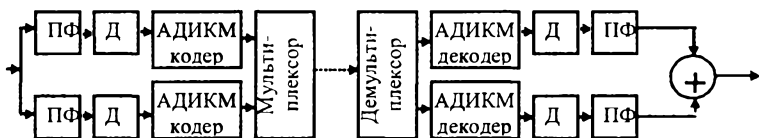


Рис. 9.16. Упрощенная схема кодера с разбиением на 2 поддиапазона

Близким к кодированию поддиапазонов является метод сжатия, основанный на применении к сигналу линейных преобразований, к примеру дискретного косинусного или синусного преобразования. Для кодирования речи используется так называемая технология АТС (*Adaptive Transform Coding*), при

которой сигнал разбивается на блоки, к каждому блоку применяется дискретное косинусное преобразование и полученные коэффициенты адаптивно, в соответствии с характером спектра сигнала, квантуются. Чем более значимыми являются коэффициенты преобразования, тем большим числом бит они кодируются. Техника очень похожа на JPEG, но применяется к речевым сигналам. Достижимые при таком кодировании скорости кодов составляют 12–16 *кбит/с* при вполне удовлетворительном качестве сигнала. Широкого распространения для сжатия речи этот метод не получил.

9.7.3. Дельта-модуляция

Для успешного восстановления формы аналогового сигнала принимающей стороной достаточно информации об относительном изменении амплитуды двух соседних импульсов. Поэтому для реализации метода, называемого дельта-модуляцией, требуется только один бит, указывающий на направление изменения амплитуды. При дельта-модуляции каждому биту соответствует целое значение амплитуды. В результате при передаче каждого значения амплитуды передается относительно небольшое количество информации об этом значении, что является основной причиной, по которой зернистость может изменить результат. Единственный способ, позволяющий удерживать оба типа искажений в допустимых пределах, — применение достаточно малого шага дискретизации. Соответственно, такой способ предполагает использование частоты дискретизации, превышающей частоту Найквиста.

Для каждого значения амплитуды отводится 1 *бит*, поэтому битовая скорость соответствует скорости дискретизации. Этот метод разработан, чтобы сократить расход полосы пропускания, но не следует ради него отказываться от преимуществ сжатия.

Дельта-модуляция не всегда может эффективно решить поставленные перед ней задачи. Однако это под силу другой методике, основанной на дельта-модуляции. Она называется дельта-модуляцией с переменной крутизной наклона (*Continuously Variable Slope Delta Modulation, CVSDM*). Метод CVSDM позволяет наблюдать за управлением битовым потоком. Для этого можно использовать либо последователь-

ность единиц, увеличивающую размер шага для понижения наклона аналогового сигнала, либо последовательность нулей, уменьшающую размер шага для увеличения крутизны наклона и ослабления эффекта зернистости. В других адаптивных формах дельта-модуляции просто выполняется наблюдение за тенденциями посредством запоминания предыдущих направлений изменения амплитуды. Если для многих импульсов было замечено одно и то же направление изменения амплитуды, то можно предположить, что наклон сигнала является достаточно пологим. Чтобы уменьшить зернистость, нужно уменьшить шаг дискретизации. И наоборот, если чередуются отрицательные и положительные направления изменения амплитуды, это говорит о том, что наклон получился выше нормы. В этом случае шаг дискретизации увеличивается во избежание завышения наклона кривой.

В дополнение к временным методам существуют и частотные подходы. Они имеют ряд преимуществ. При использовании, например, SB-ADPCM входной сигнал разбивается на несколько частотных полос, которые кодируются независимо друг от друга при помощи метода ADPCM. На принимающей стороне биты декодируются, в результате чего восстанавливается исходный полосовой сигнал.

Преимущество этого метода в том, что погрешность в каждом диапазоне частот зависит только от кодирования. Поэтому можно использовать более точные схемы кодирования и шаги квантования для тех диапазонов частот, которые наиболее важны для воспроизведения сигнала. Таким образом, погрешность в этих диапазонах частот невелика, в то время как в других допустим более высокий уровень погрешности.

Между соседними отсчетами речевого сигнала имеется значительная корреляция, которая слабо убывает по мере увеличения интервала между отсчетами. Это означает, что речевой сигнал изменяется медленно и разность между соседними отсчетами будет иметь меньшую дисперсию, чем исходный сигнал, что позволяет применять методы *разностного квантования* речевого сигнала (рис. 9.17).

Линейная дельта-модуляция (ДМ) (рис. 9.18) использует одноразрядный (двухуровневый) квантователь и предсказатель. Восстановление аналогового сигнала из сигнала линейной ДМ осуществляется суммированием шага квантования.



Рис. 9.17. Линейная дельта-модуляция: аналоговый сигнал (кривая 1) и сигнал квантователя линейной ДМ (кривая 2)

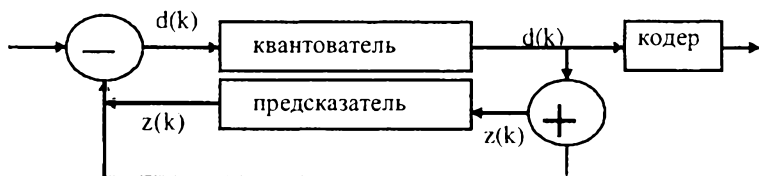


Рис. 9.18. Структурная схема кодера разностного квантования

Линейная дельта-модуляция технически реализуется относительно просто, но обладает рядом недостатков: перегрузка по крутизне, шум дробления (шум незанятого канала).

Кроме того, для обеспечения приемлемого качества восстановления речевого сигнала требуется высокая скорость преобразования (передачи) — порядка 200 кбит/с.

Адаптивная дельта-модуляция. Шаг квантования меняется в зависимости от крутизны исходного сигнала от минимального до максимального значения. Возможны различные схемы (алгоритмы) адаптивной модуляции. Отслеживается выходной поток квантователя — при чередовании 0 и 1 шаг уменьшается, при последовательных 0 или 1 шаг увеличивается. Данный вид модуляции не получил широкого применения.

Дальнейшим развитием систем разностного квантования является *адаптивная дифференциальная ИКМ*. Методы адаптации распространяются как на квантователь, так и на предсказатель. Осуществляется передача цифровых представлений адаптивного шага и коэффициентов предсказателя. Удовлетворительные результаты получаются при скорости передачи

32 и 24 кбит/с. МСЭ-Т стандартизировал данный вид модуляции в Рекомендации G.726 для скорости передачи 32 кбит/с.

9.7.4. Треллис-модуляция

Треллисное кодирование основано на сверточном кодировании и сводится к тому, что n битам сверточным кодом со скоростью кода n/m ставится в соответствие один из 2^n сигналов. Устройство, выполняющее это соответствие, в зарубежной литературе называется *mapper*. При многоуровневой модуляции переход от ансамбля сигналов с объемом канального алфавита 2^n к 2^m не требует увеличения полосы частот. В то же время сверточным кодом с числом состояний q можно формировать последовательности сигналов с достаточно большими евклидовыми расстояниями между ними. Принимая такие треллис-кодированные последовательности сигналов на основе критерия максимального правдоподобия по алгоритму Витерби, в целом можно получить энергетический выигрыш в сравнении с приемом некодированных сигналов с объемом канального алфавита 2^n . Уже найдены оптимальные коды на основе критерия максимума минимального евклидова расстояния между треллис-кодированными последовательностями сигналов. Такой критерий обеспечивает максимальную помехоустойчивость приема в гауссовском канале.

Применение многопозиционной КАМ в чистом виде сопряжено с проблемой недостаточной помехоустойчивости. Поэтому во всех высокоскоростных протоколах КАМ используется совместно с решетчатым кодированием — специальным видом сверточного кодирования. В результате появился новый способ модуляции, называемый треллис-модуляцией (*Trellis Coded Modulation*, TCM).

Выбранная определенным образом комбинация конкретной КАМ и помехоустойчивого кода в отечественной технической литературе носит название сигнально-кодовой конструкции (СКК), которая позволяет повысить помехозащищенность передачи информации одновременно со снижением требований к отношению сигнал-шум на 3–6 дБ. Это объясняется увеличением вдвое числа сигнальных точек за счет добавления к информационным битам одного избыточного, образованного путем сверточного кодирования. Расширенный таким

образом блок битов подвергается все той же КАМ. В процессе демодуляции принятый сигнал декодируется по алгоритму Витерби. Именно этот алгоритм за счет введенной избыточности и знания предыстории процесса приема позволяет по критерию максимального правдоподобия выбрать из сигнального пространства наиболее достоверную точку.

Выбор способов модуляции и кодирования сводится к поиску такого заполнения сигнального пространства, при котором обеспечивается высокая скорость и высокая помехоустойчивость. Комбинирование различных ансамблей многопозиционных сигналов и помехоустойчивых кодов порождает множество вариантов сигнальных конструкций. Согласованные определенным образом варианты, обеспечивающие улучшение энергетической и частотной эффективности, и являются сигнально-кодowymi конструкциями. Задача поиска наилучшей СКК является одной из наиболее сложных задач теории связи. Современные высокоскоростные протоколы модуляции (V.32, V.32bis, V.34 и др.) предполагают обязательное применение сигнально-кодowych конструкций.

Все применяемые сегодня СКК используют сверточное кодирование, когда при передаче одного сигнального элемента используется только один избыточный двоичный символ.

Типичный кодер, применяемый совместно с модулятором ФМ-8, представлен на рис. 9.19. Он является сверточным

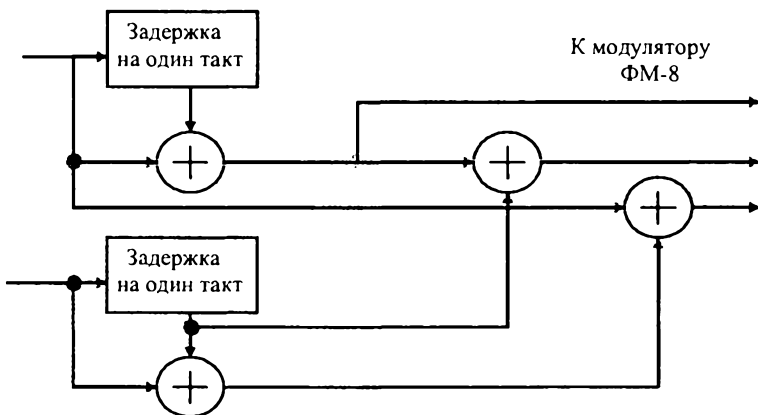


Рис. 9.19. Схема сверточного 2/3 кодера

кодером с относительной скоростью кода, равной $2/3$. Каждым двум информационным битам на входе кодер сопоставляет трехсимвольные двоичные блоки на своем выходе, которые и поступают на модулятор ФМ-8.

Применение сигналов ФМ связано с разрешением проблемы неоднозначности фазы восстановленной на приеме несущей. Данная проблема решается за счет относительного (дифференциального) кодирования, что в системах без помехоустойчивого кодирования приводит к размножению ошибок. В системах с помехоустойчивым кодированием относительное кодирование также используется. В этом случае имеет значение последовательность включения относительного и помехоустойчивого кодера.

Различают *внешнее* и *внутреннее* относительное кодирование. При внутреннем кодировании относительный кодер расположен на выходе помехоустойчивого кодера, а на приемной стороне относительный декодер включен на входе помехоустойчивого декодера. В этом случае помехоустойчивый кодер должен уметь бороться с группирующимися ошибками.

Внешнее относительное кодирование в ряде случаев является более выгодным, так как источник размножения ошибок — относительный декодер — включен на выходе помехоустойчивого декодера. Однако при этом возникают трудности декодирования, вызванные неоднозначностью фазы опорного колебания при демодуляции. При ФМ-2 неоднозначность фазы опорного колебания (0 или 1) приводит к явлению «обратной работы», заключающейся в том, что передаваемые единичные биты принимаются нулевыми, а нулевые — наоборот, единичными. При большем числе позиций фазы возможна не только инверсия, но и перестановка двоичных символов. Решение этой проблемы заключается в использовании помехоустойчивых кодов, прозрачных, т. е. нечувствительных, к неопределенности фазы опорного колебания. Известно несколько видов СКК, обеспечивающих прозрачность к неопределенности фазы восстановленной несущей. Они также основаны на сверточном кодировании, когда используется только один избыточный двоичный символ.

9.7.5. Многоуровневая модуляция с частично подавленной боковой полосой частот

Фирма Zenith запатентовала метод многоуровневой модуляции с частично подавленной боковой полосой частот (*Vestigal Side Band, VSB*). Для цифрового наземного ТВ используется 8-уровневая модуляция с треллисным кодированием, и по этой причине режим называется 8-VSB, для кабельного цифрового ТВ – режим 16-VSB без треллисного кодирования.

VSB-модуляция используется во всех системах аналогового телевидения (NTSC, PAL, SECAM). Применительно к системе NTSC (и почти так же в PAL) она реализуется следующим образом. Два цветоразностных сигнала осуществляют квадратурную амплитудную модуляцию поднесущей 3,58 МГц. Этот сигнал складывается с яркостным сигналом, полоса частот которого равна 6 МГц, производится вставка различных синхроимпульсов, и уже этим суммарным сигналом осуществляется амплитудная модуляция промежуточной частоты (ПЧ). В результате имеем две симметричные относительно ПЧ боковые полосы (БП) частот. Далее полосовой фильтрацией осуществляется частичное подавление одной БП так, чтобы несущая частота отстояла от левой граничной частоты спектра на расстоянии 1,25 МГц и была неподавленной. При приеме такого VSB-сигнала неподавленная несущая используется для фазовой и частотной синхронизации местного гетеродина с целью восстановления исходного суммарного сигнала.

10.1. Диаграммы состояний многопозиционных сигналов

При модуляции цифровым *биполярным NRZ-сигналом* на выходе *балансного модулятора* получается сигнал *двухпозиционной фазовой манипуляции (ФМ-2)*. Спектр сигнала ФМ-2 образуется переносом спектра модулирующего NRZ-сигнала на высокочастотную *несущую* f_0 . Ширина главного лепестка спектра при этом составляет $\Delta f = 2/T_c$, где T_c — длительность одного символа. С учетом того, что в случае NRZ модулирующего сигнала длительность символа совпадает с длительностью одного бита T_B , можно записать

$$\Delta f = 2/T_B. \quad (10.1)$$

Таким образом, обладая высокой помехоустойчивостью, ФМ-2 характеризуется низкой спектральной эффективностью.

Большей спектральной эффективностью обладают многопозиционные сигналы, из которых наиболее часто используют:

- четырехпозиционную фазовую манипуляцию (ФМ-4);
- шестнадцатипозиционную амплитудно-фазовую манипуляцию (АМФ-16).

Используют также сигналы ФМ-8, АФМ-8, АФМ-64 и другие. В многопозиционных сигналах каждый импульс несет информацию о нескольких битах: о двух битах в случае ФМ-4, о трех — при ФМ-8 и АФМ-8, о 6 — при АФМ-64. Сигнальные созвездия для сигналов ФМ-4 и АФМ-16 приведены на рис. 4.14 и 4.15.

Каждое из четырех возможных состояний сигнала ФМ-4 отличается фазовым сдвигом. Амплитуда же сигнала постоянна.

Состояния же сигнала АФМ-16 различаются как фазовыми сдвигами, так и амплитудами.

Постоянство огибающей при ФМ-2, ФМ-4, ФМ-8 является важным преимуществом этих сигналов, так как высокочастотные тракты передатчика и приемника при этом могут работать в нелинейном режиме, что позволяет получить максимальную выходную мощность при высоком КПД передатчика.

При использовании АФМ-8, АФМ-16 и АФМ-64 высокочастотные тракты передатчика и приемника должны быть высоколинейными. Энергетические показатели передатчика при этом существенно хуже, чем при использовании сигналов с постоянной огибающей.

10.2. Формирование и демодуляция многопозиционных радиосигналов

Несмотря на многообразие многопозиционных сигналов, их формирование в передатчиках и демодуляция в приемниках проводится с использованием общего технического решения, основанного на раздельном формировании двух независимых квадратурных составляющих модулирующего сигнала $I(t)$ и $Q(t)$ и их последующей передачи на одной несущей методом квадратурной амплитудной модуляции. Структурная схема модулятора многопозиционного сигнала приведена на рис. 10.1.

Формирователь сигналов с АИМ вырабатывает многоуровневые сигналы $I(t)$ и $Q(t)$, значения уровней которых равны проекциям векторов сигналов на вещественную и мнимую оси. Число уровней сигналов $I(t)$ и $Q(t)$ равны двум при ФМ-4,

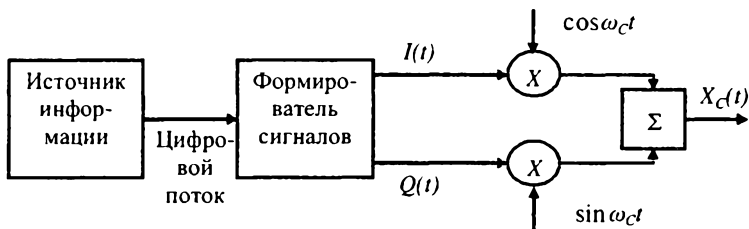


Рис. 10.1. Структурная схема модулятора многопозиционных сигналов

четырем — при АФМ-16, шести — при АФМ-64. Длительность символов равна длительности двух бит при ФМ-4, четырех бит — при АФМ-16, шести бит — при АФМ-64. Диаграммы, поясняющие процесс формирования сигнала ФМ-4, приведены на рис. 10.2, а сигнала АФМ-16 — на рис. 10.3.

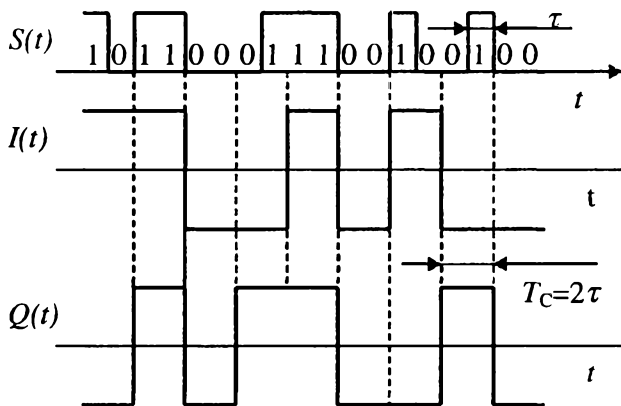


Рис. 10.2. Диаграммы, поясняющие процесс формирования сигнала ФМ-4

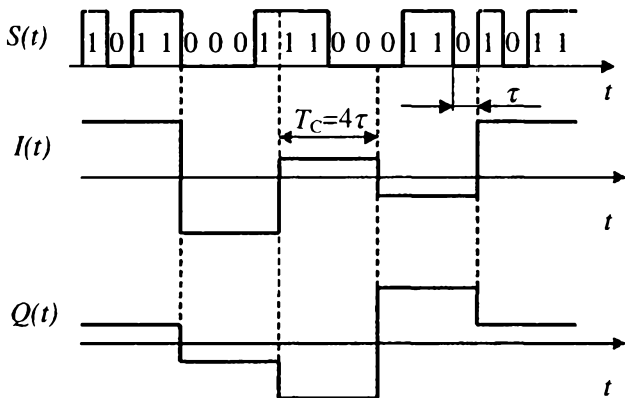


Рис. 10.3. Диаграммы, поясняющие процесс формирования сигнала АФМ-16

На рис. 10.4 показана схема демодулятора многопозиционного сигнала. Демодулятор содержит два *синхронных детектора*, которые выделяют многоуровневые сигналы $I'(t)$ и $Q'(t)$. Из-за наличия в канале шумов они отличаются от переданных $I(t)$ и $Q(t)$. Задача решающего устройства — восстановить переданную цифровую последовательность с минимальным числом ошибок. Для нормальной работы синхронных детекторов частота несущей f_c должна быть точно равна частоте несущей передатчика. Это достигается с помощью специальных методов восстановления несущей.

Прием широкополосных ФМн-сигналов осуществляется оптимальным приемником, который для сигнала с полностью известными параметрами вычисляет корреляционный интеграл

$$Z = \int_0^T x(t)u(t)dt, \quad (10.2)$$

где $x(t)$ — входной сигнал, представляющий собой сумму полезного сигнала $u(t)$ и помехи, чаще всего являющейся аддитивным гауссовским белым шумом.

Затем величина Z сравнивается с порогом Z_0 . Значение корреляционного интеграла находится с помощью согласованного фильтра или коррелятора.

Форма спектров всех упомянутых сигналов совпадает с формой спектра для сигнала ФМ-2, а ширина главного лепестка определяется тем же соотношением $\Delta f = 2/T_c$. Однако с учетом того, что длительность символа в случае многопозицион-

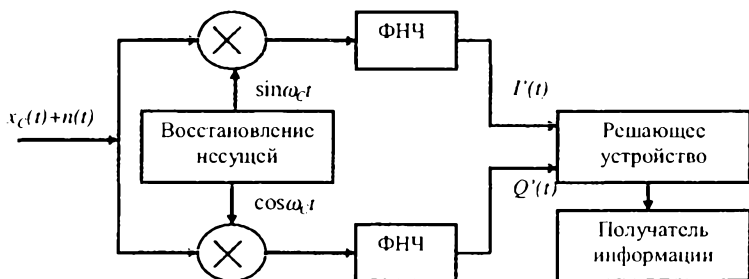


Рис. 10.4. Структурная схема демодулятора многопозиционного сигнала

ных сигналов превышает длительность бита, значения ширины полосы будут меньше: для сигналов ФМ-8 и АФМ-8 она составляет $2/3T_c$, для сигналов ФМ-16 и АФМ-16 — $1/2T_c$, для сигнала АФМ-64 — $1/3T_c$.

10.3. Помехоустойчивость многопозиционных сигналов

Помехоустойчивость различных видов модуляции удобно связать с минимальным расстоянием между возможными состояниями сигнала d_{min} , выраженную через среднюю энергию, приходящуюся на один бит E_b .

В случае ФМ-2 $d_{min} = 2A = 2\sqrt{E_b}$, где A — амплитуда сигнала.

При ФМ-4 $d_{min} = \sqrt{2}A = \sqrt{2}\sqrt{2}E_b = 2\sqrt{E_b}$.

Таким образом, при ФМ-4 минимальное расстояние между сигналами такое же, как и при ФМ-2. Следовательно, ФМ-4 обладает такой же помехоустойчивостью, что и ФМ-2. Занимаемая же полоса частот при ФМ-4 в два раза уже. Именно поэтому в действующих системах связи предпочтение отдают ФМ-4.

Остальные виды модуляции имеют меньшую помехоустойчивость, чем ФМ-2 и ФМ-4. Значения d_{min} для различных видов модуляции приведены в таблице:

$$\begin{array}{ll} \text{ФМ-8 } 1,325\sqrt{E_{bcp}}, & \text{АФМ-8 } 1,549\sqrt{E_{bcp}}, \\ \text{ФМ-16 } 0,78\sqrt{E_{bcp}}, & \text{АФМ-16 } 1,265\sqrt{E_{bcp}}, \end{array}$$

Из сравнения ФМ-8 и АФМ-8 следует, что при одинаковой полосе АФМ обладает большей помехоустойчивостью. Еще в большей степени это относится к ФМ-16 и АФМ-16. Именно поэтому ФМ-16 практически не используют, а АФМ-16 распространена очень широко.

10.4. Модуляция QPSK и S-DQPSK

Модуляция QPSK (*Quadrature или Quaternary Phase Shift Keying*) — квадратурная относительная фазовая модуляция — является решением компромисса между скоростью передачи и помехоустойчивостью и применяется как самостоятельно, так и в комбинациях с другими методами. Диаграмма состояний

модуляции QPSK и офсетной дифференциальной QPSK ($\pi/4$ DPSK) показаны на рис. 10.5. При реализации дифференциального кодирования в сочетании со сдвигом несущей на $\pi/4$ сигнальное созвездие формируется двумя четырехточечными созвездиями QPSK, наложенными со сдвигом 45° . В результате в сигнале присутствуют 8 фазовых сдвигов, причем фазы символов выбираются поочередно то из одного созвездия, то из другого. Последовательные символы имеют относительные фазовые сдвиги, соответствующие одному из 4 углов: $\pm\pi/4$ и $\pm3\pi/4$.

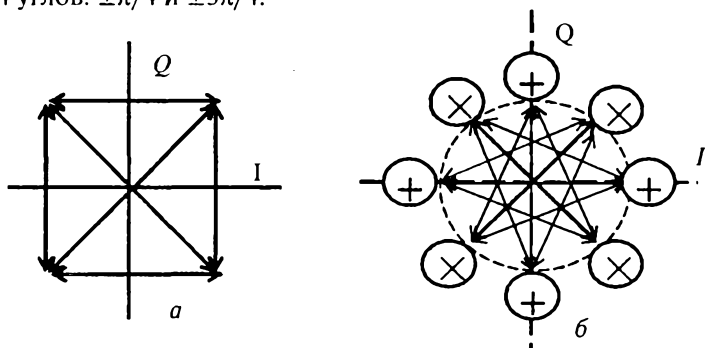


Рис. 10.5. Диаграммы состояний сигналов QPSK (а) и $\pi/4$ DQPSK (б)

В методе $\pi/4$ DQPSK все импульсы информационной последовательности модулятора разбиваются на пары, представляющие собой 2-битовые символы, и при переходе от одной пары к другой начальная фаза несущей частоты сигнала изменяется на величину D_j , которая определяется битами символа в соответствии с алгоритмом, приведенным выше.

Кружочками обозначены дискретные значения, которые может принимать фаза несущей, отсчитываемая от некоторого начального значения. Стрелками указаны возможные переходы между разрешенными значениями фазы. Оси координат соответствуют синфазной (*Inphase* — I) и квадратурной (*Quadrature* — Q) составляющим сигнала. Мы видим, что фазовая диаграмма сигнала $\pi/4$ DQPSK состоит фактически из двух диаграмм квадратурной фазовой манипуляции (QPSK): фазовые состояния одной из них помечены значком E, а другой — значком $\otimes$, и диаграммы сдвинуты одна относительно другой на угол $\pi/4$.

При переходе от одного символа к другому происходит изменение фазы от одного состояния первой диаграммы к одному из возможных состояний второй, а при переходе к следующему символу — возврат к предыдущей диаграмме, хотя, скорее всего, не к прежнему фазовому состоянию.

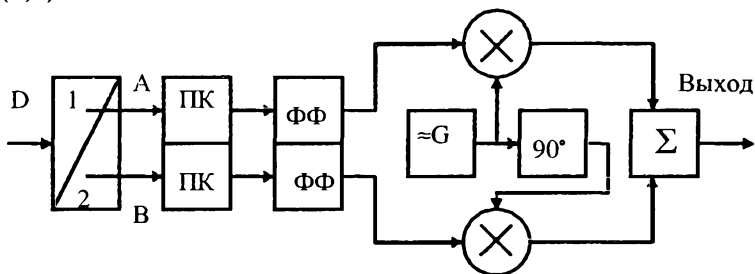
Результирующий выходной сигнал модулятора можно представить в виде: $s(t) = \cos(\omega_0 t + \varphi_k)$, где ω_0 — несущая частота, $\varphi_k = \varphi_{k-1} + \Delta\varphi$ — начальная фаза на интервале k -го импульса.

Таблица 10.1

Биты входной последовательности модулятора		Изменение фазы $\Delta\varphi$
1	1	$-3\pi/4$
0	1	$3\pi/4$
1	0	$-\pi/4$
0	0	$\pi/4$

Структурная схема модулятора показана на рис. 10.6. Входной поток данных D разделяется на два параллельных потока A и B , которые затем в преобразователе кода перекодируются в относительный код двух каналов (компонентов) I' и Q' .

Цифровые потоки I' и Q' подвергаются сглаживанию в формирующих фильтрах (ФФ), выходные сигналы которых I и Q непосредственно управляют работой четырехфазового модулятора, состоящего из двух балансных модуляторов и сумматора. Фазовый сдвиг несущих в каналах I и Q равен 90° . Правило кодирования фазовых сдвигов соответствует выполнению условий: $(0,0) - 45^\circ$; $(0,1) - 135^\circ$; $(1,0) - 315^\circ$; $(1,1) - 225^\circ$.



QPSK

Рис. 10.6. Структурная схема модулятора

10.5. Многочастотная модуляция OFDM

Для достижения необходимого качества приема в ряде систем применяют очень сложные методы модуляции и канального кодирования. К их числу относится многочастотный метод передачи OFDM (*Orthogonal Frequency Division Multiplex* — ортогональное частотное разделение мультимплекса). Под мультимплексом понимают цифровой поток, передаваемый по одному каналу связи.

Суть метода OFDM состоит в следующем. Например, при передаче телевизионного сигнала в полосе канала размещается от 2 до 8 тысяч несущих. Пример расстановки несущих в спектре OFDM показан на рис. 10.7.

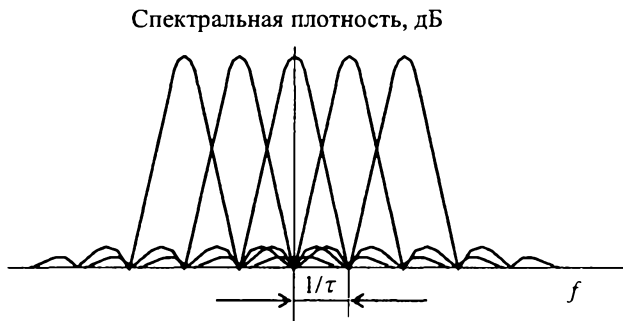


Рис. 10.7. Спектр мощности радиосигнала OFDM

Каждая несущая частота модулируется низкочастотным потоком данных, являющимся частью общего цифрового потока. В качестве первичного метода модуляции разделенных несущих допускается использовать QPSK, 16- и 64-QAM. Скорость передачи данных в каждом элементарном потоке равна общей системной скорости, деленной на число несущих. Частотное распределение несущих в полосе радиоканала производится по определенному правилу их ортогональности, допускающему перекрытие скатов смежных спектров, но при условии, что на центральной частоте каждая из несущих спектральных составляющих всех остальных модулированных несущих переходит через нуль. В результате этого за счет большо-

го числа несущих формируется групповой спектр мощности, очень близкий к прямоугольному при любом коэффициенте скругления спектра узкополосных модулирующих сигналов, что дает максимальную эффективность использования полосы радиоканала.

Каждая из N немодулированных несущих, входящих в набор OFDM, может быть выражена формулой $s_n(t) = \cos 2\pi f_n t$, где $f_n = f_0 + n/T_S$, $n = 0, 1, 2, \dots, N-1$, T_S — длительность символа, f_0 — частота самой низкой несущей, а модулированный сигнал OFDM — формулой

$$s(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} [a_k \cos 2\pi (f_0 + \frac{k}{T_S})t + b_k \sin 2\pi (f_0 + \frac{k}{T_S})t], \quad (10.3)$$

где N — число несущих, $a_k, b_k \in \{-1, +1\}$ — данные синфазного и квадратурного каналов, модулирующие k -ю несущую.

В наиболее общей форме OFDM-сигнал можно описать как параллельно передаваемый поток модулированных несущих

$$s(t) = \sum_{n=-\infty}^{\infty} (\sum_{k=0}^{N-1} C_{n,k} g_k(t - nT_S)), \quad (10.4)$$

где

$$g(t) = \begin{cases} e^{i2\pi f_k t}, & 0 \leq t < T_S \\ 0, & t < 0, t \geq T_S \end{cases}, \quad f_k = f_0 + \frac{k}{T_S}, \quad k = 0, \dots, N-1, \quad (10.5)$$

$C_{n,k}$ — амплитуда k -й поднесущей переданного символа в n -м интервале длительности T_S , N — число OFDM поднесущих, f_k — k -я поднесущая частота с f_0 в качестве самой низкой поднесущей частоты.

Определим n -е OFDM-окно (фрейм) как переданный сигнал для n -го интервала, длительность которого равна одному периоду символа T_S , и обозначим его $F_n(t)$.

Подставляя $F_n(t)$ в (4) вместо выражения, заключенного в скобки, которое соответствует n -му OFDM фрейму, соотношение (4) можно переписать как

$$s(t) = \sum_{n=-\infty}^{\infty} F_n(t). \quad (10.6)$$

Таким образом, $F_n(t)$ соответствует множеству символов $C_{n,k}$, $k = 0, \dots, N-1$, каждый из которых соответствует поднесущей частоте f_k .

Демодулятор вычисляет коэффициенты

$$C_{n,k} = \frac{1}{T_S} \int_{nT_S}^{(n+1)T_S} s(t) g_k(t) dt. \quad (10.7)$$

Структурная схема модулятора OFDM-сигнала показана на рис. 10.8, а, а демодулятора — на рис. 10.8, б, где для простоты не показаны фильтры в обоих устройствах.

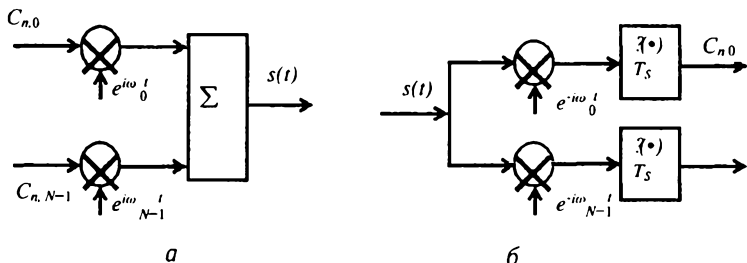


Рис. 10.8. Модулятор (а) и демодулятор (б) OFDM-сигнала

Выбирая прохождение сигнала, описываемого уравнениями (10.4) и (10.6), при скорости передачи в N раз более высокой, чем скорость передачи символов поднесущей $1/T_S$ и принимая $f_0 = 0$ для OFDM-фрейма, можем записать

$$F_n(m) = \sum C_{n,k} g_k(t - nT_S) \Big|_{t=(n+\frac{m}{N})T_S}, \quad m = 0, \dots, N-1, \quad (10.8)$$

что дает

$$F_n(m) = e^{i2\pi f_0 T_S \frac{n}{N}} \sum_{k=0}^{N-1} C_{n,k} e^{i2\pi k \frac{m}{N}} = N \cdot \text{IDFT}\{C_{n,k}\}. \quad (10.9)$$

Отметим разницу между сигналами OFDM и FDM (*frequency division multiplex*). На рис. 10.9 показаны спектры сигналов OFDM и FDM в зависимости от числа несущих.

FDM представляет собой технологию, при которой передается многочастотный сигнал по одному пути передачи информации. При этом каждый сигнал распространяется внутри уникальной полосы частот вблизи несущей, которая модулируется данными (текст, голос, видео и др.). Рассмотрим спектральную плотность мощности для двух систем с бинар-

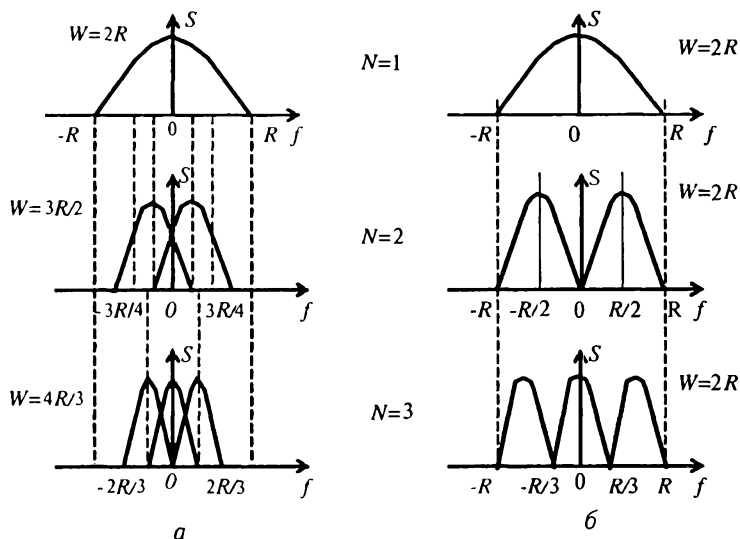


Рис. 10.9. Спектральная плотность мощности OFDM (а) и FDM (б) сигналов для разных величин числа N дискретов символа

ной фазовой модуляцией (BPSK — *Binary Phase Shift Keying*) всех поднесущих. Примем, что поток данных передается со скоростью R . Из этого рисунка легко видеть, что OFDM-сигнал требует меньшую полосу, когда число несущих увеличивается, и в пределе мы имеем:

$$\lim_{N \rightarrow \infty} W = \lim_{N \rightarrow \infty} \frac{N+1}{N} R = R = \frac{N}{T_s} \quad (10.10)$$

Это получается в результате имеющегося перекрытия спектров, разделение которых реализуется благодаря их ортогональности в соответствии с выражением (10.7).

Две боковых полосы OFDM-сигнала можно ограничить учетом главного лепестка его частотного спектра и ограничить полосой $W = N/T_s$

10.6. Уменьшение отношения пикового значения OFDM-сигнала к его среднему значению

В системах с OFDM-сигналами используется обратное быстрое преобразование Фурье сложного временного сигнала.

Важнейшим преимуществом OFDM-сигнала перед модуляцией с одной несущей является простота подавления межсимвольной интерференции (*ISI*) и фединга, не прибегая к тонким методам выравнивания спектра. Однако появляются ложные пики высокой амплитуды в композитном временном сигнале от синфазного сложения амплитуд различных спектральных составляющих. Мгновенная мощность этих пиков является очень высокой по сравнению со средней мощностью сигнала, является сильным мешающим фактором и считается одним из самых важных недостатков OFDM-сигналов.

Практические системы не используют весь диапазон изменения сигнала. Поэтому динамический диапазон системы оказывается меньше максимального уровня сигнала. К настоящему времени разработано по крайней мере три класса методов снижения отношения пикового значения мощности сигнала к средней мощности, каждый из которых имеет свои достоинства и недостатки. Это методы блочного кодирования, *clip effect transformation* и вероятностные.

10.7. Блочное кодирование в OFDM

Этот класс методов кодирования ограничивает множество возможных сигналов, которые могут быть переданы. Только сигналы с уровнем ниже определенного предела могут быть переданы. Общий анализ показывает, что только при ограниченной избыточности сигналов удастся достичь этой цели.

Рассмотрим пример BPSK модуляции с $N = 4$. В этом случае выходной сигнал модулятора описывается выражением

$$s(t) = \sum_{n=1}^{n=N} d_n(t) e^{j(2\pi f_n t + \varphi_n)}, \quad (10.11)$$

где N — число несущих, d_n , f_n , φ_n — амплитуда, частота и начальная фаза n -й несущей. Допустим, что начальная фаза равна нулю. Огибающая мощности композитного сигнала определяется произведением $s(t)s^*(t)$. Если мощность индивидуальной несущей нормализовать к 1 Bm , тогда максимум $PEP = N^2 Bm$.

PEP в течение периода символа зависит от данных d_n , которыми модулируется несущая, или данных слова d , которое является N -битовым двоичным числом. Огибающая мощно-

сти сигнала с четырьмя несущими, как функция времени для всех возможных 4 *бит* данного слова d , увеличивается последовательно от 0 до 15.

Для повышения эффективности использования полосы радиоканала для модуляции несущих применяют методы модуляции с большим числом позиций, например 16 или 64-QAM, получая при этом значения скорости 4 и 6 *бит*/(с Гц) в соответствии с требованиями ITU-R. Так как при OFDM передаваемая информация разбита на большое число низкоскоростных каналов, то длительность тактового интервала для каждой несущей намного больше типичной задержки отраженных сигналов при многолучевом распространении. Это преобразует подверженный селективному федингу широкополосный канал системы с одной несущей в большое число независимо федингующих узкополосных каналов с частотным разделением.

Более того, при OFDM небольшая группа несущих может быть полностью подавлена без потерь принимаемой информации. Для этого вводят корректирующее кодирование данных в сочетании с частотным и временным перемежением. При таком комплексном построении модема схему модуляции часто называют кодированным OFDM или COFDM. С точки зрения физических процессов модуляции и передачи радиосигналов оба эти понятия равнозначны.

Метод OFDM относится к классу методов параллельной передачи данных. В классической системе с параллельной передачей данных выделенная полоса частот канала разбивается на N неперекрывающихся частотных субканалов, для разделения которых используются фильтры. При этом возникают недостатки:

- неэффективное использование полосы, так как полосы разделительных фильтров должны быть не уже $(1 + \alpha)B_N$ (B_N — полоса частот Найквиста, α — коэффициент скругления спектра для каждого подканала);
- сложность изготовления и взаимного согласования большого числа фильтров.

Частично эти недостатки можно преодолеть методом квадратурной модуляции M-QAM с шахматной расстройкой, когда соседние спектры перекрываются по уровню — 3 дБ, а для

разделения сигналов в субканалах применяется чередующийся сдвиг данных на половину тактового интервала.

Этот метод обеспечивает получение плоского группового спектра с пониженными требованиями к разделительным фильтрам. Для устранения недостатков, связанных с низкой эффективностью системы и сложностью технической реализации, и был найден метод, получивший название OFDM.

Для получения сигнала OFDM не требуется много отдельных генераторов несущих с соответствующими узлами преобразования сигналов и субканальными разделительными фильтрами.

Спектры всех субканалов перекрываются между собой, но их разделение обеспечивается не фильтрацией, а специальной обработкой модулирующих данных в основной полосе частот.

В модеме реализована идея прямого синтеза группового спектра OFDM с использованием сигнального процессора, работающего по алгоритму ОБПФ. Обычно при практической реализации модулятора спектр сигнала OFDM формируется на некоторой промежуточной частоте, а затем конвертируется в полосу радиосигнала.

Преобразования сигналов в демодуляторе приемника OFDM отличаются использованием прямого БПФ.

Ортогональность субканалов при выделении их в приемнике посредством БПФ может быть обеспечена только в случае отсутствия межсимвольной интерференции между несущими. На практике эти условия не выполняются из-за искажений, возникающих в канале. Так как спектры несущих в субканалах теоретически не ограничены по полосе, то любое их ограничение, например общим канальным фильтром или искажение за счет многолучевого распространения, приводит к перераспределению энергии между субканалами и, следовательно, к возникновению межсимвольных искажений.

Для борьбы с этим явлением используется простое решение — общую длительность символа OFDM T_s увеличивают и часть ее в начале символа отводят под защитный интервал T_G . Поэтому полезная длительность символа $T_f = T_s - T_G$. В пределах защитного интервала передается как бы возвращенная назад во времени копия последующей части символа.

Процесс формирования полного символа OFDM, включающего защитный интервал, схематично показан на рис. 10.10.

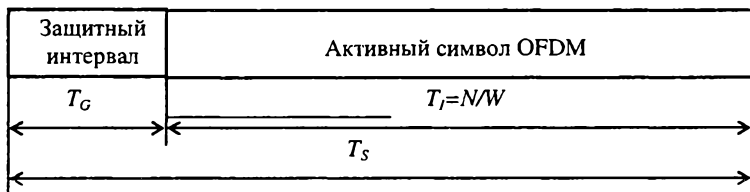


Рис. 10.10. Формирование символа OFDM с защитным интервалом

Правильный выбор длительности защитного интервала позволяет в определенных пределах устранить помехи, вызываемые эхо-сигналами. На рис. 10.11 показаны временные интервалы для основного сигнала и двух его эхо-сигналов. Задержка эхо-сигнала 1 находится в допустимых пределах, и переходный процесс из-за стыка двух символов приходится на защитный интервал основного сигнала, не искажая его полезную часть.

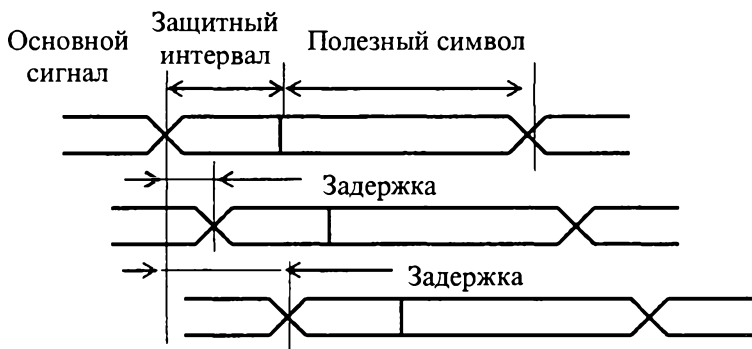


Рис. 10.11. Временные интервалы основного и двух эхо-сигналов

Напротив, если элемент 2 задержан свыше допустимых пределов, то его переходная зона приходится на полезную часть основного сигнала, т. е. защита не обеспечивается.

Так как введение защитного интервала снижает пропускную способность системы, то обычно на практике его длительность не превышает одной четверти от длительности символа. Хорошим способом увеличения длительности защитно-

го интервала и поддержания высокой пропускной способности является увеличение числа несущих и, тем самым, увеличение длительности символа при относительно уменьшенной длительности защитного интервала. В системном плане радикальным средством борьбы со всеми искажениями и помехами является переход к схеме с кодированием — COFDM.

10.8. Влияние ошибок частотной синхронизации

Значительной проблемой OFDM является ее чувствительность к ошибкам частотной синхронизации, так как демодуляция принятого сигнала, выполненная в этих условиях, приводит к высоким значениям ошибки битовой скорости (BER — *Bit Error Rate*). Это приводит к потере ортогональности между поднесущими, проявляясь в высоком уровне интерференции между поднесущими (ICI — *Inter Carrier Interference*) вследствие возникающего при этом повороте фазы принятого вектора данных. Частотная ошибка формируется в основном двумя главными причинами: уходом частоты локальных осцилляторов и доплеровским сдвигом. Любая разница между частотами приемника и передатчика проявляется в смещении частот, которое обычно компенсируется системами автоматической подстройки частоты. Движение приемника или передатчика приводит к доплеровскому сдвигу частоты Δf принимаемого сигнала, который определяется выражением:

$$\Delta f = \pm f_0(v/c),$$

где f_0 — частота источника, v — скорость движения приемника относительно источника сигнала, c — скорость света. Например, при $f_0 = 1 \text{ ГГц}$, $v = 120 \text{ км/ч} = 33,4 \text{ м/с}$ доплеровский сдвиг $\Delta f = (10^9 \cdot 33,4) / 3 \cdot 10^8 \approx 111 \text{ Гц}$.

Такой частотный сдвиг в условиях многолучевого распространения OFDM-сигнала проявляется в возникновении частотной модуляции поднесущих, которая имеет случайный характер вследствие большого числа всевозможных переотражений принимаемого сигнала. Любое изменение частоты регистрируется также в повороте фазы вектора принимаемого сигнала, причем чем больше доплеровский сдвиг, тем больший поворот фазы он вызывает.

В мобильных системах проблема обстоит еще хуже, так как передачи сигналов от каждого пользователя могут иметь различные частотные сдвиги. Даже если каждый пользователь полностью будет обеспечен синхронной работой своего приемника с базовой станцией, значительная разница частот для каждого пользователя обусловит значительный доплеровский сдвиг. Сдвиг частоты в монопользовательских системах OFDM имеет незначительные проблемы, которые устраняются усложнением схемы приемного устройства. Однако в мультипользовательских системах OFDM устранение частотных ошибок является непростой задачей.

Термин «модем» — это аббревиатура от «*МО*дуляция/*ДЕ*-*МО*дуляция». Модем позволяет двум устройствам, генерирующим цифровую информацию (например, компьютерам), взаимодействовать через общественную коммутируемую телефонную сеть. Сеть PSTN оптимизирована и ограничена частотами для передачи человеческого голоса и может переносить только звуки. Модем должен транслировать информацию из цифрового компьютерного формата и последовательность звуков, пересылаемых по телефонным линиям. Когда четкие звуки поступают в точку назначения, они демодулируются и преобразуются обратно в цифровую информацию для компьютера или другого цифрового устройства.

Все современные модемы используют сжатие и коррекцию ошибок. Алгоритмы сжатия позволяют поднять скорость пропускания от 2 до 4 раз, по сравнению с обычной скоростью обмена. Проверка ошибок служит для исследования входящего потока цифровых данных на наличие ошибок при передаче и для организации запроса на повторную пересылку ошибочного кадра.

В 1984 г. существенный прогресс в технологии модемов и в стандартах нашел отражение в «Рекомендации V.32», добавившей подавление эхо-сигнала и решетчатое кодирование (*trellis coding*). Главным было решетчатое кодирование, проложившее дорогу технологиям обеспечения *ранней* (упреждающей) *коррекции ошибок* (*Forward Error Correction*, FEC). Вскоре ITU выпустило «Рекомендацию V32bis», содержащую улучшенный вариант решетчатого кодирования и увеличивавшую скорость обмена до 14,4 Кбит/с.

Эти рекомендации усилили желание специалистов достичь еще большей скорости модемов. В 1989 и 1990 гг. разверну-

лась работа над стандартом V.34. Признавая улучшения в базовом оборудовании телефонных сетей, таком как коммутаторы, первоначальной целью V.34 стал переход от скорости 19,2 Кбит/с к 24,0 Кбит/с, а затем к 28,8 Кбит/с. Версия V.34 от 1996 г. определяет скорости модемов до 33,6 Кбит/с. Последние модели модемов достигают 10 бит на один герц полосы пропускания, что близко подходит к установленным Шенноном теоретическим пределам.

11.1. Классификация протоколов

Передача данных организуется на основе набора протоколов, каждый из которых устанавливает правила взаимодействия связываемых устройств. Протоколы, используемые в модемах, делятся на четыре основные группы:

- протоколы модуляции и передачи данных;
- протоколы коррекции ошибок;
- протоколы сжатия передаваемых данных;
- протоколы связи DTE (*Data Terminal Equipment*) и DCE (*Data Communications Equipment*).

Первые три группы относятся только к связи DCE-DCE, последняя — только к связи DCE-DTE.

Первая группа протоколов устанавливает правила вхождения модемов в связь, ее поддержания и разрыва, параметры аналоговых сигналов, правила кодирования и модуляции. Эти протоколы непосредственно относятся к сигналам, передаваемым по междемной аналоговой линии связи. Соединение двух модемов возможно только в случае поддержки ими каких-либо общих или совместимых протоколов этой группы. В семиуровневой иерархии протоколов связи OSI эта группа протоколов имеет уровень 1 (физический) и формирует канал цифровой связи в реальном времени, однако не защищенный от ошибок передачи.

Протоколы физической связи могут быть симплексными (*simplex*) — реализующими в каждый момент времени передачу только в одну сторону, и дуплексными (*duplex*) — с одновременной двунаправленной передачей. Чаще всего применяются дуплексные протоколы, которые могут быть симметричными, когда скорости передачи в обоих направлениях равны, и несимметричными, когда скорости различаются.

Несимметричный дуплекс применяется для повышения скорости передачи в одну сторону за счет ее снижения в обратную сторону, когда поток передаваемых данных имеет выраженную асимметрию.

Для определения направления передачи в физическом канале используются понятия вызывающего (инициирующего соединение) и отвечающего модемов; направление передачи определяется со стороны вызывающего модема.

Вторая группа устанавливает правила обнаружения и коррекции ошибок, возникающих на этапе передачи с помощью протоколов первой группы. Эти протоколы имеют дело только с цифровой информацией; для проверки целостности информации она разделяется на блоки (пакеты), снабжаемые контрольными избыточными кодами (*Cyclic Redundancy Check*, CRC). При несовпадении контрольного кода на приемном конце переданный пакет считается ошибочным и запрашивается его повторная передача. Эта группа протоколов формирует из ненадежного физического канала надежный (защищенный от ошибок) канал более высокого уровня, однако это приводит к потере связи в реальном времени и дается ценой определенных накладных расходов. В модели OSI эта группа соответствует уровню 2 (канальный).

Третья группа устанавливает правила сжатия передаваемых данных путем уменьшения их избыточности. При этом на передающем конце происходит их анализ и упаковка, а на приемном — распаковка в исходный вид. Сжатие позволяет повысить скорость передачи сверхфизической пропускной способности канала за счет уменьшения объема реально передаваемых данных. Реализация сжатия также требует некоторых накладных расходов на анализ информации и формирование пакетов; в случае неэффективного сжатия скорость передачи может оказаться ниже скорости физического канала.

Последняя группа протоколов задает правила взаимодействия DCE и DTE. Они подразделяются на физические, касающиеся кабелей, разъемов и сигналов взаимодействия, и информационные, относящиеся к формату и смыслу передаваемых сообщений. Посредством этих протоколов реализуется общение DTE и DCE во время подготовки к вхождению в связь, организации вызова и ответа, а также в процессе самого обмена данными.

Модемы также можно классифицировать в соответствии с реализованными в них протоколами. Все протоколы, регламентирующие те или иные аспекты функционирования модемов, могут быть отнесены к двум большим группам: *международные и фирменные*.

Протоколы международного уровня разрабатываются под эгидой ITU-T и принимаются им в качестве рекомендаций. Все рекомендации ITU-T относительно модемов относятся к серии V.

Фирменные протоколы разрабатываются отдельными компаниями — производителями модемов, с целью преуспеть в конкурентной борьбе. Часто фирменные протоколы становятся стандартными протоколами де-факто и принимаются частично либо полностью в качестве рекомендаций ITU-T, как это случилось с рядом протоколов фирмы *Microcom*. Наиболее активно разработкой новых протоколов и стандартов занимаются такие известные фирмы, как AT&T, *Motorolla*, *U.S. Robotics*, *ZyXEL* и другие.

С функциональной точки зрения модемные протоколы могут быть разделены на следующие группы:

- протоколы, определяющие нормы взаимодействия модема с каналом связи (V.2, V.25);
- протоколы, регламентирующие соединение и алгоритмы взаимодействия модема и DTE (V.10, V.11, V.24, V.25, V.25bis, V.28);
- протоколы модуляции, определяющие основные характеристики модемов в предназначенных для коммутируемых и выделенных телефонных каналов. К ним относятся такие протоколы, как V.17, V.22, V.32, V.34, HST, ZyX и большое количество других;
- протоколы защиты от ошибок (V.41, V.42, MNP1-MNP4);
- протоколы сжатия передаваемых данных, такие как MNP5, MNP7, V.42bis;
- протоколы, определяющие процедуры диагностики модемов, испытания и измерения параметров каналов связи (V.51, V.52, V.53, V.54, V.56);
- протоколы согласования параметров связи на этапе ее установления (*HandShaking*).

11.2. Протоколы взаимодействия

Для выполнения своих функций модемы должны взаимодействовать как между собой, так и с различного рода оконечным оборудованием. За исключением ранее рассмотренного набора АТ-команд, средствами такого взаимодействия выступают стандарты V.25, V.25bis и V.8.

Рекомендации V.25 и V.25bis определяют интерфейс и функции оборудования автоматического вызова и ответа. Различие этих двух стандартов определяется отличием в используемых линиях интерфейса DTE-DCE. Протокол V.8 описывает процедуры взаимодействия двух модемов на этапе установления связи между ними.

11.3. Протоколы модуляции

Способ модуляции играет основную роль в достижении максимально возможной скорости передачи информации при заданной вероятности ошибочного приема. Предельные возможности системы передачи можно оценить с помощью известной формулы Шеннона, определяющей зависимость пропускной способности C непрерывного канала с белым гауссовским шумом от используемой полосы частот F и отношения мощностей сигнала и шума P_S/P_N :

$$C = \Delta F \log_2 (1 + P_S / P_N),$$

где $P_S = E_b V$ — средняя мощность сигнала; E_b — энергия, затрачиваемая на передачу одного бита информации; V — скорость передачи информации $P_N = N_0/2$ — средняя мощность шума в полосе частот ΔF ; $N_0/2$ — спектральная плотность мощности шума.

