Р.Галлагер

ТЕОРИЯ ИНФОРМАЦИИ И НАДЕЖНАЯ СВЯЗЬ

М.: «Советское радио», 1974, 720 с.

В книге собраны, подытожены и заново переосмыслены все основные результаты теории информации. Конструкция наиболее перспективных для практического использования кодов, разнообразные методы декодирования, выражения для вероятностей ошибки, пропускная способность реальных каналов связи, методы сокращения избыточности — все это и многое другое изложено с самых современных позиций. Предлагаемые читателю результаты (вместе с изящными и полными их доказательствами) сведены в книге в единую систему. Математические рассуждения удачно сочетаются с инженерными выводами и техническими рекомендациями.

Книга предназначена для широкого круга инженеров и математиков, специализирующихся по системам связи, системам управления, вычислительным машинам и кибернетическим устройствам. Она также может служить хорошим учебным пособием для аспирантов и студентов.

ОГЛАВЛЕНИЕ

ИМЕННОЙ УКАЗАТЕЛЬ

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ПРЕДИСЛОВИЕ РЕДАКТОРОВ РУССКОГО ПЕРЕВОДА

Круг проблем, составляющих основное содержание этой книги, восходит к К. Э. Шеннону, к его первоначальной работе «Математическая теория связи», опубликованной в 1948 г. В центре внимания книги находится детальное развитие идеи о применении кодирования для помехоустойчивой передачи сообщений по каналам с шумами и для сокращения избыточности, содержащейся в сообщении.

История развития теории информации была бурной и неровной. Новизна тематики, идей и постановок, оригинальность и общность подхода, открытие новых сфер применения современного математического аппарата, а также широковещательное и, быть может, несколько неоправданное название («теория информации» вместо «теория передачи информации») произвели в начале пятидесятых годов своего рода научный бум. В эти годы теория информации привлекла внимание многих ученых за рубежом и в нашей стране, рекрутировала талантливую молодежь. Это не замедлило сказаться на достижениях теории и ее первых шагах практического использования. Именно в пятидесятые годы была выдвинута идея алгебраического кодирования и доказана его оптимальность; построены БЧХ-коды; предложены сверточные и итеративные коды, последовательное декодирование; получены границы вероятности ошибки для оптимальных кодов; исследована пропускная способность многих каналов и ε-энтропия источников.

Вместе с тем такое бурное развитие теории и шум, поднятый вокруг этого, сделали ее модной и привлекательной для ученых самых разных специальностей. Возможность перефразировать проблемы многих наук в терминах извлечения, переработки или хранения информации и перспектива получения после этого готовых решений создали около теории информации атмосферу научного Клондайка. Страницы журналов захлестнул поток легковесных, а иногда и ошибочных статей, посвященных неоправданным применениям теории информации (в большинстве случаев понятия «энтропия») к физике, психологии, кристаллографии, лингвистике, теории трафика и т. п.

Наряду с привлечением внимания к теории информации, что несомненно стимулировало ее развитие, эта волна несла в себе и скрытые опасности; в первую очередь, возможность обратного отлива и компрометации самой теории информации. Благодаря усилиям многих ученых, глубоко понимавших новую науку и обеспокоенных за ее судьбу (поучительны в этом отношении статья К. Э. Шеннона «Бандвагон» и предисловие А. Н. Колмогорова к сборнику работ К. Э. Шеннона), эти опасности были предотвращены. Правда, отлив от теории информации произошел, но это послужило ей лишь на пользу, так как ушли те, кто либо разочаровался в возможности автоматического перенесения результатов, либо почувствовал себя несостоятельным для преодоления обнажившихся трудностей.

Если вначале теорией информации занимались ученые, получившие в основном техническое образование, то впоследствии в её разви-

тие все активнее начали включаться математики различных направлений. Этот естественный процесс происходил в основном потому, что, с одной стороны, инженерам удалось четко сформулировать интересные с точки зрения приложений новые математические задачи, а с другой — эти задачи оказались настолько трудными, что без длительного и глубокого математического анализа нельзя было ожидать их решения.

Интересно отметить здесь, что среди инженеров, работающих в области радиотехники и электросвязи, существовало мнение, что все методы передачи и приема были предложены инженерами и что эти методы основываются на простых и интуитивно понятных соображениях. К таким методам относятся, например, различные классические методы модуляции, фильтрация, накопление, разнесение, корреляционный прием и т. п. Согласно этому мнению специалисты, занимающиеся теорией информации, разрабатывают лишь математически более совершенные основания этих методов, а также исследуют предельно достижимые потенциальные границы для основных параметров систем связи, что в большинстве случаев опять-таки, согласно этому мнению, сводится к доказательству, что известные методы являются оптимальными. Это мнение действительно отражало положение дел на начальных стадиях развития теории информации как науки, выросшей из потребностей радиосвязи, телефонии, телевидения, локации и других видов техники связи и, естественно, питавшейся материнским молоком их плодотворных идей. Однако с течением времени положение изменилось. Для этого потребовались годы глубоких теоретико-информационных исследований. В результате были открыты классы алгебраических, итеративных, каскадных, сверточных и других кодов, а также разработаны изящные методы их декодирования (алгоритмы декодирования циклических кодов, процедуры последовательного и порогового декодирования и др.). Для развития и понимания этих методов требуется знание идей, постановок и решений теории информации наряду с владением целым рядом современных разделов математики. Техническая интуиция здесь уже не является адекватным методом.

Далее, когда удельный вес математически сложных работ в теории информации стал превалирующим, начал ощущаться известный отрыв специалистов, работающих в области теории, от инженеров, занятых непосредственным проектированием систем связи. Причиной этого, с одной стороны, была недостаточность традиционного математического образования инженеров, а с другой стороны, типичное для математиков увлечение абстрактными задачами и формальным изложением, за которыми неискушенному читателю порой трудно было найти рациональное техническое зерно. Немаловажным обстоятельством было то, что в момент первоначальной публикации результатов их доказательства, как правило, выглядели чрезвычайно сложными и запутанными; они не позволяли легко вскрыть лежащие в их основании идеи. Этот отрыв породил некоторую обоюдную иронию в оценке имеющихся достижений. Взаимная разобщенность инженеров и теоретиков стала особенно противоестественной, когда к середине шестидесятых годов в теории был разработан ряд перспективных для практического использования методов кодирования и декодирования.

В связи с этим книга Р. Г. Галлагера, известного специалиста по теории информации, профессора Массачузетского технологического института, занимает особое место среди публикаций, появившихся в последнее время в нашей стране и за рубежом. Как указано в предисловии к русскому изданию, автор поставил себе задачу «перекинуть мост между математиками и инженерами». Излагаемый в книге материал базируется, с одной стороны, на стройных математических результатах, а с другой стороны, направлен на конкретные технические приложения. Весьма примечательно, что автор все результаты приводит с полными доказательствами на уровне строгости, отвечающем математическим публикациям, и в то же время содержание книги доступно для широкого круга читателей. Достигается это за счет того, что изложение начинается с разбора простейших случаев, затем проводится подробное обсуждение, истолкование и практическое осмысливание полученных результатов. Введение полных и строгих доказательств позволяет читателю более глубоко проникнуть в суть выводимых результатов и найти возможности их изменения в соответствии с запросами теории и практики.

Книга представляет собой методически превосходно написанный учебник по теории информации, освещающий результаты, полученные вплоть до конца шестидесятых годов. Тщательно отобранный и внутренне согласованный материал книги дает описание различных сторон проблемы передачи информации. Подробно исследуется применимость основных теорем кодирования для обширного класса каналов и источников: дискретных, непрерывных, без памяти и с памятью. Изучаются их характеристики, пропускная способность, энтропия и скорость как функция искажения. Большое место в книге уделяется построению эффективных методов кодирования и декодирования и оценке их сложности, что весьма существенно для приложений теории информации. Кодирование трактуется автором как некоторое преобразование выхода источника (включающее модуляцию); декодирование — как некоторая обработка сигнала, принятого на выходе канала (включающая демодуляцию), т. е. кодирование и декодирование рассматриваются с наиболее общих позиций, объединяющих теории кодирования, модуляции и приема сигналов. Значительное внимание уделяется источникам и каналам без памяти и гауссовским, что вполне естественно, поскольку изучение теории на примере таких источников и каналов, отражающих определенную реальную ситуацию, позволяет довольно быстро войти в существо изучаемого предмета.

Наряду с изложением известных результатов в книге впервые приведен ряд новых, полученных в последнее время автором и его коллегами. Так, например, среди последних имеются результаты, относящиеся к непрерывным каналам (гл. 7) и источникам (гл. 9). Даны также новые доказательства целого ряда опубликованных ранее результатов.

К сожалению, в книге, как отмечает и сам автор, недостаточно полно отражены результаты, полученные в Советском Союзе. Некоторые из них, непосредственно примыкающие к тексту, указаны в комментарии редакторов, приведенном в конце книги. Отметим здесь, что результаты, полученные в нашей стране, в ряде случаев приводят к сущест-

венному усилению фактов, приведенных автором. Укажем здесь на работы по последовательному декодированию, по ε-энтропии сообщений и исследованию пропускной способности и кодирования для непрерывных источников и каналов. Введены интересные новые классы кодов, исправляющих ошибки, предложены обобщения задачи кодирования источника, когда неизвестна его статистика. Весьма существен для развития теории передачи информации выдвинутый А. Н. Колмогоровым новый подход к ее основаниям.

Для чтения книги требуется знакомство с начальными курсами математического анализа, теории вероятностей, а также элементами теории случайных процессов. Кроме того, предполагается, что читатель имеет некоторую подготовку к восприятию математических доказательств.

Наличие большого числа примеров, упражнений и задач в тексте книги способствует более продуктивному ее усвоению, а также приобретению некоторых навыков к исследованию изучаемых проблем. В русское издание включены решения задач, которые в США были изданы отдельной книгой, доступ к которой разрешен лишь профессорам университетов. В библиографию добавлены публикации на русском языке, вышедшие в основном до 1969 г.

В процессе перевода и редактирования был устранен ряд опечаток, часть из которых нам сообщил Р. Г. Галлагер. Вместе с тем у редакторов остается ощущение, что некоторые опечатки исправить не удалось, особенно в тексте задач и их решений. Кроме того, следует отметить, что автор не всегда строго придерживается принятых им обозначений. Однако это не затрудняет чтения и восприятия материала книги.

Без сомнения, предлагаемая книга Р. Г. Галлагера будет полезна широкому кругу специалистов, работающих в самых различных областях, а также студентам и аспирантам, впервые приступающим к изучению предмета. Следует надеяться, что выход книги будет способствовать взаимопониманию математиков и инженеров, сближению теории и практики передачи информации. Книга может послужить твердой основой для намечающихся в последнее время серьезных тенденций расширения области приложения теории и ее методов.

Перевод книги выполнен Б. С. Цыбаковым (гл. 1—5), К. Ш. Зигангировым (гл. 6) и М. С. Пинскером (гл. 7—9).

М. С. Пинскер
Б. С. Цыбаков

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Одна из моих главных задач при написании этой книги состояла в том, чтобы перекинуть мост между американскими математиками и инженерами, работающими в теории информации. Математики считают, что техническая литература недоступна им отчасти из-за недостаточного понимания реальных проблем связи, а отчасти из-за невнимания к математической строгости, свойственной технической литературе. Инженеры считают, что математическая литература недоступна им из-за широко распространенного использования в ней незнакомых математических результатов. В связи с этим мое мнение состоит в том, что в большей части книги следует использовать лишь простейшие математические методы, ограничивая общность результатов там, где необходимо избежать затруднений, связанных с математическими тонкостями.

Более математически настроенные советские специалисты по теории информации, несомненно, найдут необычным то, что я часто получаю один и тот же результат дважды или трижды в различной степени общности, в то время как доказательство наиболее общего результата так же просто, как и доказательства в менее общих случаях. Это было вызвано желанием не отвлекать внимания студентов, специализирующихся в области техники, от основных идей использованием незнакомого математического аппарата.

Мне бы хотелось принести свои извинения советским коллегам за малочисленность ссылок на советскую литературу. Частично это произошло потому, что существует большая задержка во времени при переводе советской литературы на английский язык, а частично причина была в моем нежелании задерживать публикацию книги на достаточно долгое время, которое требуется для подробного обзора многих важных опубликованных результатов и выяснения их взаимосвязи. Мне кажется, что без этого простое расширение множества ссылок было бы бессмысленным.

Роберт Г. Галлагер

ПРЕДИСЛОВИЕ

Эта книга написана, главным образом, как учебник по теории информации для студентов старших курсов и аспирантов первого года обучения, специализирующихся в области техники или в математике. Предполагается, что читатель обладает знанием начального курса анализа и элементов теории вероятностей, а для чтения последних глав требуются некоторые начальные знания из теории случайных процессов. К сожалению, имеется еще одно требование, которое труднее выполнить. Читатель должен находиться на разумном уровне математической зрелости и обладать способностью к абстрактному мышлению. Основные результаты теории являются довольно тонкими и абстрактными и в ряде случаев они были получены с помощью, казалось бы, весьма окольных путей. К счастью, недавно достигнутые упрощения в теории позволили сделать главные результаты более доступными, чем это было ранее.

Из-за этой деликатности и абстрактности теории приходится проводить рассуждения на более строгом по сравнению с обычно принятым в технике уровне. Чтобы облегчить чтение там, где это было возможно, я старался вместе с доказательством трудных теорем давать объяснение того, почему теорема является важной, и приводить обоснование справедливости теоремы на интуитивном уровне. Была также предпринята попытка дать каждой теореме наиболее простое и элементарное доказательство; многие из приведенных здесь доказательств являются новыми. Я тщательно избегал довольно неудачную для многих элементарных учебников практику упоминания в середине доказательства некоторых нечетко сформулированных математических теорем, на основе которых и завершается доказательство.

Существует целый ряд причин для того, чтобы доказательству теорем уделить здесь особое внимание. Одна из главных причин состоит в том, что при попытке приложить теорию инженер быстро установит, что задачи, возникающие в технике, не часто можно решить с помощью непосредственного применения к ним теорем. Теоремы редко могут быть применены без всяких изменений и нужно разобраться в доказательстве для того, чтобы выяснить, дает ли эта теорема какое-либо продвижение в решении поставленной задачи. Другой причиной для выделения доказательств является то, что методы, использованные в доказательствах, часто оказываются более полезными при проведении новых исследований в этой области по сравнению с самими результатами. Последняя причина обращения особого внимания на точность формулировок результатов и на тщательность доказательств состоит в том, что эта книга задумана, скорее, как существенная часть курса по теории информации, а не как весь курс целиком. Например, философию, интуитивное толкование, примеры и приложения лучше передавать при непосредственном общении на занятиях, в то время как точные утверждения и детали лучше изложить в виде написанного учебника. Для преподавателя

и самостоятельно изучающего курс аспиранта здесь приводится вполне достаточно интуитивного материала, однако студентам последних курсов требуется давать дополнительные разъяснения на занятиях.

В конце книги приведено большое число упражнений и задач. Их диапазон простирается от простых численных примеров до существенных теоретических обобщений. В тексте книги рассмотрено лишь относительно небольшое число примеров, и читатель, который нуждается в конкретных примерах, должен часто прерывать чтение для решения некоторых наиболее простых задач, помещенных в конце книги.

Имеется целый ряд возможностей выбора помещенного здесь материала для односеместрового курса. Главу 1 следует читать вначале (а возможно, также и в конце). После этого, по моему мнению, предпочтительно чтение следующих параграфов (в указанном порядке): 2.1—2.4, 3.1—3.4, 4.1—4.5, 5.1—5.6., 6.1—6.5 и, наконец, либо 6.8—6.9, либо 6.6.—6.7, либо 8.1—8.3. Другая возможность, открытая для студентов, которые имеют некоторые знания из теории случайных процессов, состоит в том, чтобы начать с § 8.1 и 8.2 и затем проследовать по указанному выше пути, используя повсюду в качестве примера канал с гауссовым белым шумом. Следующая возможность, предлагаемая для слушателей, с серьезными намерениями применений на практике, состоит в том, чтобы начать с гл. 6 (опустив § 6.2), затем перейти к § 5.1—5.5, после этого § 6.2 и далее к гл. 2 и 4 и § 8.1 и 8.2. Иные возможности построения курса можно установить с помощью следующей таблицы используемого в тексте материала.

Таблица используемого материала

Параграфы	Используемый материал	Параграфы	Используемый материал
2.1—2.3	нет	6.2	5.6, 6.1
2.4—2.5	2.1—2.3	6.3—6.7	6.1
3.1	2.1—2.3	6.8	6.1
3.2—3.4	2.1—2.3	6.9	5.1—5.5, 6.1—6.5, 6.8
3.5—3.6	3.1—3.4	6.10	6.1—6.5, 6.8
4.1—4.5	2.1—2.3	7.1—7.5	2.1—2.5, 4.3, 5.6—5.7
4.6	3.6, 4.1—4.5	8.1—8.2	нет
5.1—5.5	нет	8.3—8.6	7.1—7.5, 8.1—8.2
5.6—5.8	4.4, 5.1—5.5	9.1—9.6	4.1—4.5
5.9	4.6, 5.1—5.6	9.7	8.5, 9.1—9.5
6.1	нет	9.8	3.5, 9.1—9.5

Как общее правило, последние темы каждой главы являются самыми трудными и излагаются в более сжатой форме по сравнению с темами, рассмотренными ранее. Они включены, главным образом, для аспиран-

тов, а также лиц, работающих в этой области, хотя большинство из них может быть прочитано в течение второго семестра. Преподавателей следует предупредить не тратить слишком много времени на гл. 3, особенно в односеместровом курсе. Материал, помещенный в § 4.1—4.5, 5.1—5.6 и 6.1—6.5, проще и имеет большее значение, чем материал, помещенный в § 3.5—3.6, несмотря на то, что он, возможно, менее знаком некоторым преподавателям.

Я приношу извинения многим авторам значительных работ в теории информации, которых я не упомянул. Я пытался привести ссылки, которые мне казались полезными при написании этой книги, наряду со ссылками на отдельные работы, содержащие дальнейшие продвижения. Многие работы, имеющие историческое значение, были опущены, и цитируемые здесь авторы не обязательно являются теми, кто внес наибольший вклад в эту область.

Мне хочется выразить признательность Исследовательской лаборатории электроники и Электротехническому факультету Массачузетского технологического института (МТИ) за терпеливую поддержку во время подготовки этой книги. Эта работа была поддержана Национальной администрацией по аэронавтике и космосу по контракту NSG-334. Я особенно благодарен Р. М. Фано, который стимулировал мой первый интерес к теории информации и кому я во многом обязан моим подходом к пониманию этого предмета. Работа над этой книгой была начата более четырех лет назад с первоначальным замыслом произвести переработку (под двойным авторством) книги Р. М. Фано *«Передача информации»*. Шли годы, текст разрастался и изменялся и стало ясно, что получилась полностью отличная книга. Однако мой долг *«Передаче информации»* очевиден всякому, кто знаком с обеими книгами.

Я также очень благодарен П. Элайсу, Дж. М. Возенкрафту и К. Э. Шеннону, за их идеи и методы изложения, которые широко использованы здесь. Другой долг я обязан отдать многим студентам и аспирантам, которые слушали курс теории информации в МТИ и делали беспристрастные замечания о многих экспериментах по различным представлениям содержащегося здесь материала. Наконец, я обязан многочисленным коллегам, которые были очень великодушны и дали детальную критику различных частей этой рукописи. В этом отношении особенно большую пользу оказал Дж. Л. Месси. Г. Д. Форни, Х. Юдкин, А. Вайнер, П. Элайс, Р. Кэн, Р. С. Кеннеди, Дж. Макс, Дж. Пинкстон, Э. Берлекэмп, А. Коленберг, И. Джекобс, Д. Сакрисон, Т. Кайлат, Л. Сейдман и Ф. Препарата все вместе сделали большое число критических замечаний, которые способствовали значительному улучшению рукописи.

Роберт Г. Галлагер

1

СИСТЕМЫ СВЯЗИ И ТЕОРИЯ ИНФОРМАЦИИ

1.1. ВВЕДЕНИЕ

Теория связи имеет главным образом дело с системами, предназначенными для передачи информации или данных из одной точки в другую. На рис. 1.1.1 приведена довольно общая блок-схема для того, чтобы наглядно представить поведение таких систем. Выход источника на рис. 1.1.1 может, например, представлять собой звуковую речь; последовательность двоичных символов, поступающих с магнитной ленты; выход ряда датчиков при зондировании космоса; сигналы, поступающие

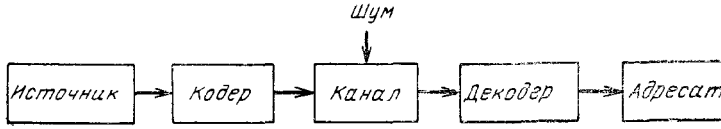


Рис. 1.1.1. Блок-схема системы связи.

к органам чувств живого организма или цель в радиолокационной системе. Каналом может быть, например, телефонная линия; высокочастотная радиолиния; линия космической связи; устройство памяти или живой организм (для случая, когда выходом источника являются сигналы, поступающие к органам чувств живого организма). В канале обычно действуют различные шумовые помехи, которые в телефонной линии, например, могут возникать из-за временных изменений частотной характеристики; из-за разговоров, проникающих из других линий; из-за теплового шума и из-за импульсного шума, источником которого являются переключательные схемы. Кодер на рис. 1.1.1 производит любую обработку выхода источника, совершаемую до передачи. Такая обработка может включать в себя, например, какую-либо комбинацию модуляций, редукции данных и внесения избыточности для борьбы с шумом в канале. Декодер производит обработку сигналов на выходе канала, целью которой является воспроизведение на приемном конце приемлемой копии (или отклика) выхода источника.

В начале 1940-х годов К. Э. Шеннон (1948) разработал математическую теорию, названную теорией информации, которая имеет дело с наиболее фундаментальными аспектами систем связи. Замечательными свойствами этой теории являются, во-первых, широкое привлечение теории вероятностей, во-вторых, указание определяющего значения кодера и декодера как с точки зрения их функциональной роли, так и с точки зрения существования (или не существования) таких кодеров и декодеров, на которых достигается заданный уровень качества передачи. За последние двадцать лет теория информации стала более

точной, получила значительное развитие и достигла того уровня, на котором возможно ее применение к практическим системам связи. Целью этой книги является представление как логической структуры этой теории, так и указание того, где и как эта теория может быть применена.

Так же как любая математическая теория, эта теория оперирует только с математическими моделями, а не с физическими источниками и физическими каналами. Можно было бы предположить поэтому, что было бы удобным начать изложение теории с обсуждения того, как построить подходящие математические модели физических источников и каналов. Однако теории строятся не так главным образом потому, что физическая реальность очень редко является достаточно простой для того, чтобы ее можно было точно представить с помощью модели, поддающейся математической обработке. Мы начнем с изучения простейших классов математических моделей источников и каналов и далее используем складывающиеся представления об этих моделях и относящиеся к ним результаты для изучения все более сложных классов моделей. Естественно, что выбор классов моделей для изучения будет навеян и обусловлен наиболее важными чертами реальных источников и каналов, но наше представление о том, какие из этих черт являются важными, будет видоизменяться на основе теоретических результатов. Наконец, после того как теория будет понята, будет установлено, что она является полезной при исследовании реальных систем связи по следующим двум причинам. Во-первых, она даст основу, на которой можно построить подробные модели реальных источников и каналов. И, во-вторых, что более важно, взаимосвязи, установленные теорией, указывают на типы обменных соотношений, возникающих при построении кодеров и декодеров для заданных систем. В то время как указаные выше замечания могут быть отнесены к почти любой математической теории, они особенно необходимы здесь потому, что должна быть разработана весьма развитая теория до того, как наиболее важные для построения систем связи рекомендации станут очевидными.

Для того чтобы произвести дальнейшие упрощения при изучении моделей источников и моделей каналов, полезно частично отделить эффекты, связанные с источником в системе связи, от эффектов, связанных с каналом. Это может быть сделано с помощью разбиения как кодера, так и декодера, изображенных на рис. 1.1.1, на две части, как это показано на рис. 1.1.2. Задачей кодера для источника является представление выхода источника с помощью последовательности двоичных символов, и один из главных вопросов, возникающих в связи с этим, является вопрос о том, как много двоичных символов в единицу времени требуется для представления выхода любой заданной модели источника. Задача кодера и декодера для канала состоит в том, чтобы надежно воспроизвести двоичные последовательности данных на выходе декодера для канала, и один из главных вопросов, возникающих в связи с этим, состоит в том, возможно ли это сделать, и если возможно, то как.

Конечно, не очевидно, приводят ли представления кодера и декодера в виде, указанном на рис. 1.1.2, к каким-либо фундаментальным

ограничениям характеристик системы связи. Однако одним из наиболее важных результатов теории является то, что при весьма широких условиях никакие такие ограничения не возникают (но этот результат не означает, что кодер и декодер вида, изображенного на рис. 1.1.2, всегда являются наиболее экономичными для достижения заданной точности передачи).

С практической точки зрения разбиение кодера и декодера, указанное на рис. 1.1.2, является особенно удобным, так как это позволяет строить кодер и декодер для канала фактически независимо от кодера и декодера для источника и использовать двоичное представление данных в качестве границы раздела. Это, конечно, облегчает использование различных источников при одном и том же канале.

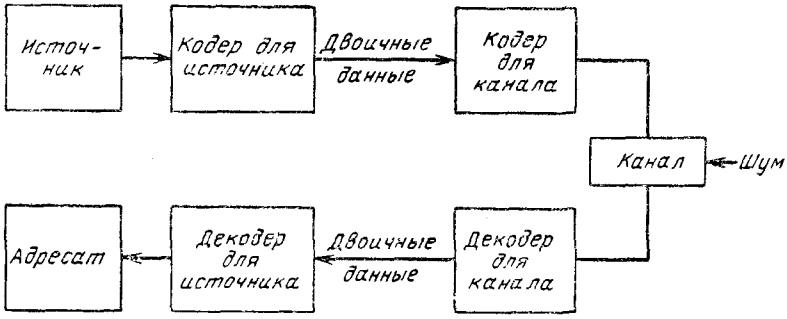


Рис. 1.1.2. Блок-схема системы связи, в которой кодер и декодер разбиты на две части.

В следующих двух параграфах будут кратко описаны классы моделей источников и моделей каналов, которые изучаются в последующих главах, а также будут описаны кодирование и декодирование этих источников и каналов. Так как основное внимание в теории информации сосредоточено, главным образом, на кодировании и декодировании, то нужно ясно понимать, что эта теория неприменима в равной мере ко всем ситуациям, возникающим в связи. Так, например, если источником является радиолокационная цель, то здесь нет возможности кодировать выход источника (если конечно, мы не хотим рассматривать выбор радиолокационных сигналов как метод кодирования), и поэтому нельзя ожидать, что эта теория дает здесь больше, чем взгляд со стороны. Подобно этому, если выходом источника являются сигналы, поступающие к органам чувств живого организма, то мы могли бы рассмотреть организм как комбинацию кодирования, канала и декодирования, но мы не смогли бы управлять кодированием и декодированием, и не совсем ясно, что такая модель является наиболее плодотворной для изучения живых организмов. Таким образом, опять-таки теория информации может дать некоторое понимание поведения таких организмов, но ее нельзя, конечно, рассматривать как магический ключ для понимания.

1.2. МОДЕЛИ ИСТОЧНИКОВ И КОДИРОВАНИЕ ДЛЯ ИСТОЧНИКОВ

Здесь мы дадим краткое описание математических моделей источников, которые будут рассматриваться в дальнейшем. Естественно, что более подробно эти модели будут рассмотрены в последующих главах. Все источники в теории информации моделируются с помощью случайных процессов или случайных последовательностей. Простейший класс моделей источников составляют дискретные источники без памяти. В этих источниках выходом является последовательность (во времени)

Способ 1	*Способ 2*
$a_1 \to 00$	$a_1 \to 0$
$a_2 \to 01$	$a_2 \to 10$
$a_3 \to 10$	$a_3 \to 110$
$a_4 \to 11$	$a_4 \to 111$

Рис. 1.2.1. Два способа преобразования алфавита из четырех букв в двоичные символы.

букв, каждая из которых выбрана из некоторого фиксированного алфавита, скажем, содержащего буквы $a_1, a_2, \ldots, a_K$. Последовательность на выходе источника состоит из этих букв, выбираемых из алфавита статистически независимо и случайно, и при этом выбор производится в соответствии с некоторым заданным распределением вероятностей $Q(a_1), \ldots, Q(a_K)$.

Несомненно, что на первый взгляд кажется довольно странным моделирование реальных источников, которые по предположению производят осмысленную информацию, с помощью случайных процессов. Следующий пример поможет пояснить причину этого. Предположим, что нужно провести некоторое измерение несколько раз подряд и что результатом каждого измерения может быть одно из четырех событий a_1, a_2, a_3 или a_4. Пусть эта последовательность измерений должна быть накоплена в двоичной форме записи, и предположим, что предлагаются два способа перехода к двоичным символам, изображенные на рис. 1.2.1.

В первом из представленных выше способов требуются два двоичных символа для представления каждой буквы источника, в то время как во втором способе требуется переменное число символов. Если известно, что в подавляющем большинстве измерений результатом будет a_1, тогда способ 2 позволит накопить длинную последовательность измерений с помощью значительно меньшего числа двоичных символов, чем способ 1. Методы кодирования выхода дискретного источника в двоичные данные будут подробно обсуждаться в гл. 3. Важным моментом здесь является то, что относительная эффективность методов, изображенных на рис. 1.2.1, критически зависит от частоты появления различных событий и что в математической модели источника последние определяются с помощью распределения вероятности на множестве букв источника. Хорошо известные, но более сложные примеры такого типа дает стенография, в которой короткие символы используются для наиболее употребительных слов, и код Морзе, в котором короткие последовательности точек и тире сопоставлены часто встречающимся буквам, а длинные последовательности — редким буквам.

С кодированием выхода источника в двоичные данные тесно связана мера информации (или неопределенности) букв алфавита источника, которая будет описана в гл. 2. Если k-я буква алфавита источника имеет вероятность $Q(a_k)$, то собственная информация этой буквы (измеренная в битах) определяется как $I(a_k) \triangleq -\log_2 Q(a_k)$. С интуитивной точки зрения (подробнее это будет рассматриваться в гл. 2) это, связанное с техникой связи определение, имеет много качественных черт, свойственных общеупотребительному понятию информации. В частности, если $Q(a_k) = 1$, то $I(a_k) = 0$ в соответствии с тем, что появление a_k не несет никакой информации, так как оно неизбежно должно произойти. Точно так же, чем меньше вероятность a_k, тем больше ее собственная информация. Вместе с тем, нетрудно заметить, что это специальное определение информации имеет некоторые качественные недостатки в сравнении с общеупотребительным понятием информации. Например, не имеет значения то, насколько редким является событие; мы не считаем это информативным (в общеупотребительном смысле), если само по себе наступление события не оказывается интересным для нас. Это не означает, что имеется какой-то недостаток в определении собственной информации; польза от того или иного определения в теории определяется тем, насколько оно дает возможность проникнуть в существо проблемы, и тем, как оно упрощает теоремы. Определение, которое дается здесь, оказывается полезным в теории, главным образом, именно потому, что оно позволяет отделить понятие неожиданности в информации от того, что представляет в информации интерес или смысл.

Среднее по буквам алфавита значение собственной информации является особенно важной величиной, которая называется энтропией буквы источника; энтропия задается выражением

$$\sum_{k=1}^{K} -Q(a_k) \log_2 Q(a_k).$$

Значение энтропии буквы источника определяется, главным образом, теоремой кодирования для источника, которая рассматривается в гл. 3. Она утверждает, что если H — энтропия буквы источника в дискретном источнике без памяти, то последовательность на выходе источника не может быть представлена двоичной последовательностью, использующей в среднем меньше чем H двоичных символов на букву источника, но она может быть представлена двоичной последовательностью, использующей в среднем сколь угодно близкое к H число двоичных символов на букву источника. Некоторое ощущение справедливости этого результата может быть получено, если заметить, что в случае, когда для некоторого целого L источник имеет алфавит из 2^L равновероятных букв, то энтропия буквы источника равна L бит. Вместе с тем, если заметить, что всего имеется 2^L различных последовательностей из L двоичных символов, то можно понять, что каждая из этих последовательностей может быть сопоставлена различным буквам алфавита источника, представляя, таким образом, выход источника с помощью L двоичных символов на букву источника. Этот пример дает

кое-что для понимания того, почему в определении собственной информации и энтропии появляется логарифм.

Часто энтропия также выражается в битах в секунду. Если для дискретного источника без памяти энтропия буквы источника равна H и если источник производит одну букву за каждые τ_s секунд, то энтропия в битах в секунду равна H/τ_s и теорема кодирования для источника указывает, что выход источника может быть представлен двоичной последовательностью, в которой число двоичных символов в секунду сколь угодно близко к H/τ_s.

В качестве более сложного класса моделей источников можно рассмотреть дискретные источники с памятью, в которых последовательные буквы источника статистически зависимы. В § 3.5 аналогичным, но более сложным образом определяется энтропия этих источников (в битах на букву или в битах в секунду) и доказывается, что теорема кодирования для источников справедлива, если источник является эргодическим.

Наконец, в гл. 9 будут рассмотрены недискретные источники. Наиболее известным примером недискретного источника является такой, у которого выходом источника является случайный процесс. При попытке закодировать случайный процесс случайной последовательностью возникает ситуация, по своей природе сильно отличающаяся от кодирования дискретных источников. Случайный процесс можно закодировать в двоичные данные, например, следующим образом: взять выборки случайной функции, затем проквантовать их и после этого закодировать проквантованные выборки в двоичные символы. Различие между этим кодированием и двоичным кодированием, описанным ранее, состоит в том, что выборочные функции не могут быть точно восстановлены по двоичной последовательности и, таким образом, это кодирование следует описывать как в терминах числа двоичных символов в секунду, так и в терминах некоторой меры искажения функции на выходе источника при представлении ее функцией, восстановленной по двоичной последовательности символов. В гл. 9 рассматривается проблема отыскания минимального числа двоичных символов в секунду, достаточного для того, чтобы закодировать выход источника так, чтобы среднее искажение выхода источника при его воспроизведении по двоичной последовательности находилось в заданных пределах. Основное здесь состоит в том, что недискретный источник может быть закодирован с некоторыми искажениями в двоичную последовательность и что требуемое число двоичных символов в единицу времени зависит от допустимого искажения.

1.3. МОДЕЛИ КАНАЛОВ И КОДИРОВАНИЕ ДЛЯ КАНАЛОВ

Для того чтобы описать математически модель канала, мы, во-первых, определим множество возможных сигналов на входе канала (или просто входов канала), во-вторых, множество возможных сигналов на выходе (или выходов канала) и, в-третьих, для каждого сигнала на входе вероятностную меру на множестве сигналов на выходе. Про-

стейший класс моделей каналов образуют дискретные каналы без памяти; они определяются следующим образом. Входом является последовательность букв из конечного алфавита, пусть $a_1, \ldots, a_K$, выходом — последовательность букв из того же самого или другого алфавита, скажем $b_1, \ldots, b_J$. Наконец, каждая буква выходной последовательности зависит статистически только от буквы, стоящей на соответствующей позиции во входной последовательности, и определяется заданной условной вероятностью $P(b_j | a_k)$, определенной для всех букв a_k алфавита на входе и всех букв b_j алфавита на выходе. Примером может служить двоичный симметричный канал (рис. 1.3.1), который представляет собой дискретный канал без памяти с двоичными последовательностями на входе и выходе, в котором каждый символ последовательности на входе с некоторой фиксированной вероятностью $1 - \varepsilon$ воспроизводится на выходе канала правильно и с вероятностью ε изменяется шумом на противоположный символ. В общем случае, в дискретном канале без памяти переходные вероятности исчерпывают собой все известные сведения о том, как сигнал на входе, взаимодействуя с шумом, образует сигнал на выходе. В дальнейшем будет описано, как дискретные каналы без памяти связаны с реальными каналами.

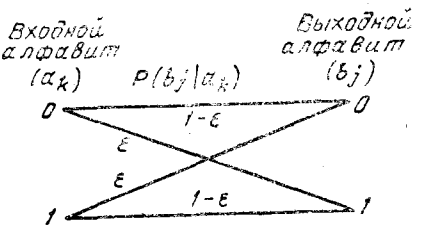


Рис. 1.3.1. Двоичный симметричный канал.

Намного более широкий класс каналов (эти каналы будут называться каналами с памятью) образуют каналы, в которых сигналами на входе снова являются последовательности букв из конечных алфавитов, но в которых каждая буква последовательности на выходе может статистически зависеть не только от соответствующей буквы входной последовательности.

Другой класс моделей каналов, которые имеют более непосредственное сходство с физическими каналами, является класс, в котором как множество входных, так и множество выходных сигналов представляют собой множества функций времени и для каждой заданной функции на входе выход — случайный процесс. Частной моделью из этого класса, которая имеет большую теоретическую и практическую важность, является канал с аддитивным белым гауссовым шумом. Множеством сигналов на входе для такой модели является множество функций времени, удовлетворяющих заданному ограничению сверху на мощность, а сигналы на выходе — сумма сигнала на входе и белого гауссового шума. При использовании этой модели для физического канала с затуханием в качестве входа в модели берется, естественно, сигнал на входе физического канала после его затухания в канале.

При передаче двоичных данных по каналу из рассмотренных выше классов часто бывает удобно разделить как кодер для канала, так и декодер для канала на две части, как показано на рис. 1.3.2. Выходом

кодера для дискретного канала на рис. 1.3.2 является последовательность букв из конечного алфавита $a_1, \ldots, a_K$. Эти буквы производятся во времени с некоторой фиксированной скоростью (одна буква, например, за каждые τ_c секунд). В каждом интервале τ_c секунд модулятор дискретных данных (МДД) производит одну функцию из заданного множества функций $s_1(t), \ldots, s_K(t)$, определенных на интервале длины τ_c. Какая именно из этих функций будет произведена, определяется буквой, поступающей на МДД в течение этого интервала: так, a_1 приводит к $s_1(t)$, а a_2 приводит к $s_2(t)$ и т. д. Таким образом, вся функция на входе канала имеет вид

$$\sum_n s_{i_n}(t - n\tau_c),$$

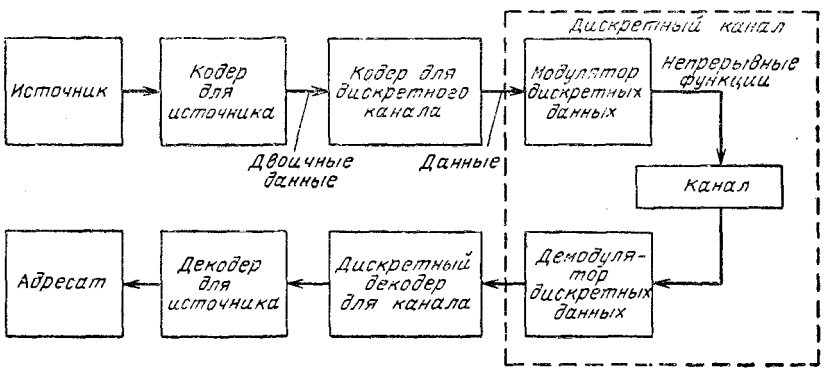


Рис. 1.3.2. Представление непрерывного канала как дискретного канала.

где последовательность i_n, $n = \ldots, -1, 0, 1, \ldots$, определяется соответствующими символами на входе МДД.

Демодулятор дискретных данных (ДДД) принимает поступающие из канала функции и преобразует их в последовательности букв конечного алфавита, $b_1, \ldots, b_J$, производя буквы вновь со скоростью одна буква за каждые τ_c секунд. В простейшем случае каждая буква, выходящая из ДДД, является решением (возможно, что неправильным) о том, какая буква поступила на МДД в соответствующем временном интервале, и в этом случае алфавит $b_1, \ldots, b_J$ будет совпадать с алфавитом на входе МДД. В более сложных случаях выход ДДД будет также содержать информацию о том, насколько правдоподобно решение; в этих случаях выходной алфавит ДДД будет больше, чем входной алфавит МДД.

Как можно заметить из рис. 1.3.2, совокупность МДД, канала, по которому передаются непрерывные сигналы, и ДДД может быть рассмотрена как дискретный канал; именно поэтому дискретные каналы играют большую роль при моделировании физических каналов. Если шум на последовательных интервалах по τ_c секунд является независимым, что имеет место в случае аддитивного белого гауссового шума, то описанный выше дискретный канал является также каналом без памяти.

Рассматривая кодирование и декодирование для класса дискретных каналов, мы, во-первых, получим некоторые результаты, касающиеся кодера и декодера для дискретного канала, входящего в систему, изображенную на рис. 1.3.2, и, во-вторых, сможем использовать эти результаты для того, чтобы в какой-то степени понять, как можно построить МДД и ДДД в такой системе.

Одним из наиболее важных параметров канала является его пропускная способность. Пропускная способность будет определена в гл. 4 и там будет показано, как ее найти для широкого класса дискретных каналов; в гл. 7 и 8 эти рассмотрения будут обобщены, так чтобы охватить недискретные каналы. Пропускная способность определяется с помощью информационной меры, подобной той, которая была использована при рассмотрении источников, и пропускная способность интерпретируется как максимальное среднее количество информации (в битах в секунду), которое может быть передано по каналу. Оказывается, что пропускная способность недискретного канала может быть сколь угодно точно приближена пропускной способностью дискретного канала, который получается из исходного недискретного канала при соответствующем выборе модулятора дискретных данных и демодулятора дискретных данных.

Важность понятия пропускной способности канала основана прежде всего на теореме кодирования для канала с шумами и ее обращении. Грубо говоря, эта теорема кодирования, справедливая для широкого класса каналов, утверждает, что если пропускная способность канала равна C бит в секунду и если двоичные данные поступают на вход кодера этого канала (см. рис. 1.1.2) со скоростью (в двоичных символах в секунду) $R < C$, то с помощью соответствующим образом построенных кодера и декодера можно воспроизводить двоичные символы на выходе декодера со сколь угодно малой вероятностью ошибки. Этот результат точно сформулирован и доказан в гл. 5 для дискретного канала и в гл. 7 и 8 для недискретных каналов. Далеко идущее значение этой теоремы будет обсуждаться ниже в этом параграфе, однако до гл. 5 на интуитивном уровне можно сказать не так уж много. Если объединить этот результат с теоремой кодирования для источников, которая была указана в предыдущем параграфе, то найдем, что если дискретный источник имеет энтропию (в битах в секунду) меньшую, чем C, то выход источника может быть воспроизведен на приемном конце с произвольно малой вероятностью ошибки с помощью использования соответствующего кодирования и декодирования. Аналогично для недискретного источника, если R является минимальным числом двоичных символов в секунду, требующихся, чтобы воспроизвести выход источника с данным уровнем среднего искажения, и если $R < C$, то выход источника может быть передан по каналу и воспроизведен с этим уровнем искажения.

Обращение теоремы кодирования формулируется и доказывается при различной степени общности в гл. 4, 7 и 8. В не очень строгой формулировке она утверждает, что если энтропия дискретного источника в битах в секунду больше, чем C, то независимо от кодирования и декодирования, использованных при передаче выхода источника по каналу,

вероятность ошибки при воспроизведении выхода источника на приемном конце не может быть меньше, чем некоторое положительное число, которое зависит от источника и от C. Так же как показано в гл. 9, если R является минимальным числом двоичных символов в секунду, требуемых для воспроизведения источника с заданным уровнем среднего искажения, и если $R > C$, то независимо от кодирования и декодирования выход источника не может быть передан по каналу и воспроизведен с этим заданным уровнем среднего искажения.

Наиболее удивительным и важным среди указанных выше результатов является теорема кодирования для канала с шумами, которую мы обсудим сейчас более детально. Предположим, что требуется передать данные по дискретному каналу и что по каналу передается одна входная буква за каждые τ_c секунд. Предположим также, что двоичные данные поступают на кодер для канала со скоростью R двоичных символов в секунду. Рассмотрим частный вид кодеров для канала, которые называются блоковыми кодерами; блоковый кодер работает следующим образом. Кодер накапливает двоичные символы на входе кодера в течение некоторого фиксированного интервала T секунд, где T является конструктивным параметром кодера. Во время этого интервала TR двоичных символов поступают на кодер (для простоты мы пренебрегаем здесь тем, что TR может не быть целым числом). Кодер можно представить себе как устройство, которое имеет список всех 2^{TR} возможных последовательностей TR двоичных символов и сопоставленного каждой из этих последовательностей кодового слова, состоящего из последовательности $N = T/\tau_c$ букв на входе канала. При получении некоторой отдельной последовательности TR двоичных символов кодер отыскивает эту последовательность в списке и передает по каналу соответствующее кодовое слово из списка. Требуется T секунд, чтобы передать N — буквенное кодовое слово по каналу, и за это время другая последовательность TR двоичных символов поступит на кодер и начнется передача следующего кодового слова. Простой пример такого кодера представлен на рис. 1.3.3. В этом примере, когда двоичная последовательность 0011... поступает на кодер, то 00 является входом кодера на первом интервале в T секунд, и в конце этого интервала формируется кодовое слово $a_1a_1a_1$ и передается за интервал времени в T секунд. Аналогично 11 является входом кодера на втором интервале времени в T секунд, и $a_1a_2a_3$—соответствующим кодовым словом, передаваемым в течение третьего интервала времени.

Декодер для такого блокового кодера работает аналогичным образом. Декодер накапливает N принятых символов, поступающих из канала и соответствующих переданному кодовому слову, и строит решения (возможно неправильные) относительно соответствующих TR двоичных символов, которые поступили на кодер. Можно считать, что эта процедура решения выполняется декодером с помощью списка всех возможных принимаемых последовательностей из N символов и соответствующей каждой из этих последовательностей последовательности из TR двоичных символов.

Для данного дискретного канала и данной скорости R (в двоичных символах в секунду) поступления символов на кодер имеется свобода

в выборе, во-первых, T (или, что эквивалентно, свободе в выборе $N = T/\tau_c$), во-вторых, множества 2^{TR} кодовых слов и, в-третьих, правила решения. Вероятность ошибки в декодированных двоичных данных, сложность системы и задержка при декодировании зависят от этих выборов. В гл. 5 будет установлено следующее соотношение между параметром T и вероятностью P_e ошибочного декодирования блока TR двоичных символов. Будет показано, что для широкого класса каналов можно выбрать 2^{TR} кодовых слов и правило решения таким образом, что

$$P_e \leqslant \exp\,[-TE\,(R)].$$

Функция $E\,(R)$ является функцией R (числа двоичных символов в секунду, поступающих на кодер) и зависит от модели канала, но не зависит от T. Показывается, что $E\,(R)$ убывает с ростом R, но остается по-

Двоичная последовательность на входе кодера		Кодовое слово на выходе кодера
00	→	$a_1 a_1 a_1$
01	→	$a_2 a_3 a_1$
10	→	$a_3 a_1 a_2$
11	→	$a_1 a_2 a_3$

Рис. 1.3.3. Пример кодера для дискретного канала, $TR=2$, $N=3$.

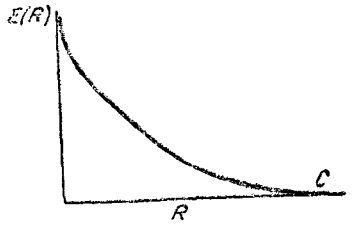


Рис. 1.3.4. График функции $E(R)$ для типичной модели канала.

ложительной при всех R меньших, чем пропускная способность (рис. 1.3.4). Оказывается, что приведенная выше граница для P_e является довольно точной, и не является нецелесообразным рассмотрение $\exp\,[-TE\,(R)]$ в качестве оценки минимальной вероятности ошибки (по всем выборам 2^{TR} кодовых слов и всем правилам решения), которая может быть достигнута при использовании блокового кодера с заданным временем T. Таким образом, чтобы сделать P_e малой, необходимо выбрать T большим, и чем R ближе к C, тем больше должно быть T.

В гл. 6 будут рассмотрены способы построения кодеров и декодеров для канала. Трудно дать простые утверждения, касающиеся сложности и вероятности ошибки этих устройств. Однако, грубо говоря, не трудно заметить, что сложность увеличивается с ростом времени T (для наилучших способов, приближенно линейно с T), что P_e убывает с ростом T при фиксированной R и что T должно возрастать вместе с ростом R для достижения фиксированного значения P_e. Следовательно, грубо говоря, имеется обменное соотношение между сложностью, скоростью и вероятностью ошибки. Чем ближе R к пропускной способности и чем меньше P_e, тем требуется большая сложность кодера и декодера.

Имея в виду указанное выше обменное соотношение, на рис. 1.3.2 можно увидеть с большей ясностью практические преимущества разде-

ления кодера и декодера на две части. В последние годы стоимость цифровых логических устройств постоянно снижалась, в то время как такой революции не было в технике аналоговых устройств. Таким образом, в сложной системе желательно выполнить самые сложные операции в цифровой части системы. Это не говорит, конечно, о том, что аналоговые системы связи полностью вышли из моды, но просто говорит, что, когда отдается предпочтение цифровым системам, возникают многие преимущества не существовавшие еще десять лет тому назад.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Многое в современной теории связи исходит из работ Шеннона (1948), Винера (1949) и Котельникова (1947). Все они ясно понимали фундаментальную роль шума в ограничении точности передачи в системах связи, а также желательность моделирования как сигнала, так и шума с помощью случайных процессов. Винер интересовался отысканием наилучшего линейного фильтра для разделения сигнала и аддитивного шума при заданной задержке, и его работа оказала важное влияние на последующие исследования в теории модуляции. Кроме того, интерес Винера к приему с отрицательной задержкой (т. е. к предсказанию) вместе с работой Колмогорова (1941) по предсказанию в отсутствии шума дали важный толчок в развитии теории управления. Аналогично Котельников интересовался обнаружением и оценкой сигналов на приемном конце. Хотя его работа не так широко известна и используется в Соединенных Штатах (как должно было быть), она внесла значительное понимание как аналоговой модуляции, так и дискретной модуляции.

Работа Шеннона много больше, чем остальные, связана с дискретной техникой и, что более важно, сфокусирована на кодере и декодере в совокупности. Благодаря этому совместному рассмотрению и благодаря тому, что она не ограничивается частными типами приемных устройств, теория Шеннона дает наиболее общие из известных концепций, которые можно заложить в основу при изучении эффективной и надежной передачи.

Хорошее теоретическое введение в теорию связи дают учебники Возенкрафта и Джекобса (1965) и Сакрисона (1968).

2

МЕРА ИНФОРМАЦИИ

Понятия информации и связи в нашем мире являются слишком широкими и емкими, чтобы можно было ожидать какую-либо универсально применимую количественную меру информации. Однако, как это было объяснено в предыдущей главе, имеется множество ситуаций в связи (в особенности таких, которые включают в себя передачу и обработку данных), для которых информация (или данные) и канал адекватно представляются вероятностными моделями. Меры информации, которые будут определены в этой главе, соответствуют этим вероятностным ситуациям, и вопрос о том, насколько адекватны эти меры, зависит в общем от адекватности вероятностной модели.

2.1. ДИСКРЕТНЫЕ ВЕРОЯТНОСТИ; ОБЗОР И ОБОЗНАЧЕНИЯ

Можно представить себе вероятностную модель как эксперимент, исход которого выбирается из множества возможных исходов с вероятностной мерой, заданной на этих возможных исходах. Множество возможных исходов называется выборочным пространством. Для дискретного множества возможных исходов вероятностная мера просто означает приписывание вероятности каждому исходу. Вероятности, конечно, не отрицательны и сумма их равна единице. Выборочное пространство и его вероятностная мера называются ансамблем*); ансамбль будет обозначаться заглавной буквой; исход эксперимента — той же самой, но строчной буквой. Для ансамбля U с выборочным пространством $\{a_1, a_2, \ldots, a_K\}$ вероятность того, что исходом u будет некоторый заданный элемент a_k выборочного пространства, будет обозначаться $P_U(a_k)$. Вероятность того, что исходом будет произвольный элемент u, обозначается через $P_U(u)$. В этом выражении нижний индекс U используется для того, чтобы отметить, какой ансамбль рассматривается, а аргумент u используется как переменная, которая принимает значения из выборочного пространства. Когда это не вызовет путаницы, нижний индекс будет опускаться.

Например, ансамбль U может представлять выход источника в некоторый заданный момент времени; в этом случае алфавит источника есть множество букв $\{a_1, \ldots, a_K\}$ и $P_U(a_k)$ является вероятностью того, что выходом будет буква a_k. Обычно мы будем иметь дело с экспериментами не с одиночным исходом, а с несколькими. Например, нас

*) Почти везде в математической литературе то, что мы называем здесь ансамблем, называется вероятностным пространством.

может интересовать последовательность букв источника, или вход и выход канала, или последовательность входов и выходов канала.

Обозначим исходы эксперимента с парой исходов через x и y, и пусть x принимает значения на множестве исходов $a_1, \ldots, a_K$, а y выбирается на множестве исходов $b_1, \ldots, b_J$. Множество $\{a_1, \ldots, a_K\}$ называется выборочным пространством X; множество $\{b_1, \ldots, b_J\}$ называется выборочным пространством Y, а множество пар $\{a_k b_j\}$, $1 \leqslant k \leqslant K$, $1 \leqslant j \leqslant J$ называется совместным выборочным пространством. Вероятностная мера на совместном выборочном пространстве задается совместной вероятностью $P_{XY}(a_k, b_j)$, определенной для $1 \leqslant k \leqslant K$, $1 \leqslant j \leqslant J$. Совокупность совместных выборочного пространства и вероятностной меры для исходов x и y называется совместным XY-ансамблем.

В ансамбле или совместном ансамбле событие определяется как подмножество элементов выборочного пространства. Для дискретного ансамбля вероятность события равна сумме вероятностей элементов выборочного пространства, содержащихся в этом событии. В рассматриваемом XY-ансамбле событие, состоящее в том, что x принимает некоторое частное значение a_k, соответствует подмножеству пар $\{a_k b_1; a_k b_2; \ldots; a_k b_J\}$. Таким образом, вероятность этого события равна

$$P_X(a_k) = \sum_{j=1}^{J} P_{XY}(a_k, b_j). \tag{2.1.1}$$

В более сокращенной записи то же равенство имеет вид

$$P(x) = \sum_{y} P(x, y). \tag{2.1.2}$$

Подобно этому вероятность данного исхода y равна

$$P(y) = \sum_{x} P(x, y). \tag{2.1.3}$$

Следует соблюдать определенную осторожность при обращении с обозначениями $P(x)$ и $P(y)$ в равенствах (2.1.2) и (2.1.3). Символы x и y играют двойную роль: они обозначают как тот исход, который рассматривается, так и переменную. В частности, если выборочные пространства x и y одни и те же, мы не можем подставить элементы выборочного пространства вместо x и y, не вызвав неопределенности.

Если $P_X(a_k) > 0$, то условная вероятность того, что исходом y является b_j при условии того, что исходом x является a_k, определяется равенством

$$P_{Y|X}(b_j | a_k) = \frac{P_{XY}(a_k, b_j)}{P_X(a_k)}. \tag{2.1.4}$$

В сокращенной записи оно имеет вид

$$P(y \mid x) = P(x, y)/P(x). \tag{2.1.5}$$

Подобно этому

$$P(x \mid y) = P(x, y)/P(y). \tag{2.1.6}$$

События $x = a_k$ и $y = b_j$ по определению статистически независимые, если

$$P_{XY}(a_k, b_j) = P_X(a_k) P_Y(b_j). \tag{2.1.7}$$

Если $P_X(a_k) > 0$, то последнее равенство эквивалентно

$$P_{Y|X}(b_j | a_k) = P_Y(b_j), \tag{2.1.8}$$

т. е. условие не меняет вероятность того, что $y = b_j$. Ансамбли X и Y являются статистически независимыми, если условие (2.1.7) удовлетворяется для всех пар $a_k b_j$ из совместного выборочного пространства.

Рассмотрим далее эксперимент со многими исходами, скажем u_1, u_2, ..., u_N, каждый из которых выбирается из некоторого множества возможных исходов. Множество возможных исходов для u_n называется *выборочным пространством* для U_n, $1 \leqslant n \leqslant N$, а множество возможных исходов для последовательности u_1, ..., u_N называется совместным выборочным пространством эксперимента. Для дискретных выборочных пространств вероятностная мера задается совместной вероятностью $P_{U_1 \ldots U_N}(u_1, \ldots, u_N)$, определенной для каждой последовательности исходов $u_1 \ldots, u_N$ в аргументе. Совокупность совместного выборочного пространства и совместного распределения вероятности называется *совместным ансамблем* $U_1, \ldots, U_N$.

Распределение вероятностей частных исходов и совокупностей исходов определяется с помощью $P(u_1, \ldots, u_N)$ так же, как в случае двух исходов. Например,

$$P_{U_n}(u_n) = \sum_{u_1} \ldots \sum_{\substack{u_i \\ i \neq n}} \ldots \sum_{u_N} P(u_1, \ldots, u_i, \ldots, u_N), \tag{2.1.9}$$

где суммирование распространяется по всем возможным исходам для каждого исхода эксперимента, отличного от n-го. Точно так же

$$P_{U_n U_m}(u_n, u_m) = \sum_{u_1} \ldots \sum_{\substack{u_i \\ i \neq n}} \ldots \sum_{\substack{u_N \\ i \neq m}} P(u_1, \ldots, u_i, \ldots, u_N). \tag{2.1.10}$$

Ансамбли U_1, U_2, ..., U_N называются статистически независимыми, если для всех u_1, ..., u_N

$$P_{U_1 \ldots U_N}(u_1, \ldots, u_N) = \prod_{n=1}^{N} P_{U_n}(u_n). \tag{2.1.11}$$

В качестве примера использования этих обозначений рассмотрим последовательность N букв источника с двоичным алфавитом **0**, **1**. Выборочным пространством для каждого u_n является множество {**0**, **1**}. Совместным выборочным пространством эксперимента является множество всех 2^N последовательностей из N двоичных цифр. Вероятностная мера задает вероятность каждой из этих последовательностей. В частном случае, когда буквы источника статистически независимы, эти вероятности имеют вид (2.1.11). Если N букв имеет одинаковое рас-

пределение, скажем $P_{U_n}(1) = q$ и $P_{U_n}(0) = 1 - q$, то вероятность последовательности зависит только от числа единиц, содержащихся в ней. Вероятность каждой из

$$\binom{N}{j} = \frac{N!}{j!\,(N-j)!}$$

последовательностей, содержащих по j символов **1** и $N - j$ символов **0**, равна $q^j (1 - q)^{N-j}$.

2.2. ОПРЕДЕЛЕНИЕ ВЗАИМНОЙ ИНФОРМАЦИИ

Пусть $\{a_1, \ldots, a_K\}$ будет выборочным пространством X, а $\{b_1, \ldots, \ldots, b_J\}$ будет выборочным пространством Y в XY совместном ансамбле с распределением вероятностей $P_{XY}(a_k, b_j)$. Например, x можно интерпретировать как вход дискретного канала с шумом, а y как его выход. Мы хотим количественно измерить, как много говорит нам о возможности появления некоторого возможного исхода, скажем a_k из ансамбля X, появление некоторого возможного исхода, скажем b_j, из ансамбля Y. На вероятностном языке, появление $y = b_j$ изменяет вероятность $x = a_k$ от априорной вероятности $P_X(a_k)$ до апостериорной вероятности $P_{X|Y}(a_k | b_j)$. Количественной мерой этого изменения (которая оказывается полезной) является логарифм отношения апостериорной вероятности к априорной. Это приводит нас к следующему фундаментальному определению: *информация о событии* $x = a_k$, *содержащаяся в событии* $y = b_j$, *равна*

$$I_{X;Y}(a_k; b_j) = \log \frac{P_{X|Y}(a_k | b_j)}{P_X(a_k)}. \tag{2.2.1}$$

Основание логарифма в этом определении определяет шкалу, по которой измеряется информация. Наиболее часто употребляются основания 2 и e. При основании логарифмов 2 значение выражения (2.2.1) называется числом бит (двоичных единиц) информации, а при натуральных логарифмах значение выражения (2.2.1) называется числом нат (натуральных единиц) информации. Таким образом, число нат равно числу бит, умноженному на $\ln 2 \approx 0{,}693$. Так как большинство положений теории и результатов остаются справедливыми при любом основании логарифмов, то основание будет указываться только в случае необходимости.

Если в равенстве (2.2.1) поменять местами x и y, то получаем, что информация о событии $y = b_j$, содержащаяся в событии $x = a_k$, равна

$$I_{Y;X}(b_j; a_k) = \log \frac{P_{Y|X}(b_j | a_k)}{P_Y(b_j)}. \tag{2.2.2}$$

Покажем теперь, используя определение условных вероятностей, что правые части равенств (2.2.1) и (2.2.2) совпадают. Из-за этой симмет-

рии $I_{X;\,Y}(a_k;\ b_j)$ называется взаимной информацией между событиями $x = a_k$ и $y = b_j$:

$$I_{Y;\,X}(b_j;\ a_k) = \log \frac{P_{XY}(a_k,\ b_j)}{P_X(a_k)P_Y(b_j)} = \log \frac{P_{X|Y}(a_k \mid b_j)}{P_X(a_k)} = I_{X;\,Y}(a_k;\ b_j). \quad (2.2.3)$$

Если не будет возникать недоразумений, мы будем пользоваться сокращенным обозначением для информации о событии x, содержащейся в некотором событии y:

$$I(x;\ y) = \log \frac{P(x \mid y)}{P(x)}. \quad (2.2.4)$$

Полное оправдание определения информации равенством (2.2.1) станет ясным только в ходе развития теории. Однако следующий пример может дать некоторое интуитивное понимание этого определения.

Пример 2.1. Канал, изображенный на рис. 2.2.1, называется *двоичным симметричным каналом*. С вероятностью $1 - \varepsilon$ выходная буква совпадает с входной, и с вероятностью ε она отлична от входной буквы.

В предположении, что входы являются равновероятными $P_X(a_1) = P_X(a_2) = {}^1/_2$, совместные вероятности задаются равенствами

$$P_{XY}(a_1,\ b_1) = P_{XY}(a_2,\ b_2) = \frac{1-\varepsilon}{2},$$

$$P_{XY}(a_1,\ b_2) = P_{XY}(a_2,\ b_1) = \frac{\varepsilon}{2}.$$

Замечая из этих равенств, что выходные буквы равновероятны, получаем

$$\begin{aligned} P_{X|Y}(a_1 \mid b_1) &= P_{X|Y}(a_2 \mid b_2) = 1-\varepsilon, \\ P_{X|Y}(a_1 \mid b_2) &= P_{X|Y}(a_2 \mid b_1) = \varepsilon. \end{aligned} \quad (2.2.5)$$

Взаимная информация тогда равна

$$\begin{aligned} I_{X;\,Y}(a_1;\ b_1) &= I_{X;\,Y}(a_2;\ b_2) = \log(2(1-\varepsilon)), \\ I_{X;\,Y}(a_1;\ b_2) &= I_{X;\,Y}(a_2;\ b_1) = \log(2\varepsilon). \end{aligned} \quad (2.2.6)$$

При $\varepsilon = 0$ канал на рис. 2.2.1 является бесшумным; его выход полностью определяет вход. При $\varepsilon = {}^1/_2$ канал полностью зашумлен; его вход и выход являются статистически независимыми. Предположим теперь, что ε достаточно мало, много меньше чем ${}^1/_2$, и предположим, что $x = a_1$, и $y = b_1$. На выходе канала прием буквы b_1 делает вероятность того, что a_1 была послана много большей соответствующей вероятности для a_2, и из соотношений (2.2.6) видно, что информация, содержащаяся в $y = b_1$ относительно $x = a_1$, является в этом случае положительной. Для $\varepsilon = 0$ эта информация равна 1 бит в соответствии с тем, что $y = b_1$ однозначно определяет на приемнике, какая из двоичных букв была послана. Когда ε увеличивается, эта взаимная информация уменьшается, соответствуя увеличению на приемнике недостатка определенности в том, что был передан $x = a_1$.

Рассмотрим далее случай, в котором передается $x = a_2$ и принимается $y = b_1$. Информация, определяемая равенствами (2.2.6), в этом случае отрицательна (при $\varepsilon < 1/2$), что соответствует тому, что прием b_1 приводит к заблуждению, давая приемнику некоторую степень уверенности в том, что был послан $x = a_1$, а не a_2. В одном из последующих примеров будет видно, как некоторая последующая положительная информация может исправить неправильное впечатление на приемном конце, вызванное первоначальной отрицательной информацией. Интересно заметить, что при ε, стремящемся к 0, эта отрицательная информация стремится к $-\infty$, соответствуя тому, что приемник не только будет находиться в заблуждении, но будет заблуждаться с абсолютной определенностью. К счастью, если $\varepsilon = 0$, то это событие не может произойти.

Как можно заметить из определения (2.2.1), взаимная информация является случайной величиной, т. е. числовой функцией элементов выборочного пространства. Взаимная информация является довольно необычной случайной величиной, так как ее значение зависит от вероятностной меры, однако с ней можно обращаться так же, как с любой другой случайной величиной. В частности, взаимная информация имеет среднее значение, дисперсию, моменты всех порядков и производящую функцию моментов. Среднее значение, которое называется средней взаимной информацией и обозначается $I(X; Y)$, задается равенством*)

$$I(X; Y) = \sum_{k=1}^{K} \sum_{j=1}^{J} P_{XY}(a_k, b_j) \log \frac{P_{X|Y}(a_k \mid b_j)}{P_X(a_k)}. \qquad (2.2.7)$$

В сокращенной записи это равенство имеет вид

$$I(X; Y) = \sum_x \sum_y P(x, y) \log \frac{P(x \mid y)}{P(x)}. \qquad (2.2.8)$$

Отсюда видно, что средняя взаимная информация является функцией только XY-ансамбля, в то время как взаимная информация, которая является случайной величиной, — функцией частных исходов x и y. В примере 2.1 взаимная информация принимает значение $\log[2 \times (1-\varepsilon)]$ с вероятностью $1-\varepsilon$ и значение $\log(2\varepsilon)$ с вероятностью ε. Средняя взаимная информация при этом равна $(1-\varepsilon)\log[2(1-\varepsilon)] + \varepsilon \log(2\varepsilon)$.

Интересным частным случаем взаимной информации является тот, в котором появление данного исхода y, скажем $y = b_j$, однозначно определяет, что исходом x будет данный элемент a_k. В этом случае

$$P_{X|Y}(a_k \mid b_j) = 1 \quad \text{и} \quad I_{X;Y}(a_k; b_j) = \log \frac{1}{P_X(a_k)}. \qquad (2.2.9)$$

Так как это выражение представляет собой взаимную информацию, требуемую для определения $x = a_k$, то *оно определяет собственную*

*) В этом выражении и далее во всей книге принимается, что $0 \log 0$ равно 0. Это соответствует пределу $W \log W$ при W, стремящемся к 0 сверху.

информацию, содержащуюся в событии $x = a_k$, которая обозначается

$$I_X(a_k) = \log \frac{1}{P_X(a_k)}. \qquad (2.2.10)$$

В сокращенной записи это равенство имеет вид: $I(x) = -\log P(x)$.

Собственная информация, содержащаяся в событии $x = a_k$, является, очевидно, функцией только ансамбля X. Собственная информация, содержащаяся в $x = a_k$, всегда неотрицательна и увеличивается с уменьшением $P_X(a_k)$. Она может быть интерпретирована либо как априорная неопределенность события $x = a_k$, либо как информация, требуемая для разрешения этой неопределенности. Собственная информация сперва казалась более простым понятием, чем взаимная информация, так как она определяется с помощью отдельного, а не совместного ансамбля. Мы определили вначале взаимную информацию отчасти потому, что она естественно обобщается на случай недискретных выборочных пространств, в то время как собственная информация не обобщается, а частично потому, что интуитивное понимание собственной информации фактически невозможно, в терминах отдельного ансамбля. Многие попытки, предпринятые в литературе для эвристической интерпретации собственной информации с помощью индивидуального ансамбля, привели к большой путанице. В частности, исходя из отдельного ансамбля, трудно понять, почему информация и неопределенность не должны быть связаны обратной зависимостью, а должны быть двумя различными взглядами на одну и ту же вещь.

Пример 2.2. Рассмотрим ансамбль X, для которого выборочное пространство является множеством всех двоичных последовательностей заданной длины m. Предположим, что все последовательности равновероятны так, что имеются 2^m элементов в выборочном пространстве, каждый с вероятностью 2^{-m}. Собственная информация любого заданного исхода равна при этом

$$I(x) = -\log P(x) = \log 2^m = m \text{ бит}. \qquad (2.2.11)$$

Как и должно быть, согласно интуитивному представлению, требуется m бит собственной информации для определения последовательности m двоичных цифр; этот пример делает ясной причину появления логарифма в мерах информации.

На совместном XY-ансамбле *определим условную собственную информацию, содержащуюся в событии $x = a_k$ при условии появления $y = b_j$, следующим образом*:

$$I_{X|Y}(a_k \mid b_j) = \log \frac{1}{P_{X|Y}(a_k \mid b_j)}. \qquad (2.2.12)$$

Или просто

$$I(x \mid y) = -\log P(x \mid y).$$

Это является собственной информацией, содержащейся в событии $x = a_k$ ансамбля при условии, что $y = b_j$. Ее можно интерпретировать как информацию, которую нужно сообщить наблюдателю для определения $x = a_k$, после того как наблюдатель установил, что произош-

ло событие $y = b_j$. Объединяя определения (2.2.1), (2.2.10) и (2.2.12), получаем

$$I(x; y) = I(x) - I(x|y), \tag{2.2.13}$$

т. е. информация об исходе x, содержащаяся в исходе y, равна собственной информации, требуемой для определения исхода x, уменьшенной на неопределенность этого исхода x при заданном y.

Точно так же, как и взаимная информация, собственная информация тоже является случайной величиной. *Энтропия ансамбля определяется как среднее значение собственной информации и задается равенством*

$$H(X) = \sum_{k=1}^{K} P_X(a_k) \log \frac{1}{P_X(a_k)} = \tag{2.2.14}$$

$$= -\sum_x P(x) \log P(x). \tag{2.2.15}$$

Имеется некоторое дополнительное основание для использования здесь символа H, кроме того, что в теории информации это обозначение используется почти всегда. Энтропия ансамбля тесно связана с энтропией, используемой в статистической термодинамике, и фактически является таковой (с точностью до аддитивной постоянной) при интерпретации множества a_k как множества элементов фазового пространства, имеющих бесконечно малые равные объемы*). К счастью, энтропия в теории информации является понятием значительно более простым, чем в термодинамике.

Условная собственная информация также является случайной величиной на совместном ансамбле XY и имеет среднее значение, задаваемое равенством

$$H(X|Y) = \sum_{x,y} P(x, y) I(x|y) = -\sum_{x,y} P(x, y) \log P(x|y). \tag{2.2.16}$$

Ее можно интерпретировать как среднюю информацию (по x и y), которая требуется для того, чтобы определить x, если известно y.

Если равенство (2.2.13) усреднить по ансамблю XY, то можно найти, что средняя взаимная информация между x и y равна разности между энтропией X и условной энтропией X при заданном Y

$$I(X; Y) = H(X) - H(X|Y). \tag{2.2.17}$$

Это равенство показывает, что $I(X; Y)$ можно интерпретировать как среднюю неопределенность X, которая снимается после наблюдения исхода ансамбля Y; $H(X|Y)$ представляет собой среднюю оставшуюся неопределенность X после наблюдения.

Можно получить еще некоторое соотношение между собственной и взаимной информацией, если рассмотреть совместный ансамбль XY как единый ансамбль, элементами которого являются пары x, y сов-

*) См., например, R. C. Tolman, The Principles of Statistical Mechanics, стр. 168 (величина H у Толмана является отрицательной энтропией).

местного выборочного пространства. Собственная информация, содержащаяся в паре x, y, равна

$$I(x, y) = -\log P(x, y). \tag{2.2.18}$$

Так как $P(x, y) = P(x)P(y|x) = P(y)P(x|y)$, то получаем

$$I(x, y) = I(x) + I(y|x) = I(y) + I(x|y). \tag{2.2.19}$$

Взаимная информация может быть также выражена через $I(x, y)$ следующим образом:

$$I(x; y) = I(x) + I(y) - I(x, y). \tag{2.2.20}$$

Усредняя эти выражения по совместному ансамблю XY, находим

$$H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y), \tag{2.2.21}$$

$$I(X; Y) = H(X) + H(Y) - H(XY). \tag{2.2.22}$$

Пусть теперь $u_1, \ldots, u_N$ будут исходами совместного ансамбля $U_1 \ldots U_N$. *Условная взаимная информация между* u_1 *и* u_2 *при условии, что задано* u_3, *определяется в соответствии с* (2.2.1) *как*

$$I(u_1; u_2|u_3) = \log \frac{P(u_1|u_2, u_3)}{P(u_1|u_3)} = \tag{2.2.23}$$

$$= I(u_1|u_3) - I(u_1|u_2, u_3). \tag{2.2.24}$$

Для средней условной взаимной информации теперь получаем

$$I(U_1; U_2|U_3) = \sum_{u_1}\sum_{u_2}\sum_{u_3} P(u_1, u_2, u_3) \log \frac{P(u_1|u_2, u_3)}{P(u_1|u_3)} = \tag{2.2.25}$$

$$= H(U_2|U_3) - H(U_1|U_2U_3). \tag{2.2.26}$$

Мы могли бы здесь получить неограниченное число соотношений между условной и безусловной взаимной и собственной информациями, используя совместные исходы вместо отдельных исходов в этих выражениях. Одно из соотношений, представляющее определенный интерес, состоит в том, что взаимная информация о некотором частном исходе u_1, содержащаяся в некоторой частной паре исходов u_2u_3, равна информации о u_1, содержащейся в u_2, сложенной с информацией о u_1, содержащейся в u_3 при условии, что задан u_2. Чтобы показать это, рассмотрим

$$I(u_1; u_2) + I(u_1; u_3|u_2) = \log \frac{P(u_1, u_2)}{P(u_1)} +$$

$$+ \log \frac{P(u_1|u_2, u_3)}{P(u_1|u_2)} = \log \frac{P(u_1|u_2, u_3)}{P(u_1)} = I(u_1; u_2, u_3). \tag{2.2.27}$$

Второе соотношение, следующее из цепной формулы для вероятности

$$P(u_1, u_2, \ldots, u_N) = P(u_1)\,P(u_2|u_1) \ldots P(u_N|u_1, \ldots, u_{N-1}),$$

имеет вид

$$I(u_1, u_2, \ldots, u_N) = I(u_1) + I(u_2|u_1) + \ldots + I(u_N|u_1, \ldots, u_{N-1}). \tag{2.2.28}$$

Усредняя (2.2.27) и (2.2.28) по совместному ансамблю, получаем

$$I(U_1; U_2 U_3) = I(U_1; U_2) + I(U_1; U_3 | U_2), \qquad (2.2.29)^*$$

$$H(U_1, U_2, \ldots, U_N) = H(U_1) + H(U_2 | U_1) + \ldots + H(U_N | U_1 \ldots U_{N-1}). \qquad (2.2.30)$$

Пример 2.3. Рассмотрим опять канал, изображенный на рис. 2.2.1, но будем использовать его три раза подряд, так что входом будет последовательность $x_1x_2x_3$ трех двоичных символов, а выходом — последовательность $y_1y_2y_3$ трех двоичных символов. Предположим также, что входная последовательность строится так, что один и тот же символ повторяется трижды: последовательность $a_1a_1a_1$ используется с вероятностью $1/2$, и последовательность $a_2a_2a_2$ с вероятностью $1/2$. Будем считать, наконец, что канал действует независимо на каждый из символов, другими словами, что

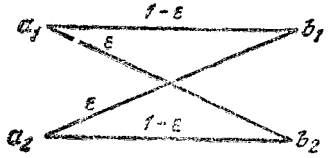


Рис. 2.2.1. Двоичный симметричный канал.

$$P(y_1\, y_2\, y_3 \,|x_1\, x_2\, x_3) = P(y_1 \,|\, x_1)\, P(y_2 \,|\, x_2)\, P(y_3 |\, x_3). \qquad (2.2.31)$$

Исследуем взаимную информацию, когда посылается последовательность $a_1a_1a_1$, а принимается последовательность $b_2b_1b_1$. Покажем, что первый выходной символ содержит отрицательную информацию о входе, однако последующие два выходных символа содержат достаточное количество положительной информации, чтобы перекрыть первоначальное заблуждение. Так же как в равенстве (2.2.6), имеем

$$I_{X_1;\, Y_1}(a_1; b_2) = \log(2\varepsilon), \qquad (2.2.32)$$

$$I_{X_1;\, Y_2 | Y_1}(a_1; b_1 | b_2) = \log \frac{P_{X_1 | Y_1 Y_2}(a_1 | b_2 b_1)}{P_{X_1 | Y_1}(a_1 | b_2)} =$$

$$= \log \frac{1/2}{\varepsilon} = -\log(2\varepsilon). \qquad (2.2.33)$$

Отсюда видно, что условная информация, содержащаяся во втором выходном символе в точности балансирует отрицательную информацию от первого выходного символа. Наглядное объяснение этого состоит в том, что после приема b_2b_1 приемник имеет в точности такую же неопределенность относительно символов на входе, какую он имел вначале. Условная информация, содержащаяся в третьем принятом символе, равна

$$I_{X_1;\, Y_3 | Y_1 Y_2}(a_1; b_1 | b_2\, b_1) = \log[2(1-\varepsilon)]. \qquad (2.2.34)$$

*) Все равенства и теоремы, помеченные звездочкой в этом и последующем параграфах, остаются справедливыми для недискретных ансамблей (см. §§ 2.4 и 2.5).

Общая информация о входе, содержащаяся в трех принятых символах, является теперь положительной в соответствии с тем, что апостериорная вероятность входа $a_1a_1a_1$ больше, чем априорная вероятность $a_1a_1a_1$.

2.3. СРЕДНЯЯ ВЗАИМНАЯ ИНФОРМАЦИЯ И ЭНТРОПИЯ

В этом параграфе будет получен ряд неравенств, для энтропии и средней взаимной информации.

Теорема 2.3.1. Пусть X — ансамбль с выборочным пространством, состоящим из K элементов. Тогда

$$H(X) \leqslant \log K \tag{2.3.1}$$

с равенством тогда и только тогда, когда все элементы равновероятны.

Доказательство. Эта теорема и ряд последующих неравенств могут быть доказаны с помощью соотношений

$$\ln z < z-1; \quad z>0, \quad z \neq 1,$$
$$\ln z = z-1; \quad z=1. \tag{2.3.2}$$

Они проиллюстрированы на рис. 2.3.1 и могут быть проверены аналитически, если заметить, что разность $\ln z - (z-1)$ имеет отрицательную вторую производную и стационарную точку при $z=1$.

Покажем теперь, что $H(X) - \log K \leqslant 0$.

$$H(X) - \log K = \sum_x P(x) \log \frac{1}{P(x)} - \sum_x P(x) \log K =$$
$$= (\log e) \sum_x P(x) \ln \frac{1}{KP(x)}. \tag{2.3.3}$$

Рассматривая сумму только по тем x, для которых $P(x) > 0$, можно применить (2.3.2) и каждому слагаемому; в результате получим

$$H(X) - \log K \leqslant (\log e) \sum_x P(x) \left[\frac{1}{KP(x)} - 1\right] =$$
$$= \log e \left[\sum_x \frac{1}{K} - \sum_x P(x)\right] \leqslant 0. \tag{2.3.4}$$

Последнее неравенство следует из того, что сумма по x имеет не более K слагаемых. Оба неравенства обращаются в равенства тогда и только тогда, когда $1/[KP(x)] = 1$ при всех x; это эквивалентно тому, что элементы равновероятны.

Так как энтропия ансамбля максимальна, когда элементы равновероятны, можно предположить, что энтропия ансамбля увеличится, если вероятность некоторого элемента увеличится за счет другого, более вероятного элемента, этот результат составляет содержание задачи 2.15.

Следующая теорема показывает, что несмотря на то, что взаимная информация как случайная величина может принимать отрицательные значения, средняя взаимная информация всегда неотрицательна.

Теорема 2.3.2*. Пусть XY дискретный совместный ансамбль. Для средней взаимной информации между X и Y справедливо

$$I(X;\ Y) \geqslant 0. \tag{2.3.5}$$

Знак равенства имеет место тогда и только тогда, когда X и Y статистически независимы.

Доказательство. Покажем, что $-I(X;\ Y) \leqslant 0$.

$$-I(X;\ Y) = (\log e) \sum_{x,\ y} P(x,\ y) \ln \frac{P(x)}{P(x \mid y)}. \tag{2.3.6}$$

Сумма в (2.3.6) берется только по тем xy, для которых $P(x,\ y) > 0$. Для этих слагаемых $P(x) > 0$, $P(x \mid y) > 0$ и (2.3.2) можно применить для каждого слагаемого

$$-I(X;\ Y) \leqslant (\log e) \sum_{x,\ y} P(x,\ y) \left[\frac{P(x)}{P(x \mid y)} - 1\right] = \tag{2.3.7}$$

$$= (\log e) \left[\sum_{x,\ y} P(x)\, P(y) - \sum_{x,\ y} P(x,\ y)\right] \leqslant 0. \tag{2.3.8}$$

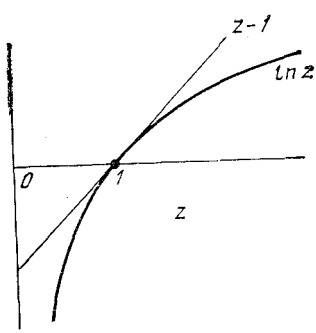


Рис. 2.3.1. Графики функций $\ln z$ и $z - 1$.

Неравенство (2.3.7) переходит в равенство тогда и только тогда, когда $P(x) = P(x \mid y)$, при $P(x,\ y) > 0$. Так как суммирование в (2.3.8) происходит только по тем парам xy, для которых $P(x,\ y) > 0$, то (2.3.8) переходит в равенство только тогда, когда $P(x)\, P(y) = 0$, при $P(x,\ y) = 0$. Таким образом, оба неравенства удовлетворяются вместе с равенством и, следовательно, $I(X;\ Y) = 0$ тогда и только тогда, когда X и Y статистически независимы. |

Непосредственным следствием этой теоремы и равенства $I(X;\ Y) = H(X) - H(X \mid Y)$ является неравенство

$$H(X) \geqslant H(X \mid Y). \tag{2.3.9}$$

Знак равенства имеет место тогда и только тогда, когда X и Y статистически независимы. Таким образом, наложение любого условия на ансамбль может только привести к уменьшению энтропии ансамбля. Важно отметить, что (2.3.9) включает усреднение по обоим ансамблям X и Y. Значение выражения

$$-\sum_{x} P(x \mid y) \log P(x \mid y)$$

может быть как больше, так и меньше $H(X)$ (см. задачу 2.16).

Применяя неравенство (2.3.9) к каждому слагаемому равенства (2.2.30) и подставляя U_n вместо X, а $U_1, \ldots, U_{n-1}$ вместо Y, будем иметь

$$H(U_1, \ldots, U_N) \leqslant \sum_{n=1}^{N} H(U_n). \tag{2.3.10}$$

Знак равенства будет тогда и только тогда, когда ансамбли статистически независимы.