Пропускная способность определяется как верхняя граница реальной скорости передачи информации V . Приведенное выше выражение позволяет найти максимальное значение скорости передачи, которое может быть достигнуто в гауссовском канале с заданными значениями ΔF и P_S/P_N .



Пример 11.1. Рассчитать максимально достижимую скорость передачи информации по телефонному каналу связи с нормальным шумом при мощности сигнала на входе модема, на 20 дБ превышающей мощность шума.

Решение. Поскольку полная полоса телефонного канала составляет 3100 Гц и отношение сигнал/шум равно 20 дБ, то максимально достижимая скорость не может превышать $C = 3100 \cdot \log_2 101 \approx 20640$ бит/с.

Вероятность ошибочного приема бита в конкретной системе передачи определяется отношением E_b/N_0 . Из формулы Шеннона следует, что возрастание удельной скорости передачи $V/\Delta F$ требует увеличения энергетических затрат (E_b) на один бит (рис. 11.1).

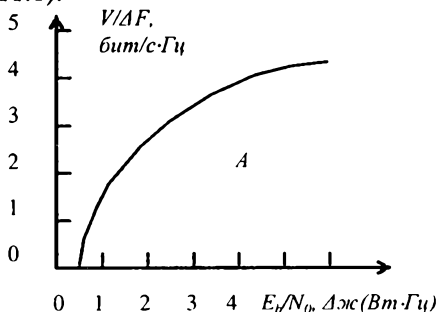


Рис. 11.1. Зависимость удельной скорости передачи информации от отношения сигнал/шум

Любая система передачи может быть описана точкой, лежащей ниже приведенной на рис. 11.1 кривой. Эту кривую часто называют границей или пределом Шеннона. Для любой точки в области *A* можно создать такую систему связи, вероятность ошибочного приема у которой может быть настолько малой, насколько это требуется. История развития как систем связи в целом, так и модемной техники в частности представляет собой непрекращающуюся серию попыток приблизить их к границе Шеннона, сохраняя при этом низкую вероятность ошибочного приема информационного бита (такие системы используют современные способы модуляции и кодирования).

Современные системы передачи данных требуют, чтобы вероятность необнаруженной ошибки была не выше величины $10^{-7} \dots 10^{-12}$. Эти значения обеспечивают протоколы исправления ошибок типа MNPI—MNP4 и V.42, которые будут рассмотрены ниже.

11.4. Протокол V.32bis

Протокол модуляции V.32bis разработан для обеспечения передачи данных со скоростью до 14400 *бит/с* по двухпроводным коммутируемым и выделенным телефонным каналам. Данный протокол принят в качестве стандарта ITU-T в 1991 году. Основные характеристики модемов, поддерживающих данный протокол, следующие:

- дуплексный режим работы по коммутируемым каналам телефонных сетей общего пользования и арендуемым двухпроводным линиям передачи;
- реализация эхоподавления;
- применение КАМ для режимов синхронной передачи со скоростью модуляции 2400 *Бод*;
- частота несущей равна 1800 *Гц*;
- приемник модема должен обеспечивать бесперебойную работу при нестабильности частоты принимаемого сигнала не более ± 7 *Гц*;
- скорости передачи данных:
- 14400, 1200, 9600, 7200 *бит/с* с треллис-кодированием;
- 4800 *бит/с* без кодирования;
- совместимость с модемами V.32 на скоростях 9600 и 4800 *бит/с*;
- обмен управляющими последовательностями и выбор скорости передачи в течение процедуры установления связи;
- процедура смены скорости передачи в течение сеанса связи без разрыва соединения;
- режим асимметричной передачи не поддерживается; другими словами, скорости передачи и приема каждого взаимодействующего модема должны быть одинаковы;
- спектр сигнала ограничен полосой частот от 600 *Гц* до 3000 *Гц*.

Устройство кодирования по протоколу V.32bis показано на рис. 11.2. При скорости передачи 14400 *бит/с* на вход кодера подаются все шесть битов $Q_{1n} - Q_{6n}$ в параллельном коде. При скорости 12000 *бит/с* входная информационная последовательность разделяется на блоки по пять битов $Q_{1n} - Q_{5n}$. Аналогично при скоростях 9600, 7200 и 4800 *бит/с* задей-

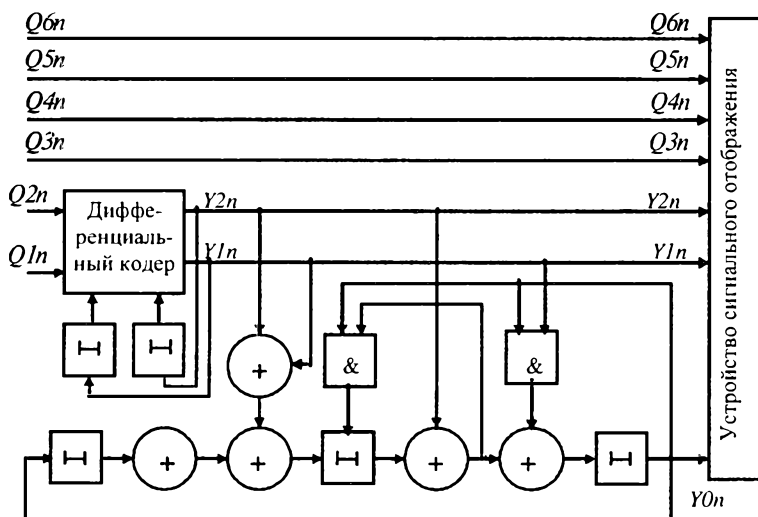


Рис. 11.2. Схема кодирования в модемах протокола v.32bis

ствуются четыре ($Q1n - Q4n$), три ($Q1n - Q3n$) и два ($Q1n - Q2n$) входа соответственно.

Дифференциальные биты $Y1n$ и $Y2n$ используются в качестве входных для систематического сверточного кодера, который генерирует избыточный бит $Y0n$. Этот избыточный бит и шесть информационных бит $Y1n$, $Y2n$, $Q3n$, $Q4n$, $Q5n$, $Q6n$ поступают на устройство сигнального отображения, которое формирует элементы сигнального созвездия, представленного на рис. 11.3.

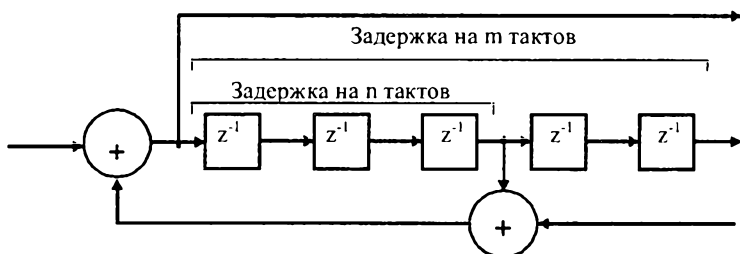


Рис. 11.3. Структурная схема скремблера

Двоичные числа на рис. 11.3 соответствуют последовательности битов Y_{0n} , Y_{1n} , Y_{2n} , Q_{3n} , Q_{4n} , Q_{5n} , Q_{6n} , а обозначения A , B , C , D — синхронизирующим сигнальным элементам.

Стандарт V32bis — это относительно недавно появившееся расширение V.32 со скоростью передачи данных 14 400 *бит/с*. В нем применяется амплитудно-фазовая модуляция (АФМ) со скоростью 2400 *Бод*, когда каждому состоянию соответствует шесть бит. Благодаря ячеистому кодированию связь получается весьма надежной. Протокол V32bis обеспечивает дуплексную связь. Если качество телефонной линии невысокое, модемы переключаются в обычный режим V.32. Несмотря на свою новизну, этот стандарт, благодаря своей высокой производительности и помехоустойчивости, уже становится общепринятым при работе в современных высококачественных телефонных сетях.

Входным сигналом модема является последовательный битовый поток данных, для кодирования и модуляции которого в протоколе V.32bis используется следующая последовательность процедур.

Входной битовый поток данных проходит через *скремблер*, структурная схема которого показана на рис. 11.3.

Различные алгоритмы скремблирования отличаются количеством слагаемых, дающих цифру результирующего кода, и сдвигом между слагаемыми. Так, в сетях ISDN при передаче данных от сети к абоненту используется преобразование со сдвигами в 5 и 23 позиций, а при передаче данных от абонента в сеть — со сдвигами 18 и 23 позиций. Существуют и более простые методы борьбы с последовательностями единиц, также относящиеся к классу скремблирования.

В нашем случае рассмотрим скремблер, который представляет собой рекурсивный фильтр, операции сложения в котором производятся по модулю два.

Порядок фильтра $m = 23$, положение отвода: для вызывающего модема $n = 18$, для отвечающего модема $n = 5$. Назначение скремблера — сделать сигнал псевдослучайным.

Скремблированный битовый поток разбивается на символы длиной по 6 бит. Два первых (во времени) бита сформированного символа подвергаются *дифференциальному кодированию*. Это кодирование математически описывается следую-

щим образом: $Y(n) = (Q(n) - Y(n - 1)) \bmod 4$. Здесь $Q(n)$ — целое число, представляемое двумя кодируемыми битами, $Y(n)$ — результирующее двухбитовое целое число, $Y(n - 1)$ — результат дифференциального кодирования предыдущего символа. Дифференциально-кодированные биты $Y_1(n)$ и $Y_2(n)$ подаются на *сверточный кодер*, структурная схема которого приведена на рис. 11.2. Кодер формирует избыточный бит $Y_0(n)$, пропуская на выход биты $Y_1(n)$ и $Y_2(n)$ без изменений. Следует обратить внимание на то, что избыточный бит $Y_0(n)$ зависит только от *предыдущих*, но не от текущих значений входных битов $Y_1(n)$ и $Y_2(n)$.

Избыточный бит $Y_0(n)$, дифференциально-кодированные биты $Y_1(n)$ и $Y_2(n)$ и четыре бита Q_3-Q_6 , не подвергавшиеся дифференциальному кодированию, образуют символ, передаваемый в течение одного символьного такта (символьная скорость — 2400 *символов/с*). Этот символ отображается на комплексную плоскость с использованием созвездия квадратной манипуляции, показанного на рис. 11.4.

При скорости передачи 14400 бит/с на вход кодера подаются все шесть битов Q_1n-Q_6n в параллельном коде. При скорости 12000 *бит/с* входная информационная последовательность разделяется на блоки по пять битов Q_1n-Q_5n . Аналогично, при скоростях 9600, 7200 и 4800 *бит/с* задействуются четыре (Q_1n-Q_4n), три (Q_1n-Q_3n) и два (Q_1n-Q_2n) входа соответственно.

На скорости передачи 12000 *бит/с* входной проскремблированный поток данных делится на группы по пять бит. Процесс дифференциального кодирования и кодирования сверточным кодом принципиально ничем не отличается от кодирования при скорости 14400 *бит/с*. При скорости 12000 *бит/с* формируются элементы сигнального созвездия, приведенного на рис. 11.5.

Двоичные числа на рис. 11.5 соответствуют последовательности шести битов $Y_0n, Y_1n, Y_2n, Q_3n, Q_4n, Q_5n$, а обозначения A, B, C, D , как и ранее, соответствуют синхронизирующим сигнальным элементам.

При скорости передачи 9600 *бит/с* входной проскремблированный поток данных делится уже на блоки по четыре бита Q_1n, Q_2n, Q_3n, Q_4n . В результате этого схема кодера V.32bis при скорости 9600 *бит/с* соответствует кодеру V.32. Сигнальное

созвездие соответствует протоколу V.32 при той же скорости передачи.

Двоичные числа на диаграмме соответствуют последовательности пяти битов $Y_0n, Y_1n, Y_2n, Q_3n, Q_4n$, поступающих на вход устройства сигнального отображения.

При скорости передачи 7200 бит/с входной проскремблированный поток данных делится на блоки по три бита Q_1n, Q_2n, Q_3n . Пространственная сигнальная диаграмма для такой скорости передачи приведена на рис. 11.6.

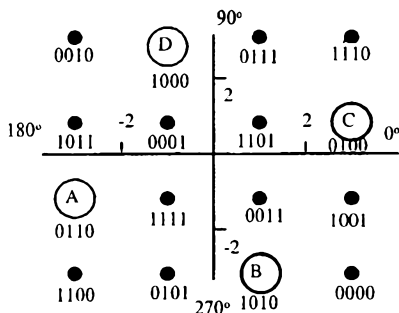


Рис. 11.6. Сигнальное созвездие для скорости 7200 бит/с

В этом случае двоичные числа соответствуют последовательности четырех бит Y_0n, Y_1n, Y_2n, Q_3n , поступающих на вход устройства сигнального отображения.

При скорости 4800 бит/с скремблированный входной поток данных разбивается на блоки по два бита Q_1n и Q_2n , которые и поступают на вход относительного кодера.

С выхода относительного кодера биты Y_1n и Y_2n отображаются в передаваемые сигнальные элементы. Таким образом, при скорости 4800 бит/с кодирования избыточным сверточным кодом не происходит.

11.5. Протоколы исправления ошибок

Одним из первых протоколов исправления ошибок стал протокол MNP (*Microcom Networking Protocol*), разработанный фирмой *Microcom*. Он оказался настолько удачным, что пре-

терпел девять модификаций и расширений, которые получили название Классов протоколов MNP. Классы 1–4 обеспечивают исправление ошибок, классы 6, 9, 10 — кроме исправления ошибок выполняют и другие функции.

MNP1 используется для асинхронного побайтного полудуплексного обмена данными. Он был разработан для того, чтобы устройства с минимальными аппаратными ресурсами могли осуществлять контроль ошибок. Из-за своей чрезвычайно низкой эффективности протокол в современных модемах больше не используется.

MNP2 позволяет исправлять канальные ошибки при асинхронной полнодуплексной передаче данных и реализуется в модемах с микропроцессорным управлением. Протокол также характеризуется повышенной избыточностью, поскольку в нем при асинхронном режиме передачи в каждый передаваемый символ включаются стартовые и стоповые биты.

MNP3 обеспечивает обмен данными между модемами по протоколу SDLC (*Synchroneuse Data Link Control*) в синхронном режиме, в то время как обмен данными с компьютером остается асинхронным. Из байт данных, принимаемых от DTE, формируются блоки данных (кадры), называемые в терминах MNP *пакетами*. Каждый пакет передается как один синхронный кадр второго канального уровня модели OSI. Скорость передачи информации при использовании MNP3 повышается за счет того, что уже не требуется передавать дополнительные стартовые и стоповые биты для каждого байта.

MNP4 предусматривает возможность изменения размера пакета в процессе процедуры согласования параметров передачи, называемой также процедурой *адаптивной сборки пакетов* (*Adaptive Packet Assembly*). Пакет может содержать 32, 64, 128, 192 или 256 *байт*. При большом уровне шумов передаются пакеты меньших размеров. В результате этого увеличивается вероятность безошибочной передачи пакета данных. По высококачественным каналам пересылаются пакеты больших размеров; при этом уменьшается количество избыточной служебной информации. Управление размером пакета со стороны пользователя часто возможно при помощи АТ-команды \An.

Протокол MNP4 позволяет повысить скорость передачи за счет оптимизации *фазы* (режима) *передачи данных* (*Data Phase*

Optimization), поскольку не требует передавача неизменяющегося заголовка для каждого нового пакета.

Благодаря этому большая часть информационной пропускной способности канала используется для передачи данных.

MNP6 рассчитан на работу со скоростями от 300 до 9600 *бит/с*. Модем начинает работу на скорости 2400 *бит/с* и затем изменяет ее в зависимости от типа удаленного модема. Этот протокол предусматривает возможность автоматического переключения из полудуплексного режима в дуплексный и обратно.

MNP9 обеспечивает совместимость с протоколом модуляции V.32 и предусматривает процедуру сжатия, а также повышает эффективность передачи за счет реализации режима селективного повтора искаженных пакетов (ARQ типа SR).

MNP10 предназначен для обеспечения передачи данных при неблагоприятных или изменяющихся условиях на линии связи, характерных, например, для сотовых систем связи. Протокол включает возможность многократных попыток установления связи, динамическую подстройку уровня передачи и размера передаваемого пакета. MNP10 также имеет возможность изменения скорости передачи не только в сторону ее уменьшения, но и в сторону увеличения.

11.6. Протоколы сжатия данных

Как известно, применение сжатия данных позволяет более эффективно использовать емкость дисковой памяти. Не менее полезно применение сжатия при передаче информации в любых системах связи. В последнем случае появляется возможность передавать значительно меньшие (как правило, в несколько раз) объемы данных, следовательно, требуются значительно меньшие ресурсы пропускной способности каналов для передачи той же самой информации. Выигрыш может выражаться в сокращении времени занятия канала и, соответственно, в значительной экономии арендной платы.

Известные методы сжатия направлены на снижение избыточности, вызванной как неравной априорной вероятностью символов, так и зависимостью между порядком поступления символов. В первом случае для кодирования исходных символов используется неравномерный код. Часто появляющиеся

символы кодируются более коротким кодом, а менее вероятные (редко встречающиеся) — более длинным кодом.

Устранение избыточности, обусловленной корреляцией между символами, основано на переходе от кодирования отдельных символов к кодированию групп этих символов. За счет этого происходит укрупнение алфавита источника, так как число N тоже растет. Общая избыточность при укрупнении алфавита не изменяется. Однако уменьшение избыточности, обусловленной взаимными связями символов, сопровождается соответствующим возрастанием избыточности, обусловленной неравномерностью появления различных групп символов, т. е. символов нового укрупненного алфавита. Происходит как бы конвертация одного вида избыточности в другой.

Таким образом, процесс устранения избыточности источника сообщений сводится к двум операциям — декорреляции (укрупнению алфавита) и кодированию оптимальным неравномерным кодом.

11.7. Сжатие данных в протоколах MNP

Расширяемость MNP при сохранении совместимости с существующими реализациями ярко продемонстрирована в его поддержке Рекомендации ITU-T V.42bis.

В процессе установления соединения передатчик и приемник «оговаривают» использование сжатия данных в процессе. Это выполняется с помощью параметра 9 или 14 блока PDU LR. Параметр 9, который специфицирует сжатие данных MNP5 или MNP7, был расширен, чтобы обеспечить «краткую» форму спецификации V.42bis. Параметр 14 является новым параметром, применяемым для детализации особенностей V.42bis, используемого в данном канале.

Если существует возможность поддерживать MNP5 и (или) MNP7 и V.42bis, передатчик может включить как параметр 9 (сжатие MNP), так и параметр 14 (сжатие V.42bis). Ответственность за выбор типа сжатия данных, который будет использоваться, в этом случае несет приемник. Он возвращает PDU LR, который указывает выбранный тип сжатия данных. Если передатчик и приемник поддерживают несколько методов сжатия, то приемник делает свой выбор в соответствии со следующим приоритетом.

Приемник не включает информацию о поддержке V.42bis в свой PDU LR, если он не принял запрос на V.42bis в LR от передатчика. Если передатчик включил такой запрос в свой PDU LR, но не получил подтверждения, он отказывается от использования сжатия по протоколу V.42bis.

11.8. Протоколы передачи файлов

Наиболее часто используемой функцией коммуникационного программного обеспечения является функция передачи файлов. Она осуществляется с помощью специальных протоколов передачи файлов. Выбор и использование протокола передачи файлов может производиться пользователем в явном виде, как это делается в терминальных программах, так и в неявном, например в игровых программах, поддерживающих модемную связь.

Основными задачами протоколов передачи файлов являются:

- обеспечение безошибочной передачи данных;
- управление потоком передаваемых данных;
- передача вспомогательной информации;
- защита соединения.

Первые протоколы передачи файлов появились задолго до модемов, поддерживающих аппаратное исправление ошибок. По этой причине задача обеспечения безошибочной передачи по сегодняшний день остается одной из основных. Для ее реализации применяются в основном те же методы, что и в современных протоколах исправления ошибок. Передаваемые данные разбиваются на блоки (кадры) определенной длины, и в каждый из них включается проверочная комбинация (CRC) для обнаружения ошибок. Эта комбинация формируется по определенному правилу на основе передаваемых информационных битов блока. На приемной стороне производится повторное вычисление проверочной комбинации по тому же правилу и сравнение ее с принятой. При совпадении проверочных комбинаций принимающая сторона посылает подтверждение правильного приема блока (*Acknowledgement*, ACK), а при несовпадении — запрос на повторную передачу данного блока (*negative Acknowledgement*, NACK). Таким образом реализуется механизм автоматического запроса повторения

(*Automatic Repeat Request*, ARQ), аналогичный механизму ARQ в протоколах исправления ошибок типа MNP классов 1–4 и V.42. При этом ARQ также может быть старт-стопного типа (*Stop and Wait*, SAW), с возвратом на N шагов (Go Back N, GBN) или селективного повторения (*Selective Repeat*, SR).

При использовании ARQ типов GBN и SR непрерывная передача неподтвержденных блоков данных может привести к перегрузке буферов как приемника, так и передатчика. Чтобы этого не происходило, используется управление потоком передаваемых данных.

Перед непосредственной передачей файла необходимо установить соединение на уровне канала данных (уровень 2 модели OSI), передать информацию об имени файла, его размере, дате последней его модификации и т. п., а после передачи — произвести разъединение канала данных. Все это осуществляется при помощи вспомогательной служебной информации, передаваемой по каналу связи.

В последние годы в функции протоколов передачи файлов включают защиту соединения, например проверку пароля.

Среди протоколов, рассчитанных на отсутствие аппаратной защиты от ошибок, можно выделить широко распространенные протоколы XModem, XModem-CRC, XModem-1K, YModem, Kermit, ZModem и ряд других.

Если же применяются модемы с аппаратной коррекцией ошибок (поддерживающие протоколы типа MNP или V.42), то предпочтительнее использовать протоколы передачи файлов типа YModem-g и ZModem. В этом случае исключается потеря времени на повторный запрос данных, переданных с ошибками. Протокол Zmodem допускает оба варианта применения.

Известны специализированные протоколы, предназначенные для определенных служб и сетей, — такие как SEALink, Telnet, CompuServe Quick B. Практически все они являются модификациями протокола XModem.

11.9. Фирменные протоколы модуляции

11.9.1. Протокол V.32terbo

Протокол V.32terbo был разработан фирмой AT&T. Он опубликован и доступен разработчикам модемов. Кроме мо-

демов фирмы AT&T данный протокол реализован в большинстве модемов корпорации MultiTech и некоторых модемах фирмы U.S. Robotics. Протокол V.32terbo предусматривает более развитые технические решения, чем те, которые предусмотрены в V.32bis: технология эхоподавления, модуляция с решетчатым кодированием (СКК). В данном протоколе скорость модуляции принята равной 2400 *Бод*, а несущая частота 1800 *Гц*. В отличие от V.32bis, за счет применения СКК-256 и СКК-512 обеспечиваются скорости передачи 16800 и 19200 *бит/с*.

Модемы V.32terbo предъявляют жесткие требования к качеству используемого телефонного канала. Так, для устойчивой работы на скорости 19200 *бит/с* необходимо, чтобы отношение сигнал/шум в канале было не менее 30 дБ.

11.9.2. Протоколы ZyX, ZyCELL

Протокол ZyX разработан корпорацией ZyXEL Communications и реализован в ее собственных модемах. Данный протокол, так же как и V.32terbo, обеспечивает скорости передачи 16800 и 19200 *бит/с*. Предусмотрено применение технологии эхоподавления и модуляции решетчатым кодом несущего колебания с частотой 1800 *Гц*. Скорость модуляции 2400 *Бод* сохраняется лишь для скорости передачи 16800 *бит/с*. Скорость 19200 *бит/с* обеспечивается за счет повышения модуляционной скорости до 2743 *Бод* при сохранении 256-позиционной СКК для обеих скоростей передачи. Такое решение позволило снизить требования к качеству канала связи на 2,4 дБ. Однако расширение спектра сигнала за счет увеличения скорости модуляции может негативно сказаться на качестве связи в каналах с большой неравномерностью амплитудно-частотной характеристики.

Протокол ZyCELL разработан специально для передачи информации по телефонным каналам низкого качества (с высоким уровнем помех), какими также являются каналы сотовых сетей связи.

Технология ZyCELL включает в себя 5 процедур физического уровня и 2 процедуры канального, улучшающие реализацию протокола исправления ошибок V.42. Поэтому протокол

ZyCELL является не только протоколом модуляции, но и протоколом исправления ошибок.

Процедуры физического уровня:

- быстрая повторная синхронизация;
- кодирование с прямым исправлением ошибок;
- автоматическое регулирование уровня передачи;
- динамический выбор скорости передачи;
- улучшенная процедура вхождения в связь.

Процедуры канального уровня:

- модифицированная процедура адаптивного изменения размера передаваемого кадра;
- процедура селективного повтора кадра (ARQ типа SR или SREJ).

При перемещении мобильного телефона из одной соты в другую происходит переключение с одного радиоканала на другой. При изменении расстояния от сотового радиотелефона до базовой станции производится автоматическое переключение мощности передатчика. В результате таких переключений радиосвязь, а значит и несущая частота модема, прерывается на 0,2...1,2 с. Обычный модем реагирует на такой перерыв связи процедурой повторного соединения, которая обычно продолжается около 10 с, или даже рассоединением. При смене сот или изменении мощности сигнала протокол ZyCELL предусматривает применение процедуры быстрой повторной синхронизации (*Fast ReSynchronization*, FRS).

Большой проблемой при обеспечении безошибочной передачи данных через сотовые системы связи являются замирания сигнала, вызванные его многократным отражением. Из-за различия фаз сигналов, пришедших к приемнику разными путями, возникает интерференция, которая в зависимости от места расположения приемника влияет на мощность принятого сигнала. В результате колебаний амплитуды несущей при передаче данных возникают ошибки и нарушается режим работы модема. Для уменьшения влияния эффекта замирания сигнала в протоколе ZyCELL применяется кодирование с прямым исправлением ошибок (*Forward Error Correction*, FEC), названное фирмой сотовым FEC (*Cellular FEC*, CFEC). Применение такого кода позволяет на приемной стороне исправить ряд ошибок, не запрашивая повторной передачи искаженных кадров. Повторная передача запрашивается только в

том случае, когда корректирующей способности кода FEC для исправления всех ошибок оказывается недостаточно. Таким образом, протокол ZyCELL предусматривает гибридное кодирование как кодом с обнаружением ошибок для формирования контрольного поля кадра в рамках протокола канального уровня типа V.42, так и кодом с исправлением ошибок. При очень низком качестве канала это позволяет резко снизить число повторных передач и, следовательно, повысить реальную скорость передачи.

Аналоговые системы сотовой связи первоначально были разработаны для голосовой связи. Они используют компрессирование и предварительную коррекцию, которые вносят дополнительные искажения в передаваемый сигнал. Очень сильный сигнал может исказиться компрессором при его ограничении по амплитуде. Слишком слабый сигнал даст низкое отношение сигнал/шум на приемной стороне. Поэтому для установки оптимального уровня передаваемого сигнала используется процедура *автоматической регулировки уровня передачи* (*Automatic Transmit Power Adjustment, АТРА*).

Протокол ZyCELL для быстрого выбора рабочей скорости передачи и режимов кодирования с целью максимизации пропускной способности предусматривает динамический *сотовый выбор скорости* (*Dynamic Cellular Speed Selection, DCSS*).

Улучшенная процедура квитирования (*Reliable Handshake Enhancements, RHE*) обеспечивает надежное вхождение в связь без повторных попыток даже на линиях с очень высоким уровнем шума.

Протокол ZyCELL обеспечивает скорость передачи данных в диапазоне от 2400 до 14400 *бит/с*. В нем используется два новых фирменных метода модуляции: ZyCELL-T и ZyCELL-C.

Метод ZyCELL-T используется при относительно низком уровне помех и обеспечивает передачу данных со скоростями 14400, 12000, 9600, 7200 и 4800 *бит/с*. При этом используются СКК на 128, 64, 32, 16 и 8 сигнальных позиций соответственно. В режиме ZyCELL-T на скорости 4800 *бит/с* реализация СКК-8 дает более высокие показатели допустимого отношения сигнал/шум по сравнению с режимом ДОФМ протокола V.32 для такой же скорости.

Метод ZyCELL-C применяется при относительно высоком уровне помех и обеспечивает передачу данных со скоростями

4800, 3600 и 2400 *бит/с* при использовании 8-позиционной СКК. В этом режиме используется кодирование CFEC.

При работе по каналам низкого качества модем может автоматически переключаться между режимами ZyCELL-T и ZyCELL-C, выбирая наилучший из них для текущих условий передачи. В обоих режимах при всех скоростях передачи скорость модуляции равна 2400 *Бод*.

Сотовый режим работы можно рекомендовать не только для использования с мобильным сотовым радиотелефоном, но и для работы на обычных телефонных линиях с повышенным уровнем помех. Особенно эффективно его использование на фоне преобладания импульсных помех, кратковременных пропадания сигнала и повышенном уровне шума.

По результатам тестов ряда авторитетных специализированных журналов ZyCELL признан лучшим по сравнению с другими протоколами для сотовых систем связи, такими как ETC фирмы AT&T и MNP10.

11.9.3. Протоколы HST, RHST

Протокол HST (*High Speed Technology*) разработан компанией U.S Robotics и реализован в ее модемах серии Courier. Это асимметричный дуплексный протокол с частотным разделением каналов. Скорость передачи по обратному каналу может составлять 300 или 450 *бит/с*. Основной канал обеспечивает скорость передачи 4800, 7200, 9600, 1200, 14400 и 16800 *бит/с*. Скорость модуляции равна 2400 *Бод*. Применены сигнально-кодовые конструкции на основе решетчатого кодирования. Протокол HST относительно прост и помехоустойчив, так как отсутствует взаимное влияние встречных каналов передачи и не требуется эхо-компенсация.

За счет использования процедуры ASL (*Adaptive Speed Leveling*) протокол HST позволяет подстраивать скорость передачи под текущее качество используемого телефонного канала. Данная процедура позволяет не только снижать скорость передачи при ухудшении качества канала связи, но и повышать ее при улучшении параметров телефонного канала. Благодаря этому в каждый момент времени передача данных происходит с максимально возможной скоростью для текущего состояния канала.

Отечественное расширение протокола HST, носящее название RHST, отличается от базового варианта следующими характеристиками:

- максимальная скорость передачи повышена до 21600 *бит/с*;
- на скоростях 16800 *бит/с* и ниже используется более устойчивый вид модуляции сигнала;
- введен режим расширенного управления скоростью передачи. Протокол RHST поддерживается только модемами Русский Курьер 21600.

11.9.4. Протоколы PEP, TurboPEP

Протоколы семейства PEP (*Packetized Ensemble Protocol*) разработаны фирмой *Telebit* и реализованы в ее модемах серий *TrailBlazer* (PEP) и *WorldBlazer* (TurboPEP). Данные протоколы являются полудуплексными. Согласно этим протоколам для обеспечения высокоскоростной передачи полоса пропускания канала тональной частоты разбивается на множество узкополосных частотных подканалов, в каждом из которых происходит независимая передача порции бит из общего двоичного потока. Такие протоколы называют многоканальными, параллельными или протоколами с множеством несущих (*multicarrier*).

В протоколе PEP полоса тонального канала разбивается на 511 подканалов. В каждом подканале шириной около 6 *Гц* данные передаются со скоростью модуляции от 2 до 6 *Бод*. С помощью КАМ один сигнальный элемент переносит от 2 до 6 *бит*. Максимальная скорость передачи по протоколу PEP составляет 19200 *бит/с*.

При установке соединения каждый отдельный подканал тестируется на предмет возможности его использования, выбора скорости модуляции и скорости передачи данных в нем. Протокол предусматривает возможность изменения параметров подканалов или их полного отключения в зависимости от изменяющейся помеховой обстановки в течение сеанса связи. При этом дискретность изменения скорости передачи не превышает 100 *бит/с*.

Протокол TurboPEP за счет увеличения числа частотных подканалов и количества кодируемых бит одним сигнальным

элементом может обеспечивать скорость передачи до 23000 *бит/с*. В отличие от PER, протокол TurboPER предусматривает применение СКК с решетчатым кодом, что увеличивает его помехоустойчивость.

Преимущество данных протоколов по сравнению с традиционными, основанными на модуляции одной несущей, заключается в малой чувствительности к неравномерности АЧХ канала и влиянию импульсных помех.

Современное радиовещание, как и другие области техники, связанные с передачей информации, переживают на рубеже веков техническую революцию. Подтверждением этому может служить создание и внедрение систем цифрового звукового радиовещания.

Существующие системы аналогового радиовещания уже давно исчерпали свои возможности. УКВ-вещание с частотной модуляцией обеспечивает достаточно высокое качество передачи, не подвержено атмосферным помехам, но в движущемся автомобиле в результате многолучевого приема и постоянно меняющихся углов отражения от препятствий возможны искажения сигнала. Кроме того, зона уверенного приема не превышает 50 км от передатчика. Для расширения зоны вещания требуется синхронная сеть передатчиков, работающих на одной частоте. Вещание с амплитудной модуляцией (АМ) охватывает большую территорию, но подвержено воздействию атмосферных и промышленных помех. Кроме того, коэффициент полезного действия АМ-передатчика составляет в среднем 4% — почти как у паровоза. Львиная доля мощности тратится на излучение несущей частоты, которое в информационном смысле является бесполезным.

Кроме того, выделенные для радиовещания диапазоны катастрофически перенаселены. Шаг сетки частот составляет в Европе 9 кГц (ДВ и СВ), в США — 10 кГц (СВ). Это дает чуть меньше 120 частотных каналов с полосой вещания $2 \cdot 4,5$ кГц в диапазоне СВ. Один частотный канал без риска взаимных помех может выделяться удаленным друг от друга передатчикам, поэтому возможное их число в несколько раз больше названной цифры. На коротковолновых диапазонах шаг сетки

частот всего 5 $\kappa\text{Гц}$, что уже меньше необходимой полосы частот. С учетом дальнего прохождения радиоволн на КВ помеховая обстановка ухудшается и качество вещания становится неприемлемым.

Переход к однополосной модуляции позволил бы увеличить число каналов в два раза и в несколько раз повысить экономичность передатчиков, но не может решить проблему помех. Кроме того, однополосную передачу невозможно принять на существующие приемники. Поэтому запланированный переход на однополосное радиовещание вряд ли состоится — на сегодняшний день регулярно вещает лишь несколько КВ-радиостанций. И это уже вопрос не техники, а экономики и социологии. В таких условиях отказ от старой системы вещания возможен только в том случае, если потребителю будет предложено не просто повышение качества, а что-то новое, недоступное ранее.

Цифровое вещание открывает принципиально новые возможности в передаче звуковых программ и программ «радиомультимедиа», сочетающих звуковую, видео, графическую, текстовую и другие виды информации. Возможность пересылки разнообразной сопутствующей информации существенно расширяет спектр сервисных услуг, которые могут быть предоставлены вещателями. Так, сопровождение музыкальных передач информацией о произведении и исполнителе стало в цифровом радио уже традиционным. Кроме того, в испытательных проектах практиковалась выдача на дисплей приемников текстового содержания передачи, фотографий эстрадных исполнителей и картинок, иллюстрирующих содержание новостей.

12.1. Выбор системы вещания

Существующие системы ЦЗРВ можно разделить на две категории:

- системы, требующие выделения отдельного частотного диапазона;
- системы, которые могут использовать диапазон совместно с существующими радиослужбами.

Наиболее совершенной из разработанных к настоящему времени систем ЦЗРВ (*Digital Audio Broadcasting*), которые от-

носятся к первой категории, является «Эврика-147/DAB». Она принята Европейским Радиовещательным Союзом (EBU) в качестве общеевропейской и рекомендована для внедрения во всем мире Межсоюзной технической комиссией всемирной конференции радиовещательных союзов (*Inter-Union Technical Committee of the World Conference of Broadcasting Unions*). Ее приняли не только государства Европы, но и Канада, Китай, Индия, Австралия и другие. Это открывает возможность беспрепятственного обмена радиопрограммами и информацией на международном уровне.

Ко второй категории можно отнести системы ЦЗРВ, разработанные в США. Эти системы предназначены для работы в УКВ диапазоне 87,5–108 МГц и СВ диапазоне 525–1608 кГц одновременно с существующими аналоговыми АМ- и ЧМ-радиостанциями.

12.2. Система ЭВРИКА-147/DAB

Европейские фирмы в 1987 году основали консорциум Eureka-147 с целью разработки принципиально новой системы цифрового радиовещания DAB.

Участниками этого проекта являются около 50 фирм и организаций из Великобритании, Германии, Франции, Голландии, Италии, Швеции, Швейцарии, Норвегии, Финляндии, Японии, Канады, США и ряда других стран. В участники проекта от России, по представлению институтов — лидеров проекта — IRT (Германия) и ССЕТТ (Франция), был в 1995 г. принят ИРПА им. А.С. Попова. В 1992 году на основе всемирного соглашения для DAB были выделены L- и S-диапазоны. Первые приемники, в основном для измерительных целей, были созданы в 1988 году. С 1990 года ряд членов проекта Эврика-147 принял участие в проекте JESSI, в рамках которого была разработана первая интегральная микросхема для коммерческих DAB-приемников.

Принципы построения и параметры системы цифрового радиовещания «Эврика-147» регламентированы принятым в конце 1994 г. европейским телекоммуникационным стандартом ETS 300401. Система прошла испытания в ряде европейских стран (Германии, Англии, Франции и др.) и в Канаде в диапазоне частот от 50 МГц до 1,5 ГГц. Они подтвердили

высокие технические и эксплуатационные характеристики «Эврики-147» и ее конкурентоспособность по сравнению с другими системами ЦРВ, особенно при приеме на подвижных объектах.

Первый DAB-приемник потребительского типа был представлен на выставке в 1995 году в Берлине. Миниатюризация приемников продолжается, в настоящее время их серийным выпуском занимаются фирмы *Grundig*, *Philips* и др. В европейских странах эксплуатируется уже несколько десятков тысяч приемников.

В настоящее время в Европе ведутся работы по внедрению системы «Эврика-147» в практику радиовещания. *Создан комплект специализированных больших интегральных схем СБИС*, разработана передающая и приемная аппаратура, ведется подготовка к регулярному наземному и спутниковому вещанию. НИИРПА им. А.С. Попова также работает над проблемами внедрения ЦРВ по системе «Эврика-147» в России.

Система «Эврика-147/ DAB» пригодна для организации наземного, спутникового и кабельного вещания и обладает следующими техническими преимуществами:

- высокое качество звуковоспроизведения (на уровне проигрывателей компакт-дисков);
- передача цифрового сигнала требует меньшей мощности, чем передача аналогового, а качество звучания остается неизменным, начиная с момента, когда прием сигнала вообще становится возможным;
- возможность передачи в узкой полосе частот шириной 1,54 МГц шести стереопрограмм (с качеством, характерным для проигрывателя компакт-дисков) совместно с разнообразной дополнительной информацией;
- для каждого канала выделяется своя полоса частот и качество передачи можно динамически регулировать в месте, откуда ведется передача.
- возможность охвата вещанием больших территорий (в том числе территории всей России) путем организации одночастотных сетей (т. е. сетей, состоящих из передатчиков, работающих в синхронном режиме на одной и той же частоте) или непосредственно вещания с искусственных спутников Земли;

- возможность передачи электронных газет, факсов, изображений, бизнес-информации, телевизионных сигналов и т. д. («радиомультимедиа»);
- возможность приема программ наземного и непосредственно спутникового вещания на радиоприемники с ненаправленными штыревыми антеннами в домашней обстановке, в движущемся автомобиле или в походных условиях;
- высокая устойчивость к воздействию помех и, в частности, помех многолучевого распространения, позволяющая добиться стабильного приема даже при быстром перемещении автомобилей в городских районах со сложной многоэтажной застройкой;
- наличие специальных каналов для передачи информации ограниченному кругу пользователей (закрытые каналы с паролем или пейджинг);
- возможность использования универсального приемника при реализации наземного, спутникового, гибридного и кабельного вариантов вещания.

12.3. Краткое описание стандарта DAB

Комплекс стандартов DAB определяет способы передачи информации, области занимаемых частот, набор сервисных услуг и методы их предоставления и т. д. Системное построение и технические характеристики системы регламентированы в принятом в 1995 г. и дополненном в 1997 г. европейском телекоммуникационном стандарте ETS 300401. Принцип транслирования DAB-радиопрограмм принципиально отличается от рассмотренных систем радиовещания. Уже не одна, а несколько различных программ объединяются в единый блок, называемый ансамблем (ENSEMBLE, или MULTIPLEX) и передаются на одной несущей частоте. Каждый ансамбль может состоять из 6 основных программ, а также дополнительных данных, связанных с программами. Для передачи цифровой эфирной информации определено использование системы COFDM (*Coded Orthogonal Frequency Division Multiplexing*). Эта система предполагает дискретизацию аналогового потока не только по времени, но и по частоте. Полученные элементы оцифровываются, определенным образом переме-

шиваются и затем модулируют совокупность несущих, количество которых определяется мощностью передаваемого информационного потока (в стандарте DAB может использоваться от 192 до 1536 несущих). Расстояние между несущими выбирается таким образом, чтобы пересечение спектров соседних несущих не создавало помех при демодуляции, т. е. чтобы выполнялось условие их ортогональности.

Дополнительная дискретизация по частоте позволяет применять перемежение и сверточные коды, повышающие устойчивость информации к помехам.

Еще одной существенной особенностью системы COFDM является использование защитного интервала между передачей отдельных символов. Этот интервал перекрывает предполагаемое время появления отраженного сигнала символа, что снижает чувствительность к помехам, вызванным многолучевым распространением радиоволн.

Количество несущих, расстояние между ними, длина защитного интервала и значение FEC (степени защиты за счет применения сверточного кодирования) являются переменными величинами. Использование такой системы передачи обеспечивает возможность чистого приема в местах плотной городской застройки.

Еще одним важным плюсом применения COFDM является высокое качество мобильного приема, одной из основных проблем которого является сложность адаптации приемника к постоянному изменению мощности сигнала из-за изменения конфигурации отраженных сигналов.

Специфика компоновки и передачи цифрового материала позволяет пересылать в общем потоке любые виды оцифрованной информации, при условии, что они не требуют широкой полосы частот. Это может быть текст, графика или даже движущиеся картинки.

Система Эврика-147/DAB подвергалась лабораторным и полевым испытаниям при работе в диапазоне частот 1452–1492 МГц (L-диапазон). При испытаниях применялся режим II, что соответствует наличию 384 несущих в передаваемом радиосигнале. Кодирование цифровых звуковых сигналов в данной системе производилось кодерами при скоростях цифровых потоков на выходах кодеров от 192 кбит/с до 256 кбит/с на один стереосигнал. Система проверялась на способность

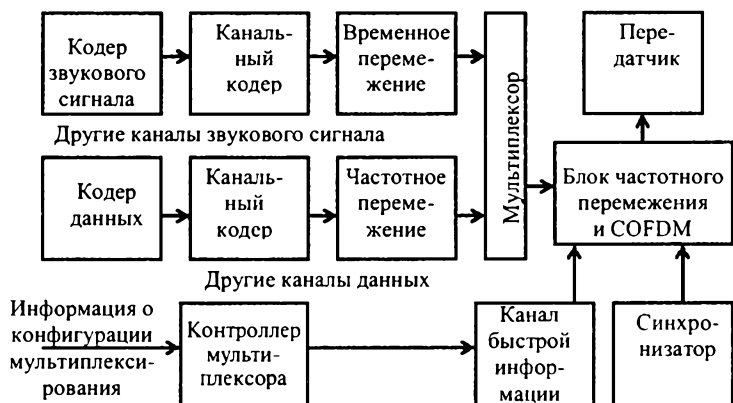


Рис.12.1. Структурная схема передающей части системы «Эврика-147»

передавать пять стереоканалов: один — со скоростью 256 кбит/с , два — со скоростью 224 кбит/с , два — со скоростью 192 кбит/с . Имелась возможность, в дополнение к пяти стереоканалам, передать один моноканал со скоростью 64 кбит/с и два канала данных со скоростями 64 и 24 кбит/с , соответственно.

В системе *DAB* сигнал, прежде чем поступить на передатчик, проходит несколько стадий обработки. Структура передающего тракта такова: сигналы отдельных каналов поступают на входы кодеров, где происходит оцифровка и сжатие информации по стандартам MPEG1 или MPEG2. Сигналы с кодеров каналов поступают на групповой мультиплексор, который формирует последовательный поток битов с временным мультиплексированием, состоящий из кадров длительностью 24 , 48 или 96 мс (модели DAB I, II, III, IV). В этот поток включаются данные скоростного канала информации (FIC), которые используются приемником для идентификации каналов.

«Эврика-147» — это принципиально новая универсальная система ЦРВ, позволяющая вывести звуковое радиовещание на высокий технический уровень. Она обеспечивает передачу, прием и распределение монофонических и стереофони-

ческих программ при наземном, спутниковом и кабельном вещании. Прием программ возможен на радиоприемники с ненаправленной штыревой антенной в домашней обстановке, в движущемся автомобиле или в походных условиях. Высокая устойчивость системы «Эврика-147» к воздействию помех, в частности помех многолучевого распространения, позволяет добиться стабильного приема даже в городских районах с многоэтажной застройкой. Достоинством «Эврики-147» является высокое качество звуковоспроизведения, сравнимое с качеством, гарантированным проигрывателями компакт-дисков. Для этой системы характерно эффективное использование радиочастотного спектра.

Например, в полосе частот $1,54 \text{ МГц}$ с ее помощью может передаваться шесть высококачественных стереофонических программ и разнообразная дополнительная информация. «Эврика-147» дает возможность создавать одночастотные сети на очень больших территориях, обеспечивая при этом десятикратную экономию частотного спектра. Она позволяет оперативно изменять параметры мультиплексирования передаваемого многопрограммного сигнала, что открывает широкие возможности выбора для служб радиовещания, повышает экономическую эффективность передающего оборудования, расширяет круг возможных потребителей. Большим преимуществом системы является также возможность использования универсального приемника при реализации наземного, спутникового и гибридного ее варианта, включая подсистему кабельного вещания. Для «Эврики-147» необходима значительно более низкая мощность передатчиков по сравнению с ЧМ-передатчиками, обслуживающими такую же территорию.

Широкий диапазон частот реализации системы (от 30 МГц до 3 ГГц) позволяет обеспечить как обслуживание больших территорий при использовании наземных одночастотных сетей или спутниковых систем непосредственного вещания, так и местное эфирное и кабельное вещание.

Система ЦРВ «Эврика-147» проста в эксплуатации, в том числе за счет использования меню выбора программ владельцем массового бытового радиоприемника. Она обеспечивает полную и точную идентификацию программ и станций, передачу текстовой информации и информации для водителей автотранспорта. Возможна даже передача изображений га-

зет в оригинальном цветном оформлении, географических карт и т. д.

Рассмотрим принцип работы системы ЦРВ «Эврика-147». На рис. 12.1 представлена упрощенная функциональная схема передающей части системы. Обработка сигнала происходит здесь в несколько этапов. На первом этапе сигналы, поступающие по каналам передачи звуковых программ и каналам передачи данных, подвергаются индивидуальному кодированию. Эту функцию выполняют специальные устройства, называемые кодерами звуковых сигналов и кодерами данных.

В системе ЦРВ «Эврика-147» применяется метод субполосного кодирования звуковых сигналов MUSICAM. Благодаря использованию эффектов маскировки, свойственных человеческому слуху, этот метод позволяет, например, снизить скорость цифрового потока каждого из каналов высококачественного стереофонического сигнала с 7178 (студийный стандарт 117-разрядное кодирование отсчетов при частоте дискретизации 48 кГц) до 917 кбит/с, т. е. в восемь раз при сохранении субъективного качества звучания на уровне, характерном для проигрывателя компакт-дисков. Система обеспечивает следующие скорости передачи звуковых сигналов: 32, 48, 517, 174, 80, 917, 112, 128, 1172 и 192 кбит/с на монофонический канал. Соответственно, число каналов звукового вещания в многопрограммном групповом цифровом потоке может изменяться от 20 монофонических (при невысоком качестве) до 4 стереофонических (с практически студийным качеством).

При использовании метода субполосного кодирования MUSICAM с помощью гребенки фильтров широкополосный звуковой сигнал, преобразованный в цифровую форму, разделяется на 32 субполосных сигнала. Цифровые отсчеты группируются в циклы. В каждом таком цикле выделяется один масштабный множитель, соответствующий максимальному уровню, достигаемому каждым субполосным сигналом. При этом охватывается полный динамический диапазон звукового сигнала, равный 120 дБ.

Однако субполосная фильтрация с ограниченным количеством полос не позволяет с высокой точностью оценить порог спектрального маскирования, в частности в низкочастотной области. По этой причине параллельно с фильтрацией

выполняется быстрое преобразование Фурье цифрового звукового сигнала. При этом кодируются и передаются только отсчеты субполосных сигналов.

Сочетание обеих этих операций позволяет с высокой точностью оценить пороги маскировки человеческого слуха. Для каждого из 32 субполосных сигналов вычисляется минимальный порог маскирования, который определяет максимально допустимый уровень шума квантования. При этом не возникает необходимости передавать информацию об отсчетах сигналов субполос, если они полностью маскируются намного более существенными для восприятия компонентами соседних субполос.

Масштабные множители и другая дополнительная информация, необходимая для правильного функционирования декодера звукового сигнала в приемнике, объединяется с информацией о субполосных отсчетах звукового сигнала в один уплотненный сигнал.

В уплотненный сигнал вводятся также данные, несущие информацию о передаваемой программе (*Program Associated Data*, PAD). Эти данные помещаются в конце сформированного цикла (фрейма) в месте, соответствующем стандарту. Типичные примеры таких данных — информация об управлении динамическим диапазоном, о видах передаваемых программ (джаз, лирика, речь, музыка и т. д.). Канал PAD может быть также использован для передачи текстовой и графической информации. Скорости передачи этого канала могут быть различными.

Кроме сигналов PAD, в общем многопрограммном цифровом потоке могут передаваться сигналы сервисной информации (*Service Information*, SI) и другие данные. Сигналы SI могут отражать наименование канала: вид программы (спорт, новости, музыкальный канал и др.); название географического места нахождения передатчика, сигнал которого принимается пользователем; программы передач и т. д. Примером передач и других данных может служить текстовая информация для широкого круга потребителей.

Вторым этапом обработки передаваемого звукового сигнала является сверточное кодирование и временное перемежение цифровой информации, поступающей на каналные кодеры (рис. 12.2). Сверточное кодирование обеспечивает вве-

Канал синхронизации	Канал быстрой информации (FIC), MSI, часть сервисной программы	Радиоканал 1	Радиоканал 2	Радиоканал 3	Радиоканал 4	Радиоканал 5	Радиоканал 6	Сервисная информация

Канал MSC: 6 стереопрограмм, PAD, часть сервисной информации

Рис. 12.2. Пример построения цикла

дение в передаваемый сигнал избыточной информации с целью повышения его помехоустойчивости при передаче по реальным каналам связи. Для кодирования используется сверточный код с длиной ограничения, равной 7. Средняя относительная кодовая скорость, определяемая как отношение скоростей передачи информации (количества бит в единицу времени) на входе и выходе канального сверточного кодера, колеблется от 0,35 (высший защитный уровень) до 0,75 (низший защитный уровень). Промежуточные значения кодовых скоростей выбираются для различных программ в соответствии с требованиями к уровням защиты информации.

Временное перемежение улучшает помехоустойчивость передачи информации, устраняя пакеты ошибок, что особенно важно при мобильном приеме в движущемся автомобиле в условиях быстро изменяющейся окружающей обстановки.

Третий этап обработки сигнала включает его мультиплексирование, а также его системную организацию и управление. Предварительно обработанные в канальных кодерах и устройствах временного перемежения сигналы поступают на главный служебный мультиплексор (*Main Service Multiplexer*, MUX), где собираются в циклы определенной длительности. Выходящий из этого устройства многопрограммный групповой цифровой поток поступает в так называемый главный служебный канал (*Main Service Channel*, MSC). Максимальная скорость передачи данных в этом канале составляет 2,304 Мбит/с.

В мультиплексор поступают также данные о синхронизации всех подлежащих мультиплексированию программных сигналов. Работой мультиплексора управляет контроллер MUX, причем режим мультиплексирования может при необходимости изменяться в соответствии с заданной программой. Чтобы обеспечить минимальную общую задержку доступа к некоторым или всем передаваемым сигналам при их приеме, точная информация о текущем режиме мультиплексирования (*Multiplex Configuration Information*, MCI) передается по каналу быстрой информации (*Fast Information Channel*, FIC). MCI представляет собой машиночитаемые данные. Они не подвергаются перемежению в канале FIC и потому не претерпевают задержки, возникающей при временном перемежении закодированной информации о звуковом сигнале и других данных. В то же время информация в канале FIC хорошо защищена, поскольку передается со средней относительной кодовой скоростью $1/3$ и для обеспечения отсутствия сбоев часто повторяется. Новая информация об изменении режима мультиплексирования передается в MCI через канал PIC. Некоторая часть SI, необходимая, например, для выбора требуемых программ, также передается по каналу FIC. Не требующаяся для управления приемником более объемная сервисная информация (программы передач и т. д.) может передаваться в общем многопрограммном групповом цифровом потоке.

Чтобы обеспечить синхронизацию приемника, передаваемый сигнал формируется в виде цикла (фрейма) с определенной последовательностью составных частей. Пример его построения показан на рис. 12.2.

Каждый цикл начинается с временного интервала, в котором содержится информация для канала синхронизации. Следующая часть зарезервирована для FIC, а остальная для MSC. Общая длительность цикла в зависимости от режима передачи (см. таблицу 12.1) составляет 917 или 24 мс. Каждой звуковой программе, передаваемой по каналу передачи звуковых сигналов, или информации, передаваемой по каналу данных, соответствует свой временной интервал в цикле.

Следующим этапом является формирование сигнала ЦРВ, предназначенного для последующей передачи в эфир.

В системе ЦРВ «Эврика-147» применяется метод уплотнения с ортогональным частотным разделением кодирован-

Параметры системы в зависимости от режима передачи

Параметры	Режим передачи		
	1	2	3
Номинальный частотный диапазон, МГц	375	1500	3000
Количество несущих	1536	384	192
Длительность фрейма, мс	917	24	24
Длительность защитного интервала, мкс	2417	172	31
Максимальное разнесение передатчиков при работе в одночастотной сети, км	917	24	12

ных сигналов (*Coded Orthogonal Frequency Division Multiplex, COFDM*). При этом в цифровой поток вводятся специальные сигналы синхронизации от соответствующего синхрогенератора (рис. 12.2).

Метод COFDM состоит в делении передаваемой информации на большое число потоков данных, имеющих низкую индивидуальную скорость. Эти данные используются затем для модуляции по фазе ряда несущих, так что длительность передаваемых символов становится большей, чем задержка распространения в канале передачи. За счет введения временного защитного интервала между следующими друг за другом символами многолучевое распространение не вызывает межсимвольную интерференцию.

Большое количество несущих может быть сформировано при использовании алгоритма дискретного преобразования Фурье (ДПФ). При наличии многолучевого распространения при передаче сигнала ЦРВ некоторые из несущих могут быть ослаблены или вообще исчезнуть (эффект частотно-селективного фединга). По этой причине в системе «Эврика-147», кроме временного, применяется и частотное перемежение за счет перераспределения цифровых потоков между несущими. В результате исчезновение части несущих из-за частотно-селективного фединга не приведет к появлению искажений сигнала, поскольку информация будет восстановлена по оставшимся неповрежденными модулированным несущим.

В системе «Эврика-147» предусмотрены три режима передачи, что позволяет организовать вещание в широком диапазоне частот от 30 МГц до 3 ГГц. В таблице приведены основные параметры системы в зависимости от режима передачи.

Как видно из приведенных в таблице данных, использование более высоких частот налагает большие ограничения на длительности защитных интервалов и, следовательно, на максимальное время задержки распространения сигналов.

Режим 1 наиболее подходит для организации наземного вещания и построения одночастотных сетей, поскольку позволяет обеспечить наибольшее разнесение передатчиков и, следовательно, обойтись меньшим их количеством при заданной площади обслуживания.

Режим 2 можно использовать, в частности, для местного вещания.

Режим 3 более пригоден для организации спутникового и кабельного вещания.

Спектр сигнала ЦРВ имеет примерно прямоугольную форму и занимает полосу частот около 1,54 МГц.

Упрощенная функциональная схема приемника для системы ЦРВ «Эврика-147» показана на рис. 12.3. Сигнал, принятый антенной, поступает на вход тюнера, который обеспечивает выделение сигналов определенного диапазона частот, их усиление, преобразование по частоте и фазовую демодуляцию. С выхода тюнера сигнал подается на вход аналогово-цифрового преобразователя (АЦП), а затем на вход блока, осуществляющего дискретное преобразование Фурье и дифференциальную демодуляцию.

Далее в следующем блоке производится частотное и временное дегерметирование и коррекция ошибок на основе алгоритма декодирования Витерби с мягким решением. Оригинальные закодированные данные с выхода этого блока обрабатываются затем в декодере звукового сигнала или в соответствующем декодере данных. На выходе первого декодера образуются звуковые монофонические или стереофонические сигналы, которые воспроизводятся громкоговорителями правого ВА1 и левого ВА2 каналов.

Приемник обеспечивает одновременное декодирование более чем одной канальной компоненты многопрограммного

группового цифрового потока, например звуковой программы параллельно с сервисной информацией.

Системный контроллер приемника соединен с интерфейсом пользователя и управляет приемником в соответствии с командами пользователя и информацией, передаваемой в FIC.

Сигнал с выхода мультиплексора поступает на вход COFDM-модулятора, который добавляет информацию для коррекции ошибок, защитные интервалы и данные, идентифицирующие передатчик. После этого осуществляется быстрое преобразование Фурье и кодирование I/Q символов. Кодированные данные поступают на высокоскоростной цифро-аналоговый преобразователь, который формирует модулированный сигнал ПЧ.

Частота сигнала ПЧ конвертируется в частоту, присвоенную каналу, и сигнал усиливается до необходимого для трансляции уровня мощности.

Работы по созданию цифровой системы радиовещания начались в России в начале 1980 года во ВНИИРПА им. А.С. Попова, которые завершились созданием отечественной системы ЦРВ, опытных образцов передающего и приемного оборудования и организацией экспериментального вещания в г. Новгороде в 1993 году. Однако, поскольку в 1995 году была стандартизована в качестве общеевропейской система ЦЗРВ Эврика-147/DAB, которая существенно отличается от отечественной, то начиная с 1993 года все работы были сосредоточены на внедрении в России этой системы. К сожалению, далее зафиксированы в основном не факты, а намерения.

Для экспериментального вещания в Санкт-Петербурге Министерство Связи приняло решение о выделении полосы частот в диапазоне 92–100 МГц. Опытные работы были успешно проведены, однако для внедрения указанной системы ЦЗРВ в масштабах Российской Федерации этих усилий недостаточно. До сих пор отсутствует единая федеральная программа введения цифрового стандарта на телевидении и радиоэлектронных СМИ. Для решения проблемы вещания необходимо решение множества организационных проблем, в первую очередь — выделение отдельного диапазона частот. Европейский опыт показал, что использование диапазона 87,5–108 МГц совместно с существующими ЧМ-станциями нецелесообразно.

Для разработки был рекомендован диапазон 176–230 МГц. Однако стратегия — не программа, и реальных шагов для реализации этой концепции до сих пор не сделано. Конкретные частоты в рекомендованном диапазоне могут быть выделены только после продолжительного анализа электромагнитной обстановки, который должен провести НИИ Радио. А после выделения частот комитетом Минсвязи предстоит еще выиграть конкурс на ее использование в МПТР (Министерство печати и теле- и радиовещания). После этого останется пустяк — найти средства для реализации проекта. На государственное финансирование рассчитывать не приходится, на средства населения — тем более. Недавно состоянием российского проекта заинтересовались некоторые западные компании — производители DAB-оборудования. Желая выйти на российский рынок, они готовы финансировать развертывание опытной зоны вещания. Вероятно, это наиболее реалистичный сценарий успешного развития событий.

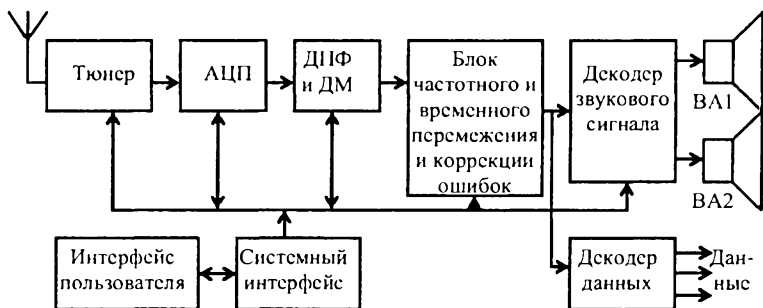


Рис. 12.3. Структурная схема приемника системы «Эврика-147»

14 мая 2002 г. начались опытные передачи по трансляции стереофонических программ звукового вещания «Радио России», «Маяк» и «Юность» в цифровом формате через спутник связи и вещания «Экспресс-6А» (точка стояния 80 градусов в.д.) на территорию России от Калининграда до Владивостока. Однако это еще не регулярное вещание.

12.4. Система DRM

Для диапазонов средних и коротких волн разработан стандарт DRM (*Digital Radio Mondiale*). Разработчики решили обратиться к этим диапазонам, поскольку они дают беспрецедентные возможности для дальнего вещания при существенно меньших затратах на его организацию, а техническое качество вещания уже не отвечает современным требованиям. Организация DRM была основана в марте 1998 г. Это некоммерческий международный консорциум, включающий более 60 участников из Европы, Азии и Америки, в том числе российский канал «Голос России». В основу нового стандарта, названного «система А», положен доработанный прототип системы *Skywave-2000*, разработанной французской фирмой *Thomcast*. Осенью 2000 года Международный Союз Электросвязи рекомендовал своим членам использовать его, после чего закрепилось название «система DRM». Формат характеризуется гибкими параметрами передачи, позволяющими использовать его во всех диапазонах ниже 30 МГц. Одновременно он может использоваться и для диапазона УКВ. Система открыта для последующего улучшения, новых методов компрессии и процессов кодирования.

12.4.1. Краткое описание стандарта DRM

DRM — многофункциональная система цифрового радиовещания, которая предназначена для применения в диапазоне частот, не превышающих 30 МГц, распределенных вещательным службам.

В системе DRM предусмотрено использование каналов радиовещания, занимающих полосы частот шириной 4,5; 5,0; 9,0; 10,0; 18,0 и 20,0 кГц.

Эта система обеспечивает слушателям возможность приема стереофонических передач.

Настройка приемника на нужную станцию осуществляется автоматически, так как настроиться на нужную станцию человек не может. Поэтому выбор нужной станции осуществляется приемником самостоятельно. Для этого передатчик DRM передает короткие импульсы, расположенные на шести несущих любой DRM-станции, отстоящих на 750, 2250 и 3000 Гц

вверх и вниз от срединной частоты. Эти импульсы играют роль опорных сигналов, которые ищет приемник при сканировании диапазона принимаемых частот.

После того, как осуществлен выбор нужной станции, информационный поток, излучаемый ею, делится на три неодинаковых по объему передаваемых станцией слоя. Сначала приемником декодируется самый простой слой, который содержит данные о частотном диапазоне, занимаемом станцией в эфире, типе модуляции, параметрах разброса данных по времени и несущим частотам, имени вещателя. Данные, передаваемые в этом слое, называемом *Fast Access Channel* (FAC — канал скоростного доступа), предельно упрощены и всегда передаются с помощью модуляции QAM-4. Для передачи информации в FAC применяется 4-QAM модуляция. Используется равная защита от ошибок EEP (*Equal Error Protection*).

Канал FAC переносит данные о параметрах радиочастотного сигнала и информацию, позволяющую выделять отдельные услуги. К параметрам сигнала относятся идентификатор потока, ширина занимаемой полосы, тип модуляции, тип кодирования, индекс глубины перемежения, количество передаваемых услуг. К параметрам, характеризующим услуги, относится указание типа сервиса (аудио/данные), флаг условного доступа, указатель языка и некоторые другие. Они передаются последовательно — в одном кадре параметры, относящиеся к одному сервису.

После декодирования FAC приемник приступает к декодированию информации, которая передается в канале называемом *Service Description Channel* (SDC — канал описания услуг). В этом канале содержится информация об условном доступе, программе передач, авторских правах, вспомогательная информация для некоторых приложений, а также список альтернативных частот, на которых передача этой станции дублируется в данный момент. Это позволяет автоматически выбрать канал, принимаемый в данный момент наилучшим образом.

В канале SDC может использоваться либо 16-QAM, либо 4-QAM. 16-QAM обеспечивает большую пропускную способность канала, а 4-QAM — более высокую устойчивость к ошибкам. Применяется равная защита от ошибок EEP, т. е. фик-

сированная скорость сверточного кода для всей передаваемой информации.

Сигнальное созвездие следует выбирать с учетом параметров MSC таким образом, чтобы обеспечить большую устойчивость SDC по сравнению с MSC.

После анализа информации, поступившей в приемник по каналам FAC и SDC, приемник начинает декодировать информацию, представляющую собой главное содержание любой DRM-передачи, которая заключена в *Main Service Channel* (MSC — главный канал передачи пользовательской информации), максимально защищенного от ошибок слоя данных, содержащего основной объем передачи.

Таким образом, процесс настройки на нужную станцию до появления первых звуков в динамической головке занимает несколько секунд.

Как и в большинстве цифровых систем, DRM-данные передаются в виде фреймов (frame) длительностью 0,4 с, которые на передатчике могут объединяться в суперфреймы (*superframe*) длительностью 1,2 с, которые определяют время, на которое приемник может терять сигнал, не давая слушателю это заметить. К сожалению, длительность суперфрейма не может превышать 1,2 с, потому что этот параметр определяется степенью непрерывности приема во время частого замирания сигнала при передаче в коротковолновом диапазоне. Чем длиннее MSC-фрейм, тем более продолжительные замирания может компенсировать радиоприемник. Но это потребует увеличения объема памяти в декодере, что сегодня уже не проблема. Проблема в другом. Задержка между микрофоном диктора и ухом слушателя равна удвоенной длине MSC-суперфрейма. Если бы разработчики не ввели жесткие границы для длительности MSC-фреймов, вещатели сразу бы выбрали фреймы оптимальной длины, которая часто превышала бы 1,2 с.

Вещание на коротких волнах ведется сразу на нескольких частотах, что увеличивает аудиторию, так как в разных местах прохождение на разных частотах оказывается разным. На практике функция смены частоты (*Alternative Frequency Switching*) происходит автоматически, поскольку приемник оценивает качество сигнала принимаемой станции и в то же время знает другие частоты, на которых эту станцию можно принимать.

Приемник часто показывает эти частоты, позволяя слушателю легко переключать их. С некоторой периодичностью приемник меняет излучаемую частоту и проверяет, есть ли на другой волне FAS-канал той же станции. Если приемник обнаруживает наличие такого сигнала, то за доли секунды оценивается уровень сигнала на новой частоте и принимается решение перейти на новую частоту или вернуться на старую. Такие проверки длятся меньше одного MSC-фрейма, поэтому звук в динамике приемника не пропадает и осуществляется непрерывное звучание. Кроме того, несмотря на короткую потерю сигнала, приемник не пропускает ни одного бита из MSC-потока и качество приема сохраняется неизменным. Этот процесс обеспечивается структурой MSC-фрейма, который начинается со служебной SDC-информации, и пока она принимается, приемник имеет возможность оценить условия приема на другой частоте и вернуться обратно. Именно к этому моменту в MSC-фрейме начинается передача звука или данных, и приемник продолжает работать так, как будто он и не проводил никакого анализа условий приема сигнала на других частотах.

В MSC может применяться либо 64-QAM, либо 16-QAM. 64-QAM обеспечивает высокую спектральную эффективность, т. е. более высокую пропускную способность радиоканала при заданной полосе занимаемых частот, в то время как 16-QAM обеспечивает большую устойчивость к ошибкам, возникающим при передаче сигналов.

В каждом случае диапазон выбранных скоростей сверточного кода должен обеспечивать наиболее приемлемый уровень коррекции ошибок для данного канала передачи. Существующие комбинации сигнальных созвездий и скоростей сверточного кода обеспечивают высокую степень гибкости в широком диапазоне вариантов каналов передачи.

Неравная защита от ошибок UEP применяется для обеспечения двух уровней защиты информации в MSC. При использовании 64-QAM и с помощью иерархической модуляции можно обеспечить третий (высший) уровень устойчивости против ошибок для части MSC.

В отличие от стандарта DAB, использующего MPEG-2, в DRM применяется более современный вариант компрессии MPEG-4. Он включает адаптивный механизм компрессии

сигнала AAC (*Advanced Audio Coding* — усовершенствованное звуковое кодирование) в моно и стереовариантах, а также CELP (*Code Excited Linear Prediction* — линейное предсказание с кодовым возбуждением) для высококачественного кодирования речи и шумоподобных сигналов и HVXC (*Harmonic Vector Excitation Coding* — кодирование с гармоническим векторным возбуждением). В MPEG-4 долговременное предсказание проводится не во временной, а в спектральной плоскости. Кодер делает предсказание, а затем кодирует либо разницу между реальным и предсказанным сигналом, либо сам входной сигнал, если его значение можно закодировать более компактно, чем разницу. Кроме того, кодер поддерживает несколько новых механизмов, связанных со способностью потока адаптироваться к изменениям параметров канала. Любой из вариантов может дополняться техникой SBR (*Spectral Band Replication*), предназначенной для повышения качества передачи верхних частот. При передаче на частотах ниже 30 МГц все форматы, кроме стереофонического, используют полосу 9/10 кГц. Использование техники SBR требует более широкой полосы.

Структурная схема передающей части системы DRM приведена на рис. 12.4.

Помимо аудиосигналов, в цифровом потоке могут передаваться данные. Мультиплексированный поток аудио и данных формируют основной сервисный канал MSC. В MSC передается до 4 потоков, каждый из которых переносит или аудио, или данные. Информация канала MSC разбивается на логические кадры по 400 мс каждый. Основной и сервисные каналы определенным образом мультиплексируются, в результате чего образуются транспортные суперкадры длительностью 1200 мс.

Логический кадр состоит из двух частей, каждая из которых имеет свой уровень защиты от ошибок. Из логических кадров всех цифровых потоков формируются кадры мультиплекса, длительность каждого из которых равна 400 мс. Эти кадры, пройдя обработку в блоках рандомизации, поступают на вход канального кодера MSC.

Информация о конфигурации мультиплексирования передается в канале SDC. При необходимости конфигурация мультиплексирования может быть изменена. В этом случае пара-

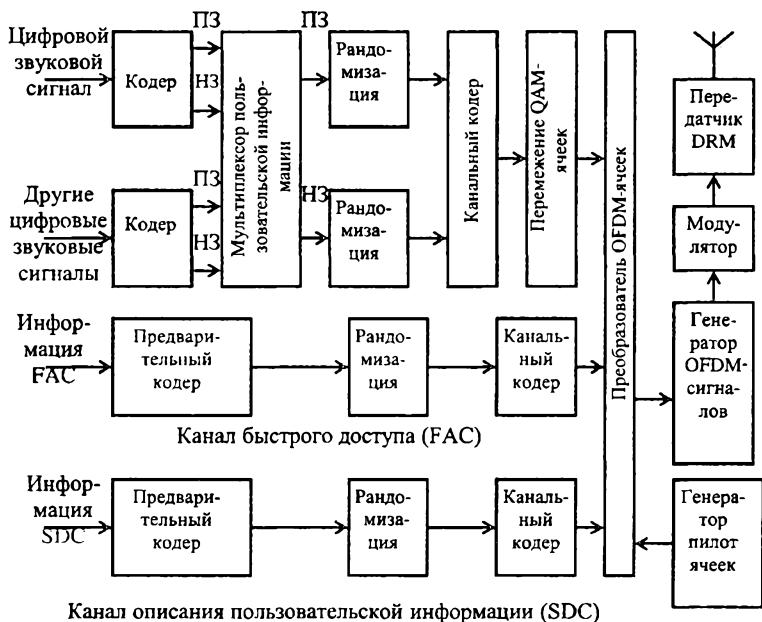


Рис. 12.4. Структурная схема передающей части системы DRM

метры новой конфигурации мультиплексирования передаются через канал SDC, а точное время изменения конфигурации указывается в FAC.

Канал MSC представляет собой последовательность кадров мультиплекса. Кадры мультиплекса формируются из логических кадров и содержат части с повышенной и нормальной защитой от ошибок, в которые помещается информация из соответствующих частей логических кадров.

12.4.2. Перемежение QAM-ячеек в канале MSC

Перемежение QAM-ячеек должно применяться после многоуровневого кодирования с возможностью выбора малой или большой глубины перемежения, т. е. интервалов времени, в течение которых производится процедура перемежения (обозначим их как «короткое» и «длинное» перемежение), в соответствии с ожидаемыми условиями распространения сигнала DRM. Базовые параметры перемежения адаптированы к раз-

меру кадра мультиплекса. Для радиоканалов с умеренными замираниями, что типично для передачи сигнала посредством земной волны в диапазонах ДВ, СВ, короткое перемежение обеспечивает приемлемые условия для правильной работы декодера в приемнике.

При сложных условиях распространения радиосигналов DRM, т. е. существенных замираниях, что характерно для отраженных от ионосферы волн в диапазонах КВ, глубина перемежения может быть увеличена.

Общая задержка информации, передаваемой в MSC, при перемежении на передающей и приемной сторонах примерно равна 800 мс для короткого и 2,4 с для длинного перемежения.

Перемежение QAM-ячеек в каналах FAC и SDC не производится в целях исключения временных задержек при передаче системной информации.

12.4.3. Построение OFDM-сигналов

После перемежения QAM-ячейки преобразуются в OFDM-символы. Каждый OFDM-символ передается за время T_s и представляет собой совокупность k несущих, равномерно распределенных в заданном частотном диапазоне. OFDM-символ — это передаваемый сигнал для такого промежутка времени, в течение которого значения амплитуд и фаз всех QAM-ячеек, модулирующих несущие OFDM-сигнала, остаются постоянными.

Из определенного количества OFDM-символов формируются кадры передачи. Первый OFDM-символ каждого кадра передачи содержит ячейку опорного времени. Длительность кадра передачи 400 мс. Из трех кадров передачи формируется сверхкадр передачи. Длительность сверхкадра передачи 1200 мс. В начале сверхкадра передачи размещается SDC-блок.

Длительность символа состоит из двух частей: полезной части T_u и защитного интервала T_g . Расстояние между несущими OFDM-сигнала равно $1/T_u$. Защитный интервал располагается перед полезной частью символа. Введение защитных интервалов в OFDM-символы позволяет бороться с межсимвольной интерференцией в радиоканалах с многолучевым распространением и подверженных эффекту Допплера.

OFDM-символы в кадре нумеруются от 0 до $N_s - 1$. Все символы содержат данные и опорную информацию.

При наличии многолучевого распространения и частотно-селективных замираний в радиоканалах DRM часть несущих, входящих в OFDM-сигнал, может быть ослаблена или вообще исчезнуть. При этом информация, содержащаяся в этих несущих, во многих случаях может быть восстановлена по оставшимся неповрежденными несущим за счет помехоустойчивого кодирования, а также перемежения битов и QAM-ячеек. В случае невозможности восстановления информации в символе или ряде символов применяется маскировка ошибок. Такие технические решения обуславливают высокую надежность и устойчивость приема сигналов в системе DRM.

Так как OFDM-сигнал состоит из множества отдельно модулированных несущих, каждый символ может быть рассмотрен как разделенный на OFDM-ячейки, соответствующие текущей модуляции одной несущей в течение времени передачи одного символа.

12.4.4. Структурная схема приемника DRM

Структурная схема радиоприемника для системы DRM показана на рис. 12.5. Радиосигнал DRM, поступающий из ненаправленной антенны, выделяется в блоке тюнера, после чего направляется на вход демодулятора OFDM. С выхода этого устройства сигнал поступает на инверсный преобразователь OFDM-ячеек, на выходе которого формируются системные каналы MSC, FAC и SDC.

В канале MSC производится деперемежение QAM-ячеек. Далее в системных каналах осуществляется канальная демодуляция, т. е. обратное преобразование QAM-ячеек и исправление ошибок, возникших при передаче сигнала DRM по реальному каналу связи.

После этого производится дерандомизация цифровых сигналов, передаваемых в системных каналах. Затем выполняется демультимплексирование MSC, т. е. разделение мультимплекса на отдельные звуковые (речевые) каналы или каналы данных. Цифровые потоки, передаваемые в них, декодируются в предназначенных для этого декодерах.

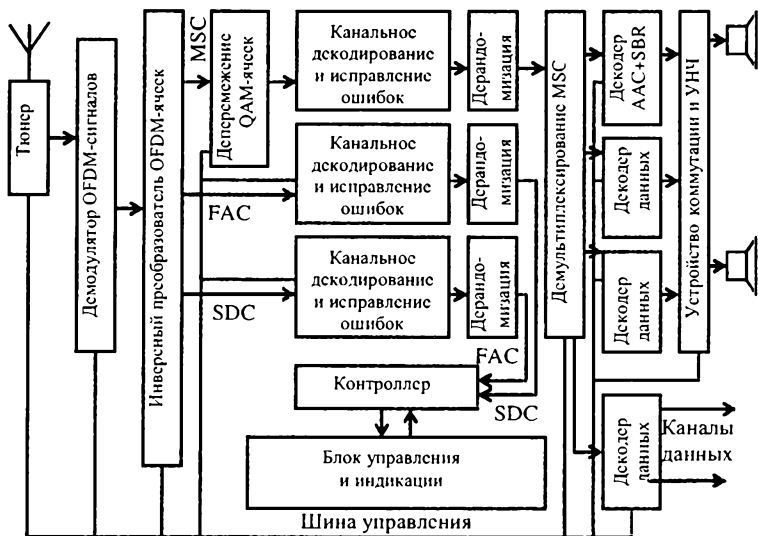


Рис. 12.5. Структурная схема радиоприемника системы DRM

Информация, передаваемая в каналах данных, может отображаться на дисплее пользователя и/или поступать на выход приемника для дальнейшего использования. Системный контроллер соединен с блоком управления и индикации. Он управляет приемником в соответствии с командами пользователя и информацией, передаваемой в FAC и SDC.

В DRM, как и в DAB, применяется система модуляции COFDM. Эта система весьма эффективна для передачи сигналов по радиоканалу с многолучевым распространением радиоволн и селективным замиранием сигнала, характерным для коротких волн. Для компенсации помех многолучевого распространения используется защитный интервал. Он не должен превышать 20% от общей длительности символа, чтобы не снизить пропускную способность канала. Количество несущих, размещаемых в полосе частот канала, ограничивается доплеровским смещением частоты сигнала, возникающим в режиме мобильного приема. С учетом этих факторов в полосе 9/10 кГц используется около 200 несущих. Их точное количество, равно как и длительность символа и защитного интервала, зависит от характера распространения радиоволн (поверх-

ностные или пространственные), предположительной дальности передачи и требуемой достоверности.

Каналы, входящие в MSC, разделяются на 2 части, различающиеся по значимости информации для правильного декодирования. Они подвергаются раздельному помехоустойчивому кодированию, характеризующемуся разной степенью помехозащищенности. В качестве помехоустойчивого кодирования применяется перемежение данных и сверточное кодирование со скоростями кода от 0,5 до 0,8. Перемежение данных в системах COFDM реализуется и по времени, и по частоте, что позволяет восстанавливать сигнал при высоком уровне селективного замирания в радиоканале. Кроме того, для борьбы с этим явлением в поток вводятся пилотные сигналы, позволяющие приемнику оценить степень затухания сигналов на каждой несущей частоте. Уровень налагаемой защиты также зависит от диапазона и предполагаемой дальности распространения сигнала. В частности, при передаче на коротких волнах глубина перемежения составляет 2,4 с, а на длинных и средних волнах — 0,8 с. Кроме того, на коротких волнах используется сверточное кодирование с более низкими скоростями кода и вводится большее количество пилотных сигналов.

12.5. Особенности, возможности, преимущества и многообразие форматов стандарта DRM

С первого взгляда кажется, что для территорий большой протяженности оптимально цифровое спутниковое вещание, например SM-Satellite Sirius (США, Африка) или WorldSpace (Азия). Через эти спутники транслируется более 100 различных радиоканалов прекрасного качества. Все бы хорошо, но для приема необходимо иметь спутниковое оборудование или специальные радиоприемники со встроенной мини-тарелкой. И даже это — полбеды. Для приема также необходимо держать «зрительную» связь между аппаратом и спутником. Такая, казалось бы, малозначительная помеха, как дерево перед окном, может нарушить прием, а что говорить о неровной горной местности. В городских условиях густой застройки пришлось бы устраиваться на крыше в надежде послушать радио. Кроме того, существует еще одна проблема — высокая

стоимость спутникового вещания. В идеале спутники необходимо обновлять каждые 10–12 лет, а это — большие средства. Неудивительно, что многие операторы связи часть программ предлагают только за абонентную плату.

Другой вариант — наземное цифровое радиовещание DAB (*Digital Audio Broadcasting*) — схож с форматом FM: приблизительно на каждые 80–100 км² необходимо расставить по передатчику. Причем в каждом канале должно одновременно транслироваться 6 программ, что чревато нежелательной конкуренцией между коммерческими станциями. Подобная система возможна в городских условиях, но не для малых населенных пунктов. Для государств с большими территориями формат DAB — слишком дорогой способ распространения программ.

Технология цифрового наземного телевещания DVB-T позволяет передавать и радиосигнал. Но, как и в случае с форматом DAB, площади покрытия маленькие, а конкуренция большая.

Новый стандарт, разработанный в США, — *HD-Radio*; он приспособлен исключительно для FM- и СВ-диапазонов с распространением сигнала по трем каналам передачи, что уже само по себе не подходит, например, для России. Ко всему прочему, данный стандарт является частной разработкой, действующей только за абонентскую плату. Самое интересное заключается в том, что формат не приспособлен для ионосферного распространения сигнала в СВ-диапазоне. Следовательно, в темное время суток трансляция на средних волнах может идти только в аналоговом режиме либо ее придется и вовсе отключить.

С цифровой трансляцией на длинных, средних и коротких волнах в формате DRM дело обстоит иначе, особенно это касается вещания на большие территории.

В 1998 г. в городе Гуанчжоу (Китай) на очередном совещании радиовещателей его участниками было принято решение об организации DRM-консорциума. Поначалу в нем насчитывалось всего двадцать единомышленников. Основателем, а потом и бессменным руководящим центром Консорциума на протяжении всех этих лет была и остается «Немецкая волна». Сейчас же в Консорциуме насчитывается почти 100 членов из более чем 30 государств. Помимо полноправ-

ного членства есть и другая возможность сотрудничества — это так называемое «supporting». Русским аналогом было бы несколько устаревшее понятие «сподвижничество», или помощь/поддержка. Таких сподвижников/помощников у Консорциума среди небольших фирм и частных лиц уже более 50, в их числе несколько представителей и из России. Число членов и помощников заметно увеличилось с весны 2005 г. — с момента, когда было принято решение развивать стандарт и в диапазоне выше 26 МГц, то есть до 108 МГц FM-диапазона (технология DRM+).

Технология DRM+ позволит разместить в 4–6 раз больше радиоканалов в диапазонах частот I и II. На 2008 г. намечено первое испытание новой расширенной разработки. Дальнейшая ее судьба будет зависеть от потребностей и интересов рынка.

Формат DRM, как и большинство цифровых технологий, отличает не только улучшенное качество аудиотрансляций, но и дополнительные сервисные удобства, которые оценит каждый радиослушатель. Параллельно к звуковому сигналу возможно передавать и текстовую информацию в качестве комментария к программе или независимо от нее.

Сервисные функции DRM включают в себя следующие возможности:

- предоставление данных о приеме всех станций в месте нахождения радиослушателя;
- выбор желаемой станции посредством маркировки ее на дисплее приемника;
- автоматический, незаметный для радиослушателя переход на лучшую по качеству частоту;
- отображение названия рубрики передачи, а также имени автора и ведущего;
- предоставление информации о названии музыкальной композиции, имени исполнителя;
- параллельная трансляция нескольких языковых программ с возможностью выбора нужного языка;
- вызов на дисплее текста самых актуальных новостей;
- запись особо важных сообщений в память приемника для последующего прослушивания;
- передача метеосводок, информации о положении на дорогах, биржевых новостей;
- возможность коммутации с мобильными средствами связи;

- возможность записи кодированных сигналов — платных сервисов, таких как аудиокниги, актуальные статьи и т. п.;
- загрузка программного обеспечения и актуализация уже имеющегося ПО, например в навигационных системах транспортных средств путем передачи данных на КВ по всему миру.

Данный список можно было бы продолжить, но это дело будущего. Сейчас на первом месте для радиовещателей стоит улучшение качества аудиотрансляций и экономия при этом средств.

Дальнейшее развитие системы будет целиком и полностью зависеть от ситуации на радиорынке.

Радиостанция «Немецкая волна» протестировала все из вышеприведенных сервисных возможностей DRM-формата. С середины 2005 г. основной упор радиостанция делает на безупречное качество трансляций и приема. При наличии соответствующего приемного оборудования DRM-трансляции можно принимать как стационарно в городских условиях, так и в движении на автомагистралях. Причем сигнал не пропадает и на высокой скорости (до 220 км/ч), и в пробках на многолюдных дорогах при сильном уровне помех.

Для коммерческих радиостанций Европы, таких как RTL и *TruckRadio* (станция для шоферов-дальнобойщиков), привлекательность нового стандарта в первую очередь заключается в том, что он позволит получать дополнительные доходы от рекламы, благодаря новым техническим возможностям. Крупные вещатели Франции, Германии, Великобритании с помощью DRM-формата смогут полностью покрыть вещанием все области Европы, до этого по разным причинам не охваченные FM- или DAB-трансляциями. DRM-формат удобен и для локального радиовещателя с небольшими мощностями как в СВ-, так и в КВ-диапазоне на 11 метрах (27 МГц).

Подобные проекты действуют в Германии, Великобритании; тестовые трансляции проходят в Мексике. Преимущество таких локальных для городских условий трансляций, аналогичных по качеству FM-сигналу, состоит в том, что необходимая мощность передатчика колеблется от 100 Вт до 1 кВт.

Последнее звено, которого не хватает в цепочке от передатчика до радиослушателя, — это доступный по цене радиоприемник. Здесь возможны два пути решения проблемы:

- *DSP/Digital System Processor* — цифровой процессор для обработки сигнала (разработка английской фирмы *Radioscape*);
- чипсет — набор микросхем (различные производители).

Фирма *Radioscape* (Великобритания) совместно с фирмой *Texas Instrument* разработала модуль, который подходит ко всем существующим сегодня на рынке DAB-приемникам. Один из уже готовых аппаратов британской фирмы *Roberts* проходит сейчас последнее тестирование и будет в ближайшее время предложен массовому потребителю. Аппарат универсален: радиослушатель сможет вести прием в формате DRM, DAB, FM с сервисом RDS (*Radio Data System*), а также и в аналоговом стандарте на длинных, средних и коротких волнах. Помимо этого у приемника есть функция записи программ для последующего прослушивания. К сожалению, пока аппарат работает на основе DSP, поэтому потребляет много энергии, что, соответственно, ведет к быстрому использованию заряда батареек. Следовательно, вести прием радиосигнала на данный аппарат лучше всего от сети. Еще несколько фирм работают по созданию радиоприемников, но тоже на основе DSP от *Radioscape*. Фирма *Analogue Devices* (США) разработала свой собственный DSP-модуль, продемонстрировав в 2005 г. собранный на его основе прототип радиоприемника.

Для портативных переносных приемников малого размера лучше всего подходит чипсет, потребляющий намного меньше энергии. Производители данной электроники уже разрабатывают необходимую часть будущих радиоаппаратов, но официальные заявления пока не делались. Хотя есть надежда, что в течение ближайших лет будет опубликована информация о новых разработках. Микросхемы сделают возможным радиоприем в DRM-формате для любого цифрового аппарата: мобильного телефона, фотокамеры, электронного переводчика. Особый интерес к данному формату проявили производители мобильных телефонов, в частности в ответ на возросший спрос на модели, имеющие опцию приема FM-радиосигнала.

12.6. Перспективы цифрового вещания

Необходимо признать, что цифровое вещание превосходит по качеству и возможностям аналоговое. Однако не всегда качественное превосходство переходит в количественное, как это получилось с DAB-стандартом. Поэтому цифровое вещание полностью заменит аналоговое еще очень не скоро, и об исчезновении традиционного радио говорить пока рано. Ожидалось, что практически во всех развитых странах Европы с 1997—1998 гг. начнется регулярное DAB-вещание, а к 2000 году будет завершен охват вещанием 80% населения Европы. Однако реальные темпы охвата заметно ниже прогнозов. Производители не торопятся наращивать выпуск соответствующих приемников, потому что не уверены в рынке сбыта. А рынок не готов принять новые приемники, поскольку большинство слушателей удовлетворено качеством аналогового вещания на УКВ и пока не видит возможностей нового формата. Существует также реальная конкуренция со стороны компаний звукозаписи, заинтересованных в том, чтобы в автомашинах слушали не радио, а записи — доходы от продаж заметно превышают авторские отчисления радиостанций.

Первоначальная ошибка была в том, что DAB рассматривался только как технический стандарт. Однако история техники наглядно демонстрирует, что одни только технические преимущества еще не являются для массового потребителя решающим аргументом при выборе товара, должны еще быть какие-то иные достоинства. При разработке DAB возможность передачи дополнительной информации была заложена изначально, однако концепция вещания поначалу сохранилась: тот же принцип построения передач, та же назойливая реклама. Существующие сегодня широковещательные программы в основном рассчитаны на безликого среднего слушателя и обращаются к общенациональной аудитории. При таком подходе нельзя учесть региональные особенности аудитории, местные интересы и традиции и т. д. Стандарт DAB рассчитан на локальное, региональное вещание и вполне отвечает интересам коммерческого радио, но в целом, смена стратегии идет достаточно медленно. Чтобы заинтересовать потребителя, а тем самым стимулировать и производителя, BBC первой создала

новую программу для цифрового формата, ее примеру последовали и другие радиовещательные компании.

DAB-радио планируется использовать не только для трансляции традиционных радиопрограмм (свободных и закодированных), но и для передачи деловой информации. Она будет выводиться на дисплей DAB-приемника или на монитор компьютера в виде телетекста. Такая возможность существует благодаря использованию протокола MOT (*Multimedia Object Transfer*), разработанного для приведения разнородных данных к стандартному формату телетекста. Кроме того, предполагается использовать DAB-службы для передачи шифрованной информации, предназначенной для узкого круга абонентов. Этой службой могут пользоваться банки, страховые конторы и другие организации для распространения информации между филиалами.

Еще одной службой, которую предполагается развить в рамках национальных DAB-проектов, является предоставление информации о состоянии на дорогах. Это особенно актуально ввиду того, что DAB-радио в сильной степени ориентировано на мобильный прием. Помимо предоставления широкого спектра сопутствующей информации цифровое вещание предполагает возможность организации передач по запросу и других интерактивных служб. В мобильных приемниках в качестве обратного канала предполагается использовать сети GSM.

На будущее положение DAB на рынке может негативно повлиять широкое распространение приемников спутникового телевизионного вещания DVB, которым отдают явное предпочтение и промышленность, и радиостанции.

Ввод в эксплуатацию цифрового мультиплексного радио *dRadio* обеспечил трансляцию через спутник *Hot Bird-5* в одном единственном канале DVB до 40 некодированных радиопрограмм в формате MPEG-2. Спасение DAB пока в том, что для DVB разработаны только стационарные приемники, рассчитанные на эксплуатацию в одном из трех режимов: DVB-S (прием сигнала со спутника), DVB-C (через кабель) и DVB-T (наземно). Однако новые системы прямого спутникового вещания позволяют принимать программы и в движущемся автомобиле.

В конце 2001 года в Америке начали регулярную работу две национальные сети спутникового непосредственного циф-

рового радиовещания: *Sirius Satellite Radio* и *XM Satellite Radio*. В эфире сразу появились десятки новых высококачественных радиопрограмм.

Специалисты отмечают не только принципиально иную бизнес-модель спутникового радиовещания, которая может существовать и без традиционной радиорекламы, на деньги подписчиков, но и ожидаемые революционные изменения в программировании и в выборе содержания прослушиваемых передач, которые станут доступными самим радиослушателям. Весь этот спектр удовольствий стоит слушателю 9 долларов 95 центов в месяц. Компании уверены, что, как и в случае с кабельным телевидением, аудитория захочет платить за то, что непосредственно удовлетворяет ее интересам и не содержит ненавистную многим рекламу. *Sirius Satellite Radio* и *XM Satellite Radio* полны оптимизма, а стоимость приемников уже сопоставима со стоимостью существующей автомобильной радиоаппаратуры. Учитывая, что за год в США продается около 24 миллионов автомобильных приемников, по прогнозам к 2007 году 15 процентов всех машин будут оснащены аппаратурой приема спутникового вещания.

Три спутника *Sirius Satellite Radio* охватывают цифровым сигналом всю территорию США. Прием ведется на антенну размером 20–30 см. В дополнение к спутникам компания развернула сеть из 100 наземных передатчиков-ретрансляторов. Со спутников транслируются 100 каналов в формате MPEG-2. *XM Satellite Radio Holdings Inc.* предлагает радиослушателям 71 музыкальный канал, их репертуар охватывает массу музыкальных направлений от классической музыки до рока. Кроме того, вниманию радиослушателей предлагаются новостные каналы, материалы для которых предоставляют CNBC, CNN *Headline News*, USA Today и C-SPAN, а также ряд других агентств. Через спутники транслируются спортивные и развлекательные передачи, ток-шоу и программы для детей. В космическом сегменте системы спутникового цифрового радиовещания работают два спутника связи под кодовым названием *Rock* и *Roll*. Сигнал принимается со спутников непосредственно на радиоприемники, а в городских районах с многоэтажной застройкой для трансляции сигнала дополнительно будут использоваться наземные ретрансляторы.

Основное преимущество прямого спутникового вещания имеют перед существующими УКВ-радиостанциями — глобальность. Однако затраты на создание такой системы достаточно велики. В странах и регионах с низкой плотностью населения или находящихся на значительном удалении от экватора спутниковое радио может оказаться неэффективным и непопулярным по чисто экономическим соображениям. И в ряде случаев наземное цифровое радиовещание имеет больше перспектив в обозримом будущем, в отличие от США, хотя бы в связи с на порядки меньшими расходами на их внедрение. Вот тут и выходит на сцену DRM.

Развитие цифрового вещания в «дальнобойных» диапазонах особенно актуально для России при ее огромной территории и средней плотности населения 9 человек на квадратный километр. В небольших городах, в которых живет более половины населения страны, коммерческое вещание УКВ-ЧМ часто оказывается нерентабельным. Поэтому для почти забытых диапазонов ниже 30 МГц в России открыты прекрасные перспективы.

Однако есть и другие проблемы. При внедрении цифрового радиовещания в диапазонах с АМ необходимо обеспечить сохранение возможности приема для более 2 млрд существующих аналоговых радиоприемников. Этого можно достичь путем одновременной передачи цифрового радиовещательного сигнала с половинной скоростью и совместимого однополосного аналогового вещательного сигнала. При этом владельцы обычных радиоприемников будут принимать аналоговую программу, а цифровая будет прослушиваться как слабый высокочастотный шум.

Радиочастотный спектр излучаемого совместимого сигнала не превышает отведенные сегодня ITU на один радиоканал 9 кГц. В чисто цифровом режиме передачи спектр передаваемого сигнала может занимать те же стандартные 9 кГц или, в случае расширенного режима, — 12 кГц, при этом достигается повышенное качество звучания и передача дополнительной мультимедийной информации.

Цифровая технология, лежащая в основе будущих приемников DAB, сделает возможным и прием сигналов DRM. Для радиостанций переход на цифровое вещание в диапазонах коротких, средних и длинных волн означает существенное (при-

мерно на три четверти) сокращение расходов на трансляцию при сохранении неизменным радиуса охвата. При этом передающие станции могут продолжать использовать имеющуюся инфраструктуру.

12.7. Передающая часть системы цифрового радиовещания DRM

Как следует из рис. 12.5, обработка сигналов, поступающих от создателей радиовещательных программ, производится в несколько этапов.

На первом этапе осуществляется устранение избыточности и последующее кодирование сигналов, осуществляемое с помощью сжатия, в результате чего снижаются скорости передачи цифровых потоков, поступающих на вход передающей части DRM. Такое кодирование не приводит к заметному ухудшению качества звуковоспроизведения на приемной стороне по сравнению с исходным звуковым сигналом и при ограниченной пропускной способности канала передачи сигналов DRM позволяет увеличить количество передаваемых программ.

В системе DRM применяются три разновидности метода кодирования цифровых звуковых сигналов MPEG-4 Audio (стандарты ISO/IEC 14496-3, ISO/IEC 14496-3/Amd1): MPEG-4 AAC, MPEG-4 CELP и MPEG-4 HVXC. Метод MPEG-4 AAC (*Advanced Audio Coding* — усовершенствованное звуковое кодирование) применяется для обработки относительно широкополосных звуковых сигналов (ЗС). При этом в системе DRM может выполняться дополнительная обработка ЗС с помощью метода *Spectral Band Replication* (SBR — копирование спектральных полос). Применение метода SBR позволяет расширить диапазон воспроизводимых частот ЗС более чем в 2 раза за счет воссоздания высокочастотных составляющих спектра ЗС. При этом используется информация, содержащаяся в более низкочастотных спектральных составляющих ЗС. Методы кодирования MPEG-4 CELP (*Code Excited Linear Prediction* — линейное предсказание с кодовым возбуждением) и MPEG-4 HVXC (*Harmonic Vector eXcitation Coding* — кодирование с гармоническим векторным возбуждением) применяются для передачи речевых сигналов с применением низкоскоростных цифровых потоков (ЦП).

Диапазон скоростей передачи ЦП после кодирования звуковых и, в частности, речевых сигналов в системе DRM — от 2 до 72 кбит/с. Скорость ЦП 2 кбит/с соответствует передаче речевого сигнала с коммуникационным качеством, а при скорости ЦП, равной 72 кбит/с, можно передать стереофонический ЗС с улучшенным качеством. В тракте передачи системы DRM формируются три системных канала:

- *Main Service Channel* (MSC — главный канал передачи пользовательской информации);
- *Fast Access Channel* (FAC — канал быстрого доступа);
- *Service Description Channel* (SDC — канал описания пользовательской информации).

Канал MSC формируется на выходе главного мультиплексора пользовательской информации MUX. На вход мультиплексора поступают подвергнутые обработке в кодерах ЗС или в предварительных кодерах цифровые потоки, соответствующие звуковым, речевым сигналам или дополнительной информации.

Эти ЦП разделяются на части с повышенной и нормальной защитой от ошибок. Повышенная защита применяется для частей ЦП, наиболее чувствительных к ошибкам.

В мультиплексоре MUX производится объединение указанных цифровых потоков. Сформированный канал MSC также содержит части с повышенной и нормальной защитой от ошибок. В блоках рандомизации производится дополнение цифровых потоков псевдослучайными последовательностями битов в целях устранения систематических повторений комбинаций двоичных символов и возникающей при этом нежелательной регулярности в передаваемом сигнале. В канальных кодерах производится помехоустойчивое кодирование информации, которое базируется на сверточном коде, перемежение битов в целях рассредоточения групповых ошибок и преобразование информации в так называемые QAM-ячейки. QAM-ячейки в канале MSC подвергаются перемежению, что позволяет повысить устойчивость приема сигналов в каналах передачи с нестабильными характеристиками (например, при приеме радиосигналов DRM, отраженных от ионосферы, в диапазоне коротких волн).

В канале FAC передается информация о полосах частот, занимаемых радиосигналами DRM, режиме модуляции, ко-

личестве и типах цифровых потоков в MSC, идентификации программ и др.

Канал SDC предназначен для передачи информации о конфигурации мультимплексирования MSC, условном доступе, частоте сигнала, районе обслуживания, языке вещания, времени, дате и др.

Перемежение QAM-ячеек в каналах FAC и SDC не применяется. Это позволяет исключить временные задержки, связанные с процедурой перемежения, и повысить оперативность работы приемника DRM. Тем не менее, информация, передаваемая в FAC и SDC, имеет достаточно высокую степень защиты от ошибок.

OFDM-преобразователь ячеек (*Orthogonal Frequency Division Multiplexing* — мультиплексирование с ортогональным частотным разделением) распределяет различные классы ячеек по времени и частоте и формирует так называемую «частотно-временную сетку».

OFDM-генератор сигналов преобразует в цифровой форме каждый ансамбль ячеек с одинаковыми временными индексами в совокупность модулированных несущих, разнесенных по частоте с определенным интервалом. Затем образуется полный OFDM-символ путем введения защитного интервала, который представляет собой повторение части символа и служит для предотвращения межсимвольной интерференции.

В модуляторе производится преобразование цифрового OFDM-сигнала в аналоговый. Эта операция включает в себя цифро-аналоговое преобразование, частотное преобразование вверх, фильтрацию в целях выполнения требований МСЭ-Р к спектрам передаваемых радиосигналов. Далее сигнал поступает на вход DRM-передатчика и затем передается в виде радиоволн.

Рассмотрим передающую часть системы DRM более подробно.

12.8. Кодирование сигналов, поступающих от создателей звуковых программ

Методы кодирования сигналов, поступающих от создателей звуковых и, в частности, речевых программ, иллюстрируются рис. 12.6, а, а методы декодирования — рис. 12.6, б.

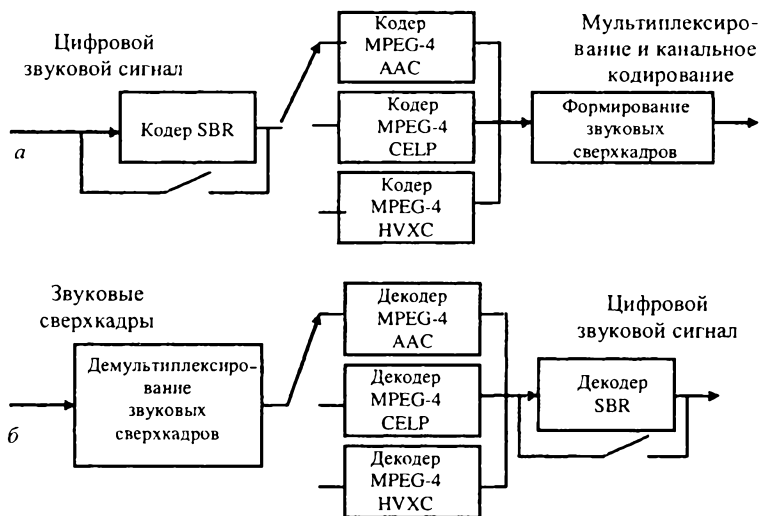


Рис 12.6. Методы кодирования (а) и декодирования (б) цифровых звуковых сигналов в системе DRM

Метод кодирования звуковых сигналов MPEG-4 AAC. Для осуществления кодирования звуковых сигналов в системе DRM применен метод MPEG-4 AAC (стандарты ISO/IEC 14496-3 и ISO/IEC 14496-3/Amd1) как оптимальный по критериям качества звуковоспроизведения, снижения скорости передачи цифрового ЗС и устойчивости функционирования всей системы.

Однако применение этого метода в системе DRM имеет свою специфику, которая заключается в следующем:

- в системе DRM применяется устойчивый к ошибкам вариант 2 MPEG-4 AAC (*Object Type ID-17*, который является частью *High Quality Audio Profile*);
- разрешены частоты дискретизации $f_d = 12 \text{ кГц}$ и $f_d = 24 \text{ кГц}$;
- длительность сформированных в кодере звуковых кадров равна 80 мс при $f_d = 12 \text{ кГц}$ и 40 мс при $f_d = 24 \text{ кГц}$, что соответствует 960 отсчетам;
- из звуковых кадров формируются звуковые сверхкадры, длительность которых равна 400 мс; таким образом, один звуковой сверхкадр содержит 5 звуковых кадров, имею-

щих длительность, равную 80 мс (на кадр), или 10 звуковых кадров с длительностью каждого 40 мс;

- один звуковой сверхкадр всегда расположен в одном логическом кадре (определение логического кадра дано ниже); это позволяет исключить дополнительную синхронизацию при кодировании и декодировании ЗС.

Как правило, применяется неравная защита от ошибок (*Unequal Error Protection*, UEP), при использовании которой наиболее чувствительная к ошибкам часть информации имеет повышенную защиту от ошибок, а оставшаяся часть — нормальную. Частота дискретизации $f_d = 24 \text{ кГц}$ применяется при кодировании стереофонических сигналов. Пример построения звукового сверхкадра MPEG-4 AAC для случая передачи стереофонического сигнала представлен на рис. 12.7.

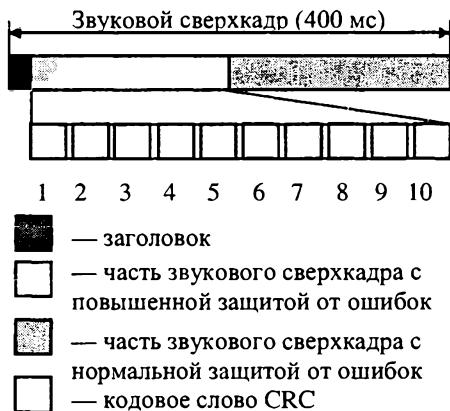


Рис. 12.7. Структура звукового сверхкадра при использовании метода MPEG4 AAC для передачи стереофонического сигнала

Метод копирования спектральных полос (SBR). Метод SBR (*Spectral Band Replication* — копирование спектральных полос) — это средство улучшения качества звуковоспроизведения при передаче ЗС с применением низкоскоростных цифровых потоков.

Известно, что подавление высокочастотных составляющих в спектре ЗС (рис. 12.8) приводит к искажению его тембра.

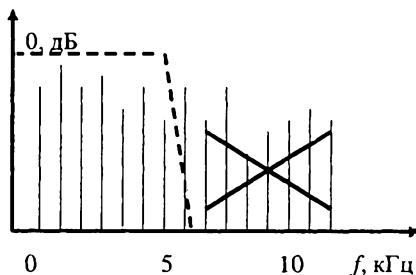


Рис. 12.8. Подавление высокочастотных составляющих спектра звукового сигнала

Тембр ЗС становится более глухим и тусклым, а речевой сигнал, кроме того, — менее разборчивым и понятным. Тем не менее, в ряде случаев разработчики вынуждены ограничивать спектр ЗС, так как, в соответствии с теоремой Котельникова, при аналого-цифровом преобразовании ЗС значение максимальной частоты спектра ЗС — $F_{з.макс}$ — должно соответствовать выражению $F_{з.макс} = f_d/2$, где f_d — частота дискретизации. Например, при $f_d = 12 \text{ кГц}$ $F_{з.макс} = 6 \text{ кГц}$. Метод SBR позволяет расширить полосу воспроизводимых частот ЗС сверх указанного предела.

Данный метод основан на том, что подавленные на передающей стороне высокочастотные составляющие спектра ЗС могут быть приблизительно воссозданы на приемной стороне при использовании имеющейся связи между низкочастотными и высокочастотными спектральными составляющими (рис. 12.9).

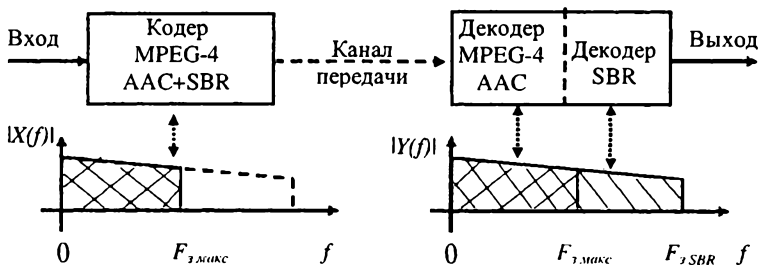


Рис. 12.9. Принципы кодирования и декодирования цифровых звуковых сигналов при использовании методов MPEG-4 AAC и SBR

На приемной стороне производится копирование части низкочастотных составляющих спектра ЗС и перенос их в высокочастотную область. Для улучшения слухового восприятия ЗС огибающая воссозданной высокочастотной части спектра не должна существенно отличаться от огибающей спектра исходного ЗС. Это достигается путем передачи от кодера к декодеру SBR дополнительной информации, позволяющей формировать огибающую высокочастотной части спектра в приемнике DRM. Для передачи такой информации требуется цифровой поток со скоростью передачи примерно 2 кбит/с на канал. При этом важно поддержание соотношений между гармоническими и шумоподобными компонентами в воссозданной высокочастотной части спектра ЗС. Поэтому на приемной стороне производится селективное дополнение высокочастотной части спектра ЗС шумоподобными компонентами. Имеются два различных протокола цифровых потоков для совместного использования: SBR и MPEG-4 AAC; SBR и MPEG-4 CELP. Блок-схема, которая иллюстрирует процесс декодирования ЗС, подвергнутого кодированию методами MPEG-4 AAC и SBR, представлена на рис. 12.10. Цифровой поток поступает с выхода декодера MPEG-4 AAC на вход 32-полосного анализирующего банка фильтров. В каждой из 32 субполос группируется по 30 отсчетов ЗС. В результате на выходе анализирующего банка фильтров формируется кадр, содержащий 960 отсчетов. Эти кадры поступают на устройство задержки, которое необходимо для согласования по времени сигналов низкочастотных и высокочастотных субполос, и на устройство воссоздания высокочастотных спектральных составляющих. На это же устройство поступает необходимая информация с блока деформатирования цифрового потока данных SBR.

На передающей стороне отсчеты огибающей ЗС и его шумоподобных компонентов квантуются и кодируются методом дельта-модуляции. Затем эта информация кодируется с помощью кода Хаффмена и передается на декодер SBR. Декодер Хаффмена преобразует принятые кодовые слова в квантованные отсчеты огибающей ЗС и шумоподобных компонентов. В кодере SBR производится путем адаптивного группирования субполосных отсчетов ЗС во временной и частотной областях. Информация о частотных диапазонах и временных



Рис. 12.10. Структурная схема, иллюстрирующая процесс декодирования звуковых сигналов при использовании методов MPEG-4 AAC и SBR

интервалах, действительных для каждого кадра, представляется как частотно-временная сетка, которая должна быть передана на декодер. Границы временных интервалов выбираются в соответствии со свойствами ЗС. Более длинные интервалы используются для квазистационарных ЗС, а более короткие — для быстро изменяющихся звуковых сигналов. Временные и частотные параметры, определяющие шумоподобные спектральные составляющие ЗС, передаются аналогичным образом. Информация с выходов декодера Хаффмена и устройства управления частотно-временной сеткой поступает на вход блока расчета коэффициентов усиления. Эти коэффициенты необходимы для формирования огибающей высокочастотной части спектра ЗС в блоке регулировки усиления. Синтезирующая фильтрация задержанных отсчетов низкочастотных субполос и высокочастотных субполосных от-

счетов, прошедших процедуру регулировки усиления, выполняется при помощи 64-канального банка фильтров. Отсчеты низкочастотных субполос поступают на низшие 32 канала синтезирующего фильтра, а высокочастотных — на остальные 32 канала, соответствующие высоким частотам. На выходе синтезирующего фильтра формируется 1920 отсчетов ЗС на каждый принятый звуковой кадр MPEG-4 AAC + SBR. Каждый такой звуковой кадр, формируемый на передающей стороне, состоит из двух частей, относящихся к AAC и SBR соответственно (рис. 12.11).