Теорема 2.3.3*. Пусть XYZ — дискретный совместный ансамбль. Тогда

$$I(X; Y \mid Z) \geqslant 0. \tag{2.3.11}$$

Это выражение равно нулю тогда и только тогда, когда при каждом заданном z ансамбли X и Y статистически независимы, т. е., когда

$$P(x, y \mid z) = P(x \mid z) P(y \mid z) \tag{2.3.12}$$

для каждого элемента совместного выборочного пространства, для которого $P(z) > 0$.

Доказательство. Доказательство этой теоремы повторяет доказательство теоремы 2.3.2, если все вероятности заменить на условные при заданном z. |

Из неравенства (2.3.11) и равенства (2.2.26) следует, что

$$H(X \mid Z) \geqslant H(X \mid ZY). \tag{2.3.13}$$

Знак равенства будет тогда и только тогда, когда справедливо (2.3.12).

Случай, когда $I(X; Y \mid Z) = 0$, имеет несколько интересных интерпретаций. Можно представить себе, что в этом случае имеется пара каналов, соединенных последовательно, как показано на рис. 2.3.2. Ансамбль X представляет собой вход первого канала; ансамбль Z представляет собой как выход первого канала, так и вход второго канала и ансамбль Y является выходом второго канала. Предположим, что выход второго канала статистически зависит только от входа второго канала, т. е., что

$$P(y \mid z) = P(y \mid z, x) \text{ для всех } x, y, z \text{ при } P(z, x) > 0. \tag{2.3.14}$$

Умножая обе стороны равенства на $P(x \mid z)$, получаем равенство (2.3.12) так, что

$$I(X; Y \mid Z) = 0. \tag{2.3.15}*$$

Для такой пары последовательных каналов разумно ожидать, что средняя взаимная информация между X и Y будет не больше, чем информация, проходящая через каждый отдельный канал. Покажем, что это справедливо на самом деле. Из равенства (2.2.29) получаем следующие равенства:

$$I(X; YZ) = I(X; Y) + I(X; Z \mid Y) = \tag{2.3.16}*$$

$$= I(X; Z) + I(X; Y \mid Z). \tag{2.3.17}*$$

Приравнивая правые части и используя (2.3.15), будем иметь

$$I(X;\ Z) = I(X;\ Y) + I(X;\ Z|Y). \qquad (2.3.18)^*$$

Из (2.3.11) следует, что $I(X;\ Z|Y) \geqslant 0$, и, таким образом, равенство (2.3.15) означает, что

$$I(X;\ Z) \geqslant I(X;\ Y). \qquad (2.3.19а)^*$$

Принимая во внимание симметрию равенства (2.3.12) относительно X и Y, получаем также, что

$$I(Y;\ Z) \geqslant I(X;\ Y). \qquad (2.3.19б)^*$$

Выражая информацию в неравенстве (2.3.19а) через энтропии, находим, что

$$H(X) - H(X|Z) \geqslant H(X) - H(X|Y),$$
$$H(X|Z) \leqslant H(X|Y). \qquad (2.3.20)$$

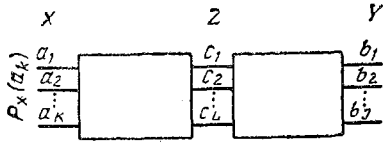


Рис. 2.3.2. Последовательное соединение каналов.

Средняя неопределенность $H(X|Z)$ относительно входа канала при заданном выходе называется *неопределенностью* для канала и, таким образом, неравенства (2.3.20) дают согласующийся с интуицией результат, что эта неопределенность для канала никогда не может уменьшаться при движении от входа по последовательно соединенным каналам.

Неравенства (2.3.19) и (2.3.20) становятся до некоторой степени более удивительными, если интерпретировать второй блок на рис. 2.3.2 как устройство обработки данных, обрабатывающее выход первого блока, который в данном случае является каналом. Независимо от того, является ли эта обработка ансамбля Z детерминированной или вероятностной, она не может уменьшить неопределенность X или увеличить взаимную информацию о X. Это не означает, что никогда не следует обрабатывать выход канала и фактически обработка обычно необходима для того, чтобы как-либо использовать выход канала. Вместе с тем это означает, что среднюю взаимную информацию следует истолковывать как среднюю меру находящихся в распоряжении статистических данных, а не в терминах полезности представления. Этот результат будет обсужден более подробно в гл. 4.

2.4. ВЕРОЯТНОСТЬ И ВЗАИМНАЯ ИНФОРМАЦИЯ ДЛЯ НЕПРЕРЫВНЫХ АНСАМБЛЕЙ

Рассмотрим ансамбль X, определяющий случайную величину x, принимающую значения из выборочного пространства, образованного множеством действительных чисел. Вероятностная мера на этом пространстве проще всего задается с помощью функции распределения

$$F_X(x_1) = \Pr[x \leqslant x_1]. \qquad (2.4.1)$$

Для каждого действительного числа x_1 функция $F_X(x_1)$ задает вероятность того, что случайная величина x будет меньше или равна x_1. Вероятность того, что случайная величина принадлежит интервалу $x_1 < x \leqslant x_2$, задается равенством

$$\Pr[x_1 < x \leqslant x_2] = F_X(x_2) - F_X(x_1). \tag{2.4.2}$$

Так как вероятность любого события должна быть неотрицательной, равенство (2.4.2) означает, что $F_X(x_1)$ является неубывающей функцией x_1. Если исключить возможность бесконечных исходов, то $F_X(x_1)$ возрастает от 0 при $x_1 = -\infty$ до 1 при $x_1 = +\infty$.

Плотность вероятности X (если она существует) задается равенством

$$p_X(x_1) = \frac{dF_X(x_1)}{dx_1} = \lim_{\Delta \to 0} \frac{F_X(x_1) - F_X(x_1 - \Delta)}{\Delta} = \tag{2.4.3}$$

$$= \lim_{\Delta \to 0} \frac{\Pr[x_1 - \Delta < x \leqslant x_1]}{\Delta}. \tag{2.4.4}$$

Таким образом, $p_X(x_1)$ является плотностью вероятности, отнесенной к единице длины на оси x. Плотность вероятности является неотрицательной; она может быть больше чем единица, но ее интеграл от $-\infty$ до $+\infty$ должен быть равен единице. Как видно из равенства (2.4.4), если $p_X(x_1)$ конечна, то вероятность того, что случайная величина x примет значение, в точности равное x_1, равна нулю. Если случайная величина x принимает значение x_1 с ненулевой вероятностью, то часто удобно считать, что p_X имеет импульсы величины $\Pr(x = x_1)$ в точке x_1.

Рассмотрим теперь совместный ансамбль XY, который определяет пару x и y случайных величин, принимающих значения из выборочного пространства, образованного множеством действительных чисел. Вероятностная мера на совместном выборочном пространстве может быть задана с помощью совместной функции распределения вероятностей

$$F_{XY}(x_1, y_1) = \Pr[x \leqslant x_1,\ y \leqslant y_1]. \tag{2.4.5}$$

Она является неубывающей функцией двух переменных и для каждой пары значений x_1 и y_1 она задает вероятность того, что случайная величина x меньше или равна x_1, а случайная величина y меньше или равна y_1. Функции распределения X и Y задаются с помощью совместных функций распределения равенствами:

$$F_X(x_1) = F_{XY}(x_1, \infty), \tag{2.4.6}$$

$$F_Y(y_1) = F_{XY}(\infty, y_1). \tag{2.4.7}$$

Совместная плотность вероятности X и Y (если она существует) задается равенством

$$p_{XY}(x_1, y_1) = \frac{\partial^2 F_{XY}(x_1, y_1)}{\partial x_1 \partial y_1}. \tag{2.4.8}$$

Функция p_{XY} является плотностью вероятности, отнесенной к единичной площади на плоскости xy, и вероятность того, что пара случайных величин x, y принадлежит некоторой области на плоскости, задается интегралом функции p_{XY} по этой области.

Отдельные плотности вероятностей, определенные равенством (2.4.3), задаются также равенствами

$$p_X(x_1) = \int_{-\infty}^{\infty} p_{XY}(x_1, y_1)\, dy_1, \tag{2.4.9}$$

$$p_Y(y_1) = \int_{-\infty}^{\infty} p_{XY}(x_1, y_1)\, dx_1. \tag{2.4.10}$$

Если $p_X(x_1)$ не равна нулю, то условная плотность распределения Y при заданном X задается равенством

$$p_{Y|X}(y_1 | x_1) = \frac{p_{XY}(x_1, y_1)}{p_X(x_1)}. \tag{2.4.11}$$

Она является плотностью вероятности, отнесенной к единице длины случайной величины y при значении y_1, при условии, что случайная величина x принимает значение x_1. Точно так же

$$p_{X|Y}(x_1 | y_1) = \frac{P_{XY}(x_1, y_1)}{P_Y(y_1)}. \tag{2.4.12}$$

Как и для дискретных ансамблей, мы часто будем опускать подстрочные символы у плотностей вероятности, если не будет возникать двусмысленности. Когда это будет делаться, нужно иметь в виду, что, если, например, $p(x)$ является плотностью вероятности случайной величины x, то она не обязательно является той же самой функцией, что и $p(y)$ — плотность вероятности случайной величины y.

Для совместных ансамблей, определяющих более чем две случайные величины, совместная функция распределения и различные совместные, отдельные и условные плотности вероятности определяются аналогичным образом.

Определим теперь взаимную информацию для непрерывного совместного ансамбля. *Пусть совместный ансамбль XY имеет выборочные пространства X и Y, состоящие из множества действительных чисел, и совместную плотность вероятности $p_{XY}(x_1, y_1)$. Взаимная информация между случайной величиной x, принимающей значение x_1, и случайной величиной y, принимающей значение y_1, определяется как*

$$I_{X;Y}(x_1; y_1) = \log \frac{p_{XY}(x_1, y_1)}{p_X(x_1)\, p_Y(y_1)}. \tag{2.4.13}$$

В принятых сокращенных обозначениях это равенство имеет вид

$$I(x; y) = \log \frac{p(x, y)}{p(x)\, p(y)}. \tag{2.4.14}$$

Используя равенства (2.4.11) и (2.4.12), это равенство можно представить как

$$I(x; y) = \log \frac{p(x \mid y)}{p(x)} = \log \frac{p(y \mid x)}{p(y)}. \qquad (2.4.15)$$

Сходство между определениями информации, предложенными здесь и для дискретных ансамблей, удобно для запоминания, но не дает реального основания для введения этого определения. Для того чтобы дать такое обоснование, проквантуем ось x на интервалы длины Δ, а ось y — на интервалы длины δ. Получаемые в результате квантованные случайные величины образуют дискретный ансамбль и взаимная информация между некоторым x-интервалом $(x_1 - \Delta, x_1)$ и некоторым y-интервалом $(y_1 - \delta, y_1)$ задается равенством

$$\log \frac{\Pr[x_1 - \Delta < x \leqslant x_1,\ y_1 - \delta < y \leqslant y_1]}{\Pr[x_1 - \Delta < x \leqslant x_1] \Pr[y_1 - \delta < y \leqslant y_1]}. \qquad (2.4.16)$$

Деля числитель и знаменатель на $\Delta\delta$, получаем

$$\log \frac{\frac{1}{\Delta\delta} \Pr[x_1 - \Delta < x \leqslant x_1,\ y_1 - \delta < y \leqslant y_1]}{\left(\frac{1}{\Delta}\right) \Pr[x_1 - \Delta < x \leqslant x_1] \left(\frac{1}{\delta}\right) \Pr[y_1 - \delta < y \leqslant y_1]}. \qquad (2.4.17)$$

Переходя к пределу при Δ и δ, стремящимся к нулю, получаем, что это выражение переходит в определенное выше выражение для $I_{X;\,Y}(x_1; y_1)$.

Так же как и в случае дискретных ансамблей, взаимная информация является случайной величиной; ее среднее значение равно

$$I(X; Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log \frac{p(x, y)}{p(x)\, p(y)}\, dx\, dy. \qquad (2.4.18)$$

Переходя к чуть более общей ситуации, предположим, что выборочное пространство X является множеством n-мерных действительных векторов, а выборочное пространство Y — множеством m-мерных действительных векторов. Если $p_{XY}(x_1, y_1)$ представляет собой совместную плотность вероятности XY на совместном $(n + m)$-мерном выборочном пространстве, а $p_X(x_1)$ и $p_Y(y_1)$ являются плотностями вероятностей на пространстве X и Y соответственно, то $I_{X;\,Y}(x_1; y_1)$ снова определяется равенством (2.4.13) и снова может быть представлена как предел при все более и более тонком квантовании каждого измерения совместного пространства. Средняя взаимная информация $I(X; Y)$ задается равенством (2.4.18), где теперь интегрирование распространяется по совместному $(n + m)$-мерному пространству.

Пусть теперь x, y и z—случайные величины с действительными конечномерными выборочными пространствами и пусть $p(x, y, z)$—их совместная плотность вероятности. *Условная взаимная информация между x и y при заданном z определяется как*

$$I(x; y \mid z) = \log \frac{p(x, y \mid z)}{p(x \mid z)\, p(y \mid z)}. \qquad (2.4.19)$$

Эту величину, так же как и $I(x; y)$, можно представить в виде предела при все более и более тонком квантовании по осям x, y и z. Средняя условная взаимная информация задается равенством

$$I(X; Y \mid Z) = \int\int\int p(x, y, z) \log \frac{p(x, y \mid z)}{p(x \mid z)\, p(y \mid z)}\, dx\, dy\, dz. \quad (2.4.20)$$

С помощью этих определений немедленно получаем все теоремы и равенства, отмеченные звездочками в § 2.2 и 2.3, если использовать приведенные там доказательства и выводы.

При рассмотрении дискретных ансамблей было ясно, что средняя взаимная информация не зависит от обозначений, принятых для элементов отдельных выборочных пространств. Эта инвариантность по отношению к обозначениям свойственна также средней взаимной информации в случае непрерывных ансамблей, хотя это менее очевидно. Чтобы показать это, рассмотрим совместный ансамбль XZ со случайными величинами x и z и пусть y будет некоторым преобразованием z, т. е. $y = f(z)$. Этот случай можно представить графически (см. рис. 2.3.2), где x — вход канала, z — выход канала и y преобразование z. Так же как и ранее*), $I(X, Y \mid Z) = 0$, и, следовательно, подобно (2.3.19)

$$I(X; Z) \geqslant I(X; Y). \quad (2.4.21)$$

Предположим, далее, что y является обратимым преобразованием z так, что $z = f^{-1}(y)$. Теперь можно рассматривать y как выход канала, а z как преобразованный выход, получая при этом

$$I(X; Y) \geqslant I(X; Z). \quad (2.4.22)$$

Объединяя эти неравенства, получаем $I(X; Y) = I(X; Z)$ и, следовательно, средняя взаимная информация между двумя ансамблями инвариантна к любому обратимому преобразованию одной из случайных величин. В точности такое же доказательство можно, конечно, применить независимо к любому обратимому преобразованию другой случайной величины.

Рассмотрим теперь вопрос о том, можно ли дать осмысленное определение собственной информации для непрерывного ансамбля. Пусть X будет ансамблем, определяющим действительную случайную величину x с конечной плотностью вероятности $p(x)$. Пусть ось x квантуется на интервалы длины Δ так, что собственная информация интервала $(x_1 - \Delta, x_1)$ равна

$$\log \frac{1}{\Pr[x_1 - \Delta < x \leqslant x_1]}. \quad (2.4.23)$$

В пределе при Δ, стремящемся к 0, $\Pr[x_1 - \Delta < x \leqslant x_1]$ стремится к величине $\Delta p_X(x_1)$, которая стремится к нулю. Таким образом собственная информация интервала стремится к ∞, при стремлении длины

*) Здесь имеются некоторые математические тонкости. Так как y однозначно определяется z, то совместная плотность вероятности $p(x, y, z)$ будет иметь импульсные функции. Это является частным случаем более общих ансамблей, которые будут рассмотрены позже, поэтому оставим эти математические тонкости до того времени.

интервала к нулю. Этот результат не является удивительным, если представлять себе действительные числа в виде десятичных дробей. Так как для точного представления произвольного действительного числа требуется бесконечная последовательность десятичных знаков, то следует ожидать, что собственная информация будет бесконечной. Трудность здесь состоит в требовании точного задания действительного числа. С физической точки зрения мы всегда удовлетворены приближенным заданием и любое приемлемое обобщение понятия собственной информации должно включать в себя некоторую желаемую аппроксимацию. Эта проблема будет исследована с фундаментальных позиций в гл. 9, но мы будем использовать термин собственная информация только для дискретных ансамблей.

Для того чтобы иметь дело с различными средними и условными взаимными информациями и производить с ними вычисления, оказывается полезным определить энтропию непрерывного ансамбля. *Если ансамбль X имеет плотность вероятности $p(x)$, определим энтропию X равенством*

$$H(X) = \int_{-\infty}^{\infty} p(x) \log \frac{1}{p(x)}\, dx. \tag{2.4.24}$$

Аналогично, условная энтропия определяется равенством

$$H(X|Y) = \int\int_{-\infty}^{\infty} p(x, y) \log \frac{1}{p(x|y)}\, dx\, dy. \tag{2.4.25}$$

С помощью этих определений подобно равенствам (2.2.17) и (2.2.22) будем иметь

$$I(X; Y) = H(X) - H(X|Y) = \tag{2.4.26}$$

$$= H(Y) - H(Y|X) = \tag{2.4.27}$$

$$= H(X) + H(Y) - H(YX). \tag{2.4.28}$$

Эти энтропии не обязательно положительны, не обязательно конечны, не инвариантны по отношению к преобразованиям случайных величин и не могут быть интерпретированы как средние собственные информации.

Пример 2.4. Следующий пример на приведенные выше определения будет полезен в дальнейшем при рассмотрении каналов с аддитивным гауссовым шумом. Пусть вход канала x будет гауссовской случайной величиной с нулевым средним значением; плотностью вероятности x будет

$$p(x) = \frac{1}{\sqrt{2\pi E}} \exp\left(-\frac{x^2}{2E}\right). \tag{2.4.29}$$

Параметр E является среднеквадратическим значением или «энергией» входа. Будем считать, что выходом канала y является сумма входа и не зависящей от него гауссовской случайной величины с нулевым средним

значением и дисперсией σ^2. Тогда условная плотность вероятности выхода при условии, что задан вход, имеет вид

$$p(y \mid x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-x)^2}{2\sigma^2}\right]. \tag{2.4.30}$$

Это значит, что при заданном x выход y имеет гауссовское распределение с дисперсией σ^2, которое сконцентрировано около точки x. Совместная плотность вероятности $p(x, y)$ равна $p(x)p(y \mid x)$ и совместный ансамбль полностью определен. Удобнее всего найти среднюю взаимную информацию $I(X; Y)$, пользуясь (2.4.27):

$$H(Y \mid X) = -\int p(x) \int p(y \mid x) \log p(y \mid x)\, dy\, dx = \tag{2.4.31}$$

$$= \int p(x) \int p(y \mid x) \left[\log \sqrt{2\pi\sigma^2} + \frac{(y-x)^2}{2\sigma^2} \log e\right] dy\, dx =$$

$$= \int p(x) \left[\log \sqrt{2\pi\sigma^2} + \frac{1}{2} \log e\right] dx = \tag{2.4.32}$$

$$= \frac{1}{2} \log (2\pi e \sigma^2). \tag{2.4.33}$$

В равенстве (2.4.32) было использовано то, что $\int p(y \mid x)(y-x)^2 dy$ равен дисперсии условного распределения, или σ^2.

Заметим теперь, что выход канала является суммой двух независимых гауссовских случайных величин и, таким образом, является гауссовской случайной величиной*) с дисперсией $E + \sigma^2$:

$$p(y) = \frac{1}{\sqrt{2\pi(E+\sigma^2)}} \exp\left[-\frac{y^2}{2(E+\sigma^2)}\right]. \tag{2.4.34}$$

Находя $H(Y)$ таким же образом, как $H(Y \mid X)$, имеем

$$H(Y) = \frac{1}{2} \log [2\pi e (E + \sigma^2)], \tag{2.4.35}$$

$$I(X; Y) = H(Y) - H(Y \mid X) = \frac{1}{2} \log \left(1 + \frac{E}{\sigma^2}\right). \tag{2.4.36}$$

Отметим, что, когда σ^2 стремится к нулю, выход y аппроксимирует вход x с возрастающей точностью и $I(X; Y)$ стремится к ∞. Это следовало ожидать, так как уже было показано, что собственная информация любого заданного выборочного значения x должна равняться ∞.

Часто нас будут интересовать совместные ансамбли, для которых некоторые случайные величины являются дискретными, а некоторые — непрерывными. Простейший способ описания вероятностной меры таких ансамблей состоит в задании совместной вероятности дискретных случайных величин, принимаемой для каждого возможного совместного исхода, и в задании условной взаимной плотности вероятности для непрерывных случайных величин при условии, что задан каждый совместный исход дискретных случайных величин. Например, если слу-

*) См. задачу 2.22.

чайная величина x имеет выборочное пространство $(a_1, \ldots, a_K)$, а выборочным пространством случайной величины y является множество действительных чисел, то определим $P_X(a_k)$ при $1 \leqslant k \leqslant K$ и $p_{Y|X}(y_1 | a_k)$ для всех действительных чисел y_1 и $1 \leqslant k \leqslant K$. Безусловная плотность вероятности случайной величины y при этом равна

$$p_Y(y_1) = \sum_{k=1}^{K} P_X(a_k)\, p_{Y|X}(y_1 | a_k). \tag{2.4.37}$$

Условная вероятность некоторого значения x при условии, что задано значение y_1 случайной величины y, для которого $p_Y(y_1) > 0$, равна

$$P_{X|Y}(a_k | y_1) = \frac{P_X(a_k)\, p_{Y|X}(y_1 | a_k)}{p_Y(y_1)}. \tag{2.4.38}$$

Взаимная информация и средняя взаимная информация между x и y задается соотношениями

$$I_{X;Y}(a_k; y_1) = \log \frac{P_{X|Y}(a_k | y_1)}{P_X(a_k)} = \log \frac{p_{Y|X}(y_1 | a_k)}{p_Y(y_1)}. \tag{2.4.39}$$

$$I(X;Y) = \sum_{k=1}^{K} \int_{-\infty}^{\infty} P_X(a_k)\, p_{Y|X}(y_1 | a_k) \log \frac{p_{Y|X}(y_1 | a_k)}{p_Y(y_1)}\, dy_1. \tag{2.4.40}$$

Условная взаимная информация определяется аналогичным образом. Все равенства, отмеченные звездочкой в §§ 2.2 и 2.3, остаются, очевидно, справедливыми для этих смешанных дискретных и непрерывных ансамблей.

2.5. ВЗАИМНАЯ ИНФОРМАЦИЯ ДЛЯ ПРОИЗВОЛЬНЫХ АНСАМБЛЕЙ

Предыдущие рассмотрения дискретных и непрерывных ансамблей с плотностями вероятностей оказываются подходящими для того, чтобы иметь дело практически со всеми задачами в теории информации, представляющими технический интерес, в особенности, если использовать некоторые разумные ограничения при рассмотрении более общих случаев. Однако для того чтобы точно сформулировать более общие результаты, не расчленяя их на множество частных случаев, часто желательна более абстрактная точка зрения. Детальное представление такой точки зрения требует использования теории меры, что выходит за рамки этой книги*). В этом параграфе будут кратко описаны те результаты, относящиеся к общему случаю, которые могут быть поняты без теории меры. Эти результаты будут использованы в гл. 7 и понадобятся только при исследовании каналов, которые не могут быть описаны с помощью достаточно хороших плотностей вероятностей. Главный результат, который будет получен в этом параграфе, состоит в том,

*) Для знакомства с этой точкой зрения см. Пинскер (1960). Примечания переводчика (Файнстейна) содержат доказательства ряда результатов, полученных в Советском Союзе в этой области, которые не являются широко доступными в переводе на английский язык.

что теоремы и равенства, отмеченные звездочками в 2.2 и 2.3, остаются справедливы в общих случаях.

В терминах теории меры ансамбль X определяется выборочным пространством — множеством событий, каждое из которых является подмножеством элементов выборочного пространства и вероятностной мерой на множестве событий. Множество событий обладает тем свойством, что любое конечное или счетное объединение или пересечение множеств событий является другим событием и что дополнение любого события является другим событием. Вероятностная мера обладает следующими свойствами: каждое событие имеет неотрицательную вероятность, все выборочное пространство имеет вероятность, равную единице, и вероятность любого конечного или счетного объединения непересекающихся событий равна сумме вероятностей отдельных событий. Для всех задач, имеющих практический интерес*), любое подмножество элементов, которое следует рассмотреть, является событием и имеет вероятность.

Совместный ансамбль XY (или $X_1 \ldots X_n$) описывается подобным же образом. Элементы совместного выборочного пространства являются парами x, y, а события представляют собой подмножества совместного выборочного пространства. Существует, однако, дополнительное ограничение, состоящее в том, что, если A является событием в выборочном пространстве X, а B является событием в выборочном пространстве Y, то совместное подмножество AB, соответствующее тому, что x принадлежит A и y принадлежит B, является событием в совместном выборочном пространстве. Отдельные вероятностные меры P_X и P_Y отдельных ансамблей определяются с помощью совместной вероятностной меры P_{XY}. Например, если B совпадает со всем выборочным пространством Y, то

$$P_X(A) = P_{XY}(AB) \text{ для каждого события } A.$$

Для того чтобы определить среднюю взаимную информацию между двумя ансамблями, рассмотрим вначале разбиение ансамбля. *Разбиение X_p ансамбля X определяется как конечный набор* $(A_1, A_2, \ldots, A_K)$, $K \geqslant 1$, *взаимно несовместных событий, объединение которых составляет все выборочное пространство.* Физически разбиение может быть истолковано как правило квантования исхода эксперимента. Разбиение X_p можно рассматривать как дискретный ансамбль с элементами $A_1, \ldots, A_K$ и вероятностями $P_X(A_1), \ldots, P_X(A_K)$. При заданном совместном ансамбле XY можно рассмотреть разбиения пространства X на $A_1, \ldots, A_K$ и пространства Y на $B_1, \ldots, B_J$, для того чтобы получить совместный дискретный ансамбль X_pY_p. Совместные вероятности задаются, конечно, с помощью $P_{XY}(A_kB_j)$. *Определим теперь среднюю*

*) Однако даже для такого простого ансамбля, как единичный интервал с равномерной плотностью вероятности, математически строго можно показать, что существуют патологические множества точек, которые не являются событиями (см. Халмош (1953) стр. 67; то что мы называем здесь событием, там названо измеримым множеством). Вероятность не может быть приписана тем подмножествам, которые не являются событиями, без нарушения аксиом вероятности.

взаимную информацию между двумя ансамблями X и Y следующим образом:

$$I(X;Y) = \sup I(X_p; Y_p), \tag{2.5.1}$$

$$I(X_p;Y_p) = \sum_{k,j} P_{XY}(A_k B_j) \log \frac{P_{XY}(A_k B_j)}{P_X(A_k) P_Y(B_j)}, \tag{2.5.2}$$

где верхняя грань берется по всем разбиениям ансамбля X и всем разбиениям ансамбля Y.

Покажем теперь, что при дальнейшем подразбиении разбиения X или Y величина $I(X_p; Y_p)$ не может уменьшиться. Для того чтобы показать это, рассмотрим рис. 2.5.1. На этом рисунке Y_{p_2} является подразбиением Y_{p_1} в том смысле, что событие B_1, принадлежащее Y_{p_1}, подразбивается на B'_1 и B'_2 из Y_{p_2}. Как мы уже отмечали при исследовании, связанном с рис. 2.3.2, $I(X_p; Y_{p_2}) \geqslant I(X_p; Y_{p_1})$, и то же самое доказательство, очевидно, может быть применено независимо от того, как подразбиваются события разбиения Y_{p_1}. Те же соображения можно применить к X, изменяя роли X и Y.

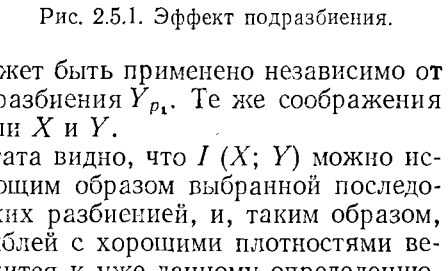


Рис. 2.5.1. Эффект подразбиения.

Из приведенного выше результата видно, что $I(X; Y)$ можно истолковать как предел соответствующим образом выбранной последовательности все более и более тонких разбиений, и, таким образом, для дискретных ансамблей и ансамблей с хорошими плотностями вероятностей определение (2.5.1) сводится к уже данному определению.

Так как уже было доказано, что $I(X_p; Y_p)$ является неотрицательной функцией, то из равенства (2.5.1) следует, что $I(X; Y)$ является неотрицательной. Более того, ансамбли X и Y статистически независимы тогда и только тогда, когда все разбиения ансамблей статистически независимы, и, таким образом, $I(X; Y) = 0$ тогда и только тогда, когда X и Y статистически независимы. Это доказывает теорему 2.3.2 в общем случае.

Для совместного ансамбля XYZ введем подобное определение

$$I(X; YZ) = \sup I(X_p; Y_p Z_p), \tag{2.5.3}$$

где верхняя грань берется по всем разбиениям пространства X, всем разбиениям пространства Y и всем разбиениям пространства Z. В этом определении имеется одна тонкость. Получим ли мы тот же самый результат, если будем понимать YZ как единый ансамбль, а не как совместный ансамбль? Другими словами, если разобьем совместное пространство YZ, а не раздельно пространство Y и пространство Z, изменит ли это значение верхней грани? К счастью, теорема Добрушина*)

*) Формулировку и доказательство этой теоремы см. у Пинскера (1960), стр. 10. Мы предполагаем здесь, что пространство YZ и его события представляют собой произведение индивидуальных пространств и их событий (см. Халмош (1953)).

(1959) показывает, что мы получим тот же самый результат в обоих случаях.

И, наконец, средняя условная взаимная информация определяется как

$$I(X; Z|Y) = I(X; YZ) - I(X; Y). \tag{2.5.4}$$

Это определение не является столь же общим, как наше определение $I(X; Y)$; трудность состоит в том, что $I(X; Z|Y)$ является неопределенной, если как $I(X; YZ)$, так и $I(X; Y)$ являются бесконечными. Пинскер (1960) дал общее определение, основанное на теории меры, однако нам оно здесь не понадобится. Пример, показывающий, почему $I(X; Z|Y)$ нельзя разумно определить как $\sup I(X_p; Z_p|Y_p)$, см. в задаче 2.27.

Определение (2.5.4) также оставляет некоторые сомнения относительно того, удовлетворяется ли равенство

$$I(X; Z|Y) = I(Z; X|Y). \tag{2.5.5}$$

Для того чтобы показать, что (2.5.5) удовлетворяется всегда, когда выражения являются определенными при любом заданном $\varepsilon > 0$, выберем разбиения, для которых:

$$I(X; YZ) - \varepsilon \leqslant I(X_p; Y_pZ_p) \leqslant I(X; YZ),$$
$$I(X; Y) - \varepsilon \leqslant I(X_p; Y_p) \leqslant I(X; Y),$$
$$I(Z; YX) - \varepsilon \leqslant I(Z_p; Y_pX_p) \leqslant I(Z; YX),$$
$$I(Z; Y) - \varepsilon \leqslant I(Z_p; Y_p) \leqslant I(Z; Y).$$

Очевидно, что четыре различных разбиения могут быть выбраны так, что каждое из них будет удовлетворять одному из написанных выше соотношений, и любое разбиение, которое является подразбиением каждого из этих четырех разбиений, удовлетворяет одновременно всем четырем равенствам. Из равенства (2.5.4) следует, что это разбиение будет удовлетворять условиям

$$|I(X; Z|Y) - I(X_p; Z_p|Y_p)| \leqslant \varepsilon, \tag{2.5.6}$$
$$|I(Z; X|Y) - I(Z_p; X_p|Y_p)| \leqslant \varepsilon. \tag{2.5.7}$$

С другой стороны, так как $I(X_p; Z_p|Y_p) = I(Z_p; X_p|Y_p)$, то отсюда следует, что

$$|I(X; Z|Y) - I(Z; X|Y)| \leqslant 2\varepsilon. \tag{2.5.8}$$

Так как $\varepsilon > 0$ можно выбрать произвольно малым, то из этого неравенства вытекает справедливость (2.5.5).

В качестве примера использования этих определений и для того чтобы показать, что $I(X; Y)$ может быть бесконечным, рассмотрим ансамбль, в котором x равномерно распределена на единичном интервале и $y = x$ с вероятностью единица. Разбив пространства X и Y на K равновеликих интервалов, можно убедиться, что $I(X_p; Y_p) = \log K$. Так как K можно сделать произвольно большим, получаем, что $I(X; Y) = \infty$.

Обратимся теперь к задаче определения взаимной информации как случайной величины для произвольных ансамблей. Это определение можно дать только в терминах теории множеств, но, к счастью, у нас не возникнет необходимости в его использовании и мы приводим его здесь только как интересный дополнительный материал. *При заданном ансамбле XY с совместной вероятностной мерой P_{XY} и отдельными мерами P_X и P_Y определим произведение вероятностных мер $P_{X \times Y}$ как вероятностную меру, заданную на совместном пространстве, если бы x и y были статистически независимы и имели меры P_X и P_Y. Взаимная информация $I(x; y)$ между точкой x ансамбля X и точкой y ансамбля Y определяется как логарифм производной Радона—Никодима в точке x, y совместной вероятностной меры P_{XY} по произведению вероятностных мер $P_{X \times Y}$.* Гельфанд и Яглом (1957) показали, что, если существует событие E, для которого $P_{XY}(E) > 0$ и $P_{X \times Y}(E) = 0$, то $I(X; Y) = \infty$; в других случаях $I(X; Y)$ равна математическому ожиданию $I(x; y)$.

В приведенном выше примере совместная вероятностная мера сконцентрирована на прямой $x = y$ и произведение вероятностных мер равномерно распределено на единичном квадрате. Следовательно, если E является событием $x = y$, то $P_{XY}(E) = 1$ и $P_{X \times Y}(E) = 0$, устанавливая, таким образом, справедливость результата Гельфанда—Яглома по крайней мере в этом частном примере.

ИТОГИ И ВЫВОДЫ

В этой главе были определены взаимная и собственная информации и установлены некоторые свойства информации, которые будут использованы в последующих главах. Эти свойства в своем большинстве оказались такими, которые следовало бы ожидать при разумном математическом определении информации, но действительное оправдание этих определений появится лишь в последующих главах. Привлекательное слово «информация» позволяет легко построить интуитивное понимание приведенных здесь понятий, но следует быть осторожным и не смешивать эту интуицию с математикой. В идеале наша интуиция должна предлагать новые математические результаты, а условия, требуемые для доказательства этих математических результатов, должны, в свою очередь, обострять нашу интуицию.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Для дополнительного чтения книга Феллера (1950) является замечательным учебником по теории вероятностей. Материал, содержащийся в этой главе, в основном имеется также у Фано (1961), Эбрамсона (1963) и Эша (1965). Большинство результатов и понятий, приведенных здесь (как и в других главах), развиты Шенноном (1948), оригинальные работы которого остаются до сих пор в высшей степени полезными для чтения. Пинскер (1960) рассмотрел вопросы, изложенные в § 2.5, более полно и строго.

3

КОДИРОВАНИЕ ДЛЯ ДИСКРЕТНЫХ ИСТОЧНИКОВ

В этой главе будет рассмотрено кодирование выхода дискретного источника информации в последовательность букв заданного кодового алфавита. Мы хотим выбрать правила кодирования таким образом, чтобы, по крайней мере с высокой вероятностью, последовательность на выходе источника могла быть восстановлена по закодированной последовательности, а также таким образом, чтобы число букв кода, требуемых на одну букву источника, было по возможности меньшим. Будет показано, что минимальное число двоичных букв кода на одну букву источника, требуемых для представления выхода источника, задается энтропией источника.

Как было указано в § 1.2, выходом дискретной модели источника является случайная последовательность букв дискретного алфавита. Дискретная модель источника является подходящей для реальных источников, которые производят дискретные данные, а также для непрерывных источников, выход которых превращается в дискретные данные с помощью таких операций, как выборка и квантование. В гл. 9 будет проведено фундаментальное рассмотрение проблемы кодирования непрерывного источника в дискретные данные при ограничении средних искажений. Настоящая глава посвящена задаче кодирования (и декодирования) для математических моделей дискретных источников, определяемых как случайные последовательности. Построение математической модели реального источника является задачей более трудной и ее решение связано с детальным изучением внутренних закономерностей источника и их использованием; такое построение модели не может быть грамотно выполнено абстрактным образом.

Предположим, что в каждую единицу времени источник вырабатывает одну букву из конечного множества букв источника $a_1, \ldots, a_K$. Вначале будем считать также, что эти буквы источника производятся с некоторыми фиксированными вероятностями $P(a_1), \ldots, P(a_K)$ и что последовательные буквы статистически независимы. Такие источники называются *дискретными источниками без памяти**). Это предположение о статистической независимости, или отсутствии памяти, является в некотором смысле нереальным для большинства дискретных источников. Однако это предположение дает возможность развить важные понятия,

*) В более общем случае буквы дискретного источника могут выбираться из бесконечного счетного алфавита $a_1, a_2, \ldots$ Мы будем различать случай конечного алфавита от случая счетного алфавита, указывая объем алфавита в первом случае.

связанные с кодированием источника, избегая математических трудностей, к которым приводит учет статистической зависимости.

Во многих практических кодах для источника, таких, как код Морзе и стенография, короткие кодовые слова приписываются наиболее часто возникающим буквам или сообщениям, а длинные кодовые слова — более редким буквам или сообщениям. Например, в коде Морзе часто встречающаяся буква *e* имеет кодовое слово «·», а редкая буква *q* имеет кодовое слово «· · — —». Такие коды, в которых различные кодовые слова содержат различное число кодовых символов, называются *неравномерными кодами*. Если источник производит буквы с фиксированной во времени скоростью и если необходимо передавать закодированные символы с фиксированной во времени скоростью, то неравномерный код приводит к проблеме ожидающей очереди. Когда источник выдает редкую букву, производится длинное кодовое слово и ожидающая очередь убеличивается. Наоборот, часто встречающиеся буквы порождают короткие кодовые слова, сокращая ожидающую очередь.

С практической точки зрения обычно желательно избежать эти проблемы, связанные с ожидающей очередью, с помощью кода с фиксированной длиной, т. е. кода, в котором каждое кодовое слово имеет одну и ту же длину. Число кодовых слов в таких практических кодах, в общем, довольно мало (так, например, 32 слова для телетайпа). Рассмотрением кодов с фиксированной длиной мы начнем следующий параграф, однако наибольший интерес для нас будут представлять коды очень большой длины. Такие коды имеют небольшую практическую значимость, но позволяют ясно обнаружить некоторые более глубокие свойства собственной информации и энтропии.

3.1. КОДЫ С ФИКСИРОВАННОЙ ДЛИНОЙ

Обозначим через $\mathbf{u}_L = (u_1, u_2, \ldots, u_L)$ последовательность L последовательных букв дискретного источника. Каждая буква выбирается из алфавита $a_1, \ldots, a_K$ и, таким образом, имеются K^L различных последовательностей длины L, которые могут появляться на выходе источника. Предположим, что нужно закодировать эти последовательности в слова кода с фиксированной длиной. Если кодовый алфавит состоит из D символов и если длина каждого кодового слова равна N, то существуют D^N различных последовательностей кодовых букв, которые могут быть рассмотрены как кодовые слова. Следовательно, если нужно сопоставить различные кодовые слова разным последовательностям источника (это необходимо сделать, если требуется восстановить каждую возможную последовательность источника по ее кодовому слову), то получаем

$$\frac{N}{L} \geqslant \frac{\log K}{\log D}. \tag{3.1.1}$$

Таким образом, для кодов с фиксированной длиной всегда, когда требуется декодировать последовательность источника по кодовому слову, необходимо иметь по меньшей мере $\log K / \log N$ кодовых букв на одну букву источника. Так, например, в случае телетайпа источник

имеет алфавит из $K = 32$ символов (26 английских букв и 6 специальных символов). Кодируя одну букву источника ($L = 1$) в буквы двоичного кода ($D = 2$), нужно иметь $N = 5$ двоичных символов на один символ источника для того, чтобы удовлетворить условию (3.1.1).

Если мы хотим использовать меньше чем $\log K/\log D$ кодовых букв на один символ источника, то, очевидно, нужно ослабить требование того, чтобы *всегда* было возможно декодировать последовательность источника по кодовой последовательности. Отсюда следует, что мы должны приписать кодовые слова только некоторому подмножеству последовательностей источника длины L. Далее будет показано, что для достаточно больших L вероятность получения последовательности источника, которая не соответствует никакому слову, может быть сделана произвольно малой и в то же самое время число кодовых букв на один символ источника может быть сделано сколь угодно близким к $H(U)/\log D$.

Для источника без памяти вероятность данной последовательности $\mathbf{u}_L = (u_1, \ldots, u_L)$ из L букв источника равна произведению вероятностей отдельных букв источника

$$\Pr(\mathbf{u}_L) = \prod_{l=1}^{L} P(u_l).$$

В этом равенстве каждая буква u_l выбирается из алфавита $a_1, \ldots, a_K$ с вероятностью $P(u_l)$. Так, например, если источник имеет алфавит из двух букв a_1 и a_2 с $P(a_1) = 0{,}7$ и $P(a_2) = 0{,}3$, то вероятность последовательности $\mathbf{u}_3 = (u_1, u_2, u_3)$ при $u_1 = a_2$, $u_2 = a_1$, $u_3 = a_1$ равна $0{,}3 \times 0{,}7 \times 0{,}7 = 0{,}147$. Собственная информация последовательности $\mathbf{u}_L$ имеет вид

$$I(\mathbf{u}_L) = -\log \Pr(\mathbf{u}_L) = -\log \prod_{l=1}^{L} P(u_l) =$$

$$= \sum_{l=1}^{L} -\log P(u_l) = \sum_{l=1}^{L} I(u_l). \qquad (3.1.2)$$

Каждая буква u_l представляет собой статистически независимую выборку для одного и того же источника и, следовательно, (3.1.2) утверждает, что $I(\mathbf{u}_L)$ является суммой L независимых одинаково распределенных случайных величин. Так как среднее значение каждой из случайных величин $I(u_l)$ является энтропией источника $H(U)$, то из закона больших чисел следует, что, если L велико, то $I(\mathbf{u}_L)/L$ будет с большой вероятностью близко к $H(U)$:

$$\frac{I(\mathbf{u}_L)}{L} \approx ? H(U). \qquad (3.1.3)$$

Символ $\approx ?$ в равенстве (3.1.3) используется как для того, чтобы показать приближенность равенства, так и для того, чтобы подчеркнуть, что мы не хотим указать точный смысл этого приближения. В начале

рассмотрим следствия равенства (3.1.3), а затем возвратимся, чтобы дать необходимые математические уточнения. Из (3.1.3) имеем

$$-\log_2 \Pr(\mathbf{u}_L) \approx ? LH(U), \tag{3.1.4}$$

$$\Pr(\mathbf{u}_L) \approx ? 2^{-LH(U)}. \tag{3.1.5}$$

Здесь и до конца этой главы все энтропии будут выражаться в битах, т. е. будут определены логарифмами с основанием 2. Из равенства (3.1.5) вероятность любой типичной достаточно длинной последовательности источника длины L в некотором смысле приближенно равна $2^{-LH(U)}$ и, следовательно, число таких типичных последовательностей M_T должно быть приближенно равно

$$M_T \approx ? 2^{LH(U)}. \tag{3.1.6}$$

Если требуется сопоставить двоичные кодовые слова всем этим типичным последовательностям, то следует отметить, что имеется 2^N различных двоичных последовательностей и, следовательно, требуется, чтобы N было приближенно равно $LH(U)$ для того, чтобы представить все типичные последовательности источника.

Приведенные эвристические соображения дают три различных толкования энтропии источника; одно с помощью вероятности типичных длинных последовательностей источника; другое с помощью числа типичных длинных последовательностей источника и третье с помощью числа двоичных символов, требуемых для представления типичных последовательностей источника. Эти эвристические идеи очень полезны для получения простой картины поведения источников, и они легко обобщаются на источники, в которых имеется статистическая зависимость между последовательными буквами. Однако до того как развить эти идеи, необходимо привести ряд уточнений.

Как было показано, $I(\mathbf{u}_L)$ является суммой L независимых одинаково распределенных случайных величин, каждая из которых имеет конечное математическое ожидание $H(U)$. Закон больших чисел*) утверждает при этом, что для любого $\delta > 0$ существует такое $\varepsilon(L, \delta) > 0$, что

$$\Pr\left[\left|\frac{I(\mathbf{u}_L)}{L} - H(U)\right| > \delta\right] \leqslant \varepsilon(L, \delta) \tag{3.1.7}$$

и

$$\lim_{L\to\infty} \varepsilon(L, \delta) = 0. \tag{3.1.8}$$

Это означает, что вероятность того, что выборочное среднее $I(\mathbf{u}_L)/L$ отличается от $H(U)$ более чем на произвольную фиксированную вели-

*) См. задачу 2.4 или какой-либо элементарный учебник по теории вероятностей. Для счетного бесконечного алфавита следует предположить, что $H(U)$ конечна. Предположение конечности дисперсии в задаче 2.4 не является необходимым (см. Феллер (1950), т. 1, гл. 10, § 2).

чину δ, стремится к нулю при увеличении L. Для некоторых заданных δ и L пусть T — множество последовательностей $\mathbf{u}_L$, для которых

$$\left|\frac{I(\mathbf{u}_L)}{L} - H(U)\right| \leqslant \delta; \quad \mathbf{u}_L \in T. \tag{3.1.9}$$

Это множество ранее мы называли множеством типичных последовательностей; из (3.1.7) получаем

$$\Pr(T) \geqslant 1 - \varepsilon(L, \delta). \tag{3.1.10}$$

Преобразуя (3.1.9), получаем для $\mathbf{u}_L \in T$,

$$L[H(U) - \delta] \leqslant I(\mathbf{u}_L) \leqslant L[H(U) + \delta], \tag{3.1.11}$$

$$2^{-L[H(U)-\delta]} \geqslant \Pr(\mathbf{u}_L) \geqslant 2^{-L[H(U)+\delta]}. \tag{3.1.12}$$

Можно ограничить число M_T последовательностей в T, заметив, что

$$1 \geqslant \Pr(T) \geqslant M_T \min_{\mathbf{u}_L \in T} \Pr(\mathbf{u}_L).$$

Отсюда, используя (3.1.12) в качестве нижней границы $\Pr(\mathbf{u}_L)$ для $\mathbf{u}_L$, принадлежащих T, получаем

$$M_T \leqslant 2^{L[H(U)+\delta]}. \tag{3.1.13}$$

Точно так же, используя (3.1.10), получаем

$$1 - \varepsilon(L, \delta) \leqslant \Pr(T) \leqslant M_T \max_{\mathbf{u}_L \in T} \Pr(\mathbf{u}_L)$$

и, используя (3.1.12) в качестве верхней границы $\Pr(\mathbf{u}_L)$ для $\mathbf{u}_L$, принадлежащих T, находим

$$M_T \geqslant [1 - \varepsilon(L, \delta)]\, 2^{L[H(U)-\delta]}. \tag{3.1.14}$$

Неравенства (3.1.12)—(3.1.14) дают точные утверждения, соответствующие (3.1.5) и (3.1.6).

Предположим далее, что требуется закодировать последовательности источника $\mathbf{u}_L$ в последовательности длины N, представляющие собой кодовые слова с алфавитом из D букв. Будем отображать только одну последовательность источника сообщений в каждую кодовую последовательность, при этом часто будут возникать последовательности из множества последовательностей сообщения, которым не будет сопоставлено никакое кодовое слово. Определим вероятность ошибки P_e как вероятность множества, которому не сопоставлены кодовые слова. Выберем вначале N так, чтобы удовлетворить неравенству

$$N \log D \geqslant L[H(U) + \delta]. \tag{3.1.15}$$

Тогда из (3.1.13) следует, что общее число кодовых слов D^N больше, чем M_T и можно сопоставить кодовые слова всем $\mathbf{u}_L$, принадлежащим T.

Используя этот метод, получаем

$$P_e \leqslant \varepsilon(L, \delta). \tag{3.1.16}$$

Если теперь считать, что L достаточно велико и, одновременно увеличивая N, удовлетворить $N/L \geqslant [H(U) + \delta]/\log D$, то можно увидеть из (3.1.8) и (3.1.16), что P_e стремится к 0 при любом $\delta > 0$. Покажем теперь, что, если N/L фиксировать равным любому числу, меньшему $H(U)/\log D$, то вероятность ошибки должна стремиться к 1 при L, стремящемся к ∞. Выберем теперь N так, чтобы

$$N \log D \leqslant L [H(U) - 2\delta]. \tag{3.1.17}$$

Таким образом, число кодовых слов D^N не больше чем $2^{L[H(U)-2\delta]}$. Так как любая последовательность $\mathbf{u}_L$ из T имеет вероятность, не большую чем $2^{-L[H(U)-\delta]}$, то полная вероятность последовательностей, принадлежащих T, которым можно сопоставить кодовые слова, не больше чем $2^{-L[H(U)-\delta]} \cdot 2^{L[H(U)-2\delta]} = 2^{-L\delta}$. Можно было бы также сопоставить кодовые слова некоторым последовательностям источника, не принадлежащим множеству T, в частности тем, которые имеют бо́льшую вероятность. Вместе с тем общая вероятность последовательностей, не принадлежащих T, не больше $\varepsilon(L, \delta)$. Следовательно, вероятность множества всех последовательностей, принадлежащих и не принадлежащих T, которым могут быть сопоставлены кодовые слова, ограничена сверху в случае, если удовлетворяется (3.1.7), следующим образом:

$$1 - P_e \leqslant \varepsilon(L, \delta) + 2^{-L\delta}. \tag{3.1.18}$$

Таким образом, если $L \to \infty$, а $N/L \leqslant [H(U) - 2\delta]/\log D$, то P_e должна стремиться к 1 при любом $\delta > 0$. Можно подытожить эти результаты следующей фундаментальной теоремой.

Теорема 3.1.1. (Теорема кодирования для источника.) Пусть дискретный источник без памяти имеет конечную энтропию $H(U)$. Рассмотрим кодирование последовательностей из L букв источника в последовательности из N кодовых букв, принадлежащих кодовому алфавиту объема D. Каждой кодовой последовательности может быть сопоставлена только одна последовательность источника. Пусть P_e — вероятность появления последовательности источника, которой не сопоставлена никакая кодовая последовательность. Тогда, если при каком-либо $\delta > 0$

$$N/L \geqslant [H(U) + \delta]/\log D, \tag{3.1.19}$$

то P_e можно сделать произвольно малой, выбирая L достаточно большим. Обратно, если

$$N/L \leqslant [H(U) - \delta]/\log D, \tag{3.1.20}$$

то P_e становится сколь угодно близкой к 1, когда L становится достаточно большим.

Теореме 3.1.1 можно дать очень простое и полезное эвристическое толкование. Так как кодовые слова дают попросту другое представление вероятных последовательностей источника, то они должны иметь в точности такую же энтропию, как последовательности источника.

Кроме того, как было показано в гл. 2, $\log D$ представляет собой наибольшую энтропию на одну букву, которой может обладать последовательность букв алфавита объема D; такая энтропия возникает, когда буквы равновероятны и статистически независимы. Теорема 3.1.1 утверждает, таким образом, что можно закодировать буквы произвольного дискретного источника без памяти так, что энтропия кодовых букв будет, по существу, максимальной.

В интерпретации этой теоремы можно представлять себе L как общее число сообщений, которые выходят из источника во время его работы, и представлять себе N как общее число кодовых букв, которые мы пожелали использовать для представления источника. Затем любой метод може быть использован для кодирования и теорема утверждает, что $H(U)/\log D$ является наименьшим числом кодовых букв на одну букву источника, которые можно использовать, чтобы все еще представлять источник с высокой вероятностью. То, что такое кодирование последовательностей фиксированной длины в последовательности фиксированной длины при очень больших длинах является довольно непрактичным, несущественно для рассматриваемого общего толкования.

3.2. НЕРАВНОМЕРНЫЕ КОДОВЫЕ СЛОВА

Предположим, что дискретный источник без памяти U имеет алфавит из K букв $a_1, \ldots, a_K$ с вероятностями $P(a_1), \ldots, P(a_K)$. Каждая буква источника должна быть представлена кодовым словом, состоящим из последовательности букв, принадлежащих заданному кодовому алфавиту. Обозначим через D число различных символов в кодовом алфавите, а через n_k — число букв в кодовом слове, соответствующем a_k. Позднее будут рассмотрены буквы источника, являющиеся последовательностями букв более простого источника, и, таким образом, мы получим существенное обобщение описанной выше ситуации.

В дальнейшем мы будем главным образом интересоваться величиной $\bar{n}$ — средним числом кодовых букв на одну букву источника:

$$\bar{n} = \sum_{k=1}^{K} P(a_k)\, n_k. \tag{3.2.1}$$

Согласно закону больших чисел, если кодируется очень длинная последовательность букв источника с помощью описанной выше процедуры кодирования, то число кодовых букв на одну букву источника будет с большой вероятностью близко к $\bar{n}$.

До того как изучить вопрос о том, на сколько мало может быть $\bar{n}$, рассмотрим некоторые ограничения на неравномерные коды, которые иллюстрируются рис. 3.2.1.

Заметим, что код I имеет неудачное свойство, состоящее в том, что буквы a_1 и a_2 кодируются в одно и то же кодовое слово **0**. Таким образом, это кодовое слово не может быть однозначно декодировано в букву источника, которая привела к этому слову. Так как такие коды не могут представлять буквы источника, они будут исключены из дальнейшего рассмотрения.

Буквы источника		Код I	Код II	Код III	Код IV
a_1	0,5	0	0	0	0
a_2	0,25	0	1	10	01
a_3	0,125	1	00	110	011
a_4	0,125	10	11	111	0111

Рис. 3.2.1.

Код II на рис. 3.2.1 обладает тем же самым недостатком что и код I, хотя и выраженным более тонким образом. Если источник порождает последовательность a_1a_1, то она будет закодирована в 00, что совпадает с кодовым словом для a_3. Это не вызовет затруднения при декодировании, если между последовательными кодовыми словами имеется какой-то интервал или разделение. Однако если такой интервал допустим, то он должен быть рассмотрен как отдельный символ s кодового алфавита и кодовые слова в коде II должны быть $0s$, $1s$, $00s$ и $11s$. Вводя обозначение для интервала (для ясности), когда это требуется, можно рассматривать такие коды просто как частные случаи кодов без интервала. По этой причине в дальнейшем такие коды не будут рассматриваться отдельно.

Как было отмечено, коды I и II из рис. 3.2.1 не могут быть использованы для представления источника, так как они не являются однозначно декодируемыми. Это приводит к следующему определению. *Код является однозначно декодируемым, если последовательности кодовых букв, соответствующие различным последовательностям источника конечной длины, являются различными.*

Приведенное выше определение непосредственно не дает какого-либо способа определить, является ли некоторый код однозначно декодируемым или нет*). Однако нас будет главным образом интересовать частный класс кодов, которые удовлетворяют условию, известному под названием «свойство префикса», легко показывается, что эти коды однозначно декодируемы.

Для того чтобы определить свойство префикса, допустим, что k-е кодовое слово в коде представляется как $\mathbf{x}_k = (x_{k,1}, \ldots, x_{k,n_k})$, где $x_{k,1} \ldots, x_{k,n_k}$ —отдельные кодовые буквы, составляющие кодовое слово. Любая последовательность, составленная из начальной части $\mathbf{x}_k$, т. е. $x_{k,1}, \ldots, x_{k,i}$ для некоторого $i \leqslant n_k$, называется префиксом $\mathbf{x}_k$. *Код, обладающий свойством префикса, определяется как код, в котором никакое кодовое слово не является префиксом никакого другого кодового слова.*

*) Сардинас и Паттерсон (1953) придумали критерий однозначной декодируемости. Простое изложение и доказательство этого критерия см. в задаче 3.14.

На рис. 3.2.1. можно заметить, что код I не обладает свойством префикса, так как **1**, кодовое слово для a_3, является префиксом **10**, кодового слова для a_4. Аналогично, если внимательно просмотреть определение префикса, то можно заметить, что **0**, кодовое слово для a_1, является префиксом **0**, кодового слова для a_2. Иными словами, любой код, два или более кодовых слова которого совпадают, не являются кодом, обладающим свойством префикса. Читателю предлагается проверить, что коды II и IV на рис. 3.2.1 не обладают свойством префикса, а код III обладает.

Для того чтобы декодировать последовательность кодовых слов из кода, обладающего свойством префикса, следует просто начать с начала последовательности и декодировать одно слово сразу. Когда дойдем до конца кодового слова, мы будем знать, что это конец, так как это кодовое слово не является префиксом какого-либо другого кодового слова. Таким образом, можно однозначно декодировать любую последовательность кодовых букв, соответствующую последовательности букв источника и, тем самым доказано, что любой код, удовлетворяющий свойству префикса, является однозначно декодируемым кодом. Так, например, последовательность источника $a_1a_4a_2a_1$ кодируется кодом III рис. 3.2.1 в **0111100**. Так как первая буква в кодовой последовательности есть **0** и она соответствует a_1 и не является начальным отрезком любой другой последовательности, то декодируется a_1 и остается кодовая последовательность **111100**. Как **1**, так **11** не соответствуют никакому кодовому слову, а **111** соответствует и декодируется в a_4, после чего остается **100**. Далее **10** декодируется в a_2 остается только **0**, который декодируется в a_1.

Не любой однозначно декодируемый код обладает свойством префикса. Чтобы заметить это, рассмотрим код IV на рис. 3.2.1. В нем каждое кодовое слово является префиксом каждого более длинного кодового слова. Вместе с тем единственность декодирования является тривиальной, так как символ **0** всегда означает начало нового кодового слова. Коды, обладающие свойством префикса, отличаются, однако, от других однозначно декодируемых кодов тем, что конец кодового слова всегда может быть опознан, так что декодирование может быть выполнено без задержки наблюдаемой последовательности кодовых слов. По этой причине коды, обладающие свойством префикса, называются иногда *мгновенными* кодами.

Удобное графическое представление множества кодовых слов, удовлетворяющих свойству префикса, можно получить, представляя каждое кодовое слово концевым узлом на дереве. Дерево, представляющее кодовые слова кода III рис. 2.3.1, показано на рис. 3.2.2. Начиная с основания дерева, два ребра, ведущие к узлам первого порядка, соответствуют выбору между **0** и **1**, рассматриваемым в качестве первой буквы кодовых слов. Подобно этому два ребра, исходящие из правого узла первого порядка, соответствуют выбору между **0** и **1** для второй буквы кодового слова, если первая буква была **1**; такое же представление применимо и для других ребер. Последовательность символов каждого кодового слова можно рассматривать как правило восхождения от основания дерева к концевому узлу, представляющему желае-

мую букву источника. Дерево можно также использовать для представления кодовых слов кода, который не обладает свойством префикса, однако в этом случае некоторые промежуточные узлы дерева будут соответствовать кодовым словам.

Обратим теперь наше внимание на задачу выбора кода, обладающего свойством префикса, так чтобы минимизировать $\bar{n}$. В начале приведем эвристическое рассмотрение этой задачи, затем докажем некоторые общие теоремы относительно длин кодовых слов и, наконец, дадим алгоритм построения кода, который минимизирует $\bar{n}$.

Для префиксного кода приемник можно представить себе как устройство, наблюдающее последовательность кодовых букв и прослеживающее их путь вверх по дереву, как изображено рис. 3.2.2, чтобы

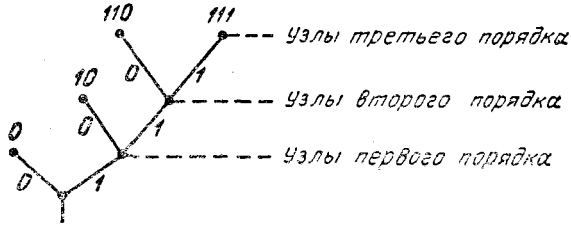


Рис. 3.2.2. Представление в виде дерева кодовых слов кода III, изображенного на рис. 3.2.1.

декодировать сообщение источника. В каждом узле этого дерева следующий кодовый символ дает информацию о том, какое нужно взять ребро. Можно заметить, что сумма взаимных информаций в последовательных узлах, ведущих к некоторому данному концевому узлу, равна в точности собственной информации символа источника, соответствующего этому узлу. Таким образом, чтобы достичь малого значения $\bar{n}$, следует достичь большого значения средней взаимной информации во всех промежуточных узлах дерева. Это в свою очередь говорит о том, что нужно попытаться выбрать кодовые слова таким образом, чтобы все ребра, исходящие из узла, в соответствующем дереве, были равновероятными. Сопоставляя дерево рис. 3.2.2 с вероятностями букв источника для кода III рис. 3.2.1, можно заметить, что каждое ребро дерева берется с вероятностью 1/2. Для этого примера $\bar{n} = 1{,}75$ и $H(U) = 1{,}75$. В длинной последовательности L букв источника значение $\bar{n}$ с большой вероятностью близко к числу кодовых букв на одну букву источника. Следовательно, если бы существовал код с $\bar{n} < H(U)$, можно было бы при больших L закодировать большинство последовательностей источника длины L с помощью меньше чем $H(U)$ кодовых букв на одну букву источника, в нарушение теоремы 3.1.1. Это означает, что 1,75 является минимальной возможной величиной $\bar{n}$ для этого кода. Это не удивительно, так как мы смогли построить код так, что каждая кодовая буква содержит в точности 1 бит информации о выходе источника.

Этот пример дает возможность произвести очевидное обобщение, позволяющее построить префиксные коды для общего множества букв источника. Разобьем вначале множество букв на D подмножеств так, чтобы вероятность каждого подмножества была по возможности наиболее близкой к $1/D$. Припишем различные начальные кодовые буквы каждому из этих подмножеств. Затем разобьем вновь каждое подмножество на D приближенно равновероятных групп и припишем вторую букву, соответствующую этому разделению. Продолжим этот процесс до тех пор, пока каждая группа не будет содержать только одну букву источника. Полученный в результате код удовлетворяет, очевидно, свойству префикса. Эта процедура не обязательно минимизирует $\bar{n}$, так как достижение большого значения средней собственной информации на одной кодовой букве может привести к обедненному выбору для последующих кодовых букв. Заметим, что если это разбиение может быть выполнено так, что группы будут в точности равновероятными на каждом этапе, то вероятности букв источника и длины кодовых слов будут связаны равенством

$$P(a_k) = D^{-n_k}. \tag{3.2.2}$$

До того как обратиться к деталям процедуры минимизации средней длины множества кодовых слов, исследуем ограничения на длины кодовых слов префиксного кода.

Теорема 3.2.1. [Неравенство Крафта (1949)]. Если целые числа n_1, n_2, ..., n_K удовлетворяют неравенству

$$\sum_{k=1}^{K} D^{-n_k} \leqslant 1, \tag{3.2.3}$$

то существует код, обладающий свойством префикса, с алфавитом объема D, длины кодовых слов в котором равны этим числам. Обратно, длины кодовых слов любого кода, обладающего свойством префикса, удовлетворяют неравенству (3.2.3). (*Замечание*. Теорема не утверждает, что любой код с длинами кодовых слов, удовлетворяющими (3.2.3), является префиксным. Так, например, множество двоичных кодовых слов **0**, **00**, **11** удовлетворяет (3.2.3), но не обладает свойством префикса. Теорема *утверждает, что существует некоторый* префиксный код с такими длинами, например, **0**, **10** и **11**).

Доказательство. Назовем *полным деревом* порядка n с алфавитом объема D дерево, содержащее D^n концевых узлов порядка n, в котором D узлов порядка i появляются из каждого узла порядка $i-1$ для каждого i, $1 \leqslant i \leqslant n$. Заметим, что доля D^{-1} узлов любого порядка $i \geqslant 1$ исходит из каждого из D узлов первого порядка. Точно так же доля D^{-2} узлов любого порядка $i \geqslant 2$ исходит из каждого из D^2 узлов порядка 2 и доля D^{-i} узлов любого порядка, большего или равного i, исходит из каждого из D^i узлов порядка i.

Пусть теперь n_1, ..., n_K удовлетворяют (3.2.3). Покажем, как построить префиксный код с этими длинами, исходя из полного дерева порядка n, равного наибольшему из n_k, и полагая различные узлы в этом полном дереве концевыми узлами кодового дерева. Так что, когда по-

строение будет закончено, кодовое дерево будет вложено в полное дерево, как показано на рис. 3.2.3. Для простоты обозначений, предположим, что n_k упорядочены в порядке возрастания $n_1 \leqslant n_2 \leqslant \ldots \leqslant n_K$. Выберем какой-либо узел порядка n_1, допустим $\mathbf{x}_1$, в полном дереве в качестве первого концевого узла кодового дерева. Все узлы полного дерева любого порядка, большего или равного n_1, еще можно использовать как концевые узлы кодового дерева, за исключением доли D^{-n_1} узлов, которые исходят из узла $\mathbf{x}_1$. Далее выберем какой-нибудь оставшийся узел порядка n_2, например $\mathbf{x}_2$, в качестве следующего концевого узла кодового дерева. Все узлы полного дерева любого порядка, большего или равного n_2, еще можно использовать как концевые узлы, за исключением доли $D^{-n_1} + D^{-n_2}$ узлов, которые исходят из узлов $\mathbf{x}_1$ и $\mathbf{x}_2$. Продолжая этот процесс, после образования k-го концевого узла кодового дерева все узлы полного дерева любого порядка, большего или равного n_k, все еще можно использовать, за исключением доли $\sum_{i=1}^{k} D^{-n_i}$, исходящей из $\mathbf{x}_1, \ldots, \mathbf{x}_k$. Согласно (3.2.3) эта доля всегда строго меньше, чем 1 при $k < K$ и, следовательно, всегда существует некоторый оставшийся узел, который может быть рассмотрен как следующий концевой узел.

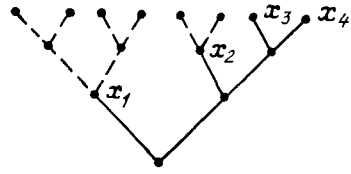


Рис. 3.2.3. Двоичное кодовое дерево (сплошные линии), дополненное до полного дерева (пунктирные линии) порядка 3.

Чтобы доказать вторую часть теоремы, заметим, что кодовое дерево, соответствующее любому префиксному коду, может быть вложено в полное дерево, порядок которого равен наибольшей длине кодового слова. Из концевого узла порядка n_k кодового дерева исходит доля D^{-nk} концевых узлов полного дерева. Так как множества концевых узлов полного дерева, исходящие из различных концевых узлов кодового дерева, являются непересекающимися, то эти доли в сумме не могут превосходить 1, давая (3.2.3). |

Докажем теперь теорему о длинах кодовых слов однозначно декодируемых кодов, которая оправдает наше рассмотрение кодов, обладающих свойством префикса.

Теорема 3.2.2*). Пусть задан код с длинами кодовых слов $n_1, \ldots, n_K$, с кодовым алфавитом из D символов. Если код однозначно декодируем, неравенство Крафта (3.2.3) удовлетворяется.

Доказательство. Пусть L — произвольное положительное целое число. Рассмотрим тождество

$$\left(\sum_{k=1}^{K} D^{-n_k}\right)^L = \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \cdots \sum_{k_L=1}^{K} D^{-[n_{k_1}+n_{k_2}+\cdots+n_{k_L}]}. \qquad (3.2.4)$$

*) Эта теорема принадлежит Макмиллану (1956). Приведенное доказательство принадлежит Карушу (1961).

Заметим, что в выражении, стоящем в правой части (3.2.4), различные слагаемые соответствуют каждой возможной последовательности из L кодовых слов. Более того, сумма $n_{k_1} + n_{k_2} + \ldots + n_{k_L}$ равна общему числу кодовых букв, соответствующему последовательности кодовых слов. Следовательно, если через A_i обозначить число последовательностей из L кодовых слов, имеющих общую длину i — число кодовых букв, то (3.2.4) можно переписать в виде

$$\left[\sum_{k=1}^{K} D^{-n_k}\right]^L = \sum_{i=1}^{Ln_{max}} A_i D^{-i}, \tag{3.2.5}$$

где n_{max} является наибольшим из n_k.

Если код однозначно декодируем, то все последовательности кодовых слов из i кодовых букв являются различными и, следовательно, $A_i \leqslant D^i$. Подставляя это неравенство в (3.2.5), получаем

$$\left[\sum_{k=1}^{K} D^{-n_k}\right]^L \leqslant \sum_{i=1}^{Ln_{max}} 1 = Ln_{max}, \tag{3.2.6}$$

$$\sum_{k=1}^{K} D^{-n_k} \leqslant (Ln_{max})^{1/L}. \tag{3.2.7}$$

Неравенство (3.2.7) справедливо для всех положительных целых чисел L; переходя к пределу при $L \to \infty$, получаем (3.2.3), что доказывает теорему.

Так как длины кодовых слов любого однозначно декодируемого кода удовлетворяют (3.2.3) и так как можно построить префиксный код для любого множества длин, удовлетворяющих (3.2.3), то любой однозначно декодируемый код можно заменить на префиксный код без изменения длин кодовых слов. Таким образом, последующие теоремы относительно средней длины кодового слова приложимы как к однозначно декодируемым кодам, так и к подклассу префиксных кодов.

3.3. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ ИСТОЧНИКА

Теорема 3.3.1. При заданных конечном ансамбле источника U с энтропией $H(U)$ и кодовом алфавите из D символов можно так приписать кодовые слова буквам источника, что будет выполняться свойство префикса и средняя длина кодового слова $\bar{n}$ будет удовлетворять условию

$$\bar{n} < \frac{H(U)}{\log D} + 1. \tag{3.3.1}$$

Более того для любого однозначно декодируемого множества кодовых слов

$$\bar{n} \geqslant \frac{H(U)}{\log D}. \tag{3.3.2}$$

Доказательство. Покажем вначале справедливость (3.3.2), установив, что

$$H(U)-\bar{n}\log D\leqslant 0. \tag{3.3.3}$$

Пусть $P(a_1)$, ..., $P(a_K)$ — вероятности букв источника и пусть n_1, ..., n_K — длины кодовых слов. Имеем

$$H(U)-\bar{n}\log D=\sum_{k=1}^{K}P(a_k)\log\frac{1}{P(a_k)}-\sum_{k=1}^{K}P(a_k)n_k\log D. \tag{3.3.4}$$

Помещая $-n_k$ под знак логарифма и объединяя слагаемые, будем иметь

$$H(U)-\bar{n}\log D=\sum_{k=1}^{K}P(a_k)\log\frac{D^{-n_k}}{P(a_k)}. \tag{3.3.5}$$

Используя неравенство $\log z\leqslant(z-1)\log e$ при $z>0$, получаем

$$H(U)-\bar{n}\log D\leqslant(\log e)\left[\sum_{k=1}^{K}D^{-n_k}-\sum_{k=1}^{K}P(a_k)\right]\leqslant 0. \tag{3.3.6}$$

Последнее неравенство в (3.3.6) следует из неравенства Крафта (3.2.3), которое справедливо для любого однозначно декодируемого кода. Это доказывает (3.3.2). Заметим, что равенство в (3.3.2) имеет место тогда и только тогда, когда

$$P(a_k)=D^{-n_k},\quad 1\leqslant k\leqslant K. \tag{3.3.7}$$

Это условие совпадает с ранее полученным условием (3.2.2) для каждой кодовой буквы и приводит к максимуму энтропии.

Покажем далее, как выбрать код, удовлетворяющий (3.3.1). Если бы длины кодовых слов не обязательно были целыми числами, то можно было бы просто подобрать n_k, чтобы удовлетворить условию (3.3.7). Однако можно приближенно удовлетворить (3.3.7), выбирая целые числа n_k так, чтобы удовлетворялись неравенства

$$D^{-n_k}\leqslant P(a_k)<D^{-n_k+1},\quad 1\leqslant k\leqslant K. \tag{3.3.8}$$

Суммирование (3.3.8) по k превращает левое неравенство в неравенство Крафта (3.2.3), и существует префиксный код с этими длинами. Логарифмируя правое неравенство в (3.3.8), получаем

$$\log P(a_k)<(-n_k+1)\log D,\quad n_k<\frac{-\log P(a_k)}{\log D}+1. \tag{3.3.9}$$

Умножая (3.3.9) на $P(a_k)$ и суммируя по всем k, получаем (3.3.1), что завершает доказательство теоремы. |

Можно получить более сильные результаты, если кодовые слова приписывать не отдельным буквам источника, а прямо последовательностям L букв источника.

Теорема 3.3.2. Для заданных дискретного источника без памяти U с энтропией $H(U)$ и кодового алфавита из D символов возможно так приписать кодовые слова последовательностям L букв источника, что

будет выполняться свойство префикса и средняя длина кодовых слов на одну букву источника $\bar{n}$ будет удовлетворять условию

$$\frac{H(U)}{\log D} \leqslant \bar{n} < \frac{H(U)}{\log D} + \frac{1}{L}. \tag{3.3.10}$$

Более того, левое неравенство справедливо для любого однозначно декодируемого множества кодовых слов.

Доказательство. Рассмотрим произведение ансамблей для последовательностей L букв источника. Энтропия произведения ансамблей равна $LH(U)$, а средняя длина кодовых слов равна $\bar{n}L$, где $\bar{n}$ означает среднюю длину на одну букву источника. Теорема 3.3.1 утверждает, что минимально достижимое значение $\bar{n}L$ (при сопоставлении кодового слова неравномерного кода каждой последовательности L букв источника) удовлетворяет неравенствам

$$\frac{LH(U)}{\log D} \leqslant \bar{n}L < \frac{LH(U)}{\log D} + 1.$$

Разделив эти выражения на L, получим результат теоремы. |

Теорема 3.3.2 очень похожа на теорему 3.1.1. Если выбрать L достаточно большим и затем применить закон больших чисел к длинной последовательности, состоящей, в свою очередь, из последовательностей L букв источника, то можно заметить, что из теоремы 3.3.2 следует первая часть теоремы 3.1.1. Однако теорема 3.3.2 немного сильнее, чем теорема 3.1.1, так как она предлагает относительно простой метод кодирования, а также альтернативу случайным ошибкам, а именно случайные длинные задержки.

3.4. ПРОЦЕДУРА ВЫБОРА ОПТИМАЛЬНОГО НЕРАВНОМЕРНОГО КОДА

В этом параграфе будет дана предложенная Д. А. Хаффманом (1952) конструктивная процедура отыскания оптимального множества кодовых слов для кодирования данного множества сообщений. Под оптимальностью будет подразумеваться то, что никакое другое однозначно декодируемое множество кодовых слов не имеет меньшую среднюю длину кодового слова, чем заданное множество. Множество длин, задаваемое (3.3.8), обычно не минимизирует $\bar{n}$, даже если на нем достигается граница в теореме 3.3.1. Вначале будут рассмотрены двоичные коды и затем будет дано обобщение на произвольный кодовый алфавит.

Пусть буквы источника $a_1, \ldots, a_K$ имеют вероятности $P(a_1), \ldots, P(a_K)$; предположим для простоты обозначений, что буквы упорядочены так, что $P(a_1) \geqslant P(a_2) \geqslant \ldots \geqslant P(a_K)$. Пусть $\mathbf{x}_1, \ldots, \mathbf{x}_K$ — множество двоичных кодовых слов для этого источника и пусть $n_1, \ldots, n_K$ — длины кодовых слов. Кодовые слова $\mathbf{x}_1, \ldots, \mathbf{x}_K$, соответствующие оптимальному коду, в общем случае не являются единственными и часто длины $n_1, \ldots, n_K$ не являются единственными (см. задачу 3.13). Далее будут получены некоторые условия, которым должен удовлетворять по крайней мере один оптимальный код, и затем будет показано, как построить код, удовлетворяющий этим условиям. Мы ограничимся рас-

смотрением префиксных кодов, так как любое множество длин, получаемых на однозначно декодируемом коде, можно получить на префиксном коде.

Лемма 1. Для любого заданного источника с $K \geqslant 2$ буквами существует оптимальный двоичный код, в котором два наименее вероятных кодовых слова $\mathbf{x}_K$ и $\mathbf{x}_{K-1}$ имеют одну и ту же длину и отличаются лишь последним символом: $\mathbf{x}_K$ оканчивается на **1**, а $\mathbf{x}_{K-1}$ на **0**.

Доказательство. Во-первых, заметим, что по крайней мере для одного оптимального кода n_K больше или равно остальным длинам кодовых слов. Чтобы показать это, рассмотрим код, в котором $n_K < n_i$ для некоторого i. Если кодовые слова $\mathbf{x}_i$ и $\mathbf{x}_K$ заменить одно другим, то $\bar{n}$ изменится на величину

$$\Delta = P(a_i)\, n_K + P(a_K)\, n_i - P(a_i)\, n_i - P(a_K)\, n_K = \\ = [P(a_i) - P(a_K)]\, [n_K - n_i] \leqslant 0. \qquad (3.4.1)$$

Следовательно, любой код может быть изменен так, чтобы сделать n_K максимальной длиной, не увеличивая $\bar{n}$. Далее заметим, что в любом оптимальном коде, для которого n_K является наибольшей длиной, должно быть другое кодовое слово, отличающееся от $\mathbf{x}_K$ лишь в последнем символе, в противном случае последний символ $\mathbf{x}_K$ мог бы быть отброшен без нарушения свойства префикса. Наконец, если $\mathbf{x}_i$ есть кодовое слово, отличающееся от $\mathbf{x}_K$ лишь в одной позиции, то должно быть $n_i \geqslant n_{K-1}$ и, как показывает (3.4.1), $\mathbf{x}_i$ и $\mathbf{x}_{K-1}$ можно поменять местами без увеличения $\bar{n}$. Таким образом, существует оптимальный код, в котором $\mathbf{x}_K$ и $\mathbf{x}_{K-1}$ отличаются лишь в последнем символе. Эти слова, если необходимо, можно поменять местами, чтобы получить $\mathbf{x}_K$ оканчивающимся на **1**. |

С помощью этой леммы задача построения оптимального кода сводится к задаче построения $\mathbf{x}_1, \ldots, \mathbf{x}_{K-2}$ и отысканию первых $n_K - 1$ символов $\mathbf{x}_K$. Определим теперь редуцированный ансамбль U' как ансамбль, состоящий из букв $a'_1, a'_2, \ldots, a'_{K-1}$ с вероятностями

$$\Pr(a'_k) = \begin{cases} P(a_k), & k \leqslant K-2, \\ P(a_{K-1}) + P(a_K), & k = K-1. \end{cases} \qquad (3.4.2)$$

Любой префиксный код для U' можно превратить в соответствующий префиксный код для U простым добавлением концевого символа **0** к $\mathbf{x}'_{K-1}$ для получения $\mathbf{x}_{K-1}$ и добавлением концевого символа **1** к $\mathbf{x}'_{K-1}$ для получения $\mathbf{x}_K$. Это устанавливает взаимно однозначное соответствие между множеством префиксных кодов для U' и множеством тех префиксных кодов для U, в которых $\mathbf{x}_K$ и $\mathbf{x}_{K-1}$ отличаются лишь последним символом; $\mathbf{x}_K$ кончается на **1**, а $\mathbf{x}_{K-1}$ кончается на **0**.

Лемма 2. Если некоторый префиксный код для U' является оптимальным, то соответствующий ему префиксный код для U является оптимальным.

Доказательство. Длины n'_k кодовых слов для U' связаны с длинами n_k соответствующего кода для U соотношениями

$$n_k = \begin{cases} n'_k; & k \leqslant K-2, \\ n'_{K-1}+1; & k = K-1,\ K. \end{cases} \tag{3.4.3}$$

Следовательно, средние длины $\bar{n}'$ и $\bar{n}$ связаны равенством

$$\bar{n} = \sum_{k=1}^{K} P(a_k)\, n_k = \sum_{k=1}^{K-2} P(a_k)\, n'_k + [P(a_{K-1}) + P(a_K)]\,(n'_{K-1}+1) =$$
$$= \sum_{k=1}^{K-2} \Pr(a'_k)\, n'_k + \Pr(a'_{K-1})\,[n'_{K-1}+1] = \bar{n}' + \Pr(a'_{K-1}). \tag{3.4.4}$$

Так как $\bar{n}$ и $\bar{n}'$ отличаются лишь на фиксированное число, не зависящее от кода для U', то можно минимизировать $\bar{n}$ в классе кодов, у которых $\mathbf{x}_K$ и $\mathbf{x}_{K-1}$ отличаются лишь в последнем символе, минимизируя $\bar{n}'$. Но согласно лемме 1 код из этого класса минимизирует $\bar{n}$ по всем кодам, обладающим свойством префикса. |

Задача отыскания оптимального кода сведена теперь к задаче отыскания оптимального кода для ансамбля, имеющего на одно сообщение меньше. Но теперь редуцированный ансамбль может иметь свои два наименее вероятные сообщения сгруппированными вместе, и может быть произведен следующий редуцированный ансамбль. Продолжая таким образом, мы постепенно достигнем того, что получится ансамбль, состоящий только из двух сообщений, и тогда оптимальный код, очевидно, получается приписыванием **1** одному сообщению и **0** другому.

Систематическая процедура для выполнения описанных выше операций приведена на рис. 3.4.1. Вначале свяжем вместе два наименее вероятных сообщения (в рассматриваемом случае a_4 и a_5), полагая, что последним символом для a_4 является **0** и последним символом a_5 является **1**. Складывая вероятности сообщений a_4 и a_5, находим, что два наименее вероятных сообщения в редуцированном ансамбле будут a_3 и a'_4. На следующем этапе двумя наименее вероятными сообщениями будут a_1 и a_2, но на этот раз остались только два сообщения. Рассматривая получающийся в результате рисунок, видим, что мы построили кодовое дерево для U, начав с наиболее удаленных ребер и спускаясь вниз с основанию. Кодовые слова считаются по этому дереву справа налево.

Имеется некоторая трудность при распространении этой процедуры на недвоичные коды. Лемма 1 остается справедливой для недвоичного кодового алфавита. А лемма 2 не остается справедливой и возникает вопрос, имеются ли еще кодовые слова помимо $\mathbf{x}_{K-1}$, которые должны отличаться от $\mathbf{x}_K$ в последнем символе.

Определим полное кодовое дерево как конечное кодовое дерево, в котором из каждого промежуточного узла исходят D узлов следующего более высокого порядка. Как можно заметить из доказательства неравенства Крафта, для полного кодового дерева неравенство Крафта удовлетворяется с равенством.

Л е м м а 3. Число концевых узлов полного кодового дерева с алфавитом объема D имеет вид $D + m(D - 1)$, где m — некоторое целое число.

Доказательство. Наименьшее полное дерево с алфавитом объема D имеет D концевых узлов. Если один из этих концевых узлов превращается в промежуточный узел, то образуется D новых концевых узлов и один узел теряется, в результате получаем приращение $D - 1$. Так как любое полное дерево может быть построено с помощью таких последовательных преобразований концевых узлов в промежуточные узлы и так как каждое такое преобразование увеличивает число узлов на $D - 1$, то окончательное число узлов должно иметь вид $D + m(D - 1)$. |

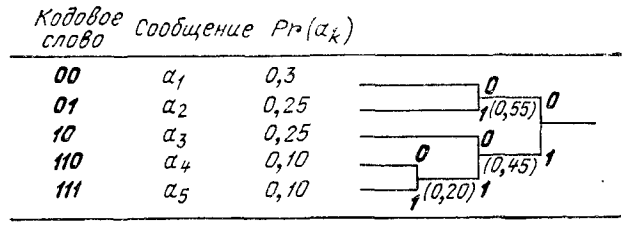


Рис. 3.4.1. Процедура кодирования Хаффмана.