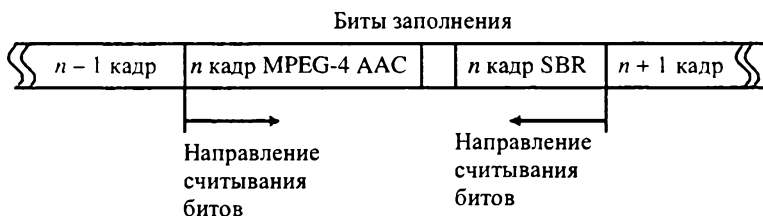


Рис. 12.11. Структура звукового кадра MPEG-4 AAC+SBR

Биты SBR расположены в конце кадра. Направления считывания битов в частях, относящихся к MPEG4 AAC и SBR, противоположны, что облегчает поиск стартовых точек обеих частей кадра. Для звуковых сигналов, скорости передачи которых равны или превышают 20 *кбит/с*, метод кодирования SBR должен быть использован. При меньших скоростях передачи ЗС метод SBR может быть использован по мере необходимости. Эффективность метода SBR можно оценить, рассмотрев конкретный пример кодирования монофонического ЗС. Для этого случая получены следующие показатели, приведенные в табл. 12.1.

В рассмотренном примере при кодировании ЗС методом MPEG-4 AAC можно обеспечить диапазон воспроизводимых частот от 0 до 6 *кГц*. Применение метода SBR позволяет расширить диапазон воспроизводимых частот с 6 до 15,2 *кГц*. При этом общая скорость передачи цифрового потока составляет примерно 22 *кбит/с*.

Метод кодирования MPEG-4 CELP. Метод кодирования MPEG-4 CELP предназначен для обработки речевых сигнала-

Показатели для кодирования монофонического ЗС

Скорость передачи ЗС	22 кбит/с
Длительность звукового кадра	40 мс
Частота дискретизации MPEG-4 AAC	24 кГц
Частота дискретизации SBR	48 кГц
Частотный диапазон ЗС при применении MPEG-4 AAC	0–6 кГц
Частотный диапазон ЗС за счет применения SBR	6–15,2 кГц
Средняя скорость цифрового потока SBR	2 кбит/с на канал

лов (РС). Устройства цифрового кодирования речи можно разделить на две категории: кодеры формы сигнала и вокодеры.

На практике применяются в основном три основных класса кодеров: кодеры формы, вокодеры и гибридные кодеры.

Кодеры формы характеризуются способностью сохранять основную форму речевого сигнала. К кодерам формы относятся кодеры с импульсно-кодовой модуляцией (ИКМ), кодеры с дифференциальной ИКМ (ДИКМ), адаптивной дифференциальной ИКМ (АДИКМ) и др. Системы передачи с подобным типом кодеров обеспечивают хорошее качество воспроизведения речевых сигналов (стандартная полоса частот которых составляет 300–3400 Гц) и более широкополосных звуковых сигналов. Однако эти кодеры малоэффективны с точки зрения снижения скоростей передачи цифровых сигналов.

Так, стандартный телефонный речевой сигнал в системе с ИКМ и мгновенным компандированием передают со скоростью 64 кбит/с. Применение АДИКМ позволяет снизить скорость передачи такого сигнала при сохранении приемлемого качества воспроизведения речи до 32 кбит/с, т. е. всего в 2 раза. Вокодеры обеспечивают значительно большее снижение скоростей передачи речевых сигналов. Сжатие на передающей стороне производится в анализаторе, выделяющем из речевого сигнала медленно меняющиеся составляющие, которые передаются по каналу связи в виде кодовых комбинаций.

На приемной стороне с помощью местных источников сигналов, управляемых с использованием принятой информа-

ции, синтезируется речевой сигнал. Работа вокодеров основана на моделировании человеческой речи с учетом ее характерных особенностей. Вокодер преобразует входной сигнал в некий другой, похожий на исходный. При этом измеряемые характеристики используются для подстройки параметров вокодера в соответствии с принятой моделью речевого сигнала. Именно эти параметры и передаются на декодер приемника, который по ним восстанавливает (синтезирует) речевой сигнал. При этом оценка качества воспроизведения речи (разборчивость, естественность, узнаваемость и др.) производится с применением субъективно-статистических экспертиз.

Наибольшее распространение получили параметрические вокодеры, в которых из речевого сигнала выделяют два типа параметров:

- параметры, характеризующие огибающую спектра речевого сигнала (фильтровую функцию);
- параметры, характеризующие источник речевых колебаний (генераторную функцию): частоту основного тона, ее изменения во времени, моменты появления и исчезновения основного тона, шумового сигнала и др.

По этим параметрам на приемной стороне синтезируют речь.

Вокодеры с линейным предсказанием (LPC— Linear Predictive Coding). В вокодерах с линейным предсказанием при анализе речевого сигнала в передающем устройстве определяются коэффициенты предсказания, а в приемном устройстве на основе этих коэффициентов с помощью рекурсивного цифрового фильтра синтезируется эквивалент голосового тракта. Принцип метода линейного предсказания состоит в том, что прогнозируемая величина речевого сигнала (PIC) в момент выборки h определяется как линейно взвешенная сумма значений сигнала в моменты предшествующих выборок.

$$\bar{\lambda}(h) = \sum_{m=1}^p \lambda(h-m) a_m,$$

где $\bar{\lambda}(h-m)$ — значения речевого сигнала в моменты предшествующих выборок; $m = 1, 2p$; p — порядок предсказания; a_m — коэффициенты предсказания. Интервалы времени между моментами выборок определяются частотой дискретизации $t_h - t_{h-1} = 1/f_d$. В момент h , когда известны не только предсказанные значения $\bar{\lambda}(h)$, но и истинное значение

речевого сигнала $\lambda(h)$, можно определить ошибку предсказания $\varepsilon_n = \lambda(h) - \bar{\lambda}(h)$ и затем подобрать коэффициенты предсказания таким образом, чтобы ошибка предсказания была минимальной. Коэффициенты предсказания, значения которых передаются по каналу связи на приемную сторону, используются в качестве переменных параметров в рекурсивном цифровом фильтре, на вход которого подаются сигналы возбуждения. При воспроизведении вокализованных звуков (гласных) — это последовательность импульсов с частотой основного тона, а при воспроизведении невокализованных звуков (согласных) — это случайная последовательность импульсов, формируемых генератором шума. При кодировании с линейным предсказанием моделируются различные параметры человеческой речи, которые передаются вместо отсчетов речевого сигнала или их разностей. Это позволяет существенно снизить скорость передачи речевого сигнала по сравнению с методами ИКМ, ДИКМ, АДИКМ.

Широко применяемый в настоящее время метод кодирования с линейным предсказанием предусматривает формирование блоков отсчетов (кадров), для каждого из которых вычисляется и передается частота основного тона, его амплитуда и информация о типе возбуждающего воздействия (гармоническое, негармоническое). Структура синтезатора речи с линейным предсказанием показана на рис. 12.12. Здесь сигналы возбуждения имеют вид последовательности импульсов на частоте основного тона (для вокализованных звуков) или случайного шума (для невокализованных звуков). Комбинации

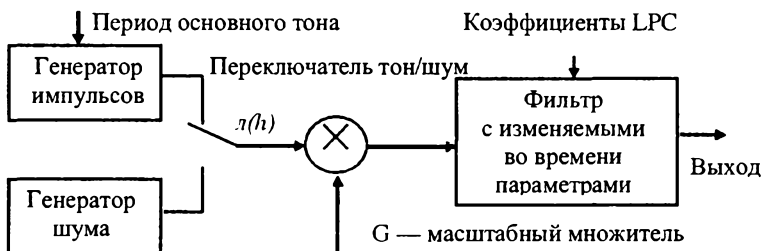


Рис. 12.12. Структура синтезатора речи с линейным предсказанием

спектральных составляющих речи, возникающей, в частности, за счет работы голосовых связок, языка и губ человека, могут быть промоделированы цифровым фильтром с изменяющимися параметрами.

При линейном предсказании обычно производится спектральный анализ речи и выполняется построение систем анализа-синтеза. Во всех случаях параметры синтезатора обновляются при смене анализируемых кадров речевого сигнала. Чтобы избежать эффектов, связанных со скачками значений параметров, необходимо плавно изменять параметры с помощью интерполяции при переходе от одного фрагмента (сегмента) речи к другому. При кодировании речевых сигналов по методу LPC обычно применяют метод анализа через синтез (*Analysis-by-Synthesis* (AbS)).

При этом синтезатор (основной элемент декодера речевого сигнала) используется как составная часть устройства кодирования (рис. 12.13). На основе формируемых данных произ-

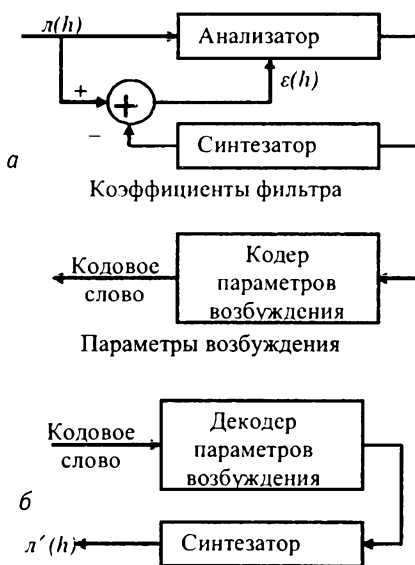


Рис. 12.13. Иллюстрация метода анализа через синтез (AbS): кодер (а) и декодер ($л'(h)$ значения РС для момента h , полученное после декодирования) (б)

водится синтез речевого сигнала, который сравнивается в процессе передачи с реальным сигналом, поступающим на вход устройства. Сигнал ошибки $\varepsilon(h)$, получаемый в результате вычитания истинного и синтезированного сигналов, используется для уточнения формируемых в кодере данных. По существу системы, использующие метод LPC, отличаются лишь способами генерирования возбуждающего воздействия и выбора параметров моделирующего фильтра.

Векторное квантование и кодовые книги. Когда набор значений амплитуд речевого сигнала, дискретизированного по времени, квантуется совместно как единый вектор, такой процесс называется векторным квантованием (*Vector Quantization, VQ*), известный также как блочное квантование. Будем считать, что $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$, N -мерный вектор с действительными значениями (индекс «Т» означает сдвиг по времени); λ_k — случайным образом меняющийся компонент с непрерывной амплитудой, где $1 = k = N$. При векторном квантовании вектору L ставится в соответствие другой N -размерный вектор y , имеющий действительные значения и дискретные амплитуды. Обычно y выбирается из конечного набора значений $Y = y_i, 1 = i = L, y_i = [y_{i1}, y_{i2}, \dots, y_{iN}]^T$. Набор Y называется кодовой книгой или шаблоном, L — размер кодовой книги, y_i — набор векторов кодовой книги.

Структурная схема простого векторного квантователя представлена на рис. 12.14. В линию связи передаются только индексы i векторов y_i входящих в кодовую книгу. Некоторые кодовые книги рассчитываются заранее и не изменяются. Они называются фиксированными кодовыми книгами. Другие кодовые книги могут обновляться в процессе работы.

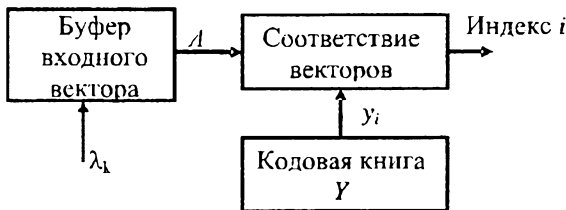


Рис. 12.14. Структурная схема векторного квантователя

Одним из способов сделать кодovou книгу следящей за характеристиками входного вектора с течением времени является ее адаптация. Такие кодовые книги называются адаптивными. При обработке речевых сигналов применяются также случайные кодовые книги. Примером такой книги может быть гауссовская кодовая книга, которая содержит случайным образом выбранные векторы, сами содержащие случайные числа. Векторное квантование может осуществляться не только с использованием значений амплитуд дискретизированного по времени сигнала. Многомерный вектор можно сформировать на основе гармонических составляющих спектра передаваемого речевого сигнала и создать соответствующую кодовую книгу, которая содержит конечное количество значений такого вектора.

Метод кодирования CELP. Кодеры речевых сигналов, использующие алгоритм CELP, относятся к классу гибридных и занимают промежуточное положение между кодерами формы, в которых сохраняется форма колебания речевого сигнала в процессе его дискретизации и квантования, и параметрическими вокодерами, основанными на процедурах оценки и кодирования небольшого числа параметров речи. При этом в кодерах CELP сохраняются преимущества обоих типов кодеров. Метод кодирования CELP основан на линейной авторегрессионной модели процесса формирования и восприятия речи и входит в группу методов анализа через синтез. Линейная авторегрессионная модель процесса формирования речевых сигналов с локально постоянными на интервалах 10–30 мс параметрами получила в настоящее время широкое распространение. Для этой модели:

$$\lambda(h) = \sum_{m=1}^M a_m \lambda(h-m) + x(h),$$

где M — порядок модели; $\lambda(h)$ — последовательность отсчетов речевого сигнала; a_m — коэффициенты линейного предсказания, характеризующие свойства голосового тракта; $x(h)$ — сигнал возбуждения голосового тракта (порождающая последовательность).

Авторегрессионная модель речевого сигнала описывает его с достаточно высокой точностью и позволяет применять развитый математический аппарат линейного предсказания. Ее

применение обеспечивает более высокое качество декодированной речи, устойчивость к входному акустическому шуму и ошибкам в канале связи по сравнению с иными принципами кодирования. При использовании метода анализа через синтез задача анализа сводится к процедуре оценки передаваемых в канал связи параметров речи, проводимой в соответствии с некоторым критерием рассогласования между исходным и декодированным (синтезированным) сигналом. Метод CELP эффективно применяется при передаче речевых сигналов со скоростями, большими или равными значению 4 кбит/с.

12.9. MPEG-4 CELP в системе DRM

В системе DRM применяется вариант 2 метода кодирования MPEG-4 CELP (стандарты ISO/IEC 14496-3 и ISO/IEC 14496-3/Amd1), обеспечивающий повышенную устойчивость против ошибок (*Object Type ID* = 24, который является частью *High Quality Audio Profile*). Метод MPEG-4 CELP в системе DRM обеспечивает кодирование и декодирование речевых сигналов при скоростях цифровых потоков на выходах коде-ров от 4 до 24 кбит/с. Для данного метода в DRM предусмотрены две частоты дискретизации: $f_d = 8$ кГц и $f_d = 16$ кГц. Соответствующие значения диапазонов воспроизводимых звуковых частот:

100 Гц – 3800 Гц для $f_d = 8$ кГц;

50 Гц – 7000 Гц для $f_d = 16$ кГц.

Структурная схема декодера MPEG-4 CELP представлена на рис. 12.15. Генератор возбуждения содержит адаптивную кодовую книгу для моделирования периодических компонентов, фиксированные кодовые книги для моделирования случайных компонентов и декодер усиления для восстановления уровня речевого сигнала. Индексы кодовых книг (повышение/понижение тона для адаптивной кодовой книги, индексы моделей для фиксированных кодовых книг, индексы усиления) используют для генерации возбуждающего сигнала. Сигнал, созданный этим генератором, поступает на вход линейного синтезирующего фильтра с предсказанием (*Linear Predictive Synthesis Filter-LP-SynthesisFilter*). Коэффициенты фильтра восстанавливаются на основе принятых LPC-индексов, которые предварительно интерполируются. Значения этих

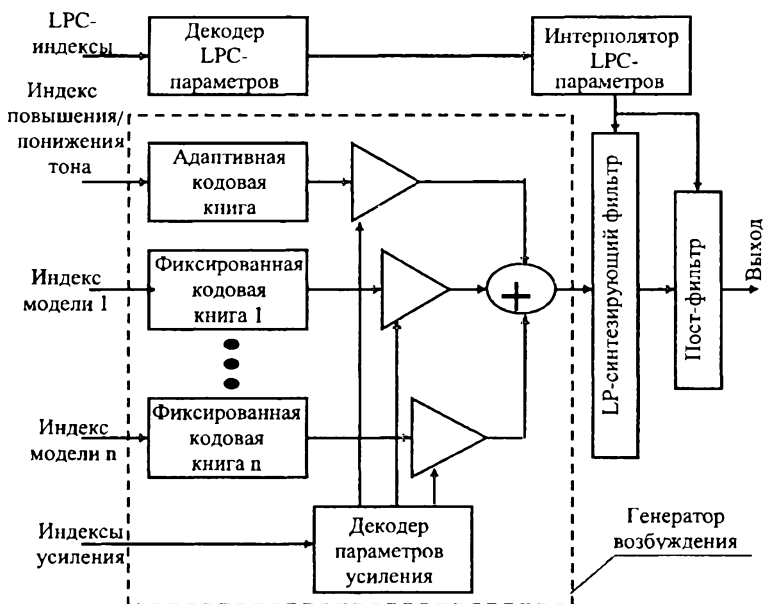


Рис. 12.15. Структурная схема декодера MPEG4 CELP

коэффициентов поступают на вход синтезирующего фильтра. На выходе декодера может быть установлен так называемый «пост-фильтр». Пост-фильтр осуществляет фильтрацию декодированного речевого сигнала в целях улучшения качества восприятия речи. Типичная схема пост-фильтра содержит три основных элемента: долговременный пост-фильтр, кратковременный пост-фильтр и устройство масштабирования усиления. Имеются также вспомогательные элементы.

Долговременный пост-фильтр, иногда называемый пост-фильтром основного тона речи, представляет собой гребенчатый фильтр, спектральные пики которого расположены на частотах, кратных частоте основного сигнала, подлежащего фильтрации. Основная задача кратковременного пост-фильтра заключается в ослаблении частотных составляющих между пиками формант. Устройство масштабирования усиления обеспечивает одинаковый уровень речевого сигнала до и после обработки в пост-фильтре. Звуковые кадры, формируемые кодером MPEG-4 CELP, имеют фиксированную длину.

Эти кадры объединяются в сверхкадры, длительность которых 400 мс. Применяется неравная защита от ошибок (UEP). Начало каждого кадра имеет повышенную защиту от ошибок; оставшиеся биты размещаются в части кадра с нормальной защитой. Индексы, указывающие скорость цифрового потока, передаются в канал SDC.

12.10. Метод кодирования MPEG-4 HVXC

В системе DRM применяется вариант MPEG4 HVXC, обеспечивающий повышенную устойчивость против ошибок (стандарты ISO/ IEC 14496-3 и ISO/IEC 14496-3/Amd1). Применение этого метода обеспечивает передачу цифровых речевых сигналов с коммуникационным качеством при скоростях ЦП 2 или 4 кбит/с. Полоса воспроизводимых частот 100–3800 Гц. Частота дискретизации $f_d = 8$ кГц. В дополнение к этому имеется функция изменения высоты звучания и скорости речи. Эти функции полезны для быстрого поиска нужного фрагмента речи в базе данных.

MPEG-4 HVXC обеспечивает различные категории устойчивости к ошибкам и может применяться в каналах передачи, подверженных влиянию ошибок.

Имеется возможность маскировки ошибок, что специально предусмотрено для системы DRM. Применяется метод равной защиты от ошибок (EER). Длина звукового кадра равна 20 мс. Из таких кадров формируются сверхкадры, длительность которых равна 400 мс. Подробное описание метода MPEG-4 HVXC применительно к системе DRM приведено ниже.

12.11. Формирование системных каналов DRM

Канал MSC. В канале MSC передается пользовательская информация, поступающая от создателей программ радиовещания. Многопрограммный групповой цифровой поток (назовем его для краткости «мультиплексом»), формируемый в канале MSC, может содержать от одной до четырех программ радиовещания, каждая из которых является либо звуковой (и, в частности, речевой) программой, либо данными. Общая скорость передачи мультиплекса зависит от ширины по-

лосы частот, занимаемых сигналом DRM, и от режима передачи. Мультиплекс содержит от одного до четырех цифровых потоков. Каждый цифровой поток разделен на логические кадры, имеющие длительность 400 мс.

Цифровой поток, в котором передается сжатый звуковой сигнал, может также содержать текстовые сообщения. Цифровой поток данных может состоять из четырех субпотоков, в каждом из которых передаются данные в пакетном режиме. Один субпоток переносит пакеты данных одной службы передачи данных. Звуковой канал DRM содержит один звуковой цифровой поток и, кроме того, может содержать один поток или один субпоток данных. Канал данных DRM содержит один поток или один субпоток данных. Логический кадр состоит из двух частей, каждая из которых имеет свой уровень защиты от ошибок. Из логических кадров всех цифровых потоков формируются кадры мультиплекса, длительность каждого из которых равна 400 мс. Эти кадры, пройдя обработку в блоках рандомизации, поступают на вход канального кодера MSC. Информация о конфигурации мультиплексирования в MSC передается в канале SDC. При необходимости конфигурация мультиплексирования может быть изменена. В этом случае параметры новой конфигурации мультиплексирования передаются через канал SDC, а точное время изменения конфигурации указывается в FAC. Канал MCS представляет собой последовательность кадров мультиплекса. Кадры мультиплекса формируются из логических кадров и содержат части с повышенной и нормальной защитой от ошибок, в которые помещается информация из соответствующих частей логических кадров.

Канал FAC. Канал FAC используется для передачи информации быстрого доступа, позволяющей приемнику без задержек находить требуемый радиоканал DRM. Канал FAC содержит информацию о параметрах радиоканала DRM (например, полоса частот, занимаемых радиоканалом, количество и типы программ, которые передаются в данном канале, виды модуляции в MSC и SDC). Канал FAC содержит также информацию о программах, подготовленных соответствующими службами, которые передаются в звуковых (речевых) ка-

налах или каналах данных, входящих в MSC. Это позволяет владельцу DRM-приемника либо декодировать этот мультиплекс, либо продолжить перестройку по частоте. Информация FAC формируется в виде FAC-блоков, которые затем используются при построении кадров передачи, рассмотренных ниже. Один FAC-блок содержит информацию о параметрах программы, подготовленной одной службой. Если в мультиплексе содержатся программы нескольких служб, то для их описания требуется несколько FAC-блоков.

Число битов L_{FAC} в SDC блоке равно 72 в каждом режиме. Общее количество N_{FAC} OFDM-ячеек в FAC блоке равно 65 в каждом режиме.

FAC-ячейки передают высокозащищенные QAM-символы, которые позволяют приемнику быстро определять тип принимаемого сигнала.

Параметры радиоканала

Параметры радиоканала включают в себя следующее:

- Флаг «Базовый/улучшенный»: 1-битовый флаг указывает уровень DRM передачи; при базовом уровне передачи радиосигнал может быть полностью декодирован обычным DRM-приемником; при улучшенном уровне передачи радиосигнал может быть полностью декодирован DRM-приемником с дополнительными возможностями.
- «Идентификация»: 2-битовое поле идентифицирует номер текущего сверхкадра передачи и применимость индекса AFS (индекс переключения альтернативных частот) в канале SDC.
- «Частотный диапазон»: в 4-битовом поле указаны частотные параметры радиосигнала DRM.
- Флаг «Глубина перемежения»: 1-битовый флаг определяет длительность процесса перемежения.
- «Режим MSC»: 2-битовое поле предназначено для указания режима модуляции канала MSC.
- «Режим SDC»: 1-битовый флаг указывает режим модуляции канала SDC.
- «Количество программ»: 4-битовое поле указывает общее количество программ радиовещания, подготовлен-

ных звуковыми службами и службами передачи данных, в радиоканале DRM.

- «Индекс реконфигурации»: 3-битовое поле указывает статус и время реконфигурации мультиплексирования.
- «gfi»: два бита зарезервированы для будущего применения; до тех пор их значения устанавливаются равными 0.

Параметры передаваемой программы

- «Идентификатор службы передачи программ радиовещания»: 24-битовое поле указывает уникальный идентификационный номер данной службы.
- «Короткий идентификатор»: 2-битовое поле идентифицирует программу, а также звуковой (речевой) канал или канал данных MSC, в котором она передается; короткий идентификатор используется как справочная информация для канала SDC.
- «Индикация условного доступа»: 1-битовый флаг указывает на наличие или отсутствие условного доступа к данной программе.
- «Язык вещания»: 4-битовое поле указывает язык вещания.
- «Флаг звук/данные»: 1-битовый флаг указывает тип программы.
- «Описание программы»: содержание этого 5-битового поля зависит от значения флага «звук/данные»; если этот флаг равен 0, то указывается тип звуковой программы (новости, погода, жанр музыки и т. д.); если названный флаг равен 1, то указывается идентификатор применения канала данных.
- «gfa»: 7 битов зарезервированы для будущих дополнений (устанавливаются равными 0, пока не определены).

Канал SDC. В канале SDC передается информация о способах декодирования MSC, о поиске альтернативных источников одной и той же информации и идентификации программ, входящих в мультиплекс. Информация SDC формируется в виде блоков. Далее эти блоки используются при построении сверхрадгов передачи.

Число битов L_{SDC} в SDC блоке зависит от режима помехоустойчивости, занятия спектра и числа сигналов M (4, 16, 64). Общее количество N_{SDC} OFDM-ячеек в блоке SDC приведено в табл. 12.2.

Таблица 12.2

Число QAM-ячеек N_{SDC} для SDC

Уровень помехозащищенности	Степень заполнения спектра					
	0	1	2	3	4	5
A	167	190	359	405	754	846
B	130	150	282	322	588	662
C				288		607
D				152		332

SDC-блок сформирован следующим образом:

- индекс AFS (*Alternative Frequency Switching*) — 4 бита;
- поле данных — n байтов;
- кодовое слово CRC (*Cyclic Redundancy Check*) — 16 битов.

4-битовый AFS-индекс — это беззнаковое двоичное число, которое указывает количество сверхкадров передачи, отделяющих данный SDC-блок от следующего с тем же содержанием. Таким образом осуществляется поиск альтернативных источников одинаковых программ. Поле данных служит для передачи данных SDC. Длина этого поля зависит от режима устойчивости радиоканала, режима SDC и ширины полосы частот, занимаемых радиоканалом DRM. CRC — это 16-битовое кодовое слово CRC, служит для проверки наличия ошибок в блоке SDC. Всего в канале SDC может передаваться 12 категорий данных:

- данные типа 1 — метки (лейблы) программ;
- данные типа 2 — условный доступ;
- данные типа 3 — информация о частотах радиосигналов DRM, AM, УКВ ЧМ, DAB;
- данные типа 4 — список частот радиосигналов для быстрого поиска;
- данные типа 5 — информация о применении; содержит описание всех передаваемых программ;
- данные типа 6 — поддержка объявлений и переключений;

- данные типа 7 — определение географических регионов;
- данные типа 8 — текущее время и дата;
- данные типа 9 — описание параметров звуковых каналов, необходимое для их декодирования;
- данные типа 10 — параметры FAC;
- данные типа 11 — данные установления связи, позволяющие находить альтернативные источники одинаковых программ;
- данные типа 12 — язык вещания и страна.

Другие типы данных зарезервированы для будущих применений.

12.12. Канальное кодирование и модуляция

В системных каналах MSC, FAC, SDC передается различная информация, и для ее кодирования и преобразования целесообразно применять различные схемы. Из-за различных требований по защите от ошибок в пределах одного сервиса или для различных сервисов в пределах одного мультиплексного потока используют различные схемы отображения и комбинаций скоростей кода: доступны неравная защита от ошибок (*Unequal Error Protection, UEP*) и равная защита от ошибок (*Equal Error Protection, EEP*) и возможно их объединение с иерархической модуляцией. Равная защита от ошибок использует одинаковую скорость кода для защиты всех данных в канале. EEP обязательна для FAC и SDC. В отличие от EEP, неравная защита от ошибок может использоваться с двумя скоростями кода, чтобы позволить разбиение данных в MSC на высокозащищенную и низкозащищенную части. При использовании иерархической модуляции применяются три стратегии отображения для MSC: стандартное отображение (*Standard mapping, SM*), симметричное иерархическое отображение (*Symmetrical Hierarchical Mapping, HMsym*) и смесь двух предыдущих отображений (*HMmix*), которая использует иерархическое отображение для реальной компоненты и стандартное отображение для мнимой компоненты совокупного сигнала. Иерархическое отображение разбивает декодируемый поток данных на две части: очень сильно защищенную часть

(*Very Strongly Protected Part*, VSPP) и стандартно защищенную часть (*Standard Protected Part*, SPP). Метод SM состоит только из SPP. В любом случае, до двух различных полных скоростей кода будет применяться для SPP MSC. Для FAC и SDC допустим только SM.

Кратко схема кодирования представлена на рис. 12.16.

Адаптивное транспортное мультиплексирование. Различные каналы (*MSC*, *SDC*, *FAC*) обрабатываются в независимых канальных кодерах. Длина вектора обработки L равна одному блоку *FAC* для *FAC*, одному блоку *SDC* для *SDC* или одному мультиплексному кадру для *MSC*.

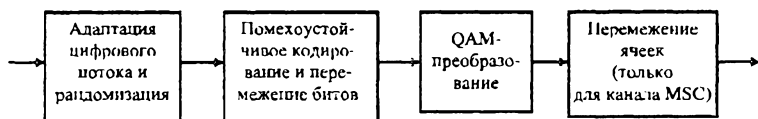


Рис. 12.16. Структурная схема процесса кодирования информации в системных каналах MSC, FAC и SDC

Число битов L_{MUX} в мультиплексном кадре зависит от режима помехоустойчивости, занятия спектра и числа сигналов M (4, 16, 64):

- при использовании одного уровня защиты (*EEP*) $L_{MUX} = L_2$;
- при использовании двух уровней защиты (*UEP*) $L_{MUX} = L_1 + L_2$, где число бит высокозащищенной части — L_1 и число бит низкозащищенной части — L_2 ;
- при использовании *HMsum* или *HMmix* число очень сильно защищенных бит обозначается через L_{VSPP} .

Для случая SM имеем:

$$L_1 = \sum_{p=0}^{P_{max}-1} 2 N_1 R_p, \quad L_2 = \sum_{p=0}^{P_{max}-1} R X_p \left\lfloor \frac{2N_2 - 12}{R Y_p} \right\rfloor \text{ и } L_{VSPP} = 0.$$

Здесь P_{max} — число уровней (для 4-QAM $P_{max} = 1$, для 16-QAM $P_{max} = 2$, для 64-QAM $P_{max} = 3$; RX_p , RY_p , R_p — числитель, знаменатель скорости и скорость кода каждого индивидуального уровня, $\lfloor \rfloor$ — означает округление к минус бесконечности.

Для YM_{sym} :

$$L_1 = \sum_{p=0}^2 2 N_1 R_p ,$$

$$L_2 = \sum_{p=0}^2 RX_p \left[\frac{2N_2 - 12}{RY_p} \right] ,$$

$$L_{VSP} = RX_0 \left[\frac{2(N_1 + N_2) - 12}{RY_0} \right] .$$

Здесь $P_{max} = 3$ — число уровней для 64-QAM, использующей HM_{sym} . Важно отметить, что иерархическая схема отображения может использоваться только с 64-QAM сигналом.

Для HM_{mix} имеем:

$$L_1 = N_1 R_0^{Im} + \sum_{p=1}^2 N_1 (R_p^{Re} + R_p^{Im}) ,$$

$$L_2 = RX_0^{Im} \left[\frac{N_2 - 12}{RY_0} \right] + \sum_{p=1}^2 RX_p^{Re} \left[\frac{N_2 - 12}{RX_p} \right] + RX_p^{Im} \left[\frac{N_2 - 12}{RY_p} \right]$$

$$L_{VSP} = RX_0^{Re} \left[\frac{N_1 + N_2 - 12}{RY_0^{Re}} \right] .$$

Здесь $P_{max} = 3$ — число уровней для 64-QAM, использующей HM_{mix} . RX_p^{Re} , RX_p^{Im} , RY_p^{Re} , RY_p^{Im} , R_p^{Re} , R_p^{Im} — знаменатели, числители скорости и скорость кода каждого индивидуального уровня соответственно.

Общее количество N_{MUX} OFDM-ячеек в мультиплексном кадре при использовании одного уровня защиты (ECP) равно N_2 . Общее количество N_{MUX} OFDM-ячеек в мультиплексном кадре при использовании двух уровней защиты (UEP) равно сумме ячеек высокозащищенной и низкозащищенной и низкозащищенной части:

$$N_{MUX} = N_1 + N_2 ,$$

где N_1 — число OFDM-ячеек, используемых для высокозащищенной и низкозащищенной части, N_2 — число OFDM-ячеек, используемых для низко защищенной части, включающей остаточные биты.

Система DRM состоит из трех разных каналов передачи цифровых данных: MSC, SDC и FAC. Вследствие различного

назначения в этих каналах применяются разные варианты кодирования и преобразования данных, которые предполагают последовательное выполнение следующих четырех операций:

- адаптивное транспортное мультиплексирование и рассеивание энергии сигнала (скремблирование);
- кодирование канала и перемежение бит;
- преобразование потока данных к виду, обеспечивающему реализацию выбранного вида многопозиционной модуляции: 4-, 16- или 64-КАМ;
- перемежение ячеек — только для канала MSC (перемежение N_f MSC-ячеек в пределах каждого мультиплексного фрейма, являющегося базовой структурой для кодирования и перемежения).

Перечисленные операции необходимы для согласования систем модуляции и кодирования. Эффект достигается при иерархическом разбиении ансамбля сигналов на набор вложенных подансамблей с монотонно возрастающими расстояниями между ними и таким подбором кодов для каждого уровня иерархии, чтобы выровнять результирующие расстояния. Принадлежность сигналов к подансамблям кодируется соответствующим числом символов, которые передаются по каналу с использованием СК с определенной кодовой скоростью. Этим реализуется разный уровень защиты бит, принадлежащих различным «созвездиям» — уровням иерархии ансамбля сигналов. Рассмотрим эти важные процедуры более подробно.

Мультиплексирование с разными схемами преобразования потока данных в сочетании с соответствующей скоростью СК позволяет сделать канальное кодирование многоуровневым. Применение нескольких уровней кодирования предполагает использование разных уровней защиты от ошибок: неравной защиты (UEP), равной (EEP) и их комбинации при иерархической модуляции (только в канале MSC), реализующих в той или иной степени согласование процедур модуляции и кодирования. Заметим, что в каналах FAC и SDC обязательна равная защита.

При использовании модуляции применяются три стратегии преобразования данных:

- стандартная модуляция (SM);

- симметричная иерархическая модуляция (HM_{sym});
- смешанная преобразованная иерархическая модуляция (HM_{mix}).

Иерархическое преобразование расщепляет кодируемый поток данных на две части: очень сильно защищенную (VSPP) и часть со стандартной защитой (SSP) и помимо мультиплексных предполагает формирование иерархических фреймов. Стандартное преобразование предполагает только стандартную защиту. Для каналов FAC и SDC приемлемо только стандартное преобразование. Сводная информация о характере мультиплексирования данных, вариантах модуляции и защиты от ошибок приведена в табл. 12.3.

Таблица 12.3

Мультиплексирование данных и защита от ошибок

Канал	Вариант модуляции	Схема мультиплексирования		Вариант защиты от ошибок
MSC	16-QAM	SM($P_{max} = 2$)		EEP+UEP/SSP
	64-QAM	Иерархическая модуляция ($P_{max} = 3$)	SM	EEP+UEP/SSP
			HM_{sym}	EEP+UEP/VSSP
			HM_{mix}	EEP+UEP/VSSP
SDC	16-QAM	SM($P_{max} = 2$)		EEP+UEP/SSP
	4-QAM	SM($P_{max} = 1$)		EEP /SSP
FAC	4-QAM	SM($P_{max} = 1$)		EEP /SSP

Помехоустойчивое кодирование в разных каналах (*MSC*, *SDC*, *FAC*) осуществляется независимо друг от друга. Вектор длины L (длина блока кода) для обработки последовательности данных в кодере канала равен одному *FAC*-блоку, одному *SDC*-блоку или одному мультиплексному фрейму *MSC*.

Число бит L_{mix} в мультиплексном фрейме зависит от уровня помехоустойчивости ($A...D$), числа сигналов M (4; 16; 64) и ширины полосы частот радиоканала. В стандарте DRM искомое число бит определяется аналитической зависимостью параметров P_{max} , RX_p и RY_p — для каждого индивидуального уровня защиты P ($P \in \{0, P_{max} - 1\}$), где P_{max} — число уровней кодирования (для 4-QAM равно единице, для 16-QAM — двум и для 64-QAM — трем), RX_p — числитель, а RY_p — знаменатель в выражении кодовой скорости $R_p = RX_p / RY_p$ для

каждого индивидуального уровня. Все эти параметры заданы таблично. У преобразования HM_{mix} они определены для реальной (Re) и мнимой (Im) компонент R_p .

Как известно, сигнал в системе DRM состоит из множества $Kh [k_{min}, k_{max}]$ модулированных несущих. Поэтому каждый OFDM-символ (ячейка) может рассматриваться как разделенный на элементарные пакеты, каждый из которых переносится одной несущей во время одного символа. Количество бит V , переносимое одной несущей за время OFDM-символа, зависит от числа позиций M при QAM-модуляции, а именно: $v = \log_2 M$, т. е. 2, 4 или 6 бит. Однако полное число N_{MUX} MSC OFDM-символов в мультиплексном фрейме — величина переменная, зависящая от числа (1...4) уровней защиты и вида преобразования (SM , HM_{sym} или HM_{mix}). Эта величина задана таблично для всех видов преобразований.

Что касается канала FAC , то число бит L_{FAC} FAC -блока равно 72, а общее число N_{FAC} FAC OFDM-ячеек FAC -блока равно 65 в каждом режиме помехоустойчивости. В этом канале используются версии EEP и SM и модуляция исключительно 4-QAM. В канале SDC число бит L_{SDC} в SDC -блоке зависит от уровня помехоустойчивости, кратности модуляции и ширины полосы частот радиоканала. Общее число N_{SDC} SDC OFDM-ячеек SDC -блока в стандарте задано таблично. Здесь используются 4-QAM и 16-QAM.

Скремблирование применяется во всех трех каналах системы. Цель скремблирования состоит в том, чтобы избежать передачи образцов сигнала, которые могли бы привести к нежелательной регулярности в передаваемом сигнале. Для SDC и FAC выход скремблера должен формировать входной поток u_i для соответствующего многоуровневого процесса кодирования. Выход скремблера, действующего в MSC мультиплексном кадре, должен формировать стандартно защищенный входной поток u_i для многоуровневого процесса кодирования для MSC. Выход скремблера, используемого в иерархическом фрейме (если такой используется) должен формировать очень сильно защищенный поток u_i для того же самого многоуровневого процесса кодирования. Скремблирование применяется на различных каналах (MSC, SDC, FAC), чтобы уменьшить вероятность того, что систематические образцы приве-

дут к нежелательной регулярности либо в передаваемом сигнале, либо в любой цифровой обработке, что обеспечивается детерминированным выборочным дополнением битов. Индивидуальные входы скремблеров, показанных на рис. 12.17, должны складываться по модулю 2 с псевдослучайной двоичной последовательностью (*Pseudo-Random Binary Sequence*, PRBS) до канального кодирования. В качестве полинома $P(X)$, задающего ПСП, используется порождающий полином $P(X) = X^9 + X^5 + 1$. ПСП генерируется регистром сдвига с обратной связью, показанным на рис. 12.17. Начальное состояние инициализации должно быть таким, чтобы первый бит PRBS формировался, когда выходы всех триггеров сдвигового регистра установлены в «1»; первые 16 битов PRBS формируются, как показано в табл. 12.4.

Таблица 12.4

Первые 16 битов последовательности PRBS

Номер бита	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Значение бита	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	0

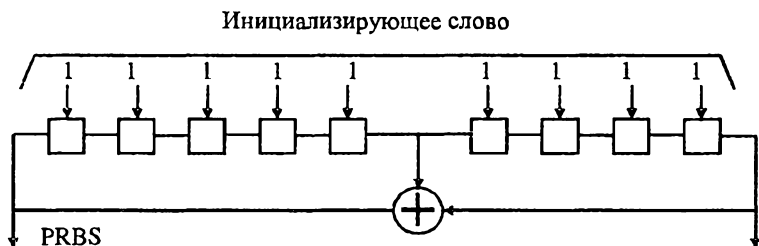


Рис. 12.17. Генератор ПСП для скремблирования цифрового потока

FAC, SDC и MSC обрабатываются скремблером следующим образом:

- длина вектора для обработки равна одному FAC-блоку для FAC, одному SDC-блоку для SDC и одному мультиплексному кадру и одному иерархическому кадру для MSC;

- каждый FAC блок состоит из 72 битов, длины блоков для SDC и MSC зависят от режима помехоустойчивости, занятия спектра и числа сигналов M (4, 16, 64);
- все четыре блока обрабатываются независимо. Входной вектор скремблируется с PRBS, первый бит вектора складывается по модулю 2 с PRBS-битом с номером 0.

Скремблеры различных каналов сбрасываются следующим образом:

- FAC: каждый FAC блок;
- SDC: каждый SDC блок;
- MSC: каждый мультиплексный кадр для стандартно защищенной части, каждый иерархический кадр для очень сильно защищенной части.

Многоуровневое кодирование

Процесс кодирования канала основан на многоуровневой схеме кодирования. Принцип многоуровневого кодирования представляет собой совместную оптимизацию кодирования и модуляции, чтобы достигнуть наилучшего исполнения передачи. Это означает, что разрядные позиции в QAM отображении, подвергаемые большему количеству ошибок, получают более высокую защиту. Различные уровни защиты достигаются различными составными кодами, которые реализованы методом выкалывания и получают от одного и того же родительского кода.

Декодирование в приемнике можно сделать либо непосредственно, либо через итерационный процесс. Следовательно, исправляющая способность декодера может быть увеличена при увеличении числа итераций и, следовательно, зависит от исполнения декодера.

В зависимости от используемого числа сигналов M (4, 16, 64) и отображения применяют пять различных схем. Схема с одним уровнем рассматривается как специальный случай многоуровневой схемы кодирования. Различные схемы отображения применимы только к 64-QAM, как это показано на рис. 12.26, 12.27 и 12.28. Для стандартного отображения и симметричной иерархической модуляции (SM и HM_{sym}), одинаковые отображения должны использоваться для реальных и мнимых компонент совокупного сигнала. Для смешанного отображения с иерархической модуляцией (HM_{mix}) отдельные

отображения будут использоваться для реальных и мнимых компонент совокупного сигнала.

Многоуровневое кодирование канала на практике предназначено для обеспечения максимальной защиты таких бит QAM-модуляции, которые наиболее подвержены ошибкам. Разные уровни защиты реализуются с помощью различных компонент сверточного кода с индивидуальными (парциальными) скоростями $R_p = m/n$. Эти компоненты образуются в кодере путем выкалывания символов в базовом (материнском) коде, характеризуемом скоростью R_B .

Одноуровневая схема рассматривается как специальный случай многоуровневого кодирования. Для стандартной и симметричной иерархической модуляции идентичные преобразования данных используются для реальных и для мнимых компонент сигнала QAM-модуляции. Для смешанной схемы HM_{mix} разные преобразования используются для реальных и мнимых компонент манипуляционного сигнала. В качестве примера на рис. 12.18–12.20 приведены схемы соответственно одно-, двух- и трехуровневого кодирования.

В стандарте описаны алгоритмы деления исходного цифрового потока на парциальные потоки для обеспечения всех видов преобразований: SM, HM_{sym} и HM_{mix} . Число парциальных потоков, а также общее (R_Σ) и парциальное (R_p) зна-

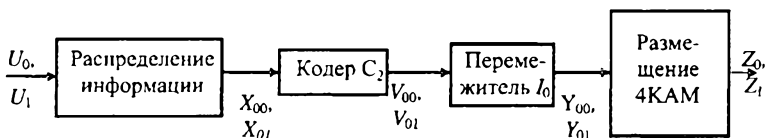


Рис. 12.18. Схема одноуровневого кодирования для SM

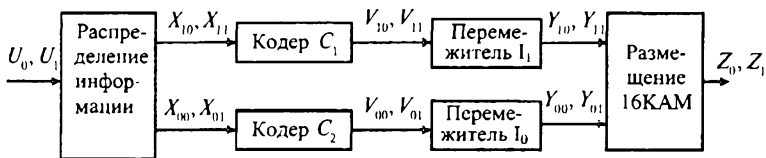


Рис. 12.19. Схема двухуровневого кодирования для SM

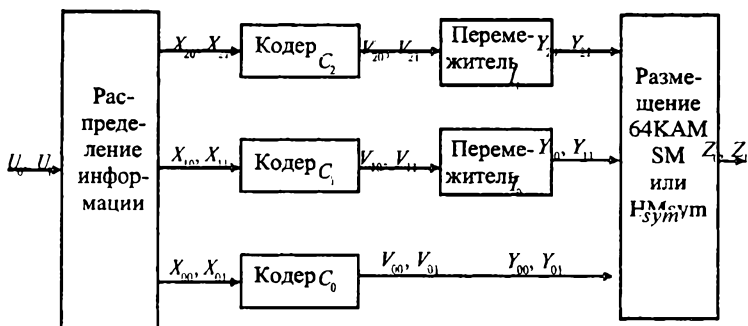


Рис. 12.20. Схема трехуровневого кодирования для SM и HMsym

чения скорости кода в кодерах каждого уровня кодирования определяются числом уровней кодирования P_{max} и защиты p .

Компоненты кода C_p в кодерах на рис. 12.18–12.21 образуются в результате сверточного кодирования с выкалыванием материнского кода со скоростью $R_b = 1/4$ и длиной кодового ограничения $k = 7$. Код задан порождающими многочленами 133, 171, 145 и 133 (в восьмеричной форме записи). Кодер (рис. 12.22) представляет собой регистр сдвига с ячейками, в котором символы кодовой последовательности формируются суммированием по модулю 2 символов с выходов некоторых ячеек.

Кодовое слово на выходе сверточного кодера определяется последовательностью процедур:

$$b_{0,i} = a_i \oplus a_{i-2} \oplus a_{i-3} \oplus a_{i-5} \oplus a_{i-6};$$

$$b_{1,i} = a_i \oplus a_{i-1} \oplus a_{i-2} \oplus a_{i-3} \oplus a_{i-6};$$

$$b_{2,i} = a_i \oplus a_{i-1} \oplus a_{i-4} \oplus a_{i-6};$$

$$b_{3,i} = a_i \oplus a_{i-2} \oplus a_{i-3} \oplus a_{i-5} \oplus a_{i-6};$$

для всех $i = 0, 1, 2, \dots, I + 5$. При этом если i не принадлежит множеству $\{0, 1, 2, \dots, I - 1\}$, то $a_i = 0$ по определению.

Коммутатор на выходе устанавливает очередность посылки кодовых символов в канал. Регистр сдвига можно рассматривать либо как регистр, содержимое которого сдвигается на один разряд вправо при введении в него слева каждого нового двоичного символа, либо как цифровую линию задержки, в которой каждый элемент задержки хранит один двоичный сим-

вол до поступления нового входного двоичного символа. Заметим, что скремблер и кодер канала идентичны одноименным устройствам, применяемым в системе *Eureka-147/DAB* (T-DAB). В стандарте представлена таблица выкалывания базовой (материнской) кодовой последовательности для реализации парциальных значений скорости кода. В канале MSC неравная защита от ошибок обеспечивает два уровня защиты, а в случае 64-QAM иерархическое преобразование обеспечивает для части канала получение также и третьего уровня защиты.

Структурная схема трехуровневого кодирования для HM_{mix} показана на рис. 12.21.

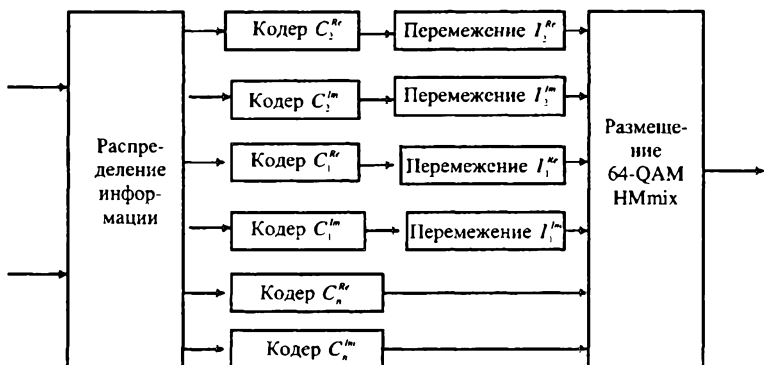


Рис. 12.21. Структурная схема трехуровневого кодирования для HM_{mix}

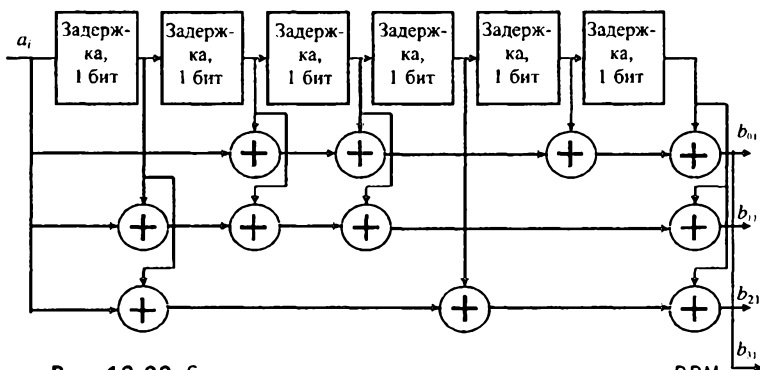


Рис. 12.22. Структурная схема кодера канала стандарта DRM

При стандартном преобразовании (SM) два уровня защиты в одном мультиплексном фрейме реализуются с помощью двух значений полной скорости кода. В табл. 12.5 и 12.6 соответственно для сигналов 16- и 64-КАМ указаны значения полной скорости R_x и парциальных скоростей кода R_p , соответствующих каждому уровню защиты: в случае 16-КАМ — для двух значений полной скорости и для $p \in \{0,1\}$, а в случае 64-КАМ — для четырех значений R_x и для $p \in \{0,1,2\}$. Информация о защищаемом уровне передается в данных канала SDC. Как отмечалось выше, стандартное преобразование SM предполагает только стандартную защиту.

Таблица 12.5

Скорости кода и уровни защиты при SM (16-QAM)

Уровень защиты	R_y	R_0	R_1
0	0,5	1/3	2/3
1	0,62	1/2	3/4

Таблица 12.6

Скорости кода и уровни защиты при SM (64-QAM)

Уровень защиты	R_y	R_0	R_1	R_2
0	0,5	1/4	1/2	3/4
1	0,6	1/3	2/3	4/5
2	0,71	1/2	3/4	7/8
3	0,78	2/3	4/5	8/9

Для преобразования HM_{sym} в табл. 12.7 представлены сведения по полным и парциальным скоростям кода для обеспечения режимов стандартной защиты (SPP) части потока и очень сильной защиты (VSPP) другой части общего потока.

Данные по числу L бит мультиплексного и иерархического фреймов, когда реализуется режим равной (EEP) защиты, приведены в стандарте для возможных значений полос канала и режимов помехоустойчивости.

В канале SDC используется фиксированная полная скорость кода. Для 16-КАМ полная и парциальные скорости (для двух уровней) приведены в табл. 12.8. В случае 4 — к QAM — уровень один: $R_x = R_0 = 0,5$. Данные по числу L бит в SDC-

Скорости кода и уровни защиты при HMsym (64-QAM)

Уровень защиты	При стандартной (SPP) защите			При очень сильной (VSPP) защите
	R_y	R_1	R_2	
0	0,45	3/10	3/5	1/2
1	0,55	4/11	8/11	4/7
2	0,72	4/7	7/8	3/5
3	0,78	2/3	8/9	2/3

Скорости кода в канале SDC (16-QAM)

R_y	R_0	R_1
0,5	1/3	2/3

блоке приведены в стандарте для возможных значений полос канала и режимов помехоустойчивости.

В канале FAC используется один уровень защиты, поэтому $R_x = R_0 = 0,6$.

Перемежение бит парциальных цифровых потоков (рис. 12.23-12.25) осуществляется квазислучайным образом в каналах FAC, SDC и MSC независимо друг от друга, с учетом схемы преобразования последовательности бит для определенного числа M позиций КАМ-модуляции. Перемежение бит представляет собой блочный процесс, то есть оно осуществляется в пределах фиксированной области данных. Размеры этой области и правила перестановки бит определены аналитически для каждого субпотока данных.

В табл. 12.9 обобщены сведения о взаимосвязи вида преобразования при мультиплексировании и модуляции с режимом системы помехоустойчивости и шириной полосы частот радиоканала (т. е. параметром «заполнение спектра»).

Адаптация цифровых потоков. Как отмечалось выше, в системе DRM для защиты информации в каналах применяются методы равной и неравной защиты от ошибок (EER) и (UEP) соответственно. При равной защите от ошибок используется помехоустойчивый код с одной и той же избыточностью для защиты всей информации в канале. При неравной защите от ошибок для кодирования наиболее чувствительной к ошибкам

Взаимосвязь основных параметров системы

Канал	Вид преобразования и модуляции/число уровней защиты	Режим помехоустойчивости	Параметр «заполнение спектра»					
			0	1	2	3	4	5
MSC	SM 64-QAM/4 16-QAM/2	A	X	X	X	X	X	X
		B	X	X	X	X	X	X
		C				X		X
		D				X		X
	HM _{sym} +HM _{mix} 64-QAM/4 для иерархических и мультимплексных фреймов	A			X	X	X	X
		B			X	X	X	X
		C				X		X
		D				X		X
SDC	16-QAM/1 4-QAM/1	A	X	X	X	X	X	X
		B	X	X	X	X	X	X
		C				X		X
		D						X

части информации применяются коды с большей избыточностью и большей исправляющей способностью (информация с повышенным уровнем защиты).

Для кодирования остальной части информации применяются коды с меньшей избыточностью (информация с нормальным уровнем защиты). В канале MSC могут применяться оба метода: UEP и EEP. В каналах FAC и SDC применяется EEP. Кроме того, в канале MSC может применяться иерархическая модуляция, для которой определены три стратегии преобразования:

Standard Mapping (SM — стандартное преобразование);

Symmetrical Hierarchial Mapping (HM_{sym} — симметричное иерархическое преобразование);

Mixture Hierarchial Mapping (HM_{mix} — смешанное иерархическое преобразование).

При иерархических преобразованиях HM_{sym} и HM_{mix} цифровой поток дополнительно разделяется на две части:

Very Strongly Protected Part (VSPP — очень сильно защищенная часть) и

Standard Protected Part (SPP — часть со стандартной защитой).

При стандартном преобразовании (*SM*) имеется только *SPP*.

Канальное кодирование цифровых потоков. При канальном кодировании информация, передаваемая в *MSC*, *FAC* и *SDC*, обрабатывается независимо. Обработке подвергаются кадры *MSC*, блоки *FAC* и *SDC*. Для канального кодирования в системе *DRM* применяются перфорированные сверточные коды, полученные из одного базового кода, скорость которого $R = 1/4$, а длина кодового ограничения $L = 7$. В зависимости от требований к уровням защиты информации от ошибок скорость кода может варьироваться от $R = 1/4$ до $R = 8/9$. Далее производится временное перемежение битов. Это позволяет повысить устойчивость системы против групповых ошибок.

Квадратурная амплитудная модуляция QAM. В системе *DRM* после канального кодирования и перемежения битов производится преобразование информации в так называемые «QAM-ячейки» (QAM — квадратурная амплитудная модуляция). Принцип такого вида модуляции можно пояснить с помощью рис. 12.23, *а, б, в*, на которых представлены векторные диаграммы и временные зависимости сигналов для 4-позиционной QAM (4-QAM). QAM-сигналы удобно представлять в виде векторов на комплексной плоскости (рис. 12.23, *а*), где на осях I и Q отмечаются действительные и мнимые величины соответственно. Для преобразования 4-QAM возможно, например, изображение, приведенное на рис. 11.23, *а*, где каждый вектор отображает гармонический сигнал с определенной амплитудой и начальной фазой, представить в виде рис. 12.23, *в*.

Кроме того, каждому из четырех положений векторов на рис. 12.23, *а* можно поставить в соответствие двоичную кодовую комбинацию, содержащую 2 бита (00, 01, 10, 11). Для удобства концы векторов изображают в виде точек. Такие и более сложные построения называются «сигнальными созвездиями» (*signal constellation*).

Поскольку в случае 4-QAM амплитуды всех сигналов (или значения длин векторов) одинаковы, то такую модуляцию можно рассматривать и как 4-позиционную фазовую модуляцию (манипуляцию) (QPSK — *Quadrature Phase Shift Keying*). В *DRM* применяются также виды модуляции 4-QAM, 16-QAM и 64-QAM, представленные на рис. 12.24, 12.25, 12.26 соответственно. В этих случаях каждому вектору на комплексной плоскости (комплексному символу модуляции z) соответствует

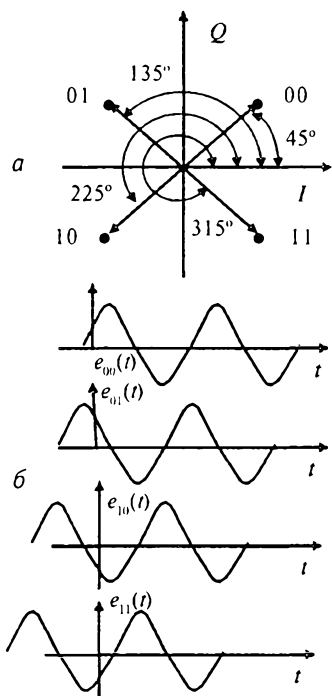


Рис. 12.23. Векторные (а) и временные диаграммы (б) сигналов 4-QAM

гармоническое колебание, имеющее одно из 16 или 64 возможных сочетаний амплитуды и начальной фазы или такое же количество кодовых комбинаций, каждая из которых соответствует определенному варианту гармонического колебания. В случае 16-QAM такие комбинации содержат по 4 бита информации, а в случае 64-QAM — по 6 битов. При увеличении числа позиций QAM увеличивается пропускная способность канала связи в пропорции $\log_2 64 : \log_2 16 : \log_2 4 = 6 : 4 : 2$, однако снижается помехоустойчивость, поскольку уменьшаются разности между смежными значениями амплитуд и фаз. Кодовые комбинации, состоящие из 2, 4 или 6 битов, в системе DRM называются «QAM-ячейками».

Кодирование и QAM-преобразование информации, передаваемой в MSC. В MSC может применяться либо 64-QAM, либо 16-

QAM. 64-QAM обеспечивает высокую спектральную эффективность, т. е. более высокую пропускную способность радиоканала при заданной полосе занимаемых частот, в то время как 16-QAM обеспечивает большую устойчивость к ошибкам, возникающим при передаче сигналов.

В каждом случае диапазон выбранных скоростей сверточного кода должен обеспечивать наиболее приемлемый уровень коррекции ошибок для данного канала передачи. Существующие комбинации сигнальных созвездий и скоростей сверточного кода обеспечивают высокую степень гибкости в широком диапазоне вариантов каналов передачи. Неравная защита от ошибок UEP применяется для обеспечения двух уровней защиты информации в MSC. При использовании 64-QAM и с

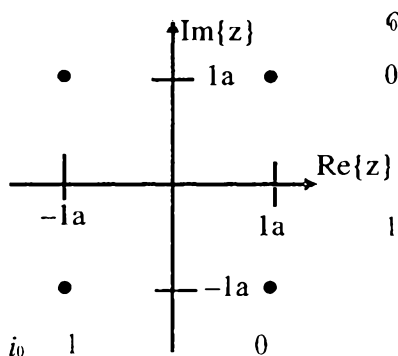


Рис. 12.24. Преобразование

4-QAM и соответствующие сочетания битов:

a — нормирующий коэффициент ($a = 1/\sqrt{2}$), z — комплексный символ модуляции, y_0', y_1' — биты, представляющие z ;
 порядок передачи битов: $\{i_0 q_0\} = \{y_0' y_1'\}$

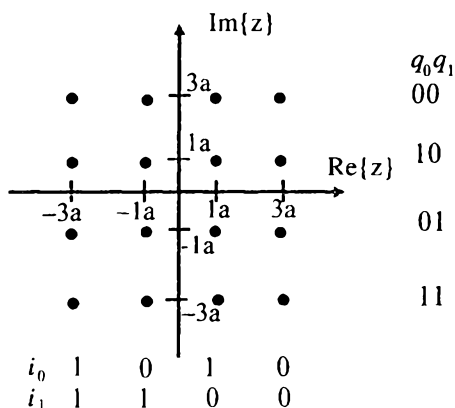


Рис. 12.25. Стандартное преобразование SM 16-QAM и соответствующие сочетания битов; a — нормирующий коэффициент ($a = 1/\sqrt{10}$);

z — комплексный символ модуляции; $y_0' \dots y_3'$ — биты, представляющие z ; порядок передачи битов:
 $\{i_0 i_1 q_0 q_1\} = \{y_0' y_1' y_2' y_3'\}$

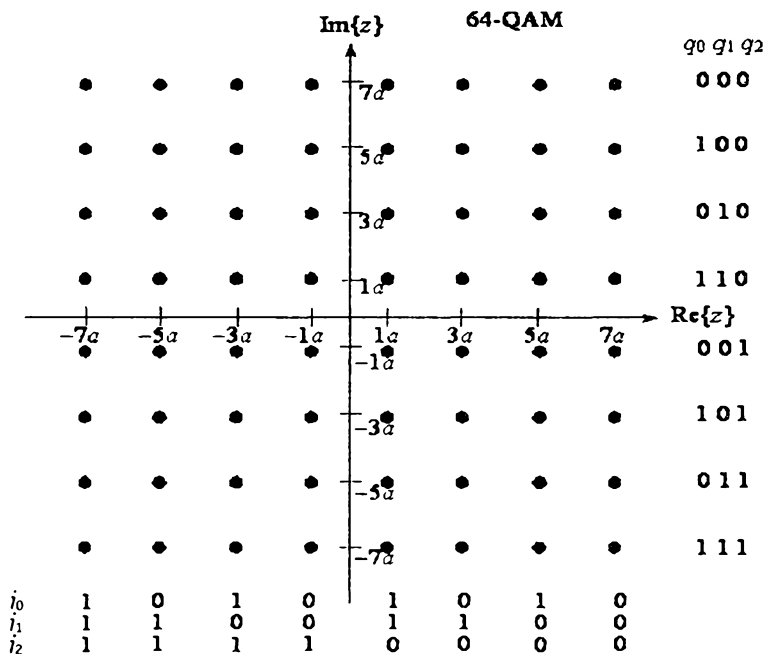


Рис. 12.26. Преобразование *SM* 64-QAM и соответствующие сочетания битов: a — нормирующий коэффициент ($a = 1/\sqrt{42}$), z — комплексный символ модуляции, y_0', y_1' — биты, представляющие z ; порядок передачи битов: $\{i_0 i_1 i_2 q_0 q_1 q_2\} = \{y_0' y_1' y_2' y_3' y_4' y_5'\}$

помощью иерархической модуляции можно обеспечить третий (высший) уровень устойчивости против ошибок для части MSC.

Кодирование и QAM-преобразование информации, передаваемой в SDC. В канале SDC может использоваться либо 16-QAM (рис. 12.15), либо 4-QAM (рис. 12.16). 16-QAM обеспечивает большую пропускную способность канала, а 4-QAM — более высокую устойчивость к ошибкам. Применяется равная защита от ошибок ЕЕР, т. е. фиксированная скорость сверточного кода для всей передаваемой информации. Сигнальное созвездие следует выбирать с учетом параметров MSC таким образом, чтобы обеспечить большую устойчивость SDC, по сравнению с MSC.

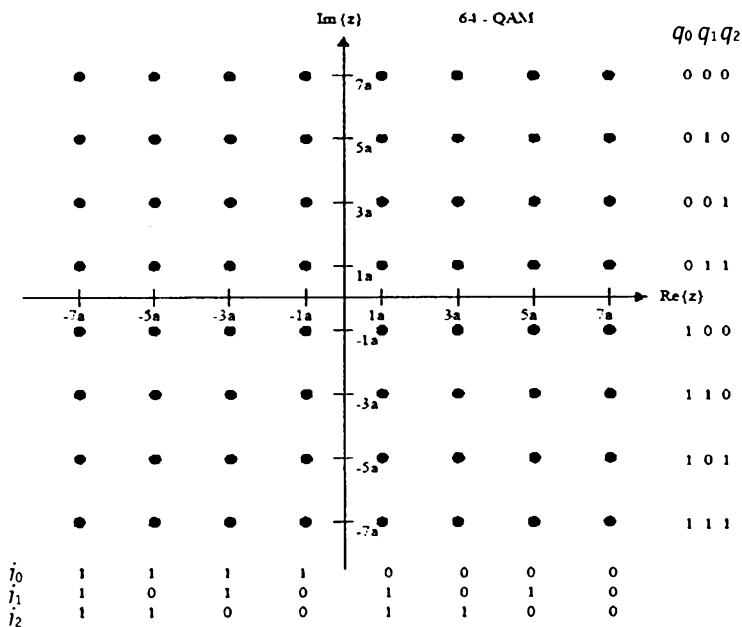


Рис. 12.27. Преобразование *HMsym* 64-QAM
и соответствующие сочетания битов:

a — нормирующий коэффициент ($a = 1/\sqrt{42}$), z — комплексный символ модуляции, y_0', y_1' — биты, представляющие z ; порядок передачи битов: $\{i_0 i_1 i_2 q_0 q_1 q_2\} = \{y_0' y_1' y_2' y_3' y_4' y_5'\}$

Созвездия сигналов и их отображение

Стратегия отображения для каждой OFDM-ячейки зависит от принадлежности к каналу (FAC, SDC, MSC) и режима помехоустойчивости. Ячейки данных могут быть 4-QAM, 16-QAM или 64-QAM. Заданный по умолчанию метод отображения должен выполняться согласно рисункам 12.18–12.22.

Поток данных на выходе перемежителя состоит из множества битовых слов. Они отображаются на одну точку сигнала в диаграмме сигнала согласно комплексному числу z . Для SM и *HMsym* 64-QAM диаграмма должна использоваться согласно рис. 12.26 и 12.27 соответственно. Для *HMmix* 64-QAM долж-на использоваться диаграмма согласно рис. 12.28. Биты

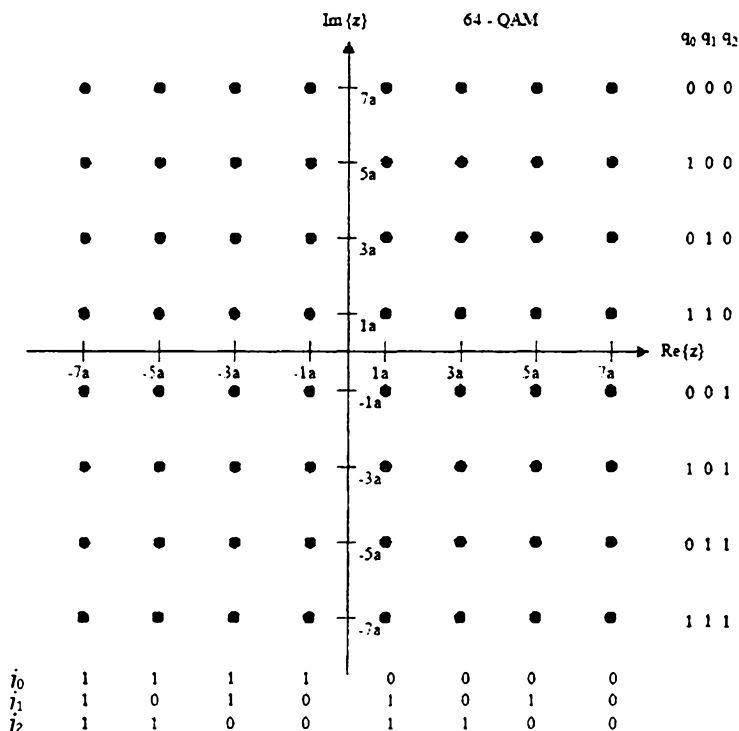


Рис. 12.28. Преобразование *HMmix* 64-QAM

и соответствующие сочетания битов:

a — нормирующий коэффициент ($a = 1/\sqrt{42}$), z — комплексный символ модуляции, y_0', y_1' — биты, представляющие z ; порядок передачи битов: $\{i_0 i_1 i_2 q_0 q_1 q_2\} = \{y_0' y_1' y_2' y_3' y_4' y_5'\}$

должны отображаться в соответствии с выражением $(y_0' y_1' y_2' y_3' y_4' y_5') = (y_{0,0}^{Re} y_{1,0}^{Re} y_{2,0}^{Re} y_{0,0}^{Im} y_{1,0}^{Im} y_{2,0}^{Im})$.

Кодирование и QAM-преобразование информации, передаваемой в FAS. Для передачи информации в FAS применяется 4-QAM преобразование (рис. 12.24). Используется равная защита от ошибок ЕЕР.

Перемежение QAM-ячеек в канале MSC. Перемежение QAM-ячеек должно применяться после многоуровневого кодирования с возможностью выбора малой или большой глубины пе-

ремежения (обозначим их как «короткое» и «длинное» перемежение) в соответствии с ожидаемыми условиями распространения сигнала DRM. Базовые параметры перемежения адаптированы к размеру кадра мультиплекса. Для радиоканалов с умеренными замираниями, что типично для передачи сигнала посредством земной волны в диапазонах ДВ, СВ, короткое перемежение обеспечивает приемлемые условия для правильной работы декодера в приемнике. При сложных условиях распространения радиосигналов DRM, т. е. существенных замираниях, что характерно для отраженных от ионосферы волн в диапазонах КВ, глубина перемежения может быть увеличена. Общая задержка информации, передаваемой в MSC, при перемежении/деперемежении на передающей и приемной сторонах соответственно, примерно равна 800 мс для короткого и 2.4 с для длинного перемежения. Перемежение QAM-ячеек в каналах FAC и SDC не производится в целях исключения временных задержек при передаче системной информации. Тем не менее, эта информация имеет достаточно высокую степень защиты от ошибок.

12.13. Структура передаваемых сигналов

Построение OFDM-сигналов. После перемежения QAM-ячейки преобразуются в OFDM-символы. Каждый OFDM-символ передается за время T_s и представляет собой совокупность K несущих, равномерно распределенных в заданном частотном диапазоне. OFDM-символ — это передаваемый сигнал для такого промежутка времени, в течение которого значения амплитуд и фаз всех QAM-ячеек, модулирующих несущие OFDM-сигнала, остаются постоянными. Из определенного количества OFDM-символов формируются кадры передачи. Первый OFDM-символ каждого кадра передачи содержит ячейку опорного времени. Длительность кадра передачи — 400 мс. Из трех кадров передачи формируется сверхкадр передачи. Длительность сверхкадра передачи — 1200 мс. В начале сверхкадра передачи размещается SDC-блок. Длительность символа представляет собой сумму двух частей:

- полезная часть, длительность которой равна T_u ;
- защитный интервал; его длительность равна T_g .

Расстояние между соседними несущими OFDM-сигнала равно $1/T_u$. Защитный интервал располагается перед полезной частью символа. В нем размещается фрагмент полезной части символа. Введение защитных интервалов в OFDM-символы позволяет бороться с межсимвольной интерференцией в радиоканалах с многолучевым распространением сигналов и подверженных эффекту Доплера. OFDM-символы в кадре передачи нумеруются от 0 до $N_s - 1$. Все символы содержат данные и опорную информацию. При наличии многолучевого распространения и частотно-селективных замираний в радиоканалах DRM часть несущих, входящих в OFDM-сигнал, может быть ослаблена или вообще исчезнуть. При этом, однако, информация, содержащаяся в таких несущих, во многих случаях может быть восстановлена по оставшимся неповрежденными несущим за счет помехоустойчивого кодирования, а также перемежения битов и QAM-ячеек. В случае невозможности восстановить информацию в символе или ряде символов применяется маскировка ошибок. Такие технические решения обуславливают высокую надежность и устойчивость приема сигналов в системе DRM. Так как OFDM-сигнал состоит из множества отдельно модулированных несущих, каждый символ может быть рассмотрен как разделенный на OFDM-ячейки, соответствующие текущей модуляции одной несущей в течение времени передачи одного символа. Кадр передачи содержит следующие виды OFDM-ячеек:

- пилот-ячейки;
- ячейки управления;
- ячейки данных.

Пилот-ячейки могут использоваться для кадровой, частотной и временной синхронизации, оценки канала передачи и режима устойчивости. Передаваемый сигнал описывается следующим выражением:

$$x(t) = \operatorname{Re} \left\{ e^{j2\pi f_R t} \sum_{r=0}^{\infty} \sum_{s=0}^{N_s-1} \sum_{k=K_{\min}}^{K_{\max}} C_{\gamma,s,k} \Psi_{\gamma,s,k}(t) k \right\},$$

где

$$\Psi_{\gamma,s,k}(t) = \begin{cases} e^{j2\pi \frac{k}{T_u}(t-T_R-sT_s-N_s r T_s)} & \text{при } (s + N_s r)T_s \leq t \leq (s + N_s r + 1)T_s, \\ 0 & \text{в других случаях} \end{cases}$$

где N_s — количество OFDM-символов в кадре передачи, k — обозначение номера несущей ($= K_{\min}, \dots, K_{\max}$), s — обозначение номера OFDM-символа ($= 0N_s - 1$), r — обозначение номера кадра передачи ($= 0$ неопределенность), T_g — длительность защитного интервала, T_u — длительность полезной части OFDM-символа, T_s — длительность OFDM-символа, K — количество передаваемых несущих ($= K_{\max} - K_{\min}$), f_R — опорная частота радиосигнала DRM, $c_{r,s,k}$ — значение комплексной QAM-ячейки для несущей k в символе s кадра r . Значения $c_{r,s,k}$ зависят от типа ячеек.

Для OFDM-ячеек данных и ячеек управления $c_{r,s,k} = z$, где z — это точка сигнального созвездия для каждой QAM-ячейки (комплексный символ модуляции).

Параметры OFDM, относящиеся к ширине полосы частот, которые занимает канал DRM. Параметры OFDM зависят от полосы частот, занимаемых каналом DRM, количества несущих и их расположения по отношению к опорной частоте f_R (т. е. частоте несущей OFDM-сигнала, которой присвоен номер $k = 0$). *Полоса частот, занимаемая спектром радиосигнала DRM, определяется номинальной шириной канала DRM.* Группа несущих, передающих информацию ФАС, всегда расположена справа (выше по частоте) по отношению к опорной частоте f_R , значение которой выбирается как целое число, кратное 1 кГц (рис. 12.29). Светлые прямоугольники означают группу несущих DRM-сигнала, занимающую полосу 4,5 или 5 кГц. Темные прямоугольники означают группу несущих DRM-сигнала, которая

Параметр ширины полосы частот DRM-сигнала, передаваемый в ФАС:

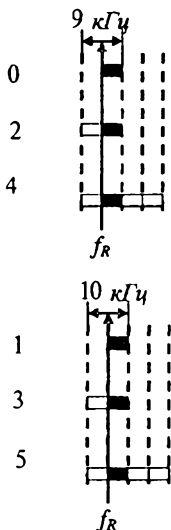


Рис. 12.29. Изображения спектров DRM-сигналов при значениях ширины занимаемых ими полос, кратных величинам 4,5 кГц и 5 кГц

содержит информацию FA, занимающих полосу частот 4,5 или 5 кГц. В табл. 12.10 приведены параметры (индексы), передаваемые в FAC, которые определяют ширину полос частот, а также соответствующие значения ширины полос частот.

В системе DRM предусмотрена возможность совместной передачи в одном канале сигнала цифрового радиовещания и аналогового вещательного сигнала с амплитудой (AM) или однополосной модуляцией (ОМ) с верхней или нижней боковой полосой (ВБП и НБП соответственно). Варианты построения таких объединенных радиосигналов представлены на рис.12.30, 12.31.

Таблица 12.10

Параметр, передаваемый в FAC, и ширина полосы частот

Ширина полосы частот, занимаемых каналом DRM						
Параметр, передаваемый в FAC	0	1	2	3	4	5
Ширина полосы частот, кГц	4.5	5	9	10	18	20

В системе DRM предусмотрена возможность совместной передачи в одном канале сигнала цифрового радиовещания и аналогового вещательного сигнала с амплитудой (AM) или однополосной модуляцией (ОМ) с верхней или нижней боковой полосой (ВБП и НБП соответственно). Варианты построения таких объединенных радиосигналов представлены на рис.12.30, 12.31.

На рис. 12.31 показаны некоторые возможности для случаев, при которых опорная частота DRM-сигнала f_R отстоит от несущей частоты f_c аналогового сигнала с AM или ОМ на расстояние, равное одному или двум каналам (т. е. ± 9 ; ± 10 ; -18 ; -20 кГц).

На рис. 12.31 представлен вариант, при котором опорная частота DRM-сигнала f_R отстоит от несущей частоты аналогового сигнала f_c на величину, равную половине ширины канала. Поскольку опорная частота f_R цифрового DRM-сигнала должна быть целой и кратной 1 кГц, в последнем случае разность частот f_R и f_c равна 4 и 5 кГц.

Параметры OFDM, относящиеся к условиям распространения радиосигналов DRM. Параметры OFDM должны выби-

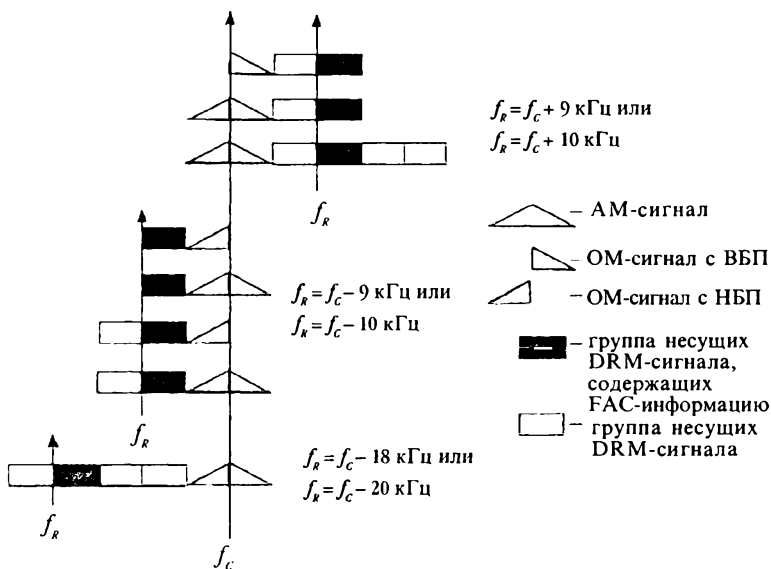


Рис. 12.30. Примеры изображений спектров, формируемых при одновременной передаче сигналов аналогового радиовещания и DRM-сигналов; сдвиг между несущими частотами аналоговых сигналов и опорными частотами DRM сигналов равен: $\pm 9, \pm 10, -18, -20 \text{ кГц}$

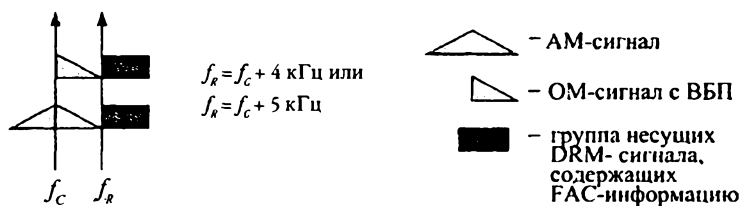


Рис. 12.31. Примеры изображений спектров, формируемых при одновременной передаче сигналов аналогового радиовещания и DRM-сигналов; сдвиг между несущими частотами аналоговых сигналов и опорными частотами DRM-сигналов равен: 4 или 5 кГц

ратся в соответствии с условиями распространения радиосигналов DRM и размерами района, который оператор планирует обслуживать. В системе DRM определены четыре категории устойчивости в зависимости от условий распространения радиосигналов (табл. 12.11).

Таблица 12.11

Типичные условия распространения радиосигналов DRM

Режимы устойчивости	Типичные условия распространения радиосигналов
A	Гауссовские каналы с малыми замираниями
B	Каналы, имеющие частотную и временную селективность, с увеличенными задержками распространения
C	Аналогично режиму B, но с повышенным эффектом Доплера
D	Аналогично режиму B, но с существенными задержками распространения и значительным эффектом Доплера

Параметры OFDM-символов в зависимости от режимов устойчивости представлены в табл. 12.12.

Таблица 12.12

Параметры OFDM-символов и режимы устойчивости

Перечень параметров		Режимы устойчивости			
		A	B	C	D
Элементарный временной период	T , мкс	$83 \frac{1}{3}$	$83 \frac{1}{3}$	$83 \frac{1}{3}$	$83 \frac{1}{3}$
Длительность полезной части символа	T_u , мс	24 ($288 \times T$)	$21 \frac{1}{3}$ ($256 \times T$)	$83 \frac{2}{3}$ ($176 \times T$)	$9 \frac{1}{3}$ ($112 \times T$)
Длительность защитного интервала	T_g , мс	$2 \frac{2}{3}$ ($32 \times T$)	$5 \frac{1}{3}$ ($64 \times T$)	$5 \frac{1}{3}$ ($64 \times T$)	$7 \frac{1}{3}$ ($88 \times T$)
	T_g / T_u	1/9	1/4	4/11	11/14
Длительность символа	$T_s = T_u + T_g$, мс	$26 \frac{2}{3}$	$26 \frac{2}{3}$	20	$16 \frac{2}{3}$
Разнос несущих	$1/T_u$, Гц	$41 \frac{2}{3}$	$46 \frac{7}{8}$	$68 \frac{2}{11}$	$107 \frac{1}{7}$

Перечень параметров		Режимы устойчивости			
		A	B	C	D
Длительность кадра передачи	$T_f, \text{мс}$	400	400	400	400
Количество символов в кадре	N_s	15	15	20	24

Количество несущих в OFDM-символах в зависимости от режимов устойчивости и значений индексов ширины полос частот, занимаемых радиосигналами DRM, указаны в табл. 12.13, где $k \in [K_{\min}, K_{\max}]$; $k = 0$ соответствует несущей с частотой, равной опорной f_R ; значения $k < 0$ означают, что несущие размещены ниже по частоте, чем f_R , а $k > 0$ соответствуют несущим, расположенным выше по частоте, чем f_R .

Таблица 12.13

Индексы ширины полосы частот

Режим устойчивости	Минимальный и максимальный номер несущей	Индекс ширины полосы частот					
		0	1	2	3	4	5
A	$K_{\min}$	2	2	– 102	– 114	–98	– 110
	$K_{\max}$	102	114	102	114	314	350
B	$K_{\min}$	1	1	–91	– 103	–87	–99
	$K_{\max}$	91	103	91	103	279	311
C	$K_{\min}$	–	–	–	–69	–	–67
	$K_{\max}$	–	–	–	69	–	213
D	$K_{\min}$	–	–	–	–44	–	–43
	$K_{\max}$	–	–	–	44	–	135

Пилот-ячейки. Некоторые ячейки в кадре передачи OFDM модулируются с заранее известными значениями частот, фаз и амплитуд. Эти ячейки называются пилот-ячейками и служат для оценки канала передачи и синхронизации. Значения

амплитуд и фаз этих ячеек тщательно выбираются в целях оптимизации характеристик системы, особенно начальной синхронизации и надежности приема сигнала. Пилот-ячейки разделяются на три типа:

- ячейки опорных частот;
- ячейки опорного времени;
- ячейки опорного усиления.

Ячейки управления. Имеется два типа ячеек управления:

- ячейки канала быстрого доступа (FAC), которые размещаются в каждом кадре передачи; они обеспечивают быструю передачу информации, которая необходима DRM-приемнику для демодуляции сигнала;
- ячейки канала описания пользовательской информации (SDC); они повторяются в каждом сверхкадре передачи и содержат информацию, которая описывает имеющиеся в наличии программы, конфигурацию мультиплексирования в канале MSC и др. Ячейки SDC используются также для автоматического переключения приемника на альтернативный источник той же программы в случае ухудшения параметров принимаемого сигнала (AFS).

Расположение ячеек FAC и FDC в сверхкадре передачи показано на рис. 12.32.

Ячейки данных. К ячейкам данных относятся такие OFDM-ячейки, которые не относятся к пилот-ячейкам и ячейкам управления и для которых $K_{\min} \leq k \leq K_{\max}$, причем k не принадлежит к подмножеству неиспользуемых ячеек. С помощью ячеек данных передается основная информация, содержащаяся в сигнале DRM.



Рис. 12.32. Временное размещение информации MSC, FAC и SDC в сверхкадре

Сотовые технологии обеспечивают связь между подвижными абонентами и стационарными серверами по радиоканалу, поэтому сотовую связь называют также мобильной связью. Основой развития мобильных сетей являются сотовые топологии.

13.1. Сотовые системы

Сотовая система подвижной радиосвязи (ССПС) использует большое число маломощных передатчиков, которые предназначены для обслуживания только сравнительно небольшой зоны, скажем, радиусом в 1...2 км. Вместо использования единственного передатчика для обслуживания территории города всю территорию можно разбить на множество небольших зон покрытия, называемых *сотами*. Если все имеющиеся в распоряжении частотные каналы могут повторно использоваться в каждой ячейке сотовой структуры, тогда, например, требуемые для 0,1% жителей территории 250 каналов можно получить разделением обслуживаемой территории радиусом в 50 км на 25 ячеек радиусом по 10 км с организацией в каждой ячейке только 10 радиоканалов с одним и тем же набором частот. Из-за недопустимо большого уровня взаимных помех ячейки с одинаковым набором частот необходимо перемежать буферными ячейками с другими наборами частот. Группа ячеек в зоне обслуживания с различными наборами частот называется *кластером*. На рис. 13.1 показан образец сотовой структуры с типичной для аналоговых сетей размерностью кластера $n = 7$.

Если, например, для обслуживания абонентов в одной ячейке требуется набор из 15 частот, то для создания сотовой

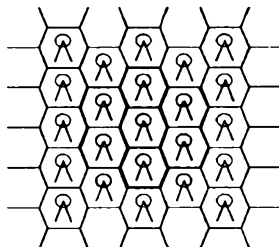


Рис. 13.1. Образец сотовой структуры

структуры с размерностью кластера $n = 7$, обслуживающей сколь угодно большую территорию, необходимо располагать набором из 105 частот.

Основной потенциал сотовой идеи заключается в том, что уровень взаимных помех зависит не от собственно расстояния между ячейками, а от отношения расстояния между ячейками к их радиусу.

Радиус ячейки зависит от мощности передатчика и, определяется разработчиком системы, который в процессе проектирования должен выбрать подходящую размерность кластера. С уменьшением радиуса ячейки возрастает количество базовых станций, приходящихся на 1 кв. км площади обслуживания и на 1 МГц используемой полосы частот.

Полномасштабное развертывание сотовой сети с самого начала ее ввода в эксплуатацию чрезвычайно дорого. Обычно ее развертывание начинается с небольшого числа крупных ячеек, которые через некоторое время постепенно трансформируются в большее число более мелких ячеек. Такой способ преобразования называется *расщеплением*. Когда в некоторой ячейке нагрузка достигает того уровня, при котором существующее в ней число каналов оказывается недостаточным для поддержания установленного качества обслуживания абонентов (т. е. вероятность *непредоставления* канала при поступлении вызова оказывается больше установленного значения, как правило, до 5%), эта ячейка разделяется на несколько более мелких с пониженной мощностью передатчиков. При этом пропускная способность сети на территории расщепленной ячейки увеличивается в число раз, равное числу вновь образованных ячеек. Эта процедура может повторяться до тех пор,

пока сотовая сеть не достигнет своей расчетной пропускной способности.

Понятно, что ячейки небольших размеров требуются только в центральной части города со значительной плотностью абонентов. Ближе к окраинам плотность снижается, и размеры ячеек могут увеличиваться. Расщепление ячеек может производиться достаточно гибко как в пространстве, так и во времени. По замыслу разработчиков сотовой системы она должна явиться чрезвычайно удобным средством в руках проектировщиков для возможности повышения пропускной способности именно там и именно в то время, где и когда это необходимо.

Использование сравнительно небольших ячеек создает *проблему поддержания непрерывности связи*. При движении по произвольному маршруту объект (абонент ССПС) в течение одного сеанса связи может миновать несколько ячеек. В этом случае непрерывность связи обеспечивается способностью системы автоматически передавать связь с объектом тем базовым станциям, в зоне действия которых он оказывается в данный момент.

Благодаря непрерывным измерениям уровней сигналов, поступающих в центр коммутации подвижной связи от базовых станций, ближайших к движущемуся объекту, система может определить момент пересечения объектом границы двух ячеек и переключить разговорный канал из первой ячейки во вторую в течение достаточно малого промежутка времени, не приводящего к нарушению непрерывности разговора. Такая процедура, получившая название *эстафетной передачи* (хэндовер — *handover*), требует весьма сложного алгоритма определения именно той ячейки из нескольких соседних, куда перемещается объект, а также быстродействующих алгоритмов и схемотехнических решений, обеспечивающих освобождение канала в первой ячейке и поиск свободного канала с восстановлением по нему связи во второй ячейке.

Реализация описанных основных принципов сотовой архитектуры (использование маломощных передатчиков с радиопокрытием небольших по размеру ячеек, повторное использование частот в пределах одной зоны обслуживания, поэтапное увеличение пропускной способности за счет расщепления ячеек, обеспечение непрерывности связи в процессе переме-

щения объекта от ячейки к ячейке) привела в начале 80-х годов к созданию в ряде промышленно развитых стран Европы и Северной Америки ССПС, которые положили начало массовому внедрению услуг подвижной связи во всем мире.

Развернутые в 80-х годах ССПС относят к *первому поколению*. К ним относятся стандарты AMPS (США), HSCMTS (Япония), NMT-450 и NMT-900 (Северная Европа), C-450 (Германия), TACS (Великобритания), ETACS (Англия, Лондон), RTMS-101H (Италия) и Radiocom-200 (Франция). Они были рассчитаны в основном на обслуживание абонентов в рамках национальных границ, использовали аналоговую ЧМ для передачи речи и *внутриполосную (in-band)* сигнализацию в процессе установления соединения между абонентскими терминалами и остальной сетью. Исключение составляла лишь система NMT-450 (NMT-900), которая была введена в эксплуатацию в 1981 году как международная система для четырех стран Северной Европы: Дании, Финляндии, Норвегии и Швеции.

Однако аналоговые ССПС уже не удовлетворяют современному уровню развития связи. Тем не менее, один из аналоговых стандартов — NMT-450 — принят в качестве федерального стандарта России. На его основе созданы ССПС в Москве («Московская сотовая связь», начало коммерческой эксплуатации — 1991 год), Санкт-Петербурге («Дельта-Телеком») и других городах. В июне 1994 года началась коммерческая эксплуатация ССПС компании «Би-Лайн», использующей стандарт AMPS. В настоящее время данная ССПС предоставляет услуги цифровой сотовой связи в стандарте D-AMPS, обслуживает более 20 тысяч абонентов в Москве и области и обеспечивает административный *роуминг* с другими сетями этого стандарта.

Системы *второго поколения* проектировались для создания крупномасштабных сетей с учетом обеспечения международного роуминга — автоматического обслуживания абонентов, приехавших со своими терминалами в другую страну.

К настоящему времени разработаны 4 основные стандарта перспективных сетей сотовой подвижной связи с макросотовой топологией сетей и радиусом соты, соответствующим максимальной дальности связи в радиальных системах (≈ 35 км):

- общеевропейский стандарт *GSM (Global System for Mobile communication)*;
- два североамериканских стандарта: *ADC(D-AMPS) (Advanced Digital Communication)* по стандарту TIA IS-54 и *CDMA* по стандарту IS-95;
- японский стандарт *JDC (Japan Digital Communication)*.

Хотя стандарты отличаются своими характеристиками, они построены на единых принципах, использованных в GSM, и отвечают требованиям современных информационных технологий. Наиболее прогрессивным является стандарт GSM, его основные характеристики подробнее рассмотрены ниже.

Стандарт D-AMPS разрабатывался в США с 1987 года. FCC не смогла выделить отдельную полосу частот в диапазоне 900 МГц для перспективной цифровой ССПС США. Ассоциация промышленности сотовой связи (CTIA) совместно с TIA приняли решение о совмещении в одной полосе частот аналоговой ССПС стандарта AMPS и будущей цифровой ССПС, сохранив используемый в AMPS разнос каналов, равный 30 кГц, при использовании речевого кодека VSELP со скоростью преобразования речи 8 кбит/с. Стандарт TIA IS-54 на ССПС ADC (D-AMPS) был принят в 1990 году. Несмотря на то, что D-AMPS представляет собою не полностью цифровое решение (используются аналоговые каналы управления), он оказался более прогрессивным, чем AMPS, и в настоящее время активно развивается в России.

ССПС, использующие кодовое разделение каналов (CDMA), были разработаны фирмой Qualcomm (США) и развиваются фирмой Motorola. На системы CDMA TIA приняла стандарт IS-95. В сентябре 1995 года в Гонконге начата коммерческая эксплуатация первой сети CDMA данного стандарта на оборудовании фирмы Motorola.

В апреле 1991 года был принят японский стандарт цифровой ССПС JDC. Стандарт JDC рассчитан на работу в диапазонах частот 800/900 МГц и 1400/1500 МГц, использует так же, как D-AMPS, временное разделение каналов с тремя временными окнами на несущую. К особенностям JDC следует отнести прямую связь с ISDN, возможность шифрования передаваемых сообщений, применение речевого кодека VSELP со скоростью преобразования речи 11,2 кбит/с, меньший, чем в D-AMPS, разнос частотных каналов — 25 кГц. В целом,

цифровая ССПС Японии во многом не уступает ССПС стандарта GSM и по некоторым параметрам превосходит американскую ССПС стандарта D-AMPS.

В 1982 году СЕРТ в целях изучения и разработки общеевропейской цифровой системы сотовой связи создала рабочую группу, получившую название GSM (*Groupe Special Mobile*). В 1989 году дело создания GSM перешло к ETSI, а в 1990 году были опубликованы спецификации первой фазы GSM. К середине 1991 года стали поддерживаться коммерческие услуги GSM, а к 1993 году функционировало уже 36 сетей GSM в 22 странах и еще 25 стран выбрали направление GSM или поставили вопрос о его принятии. Несмотря на то, что система GSM была стандартизирована в Европе, на самом деле она не является исключительно европейским стандартом. Сети GSM внедрены либо планируются к внедрению более чем в 60 странах Европы, Ближнего и Дальнего Востока, Африки, Южной Америки и в Австралии. В начале 1994 года число абонентов GSM во всем мире достигло 1,3 миллиона человек. К началу 1995 года их насчитывалось уже более 5 миллионов. Акроним GSM приобрел новое значение — *Global System for Mobile communications*.

Система GSM построена на основе новейшей технологии в виде цифровой системы с программным управлением, совместимой с цифровой телефонной сетью общего пользования интегрального обслуживания (ISDN). В ней использованы:

- ЭМ ВОС;
- система сигнализации SS7;
- принципы построения интеллектуальной сети IN/1.

Элементы этой системы способны контролировать и управлять всеми основными характеристиками сигнала в процессе передачи. Система обладает достаточным «интеллектом» для обнаружения возникшего отклонения в работе, его диагностики, принятия решения и проведения необходимой коррекции.

В ней реализована большая часть возможностей ISDN плюс дополнительные возможности, связанные с особенностями подвижной радиосети: управление по радио, слежение за местоположением подвижного объекта, обеспечение функции эстафетной передачи, защита передаваемой информации и т. п. Инфраструктура сети создает и постоянно обновляет

объемные базы данных, содержащие необходимые сведения об абонентах и их местоположении, устраняет все обнаруженные неполадки, модифицирует свою конфигурацию по мере изменения нагрузки и выполняет множество других функций по эксплуатации и обслуживанию сети, тарификации, взаимодействию с другими стационарными и подвижными сетями.

13.2. Структурная схема сети мобильной связи

Первая коммерческая сеть GSM была развернута в Германии в 1992 г.

Стандарт GSM-900 состоит из более чем 100 книг — спецификаций с техническими условиями общим объемом более 5000 страниц, объединенных в 12 серий, от 2 до 30 книг в каждой серии. Ядро стандарта GSM 1800 составляют спецификации GSM. К ним добавлено 14 дополнительных спецификаций, определяющих отличия GSM-1800. В Республике Беларусь налаживается работа систем подвижной связи стандартов GSM-900 и GSM-1800.

Рассмотрим функциональное построение системы GSM, показанное на рис. 13.2. В схему входит центр коммутации подвижной связи MSC (*Mobile Switching Centre*), оборудование базовой станции BSS (*Base Station System*), центр управления и обслуживания OMC (*Operation and Maintenance Centre*) и подвижные станции MS (*Mobile Stations*). Функциональное сопряжение элементов системы осуществляется рядом интерфейсов. Все сетевые функциональные компоненты в GSM взаимодействуют в соответствии с системой сигнализации МККТТ № 7, представляющей собой совокупность сигналов, передаваемых между элементами сети для обеспечения установления и разъединения при обслуживании вызовов, а также для передачи различной служебной информации.

MSC обслуживает группу сот и обеспечивает все виды соединений, необходимых для работы подвижной станции, а также маршрутизацию вызовов и функции управления вызовами. На MSC возложены функции коммутации радиоканалов при перемещении подвижной станции из одной зоны в другую и при появлении помех или возникновении неисправностей системы. MSC осуществляет постоянное слежение за

подвижными станциями, используя регистры положения (HLR — *Home Location Register*) и перемещения (VLR — *Visitor Location Register*). В HLR хранится та часть информации о местоположении какой-либо подвижной станции, которая позволяет MSC доставить вызов станции. Регистр HLR содержит международный идентификационный номер подвижного абонента. Его используют для опознавания подвижной станции в центре аутентификации (AUC — *Authentication Centre*).

Регистр перемещения VLR обеспечивает контроль за перемещением подвижной станции из одной зоны в другую. Он обеспечивает функционирование подвижной станции за пределами зоны, контролируемой HLR. Когда в процессе перемещения подвижная станция переходит из зоны действия одного контроллера базовой станции BSC (*Base Station Controller*) в зону действия другого контроллера BSC, ее регистрирует новый контроллер, и в VLR заносится информация о номере области связи, которая обеспечивает доставку вызовов, подвижной станции. Для сохранности данных, находящихся в HLR и VLR, в случае сбоев предусмотрена защита устройств памяти этих регистров. AUC определяет ключи шифрования абонентских станций на основе базы данных, сосредоточенных в регистре идентификации оборудования EIR (*Equipment Identification Register*).

Практически HLR представляет собой справочную базу данных о постоянно прописанных в сети абонентах. В ней содержатся опознавательные номера и адреса, а также параметры подлинности абонентов, состав услуг связи, специальная информация о маршрутизации. Ведется регистрация данных о роуминге абонента, включая данные о временном идентификационном номере подвижного абонента (TMSI) и соответствующем VLR.

К данным, содержащимся в HLR, имеют дистанционный доступ все MSC и VLR сети и, если в сети имеются несколько HLR, в базе данных содержится только одна запись об абоненте, поэтому каждый HLR представляет собой определенную часть общей базы данных сети об абонентах. Доступ к базе данных об абонентах осуществляется по номеру IMSI или MSISDN (номеру подвижного абонента в сети ISDN). К базе данных могут получить доступ MSC или VLR, относящиеся к

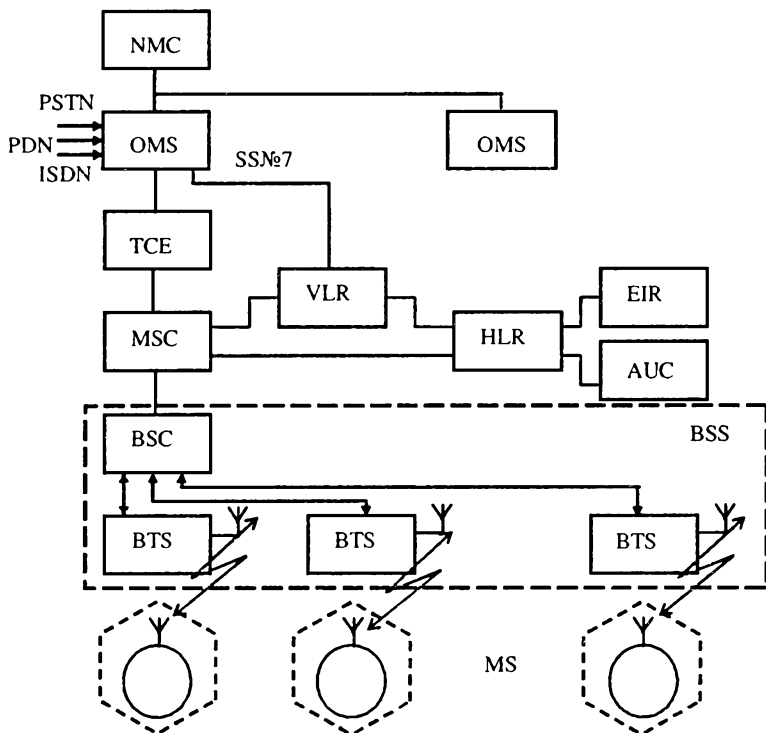


Рис. 13.2. Структурная схема цифровой мобильной связи

другим сетям, в рамках обеспечения межсетевого роуминга абонентов.

Второе основное устройство, обеспечивающее контроль за передвижением подвижной станции из зоны в зону, — регистр перемещения VLR. С его помощью достигается функционирование подвижной станции за пределами зоны, контролируемой HLR. Когда в процессе перемещения подвижная станция переходит из зоны действия одного контроллера базовой станции BSC, объединяющего группу базовых станций, в зону действия другого BSC, она регистрируется новым BSC, и в VLR заносится информация о номере области связи, которая обеспечит доставку вызовов подвижной станции. Для сохранности данных, находящихся в HLR и VLR, в случае сбоев предусмотрена защита устройств памяти этих регистров.

VLR содержит такие же данные, как и HLR, однако эти данные содержатся в VLR только до тех пор, пока абонент находится в зоне, контролируемой VLR.

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

VLR обеспечивает также присвоение номера «блуждающей» подвижной станции (MSRN). Когда подвижная станция принимает входящий вызов, VLR выбирает его MSRN и передает его на MSC, который осуществляет маршрутизацию этого вызова к базовым станциям, находящимся рядом с подвижным абонентом.

VLR также распределяет номера передачи управления при передаче соединений от одного MSC к другому. Кроме того, VLR управляет распределением новых TMSI и передает их в HLR. Он также управляет процедурами установления подлинности во время обработки вызова. По решению оператора TMSI может периодически изменяться для усложнения процедуры идентификации абонентов. Доступ к базе данных VLR может обеспечиваться через IMSI, TMSI или MSRN. В целом, VLR представляет собой локальную базу данных о подвижном абоненте для той зоны, где находится абонент, что позволяет исключить постоянные запросы в HLR и сократить время на обслуживание вызовов.

Для исключения несанкционированного использования ресурсов системы связи вводятся механизмы аутентификации — удостоверения подлинности абонента. Центр аутентификации состоит из нескольких блоков и формирует ключи и алгоритмы аутентификации. С его помощью проверяются полномочия абонента и осуществляется его доступ к сети связи. AUC принимает решения о параметрах процесса аутентификации и определяет ключи шифрования абонентских станций на осно-

ве базы данных, сосредоточенной в регистре идентификации оборудования (EIR — *Equipment Identification Register*).

Каждый подвижный абонент на время пользования системой связи получает стандартный модуль подлинности абонента (SIM), который содержит: международный идентификационный номер (IMSI), свой индивидуальный ключ аутентификации (Ki), алгоритм аутентификации (A3).

С помощью записанной в SIM информации в результате взаимного обмена данными между подвижной станцией и сетью осуществляется полный цикл аутентификации и разрешается доступ абонента к сети.

Процедура проверки сетью подлинности абонента реализуется следующим образом. Сеть передает случайный номер (RAND) на подвижную станцию. На ней с помощью Ki и алгоритма аутентификации A3 определяется значение отклика (SRES), т. е.

$$SRES = Ki * [RAND]. \quad (13.1)$$

Подвижная станция посылает вычисленное значение SRES в сеть, которая сверяет значение принятого SRES со значением SRES, вычисленным сетью. Если оба значения совпадают, подвижная станция приступает к передаче сообщений. В противном случае связь прерывается, и индикатор подвижной станции показывает, что опознавание не состоялось. Для обеспечения секретности вычисление SRES происходит в рамках SIM. Несекретная информация (например, Ki) не подвергается обработке в модуле SIM.

EIR — регистр идентификации оборудования, содержит централизованную базу данных для подтверждения подлинности международного идентификационного номера оборудования подвижной станции (IMEI). Эта база данных относится исключительно к оборудованию подвижной станции. База данных EIR состоит из списков номеров IMEI, организованных следующим образом.

Белый список — содержит номера IMEI, о которых есть сведения, что они закреплены за санкционированными подвижными станциями.

Черный список — содержит номера IMEI подвижных станций, которые украдены или которым отказано в обслуживании по другой причине.

Серый список — содержит номера IMEI подвижных станций, у которых существуют проблемы, выявленные по данным программного обеспечения, что не является основанием для внесения в *черный список*.

К базе данных EIR получают дистанционный доступ MSC данной сети, а также MSC других подвижных сетей.

Как и в случае с HLR, сеть может иметь более одного EIR, при этом каждый EIR управляет определенными группами IMEI. В состав MSC входит транслятор, который при получении номера IMEI возвращает адрес EIR, управляющий соответствующей частью базы данных об оборудовании.

IWF — межсетевой функциональный стык, является одной из составных частей MSC. Он обеспечивает абонентам доступ к средствам преобразования протокола и скорости передачи данных так, чтобы можно было передавать их между его терминальным оборудованием (DIE) сети GSM и обычным терминальным оборудованием фиксированной сети. Межсетевой функциональный стык также «выделяет» модем из своего банка оборудования для сопряжения с соответствующим модемом фиксированной сети. IWF также обеспечивает интерфейсы типа прямого соединения для оборудования, поставляемого клиентам, например для пакетной передачи данных PAD по протоколу X25.

ЕС — эхоподавитель, используется в MSC со стороны PSTN для всех телефонных каналов (независимо от их протяженности) из-за физических задержек в трактах распространения, включая радиоканал, сетей GSM. Типовой эхоподавитель может обеспечивать подавление в интервале 68 миллисекунд на участке между выходом ЕС и телефоном фиксированной телефонной сети. Общая задержка в канале GSM при распространении в прямом и обратном направлениях, вызванная обработкой сигнала, кодированием/декодированием речи, канальным кодированием и т. д., составляет около 180 мс. Эта задержка была бы незаметна подвижному абоненту, если бы в телефонный канал не была включена дифсистема с преобразованием тракта с двухпроводного на четырехпроводный режим, установка которого необходима в MSC, так как стандартное соединение с PSTN является двухпроводным. При соединении двух абонентов фиксированной сети эхо-сигналы отсутствуют. Без включения ЕС задержка от распространения

сигналов в тракте GSM будет вызывать раздражение у абонентов, прерывать речь и отвлекать внимание.

ОМС — центр эксплуатации и технического обслуживания, является центральным элементом сети GSM, который обеспечивает контроль и управление другими компонентами сети и контроль качества ее работы. ОМС соединяется с другими компонентами сети GSM по каналам пакетной передачи протокола X.25. ОМС обеспечивает функции обработки аварийных сигналов, предназначенных для оповещения обслуживающего персонала, и регистрирует сведения об аварийных ситуациях в других компонентах сети. В зависимости от характера неисправности ОМС позволяет обеспечить ее устранение автоматически или при активном вмешательстве персонала. ОМС может обеспечить проверку состояния оборудования сети и прохождения вызова подвижной станции. ОМС позволяет производить управление нагрузкой в сети. Функция эффективного управления включает сбор статистических данных о нагрузке от компонентов сети GSM, записи их в дисковые файлы и вывод на дисплей для визуального анализа. ОМС обеспечивает управление изменениями программного обеспечения и базами данных о конфигурации элементов сети. Загрузка программного обеспечения в память может производиться из ОМС в другие элементы сети или из них в ОМС.

NMC — центр управления сетью, позволяет обеспечивать рациональное иерархическое управление сетью GSM. Он обеспечивает эксплуатацию и техническое обслуживание на уровне всей сети, поддерживаемой центрами ОМС, которые отвечают за управление региональными сетями. NMC обеспечивает управление трафиком во всей сети и обеспечивает диспетчерское управление сетью при сложных аварийных ситуациях, как, например, выход из строя или перегрузка узлов. Кроме того, он контролирует состояние устройств автоматического управления, задействованных в оборудовании сети, и отражает на дисплее состояние сети для операторов NMC. Это позволяет операторам контролировать региональные проблемы и, при необходимости, оказывать помощь ОМС, ответственному за конкретный регион. Таким образом, персонал NMC знает состояние всей сети и может дать указание персоналу ОМС изменить стратегию решения региональной проблемы.

NMC концентрирует внимание на маршрутах сигнализации и соединениях между узлами с тем, чтобы не допускать условий для возникновения перегрузки в сети. Контролируются также маршруты соединений между сетью GSM и PSTN во избежание распространения условий перегрузки между сетями. При этом персонал NMC координирует вопросы управления сетью с персоналом других NMC. NMC обеспечивает также возможность управления трафиком для сетевого оборудования подсистемы базовых станций (BSS). Операторы NMC в экстремальных ситуациях могут задействовать такие процедуры управления, как «приоритетный доступ», когда только абоненты с высоким приоритетом (экстренные службы) могут получить доступ к системе.

NMC может брать на себя ответственность в каком-либо регионе, когда местный ОМС является необслуживаемым, при этом ОМС действует в качестве транзитного пункта между NMC и оборудованием сети. NMC обеспечивает операторов функциями, аналогичными функциям ОМС.

NMC является также важным инструментом планирования сети, так как NMC контролирует сеть и ее работу на сетевом уровне, а следовательно, обеспечивает планировщиков сети данными, определяющими ее оптимальное развитие.