Условимся теперь рассматривать каждое кодовое дерево как полное дерево, может быть, с некоторым числом B неиспользуемых концевых узлов, добавленных для полноты дерева. Ясно, что в оптимальном коде все неиспользуемые концевые узлы имеют ту же самую длину, что и самое длинное кодовое слово. Точно так же, меняя местами используемые и неиспользуемые концевые узлы, можно добиться того, чтобы неиспользуемые концевые узлы отличались лишь в последнем символе. Следовательно, оптимальное кодовое дерево должно иметь не более чем $D - 2$ неиспользуемых концевых узлов, так как если бы $D - 1$ неиспользуемых узлов группировались вместе, то соответствующее кодовое слово могло бы быть укорочено без нарушения свойства префикса. Число кодовых слов, сложенное с числом неиспользуемых узлов, имеет вид $D + m(D - 1)$, это выражение однозначно определяет число неиспользуемых узлов полного дерева. Например, если $D = 3$, то каждое полное дерево имеет нечетное число узлов. Если K четно, то число неиспользуемых концевых узлов B для оптимального кода равно 1, а если $K > 2$ нечетно, то $B = 0$.

Для читателя, который лучше ориентируется в языке формул, мы получим точное выражение для B, замечая, что $B + K = m(D - 1) + D$. Отсюда $K - 2 = m(D - 1) + (D - 2 - B)$. Для оптимального кода $0 \leqslant B \leqslant D - 2$ и поэтому $0 \leqslant D - 2 - B \leqslant D - 2$, следовательно, $D - 2 - B$ равно остатку от деления $K - 2$ на $D - 1$; обозначим его через $R_{D-1}(K - 2)$. Отсюда $B = D - 2 - R_{D-1}(K - 2)$. Используя те же рассуждения, что и в лемме 1, получаем, что существует оптимальный код, у которого имеются B неиспользуемых

узлов и $D - B$ наименее вероятных узлов, соответствующих кодовым словам, отличающихся лишь в последнем символе. Таким образом, первый этап процедуры декодирования состоит в группировании $D - B = 2 + R_{D-1}(K - 2)$ наименее вероятных узлов.

После этого начального этапа построение оптимального кода проходит так же, как и раньше. Производится редуцированный ансамбль с помощью объединения вероятностей предварительно сгруппированных кодовых слов. Легко проверить, что число сообщений в редуцированном ансамбле имеет вид $D + m(D - 1)$ и мы полагаем, что D менее вероятных из них отличаются лишь в последнем символе. Продолжая

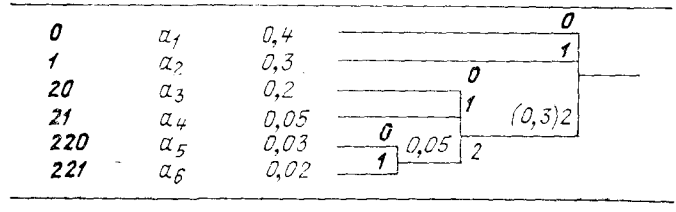


Рис. 3.4.2. Кодирование по Хаффману при $D=3$.

процедуру таким образом, в итоге получаем редуцированный ансамбль из D сообщений, для которого оптимальный код очевиден. Рис. 3.4.2 показывает построение в случае $D = 3$. Так как $K = 6$, то число сообщений, группируемых на начальном этапе, равно $2 + R_2(6 - 2) = 2$.

3.5. ДИСКРЕТНЫЕ СТАЦИОНАРНЫЕ ИСТОЧНИКИ

В предыдущих параграфах рассматривались дискретные источники без памяти. В этом параграфе будет изучен эффект статистической зависимости между буквами источника.

Пусть $\mathbf{u} = (\ldots, u_{-1}, u_0, u_1, \ldots)$ обозначает последовательность букв, производимую источником, в котором каждая буква u_l выбирается из дискретного алфавита. Полное вероятностное описание источника задается вероятностями $\Pr(u_{j+1}, u_{j+2}, \ldots, u_{j+L})$, определенными для всех L последовательностей всех начальных моментов j и всех последовательностей $u_{j+1}, \ldots, u_{j+L}$. При таком подходе источник представляет собой не что иное, как произвольный дискретный случайный процесс.

Дискретный источник называется *стационарным*, если вероятностное описание не зависит от начала отсчета времени. Точнее, источник называется стационарным, если

$$P_{U_1 U_2 \ldots U_L}(u'_1, u'_2, \ldots, u'_L) = P_{U_{j+1} U_{j+2} \ldots U_{j+L}}(u'_1, u'_2, \ldots, u'_L) \qquad (3.5.1)$$

для всех длин L, целых чисел j и последовательностей $u'_1, \ldots, u'_L$. Это означает, что вероятность того, что источник производит последо-

вательность $\mathbf{u}'$ на интервале от 1 до L, равна вероятности того, что производится в точности такая же последовательность на интервале от $j+1$ до $j+L$.

Дискретный источник называется периодическим, если (3.5.1) справедливо для всех j, которые являются кратными некоторого целого числа $m > 1$. Период — это наименьшее значение m, удовлетворяющее этому условию. Если рассмотреть m-блоки букв периодического источника с периодом m как «супербуквы» большего алфавита, то последовательность супербукв окажется стационарной. По этой причине в дальнейшем рассматриваются только стационарные источники.

Пусть $\mathbf{u}_L = (u_1, \ldots, u_L)$ — последовательность L букв дискретного стационарного источника и пусть $U_1 U_2 \ldots U_L$ — совместный ансамбль для $\mathbf{u}_L$. Определим теперь энтропию на букву источника для дискретного стационарного источника. Имеются два подхода, которые мы можем принять, и, к счастью, оба они приводят к одному и тому же результату. При первом подходе следует определить энтропию на букву в последовательности L букв как

$$H_L(U) = \frac{1}{L} H(U_1 U_2 \ldots U_L) = \overline{\frac{1}{L} \log \frac{1}{\Pr(\mathbf{u}_L)}}. \qquad (3.5.2)$$

Теперь можно было бы определить энтропию на букву источника как предел $H_L(U)$ при $L \to \infty$. При втором подходе следует определить условную энтропию L-й буквы в последовательности при условии, что заданы первые $L-1$ букв, $H(U_L \mid U_1 \ldots U_{L-1})$ и затем определить энтропию на букву источника как

$$\lim_{L \to \infty} H(U_L \mid U_1 \ldots U_{L-1}).$$

Следующая теорема утверждает, в частности, что оба этих предела существуют и равны друг другу.

Теорема 3.5.1. Для дискретного стационарного источника с $H_1(U) < \infty$ имеем

(а) $H(U_L \mid U_1 \ldots U_{L-1})$ не возрастает с L,

(б) $H_L(U) \geqslant H(U_L \mid U_1 \ldots U_{L-1})$, (3.5.3)

(в) $H_L(U)$ не возрастает с L,

(г) $\lim_{L \to \infty} H_L(U) = \lim_{L \to \infty} H(U_L \mid U_1 \ldots U_{L-1})$. (3.5.4)

Доказательство. Используя вначале то, что наложение условия не может увеличить энтропию (2.3.13), а также используя стационарность источника, получаем при $L > 2$

$$H(U_L \mid U_1 \ldots U_{L-1}) \leqslant H(U_L \mid U_2 \ldots U_{L-1}) = \\ = H(U_{L-1} \mid U_1 \ldots U_{L-2}). \qquad (3.5.5)$$

Это доказывает утверждение (а).

Используем цепное равенство (2.2.30) для разложения $H_L(U)$,

$$H_L(U)=\frac{1}{L}[H(U_1)+H(U_2|U_1)+\ldots+H(U_L|U_1\ldots U_{L-1})]. \qquad (3.5.6)$$

Согласно утверждению (а) последнее слагаемое в (3.5.6) является границей снизу каждого из L слагаемых. Применяя эту границу, получаем утверждение (б).

Согласно определению $H_L(U)$ имеем

$$H_L(U)=\frac{1}{L}H(U_1\ldots U_{L-1})+\frac{1}{L}H(U_L|U_1\ldots U_{L-1})\leqslant$$
$$\leqslant\frac{L-1}{L}H_{L-1}(U)+\frac{1}{L}H_L(U). \qquad (3.5.7)$$

После простых преобразований получаем $H_L(U)\leqslant H_{L-1}(U)$, устанавливая справедливость утверждения (в).

Так как $H_L(U)$ и $H(U_L|U_1\ldots U_{L-1})$ являются неотрицательными и невозрастающими с L, то оба предела существуют. Обозначим $\lim\limits_{L\to\infty}H_L(U)$ через H_∞. Используя опять цепное равенство, получаем

$$H_{L+j}(U)=\frac{1}{L+j}H(U_1\ldots U_{L-1})+\frac{1}{L+j}[H(U_L|U_1\ldots U_{L-1})+$$
$$+H(U_{L+1}|U_1\ldots U_L)+\ldots+H(U_{L+j}|U_1\ldots U_{L+j-1})]\leqslant$$
$$\leqslant\frac{1}{L+j}H(U_1\ldots U_{L-1})+\frac{j+1}{L+j}H(U_L|U_1\ldots U_{L-1}). \qquad (3.5.8)$$

Здесь было использовано то, что первое слагаемое в квадратных скобках является верхней границей для каждого из остальных слагаемых. Переходя в (3.5.8) к пределу при $j\to\infty$, получаем

$$H_\infty(U)\leqslant H(U_L|U_1\ldots U_{L-1}). \qquad (3.5.9)$$

Так как (3.5.9) справедлива при всех L, то будем иметь

$$H_\infty(U)\leqslant\lim_{L\to\infty}H(U_L|U_1\ldots U_{L-1}). \qquad (3.5.10)$$

Неравенства (3.5.10) и (3.5.3) устанавливают справедливость (3.5.4), что завершает доказательство теоремы. |

Теорема 3.5.2. (Теорема кодирования для источника неравномерным кодом.) Пусть $H_L(U)$ — энтропия на букву в последовательности длины L дискретного источника с алфавитом объема K. При заданном кодовом алфавите с D символами можно так закодировать последовательности из L букв источника префиксным кодом, что среднее число кодовых букв на букву источника $\bar{n}$ будет удовлетворять неравенствам

$$\frac{H_L(U)}{\log D}\leqslant\bar{n}<\frac{H_L(U)}{\log D}+\frac{1}{L}. \qquad (3.5.11)$$

Более того, левое неравенство справедливо для любого однозначно декодируемого множества кодовых слов для последовательностей L

букв источника. И, наконец, если источник является стационарным, то для любого $\delta > 0$ можно выбрать L столь большим, чтобы $\bar{n}$ удовлетворяло неравенствам

$$\frac{H_\infty(U)}{\log D} \leqslant \bar{n} < \frac{H_\infty(U)}{\log D} + \delta, \tag{3.5.12}$$

и левое неравенство для $\bar{n}$ никогда не нарушается для однозначно декодируемого кода.

Доказательство. Доказательство (3.5.11) аналогично доказательству соответствующего условия в теореме 3.3.2, за исключением того, что энтропия последовательности L букв источника равна $LH_L(U)$, а не $LH(U)$. При переходе к пределу в (3.5.11) при $L \to \infty$ $H_L(U)$ стремится к $H_\infty(U)$ и $1/L$ стремится к **0**, что доказывает (3.5.12) .|

При рассмотрении дискретных источников без памяти наш интерес к $\bar{n}$ был обусловлен законом больших чисел, который показывает, что число кодовых букв на букву источника в длинной последовательности кодовых слов стремится к $\bar{n}$. Следующий пример показывает, что это предельное поведение не обязательно имеет место для произвольных дискретных стационарных источников. Предположим, что источник с алфавитом (a_1, a_2, a_3) имеет два типа поведения, каждый из которых происходит с вероятностью $1/2$. При первом типе источник производит бесконечную последовательность повторений a_1. При втором типе источник производит бесконечную последовательность статистически независимых равновероятных выборок букв a_2 и a_3. Если закодировать последовательности L букв источника двоичным кодом, то легко увидеть, что $\bar{n}$ минимизируется отображением последовательности букв a_1 в один-единственный двоичный символ и отображением каждой из 2^L последовательностей букв a_2 и a_3 в кодовые слова длины $L + 1$. Так как тип поведения источника никогда не меняется, то либо все кодовые слова последовательности будут иметь длину 1 либо все будут иметь длину $L + 1$. Для таких источников ни $\bar{n}$, ни энтропия не являются величинами, которые играют значительную роль.

Источники, которые не могут иметь различные устойчивые типы поведения, называются *эргодическими* источниками. Для того чтобы определить эргодичность более точно, предположим, что $\mathbf{u} = \ldots, u_{-1}, u_0, u_1, \ldots$ — бесконечная последовательность букв источника и пусть $T^l\mathbf{u}$ обозначает последовательность, сдвинутую по времени на l позиций. Т. е. если обозначить $T^l\mathbf{u}$ через $\mathbf{u}'$, то имеем

$$u'_n = u_{n+l}, \quad -\infty < n < \infty.$$

Аналогично, если S — множество бесконечных последовательностей букв источника, то T^lS обозначает то же самое множество, сдвинутое на l позиций, т. е. если $\mathbf{u}' = T^l\mathbf{u}$, то $\mathbf{u}'$ принадлежит множеству T^lS тогда и только тогда, когда $\mathbf{u}$ принадлежит S. Множество последовательностей называется *инвариантным*, если $TS = S$. Легко можно заметить, что множество всех последовательностей дискретного источ-

ника инвариантно, а также то, что для любого $\mathbf{u}$ множество ..., $T^{-1}\mathbf{u}$, $\mathbf{u}$, $T\mathbf{u}$, $T^2\mathbf{u}$, ... инвариантно. *Дискретный стационарный источник называется эргодическим, если любое измеримое инвариантное множество последовательностей имеет либо вероятность* 1 *либо вероятность* 0. Можно заметить, что в предыдущем примере множества последовательностей в каждом из описанных типов поведений были инвариантными множествами и вероятности каждого из них были равны $^1/_2$. Следовательно, этот источник не был эргодическим.

Хотя приведенное выше определение является весьма изящным, с ним иногда довольно трудно работать, и оно не дает интуитивного понимания эргодичности. Следующее определение эквивалентно данному ранее. Пусть $f_n(\mathbf{u})$ является функцией бесконечной последовательности источника $\mathbf{u}$, которая зависит только от конечной последовательности u_1, ..., u_n букв источника. Дискретный стационарный источник является *эргодическим* тогда и только тогда, когда для всех $n \geqslant 1$ и всех $f_n(\mathbf{u})$, для которых $|\overline{f_n(\mathbf{u})}| < \infty$, имеет место соотношение

$$\lim_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} f_n(T^l \mathbf{u}) = \overline{f_n(\mathbf{u})} \tag{3.5.13}$$

для всех последовательностей источника $\mathbf{u}$, за исключением множества вероятности 0. Класс функций в этом определении может быть расширен до всех измеримых функций $f(\mathbf{u})$, для которых $|\overline{f(\mathbf{u})}| < \infty$, или может быть сужен до частного класса функций $f_{\mathbf{u}'_n}(\mathbf{u})$, где $\mathbf{u}'_n$ — фиксированная последовательность u'_1, ..., u'_n букв и

$$f_{\mathbf{u}'_n}(\mathbf{u}) = \begin{cases} 1, & \text{если } u_1 = u'_1,\ u_2 = u'_2, \ldots,\ u_n = u'_n, \\ 0 & \text{во всех остальных случаях.} \end{cases} \tag{3.5.14}$$

Доказательство эквивалентности этих определений эргодичности содержится у Хинчина (1956); Вольфовиц (1961), лемма 10.3.1, рассмотрел другое определение и доказал его эквивалентность приведенным здесь.

Определение (3.5.13) особенно важно, так как оно касается свойства эргодических источников, которое нам понадобится в дальнейшем. Соотношение (3.5.13) означает, что закон больших чисел применим к эргодическим источникам. Иначе это можно выразить как среднее по времени, т. е. усреднение по времени по какой-нибудь выборке выхода источника (исключение составляет множество нулевой вероятности), равно среднему по ансамблю $\overline{f_n(\mathbf{u})}$. Так как $\overline{f_{\mathbf{u}'_n}(\mathbf{u})}$ просто равно вероятности последовательности $\mathbf{u}'_n$, то (3.5.14) утверждает, что относительная частота появления $\mathbf{u}'_n$ в очень длинной последовательности источника будет приближенно равна вероятности $\mathbf{u}'_n$.

К сожалению, свойства эргодичности не хватает для того, чтобы число кодовых букв на букву источника в неравномерном коде стремилось к $\bar{n}$. Если кодировать сразу L букв источника и если через $n(u_1, \ldots, u_L)$ обозначить длину кодового слова, то среднее по времени число

кодовых букв на букву источника равно

$$\lim_{J\to\infty}\frac{1}{LJ}\sum_{j=0}^{J-1} n\,(u_{Lj+1},\ \ldots,\ u_{Lj+L}). \tag{3.5.15}$$

В задаче 3.21 приведен пример эргодического источника, в котором это среднее, рассматриваемое как случайная величина, принимает различные значения с ненулевыми вероятностями. Трудность состоит в том, что (3.5.15) представляет собой среднее по времени в ином смысле, чем (3.5.13), так как оно определено с помощью сдвигов на L букв сразу, а не сдвигов на одну букву.

К счастью, теорема 3.1.1 остается справедливой для произвольных эргодических источников. Главная трудность в доказательстве теоремы состоит в установлении справедливости закона больших чисел для собственной информации, т. е. в доказательстве того, что с большой вероятностью $I\,(\mathbf{u}_L)/L$ близко к $H_\infty\,(U)$ для больших L. Этот закон больших чисел представляет значительный математический и теоретико-информационный интерес, и он будет сформулирован в виде теоремы.

Теорема 3.5.3. [**Макмиллан (1953).**] Пусть для дискретного стационарного эргодического источника $H_1\,(U) < \infty$.

Для произвольных $\varepsilon > 0$, $\delta > 0$ существует целое число $L_0\,(\varepsilon,\ \delta)$ (которое зависит от источника), такое, что при всех $L \geqslant L_0\,(\varepsilon,\ \delta)$

$$\Pr\left[\left|\frac{I\,(\mathbf{u}_L)}{L} - H_\infty\,(U)\right| > \delta\right] < \varepsilon. \tag{3.5.16}$$

До того как доказать теорему, введем некоторые необходимые обозначения и докажем две леммы. Заметим, что

$$\frac{I\,(\mathbf{u}_L)}{L} = \frac{1}{L}\sum_{l=1}^{L} I\,(u_l \mid u_1,\ \ldots,\ u_{l-1}). \tag{3.5.17}$$

Отметим, что правая часть (3.5.17) очень похожа на среднее по времени, задаваемое (3.5.13). Отличие состоит в том, что каждое слагаемое в (3.5.17), являющееся собственной информацией, зависит от разного числа предыдущих букв источника. Центральным местом доказательства является установление того, что эта зависимость убывает достаточно быстро, когда l стремится к бесконечности. Пусть $P\,(\mathbf{u}_L) = P\,(u_1,\ \ldots,\ u_L)$ обозначает вероятность последовательности L букв источника; определим для любого целого числа $1 \leqslant m < L$ величину $Q_m\,(\mathbf{u}_L)$ как

$$Q_m\,(\mathbf{u}_L) = P\,(\mathbf{u}_m)\prod_{l=m+1}^{L} P\,(u_l \mid u_{l-1},\ \ldots,\ u_{l-m}). \tag{3.5.18}$$

Другими словами, $Q_m\,(\mathbf{u}_L)$ является приближением вероятностной меры источника, которое учитывает статистические зависимости лишь m прошлых букв. Отметим, что

$$\sum_{\mathbf{u}_L} Q_m\,(\mathbf{u}_L) = 1.$$

Это можно показать суммированием вначале по $\mathbf{u}_L$, затем по $\mathbf{u}_{L-1}$ и так далее и, наконец, по $\mathbf{u}_1$.

Лемма 1. Для дискретного стационарного эргодического источника с $H_1(U) < \infty$ и для произвольного $m \geqslant 1$

$$\lim_{L\to\infty} \frac{1}{L} \log Q_m(\mathbf{u}_L) = -H(U_{m+1} | U_m \ldots U_1) \tag{3.5.19}$$

с вероятностью 1.

Доказательство. Из (3.5.18) следует, что

$$\frac{1}{L} \log Q_m(\mathbf{u}_L) = \frac{1}{L} \log P(\mathbf{u}_m) + \frac{1}{L} \sum_{l=m+1}^{L} \log P(u_l | u_{l-1}, \ldots, u_{l-m}). \tag{3.5.20}$$

Так как $\log P(\mathbf{u}_m)$ не зависит от L и принимает конечные значения с вероятностью 1, то

$$\lim_{L\to\infty} \frac{1}{L} \log P(\mathbf{u}_m) = 0 \tag{3.5.21}$$

с вероятностью 1. Аналогично, так как $\log P(u_l | u_{l-1}, \ldots, u_{l-m})$ является функцией последовательности m букв и имеет конечное математическое ожидание $-H(U_{m+1} | U_m \ldots U_1)$, а также в силу того, что источник является эргодическим, (3.5.13) приводит к равенству

$$\lim_{L\to\infty} \frac{1}{L-m} \sum_{l=m+1}^{L} \log P(u_l | u_{l-1}, \ldots, u_{l-m}) = -H(U_{m+1} | U_m \ldots U_1) \tag{3.5.22}$$

с вероятностью 1. И, наконец, так как m фиксировано, то предел в (3.5.22) не меняется при замене $1/(L-m)$ на $1/L$. Теперь на основании равенств (3.5.20)—(3.5.22) получаем (3.5.19). |

Лемма 2. Для дискретного стационарного эргодического источника с $H_1(U) < \infty$, для произвольных $\varepsilon > 0$, $\delta > 0$, для достаточно большого m и для любого $L > m$

$$\Pr\left\{\left|\frac{1}{L} \log Q_m(\mathbf{u}_L) - \frac{1}{L} \log P(\mathbf{u}_L)\right| > \varepsilon\right\} \leqslant \delta. \tag{3.5.23}$$

Доказательство. Имеем:

$$\Pr\left\{\left|\frac{1}{L} \log Q_m(\mathbf{u}_L) - \frac{1}{L} \log P(\mathbf{u}_L)\right| > \varepsilon\right\} =$$

$$= \Pr\left\{\left|\log \frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}\right| > L\varepsilon\right\} \leqslant \frac{1}{L\varepsilon} \overline{\left|\log \frac{Q_m(\mathbf{u}_L)}{L(\mathbf{u}_L)}\right|}, \tag{3.5.24}$$

где было использовано неравенство Чебышева*) для неотрицательной случайной величины

$$\left|\log \frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}\right|.$$

Пусть теперь для любого числа y выражение $[y]_+$ обозначает «положительную часть» y, т. е.

$$[y]_+ = \begin{cases} y, & y \geqslant 0, \\ 0, & y < 0. \end{cases} \tag{3.5.25}$$

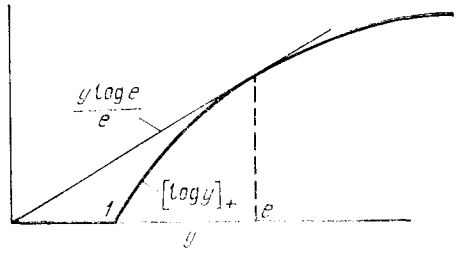


Рис. 3.5.1.

Рассматривая раздельно положительные и отрицательные значения y, легко проверить, что

$$|y| = 2[y]_+ - y. \tag{3.5.26}$$

Пользуясь рис. 3.5.1, можно также заметить, что для $y \geqslant 0$

$$[\log y]_+ \leqslant \frac{y \log \mathrm{e}}{\mathrm{e}}. \tag{3.5.27}$$

Из (3.5.26) и (3.5.27) следует, что

$$\overline{\left|\log \frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}\right|} \leqslant \frac{2 \log \mathrm{e}}{\mathrm{e}} \overline{\left[\frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}\right]} - \overline{\log \frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}}. \tag{3.5.28}$$

Для выражения, стоящего в правой части в (3.5.28), имеем

$$\overline{\left[\frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}\right]} = \sum_{\mathbf{u}_L} P(\mathbf{u}_L) \frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)} = 1, \tag{3.5.29}$$

$$-\overline{\log \frac{Q_m(\mathbf{u}_L)}{P(\mathbf{u}_L)}} = m H_m(U) + (L-m) H(U_{m+1} | U_m \cdots U_1) - L H_L(U). \tag{3.5.30}$$

Подставляя выражения (3.5.28)—(3.5.30) в (3.5.24), получим

$$\Pr\left[\left|\frac{1}{L}\log Q_m(\mathbf{u}_L) - \frac{1}{L}\log P(\mathbf{u}_L)\right| > \varepsilon\right] \leqslant$$

*) См. (5.4.5) для вывода неравенства Чебышева.

$$\leqslant \frac{1}{\varepsilon}\left[\frac{2\log e}{Le}+\frac{m}{L}H_m(U)+\right.$$
$$\left.+\left(1-\frac{m}{L}\right)H(U_{m+1}\,|\,U_m\ldots U_1)-H_L(U)\right]. \qquad (3.5.31)$$

Заметим, наконец, что из теоремы 3.5.1 следует, что выражение, стоящее в квадратных скобках в правой части (3.5.31), при $L > m$ стремится к нулю при возрастании m равномерно по $L > m$. Таким образом, для любого фиксированного ε правая часть неравенства будет меньше, чем δ для достаточно больших m. |

Доказательство теоремы. Для заданных $\varepsilon > 0$, $\delta > 0$ выберем m достаточно большим так, чтобы правая часть (3.5.31) была меньше, чем δ для всех $L > m$, и

$$|H(U_{m+1}\,|\,U_m\ldots U_1)-H_\infty(U)|<\varepsilon. \qquad (3.5.32)$$

Выберем затем достаточно большое $L_0 > m$ так, чтобы для всех $L \geqslant L_0$

$$\Pr\left[\left|\frac{1}{L}\log Q_m(\mathbf{u}_L)+H(U_{m+1}\,|\,U_m\ldots U_1)\right|>\varepsilon\right]\leqslant\delta. \qquad (3.5.33)$$

Это возможно сделать на основании леммы 1. Из неравенств (3.5.31)—(3.5.33) для $L > L_0$ имеем

$$\Pr\left[\left|\frac{1}{L}\log P(\mathbf{u}_L)+H_\infty(U)\right|>3\varepsilon\right]\leqslant 2\delta. \qquad (3.5.34)$$

Для того чтобы увидеть это, следует заметить, что, если не имеют места ни событие в левой части (3.5.31), ни событие в левой части (3.5.33), то тогда не может произойти событие в (3.5.34). Следовательно, вероятность события в (3.5.34) не больше, чем сумма вероятностей событий в (3.5.31) и (3.5.33). В силу произвольности $\varepsilon > 0$ и $\delta > 0$ это эквивалентно (3.5.16), что завершает доказательство теоремы. |

Теорема 3.1.1 с заменой $H(U)$ на $H_\infty(U)$ доказывается теперь, как и раньше при $H(U)$, если использовать теорему 3.5.3 вместо (3.1.7).

3.6. МАРКОВСКИЕ ИСТОЧНИКИ

Некоторые идеи последнего параграфа могут быть выявлены с особой ясностью, если рассмотреть частный класс источников, называемых марковскими источниками. Эти источники задаются множеством состояний, обозначаемых целыми числами 1, ..., J и алфавитом букв источника, обозначаемым, как обычно, $a_1, \ldots, a_K$. В каждую единицу времени источник производит букву и переходит в новое состояние. Последовательность букв источника обозначается через $\mathbf{u} = (u_1, u_2, \ldots)$, а последовательность состояний — через $\mathbf{s} = (s_1, s_2, \ldots)$.

Пусть Q_{ji} обозначает условную вероятность перехода в состояние i при условии, что задано предыдущее состояние j:

$$Q_{ji} = \Pr(s_l = i \,|\, s_{l-1} = j). \qquad (3.6.1)$$

Предположим, что вероятность перехода в состояние зависит только от предыдущего состояния

$$\Pr(s_l \mid s_{l-1},\ s_{l-2},\ \ldots) = \Pr(s_l \mid s_{l-1}). \qquad (3.6.2)$$

Случайная последовательность состояний, для которой выполнены (3.6.1) и (3.6.2), называется конечной однородной цепью Маркова.

Пусть $P_j(a_k)$ обозначает вероятность того, что производится буква a_k, когда источник находится в состоянии j, и предположим, что эта вероятность зависит только от текущего состояния

$$P_j(a_k) = \Pr[u_i = a_k \mid s_l = j], \qquad (3.6.3)$$

$$\Pr(u_i \mid s_l) = \Pr(u_i \mid s_l,\ u_{l-1},\ s_{l-1},\ \ldots). \qquad (3.6.4)$$

Предположим, наконец, что состояние источника однозначно определяется предыдущим состоянием и предыдущей буквой.

Рис. 3.6.1 иллюстрирует работу марковского источника. Узлы соответствуют состояниям, а направленные ребра соответствуют буквам

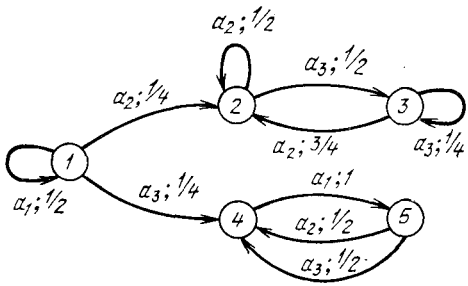


Рис. 3.6.1. Марковский источник; на ребрах указаны выход a_k и вероятность $P_j(a_k)$.

источника и переходам между состояниями. Каждое ребро, направленное от данного узла, должно соответствовать различным буквам для того, чтобы новое состояние однозначно определялось по предыдущему состоянию и букве.

Состояние марковского источника можно представить себе как тот объект, который определяет влияние истории источника на следующую букву. Так, например, стационарный источник, для которого каждая выходная буква статистически зависит лишь от l предыдущих выходных букв, является марковским источником, состояниями которого являются всевозможные последовательности из l букв. Из этого примера становится ясным, что большинство хороших стационарных источников может быть по крайней мере аппроксимировано марковскими источниками.

Поучительно рассмотреть моделирование английского текста с помощью марковского источника. Шеннон (1948) привел примеры последовательностей, порождаемых такими моделями; в первой из них буквы зависят от одной или двух предыдущих букв, а во второй — слова зависят от предыдущего слова с соответствующим выбором вероятностей.

Вот отрывок из этого последнего примера:

> The head and in frontal attack on an english writer that the character of this point is therefore another method ...*)

Хотя этот пример является чистой тарабарщиной, он довольно похож на осмысленный текст. На самом деле мы не можем моделировать текст с помощью случайного процесса достаточно точно. Очевидно, что имеется существенная разница между текстом медицинского словаря и текстом первой детской книги для чтения. Из-за этой разницы теорема кодирования для источника не может быть применена к английскому тексту и нельзя точно определить его энтропию. Вместе с тем, используя некоторые статистические зависимости английского языка, можно более эффективно описывать текст в сравнении с описанием, которое не использует эти зависимости.

Приведем здесь без доказательств некоторые свойства конечных однородных цепей Маркова**). Состояние s называется *невозвратным*, если существует некоторое состояние, которое может быть достигнуто из s за один или более переходов, но из которого невозможно никогда возвратиться в s. Например, на рис. 3.6.1 состояние 1 является невозвратным. Множество состояний называется *неразложимым*, если никакое состояние вне множества не может быть достигнуто ни из какого состояния, входящего в множество, и каждое состояние множества может быть достигнуто за один или более переходов. Так, например, состояния 2 и 3 образуют неразложимое множество, так же как и состояния 4 и 5.

Состояния любой конечной однородной цепи Маркова могут быть однозначно разбиты на одно или большее число неразложимых множеств состояний и множество (быть может пустое) невозвратных состояний. С вероятностью 1 цепь в конце концов оказывается в одном из неразложимых множеств и, конечно, остается в нем.

Число переходов, начиная из некоторого состояния s неразложимого множества, требующееся для первого возвращения в s, является случайной величиной, которая называется временем возвращения в s. *Периодом* неразложимого множества состояний называется наибольшее целое число m, такое, что все возможные времена возвращений для состояний этого множества являются кратными m. Например, период множества состояний **1** и **3** на рис. 3.6.1 равен 1, так как время возвращения для любого состояния может быть любым положительным целым числом. Период состояний **4** и **5** равен 2, так как время возвращения равно 2 для каждого состояния. Если неразложимое множество имеет период $m \geqslant 2$, то оно называется *периодическим*. Если $m = 1$, то множество называется *эргодическим*.

---

*) Русск. перев. «Голова и в фронтальной атаке на английского писателя, что характер этого места является, следовательно, другим методом....» (*Прим. перев.*).

**) Доказательства имеются, например, у Феллера (1950) или у Кокса и Миллера (1965).

Эргодическое множество состояний E имеет ассоциированное с ним множество стационарных вероятностей $q(j)$, задаваемых как решения уравнений

$$\sum_{j \in E} q(j) Q_{ji} = q(i), \quad i \in E, \tag{3.6.5}$$

$$\sum_{j \in E} q(j) = 1. \tag{3.6.6}$$

Более того, для любых i и j из E

$$\lim_{l \to \infty} \Pr(s_l = i \mid s_1 = j) = q(i), \tag{3.6.7}$$

где сходимость к пределу экспоненциальна по l.

Можно заметить, что вероятности в (3.6.1)—(3.6.4) не описывают полностью источник. Необходимо еще указать, когда источник начинает работу и каково начальное распределение вероятностей для состояний. Если состояния источника принадлежат некоторому данному эргодическому множеству состояний, начиная со сколь угодно далекого прошлого, то

$$\Pr(s_l = i) = q(i) \text{ для всех } l, \tag{3.6.8}$$

и источник является стационарным и эргодическим согласно определениям этого параграфа. Вместе с тем, если источник начинает работу с заданного конечного момента времени из заданного состояния, то он не является стационарным и эргодическим, так как у него нет прошлого и так как имеется начальная неустойчивость в вероятности.

Проблема переходного режима много серьезнее для периодических множеств состояний, чем для эргодических множеств. Для периодического множества состояний с периодом m имеется m возможных фаз, соответствующих тому, что данное состояние может наступить только в моменты $\ldots, -m, 0, m, \ldots$ или в моменты $\ldots, -m+1, 1, m+1, \ldots$ или и т. д. вплоть до моментов $\ldots, -m+(m-1), m-1, m+(m-1), \ldots$ Если источник начал свою работу в бесконечно удаленный момент прошлого в данной фазе, то результирующее случайное состояние последовательности является периодическим в смысле определения этого параграфа. Если источник начал свою работу в сколь угодно удаленном прошлом с равномерным распределением фаз, то результирующая последовательность случайных состояний удовлетворяет (3.5.13) и, следовательно, является эргодической.

Сейчас мы исследуем энтропию марковского источника. Энтропия выходной буквы источника в заданный момент времени при условии, что задано текущее состояние источника, равна

$$H(U \mid s = j) = -\sum_{k=1}^{K} P_j(a_k) \log P_j(a_k). \tag{3.6.9}$$

Затем найдем энтропию выхода источника при условии, что задано некоторое частное состояние в некоторый момент в прошлом и заданы промежуточные выходы источника.

Лемма.

$$H(U_l \mid U_{l-1} U_{l-2} \ldots U_1,\ s_1 = j) =$$

$$= \sum_{i=1}^{J} \Pr(s_l = i \mid s_1 = j)\, H(U \mid s = i). \qquad (3.6.10)$$

Доказательство. По определению марковского источника состояние s_2 в момент 2 однозначно определяется s_1 и u_1, т. е. предыдущим состоянием и выходной буквой. Подобно этому s_3 однозначно определяется s_2 и u_2 и, следовательно, определяется с помощью u_2, u_1 и s_1. Продолжая по индукции, находим, что для любого положительного l состояние s_l однозначно определяется $u_1, \ldots, u_{l-1}$ и s_1. Следовательно,

$$\Pr(u_l \mid u_1, \ldots, u_{l-1}, s_1) = \Pr(u_l \mid s_l, u_1, \ldots, u_{l-1}, s_1),$$

где s_l является состоянием, определяемым $u_1, \ldots, u_{l-1}$ и s_1. Так как u_l зависит только от s_l [см. (3.6.4)], то отсюда получаем

$$\Pr(u_l \mid u_1, \ldots, u_{l-1}, s_1) = \Pr(u_l \mid s_l). \qquad (3.6.11)$$

Логарифмируя обе части равенства (3.6.11) и усредняя по $u_1, \ldots, u_l$ и s_l, будем иметь

$$\sum_{u_1, \ldots, u_l, s_l} \Pr(u_1, \ldots, u_l, s_l \mid s_1) \log \Pr(u_l \mid u_1, \ldots, u_{l-1}, s_1) =$$

$$= \sum_{u_1, \ldots, u_l, s_l} \Pr(u_1, \ldots, u_{l-1}, s_l \mid s_1) \Pr(u_l \mid s_l) \log \Pr(u_l \mid s_l).$$

Суммируя правую часть равенства по $u_1, \ldots, u_{l-1}$, получаем (3.6.10). |

Отметим, что доказательство этой леммы существенно опирается на предположение, что состояние источника определяется предыдущим состоянием и выходной буквой. Вычисления левой части (3.6.10) для источников, не удовлетворяющих этому условию, крайне искусственны и сложны*).

Любое заданное распределение вероятностей для состояния s_1 определяет распределение вероятности состояний во все будущие моменты времени, поэтому можно усреднить (3.6.9) по s_1 и получить

$$H(U_l \mid U_{l-1} \ldots U_1 S_1) = \sum_{i=1}^{J} \Pr(s_l = i)\, H(U \mid s = i). \qquad (3.6.12)$$

Для стационарного эргодического марковского источника $\Pr(s_l = i)$ не зависит от l и из (3.6.8) имеем

$$H(U_l \mid U_{l-1} \ldots U_1 S_1) = \sum_i q(i)\, H(U \mid s = i) \text{ для всех } l \geqslant 1. \qquad (3.6.13)$$

*) Блекуэлл (1957) рассмотрел эту задачу в пределе при $l \to \infty$ и показал, что она может быть сведена к решению трудного интегрального уравнения.

Рассмотрим теперь энтропию на букву последовательности букв источника при условии, что задано начальное состояние

$$\frac{1}{L} H(U_1 \dots U_L | S_1) = \frac{1}{L} \sum_{l=1}^{L} H(U_l | U_{l-1} \dots U_1 S_1). \qquad (3.6.14)$$

Отсюда согласно (3.6.12) имеем

$$-\frac{1}{L} H(U_1 \dots U_L | S_1) = \sum_{i=1}^{J} q_{(1,L)}(i) H(U | s = i), \qquad (3.6.15)$$

где

$$q_{(1,L)}(i) = \frac{1}{L} \sum_{l=1}^{L} \Pr(s_l = i). \qquad (3.6.16)$$

Отсюда видно, что $q_{(1, L)}(i)$ в точности равна средней по времени вероятности пребывания в состоянии i. Для стационарного эргодического марковского источника $q_{(1, L)}(i)$ совпадают с $q(i)$, как показывают (3.6.5) и (3.6.6), и поэтому

$$\frac{1}{L} H(U_1 \dots U_L | S_1) = \sum_{i=1}^{J} q(i) H(U | s = i). \qquad (3.6.17)$$

В пределе при $L \to \infty$ определим

$$q_{(1,\infty)}(i) = \lim_{L \to \infty} \frac{1}{L} \sum_{l=1}^{L} \Pr(s_l = i). \qquad (3.6.18)$$

Этот предел всегда существует, хотя вообще он зависит от распределения вероятностей для s_1. Так, например, на рис. 3.6.1 это отличие становится ясным, когда s_1 задается как состояние 2 с вероятностью 1 или как состояние 4 с вероятностью 1. Вместе с тем для цепи Маркова, имеющей только одно неразложимое множество состояний, $q_{(1, \infty)}(i)$ не зависят от начального распределения вероятностей. При этом определении $q_{(1, \infty)}(i)$ получаем

$$\lim_{L \to \infty} \frac{1}{L} H(U_1 \dots U_L | S_1) = \sum_{i=1}^{J} q_{(1,\infty)}(i) H(U | s = i). \qquad (3.6.19)$$

Рассмотрим далее безусловную энтропию на букву последовательности источника. Имеем

$$H(U_1 \dots U_L) = I(S_1;\ U_1 \dots U_L) + H(U_1 \dots U_L | S_1).$$

Средняя взаимная информация в правой части приведенного выше выражения ограничена 0 и $\log J$ и, следовательно,

$$\lim_{L \to \infty} \frac{1}{L} H(U_1 \dots U_L) = \lim_{L \to \infty} \frac{1}{L} H(U_1 \dots U_L | S_1). \qquad (3.6.20)$$

Обозначая левую часть (3.6.20) через $H_\infty(U)$, получаем в силу (3.6.19), что

$$H_\infty(U) = \sum_{i=1}^{J} q_{(1,\infty)}(i) H(U | s = i). \qquad (3.6.21)$$

Следующая теорема подытоживает полученные результаты.

Теорема 3.6.1. Энтропия на букву марковского источника задается равенством (3.6.21), где $q_{(1,\infty)}(i)$ задается (3.6.18), а $H(U|s=i)$ задается (3.6.9). Если цепь Маркова имеет не больше одного неразложимого множества состояний, то $q_{(1,\infty)}(i)$ не зависит от распределения вероятностей для s_1, и если это неразложимое множество является эргодическим, то $q_{(1,\infty)}(i) = q(i)$, где $q(i)$ задается (3.6.5) и (3.6.6).

Исследование неравномерных кодов для источника, проведенное для дискретных стационарных источников, непосредственно применимо к марковским источникам, однако для марковского источника возможны некоторые упрощения. В (3.5.11) среднее число кодовых букв на букву источника для кодирования сразу L букв источника удовлетворяет условию $\bar{n} \geqslant H_L(U)/\log D$. Для того чтобы приблизить $\bar{n}$ к $H_\infty(U)/\log D$, возможно, потребуется взять L довольно большим. Для стационарных эргодических марковских источников будет показано, что, используя информацию о состоянии и кодируя сразу L букв источника, можно получить $\bar{n}$, удовлетворяющее неравенствам

$$\frac{H_\infty(U)}{\log D} \leqslant \bar{n} < \frac{H_\infty(U)}{\log D} + \frac{1}{L}. \qquad (3.6.21a)$$

Чтобы получить этот результат, используются различные коды для различных начальных состояний. Длина кодового слова, соответствующая последовательности $\mathbf{u} = u_1, \ldots, u_L$ и $s_1 = j$, может быть выбрана, как и в (3.3.6), удовлетворяющей неравенствам

$$D^{-n_j(\mathbf{u})} \leqslant \Pr(\mathbf{u} \mid s_1 = j) < D^{-n_j(\mathbf{u})+1}. \qquad (3.6.22)$$

Так же как и в теореме 3.3.[1], эти длины удовлетворяют неравенству Крафта для каждого начального состояния, и средняя длина кодового слова $\bar{n}_j L$ для данного начального состояния j удовлетворяет неравенствам $[H(U_1 \ldots U_L \mid s_1 = j)/\log D] \leqslant \bar{n}_j L < [H(U_1 \ldots U_L \mid s_1 = j)/\log D] + 1$.

Усредняя по состояниям, деля на L и используя (3.6.17), полученное неравенство можно свести к (3.6.21 а).

ИТОГИ И ВЫВОДЫ

Эта глава преследовала три цели: первая — выяснить смысл энтропии; вторая — получить некоторый навык работы с последовательностями событий и третья — узнать, как закодировать источники, используя минимальное среднее число кодовых букв на букву источника. Мы установили, что энтропию $H(U)$ дискретного источника без памяти можно истолковать в терминах последовательностей L букв источника для больших L. Грубо говоря, $H(U)$ равна умноженному на $1/L$ логарифму числа «типичных» последовательностей источника, или минус умноженному на $1/L$ логарифму вероятности типичной последовательности источника. Энтропия $H(U)$ равна также минимальному числу кодовых букв на букву источника $\bar{n}$, которое требуется для представления выхода источника. Для кода с фиксированной длиной такое $\bar{n}$ может быть достигнуто только ценой ненулевой (но убывающей вместе с L) вероят-

ности получения последовательности источника с неоднозначно приписанным кодовым словом. Для неравномерного кода достижение такого $\bar{n}$ связано с проблемами ожидания в случае, если символы источника поступают на кодирующее устройство с фиксированной скоростью и кодовые символы должны выходить из кодирующего устройства с фиксированной скоростью.

В гл. 9 будет рассмотрена намного более общая задача кодирования для источника, в которой источник не нужно будет воспроизводить точно, а лишь с некоторым заданным критерием верности.

Другая важная проблема кодирования для источника, которая не была здесь затронута, состоит в том, как конструировать коды, удовлетворяющие определенным ограничениям. Так, например, имеется значительная литература по кодам, обладающим свойством синхронизации. Это такие коды, для которых бесконечная закодированная последовательность может быть декодирована начиная с середины последовательности.

Следует помнить, что для большинства реальных источников центральные проблемы не являются теоретико-информационными проблемами представления источника, а состоят в более субъективных проблемах определения, что имеется ценного на выходе источника.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Другое рассмотрение затронутых здесь вопросов (в особенности, содержащихся в первых четырех параграфах) можно найти у Эбрамсона (1963), Фано (1961), Эша (1965) и Шеннона (1948). Теорема кодирования для источника (теорема 3.1.1) была получена Шенноном (1948), который развил большинство концепций этой главы. Шеннон также установил теорему для случая, когда буквы источника имеют неравные длительности и когда источник является марковским. Как указано в тексте, процедура построения оптимального кода, изложенная в § 2.4, принадлежит Хаффману (1952). Кодирование для источника, обладающее свойством синхронизации, было изучено Голомбом, Гордоном и Велчем (1958), Кендаллом и Ридом (1962), Истманом (1965), Шольцем (1966) и другими.

Теорема 3.5.3 была получена Макмилланом (1953) и часто называется AEP-свойством*) эргодических источников. Макмиллан доказал L_1-сходимость, которая несколько сильнее, чем сходимость, установленная здесь, но его теорема относится только к источникам с конечным алфавитом и его доказательство намного сложнее приведенного здесь. Брейман (1957) позднее доказал сходимость по вероятности для эргодического источника с конечным алфавитом. Отт (1962) разработал процедуру кодирования для марковского источника более общего типа по сравнению с рассмотренным здесь; в его источнике состояние источника не обязательно однозначно определяется предыдущим состоянием и предыдущей буквой источника.

*) AEP — первые буквы слов «Asymptotic equipartition property». В советской литературе используется термин «информационная устойчивость». (*Прим. ред.*)

4

ДИСКРЕТНЫЕ КАНАЛЫ БЕЗ ПАМЯТИ И ПРОПУСКНАЯ СПОСОБНОСТЬ

В предыдущей главе была рассмотрена задача представления выхода источника информации с помощью букв кодового алфавита. На этом пути были найдены несколько наглядных толкований собственной информации и энтропии, а также был получен ряд простых и ясных результатов, касающихся кодирования для источника. Было установлено, что трудности применения теории обусловлены не ее сложностью, а трудностью представления реальных информационных источников разумными вероятностными моделями. В этой и последующих главах будет рассмотрена передача информации по каналам с шумами. В ходе этого рассмотрения будет получено более ясное понимание природы взаимной информации и найдены некоторые глубокие и имеющие большое значение результаты, касающиеся кодирования для канала с шумами. Будет показано, что эти результаты не так просты, как результаты, относящиеся к кодированию для источников, но они имеют очень большое практическое значение. Это значение проистекает из того, что для многих реальных каналов связи могут быть построены достаточно простые и полезные вероятностные модели и что применение теории к этим моделям приводит к нетривиальному проникновению в задачи построения систем связи.

4.1. КЛАССИФИКАЦИЯ КАНАЛОВ

Каналы связи могут быть описаны в терминах множества входных сигналов, которые имеются на входе канала; множества выходных сигналов, имеющихся на выходном конце канала, и (для каждого входного сигнала) вероятностной меры на выходных событиях при условии, что задан этот входной сигнал.

Вначале рассматривается дискретный канал без памяти. Это такой канал, вход и выход которого представляют собой последовательности букв конечных алфавитов и для которого выходная буква в данный момент статистически зависит лишь от соответствующей входной буквы.

В гл. 7 рассматривается канал другого типа, а именно непрерывный по амплитуде, дискретный по времени канал без памяти. Входом и выходом в нем являются последовательности букв из алфавитов, состоящих из множества действительных чисел (или, в более общем случае, из векторов с действительными компонентами) и опять выходная буква в данный момент времени статистически зависит лишь от входа в соответствующий момент. Можно также рассмотреть каналы, в которых вход

является дискретным, а выход непрерывным или наоборот, но это оказывается тривиальным видоизменением.

Еще один тип канала, который будет рассмотрен в гл. 8, является непрерывным по времени каналом, в котором входом и выходом являются функции. Какая из описанных выше моделей канала используется для описания заданной линии связи, часто является вопросом выбора. Так, например, канал, изображенный на рис. 4.1.1, можно рассматривать и как дискретный и как непрерывный по времени. Если вначале интересоваться кодером и декодером на рис. 4.1.1, то удобно рассматривать модулятор и демодулятор дискретных данных как часть канала,

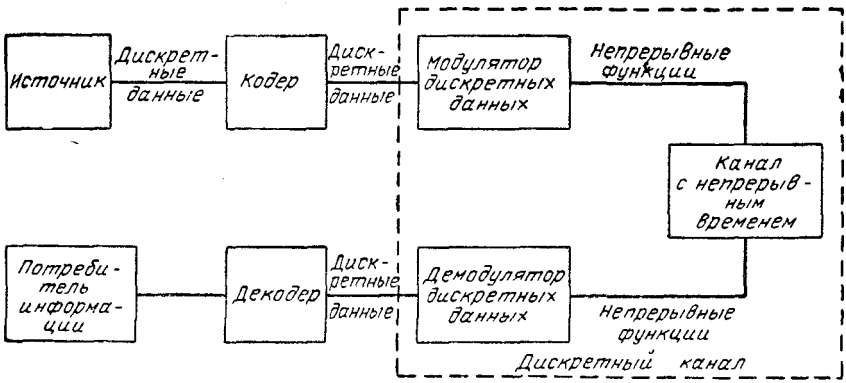


Рис. 4.1.1. Выбор модели канала.

и в этом случае канал является дискретным. С другой стороны, если интерес представляет построение как кодера, так и модулятора дискретных данных или если появляется потребность рассмотреть два этих устройства как единое целое, то соответствующий канал рассматривается, как непрерывный.

В последнем параграфе этой главы будут рассмотрены дискретные каналы с памятью; это такие каналы, в которых выход в данный момент статистически зависит как от текущего входа, так и от предыдущих входов и выходов. Память в дискретной модели канала возникает в силу целого ряда причин, действующих в реальных каналах. Одна из очевидных причин состоит в межсимвольной интерференции, происходящей из-за фильтрации в канале. В этом случае выходной символ статистически зависит от нескольких входных символов. Другой причиной являются замирания в канале. Заманчиво представлять себе каналы с замираниями как изменяющиеся со временем каналы без памяти. Однако такое представление в некотором смысле неудачно, так как замирания обычно наилучшим образом моделируются как статистическое явление и должны быть учтены в определении вероятностей выходов при заданных входах. Так, например, при передаче двоичных символов по каналу с замираниями, относительно медленными по сравнению со скоростью передачи в битах, ошибки будут иметь тенденцию

группироваться вместе под влиянием замираний, вместо того чтобы быть статистически независимыми друг от друга. Таким образом, канал обладает памятью в том смысле, что он помнит, когда он находится в плохом состоянии и имеет тенденцию оставаться в плохом состоянии в течение определенного интервала времени.

4.2. ДИСКРЕТНЫЕ КАНАЛЫ БЕЗ ПАМЯТИ

Рассмотрим дискретный канал без памяти (ДКБП), входной алфавит которого X состоит из K целых чисел 0, 1, ..., $K-1$ и выходной алфавит которого Y состоит из J целых чисел 0, 1, ..., $J-1$. Использование целых чисел в качестве входных и выходных букв упрощает обозначения в некоторых последующих рассмотрениях, а также вновь подчеркивает то обстоятельство, что обозначения, принятые для входных и выходных букв, ни на что не влияют.

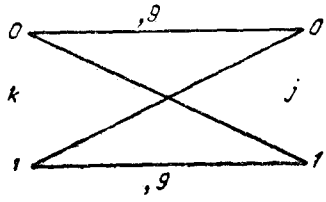


Рис. 4.2.1. Переходные вероятности в двоичном симметричном канале. (Здесь и на некоторых последующих рисунках в десятичных дробях, целая часть которых равна нулю, нуль опускается.)

Канал описывается переходными вероятностями $P(j \mid k)$, заданными для $0 \leqslant j \leqslant J-1$ и $0 \leqslant k \leqslant K-1$. По определению, $P(j \mid k)$ является условной вероятностью приема целого числа j при условии, что задан вход канала — целое число k.

Обозначим последовательность N букв на входе канала через $\mathbf{x} = (x_1, \ldots, x_n, \ldots, x_N)$, где x_n, $1 \leqslant n \leqslant N$, принимают значения из входного алфавита, т. е. значения от 0 до $K-1$. Аналогично, соответствующие последовательности выходных букв обозначим через $\mathbf{y} = (y_1, \ldots, y_N)$, где y_n принимают значения из выходного алфавита, т. е. значения от 0 до $J-1$. Вероятность y_n при условии, что задано x_n, задается описанной выше переходной вероятностью $P(y_n \mid x_n)$.

В силу того, что канал является каналом без памяти, каждая выходная буква в последовательности зависит только от соответствующей входной буквы и условная вероятность выходной последовательности $\mathbf{y} = (y_1, \ldots, y_N)$ при условии, что задана входная последовательность $\mathbf{x} = (x_1, \ldots, x_N)$, определяется равенством

$$P_N(\mathbf{y} \mid \mathbf{x}) = \prod_{n=1}^{N} P(y_n \mid x_n). \qquad (4.2.1)$$

Более формально это можно выразить следующим образом: канал является каналом без памяти, если существуют такие переходные вероятности $P(j \mid k)$, что (4.2.1) справедливо для всех N, всех $\mathbf{y} = (y_1, \ldots, y_N)$ и всех $\mathbf{x} = (x_1, \ldots, x_N)$.

Для примера использования указанных выше обозначений рассмотрим двоичный симметричный канал, изображенный на рис. 4.2.1. Переходные вероятности для рис. 4.2.1 задаются следующим образом: $P(0 \mid 0) = 0{,}9$, $P(1 \mid 0) = 0{,}1$, $P(0 \mid 1) = 0{,}1$, $P(1 \mid 1) = 0{,}9$.

Для последовательностей длины 2 равенство (4.2.1) дает: $P_2(00\,|\,00) = (0{,}9)\cdot(0{,}9) = 0{,}81$, $P_2(10\,|\,00) = (0{,}1)\cdot(0{,}9) = 0{,}09$ и т. д. Вероятность $P_2(00\,|\,00)$ обозначает вероятность приема двух нулей при условии, что были переданы два нуля.

Отметим, что при описании канала ничего не было сказано о методе использования символов на входе канала. Если задать вероятностную меру на входных целых числах, обозначая через $Q(k)$ вероятность использования числа k, то средняя взаимная информация между входом и выходом равна

$$I(X;\,Y) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} Q(k)\,P(j\,|\,k)\log\frac{P(j\,|\,k)}{\sum\limits_{i=0}^{K-1} Q(i)\,P(j\,|\,i)}. \qquad (4.2.2)$$

Вероятность приема числа j была выписана в виде $\sum\limits_i Q(i)\;P(j\,|\,i)$, чтобы подчеркнуть, что она является функцией как распределения на входе, так и переходных вероятностей канала.

Так как относительные частоты букв на входе канала могут быть соответствующим образом выбраны с помощью кодера, то не должно быть удивительным то, что максимум $I(X;\,Y)$ по входиым вероятностям является величиной, которая имеет теоретико-информационный смысл. *Пропускной способностью C дискретного канала без памяти (ДКБП) называется наибольшая средняя взаимная информация $I(X;\,Y)$, которая может быть передана по каналу при его однократном использовании, максимизированная по всем распределениям на входе*:

$$C = \max_{Q(0),\ldots,Q(K-1)} \sum_{k,j} Q(k)\,P(j\,|\,k)\log\frac{P(j\,|\,k)}{\sum Q(i)\,P(j\,|\,i)}. \qquad (4.2.3)$$

Отметим, что в то время как $I(X;\,Y)$ является функцией как канала, так и распределения на входе, C является функцией только канала. Вычисление C включает в себя максимизацию по K переменным с двумя ограничениями: одно в виде неравенства $Q(k) \geqslant 0$ и другое в виде равенства $\Sigma Q(k) = 1$. Максимальное значение существует в силу того, что функция является непрерывной и максимизация производится в замкнутой ограниченной области векторного пространства*).

В § 4.4 и 4.5 мы вернемся к задаче численного отыскания пропускной способности.

Основополагающее значение пропускной способности для ДКБП обосновывается теоремой кодирования, которая утверждает, что данные могут быть надежно переданы по каналу с любой скоростью,

*) См. любой учебник по математическому анализу, например, Бьюк (1956) (см. также Фихтенгольц Г. М. Курс математического анализа, т. 1. «Наука», 1964). Причина возникновения вопроса о существовании максимума может быть объяснена попыткой максимизировать функцию x^2 на открытом интервале $0 < x < 2$. Эта функция не имеет максимума, так как она принимает значения, сколь угодно близкие к 4, но никогда не достигает 4.

меньшей пропускной способности. Заметим, что поразительным в теореме кодирования является слово «надежно». То, что информация может быть передана со скоростью, равной пропускной способности, является очевидным, так как для этого нужно лишь просто выбрать соответствующее распределение на входе. Теорема кодирования будет рассмотрена в следующей главе. Здесь мы покажем, что пропускная способность может быть интерпретирована как максимальная средняя взаимная информация на букву, которая может быть передана для последовательности входов и выходов. Далее будет доказано обращение теоремы кодирования, т. е. что надежная передача невозможна для скоростей источника, превосходящих пропускную способность канала.

Теорема 4.2.1. Пусть $Q_N(\mathbf{x})$ — некоторое произвольное совместное распределение вероятностей, заданное на последовательностях N символов на входе ДКБП. Пусть $\mathbf{X}^N$ и $\mathbf{Y}^N$ являются ансамблями входных и выходных последовательностей и пусть $X_1, \ldots, X_N, Y_1, \ldots, Y_N$ обозначают ансамбли, соответствующие отдельным буквам. Тогда

$$I(\mathbf{X}^N; \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n; Y_n), \tag{4.2.4}$$

$$I(\mathbf{X}^N; \mathbf{Y}^N) \leqslant NC. \tag{4.2.5}$$

Равенство в (4.2.4) имеет место, если входы статистически независимы, а равенство в (4.2.5) имеет место, если входы независимы и имеют распределение вероятностей, определенное (4.2.3).

Доказательство.

$$I(\mathbf{X}^N; \mathbf{Y}^N) = H(\mathbf{Y}^N) - H(\mathbf{Y}^N \mid \mathbf{X}^N), \tag{4.2.6}$$

$$H(\mathbf{Y}^N \mid \mathbf{X}^N) = \sum_{\mathbf{x},\mathbf{y}} Q_N(\mathbf{x})\, \mathbf{P}_N(\mathbf{y} \mid \mathbf{x}) \log \frac{1}{P_N(\mathbf{y} \mid \mathbf{x})}. \tag{4.2.7}$$

Так как канал является каналом без памяти, то можно использовать (4.2.1), что дает

$$H(\mathbf{Y}^N \mid \mathbf{X}^N) = \sum_{\mathbf{x},\mathbf{y}} Q_N(\mathbf{x})\, P_N(\mathbf{y} \mid \mathbf{x}) \sum_{n=1}^{N} \log \frac{1}{P(y_n \mid x_n)}. \tag{4.2.8}$$

Величина $\log [1/P(y_n \mid x_n)]$ является случайной величиной и правая часть (4.2.8) равна среднему значению суммы N случайных величин. Она равна сумме средних значений, независимо от того, являются ли входы статистически независимыми или нет. Но среднее значение $\log [1/P(y_n \mid x_n)]$ равно $H(Y_n \mid X_n)$, так что

$$H(\mathbf{Y}^N \mid \mathbf{X}^N) = \sum_{n=1}^{N} H(Y_n \mid X_n). \tag{4.2.9}$$

Это означает, что одно из выражений в (4.2.6) равно сумме энтропий; займемся теперь другим выражением $H(\mathbf{Y}^N)$. Из (2.3.10) имеем

$$H(\mathbf{Y}^N) \leqslant \sum_{n=1}^{N} H(Y_n). \tag{4.2.10}$$

Подставляя (4.2.9) и (4.2.10) в (4.2.6), получаем

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} [H(Y_n) - H(Y_n \mid X_n)], \tag{4.2.11}$$

откуда следует (4.2.4).

Равенства в (4.2.10) и, следовательно, в (4.2.4) имеют место тогда и только тогда, когда выходные буквы статистически независимы. Если входные буквы статистически независимы, т. е.

$$Q_N(\mathbf{x}) = \prod_n Q_{X_n}(x_n),$$

то совместная вероятность равна

$$\prod_n Q_{X_n}(x_n) P(y_n \mid x_n)$$

и отсюда следует статистическая независимость выходов.

Определение C означает, что $I(X_n;\ Y_n) \leqslant C$ при всех n и, таким образом, (4.2.5) следует из (4.2.4). Более того, если входы статистически независимы и выбираются так, чтобы максимизировать каждую информацию $I(X_n;\ Y_n)$, то $I(X_n;\ Y_n) = C$ при всех n и (4.2.5) удовлетворяется с равенством. |

Из этой теоремы не нужно делать вывода о том, что следует избегать статистическую зависимость между входными буквами. В действительности все рассматриваемые ниже методы кодирования описывают способы введения статистической зависимости между входными буквами, и некоторые из таких зависимостей необходимы в общем случае для того, чтобы получить надежную передачу.

4.3. ОБРАЩЕНИЕ ТЕОРЕМЫ КОДИРОВАНИЯ

До сих пор источники и каналы обсуждались в терминах понятий различных энтропий и средних взаимных информаций. Однако в большинстве систем передачи данных взаимная информация представляет меньший интерес, чем вероятность того, что буквы источника неправильно воспроизводятся у адресата. Эта вероятность ошибки определяет содержание основной теоремы теории информации — теоремы кодирования. Для широкого класса источников и каналов эта теорема утверждает, что если энтропия источника на единицу времени меньше, чем пропускная способность канала на единицу времени, то вероятность ошибки может быть сделана сколь угодно малой с помощью использования достаточно сложных кодера и декодера. В этом параграфе нас будет интересовать обращение этого результата: если энтропия источника больше, чем пропускная способность, то нельзя достичь произвольно малой вероятности ошибки.

Рассмотрим вначале последовательность $\mathbf{u} = (u_1, \ldots, u_L)$ из L букв дискретного источника, изображенного на рис. 4.3.1. Энтропия на букву для последовательности из L букв определяется равенством

$$H_L(U) = \frac{H(\mathbf{U}^L)}{L} = -\frac{1}{L} \sum_{\mathbf{u}} \Pr(\mathbf{u}) \log \Pr(\mathbf{u}). \qquad (4.3.1)$$

Как было показано в теореме 3.5.1, для стационарного источника $H_L(U)$ не возрастает вместе с L и стремится к пределу $H_\infty(U)$ при $L \to \infty$. Для дискретного источника без памяти, очевидно, $H_L(U) = H(U)$ при всех L.

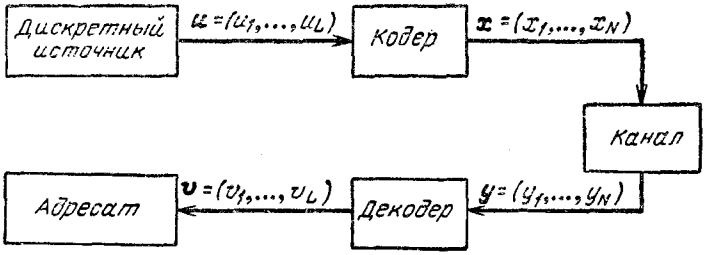


Рис. 4.3.1. Система связи.

Предположим, что выходом декодера является последовательность $\mathbf{v} = (v_1, \ldots, v_L)$ букв того же самого алфавита, что и алфавит источника. При любых заданных источнике, кодере, канале и декодере существует совместная вероятностная мера*) на множестве возможных входных последовательностей $\mathbf{u}$ и множестве возможных выходных последовательностей $\mathbf{v}$.

Целью системы связи является получение последовательности $\mathbf{v}$, воспроизводящей последовательность $\mathbf{u}$. Если $u_l \neq v_l$, то считается, что произошла ошибка в l-м переданном символе. Вероятность такой ошибки $P_{e,l}$ определяется совместным ансамблем $\mathbf{U}^L\mathbf{V}^L$. *Средняя вероятность ошибки* $\langle P_e \rangle$ *в последовательности из* L *символов определяется следующим образом*:

$$\langle P_e \rangle = \frac{1}{L} \sum_{l=1}^{L} P_{e,l}. \qquad (4.3.2)$$

Математическое ожидание числа ошибок в последовательности равно $L\langle P_e \rangle$. В последующих главах мы часто будем иметь дело с вероятностью одной или более ошибок в последовательности длины L. Здесь, однако, мы хотим показать, что надежная передача невозможна, если энтропия источника больше, чем пропускная способность. Таким образом, даже, если бы было показано, что вероятность ошибки в последовательности равна 1 для больших L, то это бы гарантировало лишь су-

*) Как будет отмечено в § 4.6, для каналов с памятью это утверждение требует некоторого разъяснения. Однако до этого времени предполагается существование такой вероятностной меры.

ществование одного ошибочного символа среди L или $\langle P_e \rangle \geqslant 1/L$. Следовательно, для того чтобы показать, что $\langle P_e \rangle$ отлично от нуля в пределе при $L \to \infty$, надо непосредственно рассмотреть $\langle P_e \rangle$ а не вероятность ошибки в последовательности.

Начнем с установления соотношений между $\langle P_e \rangle$, $H_L(U)$ и $I(\mathbf{U}^L; \mathbf{V}^L)$ в частном случае, когда $L = 1$. Затем распространим этот результат на случай произвольного L и, наконец, свяжем $I(\mathbf{U}^L; \mathbf{V}^L)$ с пропускной способностью канала.

Теорема 4.3.1. Пусть UV является совместным ансамблем, в котором выборочные пространства U и V состоят из одних и тех же M элементов $a_1, \ldots, a_M$. Пусть P_e является вероятностью того, что случайные величины u и v не равны друг другу:

$$P_e = \sum_u \sum_{v \neq u} P(u, v). \tag{4.3.3}$$

Тогда

$$P_e \log (M-1) + \mathcal{H}(P_e) \geqslant H(U|V), \tag{4.3.4}$$

где

$$\mathcal{H}(P_e) = -P_e \log P_e - (1-P_e) \log (1-P_e). \tag{4.3.5}$$

Обсуждение. Функция $P_e \log (M-1) + \mathcal{H}(P_e)$ изображена на рис. 4.3.2. Теорема утверждает, что, если $H(U|V)$ принимает заданное значение, отложенное на оси ординат на рис. 4.3.2, то P_e должно быть не меньше соответствующей абсциссы. Так как $H(U|V) = H(U) - I(U, V)$, то теорема ограничивает P_e с помощью выражения, представляющего собой превышение энтропии источника над средней взаимной информацией.

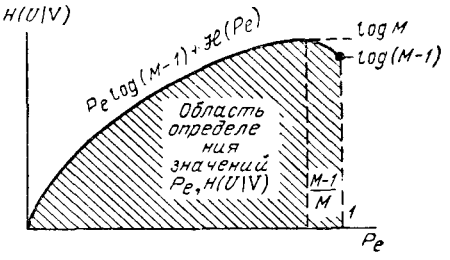


Рис. 4.3.2. Интерпретация теоремы 4.3.1.

Легко эвристически установить справедливость (4.3.4). Средняя неопределенность u при заданном v может быть разбита на два члена: первый, представляющий неопределенность, относящуюся к тому, была совершена ошибка или нет при заданном v; и второй, представляющий неопределенность переданной входной буквы в случаях, когда была сделана ошибка. Первый член ограничен сверху значением $\mathcal{H}(P_e)$, а второй член — значением, равным P_e, умноженным на максимальную неопределенность при условии, что произошла ошибка. Так как неопределенность при условии, что произошла ошибка, состоит в выборе между $M-1$ альтернативами, то второй член ограничен сверху значением $P_e \log (M-1)$.

Доказательство. Можно записать $H(U|V)$ как сумму двух слагаемых; одно, включающее пары u, v, которые приводят к ошибке,

$u \neq v$, и другое, содержащее пары u, v, для которых $u = v$:

$$H(U|V) = \sum_v \sum_{u \neq v} P(u, v) \log \frac{1}{P(u|v)} + \sum_{v,\, u=v} P(u, v) \log \frac{1}{P(u|v)}. \tag{4.3.6}$$

Согласно (4.3.3) разность между выражениями, стоящими в разных частях (4.3.4), равна

$$H(U|V) - P_e \log (M-1) - \mathcal{H}(P_e) =$$

$$= \sum_v \sum_{u \neq v} P(u, v) \log \frac{P_e}{(M-1) P(u|v)} + \sum_{v,\, u=v} P(u, v) \log \frac{1-P_e}{P(u|v)}. \tag{4.3.7}$$

С помощью неравенства $\log z \leqslant (\log e)(z-1)$ приходим к тому, что правая часть (4.3.7) меньше или равна выражению

$$(\log e) \left\{ \sum_v \sum_{u \neq v} P(u, v) \left[\frac{P_e}{(M-1) P(u|v)} - 1 \right] + \right.$$

$$\left. + \sum_{v,\, u=v} P(u, v) \left[\frac{1-P_e}{P(u|v)} - 1 \right] \right\} = \tag{4.3.8}$$

$$= (\log e) \left[\frac{P_e}{M-1} \sum_v \sum_u P(v) - \sum_v \sum_{u \neq v} P(u, v) + \right.$$

$$\left. + (1-P_e) \sum_v P(v) - \sum_{v,\, u=v} P(u, v) \right] = \tag{4.3.9}$$

$$= (\log e) [P_e - P_e + (1-P_e) - (1-P_e)] = 0. \tag{4.3.10}$$

Теорема 4.3.2. Пусть $\mathbf{U}^L \mathbf{V}^L$ обозначает совместный ансамбль последовательностей $\mathbf{u} = (u_1, \ldots, u_L)$ и $\mathbf{v} = (v_1, \ldots, v_L)$, в котором выборочные пространства для u_l и v_l состоят из одних и тех же M элементов $a_1, \ldots, a_M$. Пусть $\langle P_e \rangle$ определено в (4.3.2). Тогда

$$\langle P_e \rangle \log (M-1) + \mathcal{H}(\langle P_e \rangle) \geqslant \frac{1}{L} H(\mathbf{U}^L | \mathbf{V}^L). \tag{4.3.11}$$

Доказательство. Используя цепное правило для совместного ансамбля $\mathbf{U}^L = U_1 U_2 \ldots U_L$ [см. равенство (2.2.30)], имеем

$$H(\mathbf{U}^L | \mathbf{V}^L) = H(U_1 | \mathbf{V}^L) + H(U_2 | U_1 \mathbf{V}^L) + \ldots$$

$$\ldots + H(U_L | \mathbf{V}^L U_1 \ldots U_{L-1}) \leqslant \tag{4.3.12}$$

$$\leqslant \sum_{l=1}^{L} H(U_l | V_l). \tag{4.3.13}$$

Неравенство (4.3.13) следует из общего неравенства $H(X|Z) \geqslant H(X|ZY)$ [см. неравенство (2.3.13)].

Применяя теорему 4.3.1 к каждому слагаемому в (4.3.13), получаем

$$H(\mathbf{U}^L \mid \mathbf{V}^L) \leqslant \sum_{l=1}^{L} [P_{e,\,l} \log (M-1) + \mathcal{H}(P_{e,\,l})], \tag{4.3.14}$$

$$\frac{1}{L} H(\mathbf{U}^L \mid \mathbf{V}^L) \leqslant \langle P_e \rangle \log (M-1) + \frac{1}{L} \sum_{l=1}^{L} \mathcal{H}(P_{e,\,l}). \tag{4.3.15}$$

Чтобы завершить доказательство теоремы, следует показать, что

$$\frac{1}{L} \sum_{l=1}^{L} \mathcal{H}(P_{e,\,l}) \leqslant \mathcal{H}(\langle P_e \rangle). \tag{4.3.16}$$

Это может быть установлено с помощью неравенства $\log z \leqslant (\log e) \times (z-1)$. Однако мы не будем вдаваться в детали доказательства, так как (4.3.16) является также простым следствием свойства выпуклости энтропии, которое будет рассмотрено в следующем параграфе. |

Мы уже ограничили вероятность ошибки, связанной с источником, с помощью неопределенности $H(\mathbf{U}^L \mid V^L)$. Теперь рассмотрим канал в сочетании с источником.

По определению будем считать, что последовательность источника $\mathbf{u} = (u_1, \ldots, u_L)$ *связана с адресатом при* N*-кратном использовании канала, если совместный ансамбль* $\mathbf{U}^L\mathbf{X}^N\mathbf{Y}^N\mathbf{V}^L$ [*соответствующий выходу источника* $\mathbf{u}$, *входу канала* $\mathbf{x} = (x_1, \ldots, x_N)$, *выходу канала* $\mathbf{y} = (y_1, \ldots, y_N)$ *и декодированному выходу* (*сообщению на выходе*) $\mathbf{v} = (v_1, \ldots, v_L)$] *обладает тем свойством, что* $\mathbf{y}$ *не зависит от* $\mathbf{u}$ *при условии, что задано* $\mathbf{x}$, *а* $\mathbf{v}$ *не зависит от* $\mathbf{u}$ *и* $\mathbf{x}$ *при условии, что задано* $\mathbf{y}$.

Для дискретных ансамблей это условие означает, что $P(\mathbf{y} \mid \mathbf{x}, \mathbf{u}) = P(\mathbf{y} \mid \mathbf{x})$ и $P(\mathbf{v} \mid \mathbf{y}, \mathbf{x}, \mathbf{u}) = P(\mathbf{v} \mid \mathbf{y})$. В общем случае, как показывает рис. 4.3.1, оно означает, что выход канала статистически зависит от последовательности источника лишь через последовательность на входе канала, а декодированный выход $\mathbf{v}$ зависит от $\mathbf{u}$ и $\mathbf{x}$ лишь через выход канала $\mathbf{y}$. Другими словами, эти условия являются математическим выражением того, что нет вспомогательного «скрытого» канала, передающего декодеру информацию о $\mathbf{u}$.

Если источник обладает памятью, то это определение не столь невинно, как оно представляется. Если последовательные блоки из L символов источника передаются адресату, то может быть построен декодер, который использует блок полученных букв при декодировании следующего. Вышеприведенное условие исключает такие декодеры, но, как будет показано позже, эта проблема перестает существовать при переходе к пределу при $L \to \infty$.

Теорема 4.3.3. (Теорема переработки информации.) Пусть последовательность источника $\mathbf{u} = (u_1, \ldots, u_L)$ связана с адресатом каналом, используемым N раз. Тогда

$$I(\mathbf{U}^L;\ \mathbf{V}^L) \leqslant I(\mathbf{X}^N;\ \mathbf{Y}^N), \tag{4.3.17}$$

где $I(\mathbf{U}^L;\ \mathbf{V}^L)$ является средней взаимной информацией между последовательностью источника $\mathbf{u} = (u_1, \ldots, u_L)$ и декодированной вы-

ходной последовательностью $\mathbf{v} = (v_1, \ldots, v_L)$, а $I(\mathbf{X}^N; \mathbf{Y}^N)$ является средней взаимной информацией при N-кратном использовании канала.

Доказательство. Первое условие приведенного выше определения совместно с теоремой 2.3.3 дает: $I(\mathbf{U}^L; \mathbf{Y}^N \mid \mathbf{X}^N) = 0$. Из (2.3.19) теперь получим

$$I(\mathbf{U}^L; \mathbf{Y}^N) \leqslant I(\mathbf{X}^N; \mathbf{Y}^N). \tag{4.3.18}$$

Второе условие определения дает $I(\mathbf{U}^L; \mathbf{V}^L \mid \mathbf{Y}^N) = 0$ и, используя вновь (2.3.19), имеем

$$I(\mathbf{U}^L; \mathbf{V}^L) \leqslant I(\mathbf{U}^L; \mathbf{Y}^N). \tag{4.3.19}$$

Объединяя (4.3.18) и (4.3.19), получаем (4.3.17). |

Из доказанных двух теорем следует, что

$$\langle P_e \rangle \log(M-1) + \mathcal{H}(\langle P_e \rangle) \geqslant \frac{1}{L} H(\mathbf{U}^L \mid \mathbf{V}^L) = H_L(U) - \\ - \frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) \geqslant H_L(U) - \frac{1}{L} I(\mathbf{X}^N; \mathbf{Y}^N), \tag{4.3.20}$$

где $H_L(U) = (1/L) H(\mathbf{U}^L)$.

Если канал является ДКБП, то будем иметь $I(\mathbf{X}^N; \mathbf{Y}^N) \leqslant NC$, отсюда получаем

$$\langle P_e \rangle \log(M-1) + \mathcal{H}(\langle P_e \rangle) \geqslant H_L(U) - \frac{N}{L} C. \tag{4.3.21}$$

Свяжем теперь значения N и L с помощью интервала времени между двумя последовательными буквами источника τ_s и интервалом времени между двумя последовательными буквами канала τ_c. Предположим, что число возможных использований канала задается величиной $N = \lfloor L\tau_s/\tau_c \rfloor$, где $\lfloor x \rfloor$ означает наибольшее целое число, меньшее или равное x.

Теорема 4.3.4. (Обращение теоремы кодирования.) Пусть дискретный стационарный источник с алфавитом объема M имеет энтропию $H_\infty(U) = \lim\limits_{L \to \infty} H_L(U)$ и производит буквы со скоростью одна буква за τ_s секунд. Пусть дискретный канал без памяти имеет пропускную способность C и передача по нему ведется со скоростью одна буква за τ_c секунд. Пусть последовательность источника длины L связана с адресатом каналом, используемым N раз, где $N = \lfloor L\tau_s/\tau_c \rfloor$. Тогда для любого L вероятность ошибки на букву источника $\langle P_e \rangle$ удовлетворяет неравенству

$$\langle P_e \rangle \log(M-1) + \mathcal{H}(\langle P_e \rangle) \geqslant H_\infty(U) - \frac{\tau_s}{\tau_c} C. \tag{4.3.22}$$

Доказательство. Согласно теореме 3.5.1 $H_\infty(U) \leqslant H_L(U)$, и (4.3.22) является непосредственным следствием (4.3.21). |

Для того чтобы дать соответствующую интерпретацию доказанной теореме, рассмотрим $L\tau_s$ как общее время, в течение которого проис-

ходит передача. В течение этого времени кодер может использовать кодирование с фиксированной длиной или неравномерное кодирование для источника и блоковое или неблоковое кодирование для канала. Независимо от того, какое кодирование и какая обработка данных производились, средняя вероятность ошибки на символ источника должна удовлетворять (4.3.22) и, таким образом, она отлична от нуля, если $H_\infty(U)$ (скорость источника) больше, чем $(\tau_s/\tau_c)\,C$ (пропускная способность канала на букву источника). Следует заметить, что теорема ничего не говорит о вероятностях ошибок отдельных букв $P_{e,l}$. С помощью подходящего выбора кодера и декодера мы можем сделать $P_{e,l}$ малым для одних значений l и большим для других.

Хотя теорема в том виде, как она здесь сформулирована, приложима лишь к дискретным каналам без памяти, можно заметить, что это ограничение было использовано только при переходе от (4.3.20) к (4.3.21). Для того чтобы доказать справедливость теоремы для более общих каналов, нужно найти способ определения совместного ансамбля $\mathbf{X}^N\mathbf{Y}^N$, а также определить C таким образом, чтобы $C \to (1/N)\, I(\mathbf{X}^N; \mathbf{Y}^N)$ в пределе при $N \to \infty$. Эта задача будет рассмотрена в § 4.6.

В частном случае канала без шума с D буквами во входном и выходном алфавитах условие (4.3.22) является просто обращением теоремы кодирования для источника. Оно отличается от (3.1.20) тем, что ограничивает снизу вероятность ошибки на символ, а не вероятность ошибки в блоке.

Можно заметить, что граница, задаваемая (4.3.22), довольно слабая при большом объеме M алфавита источника. Для того чтобы показать, что эта слабость границы неизбежна, построим источник, для которого $H_\infty(U)$ произвольно велика, а $\langle P_e \rangle > 0$ произвольно мала даже при $C = 0$. Пусть источник без памяти и его алфавит составляют буквы $a_1, \ldots, a_M$. Пусть ε — сколь угодно малое положительное число и пусть $P(a_1) = 1 - \varepsilon$ и $P(a_m) = \varepsilon/(M-1)$ при $m = 2, \ldots, M$. Тогда, если декодер декодирует каждый символ в букву a_1, то ошибки появляются только тогда, когда источник вырабатывает буквы, отличные от a_1. Таким образом, вероятность ошибки $\langle P_e \rangle$ равна ε. Вместе с тем

$$H_\infty(U) = (1-\varepsilon)\log\frac{1}{1-\varepsilon} + (M-1)\,\frac{\varepsilon}{M-1}\log\frac{M-1}{\varepsilon}. \quad (4.3.23)$$

При любом $\varepsilon > 0$ можно сделать $H_\infty(U)$ сколь угодно большим, выбирая достаточно большое M. Можно заметить, что эта «система связи» удовлетворяет условию (4.3.22) с равенством, если $C = 0$.

4.4. ВЫПУКЛЫЕ ФУНКЦИИ

В этом параграфе мы возвратимся к задаче вычисления пропускной способности дискретного канала без памяти. Как можно видеть из (4.2.3), она включает в себя максимизацию нелинейной функции многих переменных с двумя условиями: одним в виде равенства и другим в виде неравенства. Эта максимизация заметно упрощается, если

воспользоваться свойством выпуклости взаимной информации*). Мы прервем здесь наше рассмотрение для того, чтобы кратко описать свойство выпуклости, которое будет полезно как для этой задачи, так и для ряда последующих подобных задач в этой книге.

Пусть $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ является K-мерным вектором с действительными компонентами, определенным в области R-векторного пространства. *Назовем область R выпуклой, если для любого вектора $\boldsymbol{\alpha}$ из R и любого вектора $\boldsymbol{\beta}$ из R вектор $\theta\boldsymbol{\alpha} + (1 - \theta)\boldsymbol{\beta}$ принадлежит R при* $0 \leqslant \theta \leqslant 1$. Геометрически, когда θ изменяется от 0 до 1, $\theta\boldsymbol{\alpha} + (1 - \theta)\boldsymbol{\beta}$ проходит отрезок прямой линии от $\boldsymbol{\beta}$ до $\boldsymbol{\alpha}$. Таким образом, область является выпуклой, если для каждой пары точек из области отрезок прямой линии между этими точками принадлежит области (рис. 4.4.1).

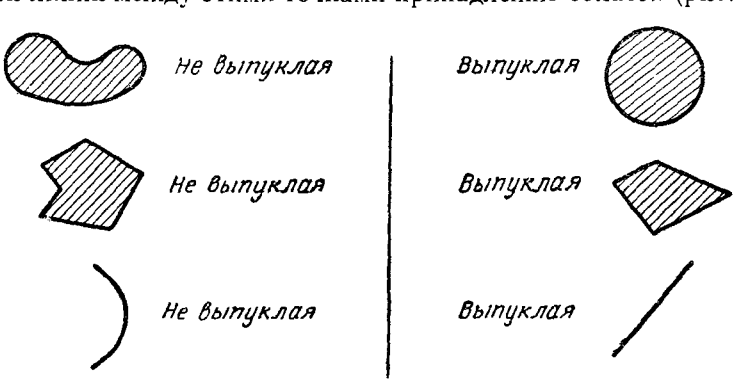


Рис. 4.4.1. Примеры выпуклых областей для двумерных векторов.

Весьма полезным для наших целей примером выпуклого множества является область векторов вероятностей. *Назовем вектор вектором вероятностей, если все его компоненты неотрицательны и в сумме равны 1.* Чтобы показать, что эта область выпукла, предположим, что $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$ являются векторами вероятностей, и положим $\boldsymbol{\gamma} = \theta\boldsymbol{\alpha} + (1 - \theta)\boldsymbol{\beta}$ при $0 \leqslant \theta \leqslant 1$. Тогда

$$\gamma_k = \theta\alpha_k + (1 - \theta)\beta_k. \tag{4.4.1}$$

Следовательно, $\gamma_k \geqslant 0$, а также

$$\sum_{k=1}^{K} \gamma_k = \theta \sum_{k=1}^{K} \alpha_k + (1 - \theta) \sum_{k=1}^{K} \beta_k = 1. \tag{4.4.2}$$

Таким образом, $\boldsymbol{\gamma}$ является вектором вероятностей и область векторов вероятностей выпукла.

Назовем действительную функцию f векторного аргумента выпуклой $\cap$ (следует читать выпуклой вверх) в выпуклой области R вектор-

*) Более полное описание свойства выпуклости можно найти, например, в книге Блекуэлла и Гиршика, Теория игр и статистические решения, гл. 2 (1954).

ного пространства, если для всех α *из* R, β *из* R *и* θ, $0 < \theta < 1$, *функция удовлетворяет условию*

$$\theta f(\alpha) + (1-\theta) f(\beta) \leqslant f[\theta\alpha + (1-\theta)\beta]. \tag{4.4.3}$$

Если имеет место неравенство, обратное (4.4.3) для всех таких α, β и θ, то $f(\alpha)$ *называется выпуклой* $\smile$ (*следует читать выпуклой вниз*). *Если неравенство может быть заменено на строгое неравенство, то* $f(\alpha)$ *называется строго выпуклой* $\frown$ *или строго выпуклой* $\smile$*).

На рис. 4.4.2 изображены выражения, стоящие в двух частях (4.4.3), как функции θ. Можно заметить, что геометрическая интерпретация (4.4.3) состоит в том, что любая хорда, соединяющая две точки на кривой, представляющей функцию, лежит ниже (или на) этой кривой. Причина, по которой область в определении взята выпуклой, состоит в том, чтобы обеспечить то, чтобы вектор, стоящий в правой части (4.4.3), принадлежал R.

Из (4.4.3) непосредственно следует, что если $f(\alpha)$ выпукла $\frown$, то $-f(\alpha)$ выпукла $\smile$, и наоборот. Поэтому мы будем иметь дело только с выпуклыми $\frown$ функциями, так как результаты могут быть легко применены к выпуклым $\smile$ функциям.

Для удобства приведем здесь ряд свойств выпуклых функций, которые часто оказываются полезными.

1) Если $f_1(\alpha)$, ..., $f_L(\alpha)$ являются выпуклыми $\frown$ функциями и если C_1, ..., C_L положительные числа, то функция

$$\sum_l C_l f_l(\alpha)$$

является выпуклой $\frown$ и выпуклость строгая, если какая-либо из $f_l(\alpha)$ строго выпукла.

2) Пусть для одномерного вектора α

$$d^2 f(\alpha)/d\alpha^2 \leqslant 0 \tag{4.4.4}$$

на всем интервале, тогда $f(\alpha)$ выпукла $\frown$ на этом интервале со строгой выпуклостью, если (4.4.4) справедливо со строгим неравенством.

3) Если $(\alpha_1, \ldots, \alpha_L)$ — множество векторов из области, в которой $f(\alpha)$ выпукла $\frown$, и если $(\theta_1, \ldots, \theta_L)$ — множество вероятностей (т. е. $\theta_l \geqslant 0$; $\Sigma\theta_l = 1$), то

$$\sum_{l=1}^{L} \theta_l f(\alpha_l) \leqslant f\left[\sum_{l=1}^{L} \theta_l \alpha_l\right]. \tag{4.4.5}$$

Считая α дискретным случайным вектором и используя черту сверху для обозначения математического ожидания, получаем, что это неравенство эквивалентно

$$\overline{f(\alpha)} \leqslant f(\bar{\alpha}). \tag{4.4.6}$$

---

*) В математической литературе выпуклая $\frown$ функция обычно называется вогнутой, а выпуклая $\smile$ функция — выпуклой. Эта терминология не используется здесь потому, что для большинства людей такое различие очень трудно для запоминания. В недавно устроенном испытании для десяти человек, которые думали, что они помнят это различие, восемь человек спутали выпуклость с вогнутостью.

Подставляя $\Sigma c_l f_l(\boldsymbol{\alpha})$ в неравенство, определяющее выпуклость (4.4.3), устанавливаем справедливость свойства 1. Свойство 2 почти очевидно геометрически (см. рис. 4.4.2) и оно доказано в задаче 4.9. Свойство 3 на геометрическом языке означает, что часть «плоскости», в которой лежат точки $f(\boldsymbol{\alpha}_1)$, ..., $f(\boldsymbol{\alpha}_L)$, находится ниже (или на) поверхности, порождаемой $f(\boldsymbol{\alpha})$ между этими точками (см. для доказательства задачу 4.9).

С помощью этих свойств легко показать, что энтропия ансамбля $-\sum_k P(a_k) \log P(a_k)$ является строго выпуклой $\frown$ функцией входящих в нее вероятностей. Свойство 2 показывает, что $-P(a_k) \times \log P(a_k)$ строго выпукла $\frown$ по $P(a_k)$, а свойство 1 показывает, что

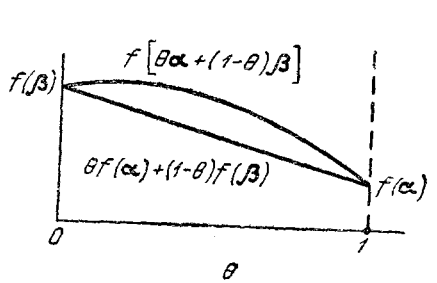


Рис. 4.4.2. Выпуклая функция.

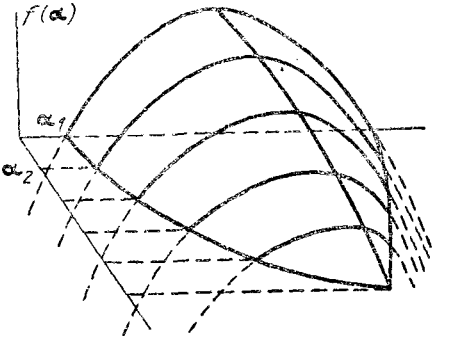


Рис. 4.4.3. Вид выпуклой функции с максимумом в области $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$, который достигается при $\alpha_2 = 0$.

сумма строго выпукла $\frown$. Неравенство (4.3.16), которое мы оставили недоказанным в предыдущем параграфе, следует из (4.4.6) и выпуклости энтропии.

Основной причиной рассмотрения здесь свойства выпуклости является то, что выпуклые $\frown$ функции относительно легко максимизировать в выпуклой области. Для того чтобы показать это, рассмотрим вначале не строго некоторые примеры. Предположим, что $f(\boldsymbol{\alpha})$ — выпуклая $\frown$ функция в области R, где $\alpha_k \geqslant 0$, $1 \leqslant k \leqslant K$. Разумно было бы попытаться найти максимум $f(\boldsymbol{\alpha})$ в R с помощью отыскания стационарной точки $f(\boldsymbol{\alpha})$, т. е. отыскания такого $\boldsymbol{\alpha}$, для которого

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = 0; \quad 1 \leqslant k \leqslant K. \tag{4.4.7}$$

Множество этих уравнений может не иметь решения, а если решения существуют, то они, возможно, не удовлетворяют ограничениям $\alpha_k \geqslant 0$. Покажем, однако, теперь, что если существует такое $\boldsymbol{\alpha}$, которое удовлетворяет как (4.4.7), так и этим ограничениям, то рассматриваемое $\boldsymbol{\alpha}$ максимизирует функцию. Чтобы показать это, предположим, что результат не верен, т. е. что имеется некоторый вектор $\boldsymbol{\beta}$, принадлежащий области, для которого $f(\boldsymbol{\beta}) > f(\boldsymbol{\alpha})$. Ордината хорды, соединяющей $f(\boldsymbol{\beta})$ и $f(\boldsymbol{\alpha})$, в этом случае увеличивается от $f(\boldsymbol{\alpha})$ к $f(\boldsymbol{\beta})$. Как показа-

но на рис. 4.4.2, скорость увеличения функции в точке $\boldsymbol{\alpha}$ в направлении $\boldsymbol{\beta}$ не меньше, чем скорость возрастания ординаты хорды. Таким образом, $\boldsymbol{\alpha}$ не может быть стационарной точкой f и мы пришли к противоречию.

Рассмотрим далее случай, когда максимум $f(\boldsymbol{\alpha})$ в R лежит на границе области, т. е. в точке, где одна или более компонент $\boldsymbol{\alpha}$ равна нулю. В этом случае, как показано на рис. 4.4.3, не следует удивляться тому, что функция является строго убывающей при изменениях аргумента, направленных внутрь области, т. е. при изменениях нулевых компонент $\boldsymbol{\alpha}$. Вместе с тем, если f дифференцируема, то можно ожидать, что максимум будет в стационарной точке, соответствующей вариациям ненулевых компонент $\boldsymbol{\alpha}$. Это предполагает замену (4.4.7) на условия

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = 0 \quad \text{при всех } k, \text{ для которых } \alpha_k > 0, \tag{4.4.8}$$

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} \leqslant 0 \quad \text{при всех } k, \text{ для которых } \alpha_k = 0. \tag{4.4.9}$$

Сейчас мы не будем касаться того, как решить эти уравнения. Важным является то, что если $\boldsymbol{\alpha}$ принадлежит области и удовлетворяет (4.4.8) и (4.4.9), то оно максимизирует f и, наоборот, если f дифференцируема и имеет максимум в области в точке $\boldsymbol{\alpha}$, то (4.4.8) и (4.4.9) удовлетворяются. Ниже будет приведено доказательство этих утверждений.

В качестве второго примера рассмотрим максимизацию выпуклой функции $f(\boldsymbol{\alpha})$ в области, в которой $\boldsymbol{\alpha}$ является вектором вероятностей, т. е. в области, где компоненты $\boldsymbol{\alpha}$ неотрицательны и в сумме равны 1. Условие $\Sigma\alpha_k = 1$ позволяет использовать метод множителей Лагранжа, который означает максимизацию $f(\boldsymbol{\alpha}) - \lambda\Sigma\alpha_k$ по области, в которой компоненты неотрицательны, и выбор λ таким образом, чтобы максимум имел место при $\Sigma\alpha_k = 1$. Применяя (4.4.8) и (4.4.9) к функции $f(\boldsymbol{\alpha}) - \lambda\Sigma\alpha_k$, будем иметь:

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} = \lambda \quad \text{при всех } k, \text{ для которых } \alpha_k > 0, \tag{4.4.10}$$

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial \alpha_k} \leqslant \lambda \quad \text{при всех } k, \text{ для которых } \alpha_k = 0. \tag{4.4.11}$$

Будет показано, что условия (4.4.10) и (4.4.11) являются в действительности необходимыми и достаточными условиями того, что вектор вероятностей $\boldsymbol{\alpha}$ максимизирует выпуклую $\frown$ дифференцируемую функцию f. Другими словами, если вектор вероятностей $\boldsymbol{\alpha}$ удовлетворяет (4.4.10) и (4.4.11) при некотором значении λ, то этот вектор $\boldsymbol{\alpha}$ максимизирует f в области, и, наоборот, если $\boldsymbol{\alpha}$ максимизирует f в области, то (4.4.10) и (4.4.11) удовлетворяются при некотором λ. Приступим теперь к доказательству этого результата.

Теорема 4.4.1. Пусть $f(\boldsymbol{\alpha})$ является выпуклой $\frown$ функцией $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ в области R, где $\boldsymbol{\alpha}$ — вектор вероятностей. Предположим, что частные производные $\partial f(\boldsymbol{\alpha})/\partial\alpha_k$ определены и непрерывны в области R с тем возможным исключением, что $\lim_{\alpha_k \to 0} \partial f(\boldsymbol{\alpha})/\partial\alpha_k$ может быть

равен $+\infty$ для некоторых k. Тогда (4.4.10) и (4.4.11) являются необходимыми и достаточными условиями того, что вектор вероятностей $\boldsymbol{\alpha}$ максимизирует f в области R.

Доказательство. Достаточность. Предположим, что (4.4.10) и (4.4.11) удовлетворяются при некотором λ и некотором векторе вероятностей $\boldsymbol{\alpha}$. Покажем теперь, что $f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})\leqslant 0$ для любого вектора вероятностей $\boldsymbol{\beta}$, что будет означать, что $\boldsymbol{\alpha}$ максимизирует f. Из определения выпуклости следует

$$\theta f(\boldsymbol{\beta})+(1-\theta) f(\boldsymbol{\alpha})\leqslant f[\theta\boldsymbol{\beta}+(1-\theta)\boldsymbol{\alpha}];\quad 0<\theta<1. \tag{4.4.12}$$

Преобразуя (4.4.12), будем иметь

$$f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})\leqslant\frac{f[\theta\boldsymbol{\beta}+(1-\theta)\boldsymbol{\alpha}]-f(\boldsymbol{\alpha})}{\theta}. \tag{4.4.13}$$

В силу того, что (4.4.13) справедливо при всех θ, $0<\theta<1$, можно перейти к пределу и получить

$$f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})\leqslant\frac{df[\theta\boldsymbol{\beta}+(1-\theta)\boldsymbol{\alpha}]}{d\theta}\Bigg|_{\theta=0}. \tag{4.4.14}$$

Выполняя дифференцирование, имеем

$$f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})\leqslant\sum_k\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_k}(\beta_k-\alpha_k). \tag{4.4.15}$$

Существование производной в (4.4.14) и эквивалентность (4.4.14) и (4.4.15) следуют из непрерывности частных производных. Эта непрерывность является следствием предположения, так как (4.4.10) и (4.4.11) исключают случай, когда $\partial f/\partial\alpha_k=+\infty$. Заметим теперь, что

$$\frac{\partial f(\boldsymbol{\alpha})}{\partial\alpha_k}(\beta_k-\alpha_k)\leqslant\lambda(\beta_k-\alpha_k). \tag{4.4.16}$$

Это следует из (4.4.10), если $\alpha_k>0$. Если $\alpha_k=0$, то $\beta_k-\alpha_k\geqslant 0$ и (4.4.16) следует из (4.4.11).

Подставляя (4.4.16) в (4.4.15), получаем

$$f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})\leqslant\lambda\left[\sum_k\beta_k-\sum_k\alpha_k\right]. \tag{4.4.17}$$

В силу того, что $\boldsymbol{\beta}$ и $\boldsymbol{\alpha}$ являются векторами вероятностей, $f(\boldsymbol{\beta})-f(\boldsymbol{\alpha})\leqslant 0$ при каждом $\boldsymbol{\beta}$ из области.

Необходимость. Пусть $\boldsymbol{\alpha}$ максимизирует f в области и предположим, на некоторое время, что частные производные непрерывны в точке $\boldsymbol{\alpha}$. Так как $\boldsymbol{\alpha}$ максимизирует f, то для любого вектора вероятностей $\boldsymbol{\beta}$ и любого θ, $0<\theta<1$, имеем

$$f[\theta\boldsymbol{\beta}+(1-\theta)\boldsymbol{\alpha}]-f(\boldsymbol{\alpha})\leqslant 0. \tag{4.4.18}$$

Разделив это выражение на θ и переходя к пределу при $\theta\to 0$, будем иметь

$$\frac{df[\theta\boldsymbol{\beta}+(1-\theta)\boldsymbol{\alpha}]}{d\theta}\Bigg|_{\theta=0}\leqslant 0, \tag{4.4.19}$$

$$\sum_k \frac{\partial f(\alpha)}{\partial \alpha_k}(\beta_k - \alpha_k) \leqslant 0. \tag{4.4.20}$$

По крайней мере одна компонента $\boldsymbol{\alpha}$ является строго положительной, предположим для простоты обозначений, что $\alpha_1 > 0$. Пусть i_k является единичным вектором с единицей в качестве k-й компоненты и остальными нулевыми компонентами; выберем $\boldsymbol{\beta}$ равным $\boldsymbol{\alpha} + \varepsilon i_k - \varepsilon i_1$. Так как $\alpha_1 > 0$, то $\boldsymbol{\beta}$ является вектором вероятностей при $0 \leqslant \varepsilon \leqslant \alpha_1$. Используя это $\boldsymbol{\beta}$ в (4.4.20), получаем

$$\varepsilon \frac{\partial f(\alpha)}{\partial \alpha_k} - \varepsilon \frac{\partial f(\alpha)}{\partial \alpha_1} \leqslant 0, \tag{4.4.21}$$

$$\frac{\partial f(\alpha)}{\partial \alpha_k} \leqslant \frac{\partial f(\alpha)}{\partial \alpha_1}. \tag{4.4.22}$$

Если $\alpha_k > 0$, то ε можно также выбрать отрицательным, и в этом случае неравенство в (4.4.22) изменяется на обратное, что приводит к

$$\frac{\partial f(\alpha)}{\partial \alpha_k} = \frac{\partial f(\alpha)}{\partial \alpha_1}, \quad \alpha_k > 0. \tag{4.4.23}$$

Выбирая, наконец, λ равной $\partial f(\boldsymbol{\alpha})/\partial \alpha_1$, приходим к эквивалентности (4.4.22) и (4.4.23) условиям (4.4.10) и (4.4.11). Для завершения доказательства теоремы рассмотрим $\boldsymbol{\alpha}$, для которого $\partial f(\boldsymbol{\alpha})/\partial \alpha_k = +\infty$ при некотором k, и покажем, что такое $\boldsymbol{\alpha}$ не может максимизировать f. Предположим для простоты обозначений, что $\alpha_1 > 0$. Будем иметь

$$\frac{f(\alpha + \varepsilon i_k - \varepsilon i_1) - f(\alpha)}{\varepsilon} = \frac{f(\alpha + \varepsilon i_k - \varepsilon i_1) - f(\alpha + \varepsilon i_k)}{\varepsilon} + \frac{f(\alpha + \varepsilon i_k) - f(\alpha)}{\varepsilon}. \tag{4.4.24}$$

В пределе при $\varepsilon \to 0$ первое слагаемое в правой части равенства (4.4.24) остается ограниченным в силу непрерывности $\partial f/\partial \alpha_1$. Второе слагаемое возрастает так, что левая часть (4.4.24) является положительной для достаточно малого ε. Это показывает, что $\boldsymbol{\alpha}$ не максимизирует f, что завершает доказательство теоремы. |

Как можно догадаться, после рассмотрения свойства выпуклости в этом параграфе мы собираемся показать, что взаимная информация является выпуклой $\frown$ функцией входных вероятностей.

Теорема 4.4.2. Пусть дискретный канал без памяти с K буквами на входе и J буквами на выходе имеет переходные вероятности $P(j \mid k)$, $0 \leqslant j \leqslant J-1$, $0 \leqslant k \leqslant K-1$. Пусть $\mathbf{Q} = [Q(0), \ldots, Q(K-1)]$ — произвольное распределение вероятностей на входе канала. Тогда

$$I(X; Y) = \sum_{j,k} Q(k) P(j \mid k) \log \frac{P(j \mid k)}{\sum_i Q(i) P(j \mid i)} \tag{4.4.25}$$

является выпуклой $\frown$ функцией $\mathbf{Q}$.

Доказательство. Пусть $\mathbf{Q}_0$ и $\mathbf{Q}_1$ — произвольные векторы вероятностей на входе канала и пусть I_0 и I_1 — соответствующие им сред-

ние взаимные информации. Пусть θ — произвольное число $0 < \theta < 1$; пусть $\mathbf{Q} = \theta\mathbf{Q}_0 + (1 - \theta)\mathbf{Q}_1$ и пусть I является средней взаимной информацией при распределении вероятностей $\mathbf{Q}$ на входе. Нужно показать, что

$$\theta I_0 + (1 - \theta) I_1 \leqslant I. \tag{4.4.26}$$

При желании можно рассматривать $\mathbf{Q}_0$ и $\mathbf{Q}_1$ как условные вероятности при условии, что задано значение двоичной случайной величины z (рис. 4.4.4), т. е.

$$Q_0(k) = Q_{X|Z}(k|0); \quad Q_1(k) = Q_{X|Z}(k|1). \tag{4.4.27}$$

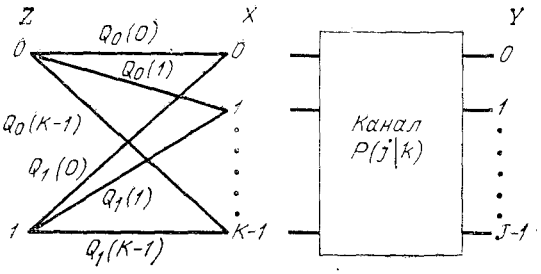


Рис. 4.4.4.

Выберем $P_Z(0) = \theta$, $P_Z(1) = 1 - \theta$ и (как показано на рис. 4.4.4) $P(y|x, z) = P(y|x)$. В обозначениях этого трехмерного ансамбля имеем, что левая часть (4.4.26) равна $I(X; Y|Z)$, а правая часть равна $I(X; Y)$.

Так же как при рассмотрении последовательных каналов в § 2.3, z и y являются независимыми при условии, что задано x. Следовательно, как и в (2.3.15),

$$I(Y; Z|X) = 0. \tag{4.4.28}$$

Кроме того, так же как в (2.3.16) и (2.3.17),

$$I(Y; ZX) = I(Y; Z) + I(Y; X|Z) = \tag{4.4.29}$$

$$= I(Y; X) + I(Y; Z|X). \tag{4.4.30}$$

Приравнивая правые части (4.4.29) и (4.4.30) и используя (4.4.28) получаем

$$I(Y; Z) + I(Y; X|Z) = I(Y; X), \tag{4.4.31}$$

$$I(Y; X|Z) \leqslant I(Y; X), \tag{4.4.32}$$

$$I(X; Y|Z) \leqslant I(X; Y). \tag{4.4.33}$$

Так как это эквивалентно (4.4.26), то доказательство теоремы закончено. |

Другое (более прямое) доказательство намечено в задаче 4.16. Приведенное здесь доказательство обладает преимуществом большей общности и оно применимо, в сущности, к любым каналам. До того как использовать этот результат при вычислении пропускной способности

канала, докажем тесно связанную с ним теорему, которая полезна как при исследовании неизвестных каналов, так и в гл. 9.

Теорема 4.4.3. Рассмотрим $I(X; Y)$ в (4.4.25) как функцию переходных вероятностей $P(j \mid k)$ и входного распределения $Q(k)$. При фиксированном входном распределении $I(X; Y)$ является выпуклой $\smile$ функцией переходных вероятностей (отметим, что она является выпуклой $\smile$, а не выпуклой $\frown$, как в теореме 4.4.2).

Доказательство. Пусть $P_0(j \mid k)$ и $P_1(j \mid k)$, $0 \leqslant k \leqslant K-1$; $0 \leqslant j \leqslant J-1$, два произвольных множества переходных вероятностей и пусть $P(j \mid k) = \theta P_0(j \mid k) + (1-\theta) P_1(j \mid k)$ для любого θ, $0 < \theta < 1$. Пусть I_0, I_1 и I являются средними взаимными информациями для этих множеств переходных вероятностей. Требуется показать, что

$$\theta I_0 + (1-\theta) I_1 \geqslant I. \tag{4.4.34}$$

Как и в последней теореме, вероятности P_0 и P_1 можно рассмотреть как условные при условии, что задана двоичная случайная величина z, т. е.

$$P_0(j \mid k) = P_{Y \mid XZ}(j \mid k, 0); \quad P_1(j \mid k) = P_{Y \mid XZ}(j \mid k, 1). \tag{4.4.35}$$

Полагая $P_Z(0) = \theta$, $P_Z(1) = 1-\theta$ и определяя z статистически независимой от x, находим, что левая часть (4.4.34) равна $I(X; Y \mid Z)$, а правая часть равна $I(X; Y)$. Продолжая далее, так же как и в последней теореме, будем иметь

$$I(X; YZ) = I(X; Z) + I(X; Y \mid Z) = \tag{4.4.36}$$
$$= I(X; Y) + I(X; Z \mid Y). \tag{4.4.37}$$

Так как x и z статистически независимы, то $I(X; Z) = 0$ и

$$I(X; Y|Z) = I(X; Y) + I(Z; X \mid Y), \tag{4.4.38}$$

$$I(X; Y \mid Z) \geqslant I(X; Y). \tag{4.4.39}$$

Это эквивалентно (4.4.34); теорема доказана. |

4.5. НАХОЖДЕНИЕ ПРОПУСКНОЙ СПОСОБНОСТИ ДИСКРЕТНОГО КАНАЛА БЕЗ ПАМЯТИ

Теорема 4.5.1. Для дискретного канала без памяти с переходными вероятностями $P(j \mid k)$ необходимые и достаточные условия того, что на входном векторе вероятностей $\mathbf{Q} = [Q(0), \ldots, Q(K-1)]$ достигается пропускная способность канала, состоят в том, что для некоторого числа C

$$I(x = k; Y) = C \text{ при всех } k, \text{ для которых } Q(k) > 0, \tag{4.5.1}$$

$$I(x = k; Y) \leqslant C \text{ при всех } k, \text{ для которых } Q(k) = 0, \tag{4.5.2}$$

где $I(x = k; Y)$ является взаимной информацией при входе k, усредненной по выходам, т. е.

$$I(x = k; Y) = \sum P(j \mid k) \log \frac{P(j \mid k)}{\sum_i Q(i) P(j \mid i)}. \tag{4.5.3}$$

Более того, число C равно пропускной способности канала.

Доказательство. Требуется максимизировать величину

$$I(X;\ Y)=\sum_{k,\ j} Q(k)\,P(j\mid k)\log\frac{P(j\mid k)}{\sum_i Q(i)\,P(j\mid i)} \tag{4.5.4}$$

по всем $\mathbf{Q}$. Взяв частные производные, получим*)

$$\frac{\partial I(X;\ Y)}{\partial Q(k)}=I(x=k;\ Y)-\log e. \tag{4.5.5}$$

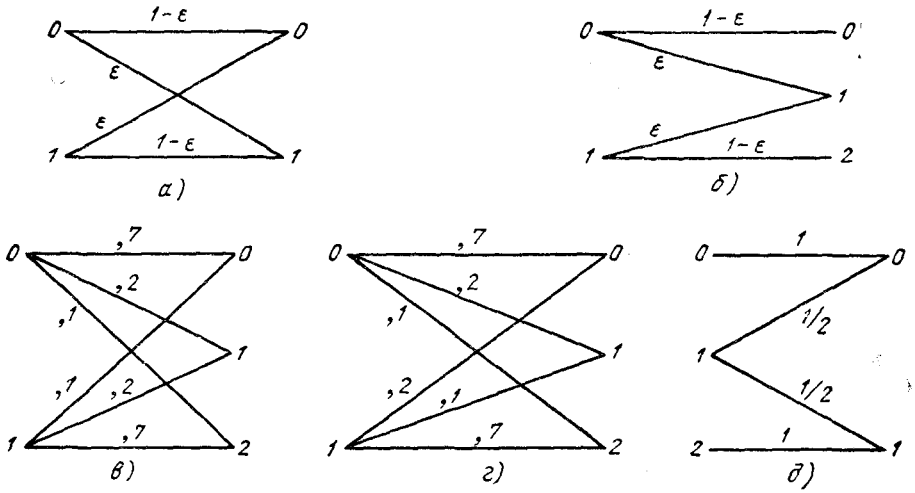


Рис. 4.5.1.

Можно применить теорему 4.4.1 для того, чтобы произвести максимизацию; это можно сделать, так как $I(X;\ Y)$ является выпуклой $\frown$ по $\mathbf{Q}$ и частные производные удовлетворяют соответствующим условиям непрерывности. Таким образом, необходимыми и достаточными условиями, которым должно удовлетворять $\mathbf{Q}$, максимизирующее $I(X;\ Y)$, являются условия

$$\frac{\partial I(X;\ Y)}{\partial Q(k)}=\lambda;\quad Q(k)>0, \tag{4.5.6}$$

$$\leqslant\lambda;\quad Q(k)=0. \tag{4.5.7}$$

Используя (4.5.5) и полагая $C=\lambda+\log e$, получаем (4.5.1) и (4.5.2). Умножая обе стороны (4.5.1) на $Q(k)$ и суммируя по всем k, для которых $Q(k)>0$, получаем слева максимальное значение $I(X;\ Y)$ и справа постоянную C, что показывает, что C действительно является пропускной способностью канала. |

*) Отметим, что сумма в знаменателе, стоящая под знаком логарифма в (4.5.4), содержит слагаемое $Q(k)\,P(j\mid k)$; это слагаемое приводит к $\log e$ в (4.5.5).

Теорема 4.5.1 допускает простую интуитивно понятную интерпретацию. Если одна входная буква приводит к большей взаимной информации, чем другая буква, то можно увеличить среднюю взаимную информацию, используя входы с большей взаимной информацией чаще. Однако любое такое изменение будет изменять взаимную информацию каждой буквы и, следовательно, после достаточного числа изменений все входы будут иметь одну и ту же информацию, за исключением, возможно, нескольких входов, которые столь малозначительны, что их вероятности сводятся к нулю.

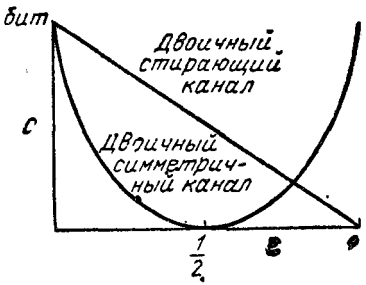


Рис. 4.5.2. Пропускная способность двоичного симметричного канала и двоичного стирающего канала.

Несмотря на изящность теоремы 4.5.1, остается открытым вопрос о том, как использовать ее при практическом вычислении пропускной способности канала. Ниже будут даны несколько методов практического вычисления пропускной способности. Начнем с рассмотрения примеров, изображенных на рис. 4.5.1.

В двоичном симметричном канале (ДСК), изображенном на рис. 4.5.1, *а* можно использовать симметрию, чтобы догадаться, что пропускная способность достигается на $Q(0) = Q(1) = 1/2$. Проверяя эту догадку с помощью (4.5.1), получаем $I(x = 0;\ Y) = I(x = 1;\ Y) = 1 - \mathcal{H}(\varepsilon)$ бит, где $\mathcal{H}(\varepsilon) = -\varepsilon \log_2 \varepsilon - (1-\varepsilon) \times \log_2 (1-\varepsilon)$ является энтропией двоичной случайной величины, принимающей значения с вероятностями ε и $(1-\varepsilon)$. Так как (4.5.1) справедливо для обоих входов, то это **Q** приводит к пропускной способности и $C = 1 - \mathcal{H}(\varepsilon)$ бит. Таким образом, одно из важных применений теоремы состоит в том, что она дает простой тест для проверки любой гипотезы относительно достижения пропускной способности на заданных входных вероятностях. Это означает также, что можно проявить математическую беззаботность при отыскании **Q**, которое приводит к пропускной способности, так как результат поддается легкой проверке. Пропускную способность двоичного стирающего канала (ДСтК), изображенного на рис. 4.5.1, *б*, можно найти подобным же образом. Можно догадаться, что $Q(0) = Q(1) = 1/2$ и проверить, что $I(x = 0;\ Y) = I(x = 1;\ Y) = C = 1 - \varepsilon$. Эти пропускные способности изображены на рис. 4.5.2. На рис. 4.5.1, *в* заметим ту же самую симметрию и опять убедимся с помощью проверки, что пропускная способность достигается на $Q(0) = Q(1) = 1/2$. Рис. 4.5.1, *г* при поверхностном рассмотрении кажется подобным рис. 4.5.1, *в*, однако в некотором отношении он является менее симметричным и если предположить, что $Q(0) = Q(1) = 1/2$, то найдем, что (4.5.1) не удовлетворяется и, следовательно, пропускная способность не достигается на этих вероятностях.

При определении того, что мы понимаем под симметричным каналом, было бы довольно неуклюже рассматривать каналы поодиночке

и говорить, что такой канал как (*в*) является симметричным, а такой канал, как (*г*) — нет. Введя определение симметрии, мы заметим, что в канале (*в*) имеются два выхода 0 и 2, которые подобны друг другу, и другой выход, который отличен. После этого будет неудивительно, что общее определение симметричных каналов будет включать в себя разбиение множества выходов на подмножества подобных выходов.

ДКБП называется симметричным, если множество выходов может быть разбито на подмножества таким образом, что для каждого подмножества матрица переходных вероятностей (используя входы как строки и выходы подмножества как столбцы) обладает тем свойством, что каждая строка является перестановкой любой другой строки и каждый столбец (если их больше чем 1) является перестановкой любого другого столбца. Так, например, разбивая выходы канала (*в*) на рис. 4.5.1 на 0, 2 и 1

		j		
		0	2	1
k	0	0,7	0,1	0,2
	1	0,1	0,7	0,2

Так как каждая из приведенных выше матриц обладает перестановочным свойством, приведенным в определении, то канал является симметричным.

Теорема 4.5.2. В симметричном дискретном канале без памяти пропускная способность достигается при использовании входных букв с равной вероятностью.

Доказательство. При равновероятных входах имеем

$$I(x=k;\ Y) = \sum_{j=0}^{J-1} P(j \mid k) \log \frac{P(j \mid k)}{(1/K)\sum_i P(j \mid i)}. \qquad (4.5.8)$$

При каком-либо разбиении выходов каждый столбец $P(j \mid k)$-матрицы является перестановкой любого другого столбца и, таким образом, выходная вероятность $(1/K)\sum_i P(j \mid i)$ одна и та же для всех выходов в разбиении. Это значит, что при разбиении выходов матрица с элементами $P(j \mid k)$ $I_{X;\,Y}(k;\ j)$ имеет те же самые перестановочные свойства, что и $P(j \mid k)$ и, таким образом, сумма этих слагаемых в (4.5.8) одна и та же для каждого входа. Следовательно, (4.5.1) справедливо и пропускная способность достигается. |

Рассмотрим далее, как найти пропускную способность канала (*д*) на рис. 4.5.1. С интуитивных позиций кажется, что вход 1 плохо выбирать при передаче информации и наше предположение будет состоять в том, что пропускная способность достигается на $Q(0) = Q(2) = {}^1/_2$. Проверяя это предположение, находим, что $I(x=0;\ Y) = I(x=2;\ Y) = 1$ и $I(x=1;\ Y) = 0$. Таким образом, (4.5.1) и (4.5.2) удовлетворяются, и доказано, что наше предположение правильно.

К сожалению, иногда возникает интерес к отысканию пропускной способности канала, который не является симметричным и для которого невозможно угадать оптимальное $\mathbf{Q}$. Самым легким является использование вычислительной машины для отыскания максимума. В силу того, что функция является выпуклой $\frown$, программа для вычислительной машины состоит просто в варьировании $\mathbf{Q}$ для увеличения $I(X; Y)$ до достижения локального (и, следовательно, глобального) максимума.

Если кто-либо особенно настаивает, однако, на отыскании традиционного решения для условий (4.5.1) и (4.5.2), то можно иногда найти ответ следующим образом. Предположим сначала, что все $Q(k)$ ненулевые. Перепишем теперь (4.5.1) в виде

$$\sum_{j=0}^{J-1} P(j \mid k) \log P(j \mid k) - \sum P(j \mid k) \log \omega(j) = C, \tag{4.5.9}$$

где выходные вероятности $\omega(j)$ равны

$$\omega(j) = \sum_{k=0}^{K-1} Q(k) P(j \mid k). \tag{4.5.10}$$

Из (4.5.9) и (4.5.10) имеем

$$\sum P(j \mid k) [C + \log \omega(j)] = \sum_{j} P(j \mid k) \log P(j \mid k). \tag{4.5.11}$$

Это дает K линейных уравнений для J неизвестных $(C + \log \omega(j))$. Если $K = J$ и если матрица $P(j \mid k)$ является невырожденной, то эти линейные уравнения могут быть решены. Если решениями будут $\beta_j = C + \log \omega(j)$, то C можно найти из условия $\Sigma \omega(j) = 1$, что приводит к равенству

$$C = \log_2 \sum_{j} 2^{\beta_j} \text{ бит}, \tag{4.5.12}$$

где все логарифмы взяты по основанию 2. К сожалению, нет уверенности в том, что C, задаваемая равенством (4.5.12), является пропускной способностью канала. Сначала нужно использовать C, чтобы найти $\omega(j)$, и затем провести решение (4.5.10) относительно входных вероятностей $Q(k)$. Если все $Q(k)$ неотрицательны, то решение правильно, в противном случае — неправильно.

Если $J > K$, то равенства (4.5.11) в общем случае имеют неединственное решение, однако только одно из них будет давать решение для (4.5.10). Нахождение C в этом случае состоит в решении системы нелинейных уравнений. Даже если решение найдено, может оказаться, что некоторые $Q(k)$ будут отрицательными, и это делает решение непригодным.

Это завершает наше рассмотрение отыскания пропускной способности. Можно догадаться использовать симметрию канала, использовать процедуру поиска с помощью вычислительной машины или решить уравнения, но решение уравнений часто является довольно громоздкой, если не невыполнимой задачей. В заключение этого параграфа приведем несколько интересных следствий теорем 4.5.1 и 4.4.2.

Следствие 1. Для любого распределения вероятностей на входе, на котором достигается пропускная способность дискретного канала без памяти, все выходные вероятности являются положительными. (Здесь предполагается, что каждый выход можно получить из некоторого входа.)

Доказательство. Пусть $\omega(j)$ является вероятностью выхода j; на основании (4.5.2) имеем

$$\sum_{j=0}^{J-1} P(j \mid k) \log \frac{P(j \mid k)}{\omega(j)} \leqslant C \text{ при всех } k, \text{ для которых } Q(k) = 0, \tag{4.5.13}$$

Если какая-либо $\omega(j) = 0$, то любой вход, который дает выход j с ненулевой переходной вероятностью, должен иметь $Q(k) = 0$. Для такого k левая часть (4.5.13) равна бесконечности, что приводит к противоречию. |

Следствие 2. Выходной вектор вероятностей, на котором достигается пропускная способность, является единственным. Все входные векторы вероятностей с соответствующими нулевыми компонентами, которые приводят к этому выходному вектору, являются такими, на которых достигается пропускная способность.

Доказательство. Пусть $\mathbf{Q}_0$ и $\mathbf{Q}_1$ — какие-либо два входных вероятностных вектора, на которых достигается пропускная способность C. При θ, лежащем между 0 и 1, входной вектор вероятностей $\theta\mathbf{Q}_0 + (1 - \theta)\mathbf{Q}_1$ также приводит к пропускной способности, так как выпуклость $\frown$ информации $I(X; Y)$ показывает, что средняя взаимная информация при $\theta\mathbf{Q}_0 + (1 - \theta)\mathbf{Q}_1$ не может быть меньше чем C. Используя те же самые условные вероятности

$$Q_0(k) = Q_{X|Z}(k \mid 0) \text{ и } Q_1(k) = Q_{X|Z}(k \mid 1),$$

как и в доказательстве теоремы 4.4.2, можно заметить, что $I(X; Y) = I(X; Y \mid Z)$. Из (4.4.31), однако, следует, что $I(Y; Z) = 0$. Поэтому y и z статистически независимы и $P_{Y|Z}(j \mid 0) = P_Y(j) = P_{Y|Z}(j \mid 1)$. Это значит, что одно и то же распределение вероятности на выходе соответствует как $\mathbf{Q}_0$, так и $\mathbf{Q}_1$ и этот выходной вектор вероятностей является единственным. Наконец, в силу того, что условия (4.5.1) и (4.5.2) зависят от $Q(k)$ только через выходные вероятности $\sum_k Q(k) P(j \mid k)$, то любое $\mathbf{Q}$ (с соответствующим образом выбранными нулевыми компонентами), приводящее к этому выходному вектору вероятностей, удовлетворяет (4.5.1) и (4.5.2) и, следовательно, приводит к пропускной способности. |

Следствие 3. Пусть m наименьшее число входов, которое может быть использовано с ненулевыми вероятностями при достижении пропускной способности и пусть A обозначает это множество входов. Тогда $m \leqslant J$, где J — объем выходного алфавита, и входное распределение вероятности на A, на котором достигается пропускная способность при использовании только входов из A, является единственным.

Доказательство. Пусть $\boldsymbol{\omega} = [\omega(0), \ldots, \omega(J-1)]$ является выходным вектором вероятностей, на котором достигается пропускная способность канала. Входные вероятности, которые считаются неравными нулю только на множестве A, должны удовлетворять равенству

$$\sum_{k \in A} Q(k) P(j \mid k) = \omega(j), \quad 0 \leqslant j \leqslant J-1. \tag{4.5.14}$$

Это представляет собой систему J уравнений с m неизвестными и по предположению она имеет по крайней мере одно решение (заметим, что любое решение удовлетворяет условию $\Sigma Q(k) = 1$). Предположим, что решение не единственно; пусть $\mathbf{Q}$ — вектор вероятностей, дающий решение, и пусть $\mathbf{h}$ — ненулевое решение однородных уравнений. Тогда $\mathbf{Q} + \theta\mathbf{h}$ удовлетворяет (4.5.14). Будем увеличивать θ, начиная от 0 до тех пор, пока некоторая компонента $\mathbf{Q} + \theta\mathbf{h}$ не достигнет 0. Это всегда можно сделать, так как $\Sigma h(k) = 0$ и $\mathbf{h}$ имеет отрицательные компоненты. Этот вектор $\mathbf{Q} + \theta\mathbf{h}$ приводит к пропускной способности при $m - 1$ ненулевых компонентах и предположение о неединственности решения является ошибочным. В силу того, что решение единственно, число неизвестных m меньше или равно числу уравнений J. |

4.6. ДИСКРЕТНЫЕ КАНАЛЫ С ПАМЯТЬЮ

Для дискретного канала с памятью каждая буква выходной последовательности статистически зависит как от соответствующего входа, так и от прошлых входов и выходов (здесь и в дальнейшем предполагается, что канал является каналом без предвосхищения, т. е. при заданном текущем входе и заданных входах и выходах в прошлом текущий выход статистически не зависит от будущих входов). Без потери общности последовательность входов и выходов вплоть до заданного момента может быть рассмотрена как состояние канала в этот момент. В этих понятиях статистическое поведение канала описывается совместной вероятностной мерой на выходной букве и состоянии в заданный момент при условии, что заданы текущая входная буква и предыдущее состояние.

При построении математической модели физических каналов с памятью часто желательно рассматривать в качестве памяти канала некоторые имеющие физический смысл параметры (такие, как уровень замирания в линии передачи с медленными замираниями). В этих случаях канал продолжает описываться вероятностной мерой на выходе и состоянии при условии, что заданы вход и предыдущее состояние, но состояние, возможно, не определяется предыдущими входами и выходами.

Для простоты исследования здесь будут рассмотрены только каналы с дискретным конечным множеством состояний, т. е. каналы с конечным множеством возможных состояний, вероятности которых не зависят от времени. Точнее, дискретный канал с конечным множеством состояний имеет на входе последовательность $\mathbf{x} = \ldots, x_{-1}, x_0, x_1, \ldots$, на выходе последовательность $\mathbf{y} = \ldots, y_{-1}, y_0, y_1, \ldots$ и

последовательность состояний $\mathbf{s} = \ldots, s_{-1}, s_0, s_1, \ldots$. Входные буквы x_n принимают значения из алфавита $\{0, 1, \ldots, K-1\}$, выходные буквы y_n принимают значения из алфавита $\{0, 1, \ldots, J-1\}$, а состояния s_n — значения из множества $\{0, 1, \ldots, A-1\}$. Статистически канал описывается условной вероятностью $P(y_n, s_n \mid x_n, s_{n-1})$. Эта вероятность не зависит от n, и P можно рассматривать просто как функцию четырех переменных, каждая из которых принимает целочисленные значения, $0 \leqslant y_n \leqslant J-1$, $0 \leqslant s_n \leqslant A-1$, $0 \leqslant s_{n-1} \leqslant A-1$, $0 \leqslant x_n \leqslant K-1$. Будем считать, что при условии, что заданы x_n и s_{n-1}, пара y_n, s_n статистически независима от всех выходов, входов и состояний, предшествующих y_n, x_n и s_{n-1} соответственно.

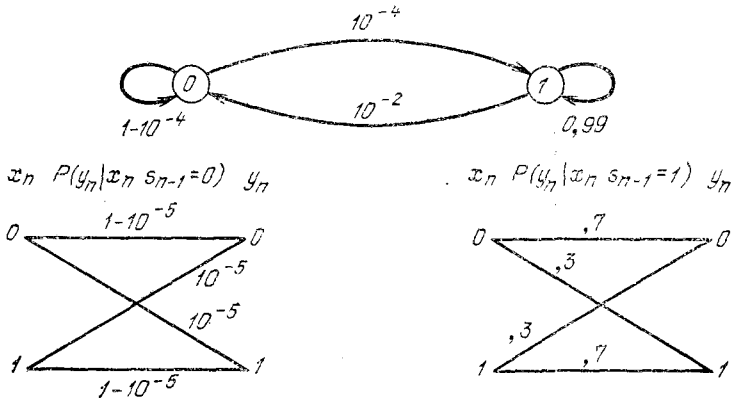
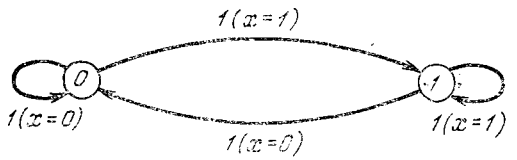


Рис. 4.6.1. Канал с конечным числом состояний; простая модель канала с замираниями (с пачками ошибок).

Последующие примеры являются частными случаями каналов с конечным числом состояний (ККЧС), в которых имеется статистическая независимость между y_n и s_n, при условии, что заданы x_n и s_{n-1}, т. е. $P(y_n, s_n \mid x_n, s_{n-1}) = P(y_n \mid x_n, s_{n-1})\, q(s_n \mid x_n, s_{n-1})$. Можно представить q с помощью графа, а P с помощью обычных диаграмм, как это показано на рис. 4.6.1 и 4.6.2. На графах состояния изображены с помощью маленьких кружков. Направленные ребра обозначают переходы из одного состояния в другое; число на каждом ребре указывает вероятность перехода. Если вероятность перехода зависит от x_n, то значение x_n, соответствующее этой вероятности, дано в круглых скобках. Например, самое верхнее ребро на рис. 4.6.1 представляет переход из состояния 0 в состояние 1. Число, написанное на ребре, 10^{-4}, является условной вероятностью перехода в состояние 1 при условии того, что задано предыдущее состояние 0 и задана 0 или 1 как текущая входная буква, т. е. $q(s_n \mid x_n, s_{n-1}) = 10^{-4}$ при $s_{n-1} = 0$, $s_n = 1$. Подобно этому самое верхнее ребро на рис. 4.6.2 указывает, что для этого ККЧС $q(s_n \mid x_n, s_{n-1}) = 1$ при $s_n = 1$, $s_{n-1} = 0$, $x_n = 1$.

Заметим, что канал, изображенный на рис. 4.6.1, имеет тенденцию пребывать в том состоянии, в котором он находится, оставаясь в состоянии 0 для типичной серии из 10^4 символов и в состоянии 1 для

типичной серии из 100 символов, Этот тип канала дает простую для понимания (но не всецело адекватную) модель для двоичной передачи данных по линиям связи с медленными замираниями. Большинство времени канал находится в состоянии 0, фактически не внося ошибки в передаваемые символы. Случайно канал переходит в состояние 1 (состояние замирания) и в течение примерно 100 символов около 3/10 принятых на выходе канала символов будут ошибочными. Этот канал можно представить себе как ДСК с зависящей от времени вероятностью ошибки, которая попеременно принимает значения 10^{-5} и 0,3. Последовательность состояний, которая определяет эту вероят-



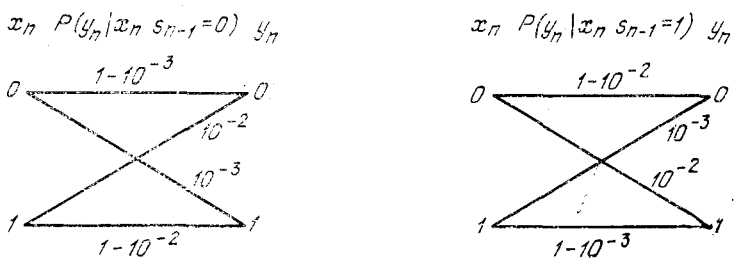


Рис. 4.6.2. ККЧС; простая модель межсимвольной интерференции.

ность ошибки, является цепью Маркова. Эта модель не является полностью удовлетворительной моделью двоичной передачи данных по линии связи с замираниями отчасти в связи с тем, что ее, как не удивительно, трудно изучать, а отчасти потому, что каждый реальный канал оказывается требует отличную модель с конечным числом состояний (обычно с более чем двумя состояниями).

Канал, изображенный на рис. 4.6.2, является простой моделью канала с межсимвольной интерференцией. Можно заметить, что значение состояния, принимаемое в какой-либо момент времени, совпадает со входом в этот момент. Таким образом, $P(y_n | x_n, s_{n-1})$ в этом случае дает вероятность текущего выхода при условии, что заданы текущий и предыдущий входы. Вероятность ошибки (т. е. того, что $y_n \neq x_n$) больше, когда $x_n \neq x_{n-1}$, чем, когда $x_n = x_{n-1}$. Подсчитывая вероятность выхода при заданном текущем входе, находим, $P(y_n | x_n) = Q(0)\ P(y_n | x_n,\ 0) + Q(1)\ P(y_n | x_n,\ 1)$, где $\mathbf{Q}$ — распределение вероятностей для x_{n-1}. Важно отметить, что $P(y_n | x_n)$ зависит от распределения на входе канала и, таким образом, не определяется полностью только каналом. Это вообще характерно для любого канала с конечным числом состояний, в котором последовательность состоя-

ний статистически зависит от входной последовательности. Таким образом, для этого класса каналов не только не выполняется равенство (4.2.1), т. е.

$$P_N(\mathbf{y} \mid \mathbf{x}) = \prod_n P(y_n \mid x_n)$$

не имеет места, *но также вероятности, входящие в это выражение, не определяются только через понятия, описывающие канал.*

Будем называть каналы, такие, как на рис. 4.6.1, где $q(s_n \mid x_n, s_{n-1})$ не зависит от x_n, каналами без межсимвольной интерференции, а каналы, такие, как на рис. 4.6.2, где $q(s_n \mid x_n, s_{n-1})$ принимает только значения 1 и 0 и зависит от x_n, каналами, в которых имеется только память, связанная с межсимвольной интерференцией. В первом случае память возникает лишь из-за шума, а во втором случае — только из-за предыдущих входов. В общем ККЧС, конечно, присутствуют оба эффекта.

В силу того, что $P_N(\mathbf{y} \mid \mathbf{x})$ не определена в общем случае для ККЧС, в основном мы будем иметь дело с $P_N(\mathbf{y}, s_N \mid \mathbf{x}, s_0)$ — вероятностью заданных выходной последовательности $\mathbf{y} = (y_1, \ldots, y_N)$ и окончательного состояния s_N в момент N при условии, что заданы входная последовательность $\mathbf{x} = (x_1, \ldots, x_N)$ и начальное состояние s_0 в момент 0. Эта вероятность может быть найдена по индукции из равенства

$$P_N(\mathbf{y}, s_N \mid \mathbf{x}, s_0) = \sum_{s_{N-1}} P(y_N, s_N \mid x_N, s_{N-1}) P_{N-1}(\mathbf{y}_{N-1}, s_{N-1} \mid \mathbf{x}_{N-1}, s_0), \quad (4.6.1)$$

где $\mathbf{x}_{N-1} = (x_1, \ldots, x_{N-1})$ и $\mathbf{y}_{N-1} = (y_1, \ldots, y_{N-1})$.

Можно просуммировать это выражение по окончательному состоянию и получить

$$P_N(\mathbf{y} \mid \mathbf{x}, s_0) = \sum_{s_N} P_N(\mathbf{y}, s_N \mid \mathbf{x}, s_0). \quad (4.6.2)$$

Пропускную способность ККЧС можно с достаточными основаниями определить несколькими различными способами. Дадим здесь два определения (которые при вычислениях в общем случае приводят к различным числовым значениям) и затем покажем на некоторых примерах смысл каждого из этих определений. Определим нижнюю пропускную способность ККЧС, $\underline{C}$, как

$$\underline{C} = \lim_{N \to \infty} \underline{C}_N, \quad (4.6.3)$$

где

$$\underline{C}_N = \frac{1}{N} \max_{\mathbf{Q}_N} \min_{s_0} I_{\mathbf{Q}}(\mathbf{X}^N; \mathbf{Y}^N \mid s_0), \quad (4.6.4)$$

$$I_{\mathbf{Q}}(\mathbf{X}^N, \mathbf{Y}^N \mid s_0) = \sum_{\mathbf{x}} \sum_{\mathbf{y}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x}, s_0) \log \frac{P_N(\mathbf{y} \mid \mathbf{x}, s_0)}{\sum_{\mathbf{x}'} Q_N(\mathbf{x}') P_N(\mathbf{y} \mid \mathbf{x}', s_0)}. \quad (4.6.5)$$

Подобно этому верхняя пропускная способность ККЧС $\overline{C}$ определяется следующим образом:

$$\overline{C} = \lim_{N\to\infty} \overline{C}_N, \tag{4.6.6}$$

где

$$\overline{C}_N = \frac{1}{N} \max_{\mathbf{Q}_N} \max_{s_0} I_{\mathbf{Q}}(\mathbf{X}^N; \mathbf{Y}^N | s_0). \tag{4.6.7}$$

Следующая теорема, доказательство которой проведено в приложении 4А, устанавливает существование этих пределов.

Теорема 4.6.1. В канале с конечным числом A состояний

$$\lim_{N\to\infty} \underline{C}_N = \sup_N \left[\underline{C}_N - \frac{\log A}{N} \right], \tag{4.6.8}$$

$$\lim_{N\to\infty} \overline{C}_N = \inf_N \left[\overline{C}_N + \frac{\log A}{N} \right]. \tag{4.6.9}$$

Из определений (4.6.4) и (4.6.7) очевидно следует, что

$$\underline{C}_N \leqslant \overline{C}_N \text{ для всех } N. \tag{4.6.10}$$

Отсюда прямым следствием теоремы является то, что для любого N

$$-\frac{\log A}{N} + \underline{C}_N \leqslant \underline{C} \leqslant \overline{C} \leqslant \overline{C}_N + \frac{\log A}{N}. \tag{4.6.11}$$

Это соотношение полезно при отыскании $\underline{C}$ и $\overline{C}$, в особенности в случае, когда $\underline{C} = \overline{C}$, так как оно дает верхнюю и нижнюю границы для пределов, которые стремятся друг к другу при возрастании N.

В качестве первого примера, который позволит понять смысл $\underline{C}$ и $\overline{C}$, рассмотрим рис. 4.6.3.

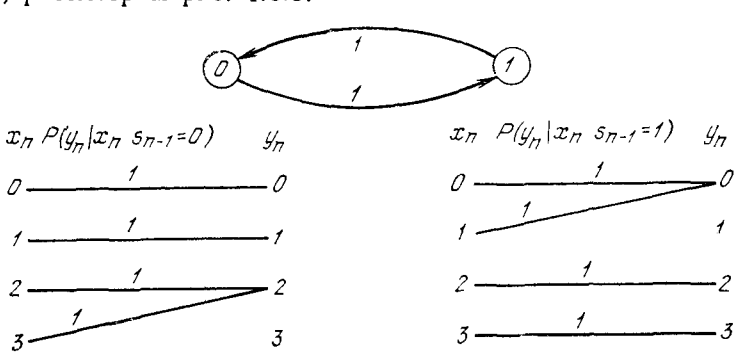


Рис. 4.6.3. ККЧС; пример канала, пропускная способность которого не определена.

Если начальным состоянием является $s_0 = 0$, то нетрудно заметить (рассматривая канал как пару параллельных каналов, как в за-

даче 4.1, с одним каналом для каждого состояния), что средняя взаимная информация максимизируется при выборе статистически независимых входов и при $Q(0) = Q(1) = Q(2) + Q(3) = {}^1/_3$ для первого и всех входов в нечетные моменты, и при $Q(0) + Q(1) = Q(2) = Q(3) = {}^1/_3$ для второго и всех входов в четные моменты. Для этих распределения на входе и начального состояния $I_{\mathbf{Q}}(\mathbf{X}^N; \mathbf{Y}^N | s_0) = N \log 3$. Точно так же, если $s_0 = 1$, то средняя взаимная информация максимизируется при перемене мест двух указанных одномерных распределений входных букв. Следовательно, $\overline{C}_N = \log 3$ при всех N. Можно заметить, однако, что при использовании соответствующего

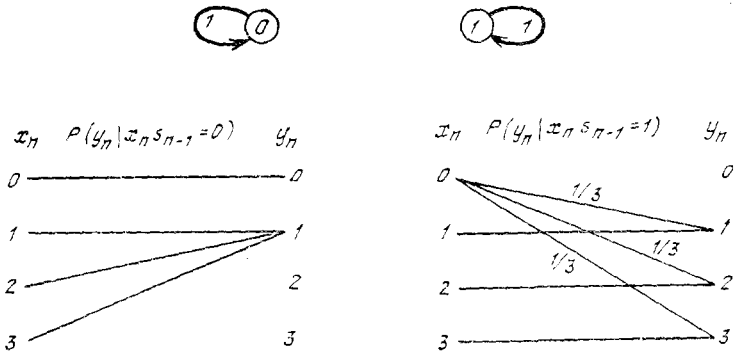


Рис. 4.6.4.

указанного выше распределения на входе, на передающем конце должно быть известным начальное состояние канала, и, таким образом, $\overline{C}$ в этом примере является максимальной средней взаимной информацией на букву, которую можно достичь, если передатчик имеет возможность выбора распределения на входе, согласованного с начальным состоянием.

Если начальное состояние не известно на передающем конце, то правильным будет так выбрать входное распределение, чтобы получить большую среднюю взаимную информацию для каждого возможного начального состояния. Нижняя пропускная способность $\underline{C}$ представляет собой наибольшую среднюю взаимную информацию на букву, которую можно достичь при фиксированном входном распределении, независимо от начального состояния. В этом примере можно показать, что $\underline{C} = \underline{C}_N = {}^3/_2$ бит при любом N и что $\underline{C}$ достигается на статистически независимых и равновероятных входах.

Для рассмотренного здесь примера нет основания для утверждения, что $\underline{C}$ или $\overline{C}$ является действительной пропускной способностью канала. Они просто применимы в близких физических ситуациях: одна к случаю, когда возможно провести некоторое измерение для определения фазы последовательности состояний, и другая — в случае, когда такое измерение невозможно.

На рис. 4.6.4 изображена задача другого типа. Канал остается все время либо в состоянии 0 либо в состоянии 1. Пропускная способность канала, соответствующая состоянию 0, равна 1 бит, а пропускная способность канала, соответствующая состоянию 1, равна $\log_2 3$ бит. При этом $\overline{C} = \log 3$ бит. С помощью несложных вычислений можно показать, что $\underline{C} \approx 0{,}965$ бит достигается на независимых входах, используемых с вероятностями $Q(0) \approx 0{,}391$, $Q(1) = Q(2) = Q(3) \approx \approx 0{,}203$. Отсюда видно, $\underline{C}$ меньше, чем пропускная способность каждого из отдельных каналов, что следует из того, что одно и то же входное распределение должно давать среднюю взаимную информацию на букву, не большую C для каждого состояния.

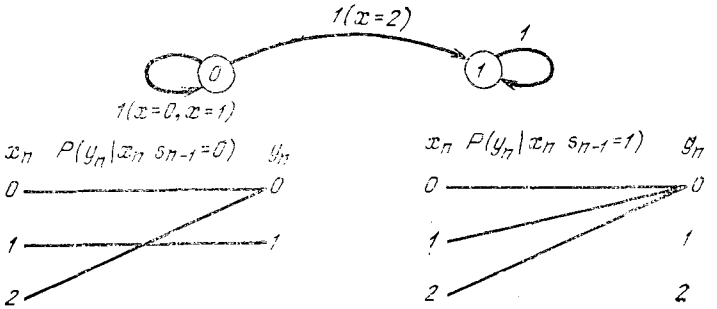


Рис. 4.6.5. «Панический» канал.

Наконец, на рис. 4.6.5 представлен «панический» канал. Входная буква 2 является панической буквой и ее использование выводит из строя канал во все будущие моменты времени. Очевидно, что $\overline{C} = 1$ бит и $\underline{C} = 0$ бит.

Предыдущие примеры были отчасти вырожденными в том смысле, что влияние начального состояния не уменьшалось с увеличением времени. Далее будет определен класс ККЧС, в котором влияние начального состояния убывает на нет со временем, и будет показано, что для этих каналов $\overline{C} = \underline{C}$. До того как приступить к этому, установим два обращения теоремы кодирования для ККЧС; одно, соответствующее $\overline{C}$, и другое — $\underline{C}$.

Так же как и в § 4.3, рассмотрим дискретный источник с алфавитом объема M, который производит буквы со скоростью одна буква за каждые τ_s секунд. Последовательность источника $\mathbf{u}$ после соответствующей обработки должна быть передана по ККЧС и воспроизведена как последовательность $\mathbf{v}$ для получателя у адресата. Пусть этот канал используется один раз каждые τ_c секунд; рассмотрим последовательности источника $\mathbf{u} = (u_1, \ldots, u_L)$ произвольной длины L и последовательности канала с длиной, которая равна наибольшему целому числу, меньшему $L\tau_s/\tau_c$:

$$N = \lfloor L\tau_s/\tau_c \rfloor. \qquad (4.6.12)$$

Будем считать, что канал находится в некотором начальном состоянии s_0 и что устройство обработки данных помещено между источником и каналом и между каналом и адресатом. Вероятности источника, переходные вероятности канала, начальное состояние и устройство обработки определяют теперь совместный ансамбль $\mathbf{U}^L$, $\mathbf{X}^N$, $\mathbf{Y}^N$, $\mathbf{V}^L$. Этот совместный ансамбль зависит от начального состояния s_0, которое в течение некоторого времени мы будем считать детерминированным. Предположим, как и в § 4.3, что последовательность источника *связана* с последовательностью, которая выдается адресату, с помощью N-кратного использования канала, в том смысле, что для всех возможных $\mathbf{u}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{v}$ имеем

$$P_N(\mathbf{y} \mid \mathbf{x}, s_0) = P_N(\mathbf{y} \mid \mathbf{x}, s_0, \mathbf{u}), \tag{4.6.13}$$

$$P(\mathbf{v} \mid \mathbf{y}, s_0) = P(\mathbf{v} \mid \mathbf{y}, s_0, \mathbf{x}, \mathbf{u}). \tag{4.6.14}$$

В этом случае применима теорема переработки информации, из которой следует

$$I(\mathbf{U}^L;\ \mathbf{V}^L \mid s_0) \leqslant I(\mathbf{X}^N;\ \mathbf{Y}^N \mid s_0). \tag{4.6.15}$$

Символ s_0 помещен в (4.6.15) просто для напоминания о том, что рассматриваемый совместный ансамбль зависит от данного начального состояния.

Далее пусть $\langle P_e(s_0)\rangle$ является средней вероятностью ошибки на букву источника, как это представлено равенством (4.3.2). Согласно теореме 4.3.2 имеем

$$\langle P_e(s_0)\rangle \log(M-1) + \mathscr{H}(\langle P_e(s_0)\rangle) \geqslant \frac{1}{L} H(\mathbf{U}^L \mid \mathbf{V}^L s_0) \geqslant \tag{4.6.16}$$

$$\geqslant \frac{1}{L}[H(\mathbf{U}^L \mid s_0) - I(\mathbf{U}^L;\ \mathbf{V}^L \mid s_0)] \geqslant \tag{4.6.17}$$

$$\geqslant \frac{1}{L}[H(\mathbf{U}^L \mid s_0) - I(\mathbf{X}^N;\ \mathbf{Y}^N \mid s_0)]. \tag{4.6.18}$$

Сделаем теперь дальнейшее предположение, состоящее в том, что вероятности источника не зависят от начального состояния канала, и применим теорему 3.5.1 для того, чтобы получить следующий результат:

$$\frac{1}{L} H(\mathbf{U}^L \mid s_0) \geqslant H_\infty(U) = \lim_{L\to\infty} \frac{1}{L} H(\mathbf{U}^L). \tag{4.6.19}$$

Используя (4.6.19) и определение N [см. (4.6.12)] совместно с (4.6.18), имеем

$$\langle P_e(s_0)\rangle \log(M-1) + \mathscr{H}(\langle P_e(s_0)\rangle) \geqslant H_\infty(U) - \frac{\tau_s}{\tau_c N} I(\mathbf{X}^N;\ \mathbf{Y}^N \mid s_0). \tag{4.6.20}$$

Теперь можно связать (4.6.20) с верхней и нижней пропускными способностями канала. В соответствии с определением (4.6.7) имеем

$$\frac{1}{N} I(\mathbf{X}^N;\ \mathbf{Y}^N \mid s_0) \leqslant \overline{C}_N \text{ для всех } s_0. \tag{4.6.21}$$

Подставляя (4.6.21) в (4.6.20) и переходя к пределу при $L\to\infty$, $N\to\infty$, получаем для всех s_0:

$$\langle P_e(s_0)\rangle \log(M-1)+\mathcal{H}(\langle P_e(s_0)\rangle)\geqslant H_\infty(U)-\frac{\tau_s}{\tau_c}\overline{C}. \quad (4.6.22)$$

Важно отметить, что при выводе (4.6.22) не было сделано предположения о независимости $\mathbf{X}^N$ и s_0. Физически это означает, что (4.6.22) остается справедливым даже, если устройство обработки входных данных знает начальное состояние канала и использует это при отображении последовательности источника в последовательности на входе канала. Точно так же неравенство (4.6.22) остается справедливым вне зависимости от того, использует или нет устройство обработки данных на выходе начальное состояние при отображении последовательностей $\mathbf{y}$ в последовательности $\mathbf{v}$.

Для того чтобы связать (4.6.20) с нижней пропускной способностью канала, нужно сделать дополнительное предположение в том, что $\mathbf{X}^N$ не зависит от начального состояния s_0. Теперь, используя определение $\underline{C}_N$, получаем, что для некоторого начального состояния s_0:

$$\frac{1}{N}I(\mathbf{X}^N;\ \mathbf{Y}^N\,|\,s_0)\leqslant \underline{C}_N. \quad (4.6.23)$$

Так как согласно теореме 4.6.1 $\underline{C}_N\leqslant \underline{C}+(\log A)/N$, то получаем, что для любого N и некоторого s_0

$$\langle P_e(s_0)\rangle \log(M-1)+\mathcal{H}(\langle P_e(s_0)\rangle)\geqslant H_\infty(U)-\frac{\tau_s}{\tau_c}\left[\underline{C}+\frac{\log A}{N}\right]. \quad (4.6.24)$$

Если имеется распределение вероятностей $q_0(s_0)$ на начальных состояниях, то $\langle P_e(s_0)\rangle$ можно усреднить и получить общую вероятность ошибки на букву источника

$$\langle P_e\rangle=\sum_{s_0} q_0(s_0)\langle P_e(s_0)\rangle.$$

Если q_{min} является наименьшей из этих вероятностей начальных состояний, то $\langle P_e\rangle\geqslant q_{min}\langle P_e(s_0)\rangle$ для каждого начального состояния, и для (4.6.24) можно получить следующую границу:

$$\frac{\langle P_e\rangle}{q_{min}}\log(M-1)+\mathcal{H}\left(\frac{\langle P_e\rangle}{q_{min}}\right)\geqslant H_\infty(U)-\frac{\tau_s}{\tau_c}\left[\underline{C}+\frac{\log A}{N}\right]. \quad (4.6.25)$$

Соотношения (4.6.24) и (4.6.25) не обязательно справедливы в случае, если устройство обработки данных на входе может использовать начальное состояние, однако они остаются в силе независимо от того, использует или нет устройство обработки данных на выходе какие-либо сведения об этом состоянии. Эти результаты можно подытожить следующей теоремой.

Теорема 4.6.2. Пусть дискретный стационарный источник с алфавитом объема M производит одну букву каждые τ_s сек и имеет предельную энтропию на букву $H_\infty(U)$. Пусть канал с конечным числом состояний с верхней пропускной способностью $\overline{C}$ и нижней пропускной

способностью $\underline{C}$ используется один раз каждые τ_c с. Если в пределе при $L \to \infty$ последовательность источника из L букв связана с последовательностью адресата с помощью $N = \lfloor L\tau_s/\tau_c \rfloor$-кратного использования канала [т. е. имеют место равенства (4.6.13) и (4.6.14)], то независимо от начального состояния канала вероятность ошибки на букву источника удовлетворяет (4.6.22). Если в дополнение ансамбль на входе канала $\mathbf{X}^N$ не зависит от начального состояния, то для каждого значения N существует некоторое начальное состояние, для которого справедливо (4.6.24). Если, кроме того, имеется распределение вероятностей начального состояния с наименьшей вероятностью q_{min}, то справедливо (4.6.25).

Неразложимые каналы

Определим теперь обширный класс каналов, известных как неразложимые ККЧС, для которых $C = \overline{C}$. Грубо говоря, неразложимым ККЧС называется такой ККЧС, для которого влияние начального состояния исчезает со временем. Точнее, пусть

$$q_N(s_N \mid \mathbf{x}, s_0) = \sum_{\mathbf{y}} P_N(\mathbf{y}, s_N \mid \mathbf{x}, s_0).$$

ККЧС является неразложимым, если для любого сколь угодно малого $\varepsilon > 0$ существует такое N_0, что для всех $N \geqslant N_0$ справедливо

$$|q_N(s_N \mid \mathbf{x}, s_0) - q_N(s_N \mid \mathbf{x}, s_0')| \leqslant \varepsilon \qquad (4.6.26)$$

при любых s_N, $\mathbf{x}$, s_0 и s_0'.

Читателю предлагается проверить, что каналы, представленные на рис. 4.6.3 и 4.6.5, не являются неразложимыми. Можно заметить, что канал на рис. 4.6.2 является неразложимым и, в действительности, левая часть (4.6.26) равна нулю для всех $N \geqslant 1$. Ниже будет показано, что канал на рис. 4.6.1 также является неразложимым.

При фиксированной последовательности на входе $(x_1, x_2, \ldots)$ можно рассматривать последовательность состояний $(s_0, s_1, \ldots)$ как неоднородную цепь Маркова. Чтобы избежать нагромождения обозначений, в последующем изложении мы опустим зависимость от входной последовательности и, например, будем использовать обозначение $q(s_N \mid s_0)$ вместо $q_N(s_N \mid \mathbf{x}, s_0)$*). Будем интересоваться зависимостью $q(s_N \mid s_0)$ от s_0 при больших N. В качестве меры этой зависимости определим расстояние $d_N(s_0', s_0'')$ следующим образом:

$$d_N(s_0', s_0'') = \sum_{s_N} |q(s_N \mid s_0') - q(s_N \mid s_0'')|. \qquad (4.6.27)$$

При $N = 0$ и $s_0' \neq s_0''$ будем считать, что $d_N(s_0', s_0'')$ равно 2.

---

*) Если быть точным, то эту вероятность следует обозначить: $q_{s_N \mid x_1, \ldots, x_N s_0}(i \mid k_1, \ldots, k_N, j)$, что означает вероятность того, что s_N принимает значение i при условии, что заданы последовательность $\mathbf{x} = (k_1, \ldots, k_N)$ и состояние $s_0 = j$.

Как показывает следующая лемма, зависимость s_N от s_0 в смысле этой вероятностной меры не увеличивается с увеличением N.

Лемма 4.6.1. Для любых заданных входов $x_1, x_2, \ldots$, любых заданных s_0', s_0'' определенное выше расстояние $d_N(s_0', s_0'')$ не увеличивается с ростом N.

Доказательство. При $N \geqslant 1$ имеем

$$d_N(s_0', s_0'') = \sum_{s_N} |q(s_N | s_0') - q(s_N | s_0'')| = \tag{4.6.28}$$

$$= \sum_{s_N} \left| \sum_{s_{N-1}} q(s_N | s_{N-1}) [q(s_{N-1} | s_0') - q(s_{N-1} | s_0'')] \right|. \tag{4.6.29}$$

Ограничивая сверху модуль суммы суммой модулей, получаем

$$d_N(s_0', s_0'') \leqslant \sum_{s_N} \sum_{s_{N-1}} q(s_N | s_{N-1}) |q(s_{N-1} | s_0') - q(s_{N-1} | s_0'')| = \tag{4.6.30}$$

$$= \sum_{s_{N-1}} |q(s_{N-1} | s_0') - q(s_{N-1} | s_0'')| = \tag{4.6.31}$$

$$= d_{N-1}(s_0', s_0''). \tag{4.6.32}$$

Следующая лемма дает условие, при котором $d_N(s_0', s_0'')$ стремится к 0 с ростом N.

Лемма 4.6.2. Предположим, что при некотором $n > 0$, некотором $\delta > 0$ и каждом $N \geqslant 0$ существует некоторое s_{N+n}, для которого

$$q(s_{N+n} | s_N) \geqslant \delta \text{ для всех значений } s_N. \tag{4.6.33}$$

Тогда $d_N(s_0', s_0'')$ стремится к 0 экспоненциально с ростом N и

$$d_N(s_0', s_0'') \leqslant 2(1-\delta)^{(N/n)-1}.$$

Доказательство. Имеем

$$d_{N+n}(s_0', s_0'') = \sum_{s_{N+n}} |q(s_{N+n} | s_0') - q(s_{N+n} | s_0'')| = \tag{4.6.34}$$

$$= \sum_{s_{N+n}} \left| \sum_{s_N} q(s_{N+n} | s_N) [q(s_N | s_0') - q(s_N | s_0'')] \right|. \tag{4.6.35}$$

Введем обозначение:

$$a(s_{N+n}) = \min_{s_N} q(s_{N+n} | s_N). \tag{4.6.36}$$

Замечая, что

$$\sum_{s_N} a(s_{N+n}) [q(s_N | s_0') - q(s_N | s_0'')] = 0,$$

представим (4.6.35) в виде

$$d_{N+n}(s_0', s_0'') = \sum_{s_{N+n}} \left| \sum_{s_N} [q(s_{N+n} | s_N) - a(s_{N+n})] [q(s_N | s_0') - q(s_N | s_0'')] \right|. \tag{4.6.37}$$

Ограничивая модуль суммы с помощью суммы модулей и замечая, что $q(s_{N+n}|s_N) - a(s_{N+n}) \geqslant 0$, будем иметь

$$d_{N+n}(s_0', s_0'') \leqslant \sum_{s_{N+n}} \sum_{s_N} [q(s_{N+n}|s_N) - a(s_{N+n})] \, |q(s_N|s_0') - q(s_N|s_0'')|. \tag{4.6.38}$$

После суммирования по s_{N+n} получаем

$$d_{N+n}(s_0', s_0'') \leqslant [1 - \sum_{s_{N+n}} a(s_{N+n})] \sum_{s_N} |q(s_N|s_0') - q(s_N|s_0'')| =$$
$$= [1 - \sum_{s_{N+n}} a(s_{N+n})] \, d_N(s_0', s_0''). \tag{4.6.39}$$

По предположению, $a(s_{N+n}) \geqslant \delta$ по крайней мере для одного значения s_{N+n} и, следовательно,

$$d_{N+n}(s_0', s_0'') \leqslant (1-\delta) \, d_N(s_0', s_0''). \tag{4.6.40}$$

Используя последнее неравенство при $N = 0$, затем при $N = n$, затем при $N = 2n$ и т. д. и вспоминая, что $d_0(s_0', s_0'') = 2$, получаем

$$d_{mn}(s_0', s_0'') \leqslant 2(1-\delta)^m. \tag{4.6.41}$$

В силу того, что $d_N(s_0', s_0'')$ не увеличивается с ростом N, это доказывает лемму. |

Следующая ниже теорема дает критерий того, является ли ККЧС неразложимым или нет.

Теорема 4.6.3. Необходимым и достаточным условием того, что ККЧС будет неразложимым, является существование при некотором фиксированном n для каждого $\mathbf{x}$ такого состояния s_n, что

$$q(s_n|\mathbf{x}, s_0) > 0 \text{ при всех } s_0 \tag{4.6.42}$$

(это s_n может зависеть от $\mathbf{x}$). Более того, если канал является неразложимым, то указанное выше n можно всегда выбрать меньшим 2^{A^2}, где A — число состояний канала.

Доказательство. Достаточность. Если (4.6.42) справедливо для некоторого n, то, так как s_0 и $\mathbf{x} = (x_1, \ldots, x_n)$ могут принимать только конечное число значений, существует некоторое $\delta > 0$, такое, что

$$q(s_n|\mathbf{x}, s_0) \geqslant \delta \tag{4.6.43}$$

при всех s_0, всех $\mathbf{x}$ и некотором s_n, зависящем от $\mathbf{x}$.

Также в силу того, что вероятности в канале не зависят от времени, имеем для любых N и некоторого s_{N+n}, зависящего от $x_{N+1}, \ldots, x_{N+n}$

$$q(s_{N+n}|x_{N+1}, \ldots, x_{N+n}, s_N) > \delta \text{ при всех } s_N. \tag{4.6.44}$$

Таким образом, условия предыдущей леммы удовлетворяются и

$$\sum_{s_N} |q(s_N|\mathbf{x}, s_0') - q(s_N|\mathbf{x}, s_0'')|$$

стремится к нулю экспоненциально при $N \to \infty$, равномерно по $\mathbf{x}$, s_0' и s_0''. Следовательно, (4.6.26) справедливо для достаточно больших N, и канал является неразложимым.

Необходимость. Выберем $\varepsilon < 1/A$, где A — число состояний, выберем N достаточно большим, чтобы удовлетворялось (4.6.26), и для заданного s_0 и $\mathbf{x}$ выберем s_N так, чтобы $q(s_N | \mathbf{x}, s_0) \geqslant 1/A$. Тогда согласно (4.6.26) $q(s_N | \mathbf{x}, s_0') > 0$ для всех s_0' и условие теоремы удовлетворяется при n, равном этому N.

Доказательство того, что $n < 2^{A^2}$. Для заданных n, $\mathbf{x}$ определим матрицу связности

$$T_{n,\mathbf{x}}(s_0, s_n) = \begin{cases} 1, & q(s_n | \mathbf{x}, s_0) > 0, \\ 0, & q(s_n | \mathbf{x}, s_0) = 0. \end{cases} \tag{4.6.45}$$

Эта матрица размера A на A из нулей и единиц, в которой индексом строки является s_0, а индексом столбца является s_n. Данный элемент равен 1, если соответствующее ему s_n может быть достигнуто из соответствующего ему s_0 при заданном $\mathbf{x}$. Так как $q(s_n | \mathbf{x}, s_0) = \sum q(s_n | x_n, s_{n-1}) q(s_{n-1} | \mathbf{x}_{n-1}, s_0)$, то можно выразить $T_{n,\mathbf{x}}(s_0, s_n)$ с помощью $T_{n-1,\, x_{n-1}}$ в виде

$$T_{n,\mathbf{x}}(s_0, s_n) = \begin{cases} 1, & T_{n-1,\, x_{n-1}}(s_0, s_{n-1}) q(s_n | x_n, s_{n-1}) > 0 \text{ для} \\ & \text{некоторого } s_{n-1}, \\ 0 & \text{во всех остальных случаях.} \end{cases} \tag{4.6.46}$$

В силу того, что существует лишь $2^{A^2} - 1$ ненулевых матриц размера A на A из двоичных элементов, то последовательность матриц*) $T_{n,\mathbf{x}}(s_0, s_n)$, $n = 1, \ldots, 2^{A^2}$, должна содержать две одинаковые матрицы, допустим с $i < j \leqslant 2^{A^2}$. Если исходя из последовательности символов $(x_1, \ldots, x_j)$ выбрать $x_{j+N} = x_{i+N}$ при всех $N \geqslant 1$, то из (4.6.46) следует, что $T_{j+N,\mathbf{x}} = T_{i+N,\mathbf{x}}$ при всех $N \geqslant 1$. Если при этом выборе $\mathbf{x}$, $T_{n,\mathbf{x}}$ не имеет столбцов из единиц при $n \leqslant j$, то она не будет иметь столбцов из единиц при всех больших n. Но это означает, что при этих $\mathbf{x}$, не существуют n, s_n, для которых $q(s_n | \mathbf{x}, s_0) > 0$ при всех s_0 и, таким образом, канал не является неразложимым. Поэтому для того чтобы канал был неразложимым, при любом $\mathbf{x}$ матрица $T_{n,\mathbf{x}}$ должна иметь единичный столбец при некотором $n \leqslant 2^{A^2}$. Наконец, если $T_{n,\mathbf{x}}$ имеет единичный столбец при некотором n, то из (4.6.46) следует, что эта матрица имеет столбец из единиц при всех больших n, и, следовательно, для неразложимого канала имеется некоторое наименьшее $n \leqslant 2^{A^2}$, для которого $T_{n,\mathbf{x}}$ имеет столбец из единиц для всех $\mathbf{x}$. |

Для канала, изображенного на рис. 4.6.1, условие (4.6.42) удовлетворяется при $n = 1$, и, следовательно, этот канал является неразложимым.