BSS — оборудование базовой станции, состоит из контроллера базовой станции (BSC) и приемопередающих базовых станций (BTS). Контроллер базовой станции может управлять несколькими приемопередающими блоками. BSS управляет распределением радиоканалов, контролирует соединения, регулирует их очередность, обеспечивает режим работы с прыгающей частотой, модуляцию и демодуляцию сигналов, кодирование и декодирование сообщений, кодирование речи, адаптацию скорости передачи для речи, данных и вызова, определяет очередность передачи сообщений персонального вызова.

BSS совместно с MSC, HLR, VLR выполняет некоторые функции, например: освобождение канала, главным образом под контролем MSC, но MSC может запросить базовую станцию обеспечить освобождение канала, если вызов не проходит из-за радиопомех. BSS и MSC совместно осуществляют приоритетную передачу информации для некоторых категорий подвижных станций.

ТСЕ — транскодер, обеспечивает преобразование выходных сигналов канала передачи речи и данных MSC (64 кбит/с ИКМ) к виду, соответствующему рекомендациям GSM по радиоинтерфейсу (Рек. GSM 04.08). В соответствии с этими требованиями скорость передачи речи, представленной в цифровой форме, составляет 13 кбит/с. Этот канал передачи цифровых речевых сигналов называется «полноскоростным». Стандартом предусматривается в перспективе использование полускоростного речевого канала (скорость передачи 6,5 кбит/с).

Снижение скорости передачи обеспечивается применением специального речепреобразующего устройства, использующего линейное предикативное кодирование (LPC), долгосрочное предсказание (LTP), остаточное импульсное возбуждение (RPE — иногда называется RELP).

Транскодер обычно располагается в MSC. Передача цифровых сообщений в направлении к контроллеру базовых станций (BSC) ведется с добавлением к потоку со скоростью передачи 13 кбит/с, дополнительных битов (стаффинг) до скорости передачи данных 16 кбит/с. Затем осуществляется уплотнение с кратностью 4 в стандартный канал 64 кбит/с. Так формируется определенная Рекомендациями GSM 30-канальная ИКМ линия, обеспечивающая передачу 120 речевых каналов. Шестнадцатый канал (64 кбит/с), «канальный интервал», выделяется отдельно для передачи информации сигнализации и часто содержит трафик SS № 7 или LAPD. В другом канале (64 кбит/с) могут передаваться также пакеты данных, согласующиеся с протоколом X.25 МСЭ-Т.

Таким образом, результирующая скорость передачи по указанному интерфейсу составляет $30 \times 64 \text{ кбит/с} + 64 \text{ кбит/с} + 64 \text{ кбит/с} = 2048 \text{ кбит/с}$.

MS — подвижная станция, состоит из оборудования, которое служит для организации доступа абонентов сетей GSM к существующим фиксированным сетям электросвязи. В рамках стандарта GSM приняты пять классов подвижных станций от модели 1-го класса с выходной мощностью 20 Вт, устанавливаемой на транспортном средстве, до портативной модели 5-го класса, максимальной мощностью 0,8 Вт. При передаче сообщений предусматривается адаптивная регулировка мощности передатчика, обеспечивающая требуемое качество связи.

Подвижный абонент и станция независимы друг от друга. Как уже отмечалось, каждый абонент имеет свой международный идентификационный номер (IMSI), записанный на его интеллектуальную карточку. Такой подход позволяет устанавливать радиотелефоны, например, в такси и автомобилях, сдаваемых на прокат. Каждой подвижной станции также присваивается свой международный идентификационный номер (IMEI). *Этот номер используется для предотвращения доступа к сетям GSM похищенной станции или станции без полномочий.*

BSS (*Base Station System*) складывается из двух частей: из базовой приемопередающей станции BTS (*Base Transceiver Station*) и контроллера базовой станции BSC (*Base Station Controller*). Интерфейс Abis (*Interface between BSC and BTS*), связывающий эти части, позволяет оперировать компонентами, созданными различными производителями. Радиопокрытие BSS делится на территории — соты — каждая покрывается одной BTS.

BTS управляет протоколами радиоканалов с MS. На крупной густонаселенной территории может располагаться много BTS, и поэтому к ним предъявляются очень строгие требования (четкость границ, надежность, переносимость и малая стоимость). BSC управляет радиоресурсами одного или нескольких BTS, контролирует предоставление радиоканала, регулировку частоты, управление перемещениями из ячейки в ячейку вызовами (хендоверами) и является связующим звеном между подвижной станцией и MSC.

Каждый подвижный абонент на время пользования сетью получает стандартный модуль подлинности абонента SIM (*Standard Identification Module*), который содержит: международный идентификационный номер IMSI (*International Mobile Station Identifier*), свой индивидуальный ключ аутентификации Ki (*Individual Subscriber Authentication Key*), алгоритм аутентификации (A3 — *Authentication Algorithm*). С помощью SIM пользователь получает доступ к ресурсам сети.

Оборудование базовой станции состоит из контроллера базовой станции (BSC), который может управлять несколькими приемопередающими блоками. Он управляет распределением радиоканалов, контролирует соединения, регулирует их очередность, обеспечивает режим работы, кодирование и декодирование сообщений, адаптацию скорости передачи и т. д.

13.3. Формирование сигналов в стандарте GSM

Все абоненты системы GSM работают в одном и том же диапазоне частот: 935965 МГц для передачи информации базовым станциям (линия «вниз») и 890915 МГц для передачи информации подвижным станциям (линия «вверх»), но при этом каждый имеет ограниченное время доступа. Каждая из этих полос разделена на 124 частотных канала, которые с разнесом 200 кГц распределены в соответствии с размещением сот. Частоты, выделенные для передачи от подвижной станции на базовую и в обратном направлении, группируются парами, организуя дуплексный режим работы с разнесом 45 МГц. Эти пары частот сохраняются и при перескоках частоты. Каждая ячейка (сота) характеризуется присвоением определенного количества пар частот от 1 до 15 (не более).

Таким образом, каждому абоненту выделен временной промежуток (кадр), в котором ему разрешено «вещание». Когда один абонент завершает свое вещание, разрешение передается другому, затем третьему и т.д. После того, как обслужены все абоненты, процесс начинается сначала. Поэтому активность абонента носит пульсирующий характер. Чем больше абонентов, тем реже каждому из них предоставляется возможность передать свои данные и тем меньше данных он может передать.

В результате анализа различных вариантов построения цифровых сетей связи в GSM принята комбинация методов множественного доступа с временным разделением каналов (TDMA) и частотным разделением каналов (FDMA). Структура временных кадров GSM показана на рис. 13.3.

Длина периода последовательности в этой структуре, которая называется гиперкадром, равна $T_F = 3 \text{ и } 28 \text{ мин. } 53 \text{ с. } 760 \text{ мс} = 12533,76 \text{ с.}$

Гиперкадр делится на 2048 суперкадров, каждый из которых имеет длительность $T_c = 12533,76/2048 = 6,12 \text{ с.}$

Суперкадр состоит из мультикадров. Для организации различных каналов связи и управления в стандарте GSM используют 2 типа мультикадров:

- 26-позиционные TDMA-мультикадры управления длительностью 235,385 мс, содержащие 51 TDMA-кадр;

- 51-позиционные TDMA-мультикадры трафика длительностью 120 мс, содержащие 26 TDMA-кадров.

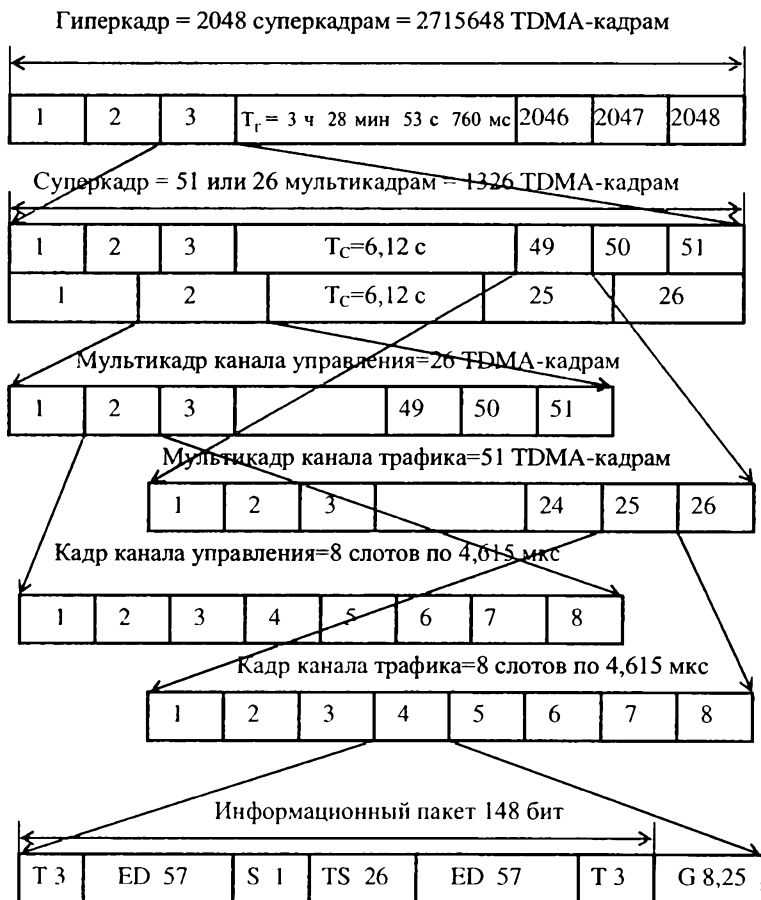


Рис. 13.3. Структура эфирного интерфейса системы GSM:
 ED (Encrypted Data) — закодированная информация, TS (Training Sequence) — обучающая последовательность, T (Tail bits) — защитный бланк — хвостовые биты, S (Stealing flag) — скрытый флажок — признак речь/управление, G (Guard period) — защитный интервал

Таким образом, в одном суперкадре содержится $51 \cdot 26 = 1326$ TDMA-кадров, а в гиперкадре — $1326 \cdot 2048 = 2715648$ TDMA-кадров. Необходимость такого большого периода гиперкадра объясняется требованиями применяемого процесса криптографической защиты, в котором номер кадра используется как входной параметр.



Пример 13.1. Оценить длительность TDMA-кадров управления и трафика.

Решение. Так как в составе TDMA-мультикадра управления длительностью 235,385 мс содержится 51 TDMA-кадр, то длительность одного TDMA-кадра равна $235,385/51 = 4,615$ мс.

Так как в составе TDMA-мультикадра трафика длительностью 120 мс содержится 26 TDMA-кадров, то длительность одного TDMA-кадра трафика равна $120/26 = 4,615$ мс.

Таким образом, TDMA-мультикадры управления и трафика состояются из одинаковых по длительности TDMA-кадров. Их длительность можно оценить также исходя из того, что в одном суперкадре длительностью $T_c = 6,12$ с содержится 1326 TDMA-кадров, откуда длительность одного TDMA-кадра составляет $6,12/1326 = 4,615$ мс.

Каждый TDMA-кадр делится на восемь временных позиций с периодом $T_0 = 4,615/8 = 60/13,8 = 576,9$ мкс (15/26 мс). Физический смысл временных позиций, которые иначе называются слотами, соответствует временным промежуткам, в течение которых осуществляется модуляция несущей частоты цифровым информационным потоком, соответствующим речевому сообщению или данным.

Цифровой информационный поток представляет собой последовательность пакетов, размещаемых в слотах. Пакеты формируются немного короче, чем интервалы, их длительность составляет 0,546 мс, что необходимо для приема сообщения при наличии временной дисперсии в канале распространения. Временной интервал TDMA-кадра содержит 156,25 бит.



Пример 13.2. Оценить длительность передачи одного бита информации и скорость передачи по информационному каналу.

Решение. Длительность передачи одного бита информации $\tau_{бит} = 576,9/156,25 = 3,69 \text{ мкс}$, а скорость передачи по информационному каналу $156,25/(15/26) = 270,833 \text{ Кбит/с}$.

Каждый временной интервал, соответствующий длительности бита, обозначается BN с номером от 0 до 155; последнему интервалу длительностью 1/4 бита присвоен номер 156.

Для передачи информации по каналам связи и управления, подстройки несущих частот, обеспечения временной синхронизации и доступа к каналу связи в структуре TDMA-кадра используются пять видов слотов:

NB (*Normal Burst*) — нормальный временной интервал;

FB (*Frequency Correction Burst*) — временной интервал подстройки частоты;

SB (*Synchronization Burst*) — интервал временной синхронизации;

DB (*Dummy Burst*) — установочный интервал;

AB (*Access Burst*) — интервал доступа.

NB используется для передачи информации по каналам связи и управления, за исключением канала доступа RACH (рис.13.4). Он состоит из 114 бит зашифрованного сообщения и включает защитный интервал (G) в 8,25 бит длительностью 30,44 мкс (см. рис. 13.3). Информационный блок 114 бит разбит на два самостоятельных блока по 57 бит, разделенных между собой обучающей последовательностью в 26 бит, которая используется для установки эквалайзера в приемнике в соответствии с характеристиками канала связи в данный момент времени.

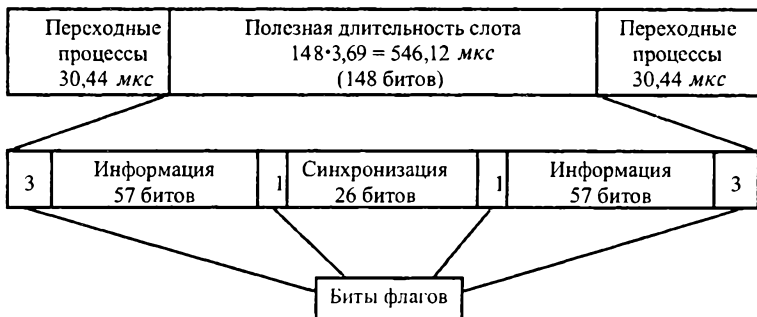


Рис. 13.4. Структура нормального слота GSM

В состав NB включены два контрольных бита S (см. рис. 13.4), которые служат признаком того, содержит ли передаваемая группа речевую информацию или информацию сигнализации. В последнем случае информационный канал (*Traffic Channel*) используется для обеспечения сигнализации.

Между двумя группами зашифрованных битов в составе NB находится обучающая последовательность из 26 битов, известная в приемнике.

С помощью этой последовательности обеспечивается:

1. Оценка частоты появления ошибок в двоичных разрядах по результатам сравнения принятой и эталонной последовательностей. В процессе сравнения вычисляется параметр RXQUAL, принятый для оценки качества связи. Конечно, речь идет только об оценке связи, а не о точных измерениях, так как проверяется только часть передаваемой информации. Параметр RXQUAL используется при вхождении в связь, при выполнении процедуры эстафетной передачи (*handover*) и при оценке зоны покрытия радиосвязью.

2. Оценка импульсной характеристики радиоканала на интервале передачи NB для последующей коррекции тракта приема сигнала за счет использования адаптивного эквалайзера в тракте приема.

3. Определение задержек распространения сигнала между базовой и подвижной станциями для оценки дальности связи. Эта информация необходима для того, чтобы пакеты данных от разных подвижных станций не накладывались при приеме на базовой станции. Поэтому удаленные на большее расстояние подвижные станции должны передавать свои пакеты раньше станций, находящихся в непосредственной близости от базовой станции.

FB предназначен для синхронизации по частоте подвижной станции. Все 142 бита в этом временном интервале — нулевые, что соответствует немодулированной несущей со сдвигом $1625/24$ кГц выше номинального значения частоты несущей. Это необходимо для проверки работы своего передатчика и приемника при небольшом частотном разносе каналов (200 кГц), что составляет около 0,022% от номинального значения полосы частот 900 МГц. FB содержит защитный интервал 8,25 бит так же, как и нормальный временной интервал. Повторяющиеся временные интервалы подстройки частоты (FB) образуют канал установки частоты (FCCN).

SB используется для синхронизации по времени базовой и подвижной станций. Он состоит из синхропоследовательности длительностью 64 бита, несет информацию о номере TDMA-кадра и идентификационный код базовой станции. Этот интервал передается вместе с интервалом установки частоты. Повторяющиеся интервалы синхронизации образуют так называемый канал синхронизации (SCH).

DB обеспечивает установление и тестирование канала связи. По своей структуре DB совпадает с NB и содержит установочную последовательность длиной 26 бит. В DB отсутствуют контрольные биты и не передается никакой информации. DB лишь информирует о том, что передатчик функционирует.

AB обеспечивает разрешение доступа подвижной станции к новой базовой станции. AB передается подвижной станцией при запросе канала сигнализации. Это первый передаваемый подвижной станцией пакет, следовательно, время прохождения сигнала еще не измерено. Поэтому пакет имеет специфическую структуру. Сначала передается концевая комбинация 8 бит, затем — последовательность синхронизации для базовой станции (41 бит), что позволяет базовой станции обеспечить правильный прием последующих 36 зашифрованных бит. Интервал содержит большой защитный интервал (68,25 бит, длительностью 251,84 мкс), что обеспечивает (независимо от времени прохождения сигнала) достаточное временное разнесение от пакетов других подвижных станций. Этот защитный интервал соответствует двойному значению наибольшей возможной задержки сигнала в рамках одной соты и тем самым устанавливает максимально допустимые размеры соты. Особенность стандарта GSM — возможность обеспечения связью подвижных абонентов в сотах с радиусом около 35 км. Время распространения радиосигнала в прямом и обратном направлениях составляет при этом $70/300000 = 233,3$ мкс.

В структуре GSM строго определены временные характеристики огибающей сигнала, излучаемого пакетами на канальном временном интервале TDMA-кадра, и спектральная характеристика сигнала. Различные формы огибающих излучаемых сигналов соответствуют разным длительностям интервала AB (88 бит) по отношению к другим указанным интервалам полного TDMA-кадра (148 бит).

Структуры слотов FB, SB, DB и AB показаны на рис. 13.5.

Одна из особенностей формирования сигналов в стандарте GSM — использование медленных скачков по частоте в процессе сеанса связи. Главное назначение таких скачков (SFH — *Slow Frequency Hopping*) — обеспечение частотного разнесения в радиоканалах, функционирующих в условиях многолучевого распространения радиоволн. SFH используется во всех подвижных сетях, что повышает эффективность кодирования и перемежения при медленном движении абонентских станций. Принцип формирования медленных скачков по частоте состоит в том, что сообщение, передаваемое в выделенном абоненту временном интервале TDMA-кадра (577 мкс), в каждом последующем кадре передается (принимается) на новой фиксированной частоте. В соответствии со структурой кадров время для перестройки частоты составляет около 1 мс.

В процессе скачков по частоте постоянно сохраняется дуплексный разнос 45 МГц между каналами приема и передачи. Всем активным абонентам, находящимся в одной соте, ставятся в соответствие ортогональные формирующие последовательности, что исключает взаимные помехи при приеме сообщений абонентами в соте. Параметры последовательности переключения частот (частотно-временная матрица и начальная частота) назначаются каждой подвижной станции в процессе установления канала. Ортогональность последовательностей переключения частот в соте обеспечивается начальным

Пакет коррекции частоты	T3	Фиксированный набор бит			T3	G 8,25
Пакет синхронизации	T3	ED 39	ETS 64	ED 39	T3	G 8,25
Холостой пакет	T3	Битовая смесь (шум)			T3	G 8,25
Пакет доступа	ET 8	TS 41	ED 36	T3	G 68,25	

Рис. 13.5. Варианты слотов канала управления системы GSM:

ETS (*Extended Training Sequence*) — расширенная обучающая последовательность, ET (*Extended Tail*) — расширенный защитный бланк

частотным сдвигом одной и той же (по алгоритму формирования) последовательности. В смежных сотах используются различные формирующие последовательности.

13.4. Организация физических и логических каналов в GSM

13.4.1. Физические каналы

Физические каналы в GSM — это комбинация временного и частотного разделения сигналов и определяется как последовательность радиочастотных каналов (с возможностью переключения по частотам) и временных окон TDMA-кадра. Каждая частотная несущая содержит 8 физических каналов, размещенных в 8 временных интервалах (слотах) в пределах TDMA-кадра. Каждый физический канал использует один и тот же слот в каждом временном TDMA-кадре.

Для формирования физического канала сообщения и данные, представленные в цифровой форме, группируются и объединяются в логические каналы двух типов:

- каналы для передачи кодированной речи и данных с полной скоростью (TCH/F — *Full Rate Traffic Channel*);
- каналы управления для передачи сигналов управления и синхронизации с половинной скоростью (TCH/H — *Half Rate Traffic Channel*).

Один физический канал может представлять собой канал передачи сообщений с полной скоростью или два канала с половинной скоростью передачи. В первом случае канал связи занимает один слот, во втором — два канала связи занимают тот же слот, но с перемежением в соседних кадрах, т. е. имеет место каждый канал, расположенный через кадр.

13.4.2. Структура логических каналов управления

Различают 4 типа каналов управления:

- BCCH (*Broadcast Control CHannels*) — каналы передачи сигналов управления;
- CCCH (*Common Control CHannels*) — общие каналы управления;
- DCCH (*Stand-alone Dedicated Control CHannels*) — индивидуальные каналы управления;

- **ACCH** (*Associated Control Channels*) — совмещенные каналы управления.

Каналы передачи сигналов управления (**BCCH**) используются только в направлении с базовой станции на все подвижные станции. Они несут информацию, необходимую подвижным станциям для работы в системе. Различают 3 вида каналов передачи сигналов управления:

- **FCCH** (*Frequency Correction Channel*) — канал подстройки частоты, используемый для синхронизации несущей в подвижной станции. По этому каналу передается немодулированная несущая с фиксированным частотным сдвигом относительно номинального значения частоты канала связи;
- **SCN** (*Synchronization Channel*) — канал синхронизации, по которому передается информация на подвижную станцию о кадровой (временной) синхронизации;
- **BCCH** (*Broadcast Control Channel*) — канал управления передачей, обеспечивающей передачу основных команд по управлению передачей (номер общих каналов управления тех из них, которые объединены с другими каналами, в том числе и с физическими и т. д.).

Используются три типа общих каналов управления **CCCH**:

- **PCN** (*Paging Channel*) — канал вызова от базовой станции к подвижной для ее вызова;
- **RACH** (*Random Access Channel*) — канал параллельного доступа от подвижной станции к базовой для запроса о назначении канала управления;
- **AGCH** (*Access Grant Channel*) — канал разрешенного доступа только для передачи с базовой станции на подвижную (для выделения специального канала управления, обеспечивающего прямой доступ к каналу связи).

Различают два вида выделенных индивидуальных каналов управления:

- **SDCCH/4** (*Stand-alone Dedicated Control Channel*) — индивидуальный канал управления, состоит из 4 подканалов;
- **SDCCH/8** (*Stand-alone Dedicated Control Channel*) — индивидуальный канал управления, состоит из 8 подканалов.

Эти каналы предназначены для установки требуемого пользователем вида обслуживания. По ним обеспечивается

запрос подвижной станции о требуемом виде обслуживания, контроль правильного ответа базовой станции и выделение свободного канала связи, если это возможно.

Различают также 2 вида каналов АССН, передающих команду управления с базовой станции и информацию о статусе подвижной станции:

- FACCH (*Fast Associated Control CHannel*) — быстрый совмещенный канал управления, служит для передачи команд при переходе подвижной станции из соты в соту;
- SACCH (*Slow Associated Control CHannel*) — медленный совмещенный канал управления, передает команды для установки выходного уровня мощности передатчика подвижной станции и данные, касающиеся уровня установленной выходной мощности мобильной станции.

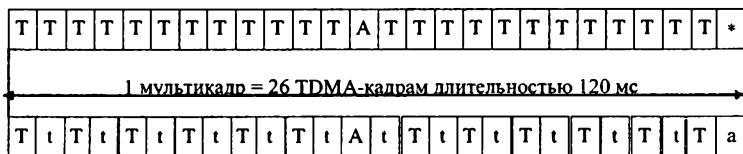
Совмещенные каналы управления всегда объединяются с каналами связи или с индивидуальными каналами управления. При этом различают 6 видов объединенных каналов управления:

- FACCH/F, объединенный с TCH/F;
- FACCH/H, объединенный с TCH/H;
- SACCH/TF, объединенный с TCH/F;
- SACCH/TH, объединенный с TCH/H;
- SACCH/C4, объединенный с SDCCH/4;
- SACCH/C8, объединенный с SDCCH/8.

Для передачи каналов связи TCH и совмещенных каналов управления FACCH и SACCH используется 26-позиционный мультикадр. Объединение каналов связи с полной и половинной скоростью передачи с медленным совмещенным каналом управления SACCH показано на рис. 13.6. В полноскоростном канале связи в каждом 3-м TDMA-кадре мультикадра передается пакет информации канала SACCH; каждый 26-й TDMA-кадр мультикадра свободен. В полускоростном канале связи пакет информации канала SACCH передается в каждом 13-м и 26-м TDMA-кадрах мультикадра.

SACCH занимает 1 TDMA-кадр в каждом мультикадре длительностью 120 мс. Каждая активная оконечная станция имеет слот в кадре SACCH. Сообщение SACCH содержит 184 бита данных. Кодеки контроля ошибок обрабатывают эту информацию и выдают 456 канальных битов, которые перемежаются и распределяются по 4 слотам. При одном слоте в мульти-

Полноскоростной канал связи (ТСН)



Полускоростной канал связи (ТСН)

T, t – TDMA-кадры для передачи канала ТСН
 A, a – TDMA-кадры для передачи канала SACCH
 * – пустой TDMA-кадр

Рис. 13.6. Объединение каналов связи с полной и половинной скоростью с медленным совмещенным каналом управления SACCH

кадре и длительности мультикадра 120 *мс* общее время передачи сообщения SACCH составляет 480 *мс*. Таким образом, скорость передачи данных равна 184 *бита* за 480 *мс* или 383 *бит/с*. Когда требуется более быстрое управление сетью, GSM создает быстродействующий канал управления FACCH, прерывая информацию пользователя на длительность 4 кадров. Время доставки сообщения FACCH составляет 4 кадра, или 18,5 *мс*, вместо 480 *мс* для SACCH.

Для одного физического канала в каждом TDMA-кадре используется 114 *бит*. Так как в мультикадре для передачи канала связи ТСН используется 24 TDMA-кадра из 26 и длительность мультикадра составляет 120 *мс*, общая скорость передачи информационных сообщений по ТСН каналу составляет 22,8 *кбит/с*. Канал SACCH занимает в полноскоростном канале связи только один TDMA-кадр, т. е. 114 *битов*, когда скорость передачи по SACCH каналу составит 950 *бит/с*. Полная скорость передачи в объединенном ТСН/SACCH канале с учетом пустого (свободного) 26-го TDMA-кадра составит $22,8 + 0,950 + 0,950 = 24,7$ *кбит/с*.

Как видно из рис. 13.6, за время 26-кадрового мультикадра (в одном физическом канале) может передаваться два полускоростных ТСН канала, каждый по 12 TDMA-кадров (*T* и *t*). Для каждого полускоростного канала управления ТСН скорость передачи составляет 11,4 *кбит/с*, а полная скорость передачи в объединенном полускоростном канале ТСН/SACCH остается прежней — 24,7 *кбит/с*.

Быстрый совмещенный канал управления FACH передается половиной информационных битов слота TDMA-кадра в канале TCH, которым он совмещается в 8 последовательных T и t кадрах.

Для передачи каналов управления (за исключением FACH и SACH) используется 51-позиционный мультикадр. Полная скорость передачи для канала BCCH, а также для канала AGCH/PCN составляет 1,94 кбит/с ($4 \cdot 114$ за 235 мс).

Существуют и другие структуры, которые могут использоваться с этим мультикадром. Отображение логических каналов на физические каналы осуществляется через процессы кодирования и шифрования передаваемых сообщений.

Для защиты логических каналов от ошибок, которые могут иметь место в процессе передачи, используют три вида кодирования: блочное для быстрого обнаружения ошибок при приеме, сверточное для исправления одиночных ошибок и перемежение для преобразования пакетов ошибок в одиночные.

Для защиты каналов от подслушивания в каналах связи и управления применяется шифрование.

13.5. Кодирование и перемежение в каналах стандарта GSM

Устройство перемежения собирает речевую информацию, генерируемую в интервале 40 мс (912 битов) и переставляет порядок последовательности битов в потоке. Цель перемежения заключается в уменьшении пакетов ошибок, возникающих при передаче, при декодировании битового потока. Когда битовый поток подвергается перемежению до передачи, а затем биты вновь ставятся на свои места, то ошибки, которые группируются в последовательности из 912 канальных битов, распределяются случайным образом в битовом потоке, поступающем на речевой декодер.

После каждых 40 мс оконечное устройство распределяет 912 кодированных речевых бит по 8 кадрам, располагая по 114 бит в каждый слот.

Для защиты от ошибок в радиоканалах подвижной связи GSM PLMN используются сверточное и блочное кодирование с перемежением. Перемежение обеспечивает преобразование пакетов ошибок в одиночные. Сверточное кодирование яв-

ляется мощным средством борьбы с одиночными ошибками. Блочное кодирование главным образом используется для обнаружения нескорректированных ошибок.

Блочный код (n, k, t) преобразует k информационных символов в n символов путем добавления символов четности $(n - k)$, а также может корректировать t ошибок символов.

Сверточные коды (СК) относятся к классу непрерывных помехоустойчивых кодов. Одной из основных характеристик СК является величина K , которая называется длиной кодового ограничения и показывает, на какое максимальное число выходных символов влияет данный информационный символ. Так как сложность декодирования СК по наиболее выгодному, с точки зрения реализации, алгоритму Витерби возрастает экспоненциально с увеличением длины кодового ограничения, то типовые значения K малы и лежат в интервале от 3 до 10. Другой недостаток СК заключается в том, что они не могут обнаруживать ошибки. Поэтому в стандарте GSM для внешнего обнаружения ошибок используется блочный код на основе сверточного кода $(2, 1, 5)$ со скоростью $r = 1/2$. Наибольший выигрыш СК обеспечивает только при одиночных (случайных) ошибках в канале.

Кодирование осуществляется следующим образом: биты класса 1 разделяются дополнительно на 50 бит класса 1а и 132 бита класса 1б (рис. 13.7). Биты класса 1а дополняются тремя битами проверки на четность (рис. 13.7). Блочный код представляет собой укороченный систематический циклический код $(53, 50)$ с формирующим полиномом вида

$$f(x) = x^3 + x + 1.$$

В соответствии с принятым правилом формирования систематического кода, ключ S_w закрыт на время первых пятидесяти тактовых импульсов, а информационные биты, поступающие на вход кодирующего устройства, одновременно поступают на блок переупорядочения и формирования бит проверки на четность. После пятидесяти тактовых импульсов переключатель S_w срабатывает, и биты проверки на четность поступают из кодирующего устройства. Биты с четными индексами собираются в первой части информационного слова, за которыми следуют три бита проверки на четность. Затем биты с нечетными индексами запоминаются в буферной

памяти и переставляются. Далее следуют четыре нулевых бита, которые необходимы для работы кодера, формирующего код, исправляющий случайные ошибки в канале. После чего 189 бит класса 1 кодируются сверточным кодом (2, 1, 5) со скоростью $s = 1/2$ (рис. 13.8). Структурная схема кодера приведена на рис. 13.9.

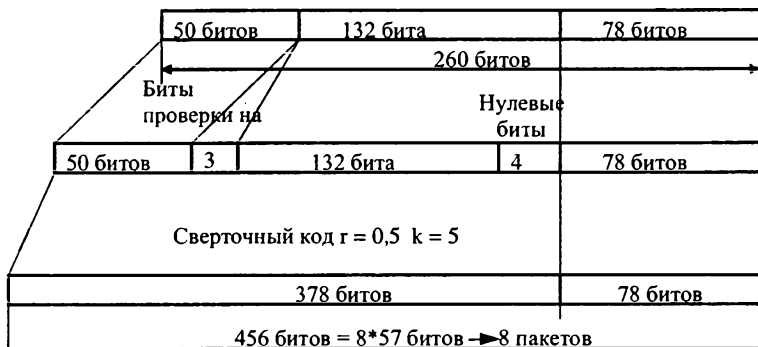


Рис. 13.7. Кодирование и перемежение в стандарте GSM



Рис. 13.8. Кодировальное устройство в стандарте GSM

После сверточного кодирования общая длина кадра составляет $2 \cdot 189 + 78 = 456$ бит. После этого кадр из 456 бит делится на восемь 57 битовых подблоков, которые подвергаются диагональному и внутрикадровому перемежению. Более точно подблоки формируются в пакеты по 114 бит, которые являются результатом блочно-диагонального перемежения, биты подблоков попарно перемежаются, образуя процесс внутрикадрового битового перемежения. В результирующий пакет

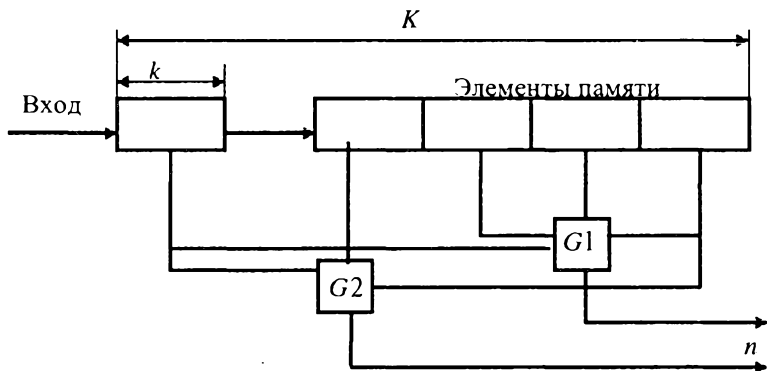


Рис. 13.9. Структурная схема кодера стандарта GSM

включаются два опережающих флага, которые используются для классификации различных пакетов передачи.

В каналах с замираниями, что имеет место в GSM PLMN, необходимо использовать СК совместно с перемежением.

В GSM PLMN основные свойства речевых каналов и каналов управления значительно отличаются друг от друга. Для речевых каналов необходима связь в реальном масштабе времени с короткими задержками при сравнительно низких требованиях к вероятности ошибки в канале. Для каналов управления требуется абсолютная целостность данных и обнаружения ошибок, но допускается более длительное время передачи и задержки.

В соответствии с общей структурой кадров в стандарте GSM передача информационных сообщений и сигналов управления осуществляется в нормальном временном интервале (NB) TDMA-кадра. Структура NB (два пакета по 57 информационных бит каждый) требует, чтобы количество кодированных бит m , соответствующих n некодированным битам в общей схеме кодирования и перемежения, равнялась бы целому числу, кратному 19. Затем эти биты зашифровываются и объединяются в I групп. Количество бит в этих группах также должно равняться 19, I групп переходят в I временных интервалов. Номер I называется степенью перемежения.

В различных логических каналах используются различные сверточные коды, поскольку скорости передачи и требования по защите от ошибок также различны. Для упрощения меха-

низмов кодирования и декодирования для формирования кодов используются только несколько полиномов. Это позволяет использовать сверточный код с одной скоростью $s = 1/2$. Однако, чтобы выполнить требования формирования полноскоростного канала связи, а также привести в соответствие структуру размещения бит со структурой кадров, необходима скорость $s = 244/456 = 0,535$. Для выравнивания скорости в речевом канале до $r = 1/2$ применяют *прореживание*, то есть периодический пропуск некоторых кодированных символов. Такая операция называется *перфорированием*, а формируемые таким образом коды называются *перфорированными*. При приеме декодер, зная алгоритм прореживания, интерполирует принимаемые данные.

При передаче логического быстрого совмещенного канала управления FACH перфорирование не используется.

13.6. Повышение эффективности сверточного кодирования в каналах стандарта GSM

Для повышения эффективности применения сверточного кодирования в полноскоростных каналах передачи данных необходим длительный период перемежения. В этих каналах внутрикадровое перемежение реализуется для степени перемежения $I = 19$, что приводит к задержке передачи данных на $19 \times 116 = 2204$ бит. Если биты I -го пакета (временного интервала) до перемежения обозначить как $C(Km)$, $m = 1 \dots 116$, то схема перемежения, то есть позиции бит после перемежения, определяются следующей формулой:

$$\begin{aligned} I(K + j, j + 19t) &= C(K, m) \text{ для всех} \\ Kj &= m \bmod 19, t = m \bmod 6. \end{aligned} \quad (13.2)$$

При перемежении кода (n, k, t) длинные кодированные слова записываются кодирующим устройством в память перемежителя строка за строкой, а затем передаются в модулятор столбец за столбцом. В приемнике после демодулятора деперемежитель обратной операцией восстанавливает первоначальный порядок символов, после чего осуществляется декодирование.

13.7. Модуляция GMSK в современных системах радиосвязи

Модуляция GMSK используется в современных системах цифровой радиосвязи GSM (DCS, PCS), GPRS и других и обеспечивает высокое качество передачи в относительно узкой полосе, занимаемой сигналом. GMSK реализуется в электронных компонентах, выпускаемых фирмами *Analog Devices*, *Texas Instruments*, *Infineon Technologies* и другими.

GMSK (*Gaussian Minimum Shift Keying*) — это гауссовская двухпозиционная частотная манипуляция с минимальным сдвигом, обладающая двумя особенностями, одна из которых — «минимальный сдвиг», другая — гауссовская фильтрация. Обе особенности направлены на сужение полосы частот, занимаемой GMSK-сигналом. Использование GMSK в системе сотовой радиосвязи GSM регламентируется стандартом ETSI (Европейский институт стандартов связи).

В общем случае, при частотной модуляции (ЧМ, FM), в том числе при манипуляции (ЧМн, FSK), спектр сигнала более широкий, чем при амплитудной модуляции (манипуляции). Расширение спектра, свойственное угловой модуляции, частным случаем которой является ЧМ (ЧМн), зависит от индекса модуляции — одного из ее основных параметров. Индекс модуляции — это величина, характеризующая изменение фазы, обусловленное модуляцией. Для ЧМ (ЧМн) индекс равен $\beta = \Delta f / F$ (в радианах), где Δf — девиация (максимальный сдвиг) частоты, а F — частота модуляции (манипуляции). Характер изменения фазы зависит от формы модулирующей функции частоты. Для обычной ЧМн функция прямоугольна, а для ЧМн с гауссовской фильтрацией, сглаживающей фронты посылок, близка к синусоидальной (при последовательности чередующихся посылок, соответствующих логическим 0 и 1). При синусоидальной модулирующей функции индекс модуляции является амплитудой изменения фазы. С учетом скорости манипуляции $\nu = 2F = 1/T$ (ν — в бит/с, а F — в Гц), где T — длительность посылок, индекс равен:

$$\beta = 2\Delta f / \nu. \quad (13.3)$$

ЧМ (ЧМн) подразделяют на узкополосную и широкополосную, зависящие от величины индекса. При узкополосной

ЧМн, характеризуемой малым индексом ($\beta = 0,5$), спектр сигнала сосредоточен, в основном, в полосе, определяемой удвоенным спектром манипулирующих посылок (практически без расширения). Отметим, что при узкополосной ЧМн частота манипуляции больше девиации частоты: $F \geq 2 \Delta f$. Широкополосная ЧМ, применяемая в основном в качестве аналоговой (например, в радиовещательном УКВ-диапазоне), характеризуется большим индексом и, соответственно, расширением спектра ЧМ-сигнала.

В основе GMSK лежит MMC (MSK) — узкополосная ЧМн «с минимальным сдвигом», характеризуемая $\beta \leq 0,5$. При MMC и, соответственно, при GMSK фаза частотно-манипулируемого колебания непрерывна, а ее «набег» в течение одной посылки, обусловленный манипуляцией частоты $\pm \Delta \omega$ ($\pm 2\pi \Delta f$), равен $\Delta \varphi = \pm \Delta \omega T$. При $\beta = 0,5$ он составляет $\pi/2$ и $-\pi/2$ для посылок 1 и 0 соответственно. Подчеркнем, что $\Delta \varphi$ — это не манипуляция фазы, а именно ее «набег», обусловленный манипуляцией частоты. При $\beta = 0,5$ скорость манипуляции $\nu = 4\Delta f$, которая для GMSK, используемой в GSM, составляет $\nu = 270,833 \text{ Кбит/с}$ при $\Delta f = 67,70833 \text{ кГц}$.

Итак, GMSK — это узкополосная ЧМн с «граничным» индексом манипуляции, который не очень мал, но спектр при нем практически еще не расширен. Можно сказать, что индекс $\beta = 0,5$ является в этом смысле оптимальным. Однако, если манипуляцию осуществлять прямоугольными посылками, в спектре которых содержатся высшие гармоники, спектр ЧМн-сигнала будет все-таки расширен, но уже за счет этих гармоник. Поэтому при формировании сигналов с GMSK используется гауссовская низкочастотная фильтрация модулирующих посылок. Гауссовской она называется потому, что в качестве импульсной характеристики фильтра используют характеристику нормального распределения Гаусса. Используют ее симметричный отрезок, взятый на конечном интервале, равном длительности посылки T . Связь импульсной характеристики с T определяют параметром B — полосой гауссовского ФНЧ, равной частоте среза его АЧХ на уровне минус 3 дБ. Параметр B определяет произведение BT , которое для GSM равно $BT = 0,3$. Отметим, что BT — это не база сигнала, которая не может быть меньше единицы.

Гауссовскую фильтрацию осуществляют обычно в цифровом процессоре (DSP), в котором формируется сигнал модуляции.

13.7.1. Структурные схемы реализации GMSK

Для GMSK обычно используют одну из двух схем модуляции — с управляемым по частоте генератором (УГ) (рис. 13.10) и квадратурным модулятором (рис. 13.11).

В модуляторе с УГ (рис. 13.10) используется система ФАПЧ, обеспечивающая высокую стабильность несущей частоты ЧМ (ЧМн) сигнала. При этом собственно модулятором является УГ, а система ФАПЧ обеспечивает получение несущей, равной $\omega_0 = (M/N)\omega_{01}$. M и N — это коэффициенты деления частоты в делителях $/M$ и $/N$, а ω_{01} — стабильная частота опорного источника. Делители используются для точной настройки на требуемую частоту ω_0 . В этом смысле модулятор по схеме на рис. 13.10 является одновременно синтезатором частот типа *Integer-N*.

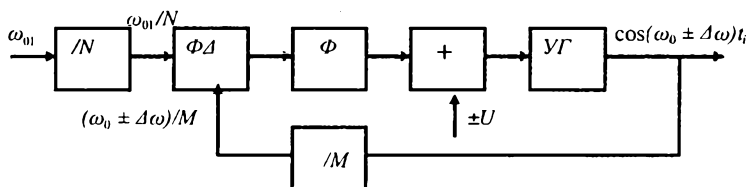


Рис. 13.10. Структурная схема GMSK-модулятора с управляемым генератором (УГ) в составе системы ФАПЧ ($\omega_{01}/N = \omega_0/M$)

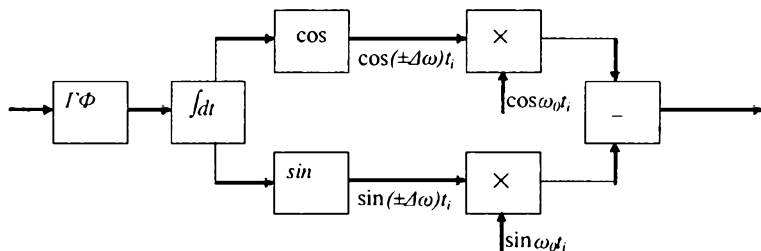


Рис. 13.11. Структурная схема GMSK-модулятора с квадратурным модулятором (с формирователем сигнала модуляции)

Модулятор характеризуется передаточной функцией

$$K_{\text{мод}}(p) = K_{\text{УГ}}/[1 + k_{\Phi}(p)/p\tau_0], \quad (13.4)$$

где $K_{\text{УГ}}$ — коэффициент преобразования УГ (с размерностью (рад/с)/В), $\tau_0 = M/K_{\text{УГ}}K_{\Phi\text{Д}}K_{\Phi}$ — постоянная времени, $K_{\Phi\text{Д}}$ — коэффициент передачи фазового детектора ФД, K_{Φ} и $k_{\Phi}(p)$ — постоянный и частотно-зависимый множители передаточной функции фильтра Φ . Согласно (13.4), модулятор обладает свойствами фильтра верхних частот. Это соответствует используемым кодовым последовательностям модулирующего сигнала (без постоянной составляющей).

Модулятор по схеме рис. 13.10, является, по существу, аналоговым, используемым для манипуляции. Обозначения, приведенные на рис. 13.10 (а также на рис. 13.12 и 13.13), например $\cos(\omega_0 \pm \Delta\omega)t_i$, являются упрощенными и, строго говоря, соответствуют установившимся значениям частоты посылок, т. е. без учета переходных процессов, связанных с изменением частоты $\pm\Delta\omega$.

Модулятор GMSK с УГ, по схеме на рис. 13.10, применяется обычно в беспроводной телефонии (в системе DECT), а в системе GSM применяется квадратурный модулятор, схема которого приведена на рис. 13.11. Собственно квадратурный модулятор содержит перемножители с опорными источниками $\cos\omega_0 t_i$ и $\sin\omega_0 t_i$ и вычитатель на выходе. На вход перемножителей поступают две составляющие квадратурного сигнала $\cos(\pm\Delta\omega)t_i$ и $\sin(\pm\Delta\omega)t_i$, формируемые обычно в DSP. При этом, если модулятор аналоговый, на выходе DSP используют ЦАП. Квадратурный модулятор может быть цифровым и иметь ЦАП на своем выходе. На рис. 13.11 показаны эквивалентные цепи DSP — гауссовский фильтр ГФ, интегратор $\int dt$ и элементы тригонометрических функций $\cos$ и $\sin$. На вход поступает нефильТРованная последовательность положительных и отрицательных значений $\pm\Delta\omega$, соответствующая 1 и 0 модулирующего кода. Указанная последовательность фильтруется в ГФ. Ее фронты сглажены, а частотный спектр, соответственно, сужен.

13.7.2. Структурные схемы квадратурных модуляторов GSM

На рис. 13.12 приведены схемы квадратурных модуляторов, совмещенных с повышающим преобразователем частоты: $\omega_0 = \omega_{01} + \omega_{02}$, где ω_0 — несущая ВЧ, ω_{01} — несущая на выходе модулятора, определяемая частотой первого гетеродина, ω_{02} — частота второго гетеродина. Показанные на входе $\cos \Omega t_i$ и $\pm \sin t \Omega t_i$, которые более наглядны, соответствуют $\cos(\pm \Delta \omega) t_i$ и $\sin(\pm \Delta \omega) t_i$ на рис. 13.11 (при $\Delta \omega = \Omega$).

Схема устройства на рис. 13.12 проще, чем на рис. 13.13, и содержит последовательно включенные модулятор и преобразователь частоты. Преобразователь построен с использованием системы ФАПЧ и дополнительно содержит смеситель (перемножитель) и фильтр $\Phi 2$. Фильтр пропускает составляющие с несущей ω_{01} , равной разности ω_0 и ω_{02} , и подавляет составляющие с $\omega_0 + \omega_{02}$. Фильтр $\Phi 2$ не является обязательным, если, например, смеситель выполнен по балансной схеме.

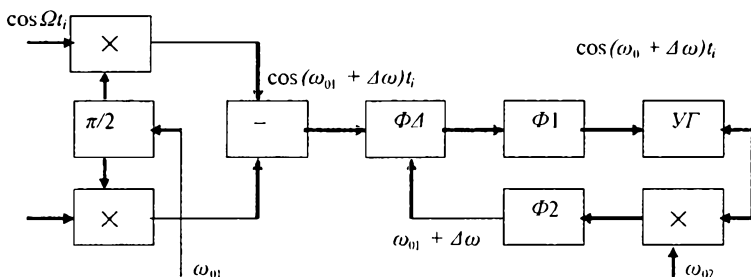


Рис. 13.12. Структурная схема преобразователя частоты с модулятором на входе

Передаточная функция преобразователя на рис. 13.12 является, в отличие от (13.1), функцией фильтра нижних частот:

$$K_{np}(p) = 1 / \{1 + p[\tau_0 / K_{\Phi 1}(p) - \tau_{зад2}]\}, \quad (13.5)$$

где $K_{\Phi 1}(p)$ — частотно-зависимый множитель функции фильтра $\Phi 1$, а $\tau_{зад2} = -\Delta \varphi_2 / \Delta \omega$ — временная задержка, вносимая

фильтром $\Phi 2$ (на рабочем участке его ФЧХ). Влияние $\Phi 2$ на передаточную функцию (13.5) видно из

$$K_{np}(p) = 1 / [1 + p(\tau_0 - \tau_{зад2}) + p^2 \tau_0 \tau_{\phi 1}] - \quad (13.6)$$

функции системы 2-го порядка с

$$K_{\phi 1}(p) = 1 / (1 + p \tau_{\phi 1}). \quad (13.7)$$

Согласно (13.6), действие $\tau_{зад2}$ эквивалентно уменьшению τ_0 при соответствующем увеличении $\tau_{\phi 1}$ (при $\tau_0 \tau_{\phi 1} = const$) и может быть компенсировано увеличением τ_0 при уменьшении $\tau_{\phi 1}$.

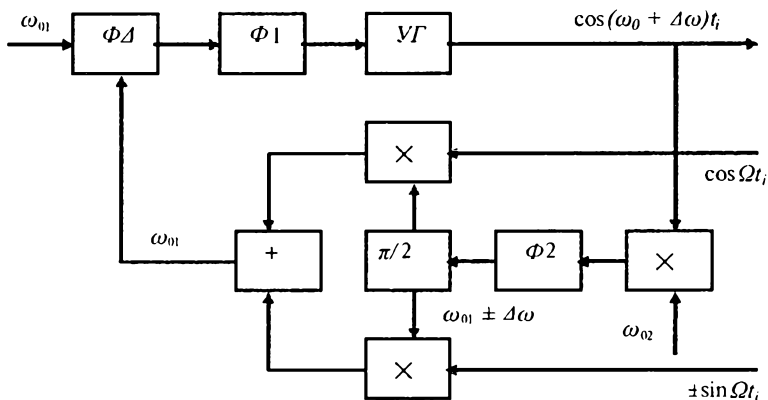


Рис. 13.13. Структурная схема преобразователя частоты со встроенным модулятором

На рис. 13.13 показана более сложная структурная схема совмещенного модулятора-преобразователя, известного как модулятор передатчика с виртуальной промежуточной частотой и используемого в новых разработках фирмы *Analog Devices*. В рассматриваемом устройстве модулятор встроен в систему ФАПЧ вместе с преобразователем. На выходе модулятора — сумматор, а выходной частотой модулятора является ω_{01} , которая равна частоте опорного колебания первого гетеродина, подаваемого на внешний вход ФД. На выходе преобразователя будут манипулируемые значения частоты

$\omega_0 \pm \Delta\omega$, где $\omega_0 = \omega_{01} + \omega_{02}$, ω_{02} — частота второго гетеродина, а манипуляция частоты $\pm\Delta\omega$ определяется входным сигналом модулятора ($\Delta\omega = \Omega$).

13.7.3. Обработка речи в стандарте GSM

Процессы обработки речи в стандарте GSM направлены на обеспечение высокого качества передаваемых сообщений, реализацию дополнительных сервисных возможностей и повышение потребительских качеств абонентских терминалов.

Обработка речи осуществляется в рамках принятой системы прерывистой передачи речи. Система прерывистой передачи речи (*Discontinuous Transmission*, DTX) обеспечивает включение передатчика только тогда, когда пользователь начинает разговор и отключает его в паузах и в конце разговора. DTX управляется детектором активности речи (*Voice Activity Detector*, VAD), который обеспечивает обнаружение и выделение интервалов передачи речи с шумом и шума без речи даже в тех случаях, когда уровень шума соизмерим с уровнем речи. В состав системы прерывистой передачи речи входит также устройство формирования комфортного шума, который включается и прослушивается в паузах речи, когда передатчик отключен. Экспериментально показано, что отключение фонового шума на выходе приемника в паузах при отключении передатчика раздражает абонента и снижает разборчивость речи, поэтому применение комфортного шума в паузах считается необходимым. Система прерывистой передачи речи в приемнике включает также интерполяцию фрагментов речи, потерянных из-за ошибок в канале.

Структурная схема процессов обработки речи в стандарте GSM показана на рис. 13.14, главным устройством в этой схеме является речевой кодек.

13.7.4. Выбор речевого кодека для стандарта GSM

Рабочей группой по разработке стандарта GSM были предъявлены следующие основные требования к речевому кодеку:

- высокое качество речи, не уступающее качеству передачи речи в лучших существующих аналоговых сотовых системах связи;

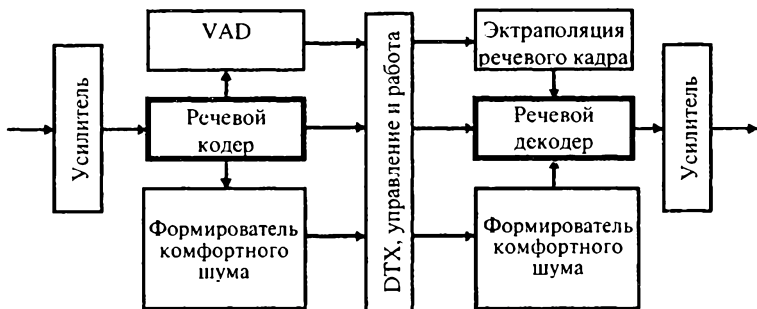


Рис. 13.14. Структурная схема обработки речи в стандарте GSM

- низкая скорость передачи речи, обеспечивающая возможность эффективного канального кодирования и результирующую скорость передачи в канале связи не выше 16 кбит/с;
- малая задержка сообщения в процессе преобразования речи;
- устойчивость к ошибкам в канале передачи;
- возможность работы в широком динамическом диапазоне входных воздействий как сигнала, так и шума;
- большой динамический диапазон выходных сигналов;
- незначительное снижение качества речи при каскадном соединении кодеков;
- прозрачность для сигналов данных;
- прямое сопряжение со смежными устройствами терминалов;
- простота реализации;
- малое потребление;
- низкая стоимость.

Для выбора речевого кодека GSM был организован конкурс проектов. Первоначально для рассмотрения было предложено 20 различных кодеков от 9 европейских стран. После международного формального тестирования это количество было сокращено до 6 из 6 стран. На следующем этапе два из четырех подполосных (SBC) кодеков (норвежский и итальянский) были сняты с рассмотрения, к окончательному этапу конкурса осталось два SBC-кодека и два кодека с предикативным кодированием:

- RPE-LPC — *Regular-Pulse Excitation/Linear Predictative Coding* (Германия, Philips) — кодек с регулярным импульсным возбуждением и линейным кодированием с предсказанием;
- MPE-LTP — *Multi-Pulse Excitation/Long-Term Prediction* (Франция, IBM) — кодек с многоимпульсным возбуждением и долговременным предсказанием.

На втором этапе происходит дальнейшее снижение динамического диапазона за счет долговременного предсказания, в процессе которого каждый сегмент выравнивается до уровня следующих друг за другом сегментов речи. В принципе, LTP-фильтр вычитает предыдущий период сигнала из текущего периода.

Этот фильтр характеризуется параметром задержки и коэффициентом усиления. Период вычисления этих параметров равен 5 мс.

Восемь коэффициентов LPC анализирующего фильтра и параметры фильтра LTP анализа кодируются и передаются со скоростью 3,6 кбит/с.

Для формирования последовательности возбуждения остаточный сигнал пропускают через фильтр нижних частот с частотой среза 3–4 кГц.

Окончательно периодическая последовательность фрагментов передается со скоростью 9,4 кбит/с. Общая скорость передачи составляет $3,6 + 9,4 = 13$ кбит/с.