Теорема 4.6.4. Для неразложимого ККЧС

$$\underline{C} = \overline{C}. \tag{4.6.47}$$

*) Здесь $\mathbf{x} = (x_1, \ldots, x_n)$ — отрезок некоторой последовательности $(x_1, \ldots, x_n, \ldots, x_{A^2})$. (*Прим. ред.*)

Доказательство. При произвольном N пусть $Q_N(\mathbf{x})$ и s_0' являются распределением на входе и начальным состоянием, которые максимизируют $I_{\mathbf{Q}}(\mathbf{X}^N; \mathbf{Y}^N | s_0)$, и пусть s_0'' обозначает начальное состояние, которое минимизирует $I_{\mathbf{Q}}$ при том же самом распределении на входе. Таким образом, по определению $\overline{C}_N$ и $\underline{C}_N$ имеем

$$\overline{C}_N = \frac{1}{N} I_{\mathbf{Q}}(\mathbf{X}^N; \mathbf{Y}^N | s_0'), \tag{4.6.48}$$

$$\underline{C}_N \geqslant \frac{1}{N} I_{\mathbf{Q}}(\mathbf{X}^N; \mathbf{Y}^N | s_0''). \tag{4.6.49}$$

Положим теперь $n + l = N$, где n и l — положительные целые числа. Пусть $\mathbf{X}_1$ обозначает ансамбль входных последовательностей $\mathbf{x}_1 = (x_1, \ldots, x_n)$, а $\mathbf{X}_2$ обозначает ансамбль последовательностей $\mathbf{x}_2 = (x_{n+1}, \ldots, x_N)$, соответствующих распределению на входе $Q_N(\mathbf{x})$. Аналогично пусть $\mathbf{Y}_1$ и $\mathbf{Y}_2$ обозначают ансамбли выходов $\mathbf{y}_1 = (y_1, \ldots, y_n)$ и $\mathbf{y}_2 = (y_{n+1}, \ldots, y_N)$ при условии, что задано s_0'. Имеем теперь

$$\overline{C}_N = \frac{1}{N}[I(\mathbf{X}_1; \mathbf{Y}_1\mathbf{Y}_2 | s_0') + I(\mathbf{X}_2; \mathbf{Y}_1 | \mathbf{X}_1, s_0') + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1\mathbf{Y}_1, s_0')]. \tag{4.6.50}$$

Вероятность элемента из произведения указанных выше ансамблей, включая и состояние в момент n, может быть выражена в виде

$$Q(\mathbf{x}_1) Q(\mathbf{x}_2 | \mathbf{x}_1) P_n(\mathbf{y}_1, s_n | \mathbf{x}_1, s_0') P_l(\mathbf{y}_2 | \mathbf{x}_2, s_n). \tag{4.6.51}$$

Отсюда можно заметить, что второе слагаемое в правой части (4.6.50) равно нулю. Также первое слагаемое ограничено сверху числом $n \log K$, так как в ансамбле $\mathbf{X}_1$ содержится K^n элементов. Наконец, из леммы 4А.1 следует, что последнее слагаемое не больше, чем $\log A + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1\mathbf{Y}_1 S_n, s_0')$. Таким образом,

$$\overline{C}_N \leqslant \frac{1}{N}[n \log K + \log A + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1\mathbf{Y}_1 S_n, s_0')]. \tag{4.6.52}$$

Оценивая подобным же образом $\underline{C}_N$ снизу, используя s_0 вместо s_0' и ограничивая снизу первое слагаемое в (4.6.50) нулем, получаем

$$\underline{C}_N \geqslant \frac{1}{N}[-\log A + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1\mathbf{Y}_1 S_n, s_0'')]. \tag{4.6.53}$$

Используя (4.6.51), заметим, что среди условий, при которых рассматривается информация, условие на $\mathbf{Y}_1$ может быть опущено в (4.6.52) и (4.6.53), будем иметь

$$\overline{C}_N - \underline{C}_N \leqslant \frac{1}{N}[n \log K + 2 \log A + I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1 S_n, s_0') - I(\mathbf{X}_2; \mathbf{Y}_2 | \mathbf{X}_1 S_n, s_0'')] = \frac{1}{N}[n \log K + 2 \log A + \tag{4.6.54}$$

$$+ \sum_{\mathbf{x}_1} Q(\mathbf{x}_1) \sum_{s_N} [q(s_n | \mathbf{x}_1, s_0') - q(s_n | \mathbf{x}_1, s_0'')] I(\mathbf{X}_2; \mathbf{Y}_2 | s_n, \mathbf{x}_1)]. \tag{4.6.55}$$

Далее можно оценить сверху это выражение, взяв модуль каждого слагаемого в сумме и оценив сверху значение I для каждого s_n и $\mathbf{x}_1$ с помощью $l \log K$. В результате получим:

$$\overline{C}_N - \underline{C}_N \leqslant \frac{1}{N} [n \log K + 2 \log A + \overline{d}_n (s_0', s_0'') \, l \log K],$$

где $\overline{d}_n$ является верхней границей для $d_n (s_0', s_0'')$, справедливой при всех $\mathbf{x}_1$. Согласно определению неразложимого канала при любом $\varepsilon > 0$ можно выбрать n так, чтобы $\overline{d}_n < \varepsilon$. Для этого фиксированного значения n

$$\lim_{N \to \infty} \overline{C}_N - \underline{C}_N \leqslant \varepsilon \log K. \tag{4.6.56}$$

В силу произвольности $\varepsilon > 0$ и того, что $\overline{C}_N \geqslant \underline{C}_N$, это завершает доказательство теоремы. |

В то время как условие неразложимости ККЧС достаточно для того, чтобы $\underline{C} = \overline{C}$, для многих разложимых ККЧС оказывается также $\underline{C} = \overline{C}$. Каналы, имеющие память, связанную только с межсимвольной интерференцией, особенно легко исследовать в этом отношении. Предположим, что такой канал является полностью связанным, т. е., что любое состояние может быть достигнуто из любого другого состояния с помощью некоторой конечной последовательности входов. Предположим далее, что имеется некоторая последовательность входов, которая приводит канал в некоторое известное состояние ($q(s_n | \mathbf{x}, s_0) = 1$ при всех s_0 и заданном $\mathbf{x}$). Тогда существует также конечная входная последовательность, которая приводит канал в какое-либо желаемое состояние и с этого места можно достичь $\overline{C}$ сколь угодно точно, так что $\overline{C} = \underline{C}$. Можно показать (см. задачу 4.26), что если канал вообще может быть переведен в известное состояние, то он может быть переведен в такое состояние с помощью самое большее 2^A входов.

ИТОГИ И ВЫВОДЫ

В этой главе была исследована вероятностная модель канала связи. Для моделей дискретного канала без памяти и моделей канала с конечным числом состояний была определена пропускная способность как максимум средней взаимной информации. Основной результат главы составляет обращение теоремы кодирования, которое утверждает, что если скорость источника (т. е. энтропия источника в битах на единицу времени) превосходит пропускную способность канала (в битах на единицу времени), то надежная передача по каналу невозможна. Было сделано некоторое введение в теорию выпуклых функций, которые являются полезным инструментом при изучении всей теории информации, и было показано, как применить эту теорию при отыскании пропускной способности дискретного канала без памяти. Для канала с конечным числом состояний были даны два имеющих смысл определения пропускной способности и показано, что они совпадают друг с другом для неразложимых каналов.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Общие идеи этой главы и их развитие принадлежат Шеннону (1948). Обращение теоремы кодирования (теорема 4.3.4) получено Галлагером (1964) и основано на теореме 4.3.1, полученной Фано (1952). Рейффен (1966) отметил, что теорема 4.3.4 может быть применена не только к источникам без памяти, но и к источникам с памятью. Теорема 4.4.1 является, по существу, частным случаем общего результата Куна и Тюкера (1951), относящегося к выпуклому программированию. Его применение к вычислению пропускной способности канала было независимо предложено Эйзенбергом (1963) и Галлагером (1962), однако необходимость условия (4.5.1) в теореме 4.5.1 была доказана Шенноном (1948).

Первая часть теоремы 4.6.1 и вторая часть теоремы 4.6.2 принадлежат Юдкину (1967), хотя Блекуэлл, Брейман и Томасян (1958) установили ранее слабое обращение теоремы кодирования для неразложимых ККЧС. (Неразложимые каналы Блекуэлла, Бреймана и Томасяна образуют тот же класс каналов, что и неразложимые каналы, рассмотренные здесь. Читатель может проверить это, если после прочтения статьи Блекуэлла, Бреймана и Томасяна он заметит, что, если цепь Маркова имеет периодическое множество состояний с периодом m, то m-я степень матрицы соответствует разложимой цепи с по крайней мере m замкнутыми множествами состояний.) Последняя часть теоремы 4.6.3 была доказана Томасяном (1963).

ПРИЛОЖЕНИЕ 4А

Начнем с двух лемм.

Лемма 1. Пусть $XYZS$ является совместным ансамблем и пусть S содержит A точек. Тогда

$$|I(X; Y|ZS) - I(X; Y|Z)| \leqslant \log A. \tag{4A.1}$$

Доказательство. Представим $I(XS; Y|Z)$ следующими способами:

$$I(XS; Y|Z) = I(X; Y|Z) + I(S; Y|ZX) = \tag{4A.2}$$

$$= I(X; Y|ZS) + I(S; Y|Z). \tag{4A.3}$$

Последнее слагаемое в (4А.2) и последнее слагаемое в (4А.3) неотрицательны и ограничены сверху величиной $H(S) \leqslant \log A$. Поэтому, приравнивая правые части (4А.2) и (4А.3), получаем (4А.1). |

Лемма 2. Пусть a_N, $N = 1, 2, \ldots$, ограниченная последовательность чисел и пусть

$$\overline{a} = \sup_N a_N \text{ и } \underline{a} = \inf_N a_N.$$

(Под ограниченной последовательностью мы понимаем такую последовательность, для которой $\overline{a} < \infty$ и $\underline{a} > -\infty$). Пусть при всех $n \geqslant 1$ и всех $N > n$

$$a_N \geqslant \frac{n}{N} a_n + \frac{N-n}{N} a_{N-n}. \tag{4A.4}$$

Тогда

$$\lim_{N\to\infty} a_N = \bar{a}. \tag{4A.5}$$

Обратно, если при всех $n \geqslant 1$ и $N > n$

$$a_N \leqslant \frac{n}{N} a_n + \frac{N-n}{N} a_{N-n}, \tag{4A.6}$$

то имеем

$$\lim_{N\to\infty} a_N = \underline{a}. \tag{4A.7}$$

Доказательство. Предположим, что справедливо (4A.4) и для любого заданного $\varepsilon > 0$ выберем n так, чтобы

$$a_n \geqslant \bar{a} - \varepsilon. \tag{4A.8}$$

Полагая $N = 2n$, из (4A.4) получаем, что

$$a_{2n} \geqslant \frac{a_n}{2} + \frac{a_n}{2} \geqslant \bar{a} - \varepsilon. \tag{4A.9}$$

Подобно этому, полагая $N = mn$ для любого целого числа $m \geqslant 2$, имеем

$$a_{mn} \geqslant \frac{a_n}{m} + \frac{(m-1)\, a_{(m-1)n}}{m}. \tag{4A.10}$$

Используя индукцию, предположим, что $a_{(m-1)n} \geqslant \bar{a} - \varepsilon$, тогда (4A.10) означает, что $a_{mn} \geqslant \bar{a} - \varepsilon$. В силу справедливости предположения индукции при $m = 2, 3$ получаем

$$a_{mn} \geqslant \bar{a} - \varepsilon \quad \text{при всех} \quad m \geqslant 1. \tag{4A.11}$$

Далее при любом $N > n$ величину N можно представить в виде $mn + j$, где $0 \leqslant j \leqslant n - 1$. Подставляя j вместо n в (4A.4), получаем

$$a_N \geqslant \frac{j}{N} a_j + \frac{N-j}{N} a_{mn} = a_{mn} + (j/N)(a_j - a_{mn}) \geqslant$$
$$\geqslant \bar{a} - \varepsilon + (n/N)(\underline{a} - \bar{a}). \tag{4A.12}$$

Отсюда следует, что для всех достаточно больших N справедливо, что $a_N \geqslant \bar{a} - 2\varepsilon$. Равенство (4A.5) следует из того, что $a_N \leqslant \bar{a}$, и того, что ε выбрано произвольным. Равенство (4A.7) получается из (4A.6), если заметить, что (4A.6) означает, что (4A.4) применимо к последовательности $-a_n$ и таким образом

$$\lim -a_n = \sup -a_n = -\inf a_n. \tag{4A.13}$$

Доказательство теоремы 4.6.1 начнем с вывода соотношения

$$\lim_{N\to\infty} \underline{C}_N = \sup_N \left(\underline{C}_N - \frac{\log A}{N} \right).$$

Пусть при произвольных положительных целых числах n и l $\mathbf{Q}_n$ и $\mathbf{Q}_l$ являются распределениями на входе, на которых достигаются $\underline{C}_n$ и $\underline{C}_l$ соответственно. Пусть $N = n + l$ и выберем $\mathbf{Q}_N$ в виде

$$Q_N(\mathbf{x}) = Q_n(\mathbf{x}_1)\, Q_l(\mathbf{x}_2), \tag{4A.14}$$

где $\mathbf{x} = (x_1, \ldots, x_N)$, $\mathbf{x}_1 = (x_1, \ldots, x_n)$ и $\mathbf{x}_2 = (x_{n+1}, \ldots, x_N)$.

Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ являются ансамблями последовательностей $\mathbf{x}_1$ и $\mathbf{x}_2$ и пусть $\mathbf{Y}_1$ и $\mathbf{Y}_2$ будут соответствующими выходными ансамблями. Так как $Q_N(\mathbf{x})$ не обязательно является входным распределением, на котором достигается пропускная способность $\underline{C}_N$, то имеем

$$N\underline{C}_N \geqslant \min_{s_0} I(\mathbf{X}_1\mathbf{X}_2;\ \mathbf{Y}_1\mathbf{Y}_2 \mid s_0) = \min_{s_0} [I(\mathbf{X}_1;\ \mathbf{Y}_1\mathbf{Y}_2 \mid s_0) + I(\mathbf{X}_2;\ \mathbf{Y}_1\mathbf{Y}_2 \mid \mathbf{X}_1, s_0)] \quad (4\text{A}.15)$$

Первое слагаемое в правой части (4А.15) ограничено снизу следующим образом:

$$I(\mathbf{X}_1;\ \mathbf{Y}_1\mathbf{Y}_2 \mid s_0) \geqslant I(\mathbf{X}_1;\ \mathbf{Y}_1 \mid s_0) \geqslant n\underline{C}_n. \quad (4\text{A}.16)$$

Последнее слагаемое в (4А.15) может быть преобразовано следующим образом:

$$I(\mathbf{X}_2;\ \mathbf{Y}_1\mathbf{Y}_2 \mid \mathbf{X}_1,\ s_0) = I(\mathbf{X}_2;\ \mathbf{Y}_1\mathbf{Y}_2\mathbf{X}_1 \mid s_0) - I(\mathbf{X}_2;\ \mathbf{X}_1) \geqslant I(\mathbf{X}_2;\ \mathbf{Y}_2 \mid s_0), \quad (4\text{A}.17)$$

где использовано то, что $\mathbf{X}_1$ и $\mathbf{X}_2$ статистически независимы [см. (4А.14)]. Из леммы 1 следует, что $I(\mathbf{X}_2;\ \mathbf{Y}_2|s_0)$ ограничена снизу значением $I(\mathbf{X}_2;\ \mathbf{Y}_2|S_n, s_0) - \log A$. Наконец,

$$I(\mathbf{X}_2;\ \mathbf{Y}_2 \mid S_n, s_0) = \sum_{s_n} q(s_n \mid s_0)\, I(\mathbf{X}_2;\ \mathbf{Y}_2 \mid s_n) \geqslant$$
$$\geqslant \min_{s_n} I(\mathbf{X}_2;\ \mathbf{Y}_2 \mid s_n) = l\underline{C}_l. \quad (4\text{A}.18)$$

Используя (4А.16)—(4А.18) совместно с (4А.15) и замечая, что граница не зависит от s_0, получаем

$$N\underline{C}_N \geqslant n\underline{C}_n + l\underline{C}_l - \log A$$

или

$$N\left[\underline{C}_N - \frac{\log A}{N}\right] \geqslant n\left[\underline{C}_n - \frac{\log A}{n}\right] + l\left[\underline{C}_l - \frac{\log A}{l}\right]. \quad (4\text{A}.19)$$

Таким образом, последовательность $\underline{C}_N - (\log A)/N$ удовлетворяет (4А.4), и по лемме 2 имеем

$$\lim_{N\to\infty} \underline{C}_N = \lim_{N\to\infty}\left[\underline{C}_N - \frac{\log A}{N}\right] = \sup_N \left[\underline{C}_N - \frac{\log A}{N}\right]. \quad (4\text{A}.20)$$

Докажем, что $\lim\limits_{N\to\infty} \overline{C}_N = \inf\limits_N [\overline{C}_N + (\log A)/N]$.

Пусть $N = n + l$ при произвольных положительных целых числах n и l и пусть $\mathbf{Q}_N$ и s_0—входное распределение и начальное состояние, на которых достигается $\overline{C}_N$. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ — получающиеся в результате ансамбли входных последовательностей $\mathbf{x}_1 = (x_1, \ldots, x_n)$ и $\mathbf{x}_2 = (x_{n+1}, \ldots, x_N)$ и пусть $\mathbf{Y}_1$, $\mathbf{Y}_2$—соответствующие выходные ансамбли. Тогда

$$N\overline{C}_N = I(\mathbf{X}_1\,\mathbf{X}_2;\ \mathbf{Y}_1\mathbf{Y}_2 \mid s_0) = \quad (4\text{A}.21)$$
$$= I(\mathbf{X}_1;\ \mathbf{Y}_1 \mid s_0) + I(\mathbf{X}_2;\ \mathbf{Y}_1 \mid \mathbf{X}_1, s_0) + I(\mathbf{X}_1\mathbf{X}_2;\ \mathbf{Y}_2 \mid \mathbf{Y}_1, s_0). \quad (4\text{A}.22)$$

Распределение вероятностей на этом ансамбле (включающем s_n-состояние в момент n) может быть представлено в виде

$$Q_n(\mathbf{x}_1)\, Q_l(\mathbf{x}_2 \mid \mathbf{x}_1)\, P_n(\mathbf{y}_1, s_n \mid \mathbf{x}_1, s_0)\, P_l(\mathbf{y}_2 \mid \mathbf{x}_2 s_n). \quad (4\text{A}.23)$$

Отсюда можно заметить, что x_2 и y_1 являются статистически независимыми при условии, что заданы x_1 и s_0. Следовательно, для второго слагаемого в (4А.22) получаем

$$I(\mathbf{X}_2;\ \mathbf{Y}_1 \mid \mathbf{X}_1, s_0) = 0. \tag{4А.24}$$

Что касается первого слагаемого в (4А. 22), то оно удовлетворяет неравенству:

$$I(\mathbf{X}_1;\ \mathbf{Y}_1 \mid s_0) \leqslant n\overline{C}_n. \tag{4А.25}$$

Последнее слагаемое в (4А.22) можно оценить сверху, используя лемму 1:

$$\begin{aligned} I(\mathbf{X}_1\mathbf{X}_2;\ \mathbf{Y}_2 \mid \mathbf{Y}_1,\ s_0) &\leqslant I(\mathbf{X}_1\mathbf{X}_2;\ \mathbf{Y}_2 \mid \mathbf{Y}_1 S_n,\ s_0) + \log A = \\ &= H(\mathbf{Y}_2 \mid \mathbf{Y}_1 S_n,\ s_0) - H(\mathbf{Y}_2 \mid \mathbf{X}_2 S_n,\ s_0) + \log A \leqslant \\ &\leqslant H(\mathbf{Y}_2 \mid S_n,\ s_0) - H(\mathbf{Y}_2 \mid \mathbf{X}_2 S_n,\ s_0) + \log A = \\ &= \sum_{s_n} q(s_n \mid s_0)\, I(\mathbf{Y}_2;\ \mathbf{X}_2 \mid s_n) + \log A \leqslant l\overline{C}_l + \log A. \end{aligned} \tag{4А.26}$$

Подставляя (4А.24), (4А.25) и (4А.26) в (4А.22), будем иметь

$$N\overline{C}_N \leqslant n\overline{C}_n + l\overline{C}_l + \log A, \tag{4А.27}$$

или

$$N\left[\overline{C}_N + \frac{\log A}{N}\right] \leqslant n\left[\overline{C}_n + \frac{\log A}{n}\right] + l\left[\overline{C}_l + \frac{\log A}{l}\right]. \tag{4А.28}$$

Из леммы 2 следует, что

$$\lim_{N\to\infty} \overline{C}_N = \lim_{N\to\infty}\left[\overline{C}_N + \frac{\log A}{N}\right] = \inf_N\left[\overline{C}_N + \frac{\log A}{N}\right]. \tag{4А.29}$$

5

ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КАНАЛА С ШУМАМИ

5.1. БЛОКОВЫЕ КОДЫ

В предыдущей главе мы показали, что если информация от заданного источника должна быть передана по заданному каналу и если энтропия источника на единицу времени *больше*, чем пропускная способность канала на единицу времени, то невозможно осуществить сколь угодно надежный прием данных этого источника. В этой главе будет показано, что если энтропия источника *меньше*, чем пропускная способность, то при определенных условиях можно осуществить сколь угодно надежный прием.

Используемый здесь подход совершенно отличен от того, который был рассмотрен в гл. 4. Там был установлен отрицательный результат: независимо от того, как кодировать и декодировать при скоростях передачи, бóльших пропускной способности, нельзя избежать некоторой вероятности ошибки. Следовательно, чтобы избежать ошибки, мы должны снизить скорость передачи или улучшить канал. Чтобы доказать это в общем случае, нельзя было накладывать никаких ограничений на кодер и декодер. Для того чтобы показать, что надежная передача возможна при скоростях, меньших пропускной способности, можно уже накладывать какие угодно ограничения на вид кодера и декодера. В действительности такие ограничения часто дают возможность проникнуть в существо методов для достижения надежной передачи.

Первое ограничение, которое будет введено, сводится к тому, что мы разобьем кодер и декодер на кодер и декодер для источника и кодер и декодер для канала (рис. 5.1.1). Кодер для источника преобразует выход источника в поток двоичных символов; кодер для канала — двоичные данные в буквы на входе канала; декодер для канала пытается преобразовать выход канала в первоначальный двоичный поток, а декодер для источника пытается восстановить исходный поток символов источника. Такое разделение имеет очевидные преимущества с практической точки зрения, так как двоичные данные позволяют стандартизировать сочленение источников и каналов. Для теории это разделение имеет даже большую важность, так как оно позволяет отделить задачу передачи в шумах от задачи представления источника. Задача представления источника уже обсуждалась в гл. 3. Поэтому в этой главе (за исключением задач) источник не будет рассматриваться и будет принято предположение, что его выход уже преобразован в двоичные данные.

Предположим, что двоичные символы поступают к кодеру для канала со скоростью 1 двоичный символ за τ_s сек и что канал является дискретным во времени и передает один символ за τ_c сек. В этой главе ограничим наше рассмотрение лишь *блоковыми кодерами*. Эти кодеры, которые делят поступающий поток двоичных данных на последовательности равной длины, скажем длины L двоичных символов. Имеются $M = 2^L$ различных двоичных последовательностей длины L и кодер сопоставляет *кодовое слово* для каждой из них. Каждое кодовое слово представляет собой последовательность фиксированной длины N букв на входе канала. Число N называется *длиной блокового кода*, и оно будет равно наибольшему целому числу, меньшему $L\tau_s/\tau_c$:

$$N = \lfloor L\tau_s/\tau_c \rfloor. \tag{5.1.1}$$

Если $L\tau_s/\tau_c$ является целым числом, то время, необходимое для того, чтобы L двоичных символов поступили в кодер, равно времени, тре-

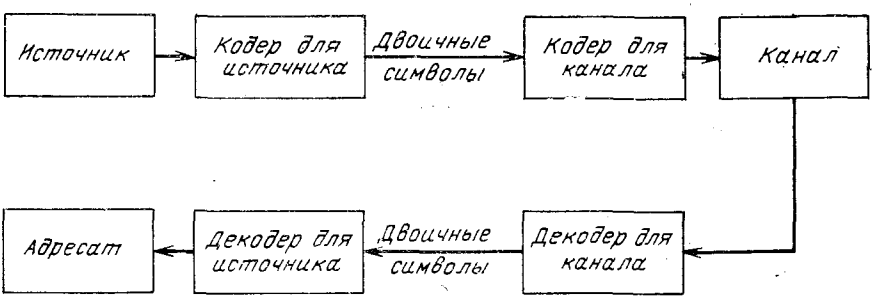


Рис. 5.1.1.

буемому для передачи кодового слова из N символов канала. Если $L\tau_s/\tau_c$ не является целым числом, то иногда нужно будет передавать по каналу «глухой символ», чтобы синхронизировать друг с другом поток двоичных данных и последовательность, передаваемую по каналу. Приемник принимает непрерывный поток символов на выходе канала. Они разделяются на последовательности длины N, соответствующие переданным последовательностям длины N. Декодер делает попытку угадать на основе принятой последовательности длины N, какая из соответствующих последовательностей L двоичных символов имела место. На практике, конечно, может возникать задача синхронизации приемника и передатчика, т. е. задача определения начала блока из N символов. Эта задача не будет здесь рассматриваться, так как ее рассмотрение здесь просто бы затемнило построение кодирования, противостоящего действию шума в канале.

Обозначим $M = 2^L$ — число кодовых слов, соответствующих 2^L двоичным последовательностям источника, через $\mathbf{x}_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,N}), \ldots, \mathbf{x}_m = (x_{m,1}, \ldots, x_{m,N}), \ldots, \mathbf{x}_M = (x_{M,1}, \ldots, \ldots, x_{M,N})$. Соответствие между целыми числами от 1 до M и двоичными последовательностями является произвольным и может быть выбрано, например, как запись этих целых чисел в двоичной системе. Однако

каким бы это соответствие ни было, оно будет считаться фиксированным, и когда двоичная последовательность, соответствующая целому числу m, поступает на кодер, то кодовое слово $\mathbf{x}_m$ передается по каналу. Рис. 5.1.2 дает пример блокового кода с длиной блока 5 для канала, входной алфавит которого состоит из трех букв. Если, например, на кодер поступает последовательность (0,1), то по каналу передается последовательность (2, 2, 1, 0, 1). Предположим здесь, что $\tau_s/\tau_c = {}^5/_2$, так что 5 символов канала могут быть переданы за время, требуемое для поступления двух двоичных символов в кодер.

Скорость R блокового кода определяется как

$$R = \frac{\log M}{N} = \frac{L \log 2}{N}. \qquad (5.1.2)$$

Если $L\tau_s/\tau_c$ является целым числом, то согласно (5.1.1) это выражение преобразуется к виду $R = (\tau_c/\tau_s) \log 2$. Таким образом, R в битах

Двоичная последовательность	m	*Кодовое слово* $\mathbf{x}_m$
00	1	0 0 0 0 0
01	2	1 1 1 1 0
10	3	2 2 1 0 1
11	4	0 2 2 2 2

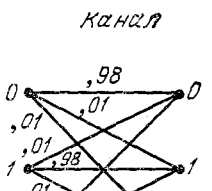


Рис. 5.1.2. Код с блоковой длиной $N=5$ и $M=4$ кодовыми словами.

(т. е. при использовании $\log_2$) является числом двоичных символов, поступающих на кодер за время передачи символа канала. Когда инженер связист говорит о скорости передачи данных, он обычно подразумевает число двоичных символов в секунду, которые поступают на передатчик. Отличие скорости в битах, которая была определена здесь, от этой скорости состоит в том, что последняя нормирована на символ в секунду, а не на символ на одно использование канала, как первая. Повсюду в этой главе будет удобно пользоваться натуральными логарифмами и в натуральных логарифмах R равна $(\ln M)/N$. Переход от скорости R (в натуральных единицах на символ) к скорости, которую используют инженеры связисты (в двоичных символах на секунду) задается соотношением

$$R = (\text{скорость передачи данных}) \cdot \tau_c \ln 2. \qquad (5.1.3)$$

Следует заметить, что R не является энтропией (хотя она может быть интерпретирована как энтропия, если символы двоичного источника независимы и равновероятны) и R не является, в общем случае, средней взаимной информацией для канала.

Если возникающие иногда «глухие символы» или другие служебные символы должны быть переданы по каналу для того, чтобы установить синхронизацию, то мы будем продолжать определять $R=(\ln M)/N$,

но (5.1.3) не будет справедливо строго и скорость передачи данных будет несколько меньше, чем указывает равенство (5.1.3).

Пусть $\mathbf{y} = (y_1, \ldots, y_N)$ является выходной последовательностью канала, соответствующей некоторому кодовому слову на входе. Задача декодера состоит в том, чтобы на основе $\mathbf{y}$ построить гипотезу о том, какая из $M = 2^L$ двоичных последовательностей поступила на кодер. Будем считать, что произошла *ошибка при блоковом декодировании*, если гипотеза, построенная декодером, отличается от последовательности, поступившей на кодер. Ошибка при блоковом декодировании означает, что произошли одна или более ошибок в последовательности L двоичных символов, но это событие не указывает, сколько ошибок произошло. В большинстве систем передачи данных интересным является как то, сколько ошибок произошло, так и то, как они распределены. Ни вероятность ошибки на символ, ни вероятность ошибки на блок не являются полностью адекватными мерами качества системы передачи. Здесь для простоты исследуется вероятность ошибки на блок. Однако, как будет показано, эта вероятность ошибки может быть сделана настолько малой при больших N и L, что различие между ошибками на блок и на символ приобретает второстепенное значение.

Так как нас будут интересовать ошибки на блок, то можно теперь не рассматривать двоичные последовательности и считать, что кодирование состоит просто в отображении целых чисел от 1 до M в кодовые слова от $\mathbf{x}_1$ до $\mathbf{x}_M$. Если сообщение m поступает на кодер, то $\mathbf{x}_m$ передается и декодер вырабатывает целое число m' на основе принятой последовательности $\mathbf{y}$. Ошибка происходит, если $m' \neq m$.

Причиной, по которой мы ограничиваемся здесь рассмотрением блоковых кодов, является не то, что они лучше, в каком-то смысле, чем другие типы кодов, а просто потому, что их легче исследовать теоретически. В следующей главе будет рассмотрен важный класс неблоковых кодов, которые называются сверточными кодами. Они имеют некоторые преимущества при реализации, но их можно понять более основательно после того, как будут поняты свойства блоковых кодов.

При построении хороших блоковых кодов главными параметрами, на которые будет обращено внимание, являются: вероятность ошибки при блоковом декодировании, обозначаемая через P_e; длина блока N и скорость R. Не удивительным является то, что уменьшением R (которое можно достичь уменьшением числа двоичных символов в секунду, поступающих на кодер) можно также уменьшить P_e. Удивительным является то, что если R меньше, чем пропускная способность канала, то можно при фиксированной R, увеличивая N, найти коды, для которых P_e экспоненциально убывает с ростом N. Это составляет существо теоремы кодирования, доказываемой в § 5.6. Существует, конечно, цена для расплаты за то, что длина блока увеличивается. Во-первых, происходит задержка в системе. Первый двоичный символ блока поступающих данных, вообще говоря, должен быть задержан на $N\tau_c$ сек до того, как будет сформировано кодовое слово и после этого еще на $N\tau_c$ сек, необходимых для передачи кодового слова, по которому этот двоичный символ будет декодирован. Во-вторых,

возникает проблема, состоящая в том, что число кодовых слов в коде задается $M = e^{NR}$ и, таким образом, следует ожидать быстрого роста сложности кодера и декодера с ростом N. Эта проблема сложности будет исследована в гл. 6, где будет показано, что этот рост сложности намного слабее, чем можно было бы ожидать, но отнюдь не является малозначительным, В большинстве систем, использующих кодирование, проблема задержки имеет существенно меньшее значение, чем сложность.

5.2. ДЕКОДИРОВАНИЕ БЛОКОВЫХ КОДОВ

Правило декодирования с *минимальной вероятностью ошибки* является правилом, которое минимизирует вероятность ошибочного декодирования для заданных ансамбля сообщений, множества кодовых слов и канала. Пусть $P_N(\mathbf{y} \mid \mathbf{x}_m)$ — вероятность приема последовательности $\mathbf{y}$ при условии, что было передано m-е кодовое слово. Для дискретного канала без памяти она выражается через переходные вероятности канала $P(y \mid x)$ как

$$P_N(\mathbf{y} \mid \mathbf{x}_m) = \prod_{n=1}^{N} P(y_n \mid x_{m,n}). \tag{5.2.1}$$

Если априорные вероятности сообщений равны $\Pr(m)$, то апостериорная вероятность сообщения m при условии, что принята последовательность $\mathbf{y}$, равна

$$\Pr(m \mid \mathbf{y}) = \frac{P_N(\mathbf{y} \mid \mathbf{x}_m) \Pr(m)}{\Pr(\mathbf{y})}, \tag{5.2.2}$$

где

$$\Pr(\mathbf{y}) = \sum_{m=1}^{M} \Pr(m) P_N(\mathbf{y} \mid \mathbf{x}_m).$$

Если декодер декодирует последовательность $\mathbf{y}$ в сообщение m, то вероятность (при заданном $\mathbf{y}$) того, что декодирование является ошибочным, равна $1 - \Pr(m \mid \mathbf{y})$. Декодер минимизирует вероятность ошибки выбором m, которое максимизирует $\Pr(m \mid \mathbf{y})$. *Таким образом, правило декодирования с минимальной вероятностью ошибки определяется следующим образом: следует декодировать принятую последовательность* $\mathbf{y}$ *в* m', *для которого*

$$\Pr(m' \mid \mathbf{y}) \geqslant \Pr(m \mid \mathbf{y}) \text{ для всех } m \neq m'. \tag{5.2.3}$$

Если при заданном $\mathbf{y}$ величина $\Pr(m \mid \mathbf{y})$ максимизируется на нескольких различных значениях m, то ясно, что нет разницы, какое из этих значений следует выбрать. Так как знаменатель в (5.2.2) не зависит от m, то правило, эквивалентное правилу декодирования с минимальной вероятностью ошибки, состоит в следующем: следует декодировать $\mathbf{y}$ в m', для которого

$$\Pr(m') P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \geqslant \Pr(m) P_N(\mathbf{y} \mid \mathbf{x}_m) \text{ для всех } m \neq m'. \tag{5.2.4}$$

Другим правилом декодирования является декодирование по максимуму правдоподобия, которое определяется следующим образом: при заданном $\mathbf{y}$ *надлежит выбрать* m', *для которого*

$$P_N(\mathbf{y}\,|\,\mathbf{x}_{m'}) \geqslant P_N(\mathbf{y}\,|\,\mathbf{x}_m) \quad \textit{для всех} \quad m \neq m'. \tag{5.2.5}$$

Очевидным преимуществом декодирования по максимуму правдоподобия является то, что оно может быть применено тогда, когда априорные вероятности сообщений не определены или не имеют смысла. Название «максимум правдоподобия» является отчасти дезориентирующим, так как соответствующее ему правило не обязательно приводит к сообщению, которое наиболее вероятно при заданном $\mathbf{y}$. Вместо этого выбирается сообщение, для которого данное $\mathbf{y}$ наиболее вероятно при заданном m [сравните (5.2.3) и (5.2.5)]. В частном случае, когда сообщения имеют равные априорные вероятности, можно заметить, что (5.2.4) и (5.2.5) эквивалентны, и в этом случае декодирование по максимуму правдоподобия минимизирует вероятность ошибки.

Другим правилом декодирования, полезным тогда, когда неравные стоимости соответствуют различным типам ошибок, является декодирование с *минимальной стоимостью*. При этом $\mathbf{y}$ декодируется в m, которое минимизирует среднюю стоимость (см. задачу 5.1). Этот класс задач рассматривается в гл. 9 с более фундаментальных позиций.

Наконец, в большинстве практических применений кодирования необходимо при выборе правил декодирования принимать во внимание простоту их реализации; связанные с этим задачи будут обсуждаться в следующей главе.

Итак, до сих пор мы рассматривали правила декодирования, по которым строится гипотеза о сообщении по заданной принятой последовательности канала. Однако, если шум особенно велик, часто при декодировании лучше отказаться от построения этой гипотезы, и в этом случае считается, что возникает *обнаруживаемая ошибка*. Способность к обнаружению ошибок, в частности, является полезной тогда, когда приемник имеет возможность передавать информацию назад к передатчику, который может вновь передать искаженные блоки.

Правило декодирования теперь можно определить формально, как отображение множества $\mathbf{Y}^N$ *последовательностей на выходе канала в множество, состоящее из* M *сообщений, и выход, соответствующий обнаруженной ошибке.* Обозначим множество последовательностей декодируемых в сообщение m через Y_m, а дополнение этого множества через Y_m^c. Когда источник вырабатывает сообщение m, передается кодовое слово $\mathbf{x}_m$ и возникает ошибка (либо необнаруживаемая, либо обнаруживаемая), если принятая последовательность $\mathbf{y}$ принадлежит множеству Y_m^c. Таким образом, вероятность ошибочного декодирования при условии, что было послано сообщение m, равна

$$P_{e,\,m} = \sum_{\mathbf{y} \in Y_m^c} P_N(\mathbf{y}\,|\,\mathbf{x}_m). \tag{5.2.6}$$

Тогда общая вероятность ошибочного декодирования, в случае, когда априорные вероятности сообщений имеют вероятности $\Pr(m)$,

равна

$$P_e = \sum_{m=1}^{M} \Pr(m) P_{e,m}. \tag{5.2.7}$$

Для примера рассмотрим код, представленный на рис. 5.2.1. В нем есть два кодовых слова $\mathbf{x}_1=(0,0,0)$ и $\mathbf{x}_2=(1,1,1)$. Правило декодирования (которое, как можно заметить, является правилом максимального правдоподобия для ДСК) состоит в том, чтобы декодировать каждую из последовательностей (0,0,0), (0,0,1), (0,1,1) и (1,0,0) в сообщение 1 и другие последовательности в сообщение 2. Следовательно, Y_1^c совпадает с Y_2 и является множеством последовательностей (0,1,1), (1,0,1) (1,1,0) и (1,1,1). Для ДСК, изображенного на рис 5.2.1, $P_3[(0,1,1) \mid (0,0,0)]=(1-\varepsilon)\varepsilon^2$, например, и подобное вычисление $P_3(\mathbf{y} \mid \mathbf{x}_1)$ для других $\mathbf{y}$ из Y_1^c дает $P_{e,1}=3(1-\varepsilon)\varepsilon^2+\varepsilon^3$.

Соотношения (5.2.5) и (5.2.7) по виду довольно безобидные. Однако, если алфавит на выходе канала состоит из J букв, то в этих суммах оказываются J^N слагаемых. Для умеренных длин блоков, таких, как $N = 50$, вычисления сумм выходят далеко за возможности современных вычислительных машин. Даже если такие вычисления могли бы быть выполнены, они дали бы лишь небольшое проникновение в задачу выбора подходящей длины блока для кода или подходящего множества кодовых слов. Развиваемый здесь подход направлен скорее не на то, чтобы вычислить P_e, а чтобы найти простые верхние границы для достижимой вероятности ошибки. Идя по этому пути, мы не только докажем теорему кодирования, но также получим значительное понимание задачи выбора подходящих параметров кодирования. Начнем с рассмотрения вероятности ошибки для множества из двух кодовых слов и затем обобщим результат на произвольно большое множество кодовых слов.

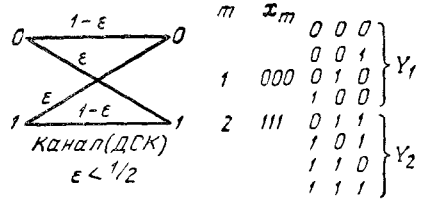


Рис. 5.2.1. Код и правило декодирования для $N=3$, $M=2$.

5.3. ВЕРОЯТНОСТЬ ОШИБКИ ДЛЯ ДВУХ КОДОВЫХ СЛОВ

Пусть $\mathbf{x}_1$ и $\mathbf{x}_2$ — два кодовых слова длины N и предположим, что используется декодирование по максимуму правдоподобия, т. е. декодируется сообщение 1, если $P_N(\mathbf{y} \mid \mathbf{x}_1) > P_N(\mathbf{y} \mid \mathbf{x}_2)$; в противном случае декодируется сообщение 2. Согласно (5.2.6) вероятность ошибки при посылке сообщения 1 равна

$$P_{e,1} = \sum_{\mathbf{y} \in Y_1^c} P_N(\mathbf{y} \mid \mathbf{x}_1).$$

При любом $\mathbf{y} \in Y_1^c$ можно оценить сверху $P_N(\mathbf{y} \mid \mathbf{x}_1)$ следующим образом:

$$P_N(\mathbf{y} \mid \mathbf{x}_1) \leqslant P_N(\mathbf{y} \mid \mathbf{x}_1)^{1-s} P_N(\mathbf{y} \mid \mathbf{x}_2)^s, \quad s \text{ — любое}, \quad 0 < s < 1. \quad (5.3.1)$$

Граница (5.3.1) следует из того, что $P_N(\mathbf{y} \mid \mathbf{x}_2) \geqslant P_N(\mathbf{y} \mid \mathbf{x}_1)$ при $\mathbf{y} \in Y_1^c$ и, следовательно, $P_N(\mathbf{y} \mid \mathbf{x}_2)^s \geqslant P_N(\mathbf{y} \mid \mathbf{x}_1)^s$. Граница (5.3.1) справедлива также при $s \geqslant 1$, но это здесь не будет использовано. Подставляя (5.3.1) в (5.2.6) и строя границу сверху с помощью суммирования по всем $\mathbf{y}$, будем иметь

$$P_{e,1} \leqslant \sum_{\mathbf{y}} P_N(\mathbf{y} \mid \mathbf{x}_1)^{1-s} P_N(\mathbf{y} \mid \mathbf{x}_2)^s, \quad s \text{ — любое}, \quad 0 < s < 1. \quad (5.3.2)$$

Оценивая $P_{e,2}$ аналогичным образом, получаем

$$P_{e,2} \leqslant \sum_{\mathbf{y}} P_N(\mathbf{y} \mid \mathbf{x}_2)^{1-r} P_N(\mathbf{y} \mid \mathbf{x}_1)^r, \quad r \text{ — любое}, \quad 0 < r < 1. \quad (5.3.3)$$

Если подставить $1-s$ вместо r в (5.3.3), то можно заметить, что мы получили одинаковые границы для $P_{e,2}$ и $P_{e,1}$. Это не приводит к потере общности, так как s все еще остается произвольным, $0 < s < 1$:

$$P_{e,m} \leqslant \sum_{\mathbf{y}} P_N(\mathbf{y} \mid \mathbf{x}_1)^{1-s} P_N(\mathbf{y} \mid \mathbf{x}_2)^s, \quad m = 1, 2; \quad 0 < s < 1. \quad (5.3.4)$$

Как будет видно из дальнейшего, если соответствующим образом выбрать s, то граница (5.3.4) будет удивительно точной. Для канала без памяти (5.3.4) можно упростить и привести к виду:

$$P_{e,m} \leqslant \sum_{y_1} \sum_{y_2} \dots \sum_{y_N} \prod_{n=1}^{N} P(y_n \mid x_{1,n})^{1-s} P(y_n \mid x_{2,n})^s. \quad (5.3.5)$$

Расписывая произведение, получаем

$$P_{e,m} \leqslant \sum_{y_1} P(y_1 \mid x_{1,1})^{1-s} P(y_1 \mid x_{2,1})^s \sum_{y_2} P(y_2 \mid x_{1,2})^{1-s} P(y_2 \mid x_{2,2})^s \times \dots$$

$$\dots \times \sum_{y_N} P(y_N \mid x_{1,N})^{1-s} P(y_N \mid x_{2,N})^s,$$

$$P_{e,m} \leqslant \prod_{n=1}^{N} \sum_{y_n} P(y_n \mid x_{1,n})^{1-s} P(y_n \mid x_{2,n})^s, \quad m = 1, 2. \quad (5.3.6)$$

Сумма по y_n в (5.3.6) имеет фундаментальное значение для последующего изложения и ей будет дана дополнительная интерпретация в следующем параграфе. Обозначим ее через

$$g_n(s) = \sum_{y_n} P(y_n \mid x_{1,n})^{1-s} P(y_n \mid x_{2,n})^s. \quad (5.3.7)$$

При этом (5.3.6) можно переписать в виде

$$P_{e,m} \leqslant \prod_{n=1}^{N} g_n(s), \quad m = 1, 2; \quad 0 < s < 1. \quad (5.3.8)$$

Функция $g_n(s)$ изображена для ряда каналов на рис. 5.3.1. Во всех случаях $x_{1,n}$ обозначает вход канала 0, а $x_{2,n}$ — вход канала 1. Для каналов, подобных тому, который представлен на рис. 5.3.1, *б* и в котором некоторые переходные вероятности равны нулю, $g_n(0)$

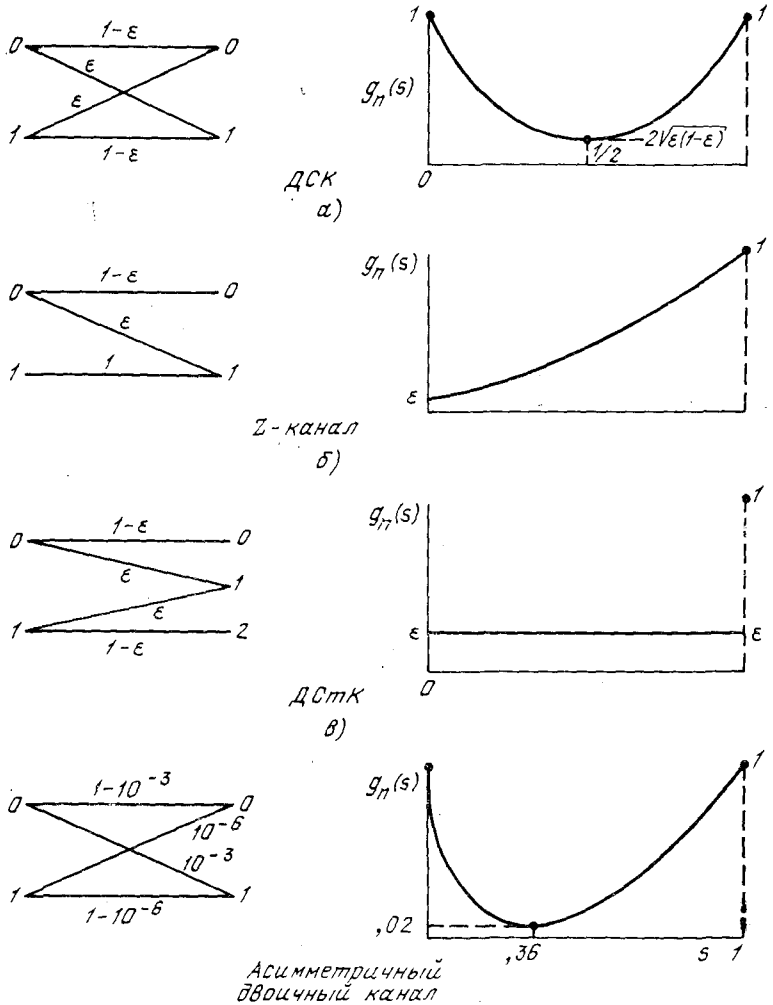


Рис. 5.3.1. Функция $g_n(s) = \sum_{j=0}^{J-1} P(j \mid 0)^{1-s} P(j \mid 1)^s$.

и $g_n(1)$ могут быть не определены. Для этих случаев определим $g_n(0)$ и $g_n(1)$ равенствами

$$g_n(0) = \lim_{s \to 0+} g_n(s) = \sum_{\substack{\text{по } y_n, \text{ для которых} \\ P(y_n \mid x_{2,n}) \neq 0}} P(y_n \mid x_{1,n}), \qquad (5.3.9)$$

$$g_n(1) = \lim_{s \to 1^-} g_n(s) = \sum_{\substack{\text{по } y_n, \text{ для которых} \\ P(y_n \mid x_{2,n}) \neq 0}} P(y_n \mid x_{2,n}). \qquad (5.3.10)$$

Можно заметить, что $g_n(0)$ и $g_n(1)$ всегда меньше или равны 1. Отсюда также следует, если взять вторую производную, что $g_n(s)$ выпукла $\cup$ при $0 \leqslant s \leqslant 1$. Следовательно, $g_n(s) \leqslant 1$ при $0 \leqslant s \leqslant 1$. В задаче 4.15 (а) показано, что $g_n(s) < 1$ при $0 < s < 1$, если $P(y_n \mid x_{1,n}) = P(y_n \mid x_{2,n})$ не имеет места одновременно для всех выходов y_n.

Используя определения (5.3.9) и (5.3.10), находим, что неравенство (5.3.8) справедливо для всех s, $0 \leqslant s \leqslant 1$. Очевидно, можно получить наилучшую границу, если провести минимизацию по s:

$$P_{e,m} \leqslant \min_{0 \leqslant s \leqslant 1} \prod_{n=1}^{N} g_n(s); \quad m = 1, 2. \qquad (5.3.11)$$

Пример. Рассмотрим двоичный симметричный канал, изображенный на рис. 5.3.1, *а*. Пусть $\mathbf{x}_1$ — последовательность N нулей, а $\mathbf{x}_2$ — последовательность N единиц. Тогда

$$g_n(s) = \varepsilon^{1-s}(1-\varepsilon)^s + \varepsilon^s(1-\varepsilon)^{1-s}; \quad 1 \leqslant n \leqslant N.$$

Это выражение минимизируется при $s = 1/2$, что дает

$$\min g_n(s) = g_n(1/2) = 2\sqrt{\varepsilon(1-\varepsilon)}, \qquad (5.3.12)$$

$$P_{e,m} \leqslant \left[2\sqrt{\varepsilon(1-\varepsilon)}\right]^N; \quad m = 1, 2. \qquad (5.3.13)$$

Для этого простого примера $P_{e,1}$ и $P_{e,2}$ можно точно подсчитать. $P_N(\mathbf{y} \mid \mathbf{x}_2)$ будет больше или равна $P_N(\mathbf{y} \mid \mathbf{x}_1)$, если принятая последовательность содержит $N/2$ или более единиц. Вероятность этого, когда сообщение 1 было передано (в предположении, что N четно), равна

$$P_{e,1} = \sum_{i=N/2}^{N} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}. \qquad (5.3.14)$$

Слагаемые этой суммы являются членами биномиального разложения, сконцентрированного около $i = \varepsilon N$. В силу того, что $\varepsilon < 1/2$, наибольшим слагаемым в сумме является первое слагаемое, в котором $i = N/2$. Применяя формулу Стирлинга для факториала $N! \approx \sqrt{2\pi N} \times N^N e^{-N}$, получаем

$$\binom{N}{N/2} \approx \sqrt{\frac{2}{\pi N}}\, 2^N, \qquad (5.3.15)$$

$$P_{e,1} \approx \sqrt{\frac{2}{\pi N}}\left[2\sqrt{\varepsilon(1-\varepsilon)}\right]^N + \text{меньшие слагаемые}. \qquad (5.3.16)$$

В задаче 5.2 (в) построена аппроксимация для меньших слагаемых в (5.3.16), что дает явную аппроксимацию для $P_{e,1}$. Вопрос, который, однако, нас интересует здесь, состоит в выяснении экспоненциальной зависимости $P_{e,1}$ от N; нам также интересно то, что эта экспоненци-

альная зависимость согласуется с той, которая представлена границей (5.3.13).

Так как, по условию, декодируется сообщение 2, если $P_N(\mathbf{y}\,|\,\mathbf{x}_2) = P_N(\mathbf{y}\,|\,\mathbf{x}_1)$, то получаем

$$P_{e,2} = \sum_{i=(N/2)+1}^{N} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}.$$

В задаче 5.2 (в) показано также, что экспоненциальная зависимость $P_{e,2}$ от N является той же самой, что и для $P_{e,1}$, и что та же самая экспоненциальная зависимость имеет место, если N нечетно.

Для более сложных каналов получение хорошей аппроксимации для $P_{e,1}$ и $P_{e,2}$ много труднее. Однако оказывается, что (5.3.11) всегда дает правильную экспоненциальную зависимость $^1/_2 (P_{e,1} + P_{e,2})$ от N (см. Шеннон, Галлагер и Берлекэмп 1 (1967), § 3). Это будет рассмотрено более детально в конце настоящей главы.

5.4. ОБОБЩЕННОЕ НЕРАВЕНСТВО ЧЕБЫШЕВА И ГРАНИЦА ЧЕРНОВА

В этом параграфе результаты § 5.3 будут выведены вновь в более общем виде, что даст тем самым дополнительное понимание использованных там методов. При передаче сообщения 1 и декодировании по максимуму правдоподобия для двух кодовых слов ошибка происходит, если $P_N(\mathbf{y}\,|\,\mathbf{x}_2) \geqslant P_N(\mathbf{y}\,|\,\mathbf{x}_1)$. Другими словами, ошибка происходит, если логарифм отношения правдоподобия $w(\mathbf{y})$ удовлетворяет условию

$$w(\mathbf{y}) \triangleq \ln \frac{P_N(\mathbf{y}\,|\,\mathbf{x}_2)}{P_N(\mathbf{y}\,|\,\mathbf{x}_1)} \geqslant 0. \tag{5.4.1}$$

Для каналов без памяти согласно (5.2.1) $w(\mathbf{y})$ можно представить в виде суммы N слагаемых

$$w(\mathbf{y}) = \sum_{n=1}^{N} z_n(y_n), \tag{5.4.2}$$

где

$$z_n(y_n) = \ln \frac{P(y_n\,|\,x_{2,n})}{P(y_n\,|\,x_{1,n})}. \tag{5.4.3}$$

При условии, что передается сообщение 1, z_n, $1 \leqslant n \leqslant N$, являются независимыми случайными величинами, каждая из которых принимает указанные значения с вероятностями $P(y_n\,|\,x_{1,n})$. Таким образом, $w(\mathbf{y})$ является суммой независимых случайных величин и значение $P_{e,1}$ задается равенством

$$P_{e,1} = \Pr[w(\mathbf{y}) \geqslant 0\,|\,\text{передано сообщение 1}]. \tag{5.4.4}$$

Отыскание эффективных границ для вероятности того, что сумма независимых случайных величин превосходит некоторое заданное число, является задачей общей как для теории информации, так и те-

ории вероятности, и поэтому она заслуживает более общего рассмотрения.

Предположим вначале, что t — случайная величина, принимающая только неотрицательные значения. Простейшей формой неравенства Чебышева является утверждение, что при любом $\delta > 0$

$$\Pr(t \geqslant \delta) \leqslant \frac{\bar{t}}{\delta}, \tag{5.4.5}$$

где $\bar{t}$ — среднее значение t. Для того чтобы доказать это, предположим, что t является дискретной случайной величиной с распределением вероятности $P(t)$. Тогда

$$\Pr(t \geqslant \delta) = \sum_{t \geqslant \delta} P(t) \leqslant \sum_{t \geqslant \delta} P(t) \frac{t}{\delta}.$$

Это неравенство возникает потому, что $t/\delta \geqslant 1$ в области суммирования. Так как t/δ является неотрицательным при всех t, то можно ограничить далее это выражение сверху, суммируя по всем t; в результате получим (5.4.5). В точности такое же доказательство лишь с небольшими изменениями в обозначениях, очевидно, применимо к случаю недискретной случайной величины. Эквивалентная запись неравенства (5.4.5) получается при замене $\delta/\bar{t}$ на α. Имеем

$$\Pr(t \geqslant \alpha\bar{t}) \leqslant 1/\alpha. \tag{5.4.6}$$

Чтобы показать на примере, насколько малоэффективно это неравенство, предположим, что t — рост случайно выбранного человека. Если $\bar{t} = 152{,}5$ см, то (5.4.5) утверждает, что вероятность того, что выбранный человек окажется выше чем 3 м 5 см, не больше 1/2, а вероятность того, что выбранный человек окажется выше чем 15 м 25 см не больше 1/10.

Предположим теперь, что w является произвольной случайной величиной со средним значением $\bar{w}$ и дисперсией σ^2. Если положить в (5.4.5) $t = (w - \bar{w})^2$, то получим

$$\Pr[(w-\bar{w})^2 \geqslant \delta] \leqslant \sigma^2/\delta. \tag{5.4.7}$$

Обозначая $\varepsilon = \sqrt{\delta}$, это неравенство можно переписать в виде, в котором обычно представляется неравенство Чебышева,

$$\Pr[|w-\bar{w}| \geqslant \varepsilon] \leqslant \sigma^2/\varepsilon^2. \tag{5.4.8}$$

Можно получить большое число других неравенств, называемых обобщенными неравенствами Чебышева, если положить t, равным другим функциям от w. Здесь будет особенно интересно неравенство, которое обычно называется границей Чернова; оно получается, если положить $t = e^{sw}$, где s — произвольное действительное число. В результате получим

$$\Pr[e^{sw} \geqslant \delta] \leqslant \overline{e^{sw}}/\delta. \tag{5.4.9}$$

Математическое ожидание e^{sw} является производящей функцией моментов случайной величины w:

$$g_w(s) = \overline{e^{sw}} = \sum_w P(w)\, e^{sw}. \tag{5.4.10}$$

Пусть δ в (5.4.9) равно e^{sA}, где A — произвольное действительное число. При $s > 0$ неравенство $e^{sw} \geqslant e^{sA}$ эквивалентно неравенству $w \geqslant A$, так что (5.4.9) принимает вид

$$\Pr[w \geqslant A] \leqslant e^{-sA} g_w(s) \text{ при любом } s > 0. \tag{5.4.11}$$

Аналогично, если $s < 0$, то неравенство $e^{sw} \geqslant e^{sA}$ эквивалентно неравенству $w \leqslant A$, что дает

$$\Pr[w \leqslant A] \leqslant e^{-sA} g_w(s), \quad s < 0. \tag{5.4.12}$$

Зависимость функции $e^{-sA} g_w(s)$ от s изображена на рис. 5.4.1. Она принимает значение 1 при $s = 0$, и ее первая производная при $s = 0$

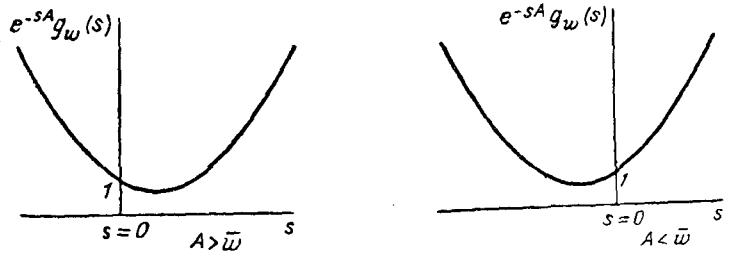


Рис. 5.4.1. Вид границы Чернова.

равна $\bar{w} - A$. Вторая производная всегда положительна, что следует из равенства $e^{-sA} g_w(s) = \overline{\exp[s(w - A)]}$. Поэтому, если $A > \bar{w}$, то граница, устанавливаемая (5.4.12), больше чем 1 при всех $s < 0$ и, следовательно, является бесполезной. Аналогично, если $A < \bar{w}$, то бесполезной является граница в (5.4.11). Другими словами, (5.4.11) и (5.4.12) могут быть использованы только при оценивании «хвостов» распределения.

В силу того, что $e^{-sA} g_w(s)$ является функцией выпуклой $\cup$ по s, то наиболее точную границу можно получить, найдя стационарную точку этой функции. Имеем

$$A = \frac{dg_w(s)}{ds} / g_w(s). \tag{5.4.13}$$

Для большинства применений более удобно оставить s свободным параметром, чем считать, что s является решением (5.4.13).

Границы (5.4.11) и (5.4.12) полезны в основном тогда, когда w является суммой статистически независимых случайных величин:

$$w = \sum_{n=1}^{N} z_n.$$

В этом случае производящая функция моментов случайной величины w может быть выражена через производящие функции моментов случайных величин z_n следующим образом. Имеем

$$g_w(s) = \overline{\exp\left(s \sum_{n=1}^{N} z_n\right)} = \overline{\prod_{n=1}^{N} \exp(sz_n)}.$$

Так как z_n статистически независимы, то математическое ожидание произведения равно произведению математических ожиданий, и, таким образом,

$$g_w(s) = \prod_{n=1}^{N} \overline{\exp(sz_n)} = \prod_{n=1}^{N} g_n(s), \tag{5.4.14}$$

где $g_n(s)$ — производящая функция моментов случайной величины z_n. Подставляя (5.4.14) в (5.4.11) и (5.4.12), получаем

$$\Pr[w \geqslant A] \leqslant e^{-sA} \prod_{n=1}^{N} g_n(s), \quad s > 0, \tag{5.4.15}$$

$$\Pr[w \leqslant A] \leqslant e^{-sA} \prod_{n=1}^{N} g_n(s), \quad s < 0. \tag{5.4.16}$$

Неравенство (5.4.15) можно теперь применить к оценке вероятности ошибки $P_{e,1}$ для кода с двумя кодовыми словами [см. (5.4.4)]. В этом случае $A = 0$, z_n задается равенством (5.4.3) и вероятностная мера берется при условии, что было передано сообщение 1. Имеем

$$g_n(s) = \sum_{y_n} P(y_n \mid x_{1,n}) \exp\left[s \ln \frac{P(y_n \mid x_{2,n})}{P(y_n \mid x_{1,n})}\right] =$$

$$= \sum_{y_n} P(y_n \mid x_{1,n})^{1-s} P(y_n \mid x_{2,n})^s, \tag{5.4.17}$$

$$P_{e,1} \leqslant \prod_{n=1}^{N} g_n(s) = \prod_{n=1}^{N} \left[\sum_{y_n} P(y_n \mid x_{1,n})^{1-s} P(y_n \mid x_{2,n})^s\right]. \tag{5.4.18}$$

Это в точности совпадает с результатом (5.3.6).

Исследуем теперь вопрос о том, насколько точны границы (5.4.15) и (5.4.16). Грубо говоря, ответ состоит в том, что если N велико, A далеко от $\overline{w}$ и s минимизирует границу, то подходящим образом выбранная граница (граница (5.4.15), если $A > \overline{w}$, и (5.4.16), если $A < \overline{w}$)) также дает хорошую оценку для $\Pr[w \geqslant A]$. Для того чтобы сформулировать это точнее, удобно слегка изменить вид границ (5.4.15) и (5.4.16). По определению, производящей функцией семиинвариантов случайной величины является натуральный логарифм ее производящей функции моментов. Таким образом, производящей функцией семиинвариантов случайной величины w является $\mu_w(s) = \ln g_w(s)$, и аналогичная функция для z_n равна $\mu_n(s) = \ln g_n(s)$.

Эти функции связаны с помощью соотношения (5.4.14) следующим образом:

$$\mu_w(s) = \ln \prod_{n=1}^{N} g_n(s) = \sum_{n=1}^{N} \mu_n(s). \tag{5.4.19}$$

Точно так же из (5.4.13) следует, что s, которое оптимизирует границу, задается с помощью производной от $\mu_w(s)$ следующим образом:

$$A = \mu_w'(s) = \sum_{n=1}^{N} \mu_n'(s). \tag{5.4.20}$$

Подставляя (5.4.19) и (5.4.20) в (5.4.15) и (5.4.16), получаем параметрические границы

$$\Pr\left[w \geqslant \sum_{n=1}^{N} \mu_n'(s)\right] \leqslant \exp\left[\sum_{n=1}^{N} \mu_n(s) - s\mu_n'(s)\right]; \quad s > 0, \tag{5.4.21}$$

$$\Pr\left[w \leqslant \sum_{n=1}^{N} \mu_n'(s)\right] \leqslant \exp\left[\sum_{n=1}^{N} \mu_n(s) - s\mu_n'(s)\right]; \quad s < 0. \tag{5.4.22}$$

В приложении 5А найдены асимптотические выражения этих вероятностей в частном случае, когда z_n одинаково распределены. В этом случае $\mu_n(s)$ не зависит от n и можно опустить индекс n. Результат зависит от того, являются ли случайные величины z_n решетчатыми или нет*). Асимптотические выражения имеют вид

$$\Pr[w \geqslant N\mu'(s)] = \left[\frac{1}{|s|\sqrt{2\pi N\mu''(s)}} + o\left(\frac{1}{\sqrt{N}}\right)\right] \times \\ + \exp\{N[\mu(s) - s\mu'(s)]\}, \quad s > 0 \tag{5.4.23}$$

для нерешетчатой случайной величины и

$$\Pr[w \geqslant N\mu'(s)] = \left[\frac{he^{-|s|\Delta}}{\sqrt{2\pi N\mu''(s)}(1 - e^{-|s|h})} + o\left(\frac{1}{\sqrt{N}}\right)\right] \times \\ \times \exp\{N[\mu(s) - s\mu'(s)]\}, \quad s > 0 \tag{5.4.24}$$

для решетчатой случайной величины.

В этих выражениях функции $o(1/\sqrt{N})$ стремятся к нулю быстрее, чем $1/\sqrt{N}$ с ростом N. При любом заданном s функцией $o(1/\sqrt{N})$ можно пренебречь при достаточно больших N, хотя, когда s стремится к 0, это требуемое N становится все больше и больше. В (5.4.24) h является расстоянием между соседними выборочными значениями (см. приложение 5А), а Δ равно расстоянию между $N\mu'(s)$ и ближайшим большим последовательным выборочным значением w. При $s < 0$ те же самые выражения справедливы для $\Pr[w \leqslant N\mu'(s)]$.

---

*) Решетчатой случайной величиной называется случайная величина, которая принимает лишь значения $\alpha + hi$, где α и h — фиксированные числа, а i принимает целочисленные значения (см. приложение 5А).

ε	α	N	Истинное значение	Граница Чернова	Асимптотическое выражение	Гауссовское приближение
0,1	0,2	20	0,1327	0,4114	0,1650	0,1318
		100	$1{,}95\times10^{-3}$	$1{,}18\times10^{-2}$	$2{,}12\times10^{-3}$	$7{,}71\times10^{-3}$
	0,3	20	$1{,}12\times10^{-2}$	$4{,}63\times10^{-2}$	$1{,}22\times10^{-2}$	$4{,}54\times10^{-3}$
		100	$2{,}50\times10^{-8}$	$2{,}12\times10^{-7}$	$2{,}54\times10^{-8}$	$4{,}02\times10^{-11}$
0,5	0,6	20	0,2517	0,6685	0,3649	0,2512
		100	$2{,}85\times10^{-2}$	0,1335	$3{,}26\times10^{-2}$	$2{,}87\times10^{-2}$
		400	$3{,}68\times10^{-5}$	$3{,}18\times10^{-4}$	$3{,}88\times10^{-5}$	$3{,}91\times10^{-3}$
		1000	$1{,}36\times10^{-10}$	$1{,}80\times10^{-9}$	$1{,}39\times10^{-10}$	$1{,}56\times10^{-10}$
	0,8	20	$5{,}91\times10^{-3}$	$2{,}12\times10^{-2}$	$6{,}29\times10^{-3}$	$6{,}95\times10^{-3}$
		100	$5{,}60\times10^{-10}$	$4{,}26\times10^{-9}$	$5{,}66\times10^{-10}$	$1{,}82\times10^{-9}$

Иллюстрация поведения различных оценок хвостов биномиального распределения Pr $(w \geqslant N\alpha)$, где w — сумма N независимых одинаково распределенных двоичных случайных величин, принимающих значение 1 с вероятностью ε. Гауссовское приближение имеет вид:

$$1-\Phi\left[\frac{N(\alpha-\varepsilon)-\frac{1}{2}}{\sqrt{N\varepsilon(1-\varepsilon)}}\right], \text{ где } \Phi(x)=\int_{-\infty}^{x}(2\pi)^{-\frac{1}{2}}\exp\left(-z^2/2\right)dz.$$

Рис. 5.4.2.

Эти выражения остаются также в силе для непрерывных случайных величин z_n, если $\overline{\exp(sz_n)} < \infty$ для s в некоторой окрестности нуля.

Для различных биномиальных распределений на рис. 5.4.2 проведено сравнение асимптотического выражения, границы Чернова, истинного значения и гауссовского приближения. Можно заметить, что граница Чернова и асимптотические выражения в (5.4.24) являются плохими приближениями при A, близких к $\overline{w}$ (при малых s), но что при больших A гауссовское приближение является плохим, а граница Чернова и асимптотическое выражение — хорошими.

5.5. СЛУЧАЙНЫЕ КОДОВЫЕ СЛОВА

В последнем параграфе было показано, что вероятность ошибочного декодирования для двух кодовых слов стремится к нулю экспоненциально с ростом длины блока. Вместе с тем в этом случае передается только один двоичный символ источника на блок, так что вероятность ошибки уменьшается только за счет скорости передачи. Ясно, что единственной возможностью для снижения вероятности ошибки без уменьшения скорости передачи является рассмотрение большего множества кодовых слов.

Мы хотим исследовать минимально достижимую вероятность ошибочного декодирования как функцию скорости R, длины блока N и канала. Будет найдена верхняя граница для этой вероятности ошибочного декодирования, которая убывает экспоненциально с длиной блока для всех скоростей, меньших пропускной способности. Эта граница выводится с помощью рассмотрения ансамбля кодов, а не одно-

го-единственного хорошего кода. Этот своеобразный подход продиктован тем, что для представляющих интерес значений N и R не известно метода отыскания кодов, которые минимизируют вероятность ошибочного декодирования; и даже, если бы такие коды могли быть найдены, прямое вычисление вероятности ошибки было бы невозможным из-за большого числа последовательностей, возможных на приемном конце. В следующей главе будут рассмотрены несколько конкретных методов блокового кодирования и декодирования, которые интересны в силу относительной простоты их реализаций. Для некоторых из этих методов можно подсчитать вероятность ошибки, но при больших значениях N эта вероятность ошибки будет на много больше, чем верхняя граница для минимальной вероятности ошибки, выведенная здесь.

Для того чтобы определить ансамбль блоковых кодов, обозначим через $Q_N(\mathbf{x})$ произвольное распределение вероятности на множестве последовательностей длины N на входе канала и будем считать, что все кодовые слова выбираются независимо с одними и теми же вероятностями. Таким образом, вероятность некоторого частного кода $\mathbf{x}_1, \ldots, \mathbf{x}_M$ в этом ансамбле кодов равна $\prod_{m=1}^{M} Q_N(\mathbf{x}_m)$. Каждый код из ансамбля имеет свою собственную вероятность ошибочного декодирования, возникающую при декодировании по максимуму правдоподобия для этого кода. Мы оценим сверху математическое ожидание (по ансамблю) этой вероятности ошибки. Так как по крайней мере один код из ансамбля должен иметь такую же вероятность ошибки, как среднее по ансамблю, это даст границу для вероятности ошибки наилучшего кода (т. е. кода с минимальной P_e).

Для того чтобы понять, почему этот подход является разумным, рассмотрим опять двоичный симметричный канал. Если выбрать для этого канала два кодовых слова длины N случайно, выбирая символы каждого слова независимо и принимающими значения 0 или 1 с равными вероятностями, то при больших N эти два кодовых слова будут с большой вероятностью отличаться приближенно в половине позиций блока. Из (5.3.12) следует, что если $x_{1,n} \neq x_{2,n}$, то $\min g_n(s) = 2\sqrt{\varepsilon(1-\varepsilon)}$. Если $x_{1,n} = x_{2,n}$, то $g_n(s) = 1$ при $0 \leqslant s \leqslant 1$. Следовательно, для двух кодовых слов, отличающихся в $N/2$ позициях, вероятность ошибки ограничена выражением

$$P_{e,m} \leqslant [2\sqrt{\varepsilon(1-\varepsilon)}]^{N/2}, \quad m = 1, 2. \tag{5.5.1}$$

Это не является средней вероятностью ошибки по ансамблю кодов; это просто вероятность ошибки для типичного кода из ансамбля. Это отличие в дальнейшем будет обсуждено более детально.

Показатель степени в (5.5.1) равен половине показателя для двух кодовых слов, отличающихся в каждой позиции, но P_e все еще стремится к 0 экспоненциально по N. В качестве компенсации за это уменьшение показателя степени появляется способ рассмотрения больших множеств кодовых слов, не требующий заботы о детальном выборе слов. Однако для больших множеств кодовых слов каждое кодовое слово не может отличаться от любого другого кодового слова во всех

позициях и, как будет показано, отмеченного уменьшения показателя не возникает.

Рассмотрим теперь произвольный дискретный канал, в котором $P_N(\mathbf{y}\,|\,\mathbf{x})$ является вероятностью получения последовательности $\mathbf{y}$ при условии, что была послана последовательность $\mathbf{x}$. Как следует из (5.3.4), для двух заданных кодовых слов $\mathbf{x}_1$ и $\mathbf{x}_2$ вероятность ошибки при передаче какого-либо слова ограничена следующим образом:

$$P_{e,m}(\mathbf{x}_1, \mathbf{x}_2) \leqslant \sum_{\mathbf{y}} P_N(\mathbf{y}\,|\,\mathbf{x}_1)^{1-s} P_N(\mathbf{y}\,|\,\mathbf{x}_2)^s,$$

$$m = 1,2; \quad s\text{—любое}, \quad 0 < s < 1. \tag{5.5.2}$$

Если рассмотреть теперь ансамбль кодов, в котором кодовые слова выбираются независимо в соответствии с распределением вероятности $Q_N(\mathbf{x})$, то вероятность кода с некоторыми частными кодовыми словами $\mathbf{x}_1$ и $\mathbf{x}_2$ равна $Q_N(\mathbf{x}_1)\,Q_N(\mathbf{x}_2)$. Следовательно, средняя вероятность ошибки по ансамблю имеет вид

$$\overline{P}_{e,m} = \sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2} Q_N(\mathbf{x}_1)\,Q_N(\mathbf{x}_2)\,P_{e,m}(\mathbf{x}_1, \mathbf{x}_2) \tag{5.5.3}$$

$$\leqslant \sum_{\mathbf{x}_1}\sum_{\mathbf{x}_2}\sum_{\mathbf{y}} Q_N(\mathbf{x}_1)\,Q_N(\mathbf{x}_2)\,P_N(\mathbf{y}\,|\,\mathbf{x}_1)^{1-s} P_N(\mathbf{y}\,|\,\mathbf{x}_2)^s. \tag{5.5.4}$$

$$\overline{P}_{e,m} \leqslant \sum_{\mathbf{y}} \Big[\sum_{\mathbf{x}_1} Q_N(\mathbf{x}_1)\,P_N(\mathbf{y}\,|\,\mathbf{x}_1)^{1-s}\Big]\Big[\sum_{\mathbf{x}_2} Q_N(\mathbf{x}_2)\,P_N(\mathbf{y}\,|\,\mathbf{x}_2)^s\Big],$$

$$m = 1,2; \quad s\text{—любое}, \quad 0 < s < 1. \tag{5.5.5}$$

Минимум (5.5.5) по s имеет место при $s = 1/2$. Для того чтобы показать это*), заметим, что $\mathbf{x}_1$ и $\mathbf{x}_2$ в (5.5.5) являются просто глухими индексами суммирования. Поэтому, если поменять местами s и $1 - s$, то функция не изменится и, следовательно, она является симметричной относительно $s = 1/2$. Так же в силу того, что (как уже было показано) правая часть (5.5.2) является выпуклой $\cup$ по s, то правая часть (5.5.5) также является выпуклой $\cup$ по s. Из симметрии и выпуклости следует, что минимум должен быть в точке $s = 1/2$ и поэтому

$$\overline{P}_{e,m} \leqslant \sum_{\mathbf{y}} \Big[\sum_{\mathbf{x}} Q_N(\mathbf{x})\sqrt{P_N(\mathbf{y}\,|\,\mathbf{x})}\Big]^2, \qquad m = 1,2. \tag{5.5.6}$$

Если канал является каналом без памяти, то

$$P_N(\mathbf{y}\,|\,\mathbf{x}) = \prod_{n=1}^{N} P(y_n\,|\,x_n),$$

В этом случае (5.5.6) можно упростить, если положить

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n),$$

---

*) Конечно, можно положить $s = 1/2$ в (5.5.5) независимо от того, минимизирует это выражение или нет. Таким образом, читатель может спокойно пренебречь этой и подобными последующими минимизациями по свободным параметрам, не беспокоясь о справедливости результата.

где $Q(k)$ — произвольное распределение вероятности для буквы. Другими словами, рассматривается ансамбль, в котором каждая буква любого кодового слова выбирается независимо с распределением вероятности $Q(k)$. При этом (5.5.6) можно представить в виде

$$\overline{P}_{e,m} \leqslant \sum_{y_1} \cdots \sum_{y_N} \left[\sum_{x_1} \cdots \sum_{x_N} \prod_{n=1}^{N} Q(x_n) \sqrt{P(y_n \mid x_n)} \right]^2. \qquad (5.5.7)$$

После раздельного суммирования по x_n каждого сомножителя в произведении [аналогично тому, как это было в (5.3.5)], получим

$$\overline{P}_{e,m} \leqslant \sum_{y_1} \cdots \sum_{y_N} \left[\prod_{n=1}^{N} \sum_{x_n} Q(x_n) \sqrt{P(y_n \mid x_n)} \right]^2. \qquad (5.5.8)$$

Изменяя порядок возведения в квадрат, умножения и суммирования по y_n точно так же, как это было сделано для x_n, получаем

$$\overline{P}_{e,m} \leqslant \prod_{n=1}^{N} \sum_{y_n} \left[\sum_{x_n} Q(x_n) \sqrt{P(y_n \mid x_n)} \right]^2. \qquad (5.5.9)$$

Так как в (5.5.9) суммирование по x_n производится по входному алфавиту $(0, 1, \ldots, K-1)$ и суммирование по y_n производится по выходному алфавиту $(0, \ldots, J-1)$, то это дает

$$\overline{P}_{e,m} \leqslant \left\{ \sum_{j=0}^{J-1} \left(\sum_{k=0}^{K-1} Q(k) \sqrt{P(j \mid k)} \right)^2 \right\}^N, \quad m = 1, 2. \qquad (5.5.10)$$

Это представляет собой границу сверху для средней вероятности ошибки по ансамблю кодов с двумя кодовыми словами длины N. Буквы кодовых слов выбираются независимо с вероятностями $Q(k)$, и канал является дискретным каналом без памяти с переходными вероятностями $P(j \mid k)$. В двоичном симметричном канале, положив $Q(0) = Q(1) = 1/2$, из (5.5.10) получаем

$$\overline{P}_{e,m} \leqslant \left\{ {}^1/_2 \left(\sqrt{\varepsilon} + \sqrt{1-\varepsilon} \right)^2 \right\}^N. \qquad (5.5.11)$$

Можно заметить, что граница в (5.5.11) отличается от границы в (5.5.1). На рис. 5.5.1 проиллюстрировано это отличие в зависимости от ε.

Причина этого отличия может быть понята с наибольшей ясностью в пределе при стремлении ε к 0. Вероятность ошибки для типичного кода с кодовыми словами, отличающимися в половине позиций, очевидно, стремится к нулю. Вместе с тем вероятность того, что в ансамбле кодов два выбранных кодовых слова будут совпадать, равна 2^{-N}; это дает границу для $\overline{P}_{e,m}$ в (5.5.11). Другими словами, при малых ε средняя вероятность ошибки $\overline{P}_{e,m}$ определяется не типичными кодами, а в высшей степени специфическими кодами, для которых $\overline{P}_{e,m}$ велика.

Приведенное рассуждение довольно просто, но оно часто используется в теории информации. Если требуется получить надежную передачу по каналу с шумами, то нужно сконцентрировать внимание

на специфических событиях, которые вызывают ошибки, а не на типичных событиях, которые не вызывают ошибок. К сожалению, этим важным принципом часто пренебрегают при построении моделей физических каналов связи.

Теперь мы почти подготовлены для отыскания верхней границы $\overline{P}_{e,m}$ для более чем двух кодовых слов, но вначале приведем другой вывод границы (5.5.4). Заметим, что $\overline{P}_{e,1}$ является средним по $\mathbf{x}_1$, $\mathbf{x}_2$ и $\mathbf{y}$. Когда сообщение 1 кодируется в $\mathbf{x}_1$, то $\mathbf{y}$ происходит с вероятностью $P_N(\mathbf{y}\,|\,\mathbf{x}_1)$, и будем считать, что ошибка возникает, если $P_N(\mathbf{y}\,|\,\mathbf{x}_2) \geqslant P_N(\mathbf{y}\,|\,\mathbf{x}_1)$. Таким образом, $\overline{P}_{e,1}$ можно записать в виде

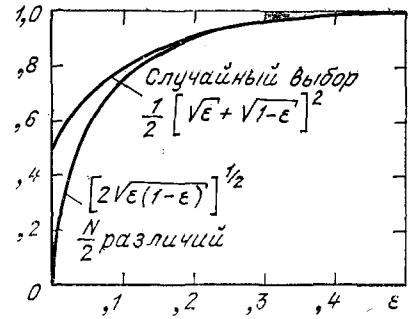


Рис. 5.5.1. Граница ошибочного декодирования для двух кодовых слов. Сравнение между случайным выбором и случаем, когда имеется отличие в $N/2$ позициях.

$$\overline{P}_{e,1} = \sum_{\mathbf{x}_1} Q_N(\mathbf{x}_1) \sum_{\mathbf{y}} P_N(\mathbf{y}\,|\,\mathbf{x}_1)\,\Pr[\text{ошибка}\,|\,m=1,\ \mathbf{x}_1,\ \mathbf{y}], \qquad (5.5.12)$$

где $\Pr[\text{ошибка}\,|\,m=1,\ \mathbf{x}_1,\ \mathbf{y}]$ — вероятность (по ансамблю выборов $\mathbf{x}_2$) того, что произойдет ошибка, при условии, что на кодер поступило сообщение 1, первым кодовым словом является $\mathbf{x}_1$ и было принято $\mathbf{y}$. Имеем

$$\Pr[\text{ошибка}\,|\,m=1,\ \mathbf{x}_1,\ \mathbf{y}] = \sum_{\mathbf{x}_2:\,P_N(\mathbf{y}\,|\,\mathbf{x}_2)\geqslant P_N(\mathbf{y}\,|\,\mathbf{x}_1)} Q_N(\mathbf{x}_2). \qquad (5.5.13)$$

При $P_N(\mathbf{y}\,|\,\mathbf{x}_1) > 0$ выражение (5.5.13) можно ограничить сверху, умножая каждое слагаемое на $[P_N(\mathbf{y}\,|\,\mathbf{x}_2)/P_N(\mathbf{y}\,|\,\mathbf{x}_1)]^s$ при любом $s > 0$. Далее, строя границу с помощью суммирования по всем $\mathbf{x}_2$, получаем

$$\Pr[\text{ошибка}\,|\,m=1,\ \mathbf{x}_1,\ \mathbf{y}] \leqslant \sum_{\mathbf{x}_2} Q_N(\mathbf{x}_2) \left[\frac{P_N(\mathbf{y}\,|\,\mathbf{x}_2)}{P_N(\mathbf{y}\,|\,\mathbf{x}_1)}\right]^s. \qquad (5.5.14)$$

Вместе с тем этот результат можно интерпретировать как границу Чернова для

$$\Pr\left\{\ln\frac{P_N(\mathbf{y}\,|\,\mathbf{x}_2)}{P_N(\mathbf{y}\,|\,\mathbf{x}_1)} \geqslant 0\right\},$$

где в качестве вероятностной меры используется $Q_N(\mathbf{x}_2)$.

Подставляя (5.5.14) в (5.5.12), получаем снова (5.5.4). Другое доказательство понадобилось здесь потому, что оно может быть легко обобщено на случай произвольного числа кодовых слов.

5.6. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КОДА С ЧИСЛОМ СЛОВ, БОЛЬШИМ ДВУХ

Теорема 5.6.1. Пусть $P_N(\mathbf{y}|\mathbf{x})$ — переходные вероятности для последовательностей длины $N \geqslant 1$ в дискретном канале. Пусть $Q_N(\mathbf{x})$ — произвольное распределение вероятности, заданное на входных последовательностях. Для данного числа $M \geqslant 2$ кодовых слов с длиной блока N рассмотрим ансамбль кодов, в котором слова выбираются независимо с вероятностной мерой $Q_N(\mathbf{x})$. Предположим, что на кодер поступает некоторое произвольное сообщение m, $1 \leqslant m \leqslant M$, и что используется декодирование по максимуму правдоподобия. Тогда средняя по этому ансамблю кодов вероятность ошибочного декодирования ограничена при любом выборе ρ, $0 \leqslant \rho \leqslant 1$, неравенством

$$\overline{P}_{e,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}\right]^{1+\rho}. \tag{5.6.1}$$

Для доказательства теоремы понадобится следующая простая лемма.

Лемма. Пусть $P(A_1)$, ..., $P(A_M)$ являются вероятностями событий A_1, ..., A_M и $P(\bigcup_m A_m)$ является вероятностью их объединения. При любом ρ, $0 < \rho \leqslant 1$, имеем

$$P\left(\bigcup_m A_m\right) \leqslant \left[\sum_{m=1}^{M} P(A_m)\right]^{\rho}. \tag{5.6.2}$$

Доказательство леммы. Имеем

$$P\left(\bigcup_m A_m\right) \leqslant \begin{cases} \sum_{m=1}^{M} P(A_m), & (5.6.3) \\ 1. & (5.6.4) \end{cases}$$

Неравенство (5.6.3) является обычной границей для вероятностей (см. задачу 2.16), а (5.6.4) является очевидной границей для вероятности. Если $\Sigma P(A_m)$ меньше чем 1, то $\Sigma P(A_m)$ увеличивается при возведении ее в степень ρ и (5.6.2) следует из (5.6.3). Обратно, если $\Sigma P(A_m) \geqslant 1$, то $[\Sigma P(A_m)]^{\rho} \geqslant 1$, так что (5.6.2) следует из (5.6.4). |

Доказательство теоремы. Имеем

$$\overline{P}_{e,m} = \sum_{\mathbf{x}_m} \sum_{\mathbf{y}} Q_N(\mathbf{x}_m) P_N(\mathbf{y}|\mathbf{x}_m) \Pr[\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}], \tag{5.6.5}$$

где $\Pr[\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}]$ — условная вероятность ошибочного декодирования при условии, во-первых, что на кодер поступило сообщение m, во-вторых, что заданная последовательность $\mathbf{x}_m$ была выбрана в качестве m-го кодового слова, и, в-третьих, что была принята последовательность $\mathbf{y}$. Суммирование производится соответственно по всем входным и всем выходным последовательностям канала длины N.

При заданных m, $\mathbf{x}_m$, $\mathbf{y}$ определим событие $A_{m'}$ для каждого $m' \neq m$ как событие, состоящее в том, что выбирается такое кодовое слово $\mathbf{x}_{m'}$, для которого

$$P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \geqslant P_N(\mathbf{y} \mid \mathbf{x}_m).$$

Теперь имеем

$$\Pr[\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}] \leqslant P\left(\bigcup_{m' \neq m} A_{m'}\right) \leqslant \tag{5.6.6}$$

$$\leqslant \left[\sum_{m' \neq m} P(A_{m'})\right]^{\rho}, \quad \rho - \text{любое}, \quad 0 < \rho \leqslant 1. \tag{5.6.7}$$

Неравенство (а не равенство) в (5.6.6) возникает потому, что декодер по максимуму правдоподобия не обязательно делает ошибку, если $P_N(\mathbf{y} \mid \mathbf{x}_{m'}) = P_N(\mathbf{y} \mid \mathbf{x}_m)$ при некотором m'.

Согласно определению $A_{m'}$ имеем

$$P(A_{m'}) = \sum_{\mathbf{x}_{m'}: P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \geqslant P_N(\mathbf{y} \mid \mathbf{x}_m)} Q_N(\mathbf{x}_{m'}) \leqslant$$

$$\leqslant \sum_{\mathbf{x}_{m'}} Q_N(\mathbf{x}_{m'}) \frac{P_N(\mathbf{y} \mid \mathbf{x}_{m'})^s}{P_N(\mathbf{y} \mid \mathbf{x}_m)^s}, \quad s - \text{любое}, \quad s > 0. \tag{5.6.8}$$

Так как $\mathbf{x}_{m'}$ является глухим переменным суммирования в (5.6.8), то индекс m' можно опустить и граница становится не зависящей от m'. В силу того, что существуют $M-1$ различных $m' \neq m$, то, подставляя (5.6.8) в (5.6.7), получаем

$$\Pr[\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}] \leqslant \left[(M-1) \sum_{\mathbf{x}} Q_N(\mathbf{x}) \frac{P_N(\mathbf{y} \mid \mathbf{x})^s}{P_N(\mathbf{y} \mid \mathbf{x}_m)^s}\right]^{\rho}. \tag{5.6.9}$$

Подставляя (5.6.9) в (5.6.5), будем иметь

$$\overline{P}_{e,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m)^{1-s\rho}\right] \times$$

$$\times \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^s\right]^{\rho}. \tag{5.6.10}$$

Отметим, что если $P_N(\mathbf{y} \mid \mathbf{x}_m) = 0$, то соответствующее слагаемое может быть опущено в сумме (5.6.5). Таким образом, $P_N(\mathbf{y} \mid \mathbf{x}_m)^{1-s\rho}$ можно положить равным нулю в (5.6.10), если $P_N(\mathbf{y} \mid \mathbf{x}_m) = 0$. Наконец, подставляя*) $s = 1/(1+\rho)$ в (5.6.10) и замечая, что $\mathbf{x}_m$ является глухим переменным суммирования, получаем (5.6.1) при $0 < \rho \leqslant 1$. Справедливость (5.6.1) в случае $\rho = 0$ следует из того, что правая часть (5.6.1) равна 1 при $\rho = 0$.|

Эта теорема обладает удивительной общностью и силой. Она применима как к каналам без памяти, так и к каналам с памятью и (как это будет показано в гл. 7) она может быть легко обобщена на недискретные каналы. Большая часть оставшегося в этой главе материала представляет собой следствия и интерпретации этой теоремы. Заметим,

*) Хотя это и не нужно для доказательства, этот выбор минимизирует (5.6.10) по s (см. задачу 5.6).

что отдельные технические детали доказательства являются очень простыми и не связаны с какими-либо предыдущими результатами. Однако доказательство этой теоремы в значительной степени связано с результатом, относящимся к двум кодовым словам. Единственным новым фактом, использованным в доказательстве, является лемма. Аддитивная граница (5.6.3) довольно точна для независимых событий, если получающаяся в результате граница мала по сравнению с 1, но является, очевидно, очень плохой, когда получающаяся в результате граница велика по сравнению с 1. Лемма дает способ уточнения аддитивной границы в последних случаях за счет первых случаев. Лемма никогда не дает границу более точную, чем наименьшая из границ (5.6.3) и (5.6.4), но, как это было использовано в теореме, она позволяет получить удобную для исследования границу для $\overline{P}_{e,\,m}$.

Используем теперь теорему 5.6.1 для случая дискретного канала без памяти, в котором

$$P_N(\mathbf{y} \mid \mathbf{x}) = \prod_n P(y_n \mid x_n).$$

Пусть $Q(k)$, $k = 0, 1, \ldots, K-1$, является некоторым произвольным распределением вероятностей на входном алфавите канала и пусть каждая буква кодовых слов выбирается независимо с этим распределением вероятностей, так что

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n),$$

$$\overline{P}_{e,\,m} \leqslant (M-1)^{\rho} \sum_{y_1} \ldots \sum_{y_N} \left\{ \sum_{x_1} \cdots \sum_{x_N} \prod_{n=1}^{N} Q(x_n) P(y_n \mid x_n)^{1/(1+\rho)} \right\}^{1+\rho} =$$

$$= (M-1)^{\rho} \prod_{n=1}^{N} \sum_{y_n} \left[\sum_{x_n} Q(x_n) P(y_n \mid x_n)^{1/(1+\rho)} \right]^{1+\rho} =$$

$$= (M-1)^{\rho} \left\{ \sum_{j=0}^{J-1} \left[\sum_{k=0}^{K-1} Q(k) P(j \mid k)^{1/(1+\rho)} \right]^{1+\rho} \right\}^{N}. \qquad (5.6.11)$$

При переходе от одного выражения к другому здесь были использованы такие же соображения, как и при переходе от (5.5.6) к (5.5.10).

Представим теперь эту границу таким образом, чтобы явно показать экспоненциальную зависимость границы от N при фиксированной скорости R. Напомним, что R по определению равна $(\ln M)/N$. Таким образом, $M = e^{NR}$ и при фиксированной скорости M экспоненциально зависит от N. К сожалению, при различных значениях N и при фиксированной R не обязательно получаются целочисленные значения e^{NR}, и это обстоятельство будет обойдено с помощью следующего определения. *При любом положительном целом значении N и любом положительном числе R (N, R)-блоковый код является кодом с $\lceil e^{NR} \rceil$ кодовыми словами длины N, где $\lceil e^{NR} \rceil$ обозначает наименьшее целое число, большее или равное e^{NR}.*

Рассматривая описанный выше ансамбль кодов как ансамбль (N, R)-блоковых кодов с $M - 1 < e^{NR} \leqslant M$, получаем

$$\overline{P}_{e,m} \leqslant e^{NR\rho} \left\{ \sum_j \left[\sum_k Q(k) P(j \mid k)^{1/(1+\rho)} \right]^{1+\rho} \right\}^N. \qquad (5.6.12)$$

Полученные результаты можно суммировать следующей теоремой, в которой также произведено дальнейшее преобразование (5.6.12).

Теорема 5.6.2. Пусть дискретный канал без памяти имеет переходные вероятности $P(j \mid k)$. При любом положительном целом значении N и положительном числе R рассмотрим ансамбль (N, R)-блоковых кодов, в котором буквы всех кодовых слов выбираются независимо с распределением вероятности $Q(k)$. Тогда для любого сообщения m, $1 \leqslant m \leqslant \lceil e^{NR} \rceil$, и любого ρ, $0 \leqslant \rho \leqslant 1$, средняя по ансамблю вероятность ошибочного декодирования при использовании декодирования по максимуму правдоподобия удовлетворяет неравенству

$$\overline{P}_{e,m} \leqslant \exp\{-N[E_0(\rho, \mathbf{Q}) - \rho R]\}, \qquad (5.6.13)$$

где

$$E_0(\rho, \mathbf{Q}) = -\ln \sum_{j=0}^{J-1} \left[\sum_{k=0}^{K-1} Q(k) P(j \mid k)^{1/(1+\rho)} \right]^{1+\rho}. \qquad (5.6.14)$$

В силу того, что (5.6.13) справедливо для любого сообщения в коде, средняя вероятность ошибки по сообщениям при произвольном наборе вероятностей сообщений $\Pr(m)$ удовлетворяет неравенству

$$\overline{P}_e = \sum_{m=1}^{M} \Pr(m) \overline{P}_{e,m} \leqslant \exp\{-N[E_0(\rho, \mathbf{Q}) - \rho R]\}. \qquad (5.6.15)$$

Наконец, так как ρ и $\mathbf{Q}$ являются произвольными в (5.6.13) и (5.6.14), наиболее точная граница получается после выбора ρ и $\mathbf{Q}$ так, чтобы максимизировать $E_0(\rho, \mathbf{Q}) - \rho R$. Таким образом, мы приходим к определению *показателя экспоненты случайного кодирования* $E_r(R)$:

$$E_r(R) = \max_{0 \leqslant \rho \leqslant 1} \max_{\mathbf{Q}} [E_0(\rho, \mathbf{Q}) - \rho R], \qquad (5.6.16)$$

где максимизация по $\mathbf{Q}$ производится по всем распределениям вероятностей $\mathbf{Q} = [Q(0), \ldots, Q(K-1)]$. Отсюда получаем следующее следствие.

Следствие 1. Для ансамбля кодов с $\mathbf{Q}$, которое максимизирует (5.6.16), имеем

$$\overline{P}_{e,m} \leqslant \exp[-NE_r(R)], \quad 1 \leqslant m \leqslant M, \qquad (5.6.17)$$

$$\overline{P}_e \leqslant \exp[-NE_r(R)]. \qquad (5.6.18)$$

На рис. 5.6.1 изображены показатели экспоненты случайного кодирования $E_r(R)$ для нескольких каналов. Как будет показано в следующем параграфе, $E_r(R) > 0$ при всех R, $0 \leqslant R < C$, где C — пропускная способность канала в натуральных единицах. Следова-

тельно, при подходящем выборе кодов вероятность ошибки можно сделать экспоненциально стремящейся к нулю с ростом длины блока при любой скорости, меньшей пропускной способности.

Так как средняя по ансамблю кодов вероятность ошибки удовлетворяет (5.6.18), то ясно, что по крайней мере один код из ансамбля должен иметь вероятность ошибки столь же малую, как и эта вероятность. Это следствие не дает способа отыскания такого кода и было бы удивительно, если бы случайно выбранный код имел вероятность ошибки гораздо большую, чем средняя вероятность. В частности, используя неравенство Чебышева (5.4.6), получаем

$$\Pr\left(P_e \geqslant \alpha \overline{P}_e\right) \leqslant \frac{1}{\alpha} \text{ при любом } \alpha > 1. \tag{5.6.19}$$

Этот результат очень важен при практическом использовании кодирования. Он говорит о том, что трудной является не проблема отыска-

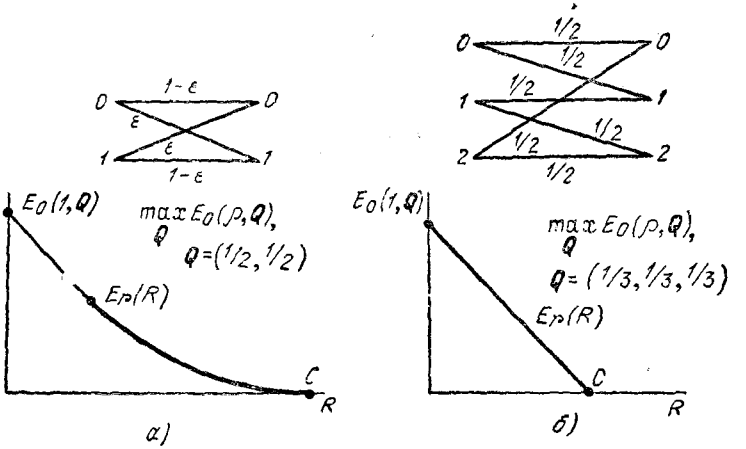


Рис. 5.6.1. Показатель экспоненты случайного кодирования $E_r(R)$ для двух каналов:

а — типичное поведение; *б* — частный случай, в котором $\partial^2 E_0/\partial\rho^2=0$.

ния хороших кодов с большой длиной блока, а проблема отыскания интересных для практики методов кодирования и декодирования таких кодов.

Приведенные рассмотрения дают возможность лучше понять поведение $P_e = \sum_m \Pr(m) P_{e,m}$ для случайно выбранного кода. К сожалению, вполне возможно (и в действительности высоко вероятно), что в таком случайно выбранном коде $P_{e,m}$ будет много больше, чем P_e для некоторых значений m и много меньше для других. Во многих системах передачи данных вероятности сообщений либо неизвестны, либо не имеют смысла. В таких ситуациях желательно обычно иметь код, для которого $P_{e,m}$ равномерно мала при всех m. Приводимое ниже следствие устанавливает существование таких кодов с помощью первоначального случайного выбора хорошего кода и последующего удаления всех слов, для которых $P_{e,m}$ слишком велико.

Следствие 2. Для любого дискретного канала без памяти, любого целого положительного N и любой положительной R существует (N, R)-блоковый код, для которого

$$P_{e,m} < 4\exp[-NE_r(R)], \text{ для всех } m,\ 1 \leqslant m \leqslant M = \lceil e^{NR} \rceil. \quad (5.6.20)$$

Доказательство. Выберем код с $2M$ кодовыми словами, для которого при равновероятных сообщениях

$$P_e = \frac{1}{2M}\sum_{m=1}^{2M} P_{e,m} \leqslant \exp\left[-NE_r\left(\frac{\ln 2M}{N}\right)\right], \quad (5.6.21)$$

Удалим M кодовых слов из этого кода, устраняя, в частности, все слова, для которых

$$P_{e,m} \geqslant 2\exp\left[-NE_r\left(\frac{\ln 2M}{N}\right)\right]. \quad (5.6.22)$$

Не может быть больше чем M слов, удовлетворяющих (5.6.22) потому, что если бы M таких слов существовали, то (5.6.21) не имело бы места. При использовании декодирования по максимуму правдоподобия декодирующие подмножества, соответствующие оставшимся словам, не могут потерять какой-либо элемент и, таким образом, для каждого оставшегося слова получаем

$$P_{e,m} < 2\exp\left[-NE_r\left(\frac{\ln 2M}{N}\right)\right].$$

Используя (5.6.16) совместно с этим результатом, будем иметь

$$P_{e,m} < 2\exp\left\{-N\left[\max_{0\leqslant\rho\leqslant 1}\left(\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) - \rho\frac{\ln M}{N} - \rho\frac{\ln 2}{N}\right)\right]\right\} \leqslant$$
$$\leqslant 2\exp\left\{-N\left[-\frac{\ln 2}{N} + \max_{0\leqslant\rho\leqslant 1}\left(\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) - \rho\frac{\ln M}{N}\right)\right]\right\} =$$
$$= 4\exp[-NE_r(R)].$$

Таким образом, это множество M кодовых слов удовлетворяет (5.6.20).

Свойства показателя экспоненты случайного кодирования $E_r(R)$

Для того чтобы понять поведение $E_r(R)$, нужно вначале исследовать $E_0(\rho, \mathbf{Q})$ как функцию ρ. На рис. 5.6.2 изображена E_0 в зависимости от ρ, и следующая теорема показывает, что график E_0 всегда имеет такой общий вид. Средняя взаимная информация

$$\mathcal{I}(\mathbf{Q}; \mathbf{P}) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} Q(k)P(j\,|\,k)\ln\frac{P(j\,|\,k)}{\sum_i Q(i)P(j\,|\,i)} \quad (5.6.23)$$

играет главную роль в описании поведения $E_0(\rho, \mathbf{Q})$. Здесь принято обозначение $\mathcal{I}(\mathbf{Q}; \mathbf{P})$, чтобы подчеркнуть, что эта информация рассматривается как математическая функция $\mathbf{Q}$ и переходных вероят-

ностей канала. $\mathcal{I}(\mathbf{Q};\mathbf{P})$ нельзя просто интерпретировать как среднюю по ансамблю кодов взаимную информацию на символ.

Теорема 5.6.3. Пусть распределение вероятности на входе $\mathbf{Q}$ и дискретный канал без памяти будут такими, что $\mathcal{I}(\mathbf{Q};\mathbf{P})>0$. Тогда $E_0(\rho,\mathbf{Q})$, определенная (5.6.14), имеет следующие свойства:

$$E_0(\rho,\mathbf{Q})\geqslant 0;\quad \rho\geqslant 0, \tag{5.6.24}$$

$$\mathcal{I}(\mathbf{Q};\mathbf{P})\geqslant\frac{\partial E_0(\rho,\mathbf{Q})}{\partial\rho}>0;\quad \rho\geqslant 0, \tag{5.6.25}$$

$$\frac{\partial^2 E_0(\rho,\mathbf{Q})}{\partial\rho^2}\leqslant 0;\quad \rho\geqslant 0. \tag{5.6.26}$$

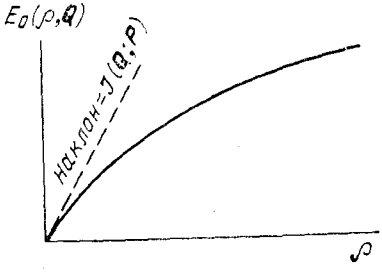


Рис. 5.6.2. График $E_0(\rho,\mathbf{Q})$.

Равенство в (5.6.24) имеет место тогда и только тогда, когда $\rho=0$; равенство в левой части (5.6.25) имеет место, когда $\rho=0$, и равенство в (5.6.26) имеет место тогда и только тогда, когда при всех j и k, таких, что $Q(k)P(j\mid k)>0$, имеем

$$\ln\frac{P(j\mid k)}{\sum\limits_i Q(i)P(j\mid i)}=\mathcal{I}(\mathbf{Q};\mathbf{P}), \tag{5.6.26а}$$

т. е., если случайная величина, представляющая собой взаимную информацию имеет нулевую дисперсию.

Можно заметить, рассматривая (5.6.14), что $E_0(0,\mathbf{Q})=0$, и легко проверить, выполняя дифференцирование, что

$$\left.\frac{\partial E_0(\rho,\mathbf{Q})}{\partial\rho}\right|_{\rho=0}=\mathcal{I}(\mathbf{Q};\mathbf{P}). \tag{5.6.27}$$

Доказательство оставшейся части теоремы содержится в приложении 5Б.

Используя эту теорему, легко максимизировать $E_0(\rho,\mathbf{Q})-\rho R$ по ρ при заданном $\mathbf{Q}$. Определим

$$E_r(R,\mathbf{Q})=\max_{0\leqslant\rho\leqslant 1}[E_0(\rho,\mathbf{Q})-\rho R]. \tag{5.6.28}$$

Уравнение для стационарной точки функции $E_0(\rho,\mathbf{Q})-\rho R$ от ρ имеет вид

$$\frac{\partial E_0(\rho,\mathbf{Q})}{\partial\rho}-R=0. \tag{5.6.29}$$

Так как $\partial^2 E_0(\rho,\mathbf{Q})/\partial\rho^2\leqslant 0$, то любое решение уравнения (5.6.29) в интервале $0\leqslant\rho\leqslant 1$ максимизирует (5.6.28). Более того, так как

$\partial E_0/\partial\rho$ является непрерывной и убывающей функцией от ρ, то решение уравнения (5.6.29) в интервале $0 \leqslant \rho \leqslant 1$ существует, если

$$\left.\frac{\partial E_0(\rho, \mathbf{Q})}{\partial\rho}\right|_{\rho=1} \leqslant R \leqslant \left.\frac{\partial E_0(\rho, \mathbf{Q})}{\partial\rho}\right|_{\rho=0} = \mathcal{I}(\mathbf{Q}; \mathbf{P}). \quad (5.6.30)$$

Значение $\partial E_0/\partial\rho|_{\rho=1}$ называется критической скоростью R_{cr} для заданного $\mathbf{Q}$.

При R, лежащей в указанном выше интервале, удобно использовать (5.6.29), чтобы связать R и $E_r(R, \mathbf{Q})$ параметрически через ρ. Получим

$$R = \partial E_0(\rho, \mathbf{Q})/\partial\rho; \quad 0 \leqslant \rho \leqslant 1 \quad (5.6.31)$$

$$E_r(R, \mathbf{Q}) = E_0(\rho, \mathbf{Q}) - \rho\partial E_0(\rho, \mathbf{Q})/\partial\rho.$$

Дифференцируя равенства (5.6.31), будем иметь $\partial R/\partial\rho = \partial^2 E_0/\partial\rho^2$ и $\partial E_r/\partial\rho = -\rho\partial^2 E_0/\partial\rho^2$. Следовательно, при изменении ρ от 0 до 1 значение R монотонно убывает от $\mathcal{I}(\mathbf{Q}; \mathbf{P})$ до $\partial E_0/\partial\rho|_{\rho=1}$, а $E_r(R, \mathbf{Q})$ монотонно возрастает от 0 до $E_0(1, \mathbf{Q}) - \partial E_0(\rho, \mathbf{Q})/\partial\rho|_{\rho=1}$. Взяв отношение производных, получим

$$\frac{\partial E_r(R, \mathbf{Q})}{\partial R} = -\rho. \quad (5.6.32)$$

Таким образом, параметр ρ можно интерпретировать как величину наклона $E_r(R, \mathbf{Q})$ к оси R.

При $R < \partial E_0/\partial\rho|_{\rho=1}$ значение $E_0(\rho, \mathbf{Q}) - \rho R$ достигает максимума (в интервале $0 \leqslant \rho \leqslant 1$) при $\rho = 1$, что дает

$$E_r(R, \mathbf{Q}) = E_0(1, \mathbf{Q}) - R. \quad (5.6.33)$$

И, наконец, в неинтересном случае, когда $R > \mathcal{I}(\mathbf{Q}; \mathbf{P})$, значение $E_0(\rho, \mathbf{Q}) - \rho R$ достигает максимума при $\rho = 0$, давая $E_r(R, \mathbf{Q}) = 0$.

Подытожим сказанное. При R из интервала, задаваемого (5.6.30), $E_r(R, \mathbf{Q})$ и R связаны (5.6.31). При меньших значениях R функции $E_r(R, \mathbf{Q})$ и R связаны линейным соотношением (5.6.33), а $E_r(R, \mathbf{Q}) = 0$ при бо́льших значениях R. В зависимости от R функция $E_r(R, \mathbf{Q})$ строго убывает и является положительной при всех $R < \mathcal{I}(\mathbf{Q}; \mathbf{P})$.

Рассмотрим теперь частный случай, в котором $\partial^2 E_0(\rho, \mathbf{Q})/\partial\rho^2 = 0$. Из равенства (5.6.26 а) теоремы 5.6.3 видно, что это соотношение должно удовлетворяться при всех $\rho \geqslant 0$, если оно удовлетворяется при каком-нибудь $\rho \geqslant 0$. В этом случае $\partial E_0(\rho, \mathbf{Q})/\partial\rho$ является постоянной и интервал, на котором справедливо (5.6.30), имеет нулевую протяженность (см. рис. 5.6.1, *б*). Этот частный случай является довольно искусственным и имеет место только для каналов без шума (для которых $H(X|Y) = 0$) и для некоторых других довольно специфических каналов, таких, как изображенный на рис. 5.6.1, *б*).

Для обычного случая, в котором $\partial^2 E_0(\rho, \mathbf{Q})/\partial\rho^2 < 0$, параметрические равенства (5.6.31) имеют силу на ненулевом интервале скоростей. Из (5.6.32) и (5.6.31) имеем $\partial^2 E_r(R, \mathbf{Q})/\partial R^2 = [\partial^2 E_0(\rho, \mathbf{Q})/\partial\rho^2]^{-1} > 0$ и, следовательно, функция $E_r(R, \mathbf{Q})$ является

строго выпуклой $\smile$ по R в этом диапазоне R. Заметим, что так как $\partial^2 E_r(R, \mathbf{Q})/\partial R^2 = 0$ вне этого диапазона, то функция $E_r(R, \mathbf{Q})$ является выпуклой $\smile$ по R при всех $R \geqslant 0$.

Показатель экспоненты случайного кодирования $E_r(R)$ можно теперь связать с $E_r(R, \mathbf{Q})$ следующим образом:

$$E_r(R) = \max_{\mathbf{Q}} E_r(R, \mathbf{Q}). \tag{5.6.34}$$

Этот максимум берется по множеству функций, которые являются выпуклыми $\smile$ и убывающими по R. Легко заметить (задача 4.12), что функция, на которой достигается максимум, также является выпуклой $\smile$ и убывающей по R. Кроме того, при $\mathbf{Q}$, на котором дости-

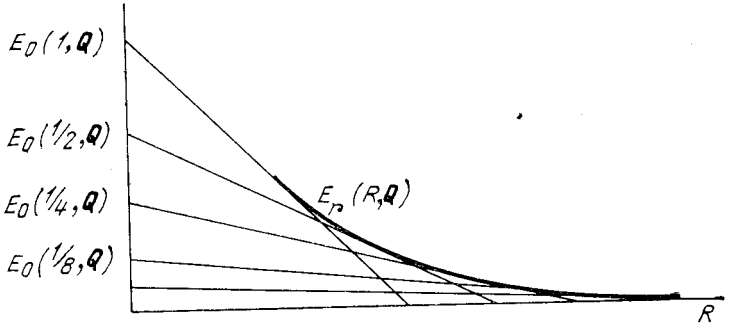


Рис. 5.6.3. Функция $E_r(R, \mathbf{Q})$ как огибающая сверху семейства линейных функций $E_0(\rho, \mathbf{Q}) - \rho R$ при ρ в качестве параметра.

гается пропускная способность канала, функция $E_r(R, \mathbf{Q})$ является положительной при $R < \mathcal{I}(\mathbf{Q}; \mathbf{P}) = C$ и, следовательно, $E_r(R)$ является положительной при $R < C$. Это доказывает следующую фундаментальную теорему.

Теорема 5.6.4. (Теорема кодирования для канала с шумами.) Для любого дискретного канала без памяти показатель экспоненты случайного кодирования $E_r(R)$ [см. (5.6.16) и (5.6.18)] является выпуклой $\smile$ убывающей положительной функцией R при $0 \leqslant R < C$.

Интересная геометрическая интерпретация процесса максимизации $E_0(\rho, \mathbf{Q}) - \rho R$ по ρ и $\mathbf{Q}$ может быть получена, если заметить, что при фиксированных ρ и $\mathbf{Q}$ функция $E_0(\rho, \mathbf{Q}) - \rho R$ является линейной от R и имеет наклон $-\rho$. Таким образом, $E_r(R, \mathbf{Q})$, как показано на рис. 5.6.3, является огибающей сверху семейства прямых линий, порожденных различными значениями ρ, $0 \leqslant \rho \leqslant 1$. Из этого построения видно, что $E_0(\rho, \mathbf{Q})$ является точкой пересечения оси ординат с касательной к кривой $E_r(R, \mathbf{Q})$, имеющей наклон $-\rho$. Из этого построения также непосредственно следует выпуклость $\smile$ функции $E_r(R, \mathbf{Q})$.

Для того чтобы максимизировать $E_0(\rho, \mathbf{Q}) - \rho R$ аналитически как по ρ, так и по $\mathbf{Q}$, целесообразно вначале провести максимизацию по $\mathbf{Q}$

$$E_r(R) = \max_{0 \leqslant \rho \leqslant 1} \left[-\rho R + \max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) \right]. \tag{5.6.35}$$

Функция $E_0(\rho, \mathbf{Q})$ не является выпуклой $\frown$ функцией $\mathbf{Q}$, однако, к счастью, она оказывается равной взятому со знаком минус логарифму выпуклой $\smile$ функции. Введем обозначение:

$$F(\rho, \mathbf{Q}) = \exp[-E_0(\rho, \mathbf{Q})] = \sum_{j=0}^{J-1} \left(\sum_{k=0}^{K-1} Q(k) P(j \mid k)^{1/(1+\rho)} \right)^{1+\rho}. \quad (5.6.36)$$

Значение $\mathbf{Q}$, на котором достигается минимум $F(\rho, \mathbf{Q})$, будет максимизировать $E_0(\rho, \mathbf{Q})$.

Теорема 5.6.5. При любом $\rho \geqslant 0$ функция $F(\rho, \mathbf{Q})$, заданная (5.6.36), является выпуклой $\smile$ функцией $\mathbf{Q}$ в области, где $\mathbf{Q}$ является вектором вероятностей. Необходимыми и достаточными условиями того, что на векторе вероятностей $\mathbf{Q}$ достигается минимум $F(\rho, \mathbf{Q})$ [или максимум $E_0(\rho, \mathbf{Q})$], являются условия

$$\sum_j P(j \mid k)^{1/(1+\rho)} \alpha_j(\mathbf{Q})^{\rho} \geqslant \sum_j \alpha_j(\mathbf{Q})^{1+\rho} \text{ при всех } k \quad (5.6.37)$$

с равенством при всех k, для которых $Q(k) > 0$. Функция $\alpha_j(\mathbf{Q})$ задается равенством

$$\alpha_j(\mathbf{Q}) = \sum_k Q(k) P(j \mid k)^{1/(1+\rho)}. \quad (5.6.38)$$

Доказательство. При $\rho \geqslant 0$ функция $\alpha_j(\mathbf{Q})^{1+\rho}$ является выпуклой $\smile$ функцией $\alpha_j(\mathbf{Q}) \geqslant 0$, так как ее вторая производная неотрицательна. В силу того, что $\alpha_j(\mathbf{Q})$ является линейной функцией $\mathbf{Q}$, из определения выпуклости следует, что $\alpha_j(\mathbf{Q})^{1+\rho}$ является выпуклой $\smile$ функцией $\mathbf{Q}$. Поэтому

$$F(\rho, \mathbf{Q}) = \sum_j \alpha_j(\mathbf{Q})^{1+\rho}$$

является выпуклой $\smile$ функцией $\mathbf{Q}$.

Согласно теореме 4.4.1 необходимыми и достаточными условиями того, что вектор вероятностей $\mathbf{Q}$ минимизирует $F(\rho, \mathbf{Q})$ являются условия

$$\frac{dF(\rho, \mathbf{Q})}{dQ(k)} \geqslant A \text{ для всех } k, \text{ с равенством, если } Q(k) > 0;\ A = \text{const}. \quad (5.6.39)$$

Отыскивая $\partial F/\partial Q(k)$ и производя деление на $(1 + \rho)$, получаем (5.6.37). Постоянная в правой части (5.6.37) вычисляется с помощью умножения каждого из соотношений на $Q(k)$ и суммированием по k.

Задача явного решения (5.6.37) и (5.6.38) для отыскания максимума $E_0(\rho, \mathbf{Q})$ почти эквивалентна задаче отыскания пропускной способности. Для некоторых каналов $\mathbf{Q}$, на котором достигается максимум, можно угадать и проверить с помощью (5.6.37). Для любого симметричного канала (определение см. в § 4.5.) легко проверить, что максимум $E_0(\rho, \mathbf{Q})$ достигается тогда, когда все $Q(k)$ равны друг другу. Далее, если число входов равно числу выходов, то иногда возможно решить (5.6.37) как систему линейных уравнений относительно $\alpha_j(\mathbf{Q})^{\rho}$

и затем решить (5.6.38) относительно $Q(k)$. Наконец, используя выпуклость $F(\rho, \mathbf{Q})$, легко найти максимум $E_0(\rho, \mathbf{Q})$ на вычислительной машине.

Так же как при отыскании пропускной способности, решение (5.6.37) и (5.6.38) относительно $\alpha_j(\mathbf{Q})$ является единственным, а решение относительно $Q(k)$ не обязательно единственно. Если входной алфавит объема K больше, чем выходной алфавит объема J, то всегда можно отыскать максимум $E_0(\rho, \mathbf{Q})$ лишь на J ненулевых значениях $Q(k)$. Единственным существенным отличием отыскания максимума $I(X; Y)$ от отыскания максимума $E_0(\rho, \mathbf{Q})$ является то, что выходные вероятности для пропускной способности всегда строго положительны, в то время как некоторые $\alpha_j(\mathbf{Q})$ могут быть нулевыми.

Задавая $\mathbf{Q}$, на котором достигается максимум $E_0(\rho, \mathbf{Q})$ при каждом фиксированном ρ, можно использовать геометрический метод, показанный на рис. 5.6.3, чтобы найти кривую $E_r(R)$. Для нахождения $E_r(R)$ можно также использовать равенства (5.6.31) и (5.6.33), взяв для каждого ρ значение $\mathbf{Q}$, на котором достигается максимум $E_0(\rho, \mathbf{Q})$. Чтобы показать, что эти равенства дают все точки на кривой $E_r(R)$, заметим, что для каждого R существуют некоторые ρ и $\mathbf{Q}$, такие, что $E_r(R) = E_0(\rho, \mathbf{Q}) - \rho R$. Для этого $\mathbf{Q}$ имеем $E_r(R) = E_r(R, \mathbf{Q})$. Но так как параметрические равенства определяют $E_r(R, \mathbf{Q})$ для заданных ρ и $\mathbf{Q}$, то они также дают $E_r(R)$. В примере 2 будет показано, однако, что эти равенства могут порождать некоторые дополнительные точки, лежащие строго ниже кривой $E_r(R)$.

Пример 1. Для двоичного симметричного канала, изображенного на рис. 5.3.1, *а*, $E_0(\rho, \mathbf{Q})$ достигает максимума по $\mathbf{Q}$ при $Q(0) = Q(1) = 1/2$. Для этого $\mathbf{Q}$ имеем

$$E_0(\rho, \mathbf{Q}) = \rho \ln 2 - (1+\rho) \ln \left[\varepsilon^{1/(1+\rho)} + (1-\varepsilon)^{1/(1+\rho)}\right]. \qquad (5.6.40)$$

Параметрические равенства (5.6.31) могут быть приведены к виду

$$R = \ln 2 - \mathcal{H}(\delta), \quad E_r(R) = T_\varepsilon(\delta) - \mathcal{H}(\delta), \qquad (5.6.41)$$

где параметр δ связан с параметром ρ, присутствующим в (5.6.31), соотношением

$$\delta = \frac{\varepsilon^{1/(1+\rho)}}{\varepsilon^{1/(1+\rho)} + (1-\varepsilon)^{1/(1+\rho)}}, \qquad (5.6.42)$$

а $\mathcal{H}(\delta)$ и $T_\varepsilon(\delta)$ задаются равенствами

$$\mathcal{H}(\delta) = -\delta \ln \delta - (1-\delta) \ln (1-\delta), \qquad (5.6.43)$$

$$T_\varepsilon(\delta) = -\delta \ln \varepsilon - (1-\delta) \ln (1-\varepsilon). \qquad (5.6.44)$$

Эти равенства справедливы лишь для δ из интервала $\varepsilon \leqslant \delta \leqslant \sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon})$. Для $R < \ln 2 - \mathcal{H}[\sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon})]$ можно использовать (5.6.33) вместе с (5.6.40), чтобы получить

$$E_r(R) = \ln 2 - 2 \ln \left(\sqrt{\varepsilon} + \sqrt{1-\varepsilon}\right) - R. \qquad (5.6.45)$$

Равенству (5.6.41) можно дать геометрическую интерпретацию, как показано на рис. 5.6.4. Можно заметить, что $T_\varepsilon(\delta)$ как функция δ является уравнением касательной к кривой $\mathcal{H}(\delta)$ с точкой касания $\delta = \varepsilon$.

Наиболее существенным в этом примере является то, что даже для такого простого канала невозможно дать другого выражения для $E_r(R)$ кроме параметрического.

Пример 2. Рассмотрим канал, матрица переходных вероятностей которого приведена на рис. 5.6.5. Можно заметить, что, если используются только первые четыре входа, то канал является симметричным,

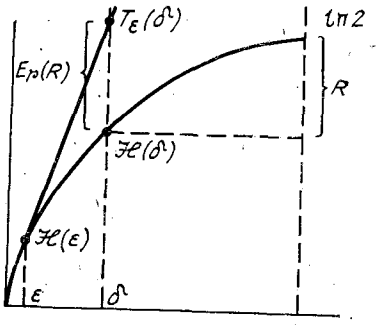


Рис. 5.6.4. Показатель экспоненты случайного кодирования для двоичного симметричного канала.

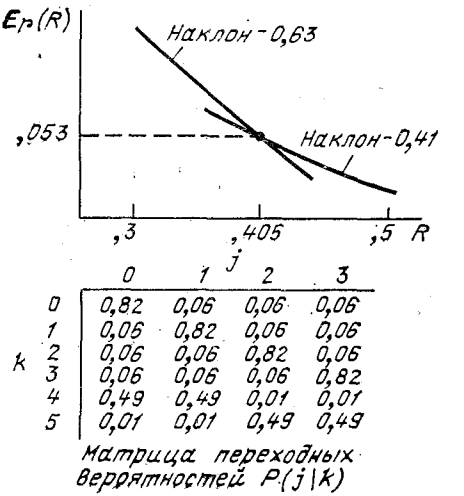


Рис. 5.6.5. Разрывность наклона $E_r(R)$.

а если используются только последние два входа, то канал сводится, по существу, к двоичному симметричному каналу. Можно догадаться, что при больших скоростях следует использовать только первые четыре входа, и что при малых скоростях следует использовать только последние два, в некотором смысле менее зашумленные входы. Проверяя эту гипотезу с помощью (5.6.37), находим, что при $\rho \leqslant 0{,}51$ функция $E_0(\rho, \mathbf{Q})$ принимает максимальное значение при $Q(0) = Q(1) = Q(2) = Q(3) = 1/4$, а при $\rho > 0{,}51$ функция $E_0(\rho, \mathbf{Q})$ достигает максимального значения при $Q(4) = Q(5) = 1/2$.

Подставляя эти значения для $\mathbf{Q}$ в (5.6.31), можно получить кривую $E_r(R)$, частично показанную на рис. 5.6.5. Имеется разрыв непрерывности наклона при переходе от $\rho = 0{,}41$ к $\rho = 0{,}63$, и для ρ из этого диапазона равенства (5.6.31) дают точки, как показано, лежащие строго ниже $E_r(R)$.

Пример 3 (каналы с очень большим шумом). Рассмотрим канал с очень большим шумом в том смысле, что вероятность получения заданного выхода почти не зависит от входа. Выведем приближенное

выражение для $E_r(R)$ для таких каналов, которое зависит только от пропускной способности. Пусть ω_j, $j = 0, \ldots, J-1$, являются набором вероятностей, определенных на выходах канала; и определим ε_{jk} с помощью равенства

$$P(j \mid k) = \omega_j (1 + \varepsilon_{jk}). \tag{5.6.46}$$

Предположим, что $|\varepsilon_{jk}| \ll 1$ при всех j и k, так что канал имеет очень большой шум в указанном выше смысле. Суммируя (5.6.46) по всем j, получаем

$$\sum_j \omega_j \varepsilon_{jk} = 0 \quad \text{при всех } k. \tag{5.6.47}$$

Найдем теперь $E_0(\rho, \mathbf{Q})$ для этого канала, разлагая E_0 в ряд по степеням ε_{jk} и отбрасывая все слагаемые порядка более высокого, чем 2. Имеем

$$E_0(\rho, \mathbf{Q}) = -\ln \sum_j \left[\sum_k Q(k)\, \omega_j^{1/(1+\rho)} (1+\varepsilon_{jk})^{1/(1+\rho)}\right]^{1+\rho}.$$

Выводя ω_j из-под внутренней суммы и разлагая $(1+\varepsilon_{jk})^{1/(1+\rho)}$, получаем

$$E_0(\rho, \mathbf{Q}) \approx -\ln \sum_j \omega_j \left\{\sum_k Q(k) \left[1 + \frac{\varepsilon_{jk}}{1+\rho} - \frac{\rho \varepsilon_{jk}^2}{2(1+\rho)^2}\right]\right\}^{1+\rho} \approx$$

$$\approx -\ln \sum_j \omega_j \left\{1 + \sum_k (1+\rho) Q(k) \left[\frac{\varepsilon_{jk}}{1+\rho} - \frac{\rho \varepsilon_{jk}^2}{2(1+\rho)^2}\right] + \right.$$

$$\left. + \frac{\rho(1+\rho)}{2} \left[\sum_k Q(k) \frac{\varepsilon_{jk}}{1+\rho}\right]^2\right\}.$$

Используя (5.6.47), это выражение приводим к виду

$$E_0(\rho, \mathbf{Q}) \approx -\ln \left\{1 - \frac{\rho}{2(1+\rho)} \sum_j \omega_j \left[\sum_k Q(k) \varepsilon_{jk}^2 - \left(\sum_k Q(k) \varepsilon_{jk}\right)^2\right]\right\} \approx$$

$$\approx \frac{\rho}{2(1+\rho)} \sum_j \omega_j \left[\sum_k Q(k) \varepsilon_{jk}^2 - \left(\sum_k Q(k) \varepsilon_{jk}\right)^2\right] = \frac{\rho}{1+\rho} f(\mathbf{Q}), \tag{5.6.48}$$

где

$$f(\mathbf{Q}) = \frac{1}{2} \sum_j \omega_j \left[\sum_k Q(k) \varepsilon_{jk}^2 - \left(\sum_k Q(k) \varepsilon_{jk}\right)^2\right]. \tag{5.6.49}$$

Параметрические равенства (5.6.31) будут иметь вид

$$R \approx \frac{f(\mathbf{Q})}{(1+\rho)^2}, \tag{5.6.50а}$$

$$E_r(R, \mathbf{Q}) \approx \frac{\rho^2 f(\mathbf{Q})}{(1+\rho)^2}. \tag{5.6.50б}$$

Средняя взаимная информация для входных вероятностей $\mathbf{Q}$ задается равенством (5.6.50 а) при $\rho = 0$. Следовательно, пропускная способность задается в виде

$$C \approx \max_{\mathbf{Q}} f(\mathbf{Q}). \tag{5.6.51}$$

Наконец, решая (5.6.50) относительно ρ и используя (5.6.51), получаем

$$E_r(R) \approx (\sqrt{C} - \sqrt{R})^2; \quad \frac{C}{4} \leqslant R \leqslant C. \tag{5.6.52}$$

Для $R < C/4$, объединяя (5.6.33), (5.6.48) и (5.6.51), будем иметь

$$E_r(R) \approx \frac{C}{2} - R; \quad 0 \leqslant R < \frac{C}{4}. \tag{5.6.53}$$

Это изображено на рис. 5.6.6.

Пример 4 (параллельные каналы). Пусть $P^*(j|k)$ и $P^{**}(l|i)$ будут переходными вероятностями двух дискретных каналов без памяти. Рассмотрим случай, когда эти каналы используются параллельно, т. е. каждую единицу времени передатчик посылает какой-либо символ k по первому каналу и какой-либо символ i по второму каналу. Будем считать, что каналы являются независимыми, т. е. что вероятность принять символ j в первом канале и символ l во втором канале при условии, что была послана пара (k, i), равна $P^*(j|k)P^{**}(l|i)$.

Эти параллельные каналы можно рассмотреть, как единый канал, входами которого являются пары (k, i), а выходами являются пары (j, l). Можно применить теорему кодирования к этому составному каналу, в котором последовательности пар на входе являются кодовыми словами. Считая, что $Q(k, i)$ — распределение вероятностей на входных парах, получаем

$$E_0(\rho, \mathbf{Q}) = -\ln \sum_{j,l} \left(\sum_{k,i} Q(k, i) [P^*(j|k) P^{**}(l|i)]^{1/(1+\rho)} \right)^{1+\rho}. \tag{5.6.54}$$

Если ограничиться такими $Q(k, i)$, для которых

$$Q(k, i) = Q^*(k) Q^{**}(i), \tag{5.6.55}$$

где $\mathbf{Q}^*$ и $\mathbf{Q}^{**}$ — произвольные распределения вероятностей на входе отдельных каналов, то $E_0(\rho, \mathbf{Q})$ упрощается следующим образом:

$$E_0(\rho, \mathbf{Q}) = -\ln \sum_{j,l} \left(\sum_k Q^*(k) P^*(j|k)^{1/(1+\rho)} \right)^{1+\rho} \times$$

$$\times \left(\sum_i Q^{**}(i) P^{**}(l|i)^{1/(1+\rho)} \right)^{1+\rho} = E_0^*(\rho, \mathbf{Q}^*) + E_0^{**}(\rho, \mathbf{Q}^{**}), \tag{5.6.56}$$

где

$$E_0^*(\rho, \mathbf{Q}^*) = -\ln \sum_j \left[\sum_k Q^*(k) P^*(j|k)^{1/(1+\rho)} \right]^{1+\rho}; \tag{5.6.57}$$

$$E_0^{**}(\rho, \mathbf{Q}^{**}) = -\ln \sum_l \left[\sum_i Q^{**}(i) P^{**}(l|i)^{1/(1+\rho)} \right]^{1+\rho}. \tag{5.6.58}$$

Таким образом, $E_0(\rho, \mathbf{Q})$ сводится к сумме функций E_0 для отдельных каналов.

Если выбрать вероятности $Q^*(k)$ так, чтобы на них достигался максимум $E_0^*(\rho, \mathbf{Q}^*)$ при заданном ρ, и выбрать вероятности $Q^{**}(i)$ так, чтобы на них достигался максимум $E_0^{**}(\rho, \mathbf{Q}^{**})$, то, как легко

следует из (5.6.37), максимум $E_0(\rho, \mathbf{Q})$ достигается при $Q(k, i) = Q^*(k)Q^{**}(i)$ и, таким образом,

$$\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) = \max_{\mathbf{Q}^*} E_0^*(\rho, \mathbf{Q}^*) + \max_{\mathbf{Q}^{**}} E_0^{**}(\rho, \mathbf{Q}^{**}). \qquad (5.6.59)$$

Этот результат имеет интересную геометрическую интерпретацию. Пусть $E_r(\rho)$ и $R(\rho)$ являются показателем экспоненты и скоростью для параллельных каналов, параметрически связанными соотношениями (5.6.31) при оптимальном $\mathbf{Q}$. Пусть $E_r^*(\rho)$, $R^*(\rho)$, $E_r^{**}(\rho)$ и $R^{**}(\rho)$ — аналогичные величины для отдельных каналов. Тогда

$$E_r(\rho) = E_r^*(\rho) + E_r^{**}(\rho), \qquad (5.6.60)$$

$$R(\rho) = R^*(\rho) + R^{**}(\rho). \qquad (5.6.61)$$

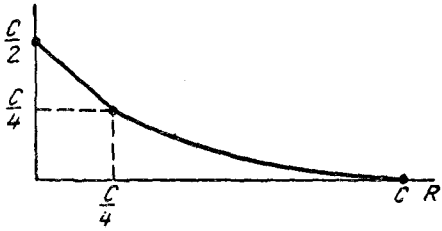


Рис. 5.6.6. Функция $E_r(R)$ для каналов с очень большим шумом.

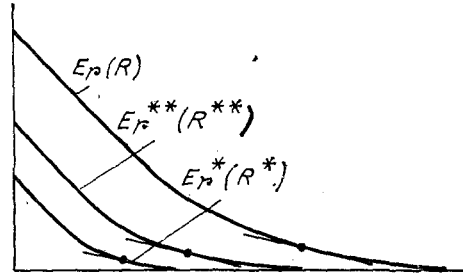


Рис. 5.6.7. Функция $E_r(R)$ для параллельных каналов.

Следовательно, пара E_r, R для системы из параллельных каналов образуется из соответствующих характеристик для отдельных каналов с помощью векторного сложения точек равного наклона кривых $E_r(\rho)$ и $R(\rho)$ (рис. 5.6.7).

5.7. ВЕРОЯТНОСТЬ ОШИБКИ ДЛЯ АНСАМБЛЯ КОДОВ С ВЫБРАСЫВАНИЕМ

В § 5.5 было показано (см. рис. 5.5.1), что средняя по ансамблю двух случайно выбранных кодовых слов вероятность ошибки сильно отличается от вероятности ошибки для двух типичных кодовых слов. Причина этого состоит в том, что хотя плохие коды в ансамбле являются маловероятными, они имеют такую большую вероятность ошибки, что определяют среднюю вероятность ошибки. То же самое явление оказывает неблагоприятное влияние на показатель экспоненты случайного кодирования при малых скоростях. Наиболее ясно это можно увидеть в двоичном симметричном канале, для которого согласно (5.6.45) при $\varepsilon \to 0$ верхняя граница вероятности ошибки стремится к $M(1/2)^N$. Это дает приближенное выражение для вероятности того, что при заданном переданном кодовом слове некоторое другое кодовое слово будет выбрано и отождествлено с ним.

В этом параграфе граница случайного кодирования будет усилена при малых скоростях с помощью выбрасывания плохих кодовых слов

из кодов, принадлежащих ансамблю. Этот подход аналогичен рассмотренному в следствии, устанавливающем границу (5.6.20). Вначале рассмотрим ансамбль, состоящий из кодов, каждый из которых содержит $M' = 2M - 1$ кодовых слов. Затем покажем, что по крайней мере в одном коде из ансамбля, имеются не меньше чем M слов, для которых $P_{e,\,m}$ удовлетворяет заданной границе. Эта граница будет выражена через произвольный параметр $s > 0$, который в дальнейшем может быть оптимизирован. На ансамбле кодов вероятность $P^s_{e,\,m}$ при каждом m, $1 \leqslant m \leqslant M'$, является случайной величиной. Применяя к этой случайной величине неравенство Чебышева (5.4.6), получаем

$$\Pr\left[P^s_{e,m} \geqslant 2\,\overline{P^s_{e,m}}\right] \leqslant 1/2. \tag{5.7.1}$$

Лемма. Для любого $s > 0$ в ансамбле кодов с $M' = 2M - 1$ кодовыми словами существует по крайней мере один код, в котором для по меньшей мере M кодовых слов выполняется неравенство

$$P_{e,m} < 2^{1/s}\,\overline{P^s_{e,\,m}}^{\,1/s}. \tag{5.7.2}$$

Доказательство. Пусть φ_m при каждом m является случайной величиной на ансамбле кодов. Пусть $\varphi_m = 1$ для кодов, для которых справедливо (5.7.2) и пусть $\varphi_m = 0$ во всех остальных случаях. Согласно (5.7.1) вероятность того, что (5.7.2) будет справедливо, не меньше 1/2 и, следовательно, $\overline{\varphi_m} \geqslant 1/2$. Число кодовых слов в случайно выбранном коде, которые удовлетворяют (5.7.2), является случайной величиной

$$\sum_{m=1}^{M'} \varphi_m.$$

Математическое ожидание числа слов, которые удовлетворяют (5.7.2), равно, следовательно,

$$\sum_{m=1}^{M'} \overline{\varphi}_m \geqslant \frac{M'}{2}.$$

Отсюда вытекает, что существует по крайней мере один код, для которого $\Sigma\varphi_m \geqslant M'/2$, и для такого кода $\Sigma\varphi_m \geqslant M$. |

Если все слова, кроме M слов, удовлетворяющих (5.7.2), будут выброшены из кода, рассмотренного в лемме, то области декодирования остающихся кодовых слов не могут быть уменьшены и, таким образом, мы получаем код с M кодовыми словами, удовлетворяющими (5.7.2). Теперь нужно найти удобную верхнюю границу для $\overline{P^s_{e,\,m}}$. Для частного кода с кодовыми словами $\mathbf{x}_1, \ldots, \mathbf{x}_{M'}$ имеем

$$P_{e,\,m} \leqslant \sum_{m' \neq m} \sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y} \mid \mathbf{x}_m)\, P_N(\mathbf{y} \mid \mathbf{x}_{m'})}. \tag{5.7.3}$$

Для того чтобы показать справедливость (5.7.3), заметим, что согласно (5.3.4) для заданного кода

$$\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y} \mid \mathbf{x}_m)\, P_N(\mathbf{y} \mid \mathbf{x}_{m'})}$$

является верхней границей вероятности того, что $P_N(\mathbf{y}\,|\,\mathbf{x}_{m'}) \geqslant$ $\geqslant P_N(\mathbf{y}\,|\,\mathbf{x}_m)$ при условии, что была передана $\mathbf{x}_m$. Так как $P_{e,\,m}$ является вероятностью объединения этих событий при $m' \neq m$, то $P_{e,\,m}$ можно оценить так, как указано в (5.7.3).

Рассмотрим теперь s из интервала $0 < s \leqslant 1$. Используя известное неравенство $(\Sigma a_i)^s \leqslant \Sigma a_i^s$ (см. задачу 4.15 (е)), будем иметь

$$P_{e,\,m}^s \leqslant \sum_{m' \neq m} \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_m)\,P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}\right]^s, \quad 0 < s \leqslant 1. \tag{5.7.4}$$

Рассмотрим ансамбль кодов, в котором кодовые слова выбираются независимо с распределением вероятности $Q_N(\mathbf{x})$. Имеем

$$\overline{P_{e,\,m}^s} \leqslant \sum_{m' \neq m} \left\{\sum_{\mathbf{x}_m} \sum_{\mathbf{x}_{m'}} Q_N(\mathbf{x}_m)\,Q_N(\mathbf{x}_{m'}) \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x}_m)\,P_N(\mathbf{y}\,|\,\mathbf{x}_{m'})}\right]^s\right\}. \tag{5.7.5}$$

В силу того, что $\mathbf{x}_{m'}$ является глухим переменным суммирования в (5.7.5), слагаемое в фигурных скобках не зависит от m' и мы имеем $M' - 1 = 2\,(M - 1)$ одинаковых слагаемых. Подставляя упростившееся таким образом выражение (5.7.5) в (5.7.2), получаем

$$P_{e,\,m} < 2^{1/s} \left\{2\,(M-1) \sum_{\mathbf{x}} \sum_{\mathbf{x}'} Q_N(\mathbf{x})\,Q_N(\mathbf{x}') \times \right.$$
$$\left. \times \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x})\,P_N(\mathbf{y}\,|\,\mathbf{x}')}\right]^s\right\}^{1/s}. \tag{5.7.6}$$

Эта граница по своему виду подобна границе, установленной в теореме 5.6.1. Это сходство может быть выявлено, если положить $\rho = 1/s$. Так как s является произвольным параметром в (5.7.6), $0 < s \leqslant 1$, то ρ является произвольным параметром, $\rho \geqslant 1$. Имеем

$$P_{e,\,m} < [4\,(M-1)]^{\rho} \times$$
$$\times \left\{\sum_{\mathbf{x}} \sum_{\mathbf{x}'} Q_N(\mathbf{x})\,Q_N(\mathbf{x}') \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\,|\,\mathbf{x})\,P_N(\mathbf{y}\,|\,\mathbf{x}')}\right]^{1/\rho}\right\}^{\rho}, \quad \rho \geqslant 1. \tag{5.7.7}$$

Неравенство (5.7.7) справедливо для любого дискретного канала как с памятью, так и без памяти, для которого можно определить $P_N(\mathbf{y}\,|\,\mathbf{x})$. Применим теперь эту границу к дискретному каналу без памяти, для которого

$$P_N(\mathbf{y}\,|\,\mathbf{x}) = \prod_n P(y_n\,|\,x_n).$$

Положим также $Q_N(\mathbf{x})$ равным произведению распределений, т. е.

$$Q_N(\mathbf{x}) = \prod_n Q(x_n).$$

После подстановки этого произведения в (5.7.7) и после преобразований, которые аналогичны проведенным при переходе от (5.5.6) к (5.5.10), получим

$$P_{e,\,m} < [4\,(M-1)]^{\rho} \prod_{n=1}^{N} \left\{\sum_{x_n} \sum_{x_n'} Q(x_n)\,Q(x_n') \times \right.$$

$$\times\left[\sum_{y_n}\sqrt{P(y_n\mid x_n)\,P(y_n\mid x_n')}\right]^{1/\rho}\Bigg\}^{\rho}=$$

$$=[4(M-1)]^{\rho}\left\{\sum_{k=0}^{K-1}\sum_{i=0}^{K-1}Q(k)\,Q(i)\left[\sum_{j=0}^{J-1}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}\right\}^{\rho N}. \quad (5.7.8)$$

В (N, R)-коде имеются $(M-1)<e^{NR}\leqslant M$ кодовых слов, поэтому (5.7.8) приводится к виду

$$P_{e,\,m}<e^{N\rho[R+(\ln 4)/N]}\left\{\sum_{k,\,i}Q(k)\,Q(i)\left[\sum_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}\right\}^{\rho N}. \quad (5.7.9)$$

Полученные результаты можно подытожить следующей теоремой.

Теорема 5.7.1. Пусть задан произвольный дискретный канал без памяти, пусть N — любое положительное целое число и R — какое-либо положительное число. Тогда существуют (N, R)-коды, для которых при всех m, $1\leqslant m\leqslant\lceil e^{NR}\rceil$,

$$P_{e,\,m}\leqslant\exp\left[-NE_{ex}\left(R+\frac{\ln 4}{N}\right)\right], \quad (5.7.10)$$

где функция E_{ex} задается равенствами

$$E_{ex}(R')=\sup_{\rho\geqslant 1}[-\rho R'+\max_{\mathbf{Q}}E_x(\rho,\mathbf{Q})], \quad (5.7.11)$$

$$E_x(\rho,\mathbf{Q})=-\rho\ln\sum_{k,i}Q(k)\,Q(i)\left[\sum_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho} \quad (5.7.12)$$

и максимум по $\mathbf{Q}$ в (5.7.11) берется по всем распределениям вероятностей для букв на входе канала.

Исследование свойств функции $E_{ex}(R')$, которая называется показателем экспоненты для процедуры с выбрасыванием, проводится почти так же, как для показателя экспоненты случайного кодирования. Ее свойства зависят от $E_x(\rho, \mathbf{Q})$ таким же образом, как свойства $E_r(R)$ зависели от $E_0(\rho, \mathbf{Q})$.

Теорема 5.7.2. Пусть $\mathbf{Q}$ — распределение вероятности на входе дискретного канала без памяти с переходными вероятностями $P(j\mid k)$; будем считать, что $\mathcal{I}(\mathbf{Q};\mathbf{P})>0$ [см. (5.6.23)]. Тогда при любом $\rho>0$ $E_x(\rho, \mathbf{Q})$ является строго возрастающей и выпуклой $\cap$ функцией ρ. Выпуклость будет строгой, за исключением случая, когда канал является каналом без шума, в том смысле, что для любой пары (i, k) используемых входов (т. е. таких входов, для которых $Q(i)>0$, $Q(k)>0$) справедливо либо $P(j\mid k)\,P(j\mid i)=0$ для всех j, либо $P(j\mid k)=P(j\mid i)$ для всех j.

Эта теорема доказана в приложении 5Б. Заметим, что согласно (5.7.11) $E_{ex}(R)$ можно понимать как верхнюю грань значений множества линейных функций R: $-\rho R+\max\limits_{\mathbf{Q}}E_x(\rho, \mathbf{Q})$

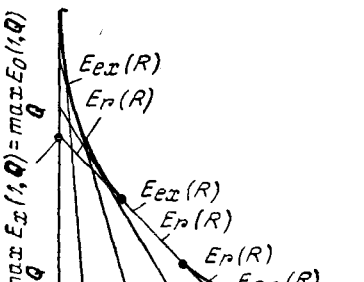


Рис. 5.7.1. Сравнение $E_{ex}(R)$ и $E_r(R)$.

с наклоном $-\rho$ при каждом $\rho \geqslant 1$. Это изображено на рис. 5.7.1. Заметим, что при $\rho = 1$ можно изменить порядок суммирования по j с суммированием по i и k в определении $E_x(\rho, \mathbf{Q})$ и после этого можно заметить, что $E_x(1, \mathbf{Q}) = E_0(1, \mathbf{Q})$. Это показывает связь между $E_{ex}(R)$ и $E_r(R)$, проиллюстрированную на рис. 5.7.1. Поскольку $E_x(\rho, \mathbf{Q})$— строго возрастающая функция ρ, то указанная геометрическая интерпретация доказывает, что (в общем случае) $E_{ex}(R) > E_r(R)$ для достаточно малых R. Однако в пределе при переходе к каналу с очень большим шумом это различие становится сколь угодно малым (см. задачу 5.31).

Заметим, что в любой области, где $E_{ex}(R) = E_r(R)$, имеем $\rho = 1$ и поэтому

$$P_{e,m} \leqslant \exp\left\{-N\left[E_{ex}\left(R + \frac{\ln 4}{N}\right)\right]\right\} = \\ = 4\exp\left\{-N\left[-R + \max_{\mathbf{Q}} E_x(1, \mathbf{Q})\right]\right\}, \tag{5.7.13}$$

что в точности совпадает с равномерной границей для $P_{e,m}$, заданной (5.6.20). Граница для процедуры с выбрасыванием (5.7.10), таким образом, строго точнее, чем граница случайного кодирования в области, где $[R + (\ln 4)/N]$ строго меньше, чем наименьшая R, для которой $E_{ex}(R) = E_r(R)$.

Может случиться, что для достаточно малой R функция $E_{ex}(R)$ принимает бесконечные значения. Для того чтобы исследовать этот случай, заметим, что координата точки пересечения оси R линейной функцией $-\rho R + E_x(\rho, \mathbf{Q})$ равна $E_x(\rho, \mathbf{Q})/\rho$. При $\rho \to \infty$ функция $-\rho R + E_x(\rho, \mathbf{Q})$ имеет вертикальную асимптоту

$$R = \lim_{\rho \to \infty} E_x(\rho, \mathbf{Q})/\rho$$

и $E_{ex}(R)$ принимает бесконечные значения при

$$R < \lim_{\rho \to \infty} E_x(\rho, \mathbf{Q})/\rho.$$

(Другими словами, для таких R функция $-\rho R + E_x(\rho, \mathbf{Q})$ стремится к ∞ при $\rho \to \infty$.) Раскрывая этот предел по правилу Лопиталя, получаем

$$\lim_{\rho \to \infty} \frac{E_x(\rho, \mathbf{Q})}{\rho} = -\ln\left[\sum_{k,i} Q(k) Q(i) \varphi_{k,i}\right], \tag{5.7.14}$$

$$\varphi_{k,i} = \begin{cases} 1 & \text{при } \sum_j \sqrt{P(j|k) P(j|i)} \neq 0, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{5.7.15}$$

Пусть $R_{x,\infty}$ обозначает максимальное значение (5.7.14) по $\mathbf{Q}$, т. е.

$$R_{x,\infty} = \max_{\mathbf{Q}} -\ln\left[\sum_{k,i} Q(k)\,Q(i)\,\varphi_{k,i}\right]. \qquad (5.7.16)$$

Из сказанного следует, что $E_{ex}(R) = \infty$ при $R < R_{x,\infty}$.

Из (5.7.14) можно увидеть, что $R_{x,\infty} = 0$, если $\varphi_{k,i} = 1$ при всех k и i, и $R_{x,\infty} > 0$ в других случаях. Если $\varphi_{k,i} = 0$ для некоторых k и i, то эти два входа не будут никогда перепутаны на приемном конце и по крайней мере один бит информации можно передать при одном использовании канала без каких-либо ошибок, используя лишь эти два входа. Если положить $Q(k) = Q(i) = 1/2$ для этой пары

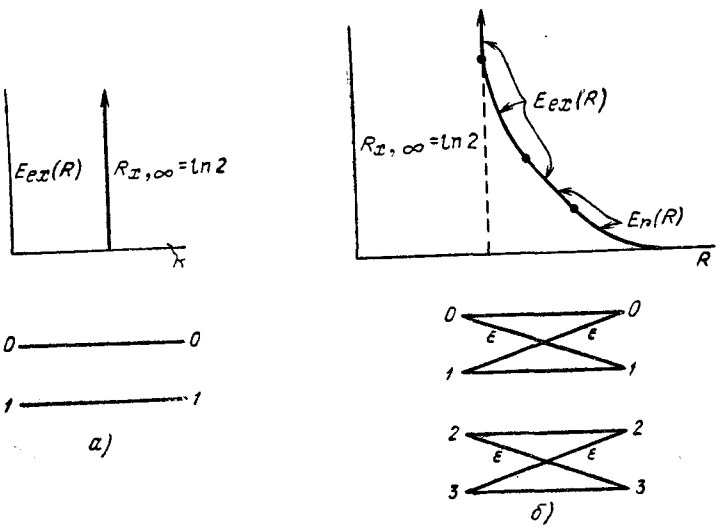


Рис. 5.7.2. Функция $E_{ex}(R)$ для каналов с $R_{x,\infty} > 0$.

входов, то правая часть (5.7.14) будет равна ln 2. Отсюда следует, что если $R_{x,\infty} \neq 0$, то $R_{x,\infty}$ должна быть по крайней мере равной ln 2. Шеннон (1956) определил пропускную способность канала с нулевой ошибкой как наибольшую скорость, с которой данные могут быть переданы по каналу с нулевой вероятностью ошибки (в противоположность к сколь угодно малой вероятности ошибки). Так как $P_{e,m} = 0$ при $R < R_{x,\infty}$, то получаем, что $R_{x,\infty}$ является нижней границей для пропускной способности канала с нулевой ошибкой. На рис. 5.7.2 изображено поведение $E_{ex}(R)$ для двух каналов с $R_{x,\infty} > 0$. В задаче 5.25 выведено простое выражение для $R_{x,\infty}$.

Нахождение максимума $E_{ex}(R)$ по ρ в значительной мере похоже на аналогичную процедуру для показателя экспоненты случайного кодирования. Если ввести обозначение

$$E_{ex}(R, \mathbf{Q}) = \sup_{\rho \geqslant 1}[-\rho R' + E_x(\rho, \mathbf{Q})].$$

то получим параметрические соотношения

$$E_{ex}(R, \mathbf{Q}) = -\rho \frac{\partial E_x(\rho, \mathbf{Q})}{\partial \rho} + E_x(\rho, \mathbf{Q}), \tag{5.7.17}$$

$$R = \frac{\partial E_x(\rho, \mathbf{Q})}{\partial \rho}, \tag{5.7.18}$$

справедливые при

$$R_{x,\infty} < R \leqslant \left.\frac{\partial E_x(\rho, \mathbf{Q})}{\partial \rho}\right|_{\rho=1}. \tag{5.7.19}$$

При больших скоростях R эта граница эквивалента прямолинейному участку границы случайного кодирования, которая была рассмотрена ранее. Для типичного канала $R_{x,\infty} = 0$ и, как показано в задаче 5.24, имеем

$$\lim_{R \to 0} E_{ex}(R, \mathbf{Q}) = -\sum_{k,i} Q(k) Q(i) \ln \left[\sum_j \sqrt{P(j|k) P(j|i)}\right]. \tag{5.7.20}$$

В настоящее время сравнительно немного известно о методах отыскания максимума этой границы по $\mathbf{Q}$. Функция $E_x(\rho, \mathbf{Q})$ не является выпуклой $\cap$ по $\mathbf{Q}$ и может иметь несколько локальных максимумов по $\mathbf{Q}$. И, что более удивительно, если попытаться оптимизировать выражение (5.7.7) для $P_{e,m}$ по $Q_N(\mathbf{x})$, не ограничиваясь произведением распределений, то иногда могут возникнуть случаи, в которых произведение распределений не оптимизирует границу (см. задачу 5.26). Джелинек (1968), однако, показал, что в случае произвольного дискретного канала с двоичным входом произведение распределений всегда оптимизирует границу и, на самом деле, оптимальное значение $\mathbf{Q}$ равно $Q(0) = Q(1) = 1/2$ (см. задачи 5.29 и и 5.30).

5.8. НИЖНИЕ ГРАНИЦЫ ДЛЯ ВЕРОЯТНОСТИ ОШИБКИ

В предыдущих параграфах были найдены верхние границы вероятности ошибочного декодирования, которые могут быть достигнуты в дискретном канале без памяти; эти границы были выражены через длину блока N и скорость R. Этот параграф посвящен отысканию минимальной вероятности ошибки, которая может быть достигнута на каком-либо (N, R)-коде. При изложении этого параграфа будем считать, что все кодовые слова являются равновероятными. В действительности без какого-либо такого предположения не существует ненулевая нижняя граница вероятности ошибки, так как если одно из кодовых слов посылается с вероятностью 1, то декодер может всегда декодировать сообщение, соответствующее этому кодовому слову и ошибки не будут происходить*).

*) Можно, однако, построить нижнюю границу вероятности ошибки для наихудшего кодового слова в коде, т. е. для $\max P_{e,m}$, не рассматривая вероятностей кодовых слов (см. задачу 5.32).

Для равновероятных сообщений вероятность ошибочного декодирования кода с M сообщениями равна

$$P_e = \frac{1}{M} \sum_{m=1}^{M} P_{e,\,m},$$

где $P_{e,\,m}$ — вероятность ошибки при условии, что было передано сообщение m. В последнем параграфе было показано, что при любой длине блока N и любой скорости $R > 0$ существуют (N, R)-блоковые коды, для которых одновременно $P_e \leqslant \exp[-NE_r(R)]$ и $P_e \leqslant \exp[-NE_{ex}(R + \ln 4/N)]$.

Вывод известных нижних границ для P_e при заданных N и R значительно тоньше и сложнее, чем вывод верхних границ. Поэтому мы только сформулируем здесь относящиеся сюда теоремы. Доказательства могут быть найдены у Шеннона, Галлагера и Берлекэмпа (1967). Доказательства для частного случая двоичного симметричного канала (ДСК) будут представлены здесь. Большая часть идей, используемых при отыскании нижних границ для P_e, проявляется в этом частном случае, но при этом удается избежать многих деталей.

Теорема 5.8.1. (Граница сферической упаковки). Для любого (N, R)-кода в дискретном канале без памяти

$$P_e \geqslant \exp(-N\{E_{sp}[R - o_1(N)] + o_2(N)\}). \tag{5.8.1}$$

где

$$E_{sp}(R) = \sup_{\rho > 0} \left[\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) - \rho R\right] \tag{5.8.2}$$

и $E_0(\rho, \mathbf{Q})$ задается равенством (5.6.14). Величины $o_1(N)$ и $o_2(N)$ стремятся к нулю с ростом N и могут быть взяты в виде

$$o_1(N) = \frac{\ln 8}{N} + \frac{K \ln N}{N}, \tag{5.8.3}$$

$$o_2(N) = \frac{\ln 8}{N} + \sqrt{\frac{2}{N} \ln \frac{e^2}{P_{min}}}, \tag{5.8.4}$$

где P_{min} является наименьшей ненулевой переходной вероятностью канала, а K — объем входного алфавита.

Как будет показано, функция $E_{sp}(R)$, называемая показателем экспоненты сферической упаковки, определяется почти так же, как показатель экспоненты случайного кодирования $E_r(R)$, и отличается только интервалом значений ρ, по которому производится максимизация. Следствием этого является то, что результаты, полученные в § 5.6, непосредственно применимы к $E_{sp}(R)$. В частности, $E_{sp}(R)$ будет положительной при $0 < R < C$, убывающей с ростом R и выпуклой $\cup$ функцией. На рис. 5.8.1 $E_{sp}(R)$ изображена для ряда каналов вместе с $E_r(R)$ для сравнения. На рис. 5.8.1, *а* показано типичное поведение этих функций, а на других рисунках изображены в некотором смысле вырожденные случаи; на них $E_{sp}(R)$ обращается в бесконечность при всех скоростях,

меньших, чем заданная постоянная, обозначаемая через R_∞. Для того чтобы найти эту постоянную, представим себе

$$\left[\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) - \rho R\right]$$

как множество линейных функций R с $\rho > 0$ в качестве параметра (см. рис. 5.6.3). Координата точки пересечения оси R с указанной выше функцией при заданном ρ равна

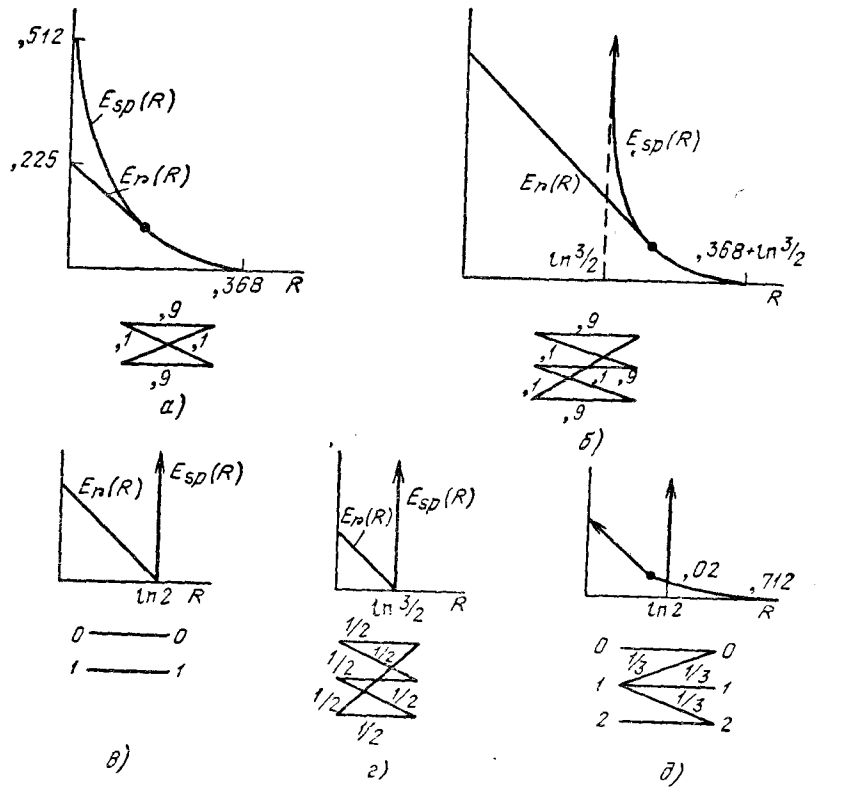


Рис. 5.8.1. Сравнение показателей экспонент сферической упаковки и случайного кодирования.

$$\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})/\rho.$$

При $\rho \to \infty$ наклоны этих прямых линий стремятся к бесконечности и так как $E_{sp}(R)$ является выпуклой оболочкой этих функций, то R_∞ задается как предел концов отрезков, отсекаемых на оси R при $\rho \to \infty$, т. е.

$$R_\infty = \lim_{\rho \to \infty} \max_{\mathbf{Q}} \frac{E_0(\rho, \mathbf{Q})}{\rho}. \tag{5.8.5}$$

Отыскивая предел либо по правилу Лопиталя, либо с помощью разложения $E_0(\rho, \mathbf{Q})$ в ряд по степеням $1/(1+\rho)$, получаем

$$R_\infty = -\ln\left[\min_{\mathbf{Q}} \max_j \sum_k Q(k)\,\varphi(j|k)\right], \tag{5.8.6}$$

где

$$\varphi(j|k) = \begin{cases} 1 & \text{при } P(j|k) \neq 0, \\ 0 & \text{при } P(j|k) = 0. \end{cases} \tag{5.8.7}$$

Это значит, что для каждого выхода берется сумма вероятностей букв на входе, из которых этот выход может быть достигнут. Входные вероятности выбираются так, чтобы получить минимум наибольшей из этих сумм, и R_∞ равно взятому со знаком минус логарифму этой минимаксной суммы. Отсюда можно увидеть, что $R_\infty = 0$, за исключением того случая, когда любой выход является недостижимым по крайней мере из одного входа.

Отметим теперь, что значение ρ, на котором достигается максимум (5.8.2), убывает с ростом R. Более того, если максимизирующее значение ρ лежит между 0 и 1, то $E_r(R) = E_{sp}(R)$. Поэтому, если $E_r(R) = E_{sp}(R)$ для некоторого значения R, то равенство также сохраняется для всех больших значений R. Определим R_{cr} как такую наименьшую скорость R, т. е. как такое значение, для которого $E_{sp}(R) = E_r(R)$ тогда и только тогда, когда $R \geqslant R_{cr}$. Другими словами, для любого канала существует интервал скоростей $R_{cr} \leqslant R \leqslant C$, в котором показатели экспонент в верхней и нижней границах вероятности ошибки совпадают.

Функция надежности канала $E(R)$ определяется равенством

$$E(R) = \lim_{N\to\infty} \sup \frac{-\ln P_e(N, R)}{N}, \tag{5.8.8}$$

где $P_e(N, R)$ — минимум P_e по всем (N, R)-кодам при заданных N и R. Таким образом, надежность определяется как наибольший показатель экспоненты, с которым может убывать вероятность ошибки с ростом N. Показатели экспонент $E_r(R)$ и $E_{sp}(R)$ являются нижней и верхней границами для $E(R)$ соответственно и, как было отмечено выше, функция $E(R)$ точно известна при $R_{cr} \leqslant R \leqslant C$. Конечно, удивительно то, что довольно грубые границы, которые были использованы в § 5.6, дают правильную функцию надежности канала в некотором интервале скоростей. Вместе с тем эти методы построения границ были выбраны с учетом того, что они дают функцию надежности. Имеется много других методов построения верхней границы вероятности ошибки, часто представляющиеся менее грубыми, которые дают, однако, более слабые результаты.

Может случиться (как в примерах (в) и (г) на рис. 5.8.1), что $R_{cr} = C$. Это означает, что для значений R, сколь угодно близких к C, выражение $[\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) - \rho R]$ не достигает максимума при ρ из интервала $0 \leqslant \rho \leqslant 1$. Это, в свою очередь, означает, что координата точки пересечения оси R с кривой $\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})/\rho$ больше или равна C при некотором $\rho \geqslant 1$, либо $\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})/\rho$ стремится к C при $\rho \to \infty$.

В любом случае, поскольку $E_0(\rho, \mathbf{Q})$ при фиксированном $\mathbf{Q}$ является выпуклой, то для некоторого $\mathbf{Q}$ должно быть $E_0(\rho, \mathbf{Q}) = \rho C$. Согласно теореме 5.6.3 это может произойти тогда и только тогда, когда (5.6.26а) удовлетворяется при этом $\mathbf{Q}$. Используя подобное рассуждение, можно показать, что эти условия являются также необходимыми и достаточными для того, чтобы $R_\infty = C$. Подытоживая сказанное, отметим, что следующие три утверждения являются эквивалентными 1) $R_{cr} = C$; 2) $R_\infty = C$; 3) равенство (5.6.26а) удовлетворяется при некотором $\mathbf{Q}$, давая пропускную способность.

Теорема 5.8.2. (Прямолинейная граница.) Пусть задан произвольный дискретный канал без памяти, для которого $E_{ex}(0) < \infty$, и пусть $E_{sl}(R)$ — линейная функция $R \geqslant 0$, которая касается кривой $E_{sp}(R)$ и вместе с тем удовлетворяет условию $E_{sl}(0) = E_{ex}(0)$. Пусть R_1 является значением R, при котором $E_{sl}(R) = E_{sp}(R)$. Пусть $o_3(N)$ и $o_4(N)$ — функции, стремящиеся к нулю с ростом N и которые могут быть представлены в виде

$$o_3(N) = \frac{\ln 2}{\sqrt{N}} + \frac{\ln 8 + K \ln N}{N}, \tag{5.8.9}$$

$$o_4(N) = \frac{2\sqrt{K} \max\limits_{i,k} \left[-2 \ln \sum\limits_j \sqrt{P(j|i) P(j|k)} \right]}{\sqrt{\lfloor \log_2 \sqrt{N} \rfloor}} + \\ + \sqrt{\frac{8}{N}} \ln \frac{e}{P_{min}} + \frac{\ln 2}{\sqrt{N}} + \frac{5 \ln 2}{N} + \frac{E_{ex}(0)}{N}. \tag{5.8.10}$$

Тогда при любом положительном целом N и любой R, $o_3(N) \leqslant R \leqslant R_1$, любой (N, R)-код имеет вероятность ошибки, которая удовлетворяет неравенству

$$P_e > \exp\left(-N\{E_{sl}[R - o_3(N)] + o_4(N)\}\right). \tag{5.8.11}$$

Это утверждение является теоремой 4 в работе Шеннона, Галлагера и Берлекэмпа (1967) и там же приведено ее доказательство. В формулировке теоремы 4 имеется небольшая ошибка, состоящая в том, что пропущено условие $R \geqslant o_3(N)$. Однако приведенное там доказательство верно для сформулированной здесь теоремы. Прямолинейный отрезок показателя экспоненты $E_{sl}(R)$ вместе с показателем экспоненты сферической упаковки дают верхнюю границу для надежности $E(R)$ при всех R, $0 < R < C$, в пределе при больших N. Эти границы изображены на рис. 5.8.2 для тех же каналов, что и на рис. 5.8.1.