В декодере речевой сигнал восстанавливается по откликам последовательности регулярного импульсного возбуждения двухступенчатым синтезирующим фильтром. При этом качество речи соответствует качеству речи, передаваемой по ISDN, и превосходит качество речи в аналоговых радиотелефонных системах. Теоретически время задержки речевого сигнала в кодеке равно длительности сегмента и составляет 20 мс. Реальное время задержки, с учетом операций канального кодирования и перемежения, а также физического выполнения рассматриваемых операций, составляет 70–80 мс.

13.7.5. Детектор активности речи

Детектор активности речи играет решающую роль в снижении потребления энергии от аккумуляторной батареи в портативных абонентских терминалах. Он также снижает интерфе-

рещионные помехи за счет переключения свободных каналов в пассивный режим. Реализация VAD зависит от типа применяемого речевого кодека. Главная задача при проектировании VAD — обеспечить надежное отличие между условиями активного и пассивного каналов. Если канал на мгновение свободен, его можно заблокировать, поскольку средняя активность речи говорящего ниже 50%, то это может привести к существенной экономии энергии аккумуляторной батареи. К устройствам VAD предъявляются следующие основные требования:

- минимизация вероятности ложной тревоги при воздействии только шума с высоким уровнем;
- высокая вероятность правильного обнаружения речи низкого уровня;
- высокое быстродействие распознавания речи, для исключения задержек включения;
- минимальное время задержки выключения.

В стандарте GSM принята схема VAD с обработкой в частотной области. Структурная схема VAD приведена на рис. 13.15. Ее работа основана на различии спектральных характеристик речи и шума. Считается, что фоновый шум является стационарным в течение относительно большого периода времени, его спектр также медленно изменяется во времени. VAD определяет спектральные отклонения входного воздействия от спектра фонового шума. Эта операция осуществляется инверсным фильтром, коэффициенты которого устанавливаются применительно к воздействию на входе только фонового шума. При наличии на входе речи и шума инверсный фильтр осуществляет подавление компонентов шума и, в целом, снижает его интенсивность. Энергия смеси сигнал + шум на выходе инверсного фильтра сравнивается с порогом, который устанавливается в период воздействия на входе только шума. Этот порог находится выше уровня энергии шумового сигнала. Превышение порогового уровня принимается за наличие на входе реализации (сигнал + шум). Коэффициенты инверсного фильтра и уровень порога изменяются во времени в зависимости от текущего значения уровня шума при воздействии на входе только шума. Поскольку эти параметры (коэффициенты и порог) используются детектором VAD для обнаружения речи, сам VAD не может на этой же основе при-

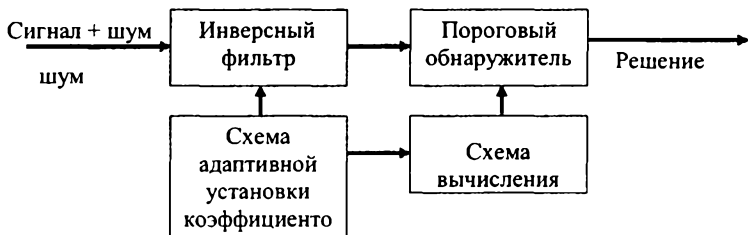


Рис 13.15. Структурная схема детектора активности речи

нимать решение, когда их изменять. Это решение принимается вторичным VAD на основе сравнения огибающих спектров в последовательные моменты времени. Если они аналогичны для относительно длительного периода времени, предполагается, что имеет место шум, и коэффициенты фильтра и шумовой порог можно изменять, т. е. адаптировать под текущий уровень и спектральные характеристики входного шума.

VAD с обработкой в спектральной области удачно сочетается с речевым RPE/LTP-LPC кодеком, так как в процессе LPC анализа уже определяется огибающая спектра входного воздействия, необходимая для работы вторичного VAD.

13.7.6. Формирование комфортного шума

Формирование комфортного шума осуществляется в паузах активной речи и управляется речевым декодером. Когда детектор активности речи (VAD) в передатчике обнаружит, что говорящий прекращает разговор, передатчик остается еще включенным в течение следующих пяти речевых кадров. Во время первых четырех из них характеристики фонового шума оцениваются путем усреднения коэффициента усиления и коэффициентов фильтра LPC-анализа. Эти усредненные значения передаются в следующем пятом кадре, в котором содержат информацию о комфортном шуме (SID-кадр).

В речевом декодере комфортный шум генерируется на основе LPC-анализа SID-кадра. Чтобы исключить раздражающее влияние модуляции шума, комфортный шум должен соответствовать по амплитуде и спектру реальному фоновому шуму в месте передачи. В условиях подвижной связи фоновый шум

может постоянно изменяться. Это значит, что характеристики шума должны передаваться с передающей стороны на приемную сторону не только в конце каждого речевого всплеска, но и в речевых паузах так, чтобы между комфортным и реальным шумом не было бы резких рассогласований в следующих речевых кадрах. По этой причине SID-кадры посылаются каждые 480 мс в течение речевых пауз.

Динамическое изменение характеристик комфортного шума обеспечивает натуральность воспроизведения речевого сообщения при использовании системы прерывистой передачи речи.

13.7.7. Экстраполяция потерянного речевого кадра

В условиях замираний сигналов в подвижной связи речевые фрагменты могут подвергаться значительным искажениям. При этом для исключения раздражающего эффекта при воспроизведении необходимо осуществлять экстраполяцию речевого кадра.

Было установлено, что потеря одного речевого кадра может быть значительно компенсирована путем повторения предыдущего фрагмента. При значительных по продолжительности перерывах в связи предыдущий фрагмент больше не повторяется, и сигнал на выходе речевого декодера постепенно заглушается, чтобы указать пользователю на разрушение канала.

То же самое происходит и с SID-кадром. Если SID-кадр потерян во время речевой паузы, то формируется комфортный шум с параметрами предыдущего SID-кадра. Если потерян еще один SID-кадр, то комфортный шум постепенно заглушается.

Применение экстраполяции речи при цифровой передаче, формирование плавных акустических переходов при замираниях сигнала в каналах в совокупности с полным DTX-процессом значительно улучшает потребительские качества связи с GSM PLMN по сравнению с существующими аналоговыми сотовыми системами связи.

13.7.8. Микросхемы для реализации GMSK

Модуляция GMSK применяется прежде всего в широко распространенной и зарекомендовавшей себя цифровой системе мобильной сотовой радиосвязи GSM-900 (*Global System*

for Mobile Communications), в более высокочастотных ее вариантах — европейском DCS-1800 (GSM-1800) и северо-американском PCS-1900 (GSM-1900), а также в новой системе GPRS (*General Packet Radio Service*).

Отметим весьма важное обстоятельство, касающееся построения современных приемных каналов в системах GSM и других, не связанное, однако, с рассматриваемым видом модуляции. Это — применение супергомодинного метода радиоприема вместо традиционного супергетеродинного. При супергомодинном методе нет промежуточного преобразования частоты (ВЧ в ПЧ), а настройка приемного канала на требуемую частоту осуществляется изменением опорной частоты демодулятора. При этом, благодаря более совершенной фильтрации на входе приемного канала, а также применению цифровой манипуляции (вместо аналоговой модуляции), обеспечивается высококачественный прием радиосигналов без недостатков, присущих супергетеродинному методу. Кроме того, применение супергомодинного метода приводит к упрощению схемы приемного канала и снижению стоимости комплектов микросхем.

Ниже дается краткое описание микросхем ведущих фирм-производителей, использующих GSMK.

Фирма *Analog Devices* выпускает чипсет *Othello* в составе двух микросхем AD6523 (модулятор/демодулятор приемопередатчика) и AD6524 (синтезатор частот) и анонсировала новый чип *Othello One* — одну микросхему AD6534, заменяющую указанные две. В чипсете *Othello* и чипе *Othello One* используется рассмотренный выше «виртуальный» модулятор (рис. 13.13). На рис. 13.13 он показан в упрощенном виде, а в указанных устройствах он имеет отличия. Дополнительно в его составе имеются делители частоты в N и M раз. Кроме того, существуют решения, в которых перемножитель смесителя выполнен с квадратурным выходом, правда, расшифровка его устройства не приводится.

Канал передачи чипа *Othello One* с дополнительными внешними элементами содержит рассмотренный модулятор с «виртуальной» ПЧ, преобразователь частоты (ПЧ в ВЧ) и два коммутируемых усилителя мощности для диапазонов частот GSM 900 МГц и GSM (DCS, PCS) 1800/1900 МГц.

Преобразователь и усилители, вместе с антенным коммутатором и входными полосовыми фильтрами приемного канала, составляют внешнюю высокочастотную часть схемы приемопередатчика. Канал приема содержит три коммутируемых входных цепи с полосовыми фильтрами и малошумящими усилителями для диапазонов 900, 1800 и 1900 МГц, демодулятор и усилители низкой частоты с ФНЧ. В приемном канале чипа *Othello One*, как и чипсета *Othello*, используется указанный супергомодинный метод.

Для работы с радиочастотными AD6523/4 и AD6534 предназначены низкочастотные чипсет AD20msp425 (содержащий микросхемы AD6421 и AD6426) и новый чипсет AD20msp430 «SoftFone» (содержащий AD6521 и AD6522). AD6421 и AD6521 — это кодеки, а AD6426 и AD6522 — процессоры.

Чипсет *Othello* и чип *Othello One* могут быть использованы также в системах GPRS и EDGE (*Enhanced Data GSM Environment*). Так как в EDGE применена 8-PSK (8-позиционная фазовая манипуляция), в передающем канале на выходе чипа *Othello One* используется микросхема AD6021, в которой осуществляется как преобразование частоты, так и преобразование GMSK в 8-PSK.

Наряду с рассмотренными чипсетом *Othello* и чипом *Othello One*, фирма *Analog Devices* выпускает микросхему более ранней разработки AD6432, представляющую собой подсистему канала промежуточной частоты приемопередатчика. В указанной микросхеме реализуется супергетеродинный метод, а для модуляции используется квадратурный модулятор. Микросхема предназначена для работы в системах радиосвязи GSM/DCS/PCS совместно с чипсетами AD20msp410 (AD7015, ADSP2178, ADPLP01) или AD20msp415 (AD6421, AD6422), в которых формируется сигнал модуляции.

В цифровых микросхемах AD9857 и AD9773/5/7 используются, соответственно, цифровые квадратурные модуляторы, в которых также предусмотрена реализация GMSK. Помимо модуляторов, указанные микросхемы содержат интерполирующие фильтры (на входе) и ЦАП (на выходе).

Для системы DECT, также использующей GMSK, выпускается микросхема приемопередатчика AD6411, модулятор которой соответствует схеме на рис. 13.10.

Фирма *Texas Instruments* также приступила к выпуску новых чипсетов: TCS1100 (содержащий микросхемы TBB1100, TWL3012 и TRF6053/TRF2253) — для системы GSM; TCS2100 (содержащий TBB2100, TWL3014 и TRF6150) — для системы GPRS [11]. TRF6053 и TRF2253 — это микросхемы приемопередатчика и синтезатора, а TRF6150 — приемопередатчика со встроенным синтезатором. В TRF6053 (для GSM) модулятор с преобразователем частоты выполнен по схеме, аналогичной на рис. 13.13, но в приемном канале используется супергетеродинный метод. В TRF6150 (для GPRS) модулятор с преобразователем выполнен по схеме, аналогичной на рис. 13.10, а в приемном канале используется супергомодинный метод.

В передающем канале приемопередатчика с TRF6150 содержатся также два коммутируемых усилителя мощности, а в приемном канале — два входных полосовых фильтра для диапазонов 900 и 1800/1900 МГц. Входной фильтр 900 МГц постоянно подключен к демодулятору 900 МГц, а фильтр 1800/1900 МГц может быть подключен к одному из демодуляторов — 1800 или 1900 МГц. Все три демодулятора — квадратурные. Микросхема TRF6150 содержит в своем составе два синтезатора — типа *Integer-N* для модулятора (ω_{01} на рис. 13.10) и типа *Fractional-N* для преобразователя ПЧ в ВЧ в передающем канале (ω_{02} на рис. 13.12) и демодуляторов приемного канала.

Предшественниками микросхем TRF6053, TRF2253 и TRF6150, являются модулятор/синтезатор TRF3040, модулятор TRF3520, приемник TRF1020, а также синтезаторы TRF2020, TRF2050.

Фирма *Infineon Technologies* также выпускает новые чипсеты для систем GSM и GPRS. В их составе: PMB 6250/3/6 (SMARTi / SMARTi+ / SMARTi DC) — микросхемы приемопередатчиков, из которых последняя — с DC (*Direct Converter*), то есть без промежуточного преобразования частоты; PMB 6850 (E-GOLD+) и PMB 8880 (M-GOLD) — микросхемы для цифровой обработки низкочастотного сигнала, в том числе для формирования сигнала модуляции; другие микросхемы, образующие с указанными комплекты, необходимые для формирования приемопередающих систем. Модулятор и преобразователь ПЧ в ВЧ выполнены по схеме, аналогичной на рис. 13.12 (но фильтр $\Phi 2$ не показан).

В чипсете для системы DECT — микросхемы PMB 6720, PMB 6610 и PMB 6818/9. В микросхеме приемопередатчика PMB 6610 используется модулятор по схеме, аналогичной на рис. 13.10.

13.8. Безопасность в мобильных сетях

Алгоритмы шифрования делят на 2 класса: классические (используют 1 ключ для шифрования и дешифрования и открытые, использующие 2 ключа — один для шифрования, другой — для дешифрования).

В стандарте GSM определены следующие механизмы безопасности:

- аутентификация;
- секретность передачи данных;
- секретность абонента;
- секретности направлений соединения абонентов.

Защита сигналов управления и данных пользователя осуществляется по радиоканалу.

13.8.1. Механизм аутентификации

Для исключения несанкционированного использования ресурсов системы связи определены механизмы аутентификации — удостоверение подлинности абонента. Каждый подвижный объект на время пользования системой получает стандартный модуль подлинности абонента SIM (*Standard Identification Module*), который содержит: международный идентификационный номер подвижного абонента (IMSI), свой индивидуальный ключ аутентификации (K_i), алгоритм идентификации (A3).

С помощью заложенной в SIM информации в результате взаимного обмена данными между подвижной станцией и сетью осуществляется полный цикл аутентификации и разрешение проблемы доступа к сети.

Процедура проверки сетью подлинности абонента работает следующим образом. Сеть передает случайный номер (RAND) на подвижную станцию. На ней с помощью ключа K_i и алгоритма аутентификации A3 определяется значение отклика SRES (*Signed RESponse*), т. е. $SRES = K_i(RAND)$. Подвижная станция посылает вычисленное значение SRES в сеть, которая

сверяет значение принятого SRES со значением SRES, вычисленного сетью. Если оба значения совпадают, подвижная станция приступает к передаче сообщений. В противном случае связь прерывается, и индикатор подвижной станции показывает, что опознавание не состоялось. Для обеспечения секретности вычисление SRES происходит в рамках SIM.

13.8.2. Секретность передачи данных

Для обеспечения секретности передачи по радиоканалу информации вводится следующий механизм защиты. Все конфиденциальные сообщения можно передавать в рамках защиты информации. Алгоритм формирования ключей шифрования (A8 — *Ciphering Key Generation Algorithm*) хранится в модуле SIM. После приема следующего номера RAND подвижная станция вычисляет, кроме отклика SRES, также и ключ шифрования K_c (*Ciphering Key*), используя RAND, K_i и алгоритм A8:

$$K_c = K_i(RAND). \quad (13.8)$$

Ключ K_c не передается по радиоканалу. Как подвижная станция, так и сеть вычисляют ключ шифрования, который используют другие подвижные абоненты. По причине секретности вычисление K_c происходит в SIM.

13.9. Беспроводные системы третьего поколения

В 1990-х годах акценты в сфере мобильной связи сместились в сторону цифровых технологий, что позволило значительно повысить пропускную способность систем беспроводной связи. Речевой сигнал переводится в цифровую форму и передается в течение определенного интервала времени по каналу связи. Этот процесс позволяет множественным абонентам одновременно пользоваться одним и тем же каналом связи, который в аналоговой системе связи использовался бы только одним из них. В настоящее время в сотовой связи применяются такие технологии, как TDMA, FDMA, CDMA. Сервисы включают передачу речи, а также комбинацию передачи данных по коммутируемым каналам и некоторых пакетно-коммутируемых данных.

Система GSM предоставляет пользователям широкий ассортимент услуг, как речевых, так и неречевой природы.

Помимо телефонии к речевым услугам относят вызовы спецслужб (полиция, скорая помощь, пожарные и т. п.), как правило, путем набора номера 112, который принят на Европейском континенте в качестве стандарта, и речевую почту.

Набор неречевых услуг основывается на перечне услуг ISDN и для абонентов сети GSM состоит из трех с половиной десятков наименований. Услуги по передаче данных различаются в зависимости от потенциальных корреспондентов (абоненты телефонной сети общего пользования, либо ISDN, либо специализированных сетей), от характера передаваемой информации (данные, факсимиле, видеотекст, телетекст и пр.), от режима передачи (коммутация пакетов либо каналов, сквозной цифровой канал либо с использованием телефонных модемов и пр.), от типа терминалов и т. д. Специфическими для подвижной сети являются службы коротких сообщений (SMS — *Short Message Service* — исходящие, входящие и вещательные), которые по сути дела представляют собой разновидность службы персонального вызова (пейджинга).

Стандарт GSM принят в России в качестве федерального. С января 1996 года в Москве и области началась коммерческая эксплуатация ССПС стандарта GSM. Оператором сети является компания «Московские ТелеСистемы» (МТС). Сеть обслуживает более 12 тысяч абонентов и обеспечивает автоматический роуминг со странами Европы.

Дальнейшее развитие систем сотовой подвижной связи осуществляется в рамках проекта создания ССПС третьего поколения. В Европе работы по созданию ССПС третьего поколения, получившей название универсальная система подвижной связи (UMTS — *Universal Mobile Telecommunication System*), проводятся СЕРТ по исследовательской программе RACE. Концепция создания UMTS предусматривает объединение функциональных возможностей существующих цифровых систем связи в единую систему с предоставлением стандартизированных услуг подвижной связи (сотовой, беспроводной, персонального вызова и пр.).

Работы по созданию единой международной ССПС третьего поколения, получившей название FPLMTS, проводит МСЭ.

В XXI веке мы движемся в направлении третьего поколения (3G, 3 *Generation*) беспроводных систем связи, а именно

сетей стандарта IMT-2000 (*International Mobile Telecommunication-2000*), предложенного Международным союзом телекоммуникаций (ITU). Эфирный интерфейс будет представлен широкополосным вариантом CDMA-технологии. 3G-системы будут представлять мультимедийные сервисы. Будут передаваться данные со скоростью 2 Мбит/с. Услуги связи будут по всему миру одинаковыми. Нужно будет ввести единый стандарт.

История беспроводных технологий передачи информации началась в конце XIX века с передачей первого радиосигнала и появлением в 20-х годах XX века первых радиоприемников с амплитудной модуляцией. В 1930-е годы появилось радио с частотной модуляцией и телевидение. В 1970-е годы были созданы первые беспроводные телефонные системы. Сначала это были аналоговые сети, в начале 1980-х появился стандарт GSM, ознаменовавший начало перехода на цифровые стандарты как обеспечивающие лучшее распределение спектра, лучшее качество сигнала и большую безопасность. С 90-х годов XX века происходит укрепление позиций беспроводных сетей. Беспроводные технологии прочно входят в нашу жизнь. Развиваясь с огромной скоростью, они стимулируют создание новых устройств и услуг.

Обилие новых беспроводных технологий, таких как CDMA (*Code Division Multiple Access* — технология с кодовым разделением каналов), GSM (*Global System for Mobile Communications* — глобальная система для мобильных коммуникаций), GPRS (*General Packet Radio Service* — услуга пакетной передачи данных), *Bluetooth*, EDGE (*Enhanced Data Rates for GSM Evolution* — увеличенная скорость передачи данных для GSM), свидетельствует о том, что в области беспроводных технологий зреет революционный подъем.

Весьма перспективно и развитие сетей WLAN (*Wireless Local Area Network* — беспроводные локальные сети), которые используют для связи OFDM-сигналы или сигналы с расширенным спектром. Технологии беспроводных сетей включают в себя широкий диапазон решений, начиная от глобальных сетей передачи голоса и данных, позволяющих пользователю

устанавливать беспроводные соединения на значительных расстояниях, и заканчивая технологиями инфракрасной и радиосвязи, используемыми на небольших расстояниях. Технологии беспроводных сетей применяются в портативных и настольных компьютерах, карманных компьютерах, сотовых телефонах. Беспроводные технологии могут использоваться для самых различных целей. Например, мобильные пользователи могут использовать свои сотовые телефоны для доступа к электронной почте. Путешественники с портативными компьютерами могут подключаться к Интернету через базовые станции, установленные в аэропортах, на вокзалах и в других общественных местах. У себя дома можно подключать устройства к настольному компьютеру для синхронизации данных и передачи файлов.

Беспроводные технологии позволяют использовать многообразные устройства для доступа к данным по всему миру, а также снижают или полностью устраняют затраты на прокладку дорогостоящих оптоволоконных или кабельных каналов передачи данных, предоставляя при этом все возможности проводных сетей. Адаптеры беспроводной сети, которые бывают внутренними и внешними, позволяют подключать компьютеры к сети без помощи кабелей или каких-либо иных физических соединений. Передаваемые данные разбиваются на небольшие пакеты и транслируются между компьютером и приемопередатчиками в виде радиосигналов в специально отведенном диапазоне частот. Беспроводные сети развертываются в аэропортах, университетах, отелях, ресторанах, на предприятиях. Началом разработки стандартов беспроводных сетей является образование всемирной организацией IEEE (*Institute of Electrical and Electronics Engineers* — Институт инженеров по электричеству и электронике) комитета 802.11 в 1990 году. Значительный импульс развитию беспроводных технологий дала Всемирная Паутина и идея работы в Сети при помощи беспроводных устройств.

В конце 90-х годов пользователям был предложен новый протокол беспроводных технологий WAP (*Wireless Application Protocol* — технология просмотра информации из сети Internet на экране сотового телефона). Эта услуга сначала не вызвала интереса. Это были основные информационные услуги — новости, погода, всевозможные расписания и т. п. Также

поначалу не пользовались спросом и *Bluetooth*, и *WLAN* — в основном из-за высокой стоимости этих средств связи. Однако по мере снижения цен рос и интерес населения.

К середине первого десятилетия XXI века счет пользователей беспроводного *Internet*-сервиса пошел на десятки миллионов. С появлением беспроводной *Internet*-связи на первый план вышли вопросы обеспечения безопасности. Основные угрозы при использовании беспроводных сетей — это перехват сообщений спецслужб, коммерческих предприятий и частных лиц, перехват номеров кредитных карточек, кража оплаченного времени соединения, вмешательство в работу коммуникационных центров. Эти проблемы успешно решаются по мере совершенствования стандартов связи.

Существенна для развития беспроводных технологий и возможность их применения домашними пользователями. Чем больше устройств в домашней сети, тем сильнее загромождают дом соединяющие их провода, усложняющие процессы перестановки мебели. Повышение степени комфортности современного дома, объединение в одно целое всех его структур и объектов (компьютера, телевизора, цифровой фотокамеры, домашнего развлекательного центра, системы охраны, климатической системы, бытовой техники и т. д.) — основа идеи создания интеллектуального цифрового дома, которая также реализуется с помощью беспроводных устройств.

Несомненно, что в недалеком будущем мы будем жить в беспроводном мире.

14.1. Технология Wi-Fi

Wi-Fi (*Wireless Fidelity* — высокая точность беспроводной передачи данных) представляет собой современную беспроводную технологию соединения компьютеров в локальную сеть, подключенную к системе *Internet*. Благодаря этой технологии *Internet* становится мобильным и дает пользователю свободу перемещения по всему миру. В настоящее время под Wi-Fi понимается целое семейство стандартов передачи цифровых потоков данных по радиоканалам.

С увеличением числа мобильных пользователей возникает острая необходимость в оперативном создании коммуникаций между ними, в обмене данными, в быстром получении ин-

формации. Поэтому естественным образом происходит интенсивное развитие технологий беспроводных коммуникаций. Особенно это актуально в отношении беспроводных сетей, называемых WLAN-сетями.

Благодаря функции роуминга между точками доступа пользователи могут перемещаться по территории покрытия сети Wi-Fi без разрыва соединения. WLAN-сети имеют ряд преимуществ перед обычными кабельными сетями:

- WLAN-сеть можно очень быстро развернуть, что очень удобно при проведении презентаций или в условиях работы вне офиса;
- пользователи мобильных устройств при подключении к локальным беспроводным сетям могут легко перемещаться в рамках действующих зон сети;
- скорость современных сетей довольно высока (до 108 Мбит/с), что позволяет использовать их для решения очень широкого спектра задач;
- WLAN-сеть может оказаться единственным выходом, если невозможна прокладка кабеля для обычной сети.

Вместе с тем необходимо помнить об ограничениях беспроводных сетей. Это, как правило, все-таки меньшая скорость, подверженность влиянию помех и более сложная схема обеспечения безопасности передаваемой информации. Сегмент Wi-Fi сети может использоваться как самостоятельная сеть, либо в составе более сложной сети, содержащей как беспроводные, так и обычные проводные сегменты.

14.2. Основные элементы сети Wi-Fi

Для построения беспроводной сети используются Wi-Fi адаптеры и точки доступа. Адаптер выполняет ту же функцию, что и сетевая карта в проводной сети: он служит для подключения компьютера пользователя к беспроводной сети. Благодаря платформе *Centrino* все современные ноутбуки имеют встроенные адаптеры Wi-Fi, совместимые со многими современными стандартами. Wi-Fi адаптерами, как правило, снабжены и карманные персональные компьютеры (КПК), что также позволяет подключать их к беспроводным сетям.

Для доступа к беспроводной сети адаптер может устанавливать связь непосредственно с другими адаптерами. Такая сеть называется *беспроводной одноранговой сетью*, или *Ad Hoc*

(к случаю). Адаптер также может устанавливать связь через специальное устройство — *точку доступа*. Такой режим называется *инфраструктурой*.

В ближайшем будущем мы будем пользоваться своим компьютером, как сейчас — мобильным телефоном, т.е., взяв ноутбук, в любом месте сможем войти в *Internet*.

Через точку доступа осуществляется взаимодействие и обмен информацией между беспроводными адаптерами, а также связь с проводным сегментом сети. Таким образом, точка доступа играет роль коммутатора. Точка доступа имеет сетевой интерфейс (*uplink port*), при помощи которого она может быть подключена к обычной проводной сети. Точка доступа может использоваться как для подключения к ней клиентов (базовый режим точки доступа), так и для взаимодействия с другими точками доступа с целью построения распределенной сети WDS (*Wireless Distributed System*). Это режимы беспроводного моста «точка-точка» и «точка — много точек», беспроводный клиент и повторитель возможностей подключения клиента к глобальной сети.

Доступ к глобальной сети обеспечивается путем передачи широкополосных сигналов через эфир. Принимающая станция может получать сигналы в диапазоне работы нескольких передающих станций. Станция-приемник использует идентификатор зоны обслуживания (*Service Set Identifier, SSID*) для фильтрации получаемых сигналов и выделения того, который ей нужен.

Зоной обслуживания (*Service Set, SS*) называются логически сгруппированные устройства, обеспечивающие подключение к беспроводной сети.

Базовая зона обслуживания (*Basic Service Set, BSS*) — это группа станций, которые связываются друг с другом по беспроводной связи. Технология BSS предполагает наличие особой станции, которая называется точкой доступа (*Access Point, AP*).

14.3. Архитектура IEEE 802.11

Самый популярный стандарт беспроводных локальных сетей — IEEE 802.11. Уровень доступа к среде стандартов 802.11 достаточно сложен, но постараемся кратко и понятно его описать.

В 1990 году IEEE сформировал рабочую группу по стандартам для беспроводных локальных сетей 802.11. Эта группа занялась разработкой всеобщего стандарта для радиооборудования и сетей, работающих на частоте 2,4 ГГц, со скоростями доступа 1 и 2 Мбит/с. Работы по созданию стандарта были завершены через 7 лет, и в июне 1997 года была утверждена первая спецификация 802.11. Стандарт IEEE 802.11 явился первым стандартом для продуктов WLAN.

Рассмотрим сначала подробно архитектуру самого популярного стандарта беспроводных локальных сетей — IEEE 802.11, а затем остановимся на особенностях стандартов IEEE 802.11a, IEEE 802.11b и IEEE 802.11g.

14.3.1. Стек протоколов IEEE 802.11

Стек протоколов стандарта IEEE 802.11 соответствует общей структуре стандартов комитета 802, то есть состоит из физического уровня и канального уровня с подуровнями управления доступом к среде MAC и логической передачи данных LLC (отсылаем читателя к разделу 5 об OSI настоящего пособия). Как и у всех технологий семейства 802, технология 802.11 определяется двумя нижними уровнями, то есть физическим уровнем и подуровнем MAC, а подуровень LLC выполняет свои стандартные общие для всех технологий LAN функции (рис. 14.1).

На физическом уровне существует несколько вариантов спецификаций, которые отличаются используемым частотным диапазоном, методом кодирования и как следствие — скоростью передачи данных. Все варианты физического уровня работают с одним и тем же алгоритмом уровня MAC, но некоторые временные параметры уровня MAC зависят от используемого физического уровня.

14.3.2. Уровень доступа к среде стандарта 802.11

В сетях 802.11 уровень MAC обеспечивает два режима доступа к разделяемой среде (рис. 14.1):

- распределенный режим DCF (*Distributed Coordination Function*);
- централизованный режим PCF (*Point Coordination Function*).

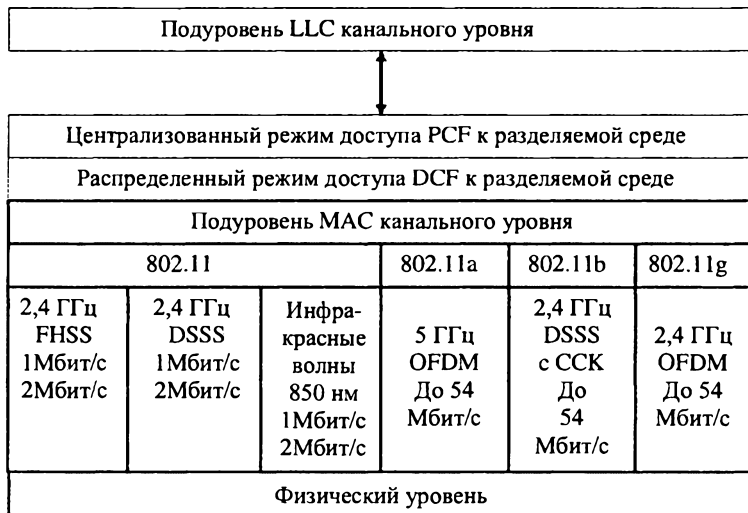


Рис. 14.1. Стек протоколов IEEE 802.11

Распределенный режим доступа DCF

Как обеспечивается доступ в распределенном режиме DCF? В этом режиме реализуется метод множественного доступа с контролем несущей и предотвращением коллизий (*Carrier Sense Multiple Access with Collision Avoidance, CSMA/CA*). Вместо неэффективного в беспроводных сетях прямого распознавания коллизий по методу CSMA/CD здесь используется их косвенное выявление. Для этого каждый переданный кадр должен подтверждаться кадром положительной квитанции, посылаемым станцией назначения. Если же по истечении оговоренного тайм-аута квитанция не поступает, станция-отправитель считает, что произошла коллизия. Режим доступа DCF требует синхронизации станций.

В спецификации 802.11 эта проблема решается так. Временные интервалы начинают отсчитываться от момента окончания передачи очередного кадра (рис. 14.2). Это не требует передачи каких-либо специальных синхронизирующих сигналов и не ограничивает размер пакета размером слота, так как слоты принимаются во внимание только при принятии реше-

ния о начале передачи кадра. Станция, которая хочет передать кадр, обязана предварительно прослушать среду.

Стандарт IEEE 802.11 предусматривает два механизма контроля присутствия в канале несущей: физический и виртуальный. Первый механизм реализован на физическом уровне и сводится к определению уровня сигнала в антенне и сравнению его с пороговой величиной. Виртуальный механизм обнаружения несущей основан на том, что в передаваемых кадрах данных, а также в управляющих кадрах ACK (*ACKnowledgment* — квитанция) и RTS (*Request To Send* — запрос на передачу)/CTS (*Clear To Send* — среда свободна для передачи) содержится информация о времени, необходимом для передачи пакета (или группы пакетов) и получения подтверждения. Все устройства сети получают информацию о текущей передаче и могут определить, сколько времени канал будет занят, т. е. устройство при установлении связи сообщает всем, на какое время оно резервирует канал. Как только станция фиксирует окончание передачи кадра, она обязана отсчитать интервал времени, равный межкадровому интервалу (*Inter Frame Spacing, IFS*). Если после истечения IFS среда все еще свободна, начинается отсчет слотов фиксированной длительности. Кадр можно передавать только в начале какого-либо из слотов при условии, что среда свободна. Станция выбирает для передачи слот на основании усеченного экспоненциального двоичного алгоритма отсрочки, аналогичного используемому в методе CSMA/CD. Номер слота выбирается как случайное целое число, равномерно распределенное в интервале $[0, CW]$, где CW означает *Contention Window* (конкурентное окно).



Рис. 14.2. Режим доступа DCF

Рассмотрим этот довольно непростой метод доступа на примере рис. 14.2. Пусть станция *A* выбрала для передачи на основании усеченного экспоненциального двоичного алгоритма отсрочки слот 3. При этом она присваивает таймеру отсрочки (назначение его будет объяснено позже) значение 3 и начинает проверять состояние среды в начале каждого слота. Если среда свободна, то из значения таймера отсрочки вычитается 1, и если результат равен нулю, начинается передача кадра. Таким образом обеспечивается условие незанятости всех слотов, включая выбранный. Это условие является необходимым для начала передачи. Если же в начале какого-нибудь слота среда оказывается занятой, то вычитания единицы не происходит, и таймер «замораживается». В этом случае станция начинает новый цикл доступа к среде, изменяя только алгоритм выбора слота для передачи. Как и в предыдущем цикле, станция следит за средой и при ее освобождении делает паузу в течение межкадрового интервала. Если среда осталась свободной, то станция использует значение «замороженного» таймера в качестве номера слота и выполняет описанную выше процедуру проверки свободных слотов с вычитанием единиц, начиная с замороженного значения таймера отсрочки.

Размер слота зависит от способа кодирования сигнала; так, для метода FHSS (*Frequency Hopping Spread Spectrum*) размер слота равен 28 мкс, а для метода DSSS (*Direct Sequence Spread Spectrum*) (мы отсылаем читателя к разделу 8 настоящего учебного пособия) — 1 мкс. Размер слота выбирается таким образом, чтобы он превосходил время распространения сигнала между любыми двумя станциями сети плюс время, затрачиваемое станцией на распознавание занятости среды. Если такое условие соблюдается, то каждая станция сети сумеет правильно распознать начало передачи кадра при прослушивании слотов, предшествующих выбранному ею для передачи слоту. Это, в свою очередь, означает следующее. Коллизия может иметь место только в том случае, когда несколько станций выбирают один и тот же слот для передачи. В этом случае кадры искажаются, и квитанции от станций назначения не приходят. Не получив в течение определенного времени квитанцию, отправители фиксируют факт коллизии и пытаются передать свои кадры снова. При каждой повторной неудачной попытке передачи кадра интервал $[0, CW]$, из которого

выбирается номер слота, удваивается. Если, например, начальный размер окна выбран равным 8 (т. е. $CW = 7$), то после первой коллизии размер окна должен быть равен 16 ($CW = 15$), после второй последовательной коллизии — 32 и т. д.

Начальное значение CW должно выбираться в зависимости от типа физического уровня, используемого в беспроводной локальной сети. Как и в методе CSMA/CD, в данном методе количество неудачных попыток передачи одного кадра ограничено, но стандарт 802.11 не дает точного значения этого верхнего предела. Когда верхний предел в N попыток достигнут, кадр отбрасывается, а счетчик последовательных коллизий устанавливается в нуль. Этот счетчик также устанавливается в нуль, если кадр после некоторого количества неудачных попыток передается успешно.

В беспроводных сетях возможна проблема скрытого терминала, когда два устройства A и B удалены и не слышат друг друга, однако оба попадают в зону действия третьего устройства C , так как в этом случае, если устройства A и B начнут передачу, то они принципиально не смогут обнаружить конфликтную ситуацию и определить, почему пакеты не проходят.

Для устранения эффекта скрытого терминала в режиме доступа DCF станция, которая хочет захватить среду и в соответствии с описанным алгоритмом начинает передачу кадра в определенном слоте, вместо кадра данных сначала посылает станции назначения короткий служебный кадр RTS. На этот запрос станция назначения должна ответить служебным кадром CTS, после чего станция-отправитель посылает кадр данных. Кадр CTS должен оповестить о захвате среды те станции, которые находятся вне зоны сигнала станции-отправителя, но в зоне досягаемости станции-получателя, то есть являются скрытыми терминалами для станции-отправителя.

Максимальная длина кадра данных 802.11 равна 2346 байт, длина RTS-кадра — 20 байт, CTS-кадра — 14 байт. Так как RTS- и CTS-кадры гораздо короче, чем кадр данных, потери данных в результате коллизии RTS- или CTS-кадров гораздо меньше, чем при коллизии кадров данных. Процедура обмена RTS- и CTS-кадрами не обязательна. От нее можно отказаться при небольшой нагрузке сети, поскольку в такой ситуации коллизии случаются редко, а значит, не стоит тратить

дополнительное время на выполнение процедуры обмена RTS-и CTS-кадрами. При помехах иногда случается, что теряются большие фреймы данных, поэтому можно уменьшить длину этих фреймов путем фрагментации.

Фрагментация фрейма — это выполняемая на уровне MAC функция, назначение которой — повысить надежность передачи фреймов через беспроводную среду. Под фрагментацией понимается дробление фрейма на меньшие фрагменты и передача каждого из них отдельно (рис. 14.3). Предполагается, что вероятность успешной передачи меньшего фрагмента через зашумленную беспроводную среду выше. Получение каждого фрагмента фрейма подтверждается отдельно; следовательно, если какой-нибудь фрагмент фрейма будет передан с ошибкой или вступит в коллизию, передавать повторно придется только его, а не весь фрейм. Это увеличивает пропускную способность среды.

Размер фрагмента может задавать администратор сети. Фрагментации подвергаются только одноадресные фреймы. Широковещательные, или многоадресные, фреймы передаются целиком. Кроме того, фрагменты фрейма передаются пакетом, с использованием только одной итерации механизма доступа к среде DCF. Хотя за счет фрагментации можно повысить надежность передачи фреймов в беспроводных локальных сетях, она приводит к увеличению «накладных расходов» MAC-протокола стандарта 802.11. Каждый фрагмент фрейма включает информацию, содержащуюся в заголовке

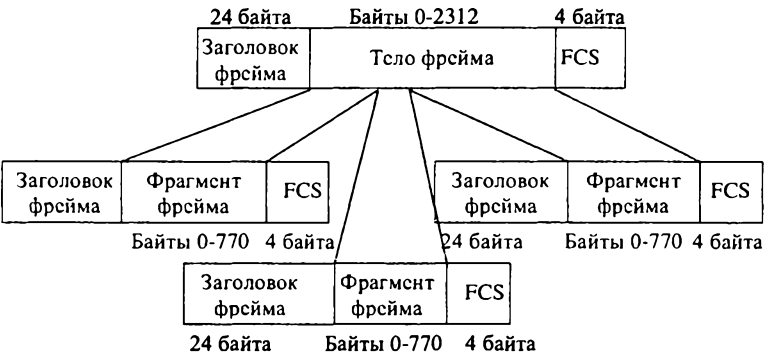


Рис. 14.3. Фрагментация фрейма

802.11 MAC, а также требует передачи соответствующего фрейма подтверждения. Это увеличивает число служебных сигналов MAC-протокола и снижает реальную производительность беспроводной станции.

Централизованный режим доступа PCF

В том случае, когда в сети имеется станция, выполняющая функции точки доступа, может также применяться централизованный режим доступа PCF, обеспечивающий приоритетное обслуживание трафика. В этом случае говорят, что точка доступа играет роль *арбитра* среды. Режим доступа PCF в сетях 802.11 сосуществует с режимом DCF. Оба режима координируются с помощью трех типов межкадровых интервалов (рис. 14.4).

После освобождения среды каждая станция отсчитывает время простоя среды, сравнивая его с тремя значениями:

- короткий межкадровый интервал (Short IFS, SIFS);
- межкадровый интервал режима PCF (PIFS);
- межкадровый интервал режима DCF (DIFS).

Захват среды с помощью механизма DCF возможен, когда среда свободна в течение времени, равного или большего, чем DIFS, т. е. в качестве IFS в режиме DCF используется интервал DIFS — самый длительный из трех возможных, что дает этому режиму самый низкий приоритет.

Межкадровый интервал SIFS имеет наименьшее значение, он служит для первоочередного захвата среды ответными CTS-

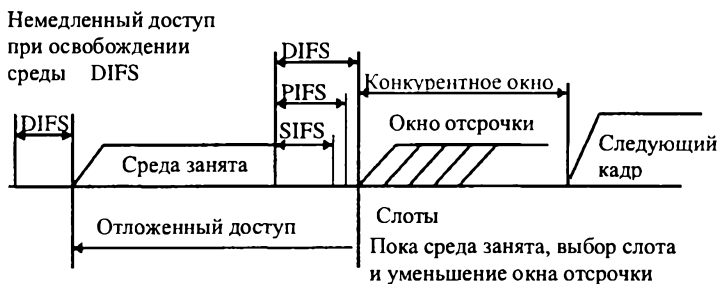


Рис. 14.4. Сосуществование режимов PCF и DCF

кадрами или квитанциями, которые продолжают или завершают уже начавшуюся передачу кадра.

Значение межкадрового интервала PIFS больше, чем SIFS, но меньше, чем DIFS. Промежутком времени между завершением PIFS и DIFS пользуется арбитр среды. В этом промежутке он может передать специальный кадр, которым оповещаются все станции о начале контролируемого периода. Приняв этот кадр, станции, которые намеревались с помощью алгоритма DCF захватить среду, не могут этого сделать, пока не закончится контролируемый период, длительность которого объявляется в специальном кадре.

Но этот контролируемый период может закончиться и раньше, если у станций нет чувствительного к задержкам трафика. В этом случае арбитр передает служебный кадр, после которого по истечении интервала DIFS начинает работать режим DCF. На управляемом интервале реализуется централизованный метод доступа PCF. Арбитр выполняет процедуру опроса, чтобы по очереди предоставить каждой такой станции право на использование среды, направляя ей специальный кадр. Станция, приняв такой кадр, может ответить другим кадром, который подтверждает прием специального кадра и одновременно передает данные (либо по адресу арбитра для транзитной передачи, либо непосредственно станции). Для того чтобы какая-то доля среды всегда доставалась асинхронному трафику, длительность контролируемого периода ограничена. После его окончания арбитр передает соответствующий кадр и начинается неконтролируемый период. Каждая станция может работать в режиме PCF, для этого она должна подписаться на данную услугу при присоединении к сети.

Кадр MAC-подуровня

На рис. 14.5 изображен формат кадра 802.11. Здесь обозначено: FC — управление кадром (указывается тип кадра и предоставляется управляющая информация), D/I — идентификатор длительности соединения (если используется поле длительности, указывается время (в мкс), на которое требуется выделить канал для успешной передачи кадра MAC, а в некоторых кадрах управления в этом поле указывается идентификатор ассоциации или соединения), SC — управление

2	2	6	6	6	2	6	0-2312	4 октета
FC	ДЛ	Адрес	Адрес	Адрес	SC	Адрес	Тело кадра	CRC

Рис. 14.5. Формат кадра MAC IEEE 802.11

очередностью (содержит 4-битовое подполе номера фрагмента, используемое для фрагментации и повторной сборки, и 12-битовый порядковый номер, используемый для нумерации кадров, передаваемых между приемником и передатчиком), CRC (*Cyclic Redundancy Code*) — циклический избыточный код. Приведенная общая структура применяется для всех информационных и управляющих кадров, хотя не все поля используются во всех случаях.

Перечислим поля кадра:

- поле управления кадром;
- идентификатор длительности соединения;
- число и значение полей адреса зависит от контекста, при этом возможны следующие адреса: источника, назначения, передающей станции, принимающей станции;
- управление очередностью;
- тело кадра, которое содержит модуль данных протокола LLC или управляющую информацию MAC;
- контрольная последовательность кадра, представляющая собой 32-битовый блок проверки четности с избыточностью.

Поле управления кадром, показанное на рис. 14.6, состоит из следующих подполей, указывающих:

- версию протокола, например для версии 802.11 (текущая версия) — 0;
- тип кадра: контроль, управление или данные;
- подтип, детализирующий дальнейшую идентификацию функций кадра. Разрешенные сочетания типов и подтипов перечислены в табл. 14.1;

2	2	4	1	1	1	1	1	1	1	1 октеты
Версия	Тип	Подтип	к DS	от DS	MF	RT	PM	MD	W	O

Рис. 14.6. Поле управления кадром

Разрешенные комбинации типа и подтипа

Значение типа	Описание типа	Значение подтипа	Описание подтипа
00	Управление	0000	Запрос ассоциации
00	Управление	0001	Ответ на запрос ассоциации
00	Управление	0010	Запрос повторной ассоциации
00	Управление	0011	Ответ на запрос повторной ассоциации
00	Управление	0100	Пробный запрос
00	Управление	0101	Ответ на пробный запрос
00	Управление	1000	Сигнальный кадр
00	Управление	1001	Объявление наличия трафика
00	Управление	1010	Разрыв ассоциации
00	Управление	1011	Аутентификация
00	Управление	1100	Отмена аутентификации
01	Контроль	1010	PS-опрос
01	Контроль	1011	Запрос передачи
01	Контроль	1100	«Готов к передаче»
01	Контроль	1101	Подтверждение
01	Контроль	1110	Без состязания (CF)-конец
01	Контроль	1111	CF-конец + CF-подтверждение
10	Данные	0000	Данные
10	Данные	0001	Данные + CF-подтверждение
10	Данные	0010	Данные + CF-опрос
10	Данные	0011	Данные + CF-подтверждение + CF-опрос
10	Данные	0100	Нулевая функция без данных
10	Данные	0101	Данные + CF-подтверждение
10	Данные	0110	Данные + CF-опрос
10	Данные	0111	Данные + CF-подтверждение + CF-опрос

- к DS — система распределения (координационная функция MAC присваивает этому биту значение 1, если кадр предназначен распределительной системе);
- от DS (координационная функция MAC присваивает этому биту значение 0, если кадр исходит от распределительной системы);
- MF — больше фрагментов (1, если за данным фрагментом следует еще несколько);
- RT — повтор (1, если данный кадр является повторной передачей предыдущего);

- PM — управление мощностью (1, если передающая станция находится в режиме ожидания);
- MD — больше данных (указывает, что станция передала не все данные, каждый блок данных может передаваться как один кадр или как группа фрагментов в нескольких кадрах);
- W — бит защиты проводного эквивалента, т. е. WEP (*Wired Equivalent Privacy*) (1, если реализован алгоритм конфиденциальности проводного эквивалента, при этом протокол WEP используется для обмена ключами шифрования при безопасном обмене данными);
- O — порядок (1, если используется услуга строгого упорядочения, указывающая адресату, что кадры должны обрабатываться строго по порядку).

Рассмотрим теперь различные типы кадров MAC.

Контрольные кадры

Контрольные кадры способствуют надежной доставке информационных кадров. Существует шесть подтипов контрольных кадров. Перечислим их.

- Опрос после выхода из экономичного режима (PS-опрос).

Данный кадр передается любой станцией станции, включающей точку доступа. В кадре запрашивается передача кадра, прибывшего, когда станция находилась в режиме энергосбережения, и в данный момент размещенного в буфере точки доступа.

- Запрос передачи (RTS). Данный кадр является первым из четверки, используемой для обеспечения надежной передачи данных. Станция, пославшая это сообщение, предупреждает адресата и остальные станции, способные принять данное сообщение, о своей попытке передать адресату информационный кадр.

- «Готов к передаче» (CTS). Второй кадр четырехкадровой схемы. Передается станцией-адресатом станции-источнику и предоставляет право отправки информационного кадра.

- Подтверждение (ACK). Подтверждение успешного приема предыдущих данных, кадра управления или кадра «PS-опрос».

- Без состязания (CF-конец). Объявляет конец периода без состязания; часть стратегии использования распределенного режима доступа.

- CF-конец + CF-подтверждение. Подтверждает кадр «CF-конец». Данный кадр завершает период без состязания и освобождает станции от ограничений, связанных с этим периодом.

Информационные кадры

Существует восемь подтипов информационных кадров, собранных в две группы. Первые четыре подтипа определяют кадры, переносящие данные высших уровней от исходной станции к станции-адресату. Перечислим эти кадры:

- Данные. Просто информационный кадр. Может использоваться как в период состязания, так и в период без состязания.
- Данные + CF-подтверждение. Может передаваться только в период без состязания. Помимо данных, в этом кадре имеется подтверждение полученной ранее информации.
- Данные + CF-опрос. Используется точечным координатором для доставки данных к мобильной станции и для запроса у мобильной станции информационного кадра, который находится в ее буфере.
- Данные + CF-подтверждение + CF-опрос. Объединяет в одном кадре функции двух описанных выше кадров.

Остальные четыре подтипа информационных кадров фактически не переносят данные пользователя. Информационный кадр «нулевая функция» не переносит ни данных, ни запросов, ни подтверждений. Он используется только для передачи точке доступа бита управления питанием в поле управления кадром, указывая, что станция перешла в режим работы с пониженным энергопотреблением. Оставшиеся три кадра (CF-подтверждение, CF-опрос, CF-подтверждение + CF-опрос) имеют те же функции, что и описанные выше подтипы кадров (данные + CF-подтверждение, данные + CF-опрос, данные + CF-подтверждение + CF-опрос), но не несут пользовательских данных.

Кадры управления

Кадры управления используются для управления связью станций и точек доступа.

Возможны следующие подтипы:

- Запрос ассоциации. Посылается станцией к точке доступа с целью запроса ассоциации с данной сетью с базовым набором услуг (*Basic Service Set* — BSS).
- Кадр включает информацию о возможностях, например будет ли использоваться шифрование, или способна ли станция отвечать при опросе.
- Ответ на запрос ассоциации. Возвращается точкой доступа и указывает, что запрос ассоциации принят.
- Запрос повторной ассоциации. Посылается станцией при переходе между BSS, когда требуется установить ассоциацию с точкой доступа в новом BSS. Использование повторной ассоциации, а не просто ассоциации, позволяет новой точке доступа договариваться со старой о передаче информационных кадров по новому адресу.
- Ответ на запрос повторной ассоциации. Возвращается точкой доступа и указывает, что запрос повторной ассоциации принят.
- Пробный запрос. Используется станцией для получения информации от другой станции или точки доступа. Кадр используется для локализации BSS стандарта IEEE 802.11.
- Ответ на пробный запрос. Отклик на пробный запрос.
- Сигнальный кадр. Передается периодически, позволяет мобильным станциям локализовать и идентифицировать BSS.
- Объявление наличия трафика. Посылается мобильной станцией с целью уведомления других (которые могут находиться в режиме пониженного энергопотребления), что в буфере данной станции находятся кадры, адресованные другим.
- Разрыв ассоциации. Используется станцией для аннулирования ассоциации.
- Аутентификация. Для аутентификации станций используются множественные кадры.
- Отмена аутентификации. Передается для прекращения безопасного соединения.

14.4. Стандарты IEEE 802.11

Из большого количества стандартов беспроводной передачи данных на практике получили распространение всего три: 802.11a, 802.11b и 802.11g. Эти стандарты различны по многим параметрам: скорости передачи данных, диапазону частот, методам модуляции сигнала и многим другим характеристикам.

Стандарт IEEE 802.11a имеет большую ширину полосы из семейства стандартов 802.11 при скорости передачи данных до 54 Мбит/с. В отличие от базового стандарта, ориентированного на область частот 2,4 ГГц, спецификациями 802.11a предусмотрена работа в диапазоне 5 ГГц. В качестве метода модуляции сигнала выбрано ортогональное частотное мультиплексирование (OFDM). К недостаткам 802.11a относятся более высокая потребляемая мощность радиопередатчиков для частот 5 ГГц, а также меньший радиус действия.

Поскольку оборудование, работающее на максимальной скорости 11 Мбит/с, имеет меньший радиус действия, чем на более низких скоростях, стандартом 802.11b предусмотрено автоматическое снижение скорости при ухудшении качества сигнала.

Стандарт IEEE 802.11g является логическим развитием 802.11b и предполагает передачу данных в том же частотном диапазоне. Он полностью совместим с 802.11b, т. е. любое устройство 802.11g должно поддерживать работу с устройствами 802.11b. Максимальная скорость передачи в стандарте 802.11g составляет 54 Мбит/с, поэтому на сегодня это наиболее перспективный стандарт беспроводной связи. При разработке стандарта 802.11g рассматривались две отчасти конкурирующие технологии: метод ортогонального частотного разделения OFDM и метод двоичного пакетного сверточного кодирования PBCC, опционально реализованный в стандарте 802.11b. В результате стандарт 802.11g содержит компромиссное решение: в качестве базовых применяются технологии OFDM и CCK, а опционально предусмотрено использование технологии PBCC. О технологиях CCK и OFDM мы рассказали в разделах 8 и 10 настоящего учебного пособия.

Набор стандартов 802.11 определяет целый ряд технологий реализации физического уровня (*Physical Layer Protocol*, PHY), которые могут быть использованы подуровнем 802.11 MAC.

Рассмотрим каждый из уровней PHY:

- Уровень PHY стандарта 802.11 со скачкообразной перестройкой частоты (FHSS) в диапазоне 2,4 ГГц.
- Уровень PHY стандарта 802.11 с расширением спектра методом прямой последовательности (DSSS) в диапазоне 2,4 ГГц.
- Уровень PHY стандарта 802.11b с комплементарным кодированием в диапазоне 2,4 ГГц.
- Уровень PHY стандарта 802.11a с ортогональным частотным мультиплексированием (OFDM) в диапазоне 5 ГГц.
- Расширенный физический уровень (*Extended Rate Physical Layer*, ERP) стандарта 802.11g в диапазоне 2,4 ГГц.

Основное назначение физических уровней стандарта 802.11 — обеспечить механизмы беспроводной передачи для подуровня MAC, а также поддерживать выполнение вторичных функций, таких как оценка состояния беспроводной среды и сообщение о нем подуровню MAC. Уровни MAC и PHY разрабатывались так, чтобы они были независимыми. Именно независимость между MAC и подуровнем PHY и позволила использовать дополнительные высокоскоростные физические уровни, описанные в стандартах 802.11b, 802.11a и 802.11g.

Каждый из физических уровней стандарта 802.11 имеет два подуровня:

- *Physical Layer Convergence Procedure* (PLCP). Процедура определения состояния физического уровня.
- *Physical Medium Dependent* (PMD). Подуровень физического уровня, зависящий от среды передачи.

На рис. 14.7 показано, как эти подуровни соотносятся между собой и с вышестоящими уровнями в модели взаимодействия открытых систем.