Приступим теперь к доказательству этих теорем в частном случае ДСК. Начнем с рассмотрения произвольного (N, R)-кода и найдем нижнюю границу для P_e, которая не зависит от кода и, таким образом, является нижней границей P_e для всех кодов с заданными N и R. Пусть $\mathbf{x}_1, \ldots, \mathbf{x}_M$ — кодовые слова кода, $M = \lceil e^{NR} \rceil$ и пусть $Y_1, \ldots, Y_M$ — области декодирования этого кода (т. е. Y_m является множеством последовательностей на выходе канала, которые декодируются в сообщение m). Если сообщение m поступает на кодер, то передается $\mathbf{x}_m$ и произойдет правильное декодирование, если будет принято $\mathbf{y} \in$

$\in Y_m$. Поскольку при условии, что задано m, это событие имеет вероятность

$$\sum_{\mathbf{y}\in Y_m} P(\mathbf{y}\,|\,\mathbf{x}_m),$$

то общая вероятность правильного декодирования равна

$$P_c = \frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in Y_m} P(\mathbf{y}\,|\,\mathbf{x}_m). \tag{5.8.12}$$

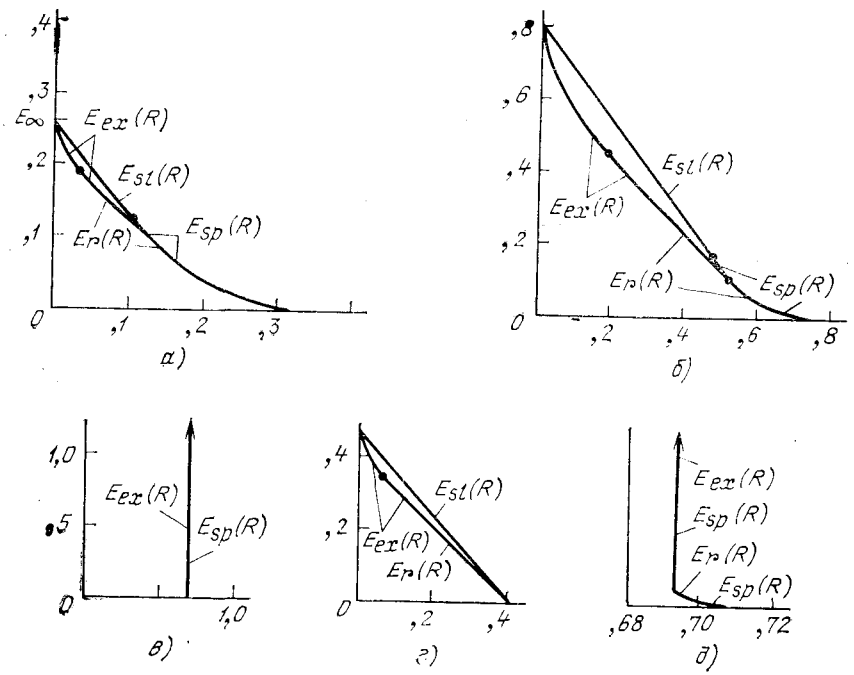


Рис. 5.8.2. Границы для функции надежности (те же каналы, что и на рис. 5.8.1).

Определим теперь *расстояние Хэмминга* $d(\mathbf{y};\mathbf{x})$ между двумя двоичными последовательностями как число позиций, в которых отличаются эти две последовательности. Например, расстояние Хэмминга между (0, 0, 1, 1, 1) и (1, 0, 1, 0, 1) равно 2. В ДСК с вероятностью ошибки ε, если $d(\mathbf{y};\mathbf{x}_m) = n$, то $P(\mathbf{y}\,|\,\mathbf{x}_m) = \varepsilon^n(1-\varepsilon)^{N-n}$. Если через $A_{n,m}$ обозначить число последовательностей $\mathbf{y}$, которые декодируются в сообщение m и расстояние которых от $\mathbf{x}_m$ равно n, то (5.8.12) можно представить в виде

$$P_c = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=0}^{N} A_{n,m}\,\varepsilon^n(1-\varepsilon)^{N-n}. \tag{5.8.13}$$

Используя равенство для биномиального распределения

$$1=\sum_{n=0}^{N}\binom{N}{n}\varepsilon^n(1-\varepsilon)^{N-n},$$

находим, что вероятность ошибки $P_e=1-P^c$ будет равна

$$P_e=\frac{1}{M}\sum_{m=1}^{M}\sum_{n=0}^{N}\left[\binom{N}{n}-A_{n,m}\right]\varepsilon^n(1-\varepsilon)^{N-n}. \qquad (5.8.14)$$

Чтобы истолковать это выражение, заметим, что если было передано сообщение m, то $A_{n,m}$ равно числу последовательностей, находящихся на расстоянии n от $\mathbf{x}_m$, прием которых приводит к правильному декодированию. В силу того, что $\binom{N}{n}$ равно общему числу последовательностей, находящихся на расстоянии n от $\mathbf{x}_m$, то $\left[\binom{N}{n}-A_{n,m}\right]$ из этих последовательностей будут приводить к ошибочному декодированию при передаче m.

Найдем теперь некоторые ограничения на множество целых чисел $\{A_{n,m}\}$, которые справедливы для всех кодов с данными N и M и затем найдем минимум правой части (5.8.14) при соблюдении этих ограничений. Ограничения, которые будут использованы, имеют вид

$$A_{n,m}\leqslant\binom{N}{n}\text{ при всех }n\text{ и }m, \qquad (5.8.15)$$

$$\sum_{m=1}^{M}\sum_{n=0}^{N}A_{n,m}\leqslant 2^N. \qquad (5.8.16)$$

Ограничение (5.8.16) связано с тем, что имеются всего 2^N выходных последовательностей; каждая последовательность декодируется не более чем в одно сообщение и расстояние от сопоставленного ей кодового слова определяется однозначно.

Минимум (5.8.14) при соблюдении этих ограничений достигается при $\varepsilon<1/2$ и для всех m, когда

$$A_{n,m}=\begin{cases}\binom{N}{n}, & 0\leqslant n\leqslant k-1,\\ 0, & k+1\leqslant n\leqslant N,\end{cases} \qquad (5.8.17)$$

где k выбирается так, что

$$M\sum_{n=0}^{k-1}\binom{N}{n}+\sum_{m=1}^{M}A_{k,m}=2^N;\quad 0<\sum_{m=1}^{M}A_{k,m}\leqslant M\binom{N}{k}. \qquad (5.8.18)$$

Отдельные значения $A_{k,m}$ не существенны, если их сумма по m удовлетворяет (5.8.18). Для того чтобы убедиться, что такой выбор приводит к минимуму (5.8.14), заметим, что при любом другом выборе можно найти n' и n, $n'<n$, такие, что $A_{n'm'}<\binom{N}{n'}$ для некоторого m' и $A_{n,m}>0$ для некоторого m. При таком выборе можно увидеть, что (5.8.14) уменьшается при увеличении $A_{n'm'}$ на 1 и уменьшении $A_{n,m}$ на 1. Подставляя (5.8.17) в (5.8.14) и замечая, что в результате

получится нижняя граница P_e для всех кодов, с заданными значениями N и M, будем иметь

$$P_e(N, M) \geqslant \left[\binom{N}{k} - \frac{1}{M}\sum_{m=1}^{M} A_{k,m}\right]\varepsilon^k (1-\varepsilon)^{N-k} + \\ + \sum_{n=k+1}^{N} \binom{N}{n} \varepsilon^n (1-\varepsilon)^{N-n}, \qquad (5.8.19)$$

где $P_e(N, M)$ определяется как минимальная вероятность ошибки по всем кодам с данной длиной блока N и с данным числом кодовых слов M.

Эта граница называется границей сферической упаковки. Множество последовательностей, находящихся на расстоянии k или меньше от кодового слова, можно интерпретировать как сферу радиуса k вокруг этого кодового слова. Граница, представленная (5.8.19), является вероятностью ошибки, которая бы имела место, если бы можно было выбрать кодовые слова так, чтобы множество сфер радиуса k, описанных вокруг различных кодовых слов, исчерпывало все пространство двоичных последовательностей длины N и сферы имели бы пересечения друг с другом только по внешним слоям радиуса k. Такие коды называются сферически упакованными кодами и код, изображенный на рис. 5.2.1, дает пример такого кода (в этом случае нет пересечений даже на внешнем слое радиуса 1). Часто при выводе этой границы сначала находится вероятность ошибки для сферически упакованного кода и затем показывается, что сферически упакованный код имеет вероятность ошибки не большую, чем любой другой код с теми же самыми N и M. В таком выводе имеется логическая ошибка, состоящая в том, что для большинства значений N и M не существует сферически упакованных кодов.

Приведем теперь (5.8.19) к аналитически более простой, но несколько более слабой форме. Существует ряд методов, чтобы сделать это, которые приводят к одной и той же экспоненциальной границе для вероятности ошибки, но с разными коэффициентами. Хотя коэффициенты здесь получаются не очень хороши, но мы избегаем некоторых запутанных деталей при доказательстве теоремы 5.8.4. Для того чтобы получить точные численные значения, в особенности при малых N, следует исходить непосредственно из (5.8.19).

Теорема 5.8.3. (Граница сферической упаковки для ДСК.) Пусть задан двоичный симметричный канал с вероятностью ошибки $\varepsilon < 1/2$ и пусть δ — произвольное число, $\varepsilon \leqslant \delta \leqslant 1/2$ и $\mathcal{H}(\delta) = -\delta \ln \delta - (1-\delta)\ln(1-\delta)$. Если число кодовых слов M удовлетворяет неравенству

$$M \geqslant \sqrt{8(N+1)}\exp\{N[\ln 2 - \mathcal{H}(\delta)]\}, \qquad (5.8.20)$$

то

$$P_e(N, M) \geqslant \frac{\varepsilon}{(1-\varepsilon)\sqrt{8(N+1)}}\exp\{N[\mathcal{H}(\delta) + \delta \ln \varepsilon + \\ + (1-\delta)\ln(1-\varepsilon)]\}. \qquad (5.8.21)$$

Доказательство. Как видно из (5.8.14), любое увеличение в какой-либо из сумм $\sum_{m=1}^{M} A_{n,m}$ до значений, больших, чем указанные в (5.8.17) и (5.8.18), приводит к дальнейшему уменьшению нижней границы P_e. Пусть $n' = \lceil \delta N \rceil$ и выберем $A_{n',m}$ так, что $\sum_{m=1}^{M} A_{n',m} = 2^N$. Для всех $n \neq n'$ выберем $A_{n,m}$ так, что $\sum_{m=1}^{M} A_{n,m} = \binom{N}{n} M$. Эти выборы, очевидно, приводят к значениям сумм, большим, чем те, которые определяются из (5.8.17) и (5.8.18); подставляя эти выражения в (5.8.14), получаем

$$P_e(N, M) \geqslant \left[\binom{N}{n'} - \frac{2^N}{M}\right] \varepsilon^{n'} (1-\varepsilon)^{N-n'}. \qquad (5.8.22)$$

Используя границу Стирлинга для факториала, можно оценить снизу $\binom{N}{n'}$ следующим образом (см. задачу 5.8(б)):

$$\binom{N}{n'} \geqslant \frac{1}{\sqrt{2N}} \exp\left[N\mathcal{H}\left(\frac{n'}{N}\right)\right]. \qquad (5.8.23)$$

Если $n' \leqslant N/2$, то $\mathcal{H}(n'/N) \geqslant \mathcal{H}(\delta)$. Так как $1/N \geqslant 1/(N+1)$, то

$$\binom{N}{n'} \geqslant \frac{1}{\sqrt{2(N+1)}} \exp[N\mathcal{H}(\delta)]. \qquad (5.8.24)$$

Так как $\delta < 1/2$ и $n' = \lceil \delta N \rceil$, то единственно возможное значение n', которое превышает $N/2$, равно $(N+1)/2$. В этом случае, используем специальную границу (см. задачу 5.8(б)):

$$\binom{N}{(N+1)/2} \geqslant \frac{1}{\sqrt{2(N+1)}} \cdot 2^N. \qquad (5.8.25)$$

Так как $2^N \geqslant \exp(N\mathcal{H}(\delta))$, то граница в (5.8.24) справедлива при всех возможных значениях n'. Далее, в силу того, что n' превышает δN не более чем на 1, будем иметь

$$\left(\frac{\varepsilon}{1-\varepsilon}\right)^{n'} \geqslant \frac{\varepsilon}{1-\varepsilon}\left(\frac{\varepsilon}{1-\varepsilon}\right)^{\delta N}.$$

Подставляя это выражение и (5.8.24) в (5.8.22) и применяя границу для M в (5.8.20), получаем (5.8.21), что завершает доказательство теоремы. |

Если найти $E_{sp}(R)$ [в соответствии с (5.8.2)], то, используя те же рассуждения, что и в примере 1 § 5.6, получим, что при $R < C$

$$E_{sp}(R) = -\delta \ln \varepsilon - (1-\delta)\ln(1-\varepsilon) - \mathcal{H}(\delta), \qquad (5.8.26)$$

$$R = \ln 2 - \mathcal{H}(\delta). \qquad (5.8.27)$$

Рассмотрим теперь некоторый произвольный (N, R)-код, для которого $(1/N)\ln[8(N+1)] < R < C$. Если выбрать δ так, чтобы удовлетво-

рялось равенство

$$R = \ln 2 - \mathcal{H}(\delta) + \frac{\ln [8(N+1)]}{N},$$

то $M = \lceil \exp NR \rceil$ должно удовлетворять (5.8.20), и из (5.8.21) вытекает результат, который эквивалентен теореме 5.8.1:

$$P_e(N, M) \geqslant \exp -N \left\{ E_{sp} \left(R - \frac{\ln \sqrt{8(N+1)}}{N} \right) + \right.$$
$$\left. + \frac{\ln [\sqrt{8(N+1)} \cdot (1-\varepsilon)/\varepsilon]}{N} \right\}. \qquad (5.8.28)$$

Для $R > C$ мы ниже выведем более сильную границу, чем граница в теореме 5.8.1.

Для того чтобы установить справедливость теоремы 5.8.2 для ДСК, нам понадобится несколько лемм, которые представляют самостоятельный интерес. В первой из них рассматривается концепция декодирования, называемая «декодирование списком». Предположим, что при заданном множестве M кодовых слов длины N декодер отображает любую принятую последовательность в список, скажем, L сообщений. Такое декодирование могло бы быть полезным, если бы планировалось использование обратной связи в системе передачи и при последующей передаче устранялась неопределенность в том, какое из L декодированных сообщений было в действительности передано. Если переданное сообщение не принадлежит списку из L декодируемых сообщений, то говорят, что произошла *ошибка при декодировании списком*. Пусть $P_e(N, M, L)$ — минимальная вероятность ошибки при декодировании списком по всем кодам с M кодовыми словами, длиной блока N и списком из L декодируемых сообщений. Можно повторить вывод границы сферической упаковки для схемы декодирования списком, обозначая через Y_m множество выходных последовательностей $\mathbf{y}$, для которых m принадлежит декодируемому списку. Равенство (5.8.14) остается справедливым, если понимать под $A_{n,m}$ число выходных последовательностей $\mathbf{y}$, для которых m принадлежит декодируемому списку и которые находятся на расстоянии n от $\mathbf{x}_m$. В этих условиях ограничение (5.8.16) для $A_{n,m}$ принимает вид

$$\sum_{m=1}^{M} \sum_{n=0}^{N} A_{n,m} = L2^N,$$

так как любое $\mathbf{y}$ декодируется в точности в L сообщений и, следовательно, дает вклад в точности в L слагаемых $A_{n,m}$. Используя это равенство, получим следующую лемму.

Лемма 1. Пусть задан ДСК с вероятностью ошибки $\varepsilon < 1/2$ и пусть δ — произвольное число, $\varepsilon < \delta < 1/2$. Если

$$\frac{M}{L} \geqslant \sqrt{8(N+1)} \exp \{N [\ln 2 - \mathcal{H}(\delta)]\}, \qquad (5.8.29)$$

то

$$P_e(N, M, L) \geqslant \frac{\varepsilon}{(1-\varepsilon)\sqrt{8(N+1)}} \exp\{N[\mathcal{H}(\delta) + \\ + \delta \ln \varepsilon + (1-\delta)\ln(1-\varepsilon)]\}. \qquad (5.8.30)$$

Доказательство. Доказательство аналогично доказательству теоремы 5.8.3, с той лишь разницей, что следует положить

$$\sum_{m=1}^{M} A_{n,m} = 2^N L$$

при всех $n \neq n'$. |

Займемся теперь получением нижней границы для вероятности ошибки, лучшей, чем граница сферической упаковки для очень малого числа кодовых слов. Определим *минимальное расстояние* двоичного кода как расстояние между двумя ближайшими кодовыми словами.

Лемма 2. (Граница Плоткина.) Минимальное расстояние любого двоичного кода с M кодовыми словами и длиной блока N удовлетворяет неравенству

$$d_{min} \leqslant \frac{NM}{2(M-1)}. \qquad (5.8.31)$$

Доказательство. Расположим кодовые слова (N, M)-кода в таблицу с N столбцами и M строками двоичных символов; m-е кодовое слово будет m-й строкой таблицы. Рассмотрим теперь сумму всех расстояний в коде, т. е.

$$\sum_{m=1}^{M}\sum_{m'=1}^{M} d(\mathbf{x}_m; \mathbf{x}_{m'}) = \sum_{n=1}^{N}\sum_{m=1}^{M}\sum_{m'=1}^{M} d(x_{m,n}; x_{m',n}). \qquad (5.8.32)$$

Пусть $Z(n)$ — число нулей в n-м столбце таблицы. Так как существует $Z(n)$ различных значений m', для которых $x_{m',n} = 0$, то при $x_{m,n} = 1$

$$\sum_{m'=1}^{M} d(x_{m,n}; x_{m',n}) = Z(n).$$

Поскольку существуют $M - Z(n)$ значений m, для которых $x_{m,n} = 1$, то

$$\sum_{m:x_{m,n}=1}\sum_{m'=1}^{M} d(x_{m,n}; x_{m',n}) = [M - Z(n)]\, Z(n). \qquad (5.8.33)$$

Точно так же существуют $Z(n)$ значений m, для которых $x_{m,n} = 0$, и $M - Z(n)$ значений m', для которых $x_{m',n} = 1$ так что

$$\sum_{m=1}^{M}\sum_{m'=1} d(x_{m,n}; x_{m',n}) = 2[M - Z(n)]\, Z(n). \qquad (5.8.34)$$

Правая часть (5.8.34) ограничена сверху числом $M^2/2$, которое является максимальным значением выражения $2Z(M-Z)$, рассматриваемого как функция Z. Этот максимум достигается при $Z = M/2$. Поэтому

$$\sum_{m=1}^{M}\sum_{m'=1}^{M} d(\mathbf{x}_m; \mathbf{x}_{m'}) \leqslant \frac{NM^2}{2}. \tag{5.8.35}$$

Вместе с тем, так как $d(\mathbf{x}_m; \mathbf{x}_{m'}) = 0$, то можно опустить те слагаемые в указанной выше сумме, для которых $m' = m$ и в результате останется $M(M-1)$ ненулевых слагаемых. Получим

$$\sum_{m=1}^{M}\sum_{m'=1}^{M} d(\mathbf{x}_m; \mathbf{x}_{m'}) =$$

$$= \sum_{m=1}^{M}\sum_{m'\neq m} d(\mathbf{x}_m; \mathbf{x}_{m'}) \geqslant M(M-1)\, d_{min}. \tag{5.8.36}$$

Объединяя (5.8.35) и (5.8.36), получаем (5.8.31). |

Определим теперь $P_{e,w}(N, M)$ как минимальную вероятность ошибки для наихудшего кодового слова в коде при выполнении минимизации по всем кодам с данной длиной блока N и с данным числом кодовых слов M, т. е.

$$P_{e,w}(N, M) = \min_{\text{по кодам}} [\max_{m} P_{e,m}].$$

Лемма 3. В ДСК с вероятностью ошибки $\varepsilon < 1/2$ при $M \geqslant N+2$ имеет место неравенство

$$P_{e,w}(N, M) \geqslant \frac{V\overline{N}\,\varepsilon}{(N+4)(1-\varepsilon)} \exp[-NE_{ex}(0)], \tag{5.8.37}$$

где

$$E_{ex}(0) = -\frac{1}{4}\ln[4\varepsilon(1-\varepsilon)]. \tag{5.8.38}$$

Отметим, что показатель экспоненты $E_{ex}(0)$ равен значению показателя экспоненты для процедуры с выбрасыванием при $R = 0$ [см. (5.7.20)].

Доказательство. Так как d_{min} должно быть целым числом, то легко проверить, что из (5.8.31) вытекает неравенство $d_{min} \leqslant N/2$ при $M \geqslant N+2$. Предположим, что в данном коде два кодовых слова $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ находятся на расстоянии $d = d_{min}$ друг от друга. Имеем

$$P_{e,m} + P_{e,m'} = \sum_{\mathbf{y}\in Y_m^c} P(\mathbf{y}\,|\,\mathbf{x}_m) + \sum_{\mathbf{y}\in Y_{m'}^c} P(\mathbf{y}\,|\,\mathbf{x}_{m'}). \tag{5.8.38а}$$

Можно получить нижнюю границу для сумм $P_{e,m} + P_{e,m'}$, расширяя Y_m и $Y_{m'}$ так, чтобы все $\mathbf{y}$ декодировались либо в m, либо в m', и далее, строя нижнюю границу с помощью преобразования Y_m и $Y_{m'}$ в области, соответствующие декодированию по максимуму правдоподобия двух кодовых слов. При этом можно пренебречь тем, что было

принято на позициях, в которых $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ совпадают, и таким образом рассмотреть только вероятность возникновения более чем $d/2$ ошибок в канале среди d символов, в которых отличаются $\mathbf{x}_m$ и $\mathbf{x}_{m'}$. Получим

$$\frac{P_{e,m}+P_{e,m'}}{2} \geqslant \sum_{i>d/2} \binom{d}{i} \varepsilon^i (1-\varepsilon)^{d-i}. \tag{5.8.39}$$

При четных d это выражение ограничено снизу следующим образом:

$$\frac{P_{e,m}+P_{e,m'}}{2} \geqslant \binom{d}{\frac{d}{2}+1} \varepsilon^{(d/2)+1} (1-\varepsilon)^{(d/2)-1} =$$

$$= \frac{d\varepsilon}{(d+2)(1-\varepsilon)} \binom{d}{d/2} \varepsilon^{d/2} (1-\varepsilon)^{d/2} \geqslant$$

$$\geqslant \frac{d\varepsilon}{(d+2)(1-\varepsilon)\sqrt{2d}} \exp\left\{\frac{d}{2} \ln [4\varepsilon(1-\varepsilon)]\right\},$$

где было использовано неравенство (5.8.23). Это выражение представляет собой убывающую функцию d при четных d, и так как $d \leqslant N/2$, то

$$\frac{P_{e,m}+P_{e,m'}}{2} \geqslant \frac{\sqrt{N}\,\varepsilon}{(N+4)(1-\varepsilon)} \exp\left\{\frac{N}{4} \ln [4\varepsilon(1-\varepsilon)]\right\}. \tag{5.8.40}$$

При нечетном d рассуждения почти совпадают; используется (5.8.25) при получении нижней границы для $\binom{d}{(d+1)/2}$. В результате получим неравенство

$$\frac{P_{e,m}+P_{e,m'}}{2} \geqslant \sqrt{\frac{\varepsilon}{(1-\varepsilon)(N+2)}} \exp\left\{\frac{N}{4} \ln [4\varepsilon(1-\varepsilon)]\right\},$$

правая часть которого ограничена снизу выражением (5.8.40); это заканчивает доказательство. |

Для того чтобы дать интерпретацию доказанной лемме, заметим, что скорость кода с $M = N + 2$ кодовыми словами стремится к нулю при N, стремящемся к ∞. В силу того, что показатель экспоненты в лемме совпадает с показателем экспоненты случайного кодирования для процедуры с выбрасыванием при нулевой скорости*), то, как было сказано, $E_{ex}(0)$ является надежностью канала в пределе при R, стремящемся к 0. Однако нужно быть всегда осторожным с этим утверждением. Лемма 3 не применима при $M < N + 2$ и в действительности, как можно показать**), при фиксированном M в ДСК

$$\lim_{N\to\infty} \frac{-\ln P_e(N,M)}{N} = \frac{M}{M-1} E_{ex}(0).$$

Одним из интересных аспектов леммы является то, что она выявляет значение минимального расстояния кода (для кода с малой скоростью) в определении его вероятности ошибки. Важность этого расстояния была также указана при рассмотрении границы случайного коди-

*) См. задачу 5.32 для установления связи между $P_{e,w}(NM)$ и $P_e(NM)$.
**) Шеннон, Галлагер и Берлекэмп (1967, II).

рования для процедуры с выбрасыванием, где мы выбрасывали кодовые слова, которые были слишком близки друг к другу. При больших скоростях, близких к пропускной способности, минимальные расстояния становятся относительно маловажными и нетрудно заметить, что в ансамбле случайных кодов большинство кодов имеют очень малые минимальные расстояния. Можно провести чистку ансамбля, значительно увеличив минимальное расстояние кодов, однако при больших скоростях это не может существенно снизить среднюю по ансамблю вероятность ошибки, так как это среднее близко к исходной границе сферической упаковки.

Л е м м а 4. *Для произвольных положительных целых чисел* N_1, N_2, M и L

$$P_e(N_1+N_2, M) \geqslant P_e(N_1, M, L)\, P_{e,w}(N_2, L+1). \qquad (5.8.41)$$

Интуитивная идея, лежащая в основе этой леммы, состоит в том, что в коде с длиной блока $N_1 + N_2$ ошибочное декодирование произойдет, если для первых N_1 принятых символов имеются L сообщений, более вероятных, чем переданное сообщение, и если для последних N_2 принятых символов одно из этих L сообщений опять является более вероятным, чем переданное сообщение. Вероятности, стоящие в правой части (5.8.41), связаны с вероятностями этих событий. Хотя здесь рассматривается только ДСК, ниже следующее доказательство применимо к произвольному дискретному каналу без памяти.

Доказательство. Пусть задан код с M кодовыми словами длины N_1+N_2, пусть $\mathbf{x}_m$ является m-м кодовым словом и пусть префиксом $\mathbf{x}_{m,1}$ будут первые N_1 символов $\mathbf{x}_m$, а суффиксом $\mathbf{x}_{m,2}$ будут последние N_2 символов. Точно так же принятую последовательность $\mathbf{y}$ разобьем на префикс $\mathbf{y}_1$ и суффикс $\mathbf{y}_2$, состоящие из N_1 и N_2 букв соответственно. Пусть Y_m при любом m, $1 \leqslant m \leqslant M$, является множеством выходных последовательностей $\mathbf{y}$, декодируемых в сообщение m, и пусть Y_m^c является дополнением Y_m. Тогда

$$P_e = \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y} \in Y_m^c} P(\mathbf{y} \mid \mathbf{x}_m). \qquad (5.8.42)$$

При любом префиксе $\mathbf{y}_1$ пусть $Y_{m,2}(\mathbf{y}_1)$ будет множеством суффиксов $\mathbf{y}_2$, для которых $(\mathbf{y}_1, \mathbf{y}_2) \in Y_m$. В канале без памяти $P(\mathbf{y}_1, \mathbf{y}_2 \mid \mathbf{x}_m) = P(\mathbf{y}_1 \mid \mathbf{x}_{m,1})\, P(\mathbf{y}_2 \mid \mathbf{x}_{m,2})$, и (5.8.42) можно переписать в виде

$$P_e = \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathbf{y}_1} P(\mathbf{y}_1 \mid \mathbf{x}_{m,1}) \sum_{\mathbf{y}_2 \in Y_{m,2}^c(\mathbf{y}_1)} P(\mathbf{y}_2 \mid \mathbf{x}_{m,2}). \qquad (5.8.43)$$

Для любого заданного $\mathbf{y}_1$ можно рассматривать множество суффиксов $\{\mathbf{x}_{m,2}\}$, $1 \leqslant m \leqslant M$, и множество областей $\{Y_{m,2}(\mathbf{y}_1)\}$, $1 \leqslant m \leqslant M$, как код и его области декодирования. Вероятности ошибок для слов в этом коде при условии, что задано $\mathbf{y}_1$, равны

$$P_{e,m}(\mathbf{y}_1) = \sum_{\mathbf{y}_2 \in Y_{m,2}^c(\mathbf{y}_1)} P(\mathbf{y}_2 \mid \mathbf{x}_{m,2}). \qquad (5.8.44)$$

Пусть $m_1(\mathbf{y}_1)$ обозначает m, для которого $P_{e,m}(\mathbf{y}_1)$ является наименьшей, $m_2(\mathbf{y}_1)$ обозначает m, для которого $P_{e,m}(\mathbf{y}_1)$ является наименьшей из оставшихся вероятностей, и т. д. Требуется показать, что для всех m, кроме, быть может, $m_1(\mathbf{y}_1), \ldots, m_L(\mathbf{y}_1)$,

$$P_{e,m}(\mathbf{y}_1) \geqslant P_{e,w}(N_2, L+1).$$

Если бы это было неверно, то множество $L+1$ кодовых слов $\{\mathbf{x}_{m,2}\}$ при $m = m_1(\mathbf{y}_1), \ldots, m_{L+1}(\mathbf{y}_1)$ и множество областей декодирования $\{Y_{m,2}(\mathbf{y}_1)\}$ при $m = m_1(\mathbf{y}_1), \ldots, m_{L+1}(\mathbf{y}_1)$ все давали бы вероятности ошибок, меньшие, чем $P_{e,w}(N_2, L+1)$, что приводит к противоречию. Теперь получаем следующую нижнюю границу:

$$\sum_{\mathbf{y}_2 \in Y^c_{m,2}(\mathbf{y}_1)} P(\mathbf{y}_2 \mid \mathbf{x}_{m,2}) \geqslant \begin{cases} 0, & \text{при } m = m_1(\mathbf{y}_1), \ldots, m_L(\mathbf{y}_1), \\ P_{e,w}(N_2, L+1) & \text{при остальных } m. \end{cases} \tag{5.8.45}$$

Изменяя порядок суммирования по m и $\mathbf{y}_1$ в (5.8.43) и подставляя (5.8.45) в полученное выражение, имеем

$$P_e \geqslant \frac{1}{M} \sum_{\mathbf{y}_1} \sum_{\substack{m_l(\mathbf{y}_1) \\ l > L}} P(\mathbf{y}_1 \mid \mathbf{x}_{m,1}) P_{e,w}(N_2, L+1). \tag{5.8.46}$$

Наконец, можно рассмотреть множество префиксов $\{\mathbf{x}_{m,1}\}$ как множество M кодовых слов с длиной блока N_1 и можно рассмотреть $m_l(\mathbf{y}_1)$, $l = 1, \ldots, L$, как правило декодирования списком для этого множества кодовых слов. Поэтому

$$\frac{1}{M} \sum_{\mathbf{y}_1} \sum_{\substack{m_l(\mathbf{y}_1) \\ l > L}} P(\mathbf{y}_1 | \mathbf{x}_{m,1}) \geqslant P(M, N_1, L). \tag{5.8.47}$$

Объединяя (5.8.47) и (5.8.46), завершаем доказательство леммы. |

Используем теперь четыре доказанные леммы для нахождения прямолинейного отрезка показателя экспоненты, представленного в теореме 5.8.2. Пусть δ, $\varepsilon < \delta < 1/2$, является вначале произвольным числом; введем обозначения

$$R_1 = \ln 2 - \mathcal{H}(\delta), \tag{5.8.48}$$

$$E_{sp}(R_1) = -\mathcal{H}(\delta) - \delta \ln \varepsilon - (1-\delta) \ln (1-\varepsilon). \tag{5.8.49}$$

Для (N, R)-кода с заданными N и R определим число λ с помощью равенства

$$R = \lambda R_1 + \frac{3 \ln [2(N+1)]}{2N}. \tag{5.8.50}$$

Будем рассматривать лишь скорости из интервала

$$\frac{3}{2N} \ln [2(N+1)] \leqslant R \leqslant R_1, \tag{5.8.51}$$

так, что $0 \leqslant \lambda < 1$. Обозначим теперь $N_1 = \lfloor \lambda N \rfloor$ и $N_2 = N - N_1$. Заметим, что $N_2 \geqslant 1$. Используя (5.8.50), можно показать, что число

кодовых слов $M = \lceil \exp NR \rceil$ удовлетворяет неравенству

$$\frac{M}{N+1} \geqslant \sqrt{8(N+1)} \exp(N\lambda R_1) \geqslant \tag{5.8.52}$$

$$\geqslant \sqrt{8(N_1+1)} \exp(N_1 R_1). \tag{5.8.53}$$

Теперь из леммы 1 следует, что

$$P_e(N_1, M, N+1) \geqslant \frac{\varepsilon}{(1-\varepsilon)\sqrt{8(N_1+1)}} \exp[-(N_1 E_{sp}(R_1)] \geqslant$$

$$\geqslant \frac{\varepsilon}{(1-\varepsilon)\sqrt{8(N+1)}} \exp[-\lambda N E_{sp}(R_1)]. \tag{5.8.54}$$

Согласно лемме 3, также имеем

$$P_{e,w}(N_2, N+2) \geqslant \frac{\sqrt{N_2}\,\varepsilon}{(N_2+4)(1-\varepsilon)} \exp[-N_2 E_{ex}(0)] \geqslant$$

$$\geqslant \frac{\varepsilon}{(N+4)(1-\varepsilon)} \exp\{-[(1-\lambda)N+1] E_{ex}(0)\} =$$

$$= \frac{\sqrt{2}\,\varepsilon^{5/4}}{(N+4)(1-\varepsilon)^{3/4}} \exp\{-(1-\lambda) N E_{ex}(0)\}. \tag{5.8.55}$$

Сочетая (5.8.54) и (5.8.55) с леммой 4, получаем

$$P_e(N, M) \geqslant \frac{\varepsilon^{9/4}}{2(1-\varepsilon)^{7/4}\sqrt{N+1}\,(N+4)} \times$$

$$\times \exp\{-N[\lambda E_{sp}(R_1) + (1-\lambda) E_{ex}(0)]\}. \tag{5.8.56}$$

Наконец, используя (5.8.50), чтобы выразить λ через R, и перенося коэффициент в показатель экспоненты, получаем

$$P_e(N, M) \geqslant \exp\left\{-N\left[E_{ex}(0) - R\left[\frac{E_{ex}(0) - E_{sp}(R_1)}{R_1}\right] + o(N)\right]\right\}, \tag{5.8.57}$$

где

$$o(N) = \frac{1}{N}\left[\frac{E_{ex}(0) - E_{sp}(R_1)}{R_1} \frac{3}{2} \ln[2(N+1)] - \right.$$

$$\left. - \ln \frac{\varepsilon^{9/4}}{2(1-\varepsilon)^{7/4}\sqrt{N+1}\,(N+4)}\right]. \tag{5.8.58}$$

Показатель экспоненты в (5.8.57) является линейной функцией R, которая изменяется от значения $E_{ex}(0)$ при $R = 0$ до значения $E_{sp}(R_1)$ при $R = R_1$. Если выбрать $R = R_1$ (т. е. δ), при которой указанная выше линейная функция имеет минимум, то, очевидно, получим наиболее точную экспоненциальную границу. Следующая теорема подытоживает полученные результаты.

Теорема 5.8.4. Пусть задан ДСК с вероятностью ошибки ε и пусть R_1 обозначает R, при которой прямая линия, проходящая через точку $R = 0$, $E = E_{ex}(0)$, является касательной к кривой $E_{sp}(R)$. Тогда при любом $N \geqslant 1$ и любой R из интервала (5.8.51) P_e для любого (N, R)-кода удовлетворяет (5.8.57).

Вероятность ошибки на блок при скоростях, больших пропускной способности

Здесь будет показано, что в любом дискретном канале без памяти и при любой фиксированной скорости, большей пропускной способности, P_e стремится к единице с ростом N*). Как было указано в § 4.3, такой результат не обязательно исключает возможность надежной передачи данных на скоростях, больших пропускной способности, так как большая вероятность ошибочного декодирования блока не означает, что будет большая вероятность ошибки в отдельном символе источника. Кроме того, такой результат ничего не говорит о вероятности ошибки для неблоковых кодов. Вместе с тем этот результат является более простым для понимания по сравнению с обращением теоремы кодирования (теоремы 4.3.4), так как он касается только канала, а не источника и канала вместе взятых, и он дает дополнительное понимание природы пропускной способности.

Теорема 5.8.5. [Вольфовиц (1957).] В произвольном дискретном канале без памяти с пропускной способностью C (в натуральных единицах) для любого (N, R)-кода при $R > C$ имеем

$$P_e \geqslant 1 - \frac{4A}{N(R-C)^2} - \exp\left[-\frac{N(R-C)}{2}\right], \tag{5.8.59}$$

где A — конечная положительная постоянная, зависящая от канала и не зависящая от N и R.

Обсуждение. Как видно из (5.8.59), при любой фиксированной скорости $R > C$ значение P_e должно стремиться к 1 при неограниченном возрастании N. Можно также заметить, что если выбрать $R = C + \delta/\sqrt{N}$ при некотором фиксированном значении $\delta > \sqrt{8A} + 2$, то P_e будет строго больше чем 0 при всех N.

Доказательство. Пусть $P(j|k)$, $0 \leqslant j \leqslant J-1$, $0 \leqslant k \leqslant K-1$, — переходные вероятности канала, а K и J — объемы входного и выходного алфавитов соответственно. Пусть $Q(0), \ldots, Q(K-1)$ будут вероятностями на входе, на которых достигается пропускная способность, и пусть $\omega(j) = \sum_k Q(k)P(j|k)$, $0 \leqslant j \leqslant J-1$, — соответствующие вероятности на выходе. Согласно теореме 4.5.1 имеем

$$I(k;Y) \triangleq \sum_{j=0}^{J-1} P(j|k) \ln \frac{P(j|k)}{\omega(j)} \leqslant C; \quad 0 \leqslant k \leqslant K-1. \tag{5.8.60}$$

---

*) Исторически, этот результат принадлежащий Вольфовицу, был назван сильным обращением теоремы кодирования, а результат теоремы 4.3.1, принадлежащий Фано,— слабым обращением теоремы кодирования. Так как результат, полученный Вольфовицем, не следует из теоремы 4.3.4, которую мы назвали обращением теоремы кодирования, то мы будем называть результаты, изложенные здесь, теоремой обращения по Вольфовицу или обращением теоремы кодирования для блокового кодирования.

Пусть $$P_N(\mathbf{y}\mid\mathbf{x})=\prod_{n=1}^{N}P(y_n\mid x_n),$$
где $\mathbf{y}=(y_1,\ldots,y_N)$ и $\mathbf{x}=(x_1,\ldots,x_N)$, и пусть
$$\omega_N(\mathbf{y})=\prod_{n=1}^{N}\omega(y_n).$$
Определим
$$I(\mathbf{x};\mathbf{y})=\ln\frac{P_N(\mathbf{y}\mid\mathbf{x})}{\omega_N(\mathbf{y})}=\sum_{n=1}^{N}I(x_n;y_n), \tag{5.8.61}$$
где
$$I(x_n;y_n)=\ln[P(y_n\mid x_n)/\omega(y_n)].$$

Рассмотрим теперь (N, R)-код с кодовыми словами $\mathbf{x}_1, \ldots, \mathbf{x}_M$ и областями декодирования $Y_1, \ldots, Y_M$. Вероятность правильного декодирования для этого кода равна
$$P_c=\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in Y_m}P_N(\mathbf{y}\mid\mathbf{x}_m). \tag{5.8.62}$$
Пусть $\varepsilon>0$ будет произвольным числом; определим для $1\leqslant m\leqslant M$
$$B_m=[\mathbf{y}:I(\mathbf{x}_m;\mathbf{y})>N(C+\varepsilon)]. \tag{5.8.63}$$
Обозначая через B_m^c дополнение B_m, представляем (5.8.62) в виде
$$P_c=\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in Y_m\cap B_m}P_N(\mathbf{y}\mid\mathbf{x}_m)+\frac{1}{M}\sum_{m=1}^{m}\sum_{\mathbf{y}\in Y_m\cap B_m^c}P_N(\mathbf{y}\mid\mathbf{x}_m). \tag{5.8.64}$$
Для $\mathbf{y}\in B_m^c$ имеем $P_N(\mathbf{y}\mid\mathbf{x}_m)/\omega_N(\mathbf{y})\leqslant\exp[N(C+\varepsilon)]$ и поэтому вторая сумма в (5.8.64) ограничена следующим образом:
$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in Y_m\cap B_m^c}P_N(\mathbf{y}\mid\mathbf{x}_m)&\leqslant\\ \leqslant\frac{1}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in Y_m\cap B_m^c}\omega_N(\mathbf{y})\exp[N(C+\varepsilon)]&\leqslant\\ \leqslant\frac{\exp[N(C+\varepsilon)]}{M}\sum_{m=1}^{M}\sum_{\mathbf{y}\in Y_m}\omega_N(\mathbf{y})&\leqslant\frac{\exp[N(C+\varepsilon)]}{M},\end{aligned} \tag{5.8.65}$$
где использовано то, что области декодирования не пересекаются и что $\omega_N(\mathbf{y})$ является распределением вероятности.

Первую сумму в (5.8.64) можно оценить сверху с помощью суммирования при любом заданном m по $\mathbf{y}\in B_m$, а не по $\mathbf{y}\in Y_m\cap B_m$. Получим
$$\sum_{\mathbf{y}\in B_m}P_N(\mathbf{y}\mid\mathbf{x}_m)=P[I(\mathbf{x}_m;\mathbf{y})>N(C+\varepsilon)\mid\mathbf{x}_m], \tag{5.8.66}$$

где при заданном $\mathbf{x}_m$ величина $I(\mathbf{x}_m; \mathbf{y})$ понимается как случайная величина, принимающая значение $I(\mathbf{x}_m; \mathbf{y})$ с вероятностью $P_N(\mathbf{y} \mid \mathbf{x}_m)$. Согласно (5.8.61) эта случайная величина является суммой независимых случайных величин $\Sigma I(x_{m,n}; y_n)$ и в соответствии с (5.8.60) среднее значение этой суммы меньше или равно NC. Отсюда согласно неравенству Чебышева имеем

$$\sum_{\mathbf{y} \in B_m} P_N(\mathbf{y} \mid \mathbf{x}_m) \leqslant \frac{\sum_{n=1}^{N} \mathbf{D}[I(x_{m,n}; y_n) \mid x_{m,n}]}{N^2 \varepsilon^2}, \qquad (5.8.67)$$

где

$$\mathbf{D}[I(x_{m,n}; y_n) \mid x_{m,n}] = \sum_{j=0}^{J-1} P(j \mid x_{m,n}) \left[\ln \frac{P(j \mid x_{m,n})}{\omega(j)}\right]^2 - \left[\sum_{j=0}^{J-1} P(j \mid x_{m,n}) \ln \frac{P(j \mid x_{m,n})}{\omega(j)}\right]^2. \qquad (5.8.68)$$

Так как $x_{m,n}$ является одной из букв на входе $0, \ldots, K-1$, то эта дисперсия всегда ограничена сверху конечным числом A, равным

$$A = \max_{0 \leqslant k \leqslant K-1} \mathbf{D}[I(k; y_n) \mid k]. \qquad (5.8.69)$$

Отсюда при всех m имеем

$$\sum_{\mathbf{y} \in B_m} P_N(\mathbf{y} \mid \mathbf{x}_m) \leqslant \frac{A}{N\varepsilon^2}. \qquad (5.8.70)$$

Подставляя (5.8.65) и (5.8.70) в (5.8.64), будем иметь

$$P_c \leqslant \frac{A}{N\varepsilon^2} + \frac{\exp[N(C+\varepsilon)}{M}.$$

В силу того, что это справедливо для всех (N, M)-кодов, имеем

$$P_e(N, M) \geqslant 1 - \frac{A}{N\varepsilon^2} - \frac{\exp[N(C+\varepsilon)]}{M}.$$

Наконец, для заданной $R > C$, выбирая $\varepsilon = (R - C)/2$ и используя равенство $M = \lceil e^{NR} \rceil$, получаем (5.8.59), что завершает доказательство теоремы. |

Некоторые обобщения этой теоремы рассматриваются в задачах 5.34—5.36. В частности, если в (5.8.67) использовать границу Чернова, а не неравенство Чебышева, то можно показать, что при фиксированной $R > C$ значение P_e стремится к 1 экспоненциально с ростом N. Точно так же, если заменить неравенство Чебышева центральной предельной теоремой, то можно получить более сильные результаты для R, близких к C.

5.9. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КАНАЛОВ С КОНЕЧНЫМ ЧИСЛОМ СОСТОЯНИЙ

В § 4.6 были описаны каналы с конечным числом состояний с помощью условной вероятностной меры $P(y_n, s_n \mid x_n, s_{n-1})$. Эта вероятностная мера определяет вероятность $P_N(\mathbf{y} \mid \mathbf{x}, s_0)$ любой последовательности на выходе $\mathbf{y} = (y_1, \ldots, y_N)$ при условии, то задана последовательность на входе $\mathbf{x} = (x_1, \ldots, x_N)$ и задано начальное состояние s_0. Вместе с тем теорема 5.6.1 является теоремой кодирования, которая справедлива для любого распределения вероятности для канала $P_N(y \mid x)$ (т. е. теорема 5.6.1 справедлива не только для каналов без памяти). Единственная трудность для прямого применения здесь этого результата состоит в том, что не ясно, как поступить с начальным состоянием. Нашей целью здесь является доказательство теоремы кодирования, которая может быть применена независимо от начального состояния. Главным результатом будет следующее утверждение: для любой скорости кода R и любом $\varepsilon > 0$ существуют коды с достаточно большой длиной блока, такие, что независимо от сообщения и начального состояния $P_e < \exp\{-N[E_r(R) - \varepsilon]\}$, где $E_r(R)$ — положительна при $R < C$. Как было показано в § 4.6, P_e нельзя сделать сколь угодно малой (независимо от начального состояния) при $R > C$.

В каналах, которые не являются неразложимыми, вероятность ошибки, достижимая при использовании кодирования (особенно при скоростях между C и $\overline{C}$), в общем случае сильно зависит как от начального состояния, так и от знания передатчиком этого начального состояния. Мы не будем подробно рассматривать какие-либо детали этой последней задачи, так как обычно она поддается изучению при малом изменении модели. Например, если «панический» канал, изображенный на рис. 4.6.5, имеет начальное состояние $s_0 = 0$, то можно просто пренебречь панической буквой (буквой 2) и не использовать ее на входе канала, рассматривая канал как двоичный канал без шума. Аналогично, если известно начальное состояние в канале с переменной фазой, изображенном на рис. 4.6.3, то его модель может быть переделана так, чтобы получилась пара параллельных каналов без памяти.

При доказательстве теоремы кодирования, не зависящей от начального состояния, возникает задача, которая формально совпадает с аналогичной задачей для составного канала. Составным каналом называется канал, который описывается множеством различных переходных распределений вероятностей $P_N^{(i)}(\mathbf{y} \mid \mathbf{x})$, где частное значение i неизвестно передатчику и приемнику. Задача состоит в том, чтобы найти код, который хорошо работает при всех значениях i. В нашем случае получается канал с конечным числом A состояний,и, следовательно, имеются A различных переходных распределений вероятностей, каждое из которых соответствует одному из возможных значений начального состояния s_0. Развитый здесь подход (который применим только, когда A конечно) для простоты использует *предположение*, что начальные состояния имеют равные вероятности. При этом предположении получаем

$$P_N(\mathbf{y}\mid\mathbf{x}) = \sum_{s_0} \frac{1}{A} P_N(\mathbf{y}\mid\mathbf{x}, s_0).$$

Теперь можно применить теорему 5.6.1, выбирая M кодовых слов независимо с распределением вероятности $Q_N(\mathbf{x})$; в результате получим

$$\bar{P}_{e,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) \left[\sum_{s_0} \frac{1}{A} P_N(\mathbf{y}\mid\mathbf{x}, s_0) \right]^{1/(1+\rho)} \right\}^{1+\rho}, \qquad (5.9.1)$$

при любом ρ, $0 \leqslant \rho \leqslant 1$. Ясно, что наиболее точная граница в (5.9.1) может быть получена с помощью минимизации по всем распределениям вероятности на входе $Q_N(\mathbf{x})$.

Используя такие же рассуждения, как и в обсуждении, следующем за теоремой 5.6.2, можно показать, что средняя вероятность ошибки по крайней мере для одного кода из ансамбля, удовлетворяет указанной выше границе и существует также код с данными N и M, для которого $P_{e,m}$, при любом m, ограничена сверху умноженной на четыре правой частью (5.9.1). Наконец, в силу того, что вероятность ошибки для такого кода является средним по A равновероятным состоянием, вероятность ошибки, при условии, что задано какое-либо начальное состояние, может не больше чем в A раз превышать среднее значение. Это дает границу для вероятности ошибки, которая в равной степени справедлива для любого начального состояния и, следовательно, больше не зависит от предположения о равновероятности состояний. Декодер, который рассматривается при выводе этой границы, декодирует сообщение m, для которого максимально значение

$$\sum_{s_0} \frac{1}{A} P_N(\mathbf{y}\mid\mathbf{x}_m, s_0),$$

и он остается хорошо определенным независимо от того, существует или нет вероятностная мера на начальных состояниях. Объединяя предыдущие рассуждения, видим, что при любой длине блока N и любом числе кодовых слов M существует код, такой, что вероятность ошибки для сообщения m при условии, что задано начальное состояние s_0, ограничена, независимо от m и s_0, следующим образом:

$$P_{e,m}(s_0) \leqslant$$
$$\leqslant 4A(M-1)^{\rho} \min_{Q_N} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) \left[\sum_{s_0} \frac{1}{A} P_N(\mathbf{y}\mid\mathbf{x}, s_0) \right]^{1/(1+\rho)} \right\}^{1+\rho} \qquad (5.9.2)$$

при всех ρ, $0 \leqslant \rho \leqslant 1$.

Для упрощения этого выражения удобно сначала вынести сумму по s_0 из квадратных скобок (5.9.2).

Используя неравенства $(\Sigma a_i)^r \leqslant \Sigma a_i^r$ при $0 < r \leqslant 1$ (см. задачу 4.15(e)), находим, что правая часть (5.9.2) будет ограничена сверху следующим выражением:

$$P_{e,m}(s_0) \leqslant$$
$$\leqslant 4A(M-1)^{\rho} \min_{Q_N} \sum_{\mathbf{y}} \left\{ \sum_{s_0} \sum_{\mathbf{x}} Q_N(\mathbf{x}) \left[\frac{1}{A} P_N(\mathbf{y}\mid\mathbf{x}) \right]^{1/(1+\rho)} \right\}^{1+\rho}. \qquad (5.9.3)$$

Умножая и деля сумму по s_0 на A, понимая $1/A$ как распределение вероятностей для s_0 и используя неравенство $(\Sigma P_i a_i)^r \leqslant \Sigma P_i a_i^r$ при $r \geqslant 1$ (см. задачу 4.15(г)), будем иметь

$$P_{e,\,m}(s_0) \leqslant$$

$$\leqslant 4A(M-1)^\rho A^\rho \min_{\mathbf{Q}_N} \sum_{s_0} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) \left[\frac{1}{A} P_N(\mathbf{y} \mid \mathbf{x},\ s_0) \right]^{1/(1+\rho)} \right\}^{1+\rho} \leqslant \tag{5.9.4}$$

$$\leqslant 4A(M-1)^\rho A^\rho \min_{\mathbf{Q}_N} \max_{s_0} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x},\ s_0)^{1/(1+\rho)} \right\}^{1+\rho}, \tag{5.9.5}$$

где сумма по s_0 была ограничена наибольшим слагаемым, умноженным на A.

Порядок, в котором разыскиваются минимум и максимум в (5.9.5), является существенным. Нетрудно заметить (см. задачу 5.37), что, если передатчик знает начальное состояние и может использовать различные коды для каждого начального состояния, то минимакс в (5.9.5) можно заменить на максимин и при этом граница, как правило, уменьшается.

Теорема 5.9.1. В произвольном канале с конечным числом A состояний при любом положительном целом числе N и любой положительной R существует (N, R)-код, у которого для всех сообщений m, $1 \leqslant m \leqslant M = \lceil e^{NR} \rceil$, всех начальных состояний и всех ρ, $0 \leqslant \rho \leqslant 1$,

$$P_{e,\,m}(s_0) \leqslant 4A \exp\{-N[-\rho R + F_N(\rho)]\}, \tag{5.9.6}$$

где

$$F_N(\rho) = -\frac{\rho \ln A}{N} + \max_{\mathbf{Q}_N} \left[\min_{s_0} E_{0,\,N}(\rho,\ \mathbf{Q}_N,\ s_0) \right], \tag{5.9.7}$$

$$E_{0,N}(\rho,\ \mathbf{Q}_N,\ s_0) = -\frac{1}{N} \ln \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x},\ s_0)^{1/(1+\rho)} \right\}^{1+\rho}. \tag{5.9.8}$$

Доказательство. Подставляя (5.9.7) и (5.9.8) в (5.9.6), видим, что (5.9.6) будет совпадать с (5.9.5), если $(M-1)$ заменить на большую величину e^{NR}.

Хотя это не понадобится нам в дальнейшем, следует отметить, что эта теорема в равной степени применима для любого составного канала с A состояниями.

Граница в (5.9.6) имеет экспоненциальный вид и теперь будет установлено, что $F_N(\rho)$ стремится к постоянной величине при $N \to \infty$. В процессе доказательства этого станет ясно, почему удобно было включить выражение $-\rho(\ln A)/N$ в определение функции $F_N(\rho)$.

Лемма 5.9.1. В любом заданном канале с конечным числом состояний функция $F_N(\rho)$, определенная равенством (5.9.7), удовлетворяет неравенству

$$F_N(\rho) \geqslant \frac{n}{N} F_n(\rho) + \frac{l}{N} F_l(\rho) \tag{5.9.9}$$

при всех положительных целых числах n и l, таких, что $N = n + l$.

Доказательство. Разобьем последовательность на входе $\mathbf{x} = (x_1, \ldots, x_N)$ на подпоследовательности $\mathbf{x}_1 = (x_1, \ldots, x_n)$ и $\mathbf{x}_2 = (x_{n+1}, \ldots, x_N)$ и разобьем $\mathbf{y} = (y_1, \ldots, y_N)$ на $\mathbf{y}_1 = (y_1, \ldots, y_n)$ и $\mathbf{y}_2 = (y_{n+1}, \ldots, y_N)$. Пусть $\mathbf{Q}_n$ и $\mathbf{Q}_l$ при заданном ρ будут распределениями вероятностей, на которых достигаются максимумы $F_n(\rho)$ и $F_l(\rho)$, соответственно; рассмотрим распределение вероятностей для $\mathbf{x} = (x_1, \ldots, x_N)$, задаваемое равенством

$$Q_N(\mathbf{x}) = Q_n(\mathbf{x}_1)\, Q_l(\mathbf{x}_2). \tag{5.9.10}$$

Пусть, наконец, s_0' является начальным состоянием, на котором достигается минимум $E_{0,N}(\rho, \mathbf{Q}_N, s_0)$. Тогда

$$F_N(\rho) > \frac{-\rho \ln A}{N} + E_{0,N}(\rho, \mathbf{Q}_N, s_0')$$

и имеем

$$\exp[-NF_N(\rho)] \leqslant$$

$$\leqslant A^\rho \sum_{\mathbf{y}} \Big\{ \sum_{\mathbf{x}} Q_N(\mathbf{x})\, P_N(\mathbf{y} \mid \mathbf{x}, s_0')^{1/(1+\rho)} \Big\}^{1+\rho} =$$

$$= A^\rho \sum_{\mathbf{y}_1} \sum_{\mathbf{y}_2} \Big\{ \sum_{\mathbf{x}_1} \sum_{\mathbf{x}_2} Q_n(\mathbf{x}_1)\, Q_l(\mathbf{x}_2) \Big[\sum_{s_n} P_n(\mathbf{y}_1, s_n \mid \mathbf{x}_1, s_0')\, P_l(\mathbf{y}_2 \mid \mathbf{x}_2, s_n) \Big]^{1/(1+\rho)} \Big\}^{1+\rho} \leqslant \tag{5.9.11}$$

$$\leqslant A^{2\rho} \sum_{s_n} \sum_{\mathbf{y}_1} \sum_{\mathbf{y}_2} \Big\{ \sum_{\mathbf{x}_1} \sum_{\mathbf{x}_2} Q_n(\mathbf{x}_1)\, Q_l(\mathbf{x}_2)\, [P_n(\mathbf{y}_1, s_n \mid \mathbf{x}_1, s_0')\, P_l(\mathbf{y}_2 \mid \mathbf{x}_2, s_n)]^{1/(1+\rho)} \Big\}^{1+\rho}. \tag{5.9.12}$$

При переходе от (5.9.11) и (5.9.12) были использованы те же неравенства для суммы по s_n, которые были использованы при переходе от (5.9.2) к (5.9.4). Преобразуя суммы [как при переходе от (5.5.7) к (5.5.9)], получаем

$$\exp[-NF_N(\rho)] \leqslant A^{2\rho} \sum_{s_n} \Big\{ \sum_{\mathbf{y}_1} \Big[\sum_{\mathbf{x}_1} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}_1, s_n \mid \mathbf{x}_1, s_0')^{1/(1+\rho)} \Big]^{1+\rho} \Big\} \times$$

$$\times \Big\{ \sum_{\mathbf{y}_2} \Big[\sum_{\mathbf{x}_2} Q_l(\mathbf{x}_2)\, P_l(\mathbf{y}_2 \mid \mathbf{x}_2, s_n)^{1/(1+\rho)} \Big]^{1+\rho} \Big\} \leqslant \tag{5.9.13}$$

$$\leqslant A^\rho \sum_{s_n} \sum_{\mathbf{y}_1} \Big[\sum_{\mathbf{x}_1} Q_n(\mathbf{x}_1)\, P_n(\mathbf{y}_1, s_n \mid \mathbf{x}_1, s_0')^{1/(1+\rho)} \Big]^{1+\rho} \exp[-lF_l(\rho)], \tag{5.9.14}$$

где последнее выражение в фигурных скобках в (5.9.13) было ограничено сверху максимальным значением этого выражения по s_n. Используя, наконец, неравенство Минковского (см. задачу 4.15 (з)) при изменении порядка суммирования по s_n и $\mathbf{x}_1$, будем иметь

$$\exp[-NF_N(\rho)] \leqslant$$

$$\leqslant A^\rho \sum_{\mathbf{y}_1} \Big\{ \sum_{\mathbf{x}_1} Q_n(\mathbf{x}_1) \Big[\sum_{s_n} P_n(\mathbf{y}_1, s_n \mid \mathbf{x}_1, s_0') \Big]^{1/(1+\rho)} \Big\}^{1+\rho} \times$$

$$\times \exp[-lF_l(\rho)] \leqslant \exp[-nF_n(\rho) - lF_l(\rho)], \tag{5.9.15}$$

где было проведено суммирование по s_n и затем найден максимум по начальному состоянию s_0'. Преобразуя (5.9.15), получаем (5.9.9), что завершает доказательство. |

Лемма 5.9.2. Пусть $F_\infty(\rho) = \sup_N F_N(\rho)$. Тогда

$$\lim_{N\to\infty} F_N(\rho) = F_\infty(\rho). \tag{5.9.16}$$

При $0 \leqslant \rho \leqslant 1$ сходимость является равномерной по ρ и $F_\infty(\rho)$ равномерно непрерывна.

Доказательство. Применяя теорему 5.6.3 с использованием $Q_N(\mathbf{x})$ вместо $Q(k)$ и $P_N(\mathbf{y}\,|\,\mathbf{x}, s_0)$ вместо $P(j\,|\,k)$, будем иметь

$$0 \leqslant \frac{\partial N E_0(\rho, \mathbf{Q}_N, s_0)}{\partial \rho} \leqslant \mathcal{J}(\mathbf{Q}_N; \mathbf{P}_N). \tag{5.9.17}$$

При входном алфавите с K буквами имеются K^N последовательностей на входе длины N и поэтому (5.9.17) может быть далее оценена следующим образом:

$$0 \leqslant \frac{\partial E_0(\rho, \mathbf{Q}_N, s_0)}{\partial \rho} \leqslant \log K. \tag{5.9.18}$$

Отсюда и из (5.9.7) при любых $0 \leqslant \rho_1 < \rho_2 \leqslant 1$ выводим

$$\frac{-(\rho_2-\rho_1)\ln A}{N} \leqslant F_N(\rho_2) - F_N(\rho_1) < (\rho_2-\rho_1)\log K. \tag{5.9.19}$$

Из этого, в частности, следует, что при любом ρ, $0 \leqslant \rho \leqslant 1$ функция $F_N(\rho)$ ограничена выражением, не зависящим от N. Таким образом, сочетая лемму 5.9.1 и лемму 2 приложения 4 А, получаем (5.9.16). Равномерная сходимость и равномерная непрерывность вытекают из того, что, как показывает (5.9.19), наклон $F_N(\rho)$ является ограниченным при всех N.]

Теорема 5.9.2. Пусть задан произвольный канал с конечным числом состояний и пусть

$$E_r(R) = \max_{0\leqslant\rho\leqslant 1} [F_\infty(\rho) - \rho R]. \tag{5.9.20}$$

Тогда при любом $\varepsilon > 0$ существует $N(\varepsilon)$, такое, что для всех $N \geqslant N(\varepsilon)$ и любых $R \geqslant 0$ существует (N, R)-код, такой, что для всех m, $1 \leqslant m \leqslant M = \lceil e^{NR} \rceil$, и всех начальных состояниях

$$P_{e,m}(s_0) \leqslant \exp\{-N[E_r(R) - \varepsilon]\}. \tag{5.9.21}$$

Более того, при $0 \leqslant R < C$ функция $E_r(R)$ является строго положительной строго убывающей с ростом R и выпуклой $\smile$.

Обсуждение. Эта теорема устанавливает экспоненциальную границу для вероятности ошибки при всех $R < C$, где C определяется согласно (4.6.3) и (4.6.4). Кажется правдоподобным, что $E_r(R)$ является надежностью канала при R, близких к C, но доказательство этого существует только в частных случаях. Эта теорема в какой-то степени

является более слабой, чем соответствующая теорема для дискретного канала без памяти, так как (5.9.21) выполняется лишь при $N \geqslant N(\varepsilon)$ и мало известно о зависимости $N(\varepsilon)$ от канала. Наконец, мало известно о том, как найти $F_\infty(\rho)$ [и, следовательно, $E_r(R)$], кроме некоторых частных случаев, наиболее важные из которых будут рассмотрены ниже. Однако функция $F_\infty(\rho)$ всегда может быть оценена снизу следующим образом:

$$F_\infty(\rho) \geqslant \frac{-\rho \ln A}{N} + \min_{s_0} E_{0,N}(\rho, \mathbf{Q}_N, s_0) \tag{5.9.22}$$

при любых N и $\mathbf{Q}_N$.

Доказательство. При любых N и R неравенство (5.9.6) можно переписать в виде

$$P_{e,m}(s_0) \leqslant \exp\left\{-N\left[-\rho R + F_N(\rho) - \frac{\ln 4A}{N}\right]\right\}. \tag{5.9.23}$$

Лемма 5.9.2 утверждает, что при любом $\varepsilon > 0$ можно выбрать $N(\varepsilon)$ так, чтобы при $N \geqslant N(\varepsilon)$

$$F_\infty(\rho) - F_N(\rho) + \frac{\ln 4A}{N} \leqslant \varepsilon, \quad 0 \leqslant \rho \leqslant 1. \tag{5.9.24}$$

Подставляя (5.9.24) в (5.9.23), получаем

$$P_{e,m}(s_0) \leqslant \exp\{-N[-\rho R + F_\infty(\rho) - \varepsilon]\}, \quad 0 \leqslant \rho \leqslant 1. \tag{5.9.25}$$

При ρ, на котором достигается максимум $-\rho R + F_\infty(\rho)$, граница (5.9.25) сводится к (5.9.21). Предположим теперь, что $R < \underline{C}$ и покажем, что $E_r(R) > 0$. Положим δ равным $\underline{C} - R$. Согласно теореме 4.6.1, N можно выбрать достаточно большим, так, чтобы

$$R + \frac{\ln A}{N} < \underline{C}_N - \frac{\delta}{2}. \tag{5.9.26}$$

Пусть $\mathbf{Q}_N$ при таком N будет распределением на входе, на котором достигается $\underline{C}_N$. Тогда согласно теореме 5.6.3 имеем

$$\left.\frac{\partial E_{0,N}(\rho, \mathbf{Q}_N, s_0)}{\partial \rho}\right|_{\rho=0} \geqslant \underline{C}_N \text{ при всех } s_0. \tag{5.9.27}$$

Так как $\partial E_{0,N}(\rho, \mathbf{Q}_N, s_0)/\partial\rho$ является непрерывной функцией ρ, то при любом s_0 существует интервал значений $\rho > 0$, такой, что

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) - \rho\left(R + \frac{\ln A}{N}\right) > 0. \tag{5.9.28}$$

В силу того, что имеется конечное число начальных состояний s_0, можно выбрать $\rho^* > 0$ достаточно малым, чтобы

$$E_{0,N}(\rho^*, \mathbf{Q}_N, s_0) - \rho^*\left(R + \frac{\ln A}{N}\right) > 0 \text{ при всех } s_0. \tag{5.9.29}$$

Но, так как

$$F_\infty(\rho^*) > F_N(\rho^*) > F_{0,N}(\rho^*,\ \mathbf{Q}_N,\ s_0) - \frac{\rho^* \ln A}{N}$$

при любых s_0, то это означает, что

$$F_\infty(\rho^*) - \rho^* R > 0, \tag{5.9.30}$$

и, следовательно, что $E_r(R) > 0$ при всех $R < C$. Выпуклость $E_r(R)$ можно установить, если заметить, что $E_r(R)$ является верхней гранью множества прямых линий $F_\infty(\rho) - \rho R$. Наконец, в силу того, что $F_\infty(\rho) - \rho R$ равна нулю при $\rho = 0$, получаем, что при $R < C$ функция $F_\infty(\rho) - \rho R$ достигает максимума на интервале $0 \leqslant \rho \leqslant 1$ при некотором $\rho > 0$. Любое уменьшение R при этом фиксированном ρ приводит к увеличению $F_\infty(\rho) - \rho R$ и, следовательно, $E_r(R)$ является строго убывающей функцией R для $R < C$. |

Состояние известно на приемном конце

Рассмотрим теперь частный случай полученных выше результатов, в котором состояние в момент n является детерминированной функцией выхода в момент n и состояния в момент $n-1$, т. е. $s_n = g(y_n, s_{n-1})$. В этом случае приемник способен проследить за состоянием канала, если оно было известно когда-либо в прошлом. Пример такого канала представлен на рис. 5.9.1 и в этом примере s_n является функцией только y_n. В каждом состоянии канал является ДСК; выходы 0 и 1 соответствуют входу 0, а выходы 2 и 3 соответствуют входу 1. Выходы 0 и 3 соответствуют состоянию 0, а выходы 1 и 2 соответствуют состоянию 1. Можно увидеть, если не обращать внимания на числа, что эта модель является моделью того самого типа, что и модель, представленная на рис. 4.6.1, за исключением того, что здесь выходной алфавит был расширен для того, чтобы учесть предположение о том, что состояние канала является известным. Эта модель является простой и довольно грубой аппроксимацией канала с замираниями, в котором используется алфавит из двух сигналов на входе и в котором приемник не только строит решения о том, что было на входе, но также измеряет уровень принятого сигнала.

Чтобы исследовать этот класс каналов, заметим, что последовательность на выходе $\mathbf{y} = (y_1, \ldots, y_N)$ и начальное состояние s_0 однозначно определяют последовательность состояний $\mathbf{s} = (s_1, \ldots, s_N)$, для которой примем обозначения $\mathbf{s}(\mathbf{y}, s_0)$. Теперь имеем

$$P_N(\mathbf{y},\ \mathbf{s} \mid \mathbf{x},\ s_0) = \begin{cases} P_N(\mathbf{y}, \mid \mathbf{x},\ s_0) \text{ при } \mathbf{s} = \mathbf{s}(\mathbf{y},\ s_0), \\ 0 \text{ во всех остальных случаях,} \end{cases} \tag{5.9.31}$$

$$F_{0,N}(\rho,\ \mathbf{Q}_N,\ s_0) =$$

$$= -\frac{1}{N} \ln \sum_{\mathbf{y}} \sum_{\mathbf{s}} \left\{ \sum_{\mathbf{x}} \mathbf{Q}_N(\mathbf{x})\, P_N(\mathbf{y},\ \mathbf{s} \mid \mathbf{x},\ s_0)^{1/(1+\rho)} \right\}^{1+\rho}. \tag{5.9.32}$$

Для того чтобы доказать (5.9.32), заметим, что для любого $\mathbf{y}$ сумма по $\mathbf{s}$ не равна нулю лишь в случае, когда $\mathbf{s} = \mathbf{s}(\mathbf{y}, s_0)$ и для этого значения $\mathbf{s}$ имеем $P_N(\mathbf{y}, \mathbf{s} | \mathbf{x}, s_0) = P_N(\mathbf{y} | \mathbf{x}, s_0)$. Таким образом, (5.9.32) эквивалентно определению $E_{0,N}(\rho, \mathbf{Q}_N, s_0)$ в (5.9.8).

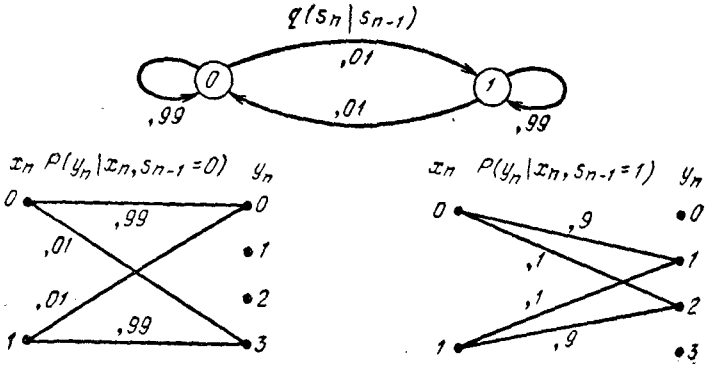


Рис. 5.9.1. Простая модель канала с замираниями.

Предположим теперь, что $Q_N(\mathbf{x})$ представляет собой произведение мер (т. е. последовательные буквы в ансамбле кодовых слоев выбираются независимо):

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n). \tag{5.9.33}$$

В этом случае (5.9.32) приводится к виду

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) =$$

$$= -\frac{1}{N} \ln \sum_{\mathbf{s}} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} \prod_{n=1}^{N} Q(x_n) P(y_n, s_n | x_n, s_{n-1})^{1/(1+\rho)} \right\}^{1+\rho} = \tag{5.9.34}$$

$$= -\frac{1}{N} \ln \sum_{\mathbf{s}} \prod_{n=1}^{N} \sum_{j=0}^{J-1} \left\{ \sum_{k=0}^{K-1} Q(k) P(j, s_n | k, s_{n-1})^{1/(1+\rho)} \right\}^{1+\rho}, \tag{5.9.35}$$

где был изменен порядок взятия произведения и суммирования, так же как и при переходе от (5.5.6) к (5.5.10). Если для заданных ρ и $\mathbf{Q}$ определить

$$\alpha(s_{n-1}, s_n) = \sum_{j=0}^{J-1} \left\{ \sum_{k=0}^{K-1} Q(k) P(j, s_n | k, s_{n-1})^{1/(1+\rho)} \right\}^{1+\rho}, \tag{5.9.36}$$

то будем иметь

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) = -\frac{1}{N} \ln \sum_{\mathbf{s}} \prod_{n=1}^{N} \alpha(s_{n-1}, s_n). \tag{5.9.37}$$

Определим теперь матрицу $A \times A$

$$[\alpha] = \begin{Bmatrix} \alpha(0,\ 0) & \dots \alpha(0,\ A-1) \\ \vdots & \vdots \\ \alpha(A-1,0) \dots \alpha(A-1,\ A-1) \end{Bmatrix}. \tag{5.9.38}$$

Обозначим через $1]$ A-мерный вектор-столбец, состоящий из одних единиц, и обозначим через $\underline{e(s_0)}$ A-мерный единичный вектор-строку с 1 в позиции, соответствующей s_0, т. е. в первой позиции для $s_0 = 0$ и в i-й позиции для $s_0 = i - 1$. Немного подумав, можно заметить, что (5.9.37) можно записать в виде

$$E_{0,N}(\rho,\ \mathbf{Q}_N,\ s_0) = -\frac{1}{N} \ln \underline{e(s_0)}\, [\alpha]^N\, 1]. \tag{5.9.39}$$

Квадратная матрица $[\alpha]$ называется *неприводимой*, если с помощью перестановки соответствующих столбцов и строк (т. е. с помощью перенумерации состояний) невозможно привести ее к виду

$$\left[\begin{array}{c|c} \alpha_1 & 0 \\ \hline \alpha_2 & \alpha_3 \end{array}\right],$$

где α_1 и α_3 — квадратные матрицы. Без труда можно показать, что $[\alpha]$ является неприводимой, тогда и только тогда, когда можно с ненулевой вероятностью достигнуть любое состояние из любого другого состояния за конечное число шагов при использовании $Q(k)$ в качестве распределения на входе. Если $[\alpha]$ неприводима, то можно применить теорему Фробениуса*), которая утверждает, что неприводимая ненулевая матрица с неотрицательными элементами имеет наибольшее положительное собственное значение λ, соответствующее правому собственному вектору $v]$ и левому собственному вектору u, которые имеют положительные компоненты. Т. е. считая, что $v]$ является вектором-столбцом, имеем

$$[\alpha]\ v] = \lambda\ v]. \tag{5.9.40}$$

Л е м м а 5.9.3. Пусть λ будет наибольшим собственным значением неприводимой матрицы $[\alpha]$ и пусть v_{max} и v_{min} будут наибольшей и наименьшей компонентами положительного правого собственного вектора $v]$, соответствующего λ.

Тогда при любом s_0

$$\frac{v_{min}}{v_{max}} \lambda^N \leqslant \underline{e(s_0)}\, [\alpha]^N\, 1] < \frac{v_{max}}{v_{min}} \lambda^N. \tag{5.9.41}$$

Доказательство. Согласно (5.9.40) имеем

$$[\alpha]^N\, v] = [\alpha]^{N-1}\, [\alpha]\, v] = \lambda\, [\alpha]^{N-1}\, v] =$$
$$= \lambda^2\, [\alpha]^{N-2}\, v] = \dots = \lambda^N\, v]. \tag{5.9.42}$$

*) См., например, Гантмахер (1967).

Так как $[\alpha]$ имеет неотрицательные элементы, то компоненты вектора-строки $\underbrace{e(s_0)}\,[\alpha]^N$ являются неотрицательными и поэтому $\underbrace{e(s_0)}\,[\alpha]^N 1]$ можно оценить сверху, оценивая сверху компоненты $1]$ в рассматриваемом случае величиной $(1/v_{min})\,v]$. Отсюда

$$\underbrace{e(s_0)}\,[\alpha]^N 1] \leqslant \frac{1}{v_{min}} \underbrace{e(s_0)}\,[\alpha]^N v] =$$

$$= \frac{\lambda^N}{v_{min}} \underbrace{e(s_0)}\, v] \leqslant \frac{v_{max}}{v_{min}} \lambda^N. \qquad (5.9.43)$$

Поступая аналогично, можно завершить доказательство неравенством

$$\underbrace{e(s_0)}\,[\alpha]^N 1] \geqslant \frac{1}{v_{max}} \underbrace{e(s_0)}\,[\alpha]^N v] \geqslant \frac{v_{min}}{v_{max}} \lambda^N. \,|$$

Подставляя (5.9.41) в (5.9.39) и используя обозначение $\lambda(\rho, \mathbf{Q})$ для того, чтобы отметить зависимость λ от ρ и $\mathbf{Q}$, получаем,

$$|E_{0,N}(\rho, \mathbf{Q}_N, s_0) + \ln \lambda(\rho, \mathbf{Q})| \leqslant \frac{1}{N} \ln \frac{v_{max}}{v_{min}}, \qquad (5.9.44)$$

где правая часть (5.9.44) не зависит от s_0.
Таким образом, доказана следующая теорема.

Теорема 5.9.3. Пусть задан канал с конечным числом состояний, в котором $s_n = g(y_n, s_{n-1})$, и пусть $\mathbf{Q}$ является распределением вероятностей, таким, что, когда входы выбираются независимо с распределением вероятностей $\mathbf{Q}$, любое состояние может быть достигнуто с ненулевой вероятностью из любого другого состояния за конечное число шагов. Тогда при любой $R > 0$ и любом положительном целом N существуют (N, R)-коды, такие, что при любом сообщении m, $1 \leqslant m \leqslant \lceil e^{NR} \rceil$, любом s_0 и всех ρ, $0 \leqslant \rho \leqslant 1$,

$$P_{e,m}(s_0) \leqslant 4A \frac{v_{max}}{v_{min}} \exp\{-N[-\ln \lambda(\rho, \mathbf{Q}) - \rho R]\}, \qquad (5.9.45)$$

где $\lambda(\rho, \mathbf{Q})$ является наибольшим собственным значением матрицы $[\alpha]$, заданной (5.9.38) и (5.9.36), а v_{max} и v_{min} являются экстремальными компонентами положительного правого собственного вектора, соответствующего $\lambda(\rho, \mathbf{Q})$.

Граница в (5.9.45), конечно, может быть оптимизирована по $\mathbf{Q}$ и ρ с целью получить более точную границу. К сожалению, в общем случае эта граница слабее, чем граница, представленная теоремой 5.9.2, так как здесь мы ограничились использованием ансамблей случайных кодов, в которых буквы кодовых слов выбираются независимо. Однако имеется важный класс каналов, для которого граница оптимизируется на независимо выбранных буквах, и в этом случае распределение $\mathbf{Q}$, которое минимизирует $\alpha(s_{n-1}, s_n)$ в (5.9.36), не зависит от значений s_{n-1} и s_n. Чтобы увидеть это, фиксируем s_0 и $\mathbf{s} = (s_1, \ldots, s_N)$ и рассмотрим минимум выражения

$$\sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) \prod_{n=1}^{N} P(y_n, s_n | x_n, s_{n-1})^{1/(1+\rho)} \right\}^{1+\rho}$$

по $Q_N(\mathbf{x})$. Используя такие же рассуждения, как в примере 4 § 4.6 с параллельными каналами, можно показать, что минимум достигается на произведении распределений

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q^{(n)}(x_n),$$

где при каждом n на распределении $Q^{(n)}(x_n)$ достигается минимум выражения

$$\sum_{y_n} \left\{ \sum_{x_n} Q^{(n)}(x_n) P(y_n, s_n \mid x_n, s_{n-1})^{1/(1+\rho)} \right\}^{1+\rho}.$$

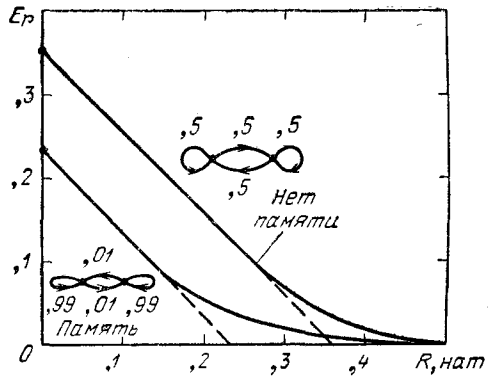


Рис. 5.9.2. Двоичный канал с двумя состояниями; состояние известно на приемном конце.

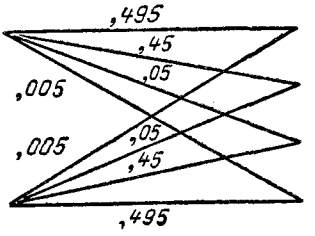


Рис. 5.9.3. Канал, изображенный на рис. 5.9.1, у которого устранена память.

Если то же самое $\mathbf{Q}$ минимизирует $\alpha = (s_{n-1}, s_n)$ при любых s_{n-1} и s_n, то $Q^{(n)}$ не зависит от n и не зависит также от $\mathbf{s}$ и s_0. Следовательно

$$Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n)$$

минимизирует указанное выше выражение при всех $\mathbf{s}$ и s_0 и, таким образом, оно максимизирует $E_0(\rho, \mathbf{Q}_N, s_0)$ при всех s_0. В этом случае $F_\infty(\rho)$, заданное равенством (5.9.16), равно $-\ln \lambda(\rho, \mathbf{Q})$ при этом минимизирующем значении $\mathbf{Q}$. Хотя отыскание $\lambda(\rho, \mathbf{Q})$ при этом минимизирующем значении является нетривиальной задачей, оно, по крайней мере, не зависит от длины блока.

Пример, изображенный на рис. 5.9.1, дает канал из указанного выше класса и для него почти очевидно, что входы нужно использовать независимо и с равными вероятностями в ансамбле кодов. Для этого канала $E_r(R)$ изображена на рис. 5.9.2. Для сравнения на нем также изображена $E_r(R)$ для канала без памяти, представленного на рис. 5.9.3. Этот канал эквивалентен каналу, представленному на рис. 5.9.1, с вероятностями $q(s_n \mid s_{n-1})$, равными 1/2 при любых s_{n-1} и s_n или, говоря более наглядно, этот канал является каналом

рис. 5.9.1, в котором устранена память. Заметим, что пропускная способность C не изменяется при устранении памяти (см. задачу 5.39), однако показатель экспоненты возрастает. Это может быть качественно объяснено тем, что среднее время, проводимое в каждом состоянии, не изменяется при устранении памяти, а вероятность пребывания в плохом состоянии (состоянии 1) значительно дольше, чем среднее время, существенно уменьшается при устранении памяти. Например, в канале, представленном на рис. 5.9.3, при $N = 100$ вероятность пребывания в плохом состоянии в течение всего блока приближенно равна $1/(2e)$ при наличии памяти и равна 2^{-100} при отсутствии памяти.

В каналах, в которых состояние не известно на приемном конце, имеется другой качественный эффект, обязанный долгодействующей памяти. Приемник может оценить состояние по выходам канала, используя знание кода. Это увеличивает пропускную способность канала по сравнению с той, которая имеет место при отсутствии памяти (см. задачу 5.38).

ИТОГИ И ВЫВОДЫ

Основным результатом этой главы была теорема кодирования для канала с шумами. Вначале была изучена простая задача проверки гипотез, в которой одно из двух кодовых слов передавалось по дискретному каналу без памяти, и затем был изучен случай большого числа кодовых слов. Для случая большого числа кодовых слов было установлено, что основная трудность состоит в том, что не ясно, как выбирать кодовые слова. Решение этой проблемы было найдено с помощью построения верхней границы для средней по ансамблю кодов вероятности ошибки и последующего указания на то, что, по крайней мере, один код из ансамбля должен иметь столь же малую вероятность ошибки, как и средняя вероятность. При исследовании этой верхней границы было найдено, что при любой скорости R, меньшей, чем пропускная способность канала, существуют коды с любой длиной блока N, для которых $P_e \leqslant \exp[-NE_r(R)]$. Скорость R здесь понимается как умноженное на $\ln 2$ число двоичных символов, поступающих на кодер за время передачи одного символа в канале. Также была найдена более точная граница вероятности ошибки при малых R. Затем было установлено, что имеется нижняя граница для вероятности ошибки наилучших кодов, с заданными N и R, для которой вероятность ошибки убывает экспоненциально по N, и было показано, что показатели этих экспонент совпадают при скоростях, близких к пропускной способности, и в пределе при $R \to 0$. Нижние границы были выведены только для случая двоичного симметричного канала. Наконец, было показано, что каналы с конечным числом состояний имеют того же типа экспоненциальную верхнюю границу для вероятности ошибки. Ни один из полученных здесь результатов не дает прямого указания на то, как строить кодеры и декодеры; это является предметом следующей главы. В гл. 7 и 8 полученные здесь результаты будут распространены на недискретные каналы.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Теорема кодирования для канала с шумами принадлежит Шеннону (1948) и, несомненно, является самым значительным результатом в теории информации. Первое строгое доказательство этой теоремы для дискретных каналов без памяти было дано Файнстейном (1954) и вскоре после этого более простые доказательства были предложены Шенноном (1957) и Вольфовицем (1957). Впервые Файнстейн (1955) показал, что P_e стремится к нулю экспоненциально по N при фиксированной скорости $R < C$. Граница случайного кодирования, граница сферической упаковки и тот факт, что они экспоненциально совпадают при скоростях, близких к пропускной способности, были впервые получены Элайсом (1955) в частных случаях двоичного симметричного канала и двоичного канала со стиранием. Фано (1961) использовал методы случайного кодирования, развитые Шенноном, и производящих функций моментов, для получения показателя экспоненты случайного кодирования $E_r(R)$ и для эвристического вывода границы сферической упаковки для общего дискретного канала без памяти. Граница случайного кодирования для процедуры с выбрасыванием и большинство свойств показателя экспоненты случайного кодирования $E_r(R)$ были получены Галлагером (1965). Нижние границы для вероятности ошибки (рассмотренные в § 5.8) в дискретном канале без памяти были получены Шенноном, Галлагером и Берлекэмпом (1967). Теорема кодирования для каналов с конечным числом состояний была впервые доказана Галлагером (1958) при, в некотором отношении, более жестких условиях, а в более сильной форме она была доказана Блекуэллом, Брейманом и Томасяном (1958). Показатель экспоненты случайного кодирования для каналов с конечным числом состояний и его исследование, проведенное в § 5.9, принадлежат Юдкину (1967). Теорема 5.9.3 была доказана Галлагером (1964). Единственная граница сферической упаковки, которая пока что получена, для каналов с конечным числом состояний, принадлежит Кеннеди (1963) и относится к одному классу двоичных каналов.