Подуровень PLCP по существу является уровнем обеспечения взаимодействия, на котором осуществляется перемещение элементов данных протокола MAC (*MAC Protocol Data Units*, MPDU) между MAC-станциями с использованием подуровня PMD, на котором реализуется тот или иной метод передачи и приема данных через беспроводную среду. Подуровни PLCP и PMD отличаются для разных вариантов стандарта 802.11. Перед тем как приступить к изучению физических уровней, рассмотрим одну из составляющих физического

уровня, до сих пор не упомянутую, а именно — скрэмлирование.

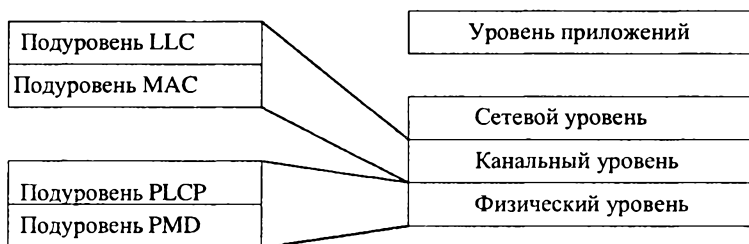


Рис. 14.7. Подуровни уровня РНУ

14.4.1. Методы передачи в IEEE 802.11

Исходный стандарт 802.11 определяет три метода передачи на физическом уровне:

- передача в диапазоне инфракрасных волн;
- технология расширения спектра путем скачкообразной перестройки частоты (FHSS) в диапазоне 2,4 ГГц;
- технология широкополосной модуляции с расширением спектра методом прямой последовательности (DSSS) в диапазоне 2,4 ГГц.

14.4.2. Передача в диапазоне инфракрасных волн

Средой передачи являются инфракрасные волны диапазона 850 нм, которые генерируются либо полупроводниковым лазерным диодом, либо светодиодом (LED — *Light Emitting Diode*). Так как инфракрасные волны не проникают через стены, область покрытия LAN ограничивается зоной прямой видимости. Стандарт предусматривает три варианта распространения излучения: ненаправленную антенну, отражение от потолка и фокусное направленное излучение. В первом случае узкий луч рассеивается с помощью системы линз. Фокусное направленное излучение предназначено для организации двухточечной связи, например между двумя зданиями.

14.4.3. Беспроводные локальные сети со скачкообразной перестройкой частоты (FHSS)

Беспроводные локальные сети FHSS поддерживают скорости передачи 1 и 2 *Мбит/с*. Устройства FHSS делят предназначенную для их работы полосу частот от 2,402 до 2,480 *ГГц* на 79 неперекрывающихся каналов (это справедливо для Северной Америки и большей части Европы). Ширина каждого из 79 каналов составляет 1 *МГц*, поэтому беспроводные локальные сети FHSS используют относительно высокую скорость передачи символов — 1 *МГц* — и намного меньшую скорость перестройки с канала на канал.

Последовательность перестройки частоты должна иметь следующие параметры: частота перескоков не менее 2,5 раз в секунду как минимум между шестью (6 *МГц*) каналами. Чтобы минимизировать число коллизий между перекрывающимися зонами покрытия, возможные последовательности перескоков должны быть разбиты на три набора последовательностей, длина которых для Северной Америки и большей части Европы составляет 26. В табл. 14.2 представлены схемы скачкообразной перестройки частоты, обеспечивающие минимальное перекрытие.

По сути, схема скачкообразной перестройки частоты обеспечивает неторопливый переход с одного возможного канала на другой таким образом, что после каждого скачка покрывается полоса частот, равная как минимум 6 *МГц*, благодаря чему в многосотовых сетях минимизируется возможность возникновения коллизий.

Таблица 14.2

Схема FHSS для Северной Америки и Европы

Набор	Схема скачкообразной перестройки частоты
1	0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75
2	1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58, 61, 64, 67, 70, 73, 76
3	2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, 68, 71, 72, 77

После того как уровень MAC пропускает MAC-фрейм, который в локальных беспроводных сетях FHSS называется также служебным элементом данных PLCP, или PSDU (PLCP *Service Data Unit*), подуровень PLCP добавляет два поля в начало фрейма, чтобы сформировать таким образом фрейм PPDU (PPDU — элемент данных протокола PLCP).

На рис. 14.8 представлен формат фрейма FHSS подуровня PLCP.

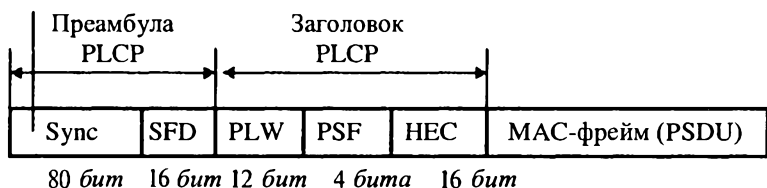


Рис. 14.8. Формат фрейма FHSS подуровня PLCP

Препамбула PLCP состоит из двух подполей.

Подполе *Sync* размером 80 бит. Строка, состоящая из чередующихся 0 и 1, начинается с 0. Приемная станция использует это поле, чтобы принять решение о выборе антенны при наличии такой возможности, откорректировать уход частоты (*frequency offset*) и синхронизировать распределение пакетов (*packet timing*).

Подполе флага начала фрейма (*Start of Frame Delimiter*, SFD) размером 16 бит. Состоит из специфической строки (0000 1100 1011 1101, крайний слева бит первый) в обеспечении синхронизации фреймов (*frame timing*) для приемной станции.

Заголовок фрейма PLCP состоит из трех подполей.

Слово длины служебного элемента данных PLCP (PSDU), PSDU Length Word (PLW) размером 12 бит. Указывает размер фрейма MAC (PSDU) в октетах.

Сигнальное поле PLCP (*Signaling Field PLCP*, PSF) размером 4 бита. Указывает скорость передачи данных конкретного фрейма.

HEC (*Header Error Check*). Контрольная сумма фрейма.

Служебный элемент данных PLCP (PSDU) проходит через операцию скремблирования с целью отбеливания (рандо-

мизации) последовательности входных битов. Получившийся в результате PSDU представлен на рис. 14.9. Заполняющие символы вставляются между всеми 32-символьными блоками. Эти заполняющие символы устраняют любые систематические отклонения в данных, например когда единиц больше, чем нулей, или наоборот, которые могли бы привести к нежелательным эффектам при дальнейшей обработке.



Рис. 14.9. Скрэмблированный PSDU в технологии FHSS

Подуровень PLCP преобразует фрейм в поток битов и передает его на подуровень PMD. Подуровень PMD технологии FHSS модулирует поток данных с использованием модуляции, основанной на гауссовой частотной модуляции (*Gaussian Frequency Shift Keying*, GFSK).

14.4.4. Беспроводные локальные сети, использующие широкополосную модуляцию DSSS с расширением спектра методом прямой последовательности

В спецификации стандарта 802.11 оговорено использование и другого физического уровня — на основе технологии широкополосной модуляции с расширением спектра методом прямой последовательности (DSSS). Как было указано в стандарте 802.11 разработки 1997 года, технология DSSS поддерживает скорости передачи 1 и 2 Мбит/с.

Аналогично подуровню PLCP, используемому в технологии FHSS, подуровень PLCP технологии DSSS стандарта 802.11 добавляет два поля во фрейм MAC, чтобы сформировать PPDU: преамбулу PLCP и заголовок PLCP. Формат фрейма представлен на рис. 14.10.

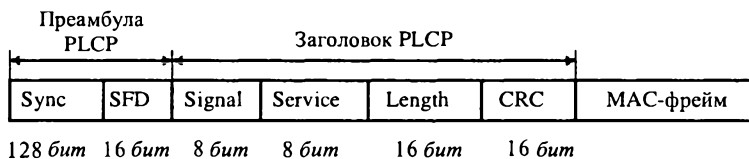


Рис. 14.10. Формат фрейма DSSS подуровня PLCP

Прембула PLCP состоит из двух подполей.

Подполе *Sync* шириной 128 бит, представляющее собой строку, состоящую из единиц. Задача этого подполя — обеспечить синхронизацию для приемной станции.

Подполе *SFD* шириной 16 бит; в нем содержится специфичная строка 0xF3A0; его задача — обеспечить тайминг (*timing*) для приемной станции.

Заголовок PLCP состоит из четырех подполей:

Подполе *Signal* шириной 8 бит, указывающее тип модуляции и скорость передачи для данного фрейма.

Подполе *Service* шириной 8 бит зарезервировано. Это означает, что во время разработки спецификации стандарта оно осталось неопределенным; предполагается, что оно пригодится в будущих модификациях стандарта.

Подполе *Length* шириной 16 бит, указывающее количество микросекунд, необходимое для передачи части MAC-фрейма.

Подполе *CRC* представляет собой 16-битную контрольную сумму.

Подуровень PLCP преобразует фрейм в поток битов и передает данные на подуровень PMD. Весь PPDU проходит через процесс скремблирования с целью рандомизации данных.

Скремблированная преамбула PLCP всегда передается со скоростью 1 Мбит/с, в то время как скремблированный фрейм MPDU передается со скоростью, указанной в подполе *Signal*. Подуровень PMD модулирует отбеленный поток битов, используя следующие методы модуляции:

- двоичная относительная фазовая модуляция (*Differential Binary Phase Shift Keying, DBPSK*) для скорости передачи 1 Мбит/с;
- квадратурная относительная фазовая модуляция (*Differential Quadrature Phase Shift Key, DQPSK*) для скорости передачи 2 Мбит/с.

14.4.5. IEEE 802.11b

На физическом уровне к MAC-кадрам (MPDU) добавляется заголовок физического уровня, состоящий из преамбулы и собственно PLCP-заголовка (рис. 14.11).

Преамбула содержит стартовую синхропоследовательность (SYNC) для настройки приемника и 16-битный код начала кадра (SFD) — число F3A016. PLCP-заголовок включает поля SIGNAL (информация о скорости и типе модуляции), SERVICE (дополнительная информация, в том числе о применении высокоскоростных расширений и PBSS-модуляции) и LENGTH (время в микросекундах, необходимое для передачи следующей за заголовком части кадра). Все три поля заголовка защищены 16-битной контрольной суммой CRC.

В стандарте IEEE 802.11b предусмотрено два типа заголовков: длинный и короткий (рис. 14.12). Они отличаются

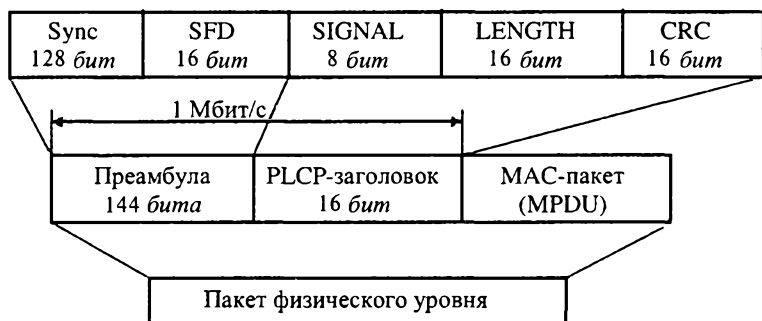


Рис. 14.11. Структура кадров сети IEEE 802.11b физического уровня

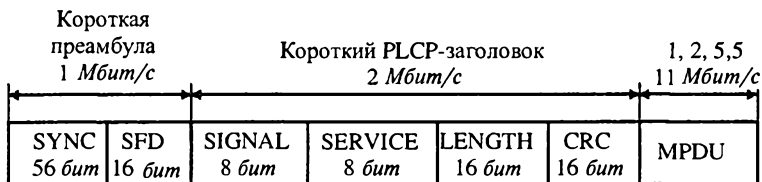


Рис. 14.12. Короткий заголовок кадров сети 802.11b

длиной синхропоследовательности (128 и 56 бит), способом ее генерации, а также тем, что символ начала кадра в коротком заголовке передается в обратном порядке. Кроме того, если все поля длинного заголовка передаются со скоростью 1 Мбит/с, то при коротком заголовке преамбула транслируется на скорости 1 Мбит/с, другие поля заголовка — со скоростью 2 Мбит/с. Остальную часть кадра можно передавать на любой из допустимых стандартом скоростей передачи, указанных в полях SIGNAL и SERVICE. Короткие заголовки физического уровня предусмотрены спецификацией IEEE 802.11b для увеличения пропускной способности сети. Из описания процедур связи сети IEEE 802.11 видно, что «накладные расходы» в этом стандарте выше, чем в проводной сети Ethernet. Поэтому крайне важно обеспечить высокую скорость передачи данных в канале.

Повысить пропускную способность канала с заданной шириной полосы частот можно, разрабатывая и применяя новые методы модуляции. По этому пути пошла группа разработчиков IEEE 802.11b.

Изначально стандарт IEEE 802.11 предусматривал работу в режиме DSSS с использованием так называемой Баркеровской (*Barker*) последовательности (здесь мы отсылаем читателя к разделу 8 настоящего пособия, там же мы говорим и о СКК-модуляции) длиной 11 бит: $B1 = (10110111000)$. Каждый информационный бит замещается своим произведением по модулю 2 (операция «исключающее ИЛИ») с данной последовательностью, т. е. каждая информационная единица заменяется на $B1$, каждый ноль — на инверсию $B1$. В результате бит заменяется последовательностью 11 чипов. Далее сигнал кодируется посредством дифференциальной двух- или четырехпозиционной фазовой модуляции (DBPSK или DQPSK, один или два чипа на символ соответственно). При частоте модуляции несущей 11 МГц общая скорость составляет в зависимости от типа модуляции 1 и 2 Мбит/с. Стандарт IEEE 802.11b дополнительно предусматривает скорости передачи 11 и 5,5 Мбит/с. Для этого используется так называемая СКК-модуляция (*Complementary Code Keying* — кодирование элементарным кодом). Механизм расширения спектра, используемый для получения скоростей 5,5 и 11 Мбит/с с применением СКК, относится к методам, которые применяются для

скоростей 1 и 2 Мбит/с. В обоих случаях применяется метод расширения, но при использовании модуляции ССК расширяющий код представляет собой код из 8 комплексных чипов, в то время как при работе со скоростями 1 и 2 Мбит/с применяется 11-разрядный код. В зависимости от скорости передачи данных 8-чиповый код определяется или 4, или 8 битами. В результате чего можно добиться скорости передачи данных 5,5 и 11 Мбит/с.

Для передачи данных со скоростью 5,5 Мбит/с нужно сгруппировать скремблированный поток битов в символы по 4 бита (b_0, b_1, b_2 и b_3). Последние два бита (b_2 и b_3) используются для определения 8 последовательностей комплексных чипов, как показано в таблице 14.3, где $\{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8\}$ представляют чипы последовательности ($j = \sqrt{-1}$).

Таблица 14.3

Последовательность чипов ССК

(b_2, b_3)	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
00	j	1	j	-1	j	1	-1	1
01	$-j$	-1	$-j$	1	j	1	$-j$	1
10	$-j$	1	$-j$	-1	$-j$	1	j	1
11	j	-1	j	1	$-j$	1	j	1

Имея последовательность чипов, определенную битами (b_2, b_3), можно использовать первые два бита (b_0, b_1) для определения поворота фазы, осуществляемого при модуляции по методу DQPSK, который будет применен к последовательности (табл. 14.4). Нужно также пронумеровать каждый 4-битовый символ PSDU, начиная с 0, чтобы можно было определить, преобразуется ли четный либо нечетный символ в соответствии с этой таблицей. Здесь речь идет об использовании DQPSK, а не QPSK, и поэтому представленные в таблице изменения фазы отсчитываются по отношению к предыдущему символу или, в случае первого символа PSDU, по отношению к последнему символу предыдущего DQPSK-символа, передаваемого со скоростью 2 Мбит/с.

Поворот фазы при модуляции ССК

b_0, b_1	Изменение фазы четных символов	Изменение фазы нечетных символов
00	0	φ
01	$\varphi/2$	$-\varphi/2$
11	φ	0
10	$-\varphi/2$	$\varphi/2$

Это вращение фазы применяется по отношению к 8 комплексным чипам символа, затем осуществляется модуляция на подходящей несущей частоте.

Чтобы передавать данные со скоростью 11 Мбит/с, скремблированная последовательность битов PSDU разбивается на группы по 8 символов. Последние 6 битов выбирают одну последовательность, состоящую из 8 комплексных чипов, из числа 64 возможных последовательностей, почти так же, как использовались биты (b_2, b_3) для выбора одной из четырех возможных последовательностей. Биты (b_0, b_1) используются таким же образом, как при модуляции ССК на скорости 5,5 Мбит/с для вращения фазы последовательности и дальнейшей модуляции на подходящей несущей частоте.

Так в чем же достоинство ССК-модуляции? Нетрудно заметить, что теоретическое операционное усиление ССК-модуляции — 3 дБ (в два раза), поскольку без кодирования QPSK-модулированный с частотой 11 Мбит/с сигнал переносит информацию со скоростью 22 Мбит/с. Как видно, ССК-модуляция представляет собой вид блочного кодирования, а потому достаточно проста при аппаратной реализации. Совокупность этих свойств и обеспечила ССК место в стандарте IEEE 802.11b в качестве обязательного вида модуляции.

Существенную роль играет равномерность распределения символов в фазовом пространстве и они должны как можно дальше отстоять друг от друга, чтобы минимизировать ошибки их детектирования. Но с этой точки зрения ССК-модуляция не выглядит оптимальной. Поэтому изначально прорабатывался другой способ модуляции — пакетное бинарное сверточное кодирование (см. раздел 8 настоящего пособия). Этот метод вошел в стандарт IEEE 802.11b как дополнительная (нео-

бязательная) опция. Механизм RBCC (рис. 14.13) позволяет добиваться в сетях 802.11b пропускной способности 5,5, 11 и 22 Мбит/с.

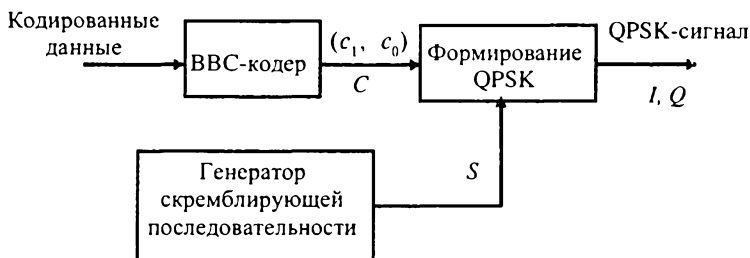


Рис. 14.13. Общая схема RBCC-модуляции

Как следует из названия, метод основан на сверточном кодировании. Для скоростей 5,5 и 11 Мбит/с поток информационных битов поступает в шестиразрядный сдвиговый регистр с сумматорами (рис. 14.14). В начальный момент времени все триггеры сдвигового регистра инициализируют нулем. В результате каждый исходный бит d заменяется двумя битами кодовой последовательности (c_0 , c_1). При скорости 11 Мбит/с c_0 и c_1 задают один символ четырехпозиционной QPSK-модуляции. Для скорости 5,5 Мбит/с используют двухпозиционную BPSK-модуляцию, последовательно передавая кодовые биты c_0 и c_1 . Если же нужна скорость 22 Мбит/с, схема кодирования усложняется (рис. 14.15): три кодовых бита (c_0 – c_2) определяют один символ в 8-позиционной 8-PSK-модуляции.

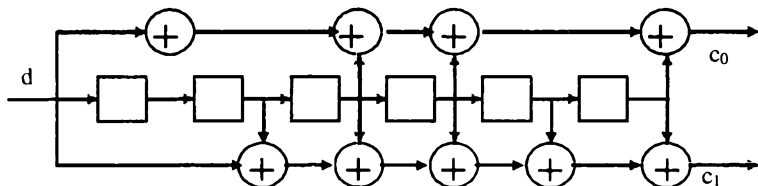


Рис. 14.14. Сверточное кодирование с двумя битами кодовой последовательности

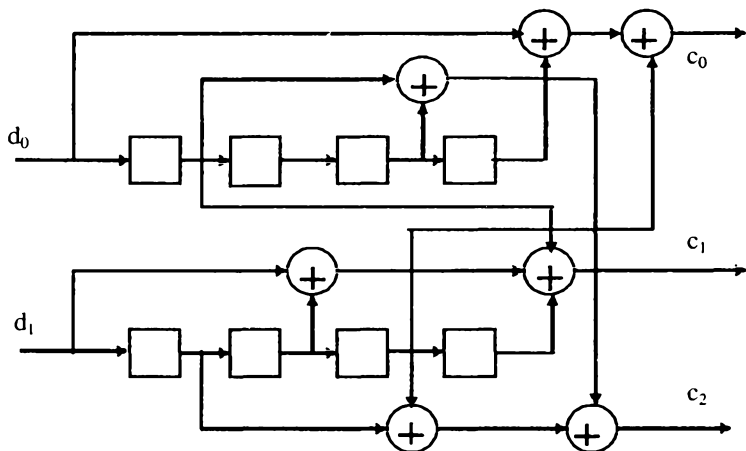


Рис. 14.15. Сверточное кодирование с тремя битами кодовой последовательности

14.4.6. IEEE 802.11a

Стандарт IEEE 802.11a появился практически одновременно с IEEE 802.11b, в сентябре 1999 года. Эта спецификация была ориентирована на работу в диапазоне 5 ГГц и основана на принципиально ином, чем описано выше, механизме кодирования данных — на частотном мультиплексировании посредством ортогональных несущих (OFDM).

Стандарт IEEE 802.11a определяет характеристики оборудования, применяемого в офисных или городских условиях, когда распространение сигнала происходит по многолучевым каналам из-за множества отражений. В IEEE 802.11a каждый кадр передается посредством 52 ортогональных несущих, каждая с шириной полосы порядка 300 кГц (20 МГц/64). Ширина одного канала — 20 МГц. Несущие модулируют посредством BPSK, QPSK, а также 16- и 64-позиционной квадратурной амплитудной модуляции (QAM). В совокупности с различными скоростями кодирования r (1/2, для 64-QAM — 2/3) образуется набор скоростей передачи 6, 9, 12, 18, 24, 36, 48 и 54 Мбит/с. В табл. 14.5 показано, как необходимая

скорость передачи данных преобразуется в соответствующие параметры узлов передатчика OFDM.

Таблица 14.5

Параметры передатчика стандарта 802.11a

Скорость передачи данных (Мбит/с)	Модуляция	Скорость сверточного кодирования	Число канальных битов на поднесущую	Число канальных битов на символ	Число битов данных на символ OFDM
6	BPSK	1/2	1	48	24
9	BPSK	3/4	1	48	36
12	QPSK		2	96	48
18	QPSK	3/4	2	96	72
24	16-QAM	1/2	4	192	96
36	16-QAM	3/4	4	192	144
48	64-QAM	2/3	6	288	192
54	64-QAM	3/4	6	288	216

Из 52 несущих 48 предназначены для передачи информационных символов, остальные 4 — служебные. Структура заголовков физического уровня отличается от принятого в спецификации IEEE 802.11b, но незначительно (рис. 14.16).

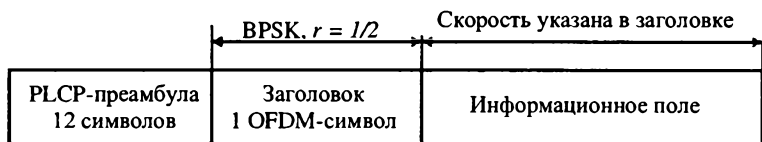


Рис. 14.16. Структура заголовка физического уровня стандарта IEEE 802.11a

Кадр включает преамбулу (12 символов синхропоследовательности), заголовок физического уровня (PLCP-заголовок) и собственно информационное поле, сформированное на MAC-уровне. В заголовке передается информация о скорости кодирования, типе модуляции и длине кадра. Преамбула и заголовок транслируются с минимально возможной скоростью (BPSK, скорость кодирования $r = 1/2$), а информационное

поле — с указанной в заголовке, как правило максимальной, скоростью, в зависимости от условий обмена. OFDM-символы передаются через каждые 4 мкс, причем каждому символу длительностью 3,2 мкс предшествует защитный интервал 0,8 мкс (повторяющаяся часть символа). Последний необходим для борьбы с многолучевым распространением сигнала — отраженный и пришедший с задержкой символ попадет в защитный интервал и не повредит следующий символ.

Естественно, формирование/декодирование OFDM-символов происходит посредством быстрого преобразования Фурье (обратного/прямого, ОБПФ/БПФ). Функциональная схема трактов приема/передачи (рис. 14.17) достаточно стандартна для данного метода и включает сверточный кодер, механизм перемежения/перераспределения (защита от пакетных ошибок) и процессор ОБПФ.

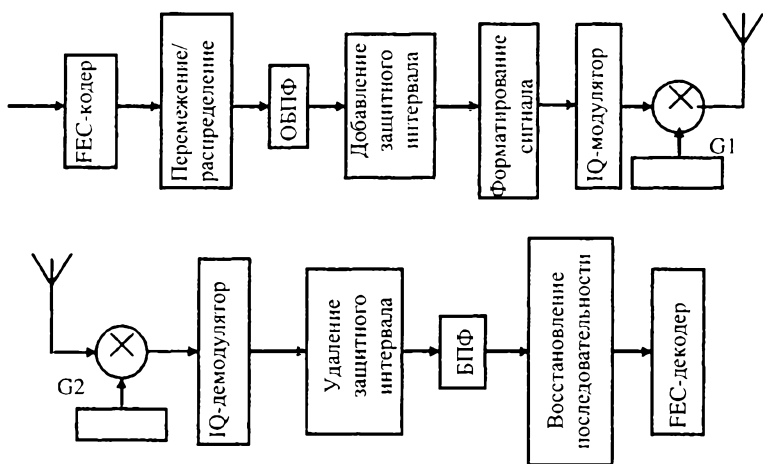


Рис. 14.17. Функциональная схема трактов приема/передачи стандарта IEEE 802.11a

14.4.7. IEEE 802.11g

Стандарт IEEE 802.11g по сути представляет собой перенесение схемы модуляции OFDM, прекрасно зарекомендовавшей себя в 802.11a, из диапазона 5 ГГц в область 2,4 ГГц при сохранении функциональности устройств стандарта 802.11b.

Это возможно, поскольку в стандартах 802.11 ширина одного канала в диапазонах 2,4 и 5 ГГц схожа — 22 МГц.

Одним из основных требований к спецификации 802.11g была обратная совместимость с устройствами 802.11b. Действительно, в стандарте 802.11b в качестве основного способа модуляции принята схема ССК (*Complementary Code Keying*), а в качестве дополнительной возможности допускается модуляция PBSS.

Разработчики 802.11g предусмотрели ССК-модуляцию для скоростей до 11 Мбит/с и OFDM для более высоких скоростей. Но сети стандарта 802.11 при работе используют принцип CSMA/CA — множественный доступ к каналу связи с контролем несущей и предотвращением коллизий. Ни одно устройство 802.11 не должно начинать передачу, пока не убедится, что эфир в его диапазоне свободен от других устройств. Если в зоне слышимости окажутся устройства 802.11b и 802.11g, причем обмен будет происходить между устройствами 802.11g посредством OFDM, то оборудование 802.11b просто не поймет, что другие устройства сети ведут передачу, и попытается начать передачу. Чтобы не допустить подобной ситуации, предусмотрена возможность работы в смешанном режиме — ССК-OFDM. Информация в сетях 802.11 передается кадрами. Каждый информационный кадр включает два основных поля: преамбулу с заголовком и информационное поле (рис. 14.18).

Преамбула/ заголовок	Информационное поле	
OFDM	OFDM	Обязательно
ССК	ССК	
ССК	OFDM	Возможно
ССК	PBSS	

Рис. 14.18. Кадры IEEE 802.11g в различных режимах модуляции

Преамбула содержит синхропоследовательность и код начала кадра, заголовок — служебную информацию, в том числе о типе модуляции, скорости и продолжительности передачи кадра. В режиме CCK-OFDM преамбула и заголовок модулируются методом CCK (реально — путем прямого расширения спектра DSSS посредством последовательности Баркера, поэтому в стандарте 802.11g этот режим именуется DSSS-OFDM), а информационное поле — методом OFDM. Таким образом, все устройства 802.11b, постоянно «прослушивающие» эфир, принимают заголовки кадров и узнают, сколько времени будет транслироваться кадр 802.11g. В этот период они «молчат». Естественно, пропускная способность сети падает, поскольку скорость передачи преамбулы и заголовка — 1 Мбит/с.

Видимо, данный подход не устраивал лагерь сторонников технологии PBSS, и для достижения компромисса в стандарт 802.11g в качестве дополнительной возможности ввели, так же как и в 802.11b, необязательный режим — PBSS, в котором заголовок и преамбула передаются так же, как и при CCK, а информационное поле модулируется по схеме PBSS и передается на скорости 22 или 33 Мбит/с. В результате устройства стандарта 802.11g должны оказаться совместимыми со всеми модификациями оборудования 802.11b и не создавать взаимных помех. Диапазон поддерживаемых им скоростей отражен в табл. 14.6, зависимость скорости от типа модуляции — на рис. 14.19.

Очевидно, что устройствам стандарта IEEE 802.11g достаточно долго придется работать в одних сетях с оборудованием 802.11b. Также очевидно, что производители в массе своей не будут поддерживать режимы CCK-OFDM и PBSS в силу их необязательности, ведь почти все решает цена устройства. Поэтому одна из основных проблем данного стандарта — как обеспечить бесконфликтную работу смешанных сетей 802.11b/g.

Основной принцип работы в сетях 802.11 — «слушать, прежде чем вещать». Но устройства 802.11b не способны услышать устройства 802.11g в OFDM-режиме. Ситуация аналогична проблеме скрытых станций: два устройства удалены настолько, что не слышат друг друга и пытаются обратиться к третьему, которое находится в зоне слышимости обоих. Для предотвращения конфликтов в подобной ситуации в 802.11 вве-

**Возможные скорости и тип модуляции
в спецификации IEEE 802.11g**

Скорость, Мбит/с	Тип модуляции	
	Обязательно	Допустимо
1	Последовательность Баркера	
2	Последовательность Баркера	
5,5	ССК	PBCC
6	OFDM	OFDM
9		OFDM, ССК-OFDM
11	ССК	PBCC
12	OFDM	ССК-OFDM
18		OFDM, ССК-OFDM
22		PBCC
24	OFDM	ССК-OFDM
33		PBCC
36		OFDM, ССК-OFDM
48		OFDM, ССК-OFDM
54		OFDM, ССК-OFDM

ден защитный механизм, предусматривающий перед началом информационного обмена передачу короткого кадра «запрос на передачу» (RTS) и получение кадра подтверждения «можно передавать» (CTS). Механизм RTS/CTS применим и к смешанным сетям 802.11b/g. Естественно, эти кадры должны транслироваться в режиме ССК, который обязаны понимать все устройства. Однако защитный механизм существенно снижает пропускную способность сети.

В табл. 14.7 представлена сводная информация по параметрам физических уровней.

Таблица 14.7

Стандарты физического уровня

Параметр	802.11 DSSS	802.11 FHSS	802.11b	802.11a	802.11g
Частотный диапазон (ГГц)	2,4	2,4	2,4	5	2,4

Параметр	802.11 DSSS	802.11 FHSS	802.11b	802.11a	802.11g
Максимальная скорость передачи данных (Мбит/с)	2	2	11	54	54
Технология	DSSS	FHSS	CCK	OFDM	OFDM
Тип модуляции (для максимальной скорости передачи)	QPSK	GFSK	QPSK	64-QAM	64-QAM
Число перекрывающихся каналов	3	3	3	15	3

14.5. Беспроводная технология WiMAX

WiMAX — очень перспективное направление в развитии беспроводных технологий. Характеристики технологии WiMAX во многом превосходят стандарт IEEE 802.11. Достаточно интересно будет ознакомиться с архитектурой работы системы WiMAX. Несомненно, существует несколько режимов работы стандарта IEEE 802.16, подробное описание которых можно найти в дополнительной литературе. Как ни странно, у новой технологии есть несколько сдерживающих факторов, ограничивающих ее быстрое распространение.

14.5.1. Цели и задачи WiMAX

При всем богатстве выбора сетевых подключений сложно одновременно соблюсти три основных требования к сетевым соединениям: реализовать высокую пропускную способность, надежность и мобильность. Решить подобную задачу может следующее поколение беспроводных технологий WiMAX (*Worldwide Interoperability for Microwave Access*), стандарт IEEE 802.16.

Для продвижения и развития технологии WiMAX был сформирован WiMAX-форум на базе рабочей группы IEEE

802.16, созданной в 1999 году. В форум вошли такие фирмы, как *Nokia, Harris Corporation, Ensemble, Crossspan и Aperto*. К маю 2005 года форум объединял уже более 230 участников. В том же году Всемирный съезд по вопросам информационного сообщества (*World Summit on Information Society, WSIS*) сформулировал следующие задачи, которые были возложены на технологию WiMAX:

1) Обеспечить при помощи WiMAX доступ к услугам информационных и коммуникационных технологий для небольших поселений, удаленных регионов, изолированных объектов, учитывая при этом, что в развивающихся странах 1,5 миллиона поселений с числом жителей более 100 человек не подключены к телефонным сетям и не имеют кабельного сообщения с крупными городами.

2) Обеспечить при помощи WiMAX доступ к услугам информационных и коммуникационных технологий более половины населения планеты в пределах досягаемости, учитывая при этом, что общее число пользователей *Internet* в 2005 году составляло приблизительно 960 млн человек, или около 14,5% всего населения Земли.

Цель технологии WiMAX заключается в том, чтобы предоставить универсальный беспроводной доступ для широкого спектра устройств (рабочих станций, бытовой техники «умного дома», портативных устройств и мобильных телефонов) и их логического объединения — локальных сетей. Надо отметить, что данная технология имеет ряд преимуществ:

1. По сравнению с проводными (xDSL или широкополосным), беспроводными или спутниковыми системами сети WiMAX должны позволить операторам и сервис-провайдерам экономически эффективно охватить не только новых потенциальных пользователей, но и расширить спектр информационных и коммуникационных технологий для пользователей, уже имеющих фиксированный (стационарный) доступ.

2. Стандарт объединяет технологии уровня оператора связи (для объединения многих подсетей и предоставления им доступа к *Internet*), а также технологии «последней мили» (конечного отрезка от точки входа в сеть провайдера до компьютера пользователя), что создает универсальность и, как следствие, повышает надежность системы.

3. Беспроводные технологии более гибки и, как следствие, проще в развертывании, так как по мере необходимости могут масштабироваться.

4. Простота установки как фактор уменьшения затрат на развертывание сетей в развивающихся странах, малонаселенных или удаленных районах.

5. Дальность охвата является существенным показателем системы радиосвязи. На данный момент большинство беспроводных технологий широкополосной передачи данных требуют наличия прямой видимости между объектами сети. WiMAX благодаря использованию технологии OFDM создает зоны покрытия в условиях отсутствия прямой видимости от клиентского оборудования до базовой станции, при этом расстояния исчисляются километрами.

6. Технология WiMAX изначально содержит протокол IP, что позволяет легко и прозрачно интегрировать ее в локальные сети.

7. Технология WiMAX подходит для фиксированных, перемещаемых и подвижных объектов сетей на единой инфраструктуре.

14.5.2. Принципы работы

Система WiMAX состоит из двух основных частей:

1. Базовая станция WiMAX, которая может размещаться на высотном объекте — здании или вышке.

2. Приемник WiMAX: антенна с приемником.

Соединение между базовой станцией и клиентским приемником производится в СВЧ диапазоне 2–11 ГГц. Данное соединение в идеальных условиях позволяет передавать данные со скоростью до 20 Мбит/с и не требует, чтобы станция находилась на расстоянии прямой видимости от пользователя. Этот режим работы базовой станции WiMAX близок широко используемому стандарту 802.11 (Wi-Fi), что допускает совместимость уже выпущенных клиентских устройств и WiMAX.

Следует помнить, что технология WiMAX применяется как на «последней миле» — конечном участке между провайдером и пользователем, так и для предоставления доступа региональным сетям: офисным и районным.

Между соседними базовыми станциями устанавливается постоянное соединение с использованием сверхвысокой час-

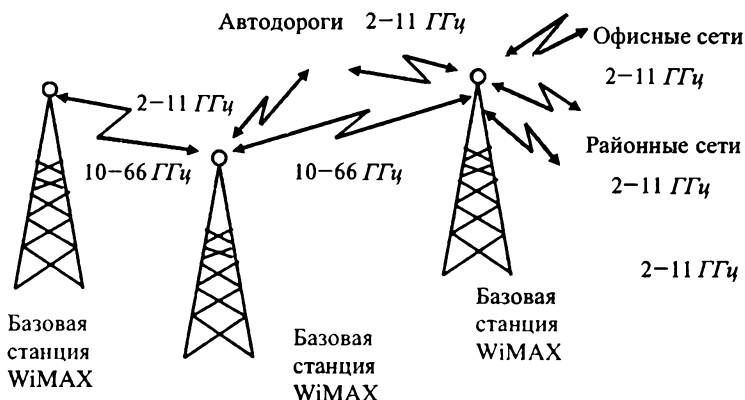


Рис. 14.19. Архитектура WiMAX

тоты 10–66 ГГц радиосвязи прямой видимости. Данное соединение в идеальных условиях позволяет передавать данные со скоростью до 120 Мбит/с. Ограничение по условию прямой видимости, разумеется, не является преимуществом, однако оно накладывается только на базовые станции, участвующие в цельном покрытии района, что вполне возможно реализовать при размещении оборудования.

Как минимум одна из базовых станций может быть постоянно связана с сетью провайдера через широкополосное скоростное соединение. Фактически, чем больше станций имеют доступ к сети провайдера, тем выше скорость и надежность передачи данных. Однако даже при небольшом количестве точек система способна корректно распределить нагрузку за счет сотовой топологии.

На базе сотового принципа разрабатываются также пути построения оптимальной сети, огибающей крупные объекты (например, горные массивы), когда серия последовательных станций передает данные по эстафетному принципу. Подобные разработки планируется включить в следующую версию стандарта. Ожидается, что эти изменения позволят существенно поднять скорость.

По структуре сети стандарта IEEE 802.16 очень похожи на традиционные сети мобильной связи: здесь тоже имеются базовые станции, которые действуют в радиусе до 50 км, при

этом их также не обязательно устанавливать на вышках. Для них вполне подходят крыши домов, требуется лишь соблюдение условия прямой видимости между станциями. Для соединения базовой станции с пользователем необходимо наличие абонентского оборудования. Далее сигнал может поступать по стандартному *Ethernet*-кабелю как непосредственно на конкретный компьютер, так и на точку доступа стандарта 802.11 Wi-Fi или в локальную проводную сеть стандарта *Ethernet*.

Это позволяет сохранить существующую инфраструктуру районных или офисных локальных сетей при переходе с кабельного доступа на WiMAX. Кроме того, это дает возможность максимально упростить развертывание сетей, используя знакомые технологии для подключения компьютеров.

14.5.3. Режимы работы

Стандарт 802.16e-2005 вобрал в себя все ранее выходившие версии и на данный момент предоставляет следующие режимы:

- Fixed WiMAX — фиксированный доступ.
- Nomadic WiMAX — сеансовый доступ.
- Portable WiMAX — доступ в режиме перемещения.

Стандарт WiMAX в настоящее время находится на стадии тестирования. Единственная конкурентоспособная версия стандарта, для которой существует лицензия на оборудование, — это Fixed WiMAX. Однако провайдеры не спешат заменять дорогостоящее, но уже работающее оборудование новым, ибо это требует существенных вложений без возможности поднять производительность (и, соответственно, цену на услуги) и быстро вернуть вложенные средства.

При развертывании WiMAX-сетей там, где доступа к *Internet* раньше не было, приходится сталкиваться с проблемой наличия в малонаселенных или удаленных регионах достаточного числа потенциальных пользователей, обладающих необходимым оборудованием или денежными средствами на его приобретение. Это касается и перехода на Mobile WiMAX после его лицензирования, так как, помимо затрат провайдеров на модернизацию операторского оборудования, следует учитывать затраты пользователей на модернизацию клиентского оборудования: приобретение WiMAX-карт и обновление портативных устройств.

Вторым сдерживающим фактором является позиция многих специалистов, которые считают недопустимым использование сверхвысоких частот радиосвязи прямой видимости, вредных для здоровья человека. Наличие вышек на расстоянии десятков метров от жилых объектов (а базовые станции рекомендуется устанавливать на крышах домов) может пагубно сказаться на здоровье жителей, особенно детей. Однако результатов медицинских экспериментов, подтверждающих наличие или высокую вероятность вреда, пока не опубликовано.

Третьим фактором является, как ни странно, быстрое развитие стандарта. Появление новых, принципиально различных версий стандарта WiMAX приводит к вопросу о неизбежной смене оборудования через несколько лет. Так, станции, сейчас работающие в режиме Fixed WiMAX, не смогут поддерживать Mobile WiMAX. При переходе на следующий стандарт потребуются обновление части оборудования, что отпугивает крупных провайдеров. На данный момент внедрение и использование Fixed WiMAX на коммерческой основе могут позволить себе только небольшие компании, которые не планируют значительного расширения (в том числе территориального) и используют новые технологии для привлечения клиентов.

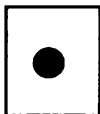
И, наконец, четвертым фактором является наличие конкурентного стандарта широкополосной связи, использующего близкие диапазоны радиочастот — WBro. Этот стандарт тоже до конца не лицензирован, однако он уже получил определенную известность. А потому всегда существует вероятность, что через несколько лет предпочтительным окажется не WiMAX, а WBro. И компании, вложившие средства в разработку и внедрение WiMAX-систем, серьезно пострадают. Впрочем, из-за схожести стандартов существует также вероятность слияния и в дальнейшем использования оборудования, поддерживающего оба стандарта одновременно.

Таким образом, при видимых преимуществах стандарта еще рано говорить о тотальном внедрении технологии или даже о возможности перехода на нее и отказа от существующих сетевых решений. Необходимо сначала получить первое лицензированное оборудование стандарта Mobile WiMAX, а также результаты полевых испытаний. Затем можно ожидать утверж-

дения стандартов версии 802.16f (Full Mobile WiMAX) и 802.16m.

Первый из них включает в себя алгоритмы обхода препятствий и оптимизацию сотовой топологии покрытия между базовыми станциями. Второй стандарт должен поднять скорость передачи данных со стационарным клиентским оборудованием до 1 Гбит/с и с мобильным клиентским оборудованием — до 100 Мбит/с . Эти стандарты планируется утвердить в 2009–2010 годах соответственно.

Далее можно ожидать лицензирования оборудования с поддержкой новых стандартов, конкуренции на рынке производства оборудования и услуг доступа через WiMAX. И только тогда можно будет говорить о действительных преимуществах и недостатках этой технологии по сравнению с ныне существующими.



В книге рассмотрены преимущественно теоретические вопросы, объясняющие основные идеи и тенденции в современном состоянии радиоэлектроники, период развития которой автором разделен на два этапа: классический и современный. Автору известны и другие подходы к классификации периодов развития радиоэлектроники. Но главную задачу этой книги автор представляет в виде описания фундаментальных особенностей состояния современной радиоэлектроники, чем и обусловлена приведенная автором классификация.

В дополнение к изложенному материалу автором разработан практикум по курсу «Основы современной радиоэлектроники», выполнение которого предполагает использование математического пакета MATLAB и его многочисленных приложений для изучения следующих вопросов:

1. Введение в систему MATLAB и *Simulink*.
2. Сигналы с амплитудной модуляцией.
3. Сигналы с амплитудной манипуляцией.
4. Сигналы с угловой модуляцией.
5. Сигналы с частотной и фазовой манипуляцией.
6. Сигналы с квадратурной модуляцией.
7. Сигналы с квадратурной манипуляцией.
8. Моделирование работы модемного протокола V.32 средствами пакета MATLAB.
9. Моделирование созвездий сигналов средствами пакета MATLAB.
10. Моделирование работы модемного протокола V.32bis средствами пакета MATLAB.
11. Моделирование передающей части системы связи с OFDM средствами пакета MATLAB.

12. Моделирование приемной части системы связи с OFDM средствами пакета MATLAB.

13. Моделирование сетевых возможностей средствами пакета MATLAB (TCP/IP, UDP).

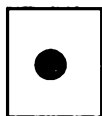
14. Моделирование работы защищенной связи между двумя сетевыми процессами в среде MATLAB.

15. Моделирование протокола SMTP средствами MATLAB.

Изучая эти вопросы, студенты более глубоко познакомятся с практически неисчерпаемыми возможностями пакетов MATHCAD и MATLAB, важность которых трудно переоценить при подготовке современного радиоинженера.

Автор использовал огромное количество литературных источников, список которых занял бы не один десяток страниц.

Насколько успешным оказался замысел автора — судить читателям. Все замечания им будут приняты с признательностью и искренней благодарностью.



ЛИТЕРАТУРА

Беспроводные сети Wi-Fi: учебное пособие / А.В. Пролетарский и др. — М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2007. — 215 с.

Волков Л.Н. Системы цифровой радиосвязи: базовые методы и характеристики: учебное пособие / Л.Н. Волков, М.С. Немировский, Ю.С.Шинаков. — М.: Эко-Трендз, 2005. — 392 с.

Гольштейн Б.С. Протоколы сети доступа T2 / Б.С.Гольштейн — 3-е изд. — СПб.: БХВ-Санкт-Петербург, 2005. — 288 с.

Голяницкий И.А. Математические модели и методы в радиосвязи / И.А. Голяницкий, под. ред. Ю.А. Громакова. — М.: Эко-Трендз, 2005. — 440 с.

Динамика радиоэлектроники / Под общ. ред. Ю.И. Борисова. — М.: Техносфера, 2007. — 400 с.

Дингес С.И. Мобильная связь: технология DECT. — М.: СОЛОН-Пресс, 2003. — 272 с.

Золотарев В.В. Помехоустойчивое кодирование: методы и алгоритмы: справочник / В.В. Золотарев, Г.В. Овечкин. — М.: Горячая линия-Телеком, 2004. — 126 с.

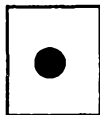
Королев А.И. Коды и устройства помехоустойчивого кодирования информации. — Мн.: Бестпринт, 2002. — 286 с.

Першин В.Т. Основы радиоэлектроники и схемотехники: учебное пособие для студентов вузов. — Ростов н/Д: Феникс, 2006. — 544 с.

Попов В.И. Основы сотовой связи стандарта GSM / В.И. Попов. — М.: Эко-Трендз, 2005. — 296 с.

Рихтер С.Г. Цифровое радиовещание. — М.: Горячая линия-Телеком, 2004. — 352 с.

Тихвинский В.О. Управление и качество услуг в сетях GPRS/UMTS: учебное пособие [доп. УМО РФ] / В.О. Тихвинский, С.В. Терентьев. — М.: Эко-Трендз, 2007. — 400 с.



СОДЕРЖАНИЕ

Предисловие	3
Классификация периодов развития радиоэлектроники	4
Развитие радиоэлектроники послевоенного времени	5
Основы помехоустойчивого кодирования	6
Новые подходы к созданию современной теории информации	7
Проблема сжатия информации	8
Современные стандарты сжатия информации	9
Кодирование речевых и аудиосигналов	9
Построение реляционных баз данных	11
Завершение разработки теории преобразований Фурье	13
Реализация методов помехоустойчивого кодирования	13
Цифровое радиовещание и телевидение	14
Европейский стандарт EN 300 744 наземного цифрового телевидения	16
1. ПОНЯТИЯ ЭНТРОПИИ И ИНФОРМАЦИИ	18
1.1. Понятие информации	18
1.2. Информация и энтропия	20
1.3. Понятие сигнала	26
1.4. Свойства энтропии	27
1.5. Условная энтропия	28
1.6. Энтропия объединения нескольких источников	31
1.7. Понятие дифференциальной энтропии	32
1.8. Взаимная информация между двумя непрерывными случайными величинами	35
1.9. Е-энтропия и е-производительность источника непрерывных сообщений	36
1.10. Связь между энтропией и информацией	39
1.11. Взаимная энтропия и информация	42
1.12. Условная взаимная информация	45
2. СИГНАЛЫ И ИХ АНАЛИЗ	50
2.1. Классификация сигналов	50
2.2. Детерминированные сигналы	53
2.3. Случайные сигналы	62
2.4. Квантование сигнала по уровню	65
2.5. Спектральный (гармонический) анализ сигналов	66
2.6. Дискретизация сигналов	81
2.7. Скорость передачи информации по каналу связи и скорость модуляции	85
2.8. Семейство преобразований Фурье	88

2.10. Дискретизация сигналов с непрерывным временем	102
2.11. ДПФ и дискретное косинусное преобразование	109
3. ИСТОЧНИКИ ИНФОРМАЦИИ И КАНАЛЫ СВЯЗИ	111
3.1. Источники сообщений без памяти и с памятью	111
3.2. Производительность источника дискретных сообщений. Скорость передачи информации	113
3.3. Коэффициент сжатия	114
3.4. Понятие двоичного симметричного канала связи	115
3.5. Пропускная способность дискретного канала	118
3.6. Однородный симметричный канал со стиранием	123
4. СОГЛАСОВАНИЕ ИСТОЧНИКА ИНФОРМАЦИИ С КАНАЛОМ СВЯЗИ	129
4.1. Согласование дискретного источника с дискретным каналом без шума. Кодирование информации	129
4.2. Первая и вторая теоремы Шеннона	133
4.3. Кодирование в дискретном канале связи без помех	137
4.4. Оптимальные коды Шеннона-Фано	139
4.4. Код Хаффмена	141
4.6. Канальное кодирование	145
4.7. Достоинства и недостатки эффективных кодов	147
5. ВЗАИМОДЕЙСТВИЕ ОТКРЫТЫХ СИСТЕМ	148
5.1. Понятие открытой системы	151
5.2. Взаимодействие открытых систем	153
5.3. Протокол. Интерфейс. Стек протоколов	156
5.4. Модель OSI	160
5.5. Физический уровень (уровень 1)	160
5.6. Канальный уровень (уровень 2)	162
5.7. Сетевой уровень (уровень 3)	165
5.8. Транспортный уровень (уровень 4)	168
5.9. Сеансовый уровень (уровень 5)	170
5.10. Представительный уровень (уровень 6)	171
5.11. Прикладной уровень (уровень 7)	172
5.12. Сетезависимые и сетезависимые уровни	173
5.13. Инкапсуляция и обработка пакетов	173
6. СИГНАЛЫ ФИЗИЧЕСКОГО УРОВНЯ	175
6.1. Требования к методам цифрового кодирования	175
6.2. Методы линейного кодирования	178
6.3. Ширина магистралей — требуемая полоса частот	192
6.4. Ширина спектра сигнала	192
6.5. Логическое кодирование	194
6.6. Выводы	199
7. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ	201
7.1. Помехоустойчивые коды и их основные параметры	201
7.2. Понятие о конечных полях	207
7.3. Способы задания линейных кодов	217
7.4. Принцип построения циклических кодов	218
7.5. Получение кодовой комбинации добавлением остатка $R(x)$	220
7.6. Построение кодовой комбинации путем умножения на образующий полином	222

7.7. Операции над циклическими кодами	223
7.8. Способы кодирования циклическим кодом	223
7.9. Матричное задание циклических кодов	224
7.10. Коды Хэмминга	225
7.11. Коды BCH	227
7.12. Коды Рида-Соломона (РС)	229
7.13. Методы описания сверточных кодов	232
7.14. Декодирование сверточных кодов	235
7.15. Пороговое декодирование	236
7.16. Турбокоды	237
7.17. Граничные соотношения между параметрами помехоустойчивых кодов	239

8. СИСТЕМЫ СВЯЗИ С УПЛОТНЕНИЕМ И РАЗДЕЛЕНИЕМ КАНАЛОВ

8.1. Уплотнение	241
8.2. Системы связи с частотным разделением каналов	249
8.3. Системы связи с временным разделением	251
8.4. Сравнение систем с частотным и временным разделением	252
8.5. Методы множественного доступа	254
8.6. Растянутый спектр	257
8.7. Технологии уширения спектра беспроводных протоколов	260

9. ОСНОВНЫЕ ВИДЫ ИМПУЛЬСНОЙ МОДУЛЯЦИИ И МАНИПУЛЯЦИИ

9.1. Основные виды импульсной модуляции	280
9.2. Основные виды манипуляции сигналами	287
9.3. Ортогональная модуляция с частотным разделением	297
9.4. Генерирование OFDM-сигналов	300
9.5. Прием-OFDM сигнала	302
9.6. Кодированное разделение с ортогональными несущими	303
9.7. Сложные виды модуляции и манипуляции	304

10. МНОГОПОЗИЦИОННЫЕ ДИСКРЕТНЫЕ СИГНАЛЫ

10.1. Диаграммы состояний многопозиционных сигналов	317
10.2. Формирование и демодуляция многопозиционных радиосигналов	318
10.3. Помехоустойчивость многопозиционных сигналов	321
10.4. Модуляция QPSK и S-DQPSK	321
10.5. Многочастотная модуляция OFDM	324
10.6. Уменьшение отношения пикового значения OFDM-сигнала к его среднему значению	327
10.7. Блочное кодирование в OFDM	328
10.8. Влияние ошибок частотной синхронизации	332

11. ОСНОВНЫЕ ПРОТОКОЛЫ ВЗАИМОДЕЙСТВИЯ РАДИОЭЛЕКТРОННЫХ УСТРОЙСТВ

11.1. Классификация протоколов	334
11.2. Протоколы взаимодействия	338
11.3. Протоколы модуляции	338
11.4. Протокол V.32bis	340
11.5. Протоколы исправления ошибок	345

11.6. Протоколы сжатия данных	347
11.7. Сжатие данных в протоколах MNP	348
11.8. Протоколы передачи файлов	349
11.9. Фирменные протоколы модуляции	350
12. ЦИФРОВОЕ РАДИОВЕЩАНИЕ	357
12.1. Выбор системы вещания	358
12.2. Система ЭВРИКА-147/DAB	359
12.3. Краткое описание стандарта DAB	361
12.4. Система DRM	373
12.5. Особенности, возможности, преимущества и многообразие форматов стандарта DRM	382
12.6. Перспективы цифрового вещания	387
12.7. Передающая часть системы цифрового радиовещания DRM	391
12.8. Кодирование сигналов, поступающих от создателей звуковых программ	393
12.9. MPEG-4 CELP в системе DRM	406
12.10. Метод кодирования MPEG-4 HVXC	408
12.11. Формирование системных каналов DRM	409
12.12. Канальное кодирование и модуляция	413
12.13. Структура передаваемых сигналов	433
13. ПРИНЦИПЫ ПОСТРОЕНИЯ ЦИФРОВЫХ СЕТЕЙ СОТОВОЙ ПОДВИЖНОЙ СВЯЗИ	441
13.1. Сотовые системы	441
13.2. Структурная схема сети мобильной связи	447
13.3. Формирование сигналов в стандарте GSM	457
13.4. Организация физических и логических каналов в GSM	464
13.5. Кодирование и перемежение в каналах стандарта GSM	468
13.6. Повышение эффективности сверточного кодирования в каналах стандарта GSM	472
13.7. Модуляция GMSK в современных системах радиосвязи	473
13.8. Безопасность в мобильных сетях	488
13.9. Беспроводные системы третьего поколения	489
14. БЕСПРОВОДНЫЕ ТЕХНОЛОГИИ ПЕРЕДАЧИ ИНФОРМАЦИИ	492
14.1. Технология Wi-Fi	494
14.2. Основные элементы сети Wi-Fi	495
14.3. Архитектура IEEE 802.11	496
14.4. Стандарты IEEE 802.11	510
14.5. Беспроводная технология WiMAX	528
Заключение	535
Литература	537

ISBN 978-5-222-14681-1



9 785222 146811