ПРИЛОЖЕНИЕ 5А

Пусть

$$w = \sum_{n=1}^{N} z_n$$

является суммой N дискретных независимых одинаково распределенных случайных величин. Семиинвариантная пронзводящая функция моментов каждой случайной величины z_n определяется с помощью распределения вероятности $P_z(z_n)$ равенством

$$\mu(s) = \ln g(s) = \ln \sum_z P_z(z)\, e^{sz*)}. \qquad (5A.1)$$

Предположим, что $\mu(s)$ существует на открытом интервале действительных значений s вокруг $s = 0$. Если выборочные значения для z_n ограничены, то ясно,

*) P_z — распределение вероятностей случайной величины z. (*Прим. ред.*).

что это условие выполняется. Первые две производные $\mu(s)$ задаются равенствами

$$\mu'(s) = \frac{\sum_z zP_z(z)\,e^{sz}}{\sum_z P_z(z)\,e^{sz}}, \tag{5A.2}$$

$$\mu''(s) = \frac{\sum_z z^2P_z(z)\,e^{sz}}{\sum_z P_z(z)\,e^{sz}} - [\mu'(s)]^2. \tag{5A.3}$$

Отметим, что $\mu'(0)$ и $\mu''(0)$ являются соответственно средним значением и дисперсией z_n.

Пусть $\mu_w(s)$ является семиинвариантной производящей функцией моментов суммы w, т. е.

$$\mu_w(s) = \ln g_w(s) = \ln \sum_w P_w(w)\,e^{sw}. \tag{5A.4}$$

Согласно (5.4.19) имеем

$$\mu_w(s) = N\mu(s). \tag{5A.5}$$

Для того чтобы оценить $\Pr(w \geqslant A)$, где $A \gg \overline{w}$, определим новую сумму случайных величин, называемых перекошенными случайными величинами, распределения вероятностей которых связаны с P_z, но для которых среднее значение суммы равно A. Затем мы применим к этой сумме перекошенных случайных величин центральную предельную теорему.[14]

Для любого заданного s и открытого интервала, в котором существует $\mu(s)$, определим скошенные случайные величины $z_{n,s}$, которые будут принимать те же самые значения, что и z_n, но с распределением вероятностей

$$Q_{z,s}(z) = \frac{P_z(z)\,e^{sz}}{\sum_z P_z(z)\,e^{sz}} = P_z(z)\,e^{sz-\mu(s)}. \tag{5A.6}$$

Из (5A.2) и (5A.3) видно, что $\mu'(s)$ и $\mu''(s)$ соответственно являются средним значением и дисперсией перекошенных случайных величин $z_{n,s}$. Отсюда следует, что $\mu''(s)$ положительна (за исключением тривиальных случайных величин, которые принимают одно-единственное значение с вероятностью 1). Таким образом, $\mu'(s)$ является строго возрастающей функцией. Из (5A.2) можно увидеть, что

$$\lim_{s\to-\infty} \mu'(s)$$

является наименьшим значением, принимаемым z, и

$$\lim_{s\to+\infty} \mu'(s)$$

является наибольшим значением.

Предположим теперь, что перекошенные случайные величины $z_{n,s}$ являются статистически независимыми, и определим перекошенную сумму w_s следующим образом:

$$w_s = \sum_{n=1}^{N} z_{n,s}. \tag{5A.7}$$

Среднее и дисперсия w_s даются равенствами

$$\overline{w}_s = N\mu'(s); \qquad \mathbf{D}(w_s) = N\mu''(s). \tag{5A.8}$$

Свяжем далее распределение вероятностей для w_s, которое будем обозначать через $Q_{w,\,s}$, с распределением вероятностей P_w первоначальной суммы. Вероятность любой заданной последовательности значений перекошенных случайных величин определяется равенством

$$\prod_{n=1}^{N} P_z(z_{n,\,s}) \exp\left[s z_{n,\,s} - \mu(s)\right].$$

Поэтому

$$Q_{w,\,s}(w_s) = \sum_{z_{1,\,s}} \cdots \sum_{z_{N,\,s}} \prod_{n=1}^{N} \left[P_z(z_{n,\,s})\, e^{s z_{n,\,s} - \mu(s)}\right],$$

где суммирование производится по тем z_{1s}, ..., $z_{N,\,s}$, которые удовлетворяют условию $\Sigma z_{n,\,s} = w_s$. Далее имеем

$$Q_{w,\,s}(w_s) = \sum_{z_{1,\,s}} \cdots \sum_{z_{N,\,s}} \prod_{n} [P_z(z_{n,\,s})]\, e^{s w_s - N\mu(s)}$$

той же самой об ластью суммирования по $z_{n,\,s}$. Теперь получаем

$$Q_{w,\,s}(w_s) = P_w(w_s) \exp\left[s w_s - N\mu(s)\right]. \tag{5A.9}$$

Заметим, что $Q_{w,\,s}$ перекошено по отношению к P_w в том же самом смысле, как $Q_{z,\,s}$ перекошено по отношению к P_z.

Если нужно найти $\Pr(w \geqslant A)$ для $A > \overline{w}$, то выберем такое однозначно определяемое значение s, для которого

$$N\mu'(s) = A. \tag{5A.10}$$

В силу того, что $\mu'(s)$ является возрастающей функцией, то s, удовлетворяющее (5A.10), должно быть больше, чем 0. Используя (5A.9), теперь получаем

$$\Pr\left[w \geqslant N\mu'(s)\right] = \sum_{w \geqslant N\mu'(s)} P_w(w) = \sum_{w_s \geqslant N\mu'(s)} Q_{w,\,s}(w_s)\, e^{-s w_s + N\mu(s)} =$$

$$= e^{N[\mu(s) - s\mu'(s)]} \sum_{w_s \geqslant N\mu'(s)} Q_{w,\,s}(w_s)\, e^{-s[w_s - N\mu'(s)]}. \tag{5A.11}$$

Отметим, что суммирование в (5A.11) проводится, начиная со среднего значения w_s, и что экспоненциально убывающий множитель по существу обрывает сумму для больших w_s. Фактически, так как стандартное отклонение w_s пропорционально $\sqrt{N}$ и так как скорость экспоненциального убывания не зависит от N, то представляет интерес только $Q_{w,\,s}$ в области, которая составляет малую долю стандартного отклонения при большом N. Здесь можно использовать центральную предельную теорему для оценки $Q_{w,\,s}$ в той форме, которая чувствительна к малым изменениям w_s. Какую теорему следует применить, зависит от того, является $z_{n,\,s}$ решетчатой случайной величиной или нет. Решетчатая случайная величина является случайной величиной, у которой принимаемые значения могут быть выражены в виде $\alpha + ih$, где α и h — заданные постоянные, а i является целым числом, которое изменяется с изменением выборочных значений случайной величины. Например, значения 0, 1 и 2 могут приниматься решетчатой случайной величиной; значения 1, $1 + \pi$ и $1 + 2\pi$ также могут приниматься решетчатой случайной величиной. Значения 0, 1 и π не могут приниматься решетчатой случайной величиной. Шаг h решетчатой случайной величины является наибольшим значением h, которое может быть использовано в приведенном выше определении. Если z_n является решетчатой случайной величиной, то $z_{n,\,s}$, w и w_s, очевидно, также являются решетчатыми случайными величинами, имеющими тот же самый шаг. Если z_n и, следовательно, $z_{n,\,s}$ не являются решетчатыми, то

w_s ведет себя существенно отличным образом; интервал между соседними принимаемыми значениями становится все меньше и меньше с ростом N.

Для решетчатого распределения с шагом h соответствующая предельная теорема*) утверждает, что в точках решетки

$$\left| Q_{w,\,s}(w_s) - \frac{h}{\sqrt{2\pi N\mu''(s)}} \exp \frac{-[w_s - N\mu'(s)]^2}{2N\mu''(s)} \right| \leqslant \frac{1}{\sqrt{N}}\,\varepsilon(N), \quad (5A.12)$$

где $\varepsilon(N)$ не зависит от выборочного значения w_s и

$$\lim_{N\to\infty} \varepsilon(N) = 0.$$

Другими словами, $Q_{w,s}(w_s)$ приближенно равно расстоянию между точками решетки h, умноженному на плотность гауссовского распределения с тем же самым средним значением и дисперсией, что и у w_s.

Так как представляют интерес только значения w_s, очень близкие к среднему значению, то можно использовать неравенства $1 \geqslant e^{-x} \geqslant 1 - x$, чтобы получить из (5A.12)

$$\left| Q_{w,\,s}(w_s) - \frac{h}{\sqrt{2\pi N\mu''(s)}} \right| \leqslant \frac{1}{\sqrt{N}}\varepsilon(N) + \frac{h\,[w_s - N\mu'(s)]^2}{2\sqrt{2\pi}\,N^{3/2}\,[\mu''(s)]^{3/2}}. \quad (5A.13)$$

Для того чтобы найти сумму в (5A.11), заменим вначале $Q_{w,s}$ на $h/\sqrt{2\pi N\mu''(s)}$. Обозначим через Δ расстояние между $N\mu'(s)$ и первым значением, принимаемым w_s, которое входит в сумму. Будем иметь после этого

$$\sum_{w_s \geqslant N\mu'(s)} \frac{h}{\sqrt{2\pi N\mu''(s)}} \exp\{-s[w_s - N\mu'(s)]\} =$$

$$= \frac{h\,e^{-s\Delta}}{\sqrt{2\pi N\mu''(s)}}\,[1 + e^{-hs} + e^{-2hs} + \ldots] = \frac{h\,e^{-s\Delta}}{\sqrt{2N\mu''(s)}\,(1 - e^{-sh})}. \quad (5A.14)$$

Аналогично можно умножить оба слагаемых для ошибки в (5A.13) на $\exp - s[w_s - N\mu'(s)]$ и провести суммирование по значениям $w_s \geqslant N\mu'(s)$. Первая сумма стремится к нулю быстрее, чем $1/\sqrt{N}$ при $N \to \infty$, а второе выражение стремится к нулю как $N^{-3/2}$. Объединяя (5A.13) и (5A.14), в результате получаем

$$\sum_{w_s \geqslant N\mu'(s)} Q_{w,\,s}(w_s) \exp\{-s[w_s - N\mu'(s)]\} =$$

$$= \frac{h\,e^{-s\Delta}}{\sqrt{2N\mu''(s)}\,(1 - e^{-sh})} + o(1/\sqrt{N}), \quad (5A.15)$$

где $o(1/\sqrt{N})$ стремится к нулю при $N \to \infty$ быстрее, чем $1/\sqrt{N}$. Используя (5A.15) совместно с (5A.11), получаем окончательный результат для решетчатых случайных величин

$$\Pr[w \geqslant N\mu'(s)] =$$

$$= \exp\{N[\mu(s) - s\mu'(s)]\} \left[\frac{h\,e^{-s\Delta}}{\sqrt{2N\mu''(s)}\,(1 - e^{-sh})} + o(1/\sqrt{N}) \right]. \quad (5A.16)$$

Равенство (5A.16) справедливо при всех $s > 0$, но, как можно заметить, рассматривая второе слагаемое для ошибки в (5A.13), сходимость по N становится мед-

*) См. Феллер (1966), т. 2, гл. XV, § 5.

ленное, когда s принимает значения, более близкие к нулю. Заметим, что для заданного s значение Δ изменяется с N, но всегда, конечно, лежит в пределах $0 \leqslant \Delta < n$.

Оценим теперь (5А.11) в нерешетчатом случае. Сумма в (5А.11)) может быть «проинтегрирована по частям», что даст

$$\sum_{w_s \geqslant N\mu'(s)} Q_{w,\,s}(w_s) \exp\{-s[w_s - N\mu'(s)]\} =$$

$$= \int_{w_s = N\mu'(s)}^{\infty} s\{F(w_s) - F[N\mu'(s)]\} \exp\{-s[w_s - N\mu'(s)]\}\, dw_s, \quad (5А.17)$$

где

$$F(w_s) = \sum_{w \leqslant w_s} Q_{w,\,s}(w)$$

— функция распределения w_s. Теперь пусть $u = [w_s - N\mu'(s)]/\sqrt{N\mu''(s)}$ является нормированной случайной величиной, соответствующей w_s, и пусть $G(u)$ будет функцией распределения u. Правая часть (5А.17) может быть записана в виде

$$s\sqrt{N\mu''(s)} \int_0^{\infty} [G(u) - G(0)] \exp[-s\sqrt{N\mu''(s)}\, u]\, du. \quad (5А.18)$$

Соответствующая этому случаю центральная предельная теорема утверждает, что*)

$$G(u) = \Phi(u) + \frac{\mu_3(1-u)^2}{6\mu_2^{3/2}\sqrt{2\pi N}} e^{-u^2/2} + o\left(\frac{1}{\sqrt{N}}\right), \quad (5А.19)$$

где

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \quad (5А.20)$$

$$\mu_2 = \overline{(z_{n,\,s} - \bar{z}_{n,\,s})^2}, \quad \mu_3 = \overline{(z_{n,\,s} - \bar{z}_{n,\,s})^3}$$

и $o(1/\sqrt{N})$ стремится к нулю при $N \to \infty$ равномерно по u и быстрее, чем $1/\sqrt{N}$.

Используя неравенства $1 \geqslant e^{-x^2/2} \geqslant 1 - x^2/2$ в (5А.20), будем иметь при $u > 0$

$$\frac{u}{\sqrt{2\pi}} \geqslant \Phi(u) - \Phi(0) \geqslant \frac{u}{\sqrt{2\pi}} - \frac{u^3}{6\sqrt{2\pi}}. \quad (5А.21)$$

Подстановка (5А.21) в (5А.19) дает

$$\left| G(u) - G(0) - \frac{u}{\sqrt{2\pi}} \right| \leqslant \frac{u^3}{6\sqrt{2\pi}} + \frac{|\mu_3|[1-(1-\mu^2)\, e^{-u^2/2}]}{6\mu_2^{3/2}\sqrt{2\pi N}} + o(1/\sqrt{N}). \quad (5А.22)$$

Аппроксимируя $G(u)$ — $G(0)$ в (5А.18) с помощью $u/\sqrt{2\pi}$, будем иметь

$$s\sqrt{N\mu''(s)} \int_0^{\infty} \frac{u}{\sqrt{2\pi}} e^{-s\sqrt{N\mu''(s)}\, u}\, du = \frac{1}{s\sqrt{2\pi N\mu''(s)}}. \quad (5А.23)$$

*) См. Феллер (1966), т. 2, гл. XVI, § 4. С помощью прямых вычислений можно проверить, что $\mu_3 = \mu'''(s)$ и является конечной величиной.

Умножая каждое слагаемое ошибки в (5А.22) на $s\sqrt{N\mu''(s)}\ e^{-s\sqrt{N\mu''(s)}}$ и интегрируя, замечаем, что каждый интеграл стремится к нулю быстрее, чем $1/\sqrt{N}$ при $N \to \infty$, что дает

$$s\sqrt{N\mu''(s)}\int_0^\infty [G(u)-G(0)]\, e^{\left[-s\sqrt{N\mu''(s)}\,u\right]}\,du = \\ = \frac{1}{s\sqrt{2\pi N\mu''(s)}} + o\left(\frac{1}{\sqrt{N}}\right). \qquad (5А.24)$$

Вспоминая, что эти выражения равны левой части (5А.17), можно подставить этот результат в (5А.11) и получить

$$\Pr[w \geqslant N\mu'(s)] = e^{N[\mu(s)-s\mu'(s)]}\left[\frac{1}{s\sqrt{2\pi N\mu''(s)}} + o\left(\frac{1}{\sqrt{N}}\right)\right]. \qquad (5А.25)$$

ПРИЛОЖЕНИЕ 5Б

В этом приложении будут доказаны теоремы 5.6.3 и 5.7.2, описывающие поведение $E_0(\rho, \mathbf{Q})$ и $E_x(\rho, \mathbf{Q})$ как функции ρ. Начнем с леммы.

Л е м м а. Пусть $\mathbf{Q} = [Q(0), \ldots, Q(K-1)]$ — вектор вероятностей и пусть $a_0, \ldots, a_{K-1}$ — множество неотрицательных чисел. Тогда функция

$$f(s) = \ln\left[\sum_{k=0}^{K-1} Q(k)\, a_k^{1/s}\right]^s \qquad (5Б.1)$$

является невозрастающей и выпуклой $\cup$ по s при $s > 0$. Более того, $f(s)$ является строго убывающей, если не *все* a_k, для которых $Q(k) > 0$, равны друг другу. Выпуклость является строгой, если не все *ненулевые* a_k, для которых $Q(k) > 0$, равны друг другу.

Доказательство. То, что $f(s)$ является невозрастающей, и условия, при которых она является строго убывающей, следуют непосредственно из известного неравенства (см. задачу 4.15 (д)).

$$\left[\sum Q(k)\, a_k^{1/s}\right]^s \geqslant \left[\sum Q(k)\, a_k^{1/r}\right]^r$$

при $0 < s < r$. Чтобы доказать выпуклость, будем считать, что s, r и θ — произвольные числа, $0 < s < r$, $0 < \theta < 1$, и определим

$$t = \theta s + (1-\theta)\, r. \qquad (5Б.2)$$

Для того чтобы показать, что $f(s)$ является выпуклой $\cup$, нужно показать, что

$$f(t) \leqslant \theta f(s) + (1-\theta) f(r). \qquad (5Б.3)$$

Определим число λ из соотношений:

$$\lambda = \frac{s\theta}{t}; \quad 1-\lambda = \frac{r(1-\theta)}{t}. \qquad (5Б.4)$$

Эти соотношения, как можно заметить (если их сложить и использовать (5Б.2)), являются совместными. Из (5Б.4) также следует, что

$$\frac{1}{t} = \frac{\theta}{t} + \frac{(1-\theta)}{t} = \frac{\lambda}{s} + \frac{1-\lambda}{r}, \qquad (5Б.5)$$

$$\sum_k Q(k)\, a_k^{1/t} = \sum_k Q(k)\, a_k^{\lambda/s} a_k^{(1-\lambda)/r} \leqslant$$

$$\leqslant \left[\sum_k Q(k)\, a_k^{1/s}\right]^{\lambda} \left[\sum_k Q(k)\, a_k^{1/r}\right]^{1-\lambda}, \qquad (5Б.6)$$

где (5Б.6) следует из неравенства Гельдера (см. задачу 4.15(в). Возводя обе части (5Б.6) в степень t и используя (5Б.4), получаем

$$\left[\sum_k Q(k)\, a_k^{1/t}\right]^t \leqslant \left[\sum_k Q(k)\, a_k^{1/s}\right]^{s\theta} \left[\sum_k Q(k)\, a_k^{1/r}\right]^{r(1-\theta)}. \qquad (5Б.7)$$

Взяв логарифм от обеих частей (5Б.7), получаем (5Б.3). Выпуклость является строгой до тех пор, пока (5Б.6) удовлетворяется со строгим неравенством; равенство в (5Б.6) имеет место тогда и только тогда, когда существует постоянная C, такая, что $Q(k)\, a_k^{1/s} = Q(k)\, a_k^{1/r} C$ при всех k (см. задачу 4.15(в)). Отсюда немедленно следует условие строгой выпуклости, указанное в лемме.

Доказательство теоремы 5.6.3. Имеем

$$E_0(\rho, \mathbf{Q}) = -\ln \sum_{j=0}^{J-1} \left[\sum_{k=0}^{K-1} Q(k)\, P(j \mid k)^{1/(1+\rho)}\right]^{1+\rho}. \qquad (5Б.8)$$

Взяв в лемме $P(j \mid k)$ вместо a_k и $1 + \rho$ вместо s, видим, что

$$\left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho)}\right]^{1+\rho}$$

является невозрастающей функцией ρ для каждого j. По предположению $\mathcal{I}(\mathbf{Q}; \mathbf{P}) > 0$, и, следовательно, $P(j \mid k)$ не зависит от k для тех k, для которых $Q(k) > 0$. Таким образом, написанное выше выражение, является строго убывающим по крайней мере для одного j; $E_0(\rho, \mathbf{Q})$ является строго возрастающей функцией ρ и $\partial E_0(\rho, \mathbf{Q})/\partial\rho > 0$ при $\rho \geqslant 0$. Так как $E_0(0, \mathbf{Q}) = 0$, то это означает также, что $E_0(\rho, \mathbf{Q}) > 0$ при $\rho > 0$. Покажем теперь, что $E_0(\rho, \mathbf{Q})$ является выпуклой $\cap$ по ρ. Пусть $\rho_1 > 0$ и $\rho_2 > 0$ являются произвольными числами и пусть θ удовлетворяет неравенствам $0 < \theta < 1$. Положим $\rho_3 = \rho_1\theta + \rho_2(1 - \theta)$. Из леммы [см. неравенство (5Б.7)] имеем

$$\sum_j \left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho_3)}\right]^{1+\rho_3} \leqslant \sum_j \left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho_1)}\right]^{(1+\rho_1)\theta} \times$$

$$\times \left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho_2)}\right]^{(1+\rho_2)(1-\theta)}. \qquad (5Б.9)$$

Используем теперь неравенство Гельдера (см. задачу 4.15(б))

$$\sum_j a_j b_j \leqslant \left[\sum_j a_j^{1/\theta}\right]^{\theta} \left[\sum_j b_j^{1/(1-\theta)}\right]^{1-\theta} \qquad (5Б.10)$$

для правой части (5Б.9) и получим

$$\sum_j \left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho_3)}\right]^{1+\rho_3} \leqslant \left\{\sum_j \left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho_1)}\right]^{1+\rho_1}\right\}^{\theta} \times$$

$$\times \left\{\sum_j \left[\sum_k Q(k)\, P(j \mid k)^{1/(1+\rho_2)}\right]^{1+\rho_2}\right\}^{1-\theta}. \qquad (5Б.11)$$

Взяв логарифм от обеих частей (5Б.11), будем иметь

$$-E_0(\rho_3, \mathbf{Q}) \leqslant -\theta E_0(\rho_1, \mathbf{Q}) - (1-\theta) E_0(\rho_2, \mathbf{Q}). \qquad (5Б.12)$$

Это означает, что $E_0(\rho, \mathbf{Q})$ является выпуклой $\cap$ по ρ. Выпуклость перестает быть строгой тогда и только тогда, когда как (5Б.9), так и (5Б.10) удовлетворяются с равенством. Согласно лемме (5Б.9) удовлетворяется с равенством тогда и только тогда, когда $P(j \mid k)$ не зависит от k для всех j, k, удовлетворяющих $Q(k)P(j \mid k) >$

>0. Условие равенства в (5Б.10) (см. задачу 4.15(б)) состоит в том, что существует постоянная C, такая, что для всех j

$$\left[\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho_1)}\right]^{1+\rho_1} = C\left[\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho_2)}\right]^{1+\rho_2}.$$

Если (5Б.9) удовлетворяется с равенством, то ненулевые $P(j\mid k)$ могут быть удалены из написанного выше равенства, что даст

$$\left[\sum_{k\,:\,P(j\mid k)>0} Q(k)\right]^{1+\rho_1} = C\left[\sum_{k\,:\,P(j\mid k)>0} Q(k)\right]^{1+\rho_2} \qquad (5Б.13)$$

при всех j. Это означает, что выражение в квадратных скобках является некоторой постоянной α, не зависящей от j, и поэтому для любых j, k, для которых $Q(k)P(j\mid k) > 0$, будем иметь

$$\frac{\sum_i Q(i)\,P(j\mid i)}{P(j\mid k)} = \alpha. \qquad (5Б.14)$$

Доказательство теоремы 5.7.2. Имеем

$$E_x(\rho, \mathbf{Q}) = -\ln\left\{\sum_k \sum_i Q(k)\,Q(i)\left[\sum_j \sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}\right\}^{\rho}. \qquad (5Б.15)$$

Можно применить лемму непосредственно к (5Б.15), сопоставляя двойную сумму в (5Б.15) однократной сумме в (5Б.1), сопоставляя $Q(k)Q(i)$ функции $Q(k)$ в (5Б.1) и сопоставляя

$$\sum \sqrt{P(j\mid k)\,P(j\mid i)}$$

числам a_k в (5Б.1). Таким образом, $E_x(\rho, \mathbf{Q})$ является возрастающей и выпуклой $\cap$ функцией ρ. Выпуклость является строгой, если не все ненулевые значения

$$\sum_j \sqrt{P(j\mid k)\,P(j\mid i)},$$

для которых $Q(k)Q(i) > 0$, равны друг другу. Эта сумма равна 1 при $k = i$ и (как следует из задачи 4.15(а)) она равна 1 при $k \neq i$ тогда и только тогда, когда $P(j\mid k) = P(j\mid i)$ для всех j. Подобно этому указанная выше сумма равна 0 тогда и только тогда, когда $P(j\mid k)P(j\mid i) = 0$ для всех j, что доказывает указанные в теореме условия строгой выпуклости. |

МЕТОДЫ КОДИРОВАНИЯ И ДЕКОДИРОВАНИЯ

6.1. КОДЫ С ПРОВЕРКОЙ НА ЧЕТНОСТЬ

В предыдущей главе было рассмотрено использование блокового кодирования для надежной передачи данных по дискретным каналам. Было доказано, что для соответствующим образом выбранных кодов и при любой заданной скорости передачи, меньшей пропускной способности, вероятность ошибочного декодирования ограничена неравенством $P_e \leqslant \exp[-NE_r(R)]$, где N — длина блока, а $E_r(R) > 0$ определяется соотношением (5.6.16). Вместе с тем число кодовых слов и число возможных принятых последовательностей являются экспоненциально растущими функциями N; поэтому при больших N практически невозможно осуществить хранение в памяти кодера всех кодовых слов и хранение в памяти декодера способа отображения принятых последовательностей в сообщения.

В настоящей главе будут рассмотрены методы кодирования и декодирования, позволяющие избежать упомянутой проблемы памяти путем использования алгоритмов для построения кодовых слов по сообщениям и сообщений по принятым последовательностям. Почти все известные методы кодирования и декодирования основаны на идеях, лежащих в основе кодов с проверкой на четность, и поэтому мы начнем изучение с этих кодов. Полезно начать описание таких кодов применительно к двоичному симметричному каналу (ДСК), хотя позднее будет показано, что это ограничение не является обязательным.

Код с проверкой на четность является частным случаем отображения двоичной последовательности длины L в двоичную последовательность некоторой большей длины N. Прежде чем определять коды с проверкой на четность, приведем довольно простой, но широко используемый пример проверки на четность. Допустим, что последовательность двоичных символов кодируется с помощью простого добавления одного двоичного символа в конце последовательности; этот последний символ выбирается таким образом, чтобы общее число единиц в кодовой последовательности было четным. Назовем первоначальные символы информационными символами, а добавленный в конце последовательности символ — проверочным символом. Легко видеть, что если затем изменить один из символов последовательности, то общее число единиц станет нечетным, что обнаруживает факт наступления ошибки. Однако, если произойдут две ошибки, число единиц вновь станет четным и ошибки не будут обнаружены. Кодирование такого типа широко используется при записи на магнитные ленты, а также и в других случаях, когда желательна некоторая небольшая возможность обнаружения ошибок при минимальных затратах на оборудование.

Удобно рассматривать эти проверки на четность или нечетность в терминах арифметических действий по модулю 2. В арифметике по модулю 2 существуют лишь два числа **0** и **1**; сложение (обозначаемое $\oplus$) определяется по правилу

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

Нетрудно заметить, что это сложение совпадает с обычным, если не считать того, что $1 \oplus 1 = 0$. Умножение аналогично умножению в обычной арифметике и обозначается тем же знаком.

Легко убедиться (кто сомневается, может это непосредственно проверить), что обычные ассоциативный, коммутативный и дистрибутивный арифметические законы справедливы в арифметике по модулю 2. Это значит, что если a, b и c — двоичные числа, то

$$\left.\begin{aligned}(a \oplus b) \oplus c &= a \oplus (b \oplus c)\\ (ab)\,c &= a\,(bc)\end{aligned}\right\}\text{— ассоциативность,}$$

$$a \oplus b = b \oplus a; \quad ab = ba \text{— коммутативность,}$$

$$(a \oplus b)\,c = ac \oplus bc \text{— дистрибутивность.}$$

Существует несколько интерпретаций арифметических действий по модулю 2. Если интерпретировать **0** как четное, а **1** как нечетное число, то эти действия сложения и умножения задают правила комбинирования четных и нечетных чисел. Кроме того, можно интерпретировать сумму по модулю 2 последовательности двоичных чисел как остаток от деления на 2 их суммы в обычном смысле. Тогда легко убедиться, что проверочный символ в рассмотренном выше примере выбирается таким образом, чтобы сумма по модулю 2 всех символов равнялась **0**. Если сумма по модулю 2 информационных символов равна **0**, то проверочный символ выбирается равным **0**; если сумма по модулю 2 информационных символов равна **1**, то проверочный символ выбирается равным **1**. Поэтому проверочный символ равен сумме по модулю 2 информационных символов.

В качестве простого (но чрезвычайно сильного) обобщения использования одной проверки на четность применительно к проверке последовательности информационных символов рассмотрим такое использование множества символов проверки на четность, при котором каждый из символов проверяет некоторое предварительно определенное множество информационных символов. Точнее, пусть $\mathbf{u} = (u_1, \ldots, u_L)$ обозначает последовательность L двоичных информационных символов; рассмотрим образование кодового слова $\mathbf{x}$ с длиной блока $N > L$ из последовательности $\mathbf{u}$ по правилу

$$x_n = u_n; \quad 1 \leqslant n \leqslant L, \tag{6.1.1}$$

$$x_n = \sum_{l=1}^{L} u_l\, g_{l,\,n}; \qquad L+1 \leqslant n \leqslant N, \tag{6.1.2}$$

где $\sum$ означает здесь сумму по модулю 2.

Элементы $g_{l,\,n}$ при $1 \leqslant l \leqslant L$ и $L+1 \leqslant n \leqslant N$ в соотношении (6.1.2) представляют собой фиксированные двоичные символы, не за-

	Информационные последовательности	Кодовые слова
$x_1 = u_1$	000	000000
$x_2 = u_2$	001	001110
	010	010101
$x_3 = u_3$	011	011011
	100	100011
$x_4 = u_2 \oplus u_3$	101	101101
$x_5 = u_1 \oplus u_3$	110	110110
$x_6 = u_1 \oplus u_2$	111	111000

Рис. 6.1.1.

висящие от $\mathbf{u}$; поэтому соотношения (6.1.1) и (6.1.2) задают отображение множества 2^L возможных информационных последовательностей в множество 2^L кодовых слов с длиной блока N. Назовем первые L символов в каждом кодовом слове *информационными символами*, а последние $N-L$ символов *проверочными символами*.

Систематическим кодом с проверкой на четность называется двоичный блоковый код с произвольной длиной блока N, для которого множество сообщений представляет собой множество 2^L двоичных последовательностей некоторой фиксированной длины $L < N$ и в котором каждому сообщению $\mathbf{u} = (u_1, \ldots, u_L)$ сопоставлено кодовое слово $\mathbf{x} = (x_1, \ldots, x_N)$, определяемое соотношениями (6.1.1) *и* (6.1.2), *где множество двоичных символов $\{g_{l,n}\}$ для $1 \leqslant l \leqslant L$, $L+1 \leqslant n \leqslant N$ является произвольным, но фиксированным и независимым от $\mathbf{u}$.* Разумеется, при каждом выборе совокупности $\{g_{l,n}\}$ получаются различные систематические коды с проверкой на четность.

Например, пусть для $L = 3$, $N = 6$ вектор $\mathbf{u} = (u_1, u_2, u_3)$ обозначает информационную последовательность, а вектор $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)$ — соответствующее кодовое слово, образуемое так, как показано на рис. 6.1.1.

Общий код с проверкой на четность определяется так же, как систематический код с проверкой на четность, за исключением того, что при вычислении в нем кодовых слов по последовательностям сообщений используются не соотношения (6.1.1) и (6.1.2), а соотношение

$$x_n = \sum_{l=1}^{L} u_l g_{l,n}; \quad 1 \leqslant n \leqslant N, \tag{6.1.3}$$

где $\{g_{l,n}\}$ — произвольное, но фиксированное множество двоичных чисел $1 \leqslant l \leqslant L$, $1 \leqslant n \leqslant N$. Из сравнения (6.1.1) и (6.1.3) видно, что систематический код с проверкой на четность представляет собой частный случай общего кода с проверкой на четность, в котором

$$\begin{aligned} g_{l,n} &= \mathbf{1}; \quad l = n, \\ g_{l,n} &= \mathbf{0}; \quad 1 \leqslant n \leqslant L, \quad l \neq n. \end{aligned} \tag{6.1.4}$$

В тех случаях, когда при использовании некоторого кода с проверкой на четность нужно будет специально обратить внимание на длину

блока N и на длину последовательности сообщения L, код будем называть (N, L)-кодом с проверкой на четность (или систематическим (N, L)-кодом с проверкой на четность).

Одну из причин рассмотрения кодов с проверкой на четность как систематических, так и несистематических, можно понять, если рассмотреть реализации кодера. Для кодера с проверкой на четность необходимы регистр для запоминания последовательности сообщения $\mathbf{u}$, регистр для запоминания кодового слова $\mathbf{x}$ и сумматоры по модулю 2, число которых пропорционально NL. Поэтому использование кодеров с проверкой на четность позволяет избежать экспоненциального роста объема памяти с ростом L, неизбежного при использовании произвольного блокового кода с 2^L кодовыми словами, структура которого не выбиралась специальным образом.

Порождающие матрицы

Соотношение (6.1.3) можно выразить в более компактной форме, если ввести понятие *порождающей матрицы* G для (N, L)-кода с проверкой на четность. Эта матрица, как показано на рис. 6.1.2, представляет собой двоичную матрицу с размерами L на N и компонен-

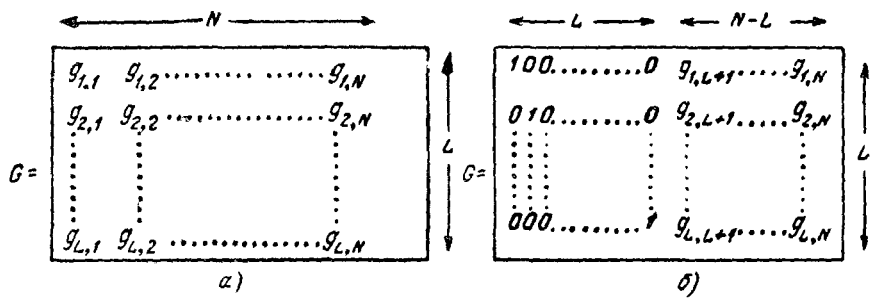


Рис. 6.1.2. Порождающая матрица:
а — произвольный код с проверкой на четность; *б* — систематический код с проверкой на четность.

тами $g_{l,n}$, определяемыми соотношениями (6.1.3). Из (6.1.4) следует, что в случае систематического кода с проверкой на четность подматрица, соответствующая первым L столбцам, является единичной матрицей.

Рассматривая $\mathbf{u}$ и $\mathbf{x}$ как вектор-строки, получаем

$$\mathbf{x} = \mathbf{u}G, \tag{6.1.5}$$

где $\mathbf{u}G$ представляет собой матричное произведение, использующее сложение по модулю 2 [это означает, что $\mathbf{u}G$ определяется таким образом, чтобы (6.1.5) было эквивалентно (6.1.3)].

Пусть теперь $\mathbf{u}'$ и $\mathbf{u}''$ — две информационные последовательности; определим сумму по модулю 2 двоичных векторов соотношением

$\mathbf{u} = \mathbf{u}' \oplus \mathbf{u}'' = u(u'_1 \oplus u''_1, \ldots, u'_L \oplus u''_L)$. Тогда, если $\mathbf{x}' = \mathbf{u}'G$ и $\mathbf{x}'' = \mathbf{u}''G$, то

$$\mathbf{x}' \oplus \mathbf{x}'' = \mathbf{u}'G \oplus \mathbf{u}''G = (\mathbf{u}' \oplus \mathbf{u}'')\,G = \tag{6.1.6}$$

$$= \mathbf{u}G. \tag{6.1.7}$$

Последнее равенство в (6.1.6) вытекает из ассоциативного, коммутативного и дистрибутивного законов сложения по модулю 2 и может быть доказано с помощью соотношения (6.1.3). Из (6.1.7) следует, что сумма по модулю 2 двух кодовых слов равна другому кодовому слову, соответствующему информационной последовательности $\mathbf{u} = \mathbf{u}' \oplus \mathbf{u}''$.

Заметим далее, что если в информационной последовательности имеется лишь один символ **1**, скажем на l-й позиции, то результирующее кодовое слово равно l-й строке матрицы G, которую обозначим через $\mathbf{g}_l$. Отсюда, учитывая соотношение (6.1.6), получаем, что произвольное кодовое слово можно представить в виде

$$\mathbf{x} = \sum_l u_l\, \mathbf{g}_l. \tag{6.1.8}$$

Другими словами, множество кодовых слов является *пространством строк матрицы* G, т. е. совокупностью линейных по модулю 2 комбинаций строк G, как это определяется соотношением (6.1.8).

Проверочные матрицы систематических кодов с проверкой на четность

Временно ограничимся рассмотрением систематических (N,L)-кодов с проверкой на четность и введем в рассмотрение новую матрицу H, называемую *проверочной матрицей*. Она представляет собой матрицу с размерами N на $N-L$ и определяется, как показано на рис. 6.1.3, через коэффициенты $g_{l,\,n}$, введенные в (6.1.2).

Чтобы понять значение матрицы H, перепишем (6.1.2), используя соотношение (6.1.1), в виде

$$x_n = \sum_{l=1}^{L} x_l\, g_{l,\,n}; \quad L < n \leqslant N. \tag{6.1.9}$$

Прибавляя x_n к обеим частям (или вычитая x_n, что эквивалентно в арифметике по модулю 2), получим для каждого кодового слова $\mathbf{x}$:

$$\mathbf{0} = \sum_{l=1}^{L} x_l\, g_{l,\,n} \oplus x_n; \quad L < n \leqslant N. \tag{6.1.10}$$

Путем сравнения этого равенства с матрицей на рис. 6.1.3 нетрудно убедиться, что оно может быть переписано в следующей матричной форме: для каждого кодового слова $\mathbf{x}$

$$\mathbf{x}H = \mathbf{0}. \tag{6.1.11}$$

Обратно, если произвольная последовательность $\mathbf{x}$ удовлетворяет соотношению (6.1.11), она также удовлетворяет (6.1.9). Эта последовательность должна быть кодовым словом, соответствующим инфор-

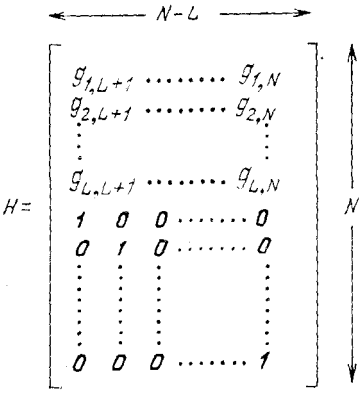


Рис. 6.1.3.

мационной последовательности, определяемой первыми L символами вектора $\mathbf{x}$. Таким образом, множество кодовых слов — это множество последовательностей $\mathbf{x}$, для которых $\mathbf{x}H = \mathbf{0}$ (т. е. нуль в пространстве столбцов матрицы H); множество кодовых слов является также множеством линейных комбинаций строк G (т. е. пространством строк матрицы G).

Матрица H используется в основном для декодирования. Пусть при передаче по двоичному каналу используется некоторый заданный систематический код с проверкой на четность, и $\mathbf{y}$ — принятая последовательность; определим соответствующий $\mathbf{y}$ синдром $\mathbf{S}$ равенством

$$\mathbf{S} = \mathbf{y}H. \tag{6.1.12}$$

Синдром $\mathbf{S}$ — это вектор-строка $(S_1, \ldots, S_{N-L})$ с $N-L$ компонентами, по одной для каждого проверочного символа. Поскольку

$$S_i = \sum_{l=1}^{L} y_l g_{l,\,L+i} \oplus y_{L+i},$$

то нетрудно убедиться, что S_i равно $\mathbf{1}$ тогда и только тогда, когда принятый i-й проверочный символ y_{L+i} отличается от i-го проверочного символа, вычисленного по принятым информационным символам.

Отметим, что если $\mathbf{x}_m$ является каким-либо кодовым словом, то

$$(\mathbf{y} \oplus \mathbf{x}_m)\, H = \mathbf{y}H \oplus \mathbf{x}_m H = \mathbf{S}. \tag{6.1.13}$$

Если $\mathbf{x}_m$—переданный, а $\mathbf{y}$ — принятый вектор, то последовательность $\mathbf{y} \oplus \mathbf{x}_m$, содержит $\mathbf{1}$ на тех позициях, на которых отличаются $\mathbf{y}$ и $\mathbf{x}_m$. Эта последовательность называется *шумовой последовательностью* $\mathbf{z}_m$, соответствующей $\mathbf{x}_m$; из (6.1.13) получаем

$$\mathbf{z}_m H = \mathbf{S}. \tag{6.1.14}$$

Так как $\mathbf{x}H = \mathbf{0}$, тогда и только тогда, когда $\mathbf{x}$ кодовое слово, из (6.1.13) следует, что для выполнения соотношения $\mathbf{z}H = \mathbf{S}$ необходимо и достаточно, чтобы $\mathbf{z}$ была одной из определенных выше шумовых последовательностей.

Нужно отметить, что соотношение (6.1.14) не позволяет точно установить, какая шумовая последовательность действительно имела место при передаче. Оно представляет собой соотношение, справедливое для всех $M = 2^L$ возможных последовательностей ошибок, соответствующих M кодовым словам. Выбор $\mathbf{z}_m$ (или выбор кодового слова) зависит от канала.

Предположим теперь, что при передаче по ДСК (см. рис. 5.3.1, *а*) с переходной вероятностью $\varepsilon < 1/2$ используется систематический код с проверкой на четность. Тогда, если кодовое слово $\mathbf{x}_m$ и принятая последовательность $\mathbf{y}$ отличаются в e позициях, то $\Pr(\mathbf{y} \mid \mathbf{x}_m) = \varepsilon^e (1 - \varepsilon)^{N-e}$. Число позиций, в которых отличаются две двоичные последовательности, называется *расстоянием Хэмминга* между этими последовательностями. Нетрудно видеть, что $\Pr(\mathbf{y} \mid \mathbf{x}_m)$ — убывающая функция от расстояния Хэмминга между $\mathbf{y}$ и $\mathbf{x}_m$, и поэтому декодирование по максимуму правдоподобия эквивалентно выбору кодового слова, находящегося на минимальном расстоянии от $\mathbf{y}$. Отметим также, что расстояние между $\mathbf{y}$ и $\mathbf{x}_m$ равно числу единиц (называемому *весом*) в шумовой последовательности $\mathbf{z}_m = \mathbf{y} \oplus \mathbf{x}_m$. Декодирование по максимуму правдоподобия состоит в выборе кодового слова $\mathbf{x}_m$, для которого $\mathbf{z}_m = \mathbf{y} \oplus \mathbf{x}_m$ имеет минимальный вес. Эти результаты можно сформулировать в виде следующей теоремы.

Теорема 6.1.1. Пусть систематический код с проверкой на четность имеет проверочную матрицу H и пусть этот код используется в ДСК с переходной вероятностью $\varepsilon < 1/2$. Тогда при заданной принятой последовательности $\mathbf{y}$ декодирование по максимуму правдоподобия состоит в вычислении синдрома $\mathbf{S} = \mathbf{y}H$, нахождении среди последовательностей, удовлетворяющих соотношению $\mathbf{S} = \mathbf{z}H$, последовательности $\mathbf{z}$ с минимальным весом и декодировании кодового слова $\mathbf{x} = \mathbf{z} \oplus \mathbf{y}$.

Таблицы декодирования

Доказанная теорема оставляет нерешенной задачу нахождения решения уравнения $\mathbf{S} = \mathbf{z}H$, которое обладает минимальным весом. Один из путей решения этой задачи основан на таблице декодирования. Мы можем составить список из 2^{N-L} возможных значений $\mathbf{S}$ и каждому из них поставить в соответствие последовательность $\mathbf{z}$ с минимальным весом. Проще всего реализовать эту идею, если начать с последовательностей нулевого веса, затем составить список всех $\mathbf{z}$, обладающих единичным весом, затем весом 2 и т. д. Для каждого из $\mathbf{z}$ можно вычислить синдром $\mathbf{S} = \mathbf{z}H$ и, если в дальнейшем в списке появится уже вычисленный синдром $\mathbf{S}$, то соответствующий ему новый вектор $\mathbf{z}$ опускается. Составление таблицы заканчивается, как только все возможные 2^{N-L} векторов $\mathbf{S}$ появятся в таблице. На рис. 6.1.4 представлена такая таблица декодирования для кода, приведенного на рис. 6.1.1.

Если при использовании этого кода принята последовательность $\mathbf{y} = \mathbf{010011}$, то $\mathbf{S} = \mathbf{y}H = \mathbf{110}$. Согласно приведенной таблице $\mathbf{z} = \mathbf{001000}$ и наиболее вероятное кодовое слово равно $\mathbf{x} = \mathbf{y} \oplus \mathbf{z} = \mathbf{011011}$. Поэтому декодированным сообщением будет **011**.

Из дальнейшего ясно, что приведенное в таблице множество шумовых последовательностей точно совпадает с множеством ошибок, которые будут исправлены (независимо от переданного кодового слова). В тех случаях, когда в канале появляется любая из этих шумовых

Синдром	Шумовая последовательность	Проверочная матрица
S	**z**	H
000	000000	
011	100000	011
101	010000	101
110	001000	110
100	000100	100
010	000010	010
001	000001	001
111	100100	

Рис. 6.1.4. Таблица декодирования для кода, приведенного на рис. 6.1.1.

последовательностей, декодер вычисляет соответствующий синдром и отыскивает в таблице декодирования эту самую шумовую последовательность. В тех случаях, когда появляется шумовая последовательность, не содержащаяся в таблице, декодер выбирает по таблице некоторую шумовую последовательность, и результат декодирования получается ошибочным. Для кода, представленного на рис. 6.1.1, и для декодирующей таблицы, представленной на рис. 6.1.4, исправляются все единичные ошибки, а также одна из двойных ошибок. Синдром **111** в таблице на рис. 6.1.4 имеет несколько конфигураций двойных ошибок, однако в таблицу может быть включена лишь одна из них; совершенно ясно, что в случае двоичного симметричного канала безразлично, какую из них включать.

Рассуждая подобным образом, можно убедиться, что для систематического кода с проверкой на четность, использующего таблицу декодирования, вероятность ошибочного декодирования в двоичном симметричном канале равна вероятности появления шумовой последовательности, не включенной в таблицу. Поскольку в таблице на рис. 6.1.4 содержатся одна безошибочная последовательность, шесть последовательностей, содержащих единичные ошибки, и одна последовательность, содержащая двойную ошибку, то вероятность ошибки для этого случая дается соотношением

$$P_e = 1 - (1-\varepsilon)^6 - 6(1-\varepsilon)^5\varepsilon - (1-\varepsilon)^4\varepsilon^2. \qquad (6.1.15)$$

Коды Хэмминга

Как было выяснено, синдром, соответствующий шумовой последовательности $\mathbf{z}$, определяется соотношением $\mathbf{S} = \mathbf{z}H$. Если $\mathbf{z}$ содержит лишь одну ошибку, скажем на n-й позиции, то $\mathbf{z}H$ совпадает с n-й строкой матрицы H. Если все строки матрицы H ненулевые и различные, то безошибочная последовательность и все последовательности, содержащие единичные ошибки, имеют различные синдромы и поэтому код может исправлять все единичные ошибки. Для кодов с $N-L$ проверочными символами существует $2^{N-L}-1$ различных ненулевых последовательностей, которые могут быть выбраны в качестве

N	L
3	1
7	4
15	11
31	26

$$H = \begin{bmatrix} 011 \\ 101 \\ 110 \\ 111 \\ 100 \\ 010 \\ 001 \end{bmatrix}$$

Набор значений N, L для кодов Хэмминга

Проверочная матрица для $N=7$, $h=4$

Рис. 6.1.5.

строк матрицы H; поэтому при $N \leqslant 2^{N-L} - 1$ строки H можно выбирать ненулевыми и различными.

Коды Хэмминга — это коды, для которых строки матрицы H различны и включают все ненулевые последовательности длины N—L. Поэтому ясно, что для таких кодов N и L связаны соотношением

$$N = 2^{N-L} - 1. \tag{6.1.16}$$

На рис. 6.1.5 приведены небольшая таблица значений N и L, удовлетворяющих соотношению (6.1.16), и матрица H для случая $N = 7$, $L = 4$.

Так как таблица декодирования для кодов Хэмминга не содержит ничего, кроме безошибочной последовательности и всех последовательностей с единичными ошибками, то ясно, что любая возможная принятая последовательность является либо кодовым словом, либо лежит на расстоянии 1 от одного и только от одного из кодовых слов.

Эти коды являются примером *сферически упакованных* кодов. Сферой радиуса e вокруг некоторой последовательности назовем множество всех последовательностей, лежащих на расстоянии e или меньшем от этой последовательности. Сферически упакованным кодом с исправляющей способностью e назовем код, у которого сферы радиуса e вокруг кодовых слов взаимно не пересекаются и любая последовательность лежит на расстоянии, не большем $e + 1$ от какого-либо кодового слова. При декодировании принятой последовательности в ближайшее кодовое слово исправляются все конфигурации не более чем e ошибок и некоторые конфигурации $e + 1$ ошибок; ни одна из конфигураций большего числа ошибок не исправляется. Легко видеть (см. § 5.8), что для двоичного симметричного канала сферически упакованный код с такой схемой декодирования обладает минимальной вероятностью ошибки среди всех кодов с той же самой длиной блока и с тем же самым числом кодовых слов.

Если сферически упакованный код, исправляющий e ошибок, удовлетворяет более сильному условию, состоящему в том, что любая последовательность удалена от некоторого кодового слова на расстояние, не большее e, то он называется *совершенным* кодом. В этом случае рассмотренный выше декодер исправляет все конфигурации e ошибок и не исправляет ни одной конфигурации большего числа ошибок.

Единственными найденными двоичными совершенными кодами являются: коды Хэмминга (исправляющие единичные ошибки), все коды из двух кодовых слов с нечетной длиной блока, у которых кодовые слова отличаются во всех позициях, и код с проверкой на четность, исправляющий тройные ошибки с $N = 23$ и $L = 12$, открытый Голеем (1949). Как показал Элайс*), для любой скорости R, $0 < R < 1$, выраженной в битах, не существуют двоичные сферически упакованные коды с длиной блока, большей N, где N зависит от R.

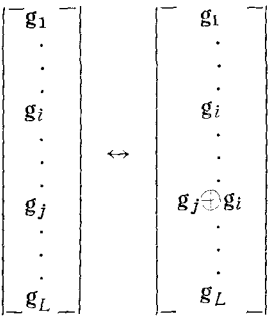


Рис. 6.1.6.

Проблема отыскания сферически упакованных кодов является, к сожалению, проблемой отыскания оптимальных (обладающих минимальной вероятностью ошибки) кодов среди кодов с проверкой на четность и среди произвольных кодов. Питерсон (1961) для двоичного симметричного канала составил таблицу известных оптимальных кодов с проверкой на четность, но ее расширение на большие длины блоков при произвольных скоростях не представляется возможным. Вместе с тем с практической точки зрения решение этой проблемы нельзя считать настоятельно необходимым. Известно, что вероятность ошибки при заданной скорости может быть сделана произвольно малой с помощью увеличения длины блока. Более важной чем проблема отыскания *наилучшего* кода при заданной длине блока является проблема отыскания наиболее легко практически реализуемого кода с *какой-либо* длиной блока, дающего требуемую вероятность ошибки.

Далее изучим здесь связь между различными кодами с проверкой на четность и, в частности, связь между систематическими и несистематическими кодами. Если две порождающие матрицы имеют одно и то же пространство строк, то они порождают одно и то же множество кодовых слов, хотя и с различными отображениями информационных последовательностей на кодовые слова. Назовем такие порождающие матрицы *эквивалентными*. Отметим, что если две строки $\mathbf{g}_i$ и $\mathbf{g}_j$ порождающей матрицы переставить местами, то результирующая матрица будет эквивалентна первоначальной. Аналогично, если строку $\mathbf{g}_i$

*) Доказательство результата Элайса приведено в работе Шеннона, Галлагера и Берлекэмпа (1967).

прибавить к $\mathbf{g}_j$, как показано на рис. 6.1.6, то результирующая матрица также будет эквивалентна первоначальной.

Произвольная порождающая матрица G может быть приведена к эквивалентной матрице в «приведенно-ступенчатой форме» следующим образом. Выбирается первый ненулевой столбец, затем строки переставляются таким образом, чтобы первая строка имела в этом столбце **1**, и, наконец, эта строка добавляется ко всем другим строкам, имеющим единицы в этом столбце, после чего единица остается лишь в первой строке рассматриваемого столбца. Аналогичная процедура может быть проделана со следующим столбцом, имеющим хотя бы одну единицу в оставшихся $L-1$ строках, после чего в этом столбце единица остается лишь во второй строке, и т. д. Процесс заканчивается либо тем, что в каждом из L столбцов останется по одной единице в различных строках, либо тем, что в нижней части матрицы останутся одна или несколько строк, целиком состоящих из нулей. Последний случай, который произойдет, если строки G линейно зависимы, неинтересен, так как означает, что имеются меньше чем 2^L различных кодовых слов и для каждого сообщения существует по крайней мере одно другое сообщение с тем же кодовым словом (см. задачу 6.10). В первом же случае каждый из столбцов, содержащих одну единицу, можно интерпретировать как столбец, соответствующий информационным символам, а остальные $N-L$ столбцов — как соответствующие проверочным символам.

Таким образом, для любой порождающей матрицы G существует эквивалентная матрица G', которая, если не принимать во внимание расположение информационных символов, соответствует систематическому коду. Проверочная матрица H для эквивалентного систематического кода имеет вид, аналогичный представленному на рис. 6.1.3, с той разницей, что единичная подматрица занимает те $(N-L)$ строк, которые соответствуют положению проверочных символов, и эти строки не обязательно являются последними $N-L$ строками. Так как первоначальный код имеет то же самое множество кодовых слов, что и эквивалентный систематический код, кодовые слова в обоих случаях удовлетворяют соотношению $\mathbf{x}H = \mathbf{0}$. Синдром принятой последовательности $\mathbf{y}$, как и прежде, определяется равенством $\mathbf{S} = \mathbf{y}H$ и можно, как и прежде, проводить декодирование на основе синдромной таблицы декодирования.

Нетрудно убедиться, что имеет место следующее полезное свойство проверочной матрицы: последовательность $\mathbf{x}$ является кодовым словом тогда и только тогда, когда $\mathbf{x}H = \mathbf{0}$. Было показано, как найти одну такую проверочную матрицу H для любой порождающей матрицы с линейно независимыми строками. Легко проверить (задача 6.11), что для всякой матрицы H', столбцы которой порождают то же пространство, что и столбцы матрицы H, $\mathbf{x}$ является кодовым словом тогда и только тогда, когда $\mathbf{x}H = \mathbf{0}$. По этой причине будем называть любую такую матрицу проверочной матрицей для данного кода.

6.2. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КОДОВ С ПРОВЕРКОЙ НА ЧЕТНОСТЬ

При доказательстве теоремы кодирования для кодов с проверкой на четность удобно ввести в рассмотрение несколько более широкий класс кодов, которые по причинам, объясняемым в следующем параграфе, назовем смежными кодами. *Смежным (N, L)-кодом называется код с 2^L кодовыми словами, имеющими длину блока $N > L$, в котором сообщения являются двоичными последовательностями длины L, а отображение сообщения* $\mathbf{u}$ *в кодовое слово* $\mathbf{x}$ *дается равенством*

$$\mathbf{x} = \mathbf{u}G \oplus \mathbf{v}, \tag{6.2.1}$$

где G — фиксированная, но произвольная двоичная матрица размера L на N, а $\mathbf{v}$ — фиксированная, но произвольная последовательность N двоичных символов. Кодовые слова смежного кода образуются из кодовых слов соответствующего кода с проверкой на четность $\mathbf{x}' = \mathbf{u}G$ путем добавления фиксированной последовательности $\mathbf{v}$ к каждому кодовому слову. В случае ДСК эта фиксированная последовательность может быть исключена из принятой последовательности перед декодированием и тем самым ее влияние нейтрализуется. Точнее, если принята последовательность $\mathbf{y} = \mathbf{x} \oplus \mathbf{z}$, то после вычитания из нее $\mathbf{v}$ получим $\mathbf{y}' = \mathbf{y} \oplus \mathbf{v}$, что в точности равно $\mathbf{x}' \oplus \mathbf{z}$. Так как в случае ДСК шумовая последовательность не зависит от переданной последовательности, то декодер максимального правдоподобия правильно декодирует то же самое множество шумовых последовательностей, что и при соответствующем коде с проверкой на четность, и вероятность ошибки будет та же самая.

Теорема 6.2.1. Рассмотрим ансамбль смежных (N, L)-кодов, в котором все символы G и $\mathbf{v}$ выбираются случайно и независимо равными **0** или **1**, причем вероятности этих значений одинаковы. Средняя по этому ансамблю кодов вероятность ошибки для каждого из сообщений при передаче их по ДСК и декодированию по максимуму правдоподобия ограничена неравенством

$$\overline{P}_{e,\,m} \leqslant \exp\left[-NE_r(R)\right], \tag{6.2.2}$$

где $E_r(R)$ выражается через переходную вероятность канала соотношениями (5.6.41) и (5.6.45) и $R = (L \ln 2)/N$.

Доказательство. Пусть $\mathbf{u}_m$ — произвольная информационная последовательность и $\mathbf{x}_m$ — соответствующее ей кодовое слово. Для рассматриваемого ансамбля кодов вероятность того, что $\mathbf{u}_m$ будет отображена в данную последовательность $\mathbf{x}_m$, равна

$$Q_N(\mathbf{x}_m) = 2^{-N}. \tag{6.2.3}$$

Чтобы доказать это, заметим, что существуют $2^{N(L+1)}$ способов выбора G и $\mathbf{v}$, каждый из которых имеет вероятность $2^{-N(L+1)}$. Для каждой фиксированной G существует один способ выбора $\mathbf{v}$,

при котором $\mathbf{x}_m$ принимает любое фиксированное значение. Поскольку существует 2^{NL} способов выбора G, то

$$Q_N(\mathbf{x}_m) = 2^{NL}\, 2^{-N(L+1)} = 2^{-N}.$$

Пусть далее $\mathbf{u}_{m'}$ — информационная последовательность, отличная от $\mathbf{u}_m$ и пусть $\mathbf{x}_{m'}$ — соответствующее ей кодовое слово. Покажем, что $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ статистически независимы в рассматриваемом ансамбле кодов. Имеем

$$\mathbf{x}_m \oplus \mathbf{x}_{m'} = (\mathbf{u}_m \oplus \mathbf{u}_{m'})\, G. \tag{6.2.4}$$

Предположим, что $\mathbf{u}_m$ и $\mathbf{u}_{m'}$ отличаются в j-й позиции. Тогда при любом выборе $\mathbf{g}_1, \ldots, \mathbf{g}_{j-1}, \mathbf{g}_{j+1}, \ldots, \mathbf{g}_L$ существует один способ выбора $\mathbf{g}_j$, при котором $\mathbf{x}_m \oplus \mathbf{x}_{m'}$ принимает любое фиксированное значение. Поэтому существуют $2^{N(L-1)}$ способов выбора G и $\mathbf{v}$, при которых пара $(\mathbf{x}_m, \mathbf{x}_{m'})$ принимает любое наперед заданное значение и $P(\mathbf{x}_m, \mathbf{x}_{m'}) = 2^{-2N}$. Отсюда и из (6.2.3) следует, что $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ независимы.

Теперь заметим, что в теореме 5.6.1 предполагалось, что кодовые слова выбирались независимо. Вместе с тем, если внимательно просмотреть доказательство, то можно установить, что в нем использовалась лишь попарная независимость. Таким образом, теорема 5.6.1 применима и к нашему ансамблю. Следовательно, также применима и теорема 5.6.2, из которой непосредственно вытекают выражения $E_r(R)$ для ДСК, приведенные в (5.6.41) и (5.6.45). |

Как и в § 5.6, этот результат доказывает существование смежного (N, L)-кода со средней вероятностью ошибки, по большей мере равной $\exp[-NE_r(R)]$, и в силу сделанного замечания, соответствующий код с проверкой на четность, получающийся после устранения фиксированной последовательности $\mathbf{v}$, обладает той же самой вероятностью ошибки. Так как множество правильно декодируемых шумовых последовательностей не зависит от сообщения, то точно так же $P_{e,\,m} \leqslant \exp[-NE_r(R)]$ для всех сообщений, закодированных этим кодом. Наконец, как было доказано в § 6.1, матрица G может быть преобразована в систематическую порождающую матрицу с помощью элементарных операций над строками и, быть может, перестановки некоторых столбцов*). Это доказывает следующее следствие.

Следствие. При всех положительных целых L и N, $L < N$ существуют систематические (N, L)-коды с проверкой на четность, для которых при использовании их в ДСК

$$P_{e,\,m} \leqslant \exp[-NE_r(R)] \text{ для всех } m,\ 1 \leqslant m \leqslant 2^L, \tag{6.2.5}$$

где $R = (L \ln 2)/N$, а $E_r(R)$ определяется равенствами (5.6.41) и и (5.6.45).

*) Для этого необходимо, чтобы строки матрицы G были линейно независимы. В коде с линейно зависимыми строками каждому кодовому слову соответствуют по меньшей мере два сообщения, и декодер максимального правдоподобия всегда будет принимать решение в условиях неопределенности. Так как все подобного рода события теорема 5.6.1 классифицирует как ошибки, то коды, для которых выполняется неравенство $P_e \leqslant \exp[-NE_r(R)]$, должны существовать среди кодов, для которых строки матрицы G линейно независимы.

Рассмотрим теперь использование кодов с проверкой на четность в произвольных дискретных каналах без памяти. В качестве первого примера рассмотрим случай, когда входной алфавит канала состоит из трех букв, и допустим, что нужно кодировать двоичные входные последовательности длины L кодовыми словами длины N. Скорость передачи в данном случае равна $R = (L \ln 2)/N$. Предположим, что для данного канала и данной скорости передачи значения входных вероятностей, максимизирующие величину $E_r(R)$ в (5.6.16), равны $Q(0) = {}^3/_8$, $Q(1) = {}^3/_8$, $Q(2) = {}^2/_8$. Первой нашей задачей будет

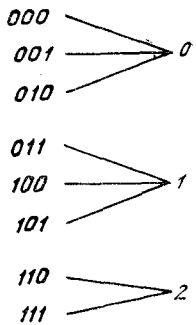


Рис. 6.2.1. Отображение двоичных последовательностей во входные буквы канала.

построение смежного кода с длиной блока $3N$ путем использования правила кодирования (6.2.1), где G — двоичная матрица размера L на $3N$, а $\mathbf{v}$ — двоичная последовательность длины $3N$. Можно рассматривать двоичные кодовые слова как последовательности N троек двоичных символов. Затем эти тройки двоичных символов кодируются во входные буквы канала по правилу, представленному на рис. 6.2.1. Это отображает каждое кодовое слово, представляющее собой последовательность $3N$ двоичных символов, в кодовое слово канала, представляющее собой последовательность N символов канала.

Для ансамбля кодов, рассмотренных в теореме 6.2.1, каждое кодовое слово является последовательностью $3N$ независимых равновероятных двоичных символов. Поэтому каждое кодовое слово канала представляет собой последовательность N троичных независимых символов с вероятностями $Q(0) = {}^3/_8$, $Q(1) = {}^3/_8$, $Q(2) = {}^1/_4$. Более того, кодовые слова попарно статистически независимы и поэтому снова можно применять теоремы 5.6.1 и 5.6.2. В результате получаем, что существует код рассматриваемого типа, для которого $P_e \leqslant \exp[-NE_r(R)]$.

В общем случае дискретного канала без памяти можно применить аналогичный метод доказательства, отличающийся лишь тем, что конкретное преобразование, представленное на рис. 6.2.1, будет другим. Необходимо лишь аппроксимировать нужные входные вероятности $\mathbf{Q}$ в теореме кодирования следующим образом:

$$Q(k) \approx \frac{i_k}{2^j}; \quad \sum_k i_k = 2^j. \tag{6.2.6}$$

Тогда для любого k, i_k двоичных последовательностей длины j отображаются во входной символ канала k, аналогично тому, как на рис. 6.2.1. Как велико должно быть j в соотношении (6.2.6), аппроксимирующем $Q(k)$, зависит от $\mathbf{Q}$ и от того, как близко нужно приблизиться к показателю экспоненты $E_r(R)$. При любых заданных j, длине сообщения L и длине кодового слова в канале N нужно использовать рассмотренный в теореме 6.2.1 ансамбль двоичных кодов с длиной двоичного блока jN. После отображения при помощи описанного выше преобразования двоичных кодовых слов в кодовые слова канала длины N и после использования теоремы 5.6.2 получаем, что существует код, для которого

$$P_e \leqslant \exp_2\left\{-N\left[\max_{0<\rho<1}\left(E_0(\rho, \mathbf{Q})-\rho R\right)\right]\right\}, \tag{6.2.7}$$

где $\mathbf{Q}$ определяется соотношением $Q(k) = i_k/2^j$.

Таким образом, мы описали простой алгоритм генерирования кодовых слов, при помощи которого можно достаточно хорошо приблизиться к границам, задаваемым теоремой кодирования. К сожалению, проблема нахождения алгоритмов *декодирования* является не такой простой.

6.3. ТЕОРИЯ ГРУПП

В предыдущих двух параграфах было широко использовано сложение по модулю 2. Для этого вначале было введено множество элементов **0** и **1**, а затем определена операция $\oplus$ над этими элементами. Под операцией здесь будем понимать правило, ставящее что-то в соответствие комбинации двух элементов множества. Ниже также будет рассматриваться ряд других множеств элементов и операций. Так как все они будут иметь одну и ту же математическую природу (образовывать группу), то полезно здесь прервать изложение и привести некоторые элементарные результаты теории групп, которые потребуются в дальнейшем.

*Группой называется множество элементов a, b, c, ... с операцией, обозначаемой символом $\cdot$, обладающей следующими свойствами**).*

1. Для любых элементов a, b, принадлежащих множеству, $a \cdot b$ также принадлежит множеству.

2. Выполняется ассоциативный закон, т. е. для любых a, b, c, принадлежащих множеству,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c. \tag{6.3.1}$$

3. В множестве имеется *нейтральный* элемент e, такой, что $a \cdot e = e \cdot a = a$ для всех a, принадлежащих множеству. (6.3.2)

*) Эта совокупность аксиом не является минимальной среди тех, которые могли бы быть использованы (см., например, книгу Биркгофа и Маклейна (1941)), но она наиболее полезна для нас.

4. Для любого элемента a из множества существует принадлежащий множеству обратный элемент a^{-1}, удовлетворяющий соотношениям

$$a \cdot a^{-1} = a^{-1} \cdot a = e. \tag{6.3.3}$$

Абелева (или коммутативная) группа определяется как группа, для которой также выполняется коммутативный закон:

$a \cdot b = b \cdot a$ для всех a, b, принадлежащих множеству. (6.3.4)

Ниже наибольший интерес будут представлять абелевы группы.

Введенная выше операция $\cdot$ во многом подобна обычному умножению, и нетрудно убедиться, что множество обычных ненулевых действительных чисел с операцией обычного умножения удовлетворяет всем аксиомам группы. С равным успехом, однако, операцией в группе может быть сложение или любая другая произвольным образом введенная операция. Например, целые числа с операцией сложения образуют группу. В этом случае 0 является нейтральным элементом, а обратным элементом для a служит $-a$. Аналогично, элементы **0** и **1** образуют группу, в которой операцией является сложение по модулю 2. Во всех случаях, когда операция в группе обозначается знаком $+$ или $\oplus$, условимся обозначать нейтральный элемент знаком **0**, а элемент, обратный для элемента a, через $-a$.

Следующие свойства полезно знать при работе с группами. Пусть e — нейтральный элемент, определяемый соотношением (6.3.2), a — произвольный элемент группы и a^{-1} — обратный элемент для элемента a [см. (6.3.3)]. Первые два приведенных ниже свойства утверждают единственность элементов e и a^{-1}.

1. Единственным элементом x, для которого $a \cdot x = a$, является e. (6.3.5)

2. Единственным x, для которого $a \cdot x = e$, является a^{-1}. (6.3.6)

3. Если $a \cdot b = a \cdot c$, то $b = c$. (6.3.7)

4. Уравнение $a \cdot x = b$ имеет единственное решение

$$x = a^{-1} \cdot b. \tag{6.3.8}$$

Чтобы проверить свойство 1, умножим обе части равенства $a \cdot x = a$ на a^{-1}, получим $a^{-1} \cdot a \cdot x = a^{-1} \cdot a$, или $e \cdot x = e$, или $x = e$. Свойства 2, 3 и 4 доказываются аналогично.

Подгруппы

Подмножество S элементов группы G, удовлетворяющее аксиомам группы при том же определении операции, что и в группе G, называется *подгруппой* группы G. В качестве примера рассмотрим множество всех двоичных последовательностей длины N. Это множество с операцией сложения последовательностей по модулю 2 образует группу. Последовательность **0** является нейтральным элементом и каждая последовательность является обратным элементом для самой себя.

Кодовые слова любого кода с проверкой на четность, имеющего длину блока N, образуют подгруппу этой группы. Чтобы показать это, заметим, что **0** всегда является кодовым словом, обратным элементом для кодового слова служит само это слово, а сумма по модулю 2 любых двух кодовых слов также является кодовым словом.

Порядком группы или подгруппы называется число элементов в группе или подгруппе. Следующая теорема Лагранжа содержит важный результат.

Теорема 6.3.1. (Лагранж.) Порядок группы, если он конечен, кратен порядку каждой подгруппы.

Нам потребуется несколько вспомогательных результатов. *Правым смежным классом* (левым смежным классом) группы G по подгруппе S называется подмножество $s_1 \cdot a$, $s_2 \cdot a$, ..., $(a \cdot s_1, a \cdot s_2, \ldots)$ элементов G, где a — произвольный заданный элемент группы G, а s_1, s_2 ... — все элементы S. Непосредственно можно убедиться, что для подгруппы конечного порядка число элементов в смежном классе равно числу элементов в подгруппе, так как если $s_i \neq s_j$, то $s_i \cdot a \neq s_j \cdot a$. Кроме того, если два правых смежных класса (левых смежных класса) одной и той же подгруппы имеют какой-либо общий элемент, то эти два подмножества совпадают. Чтобы доказать это, предположим, что одно подмножество порождается элементом a, а другое — элементом b. Если $s_i \cdot a = s_j \cdot b$, то $s_j^{-1} \cdot s_i \cdot a = b$ и b принадлежит смежному классу, порождаемому a. Поэтому для любого s_n, принадлежащего S, $s_n \cdot b = s_n \cdot s_j^{-1} \cdot s_i \cdot a$ и любой элемент смежного класса, порождаемого b, принадлежит смежному классу, порождаемому a.

В коде с проверкой на четность, как было показано, для любой фиксированной последовательности **y** сумма $\mathbf{x} \oplus \mathbf{y}$ имеет тот же самый синдром, что и кодовое слово **x**. Поэтому множество последовательностей с данным синдромом является смежным классом по подгруппе, образуемой кодовыми словами. Тогда правило декодирования по максимуму правдоподобия в двоичном симметричном канале может быть переформулировано следующим образом: по заданному **y** найти шумовую последовательность **z**, полагая, что она является последовательностью минимального веса в смежном классе, которому принадлежит **y**.

Теперь можно доказать теорему 6.3.1. Предположим, что группа G порядка n содержит подгруппу S порядка m. Если $n > m$, то выберем элемент группы G, не принадлежащий S, и образуем правый смежный класс. Подгруппа и смежный класс вместе содержат $2m$ элементов; если $n > 2m$, то можно выбрать другой элемент в G, не принадлежащий ни подгруппе, ни смежному классу, и образовать новый правый смежный класс, что даст нам в совокупности $3m$ элементов. Продолжая таким образом, в конце концов достигнем того, что все элементы G будут принадлежать либо S, либо одному из смежных классов; если кроме подгруппы существуют u смежных классов, то получим $n = m(u + 1)$, что завершает доказательство. |

Циклические подгруппы

Пусть a — элемент конечной группы G. Рассмотрим последовательность элементов

$$a,\ a^2,\ a^3,\ \ldots, \tag{6.3.9}$$

где a^2 означает $a \cdot a$; a^3 означает $a \cdot a \cdot a$ и т. д. Так как группа конечна, должны существовать два значения показателя степени i и j, $j > i$, для которых

$$a^i = a^j = a^i \cdot a^{j-i}. \tag{6.3.10}$$

Согласно (6.3.5) отсюда следует, что $a^{j-i} = e$. *Порядком* элемента a группы называется наименьшее положительное целое число m, для которого $a^m = e$. Приведенные выше рассуждения показывают, что каждый элемент конечной группы имеет конечный порядок. Более того, элементы $a,\ a^2,\ \ldots,\ a^m = e$ должны быть различными, поскольку равенство $a^i = a^j$, где $j > i$, может выполняться лишь при $j - i \geqslant m$. Поэтому последовательность (6.3.9) степеней элемента a имеет следующее циклическое свойство:

$$a,\ a^2,\ \ldots,\ a^m = e,\ a,\ a^2,\ \ldots,\ a^m = e,\ a,\ \ldots \tag{6.3.11}$$

Из (6.3.11) следует, что $a^n = e$ тогда и только тогда, когда n кратно m.

Теорема 6.3.2. Пусть элемент a конечной группы G имеет порядок m. Тогда элементы $a,\ a^2,\ \ldots,\ a^m$ образуют подгруппу группы G и m является делителем для порядка группы G.

Доказательство. Подмножество $a,\ a^2,\ \ldots\ a^m$ содержит нейральный элемент $e = a^m$. Тогда для любого a^i, принадлежащего этому подмножеству, $a^i \cdot a^{m-i} = a^m = e$, откуда следует, что каждый элемент подмножества имеет обратный элемент в том же подмножестве. Для завершения доказательства необходимо показать, что если a^i и a^j принадлежат подмножеству, то этому подмножеству принадлежит и $a^i \cdot a^j$. Имеем $a^i \cdot a^j = a^{i+j}$ и при $i + j \leqslant m$ произведение $a^i \cdot a^j$ принадлежит подмножеству. Если $i + j > m$, то имеем $a^{i+j} = a^m \cdot a^{i+j-m} = a^{i+j-m}$. Так как $i + j - m \leqslant m$, то $a^i \cdot a^j$ принадлежит подмножеству. Поэтому это подмножество образует подгруппу порядка m. Из теоремы 6.3.1 следует, что m является делителем порядка группы G. |

Группа или подгруппа называется *циклической*, если в этой группе или подгруппе существует элемент, степени которого образуют всю группу или подгруппу. Поэтому подгруппы, входящие в формулировку теоремы 6.3.2, являются циклическими.

Следующие результаты, касающиеся порядка элементов конечной абелевой группы, будут использоваться в § 6.6.

Л е м м а. Пусть a — элемент порядка m и b — элемент порядка n в абелевой группе. Тогда, если m и n взаимно простые*), то порядок элемента $a \cdot b$ равен mn.

*) Два положительных целых числа называются взаимно простыми, если их наибольший общий делитель равен 1, т. е. разложения этих чисел на множители не содержат ни одного общего множителя, большего 1.

Доказательство. Имеем

$$(a \cdot b)^{mn} = (a^m)^n \cdot (b^n)^m = e. \tag{6.3.12}$$

Поэтому, обозначив порядок элемента $a \cdot b$ через l, получим $l \leqslant mn$. Кроме того,

$$e = (a \cdot b)^{ln} = a^{ln} \cdot (b^n)^l = a^{ln}. \tag{6.3.13}$$

Отсюда следует, что ln кратно порядку m элемента a. Так как m и n взаимно простые, то l кратно m. Поменяв ролями m и n в приведенных выше рассуждениях, получим, что l также кратно n и в силу того, что m и n взаимно простые, $l \geqslant mn$, что завершает доказательство. |

Теорема 6.3.3. Пусть m — максимальный порядок элементов конечной абелевой группы. Тогда m кратно порядку каждого элемента группы.

Доказательство. Пусть a— элемент максимального порядка m и пусть n—порядок какого-либо другого элемента b. Обозначим через $p_1, p_2, \ldots, p_r$ все простые числа, которые служат делителями либо для m, либо для n, и представим m и n в виде

$$m = p_1^{m_1} p_2^{m_2} \ldots p_r^{m_r}, \tag{6.3.14}$$

$$n = p_1^{n_1} p_2^{n_2} \ldots p_r^{n_r}, \tag{6.3.15}$$

где $m_1, \ldots, m_r$ и $n_1, \ldots, n_r$ — целые неотрицательные числа. Если m не кратно n, то для некоторого i, $1 \leqslant i \leqslant r$, должно выполняться $n_i > m_i$. Пусть далее для любого j через a_j обозначен элемент порядка $p_j^{m_j}$ (такой элемент равен a в степени $m/p_j^{m_j}$). Аналогично, пусть b_i есть элемент порядка $p_i^{n_i}$. Рассмотрим теперь элемент $c = a_1 \cdot a_2 \ldots a_{i-1} \cdot b_i \cdot a_{i+1} \ldots a_r$. Последовательно применяя предыдущую лемму при каждом включении нового сомножителя в это произведение, убеждаемся, что c имеет порядок $mp_i^{n_i}/p_i^{m_i} > m$. Это привело к противоречию, поскольку m равно наибольшему порядку элементов; поэтому m кратно n. |

6.4. ПОЛЯ И МНОГОЧЛЕНЫ

Значительная часть алгебраической теории кодирования основана на теории конечных полей. Грубо говоря, поле — это множество элементов, в котором сложение, вычитание, умножение и деление могут трактоваться как обычные правила арифметики. *Более точной является следующая формулировка: поле — это множество по меньшей мере двух элементов, замкнутое**) *по двум операциям, называемым сложением* (+) *и умножением* (·), *и удовлетворяющее следующим аксиомам:*

---

*) Множество элементов называется замкнутым по операции ·, если для любой пары элементов (a, b), принадлежащих множеству, $a \cdot b$ также принадлежит этому множеству.

1. Множество элементов образует абелеву группу по операции сложения.

2. Множество ненулевых элементов (где **0** является нейтральным элементом группы по операции сложения) образует абелеву группу по операции умножения.

3. Выполняется дистрибутивный закон

$$(a + b) \cdot c = a \cdot c + b \cdot c \text{ для всех } a, b, c. \qquad (6.4.1)$$

Легко непосредственно убедиться, что множество действительных чисел с обычными операциями сложения и умножения удовлетворяет этим аксиомам. Множество двоичных элементов **0** и **1** с операцией сложения по модулю 2 и обычным умножением также удовлетворяет этим аксиомам. Однако множество целых чисел не удовлетворяет этим аксиомам, поскольку любое целое число, большее 1, не имеет обратного по операции умножения элемента, который являлся бы целым числом.

Условимся всюду обозначать нейтральный по операции сложения элемент символом **0**, а нейтральный по операции умножения элемент — символом **1**. Будем всюду обозначать обратные элементы по операциям сложения и умножения как $-a$ и a^{-1} соответственно.

Следующие соотношения выражают некоторые элементарные свойства полей:

$$a \cdot \mathbf{0} = \mathbf{0} \cdot a = \mathbf{0} \text{ для всех } a; \qquad (6.4.2)$$

$$a \cdot b \neq \mathbf{0} \text{ для всех ненулевых } a, b; \qquad (6.4.3)$$

$$-(a \cdot b) = (-a) \cdot b = a \cdot (-b) \quad \text{для всех } a, b; \qquad (6.4.4)$$

$$a \cdot b = a \cdot b' \Rightarrow b = b' \text{ для } a \neq \mathbf{0}. \qquad (6.4.5)$$

Чтобы проверить справедливость (6.4.2), допустим, что a и b произвольны и заметим, что $a \cdot b = a \cdot (b + \mathbf{0}) = a \cdot b + a \cdot \mathbf{0}$. Отсюда следует, что $a \cdot \mathbf{0}$ является нейтральным элементом в группе по сложению [см. (6.3.5)] и $a \cdot \mathbf{0} = \mathbf{0}$. Соотношение (6.4.3) утверждает, что множество ненулевых элементов замкнуто по умножению; это следует из аксиомы 2. Чтобы проверить (6.4.4), запишем следующее соотношение

$$\mathbf{0} = \mathbf{0} \cdot b = [a + (-a)] \cdot b = a \cdot b + (-a) \cdot b.$$

Поэтому $(-a) \cdot b$ является обратным элементом по операции сложения для $a \cdot b$. Вторая часть равенства (6.4.4) доказывается аналогично. Из (6.4.4) следует, что, не вызывая недоразумений, знак минус может использоваться без скобок; в дальнейшем мы так и будем поступать. Чтобы доказать математическое правило сокращения (6.4.5), заметим, что из $a \cdot b = a \cdot b'$ следует $\mathbf{0} = a \cdot b - a \cdot b' = a \cdot (b - b')$. Так как $a \neq \mathbf{0}$, то отсюда следует, что $b - b' = \mathbf{0}$ и $b = b'$.

В дальнейшем будут рассматриваться лишь *поля Галуа*, которые по определению являются полями с конечным числом элементов. Следующая теорема будет часто весьма полезной при решении вопроса о том, образует ли поле некоторое конечное множество элементов.

Теорема 6.4.1. Аксиома 2 в данном выше определении поля для случая конечного множества элементов может быть заменена более

слабым условием (2′): операция умножения является коммутативной и ассоциативной и выполняется (6.4.3).

Доказательство. Легко видеть, что (6.4.2), (6.4.4.) и (6.4.5) по-прежнему верны, поскольку их доказательства не опирались на аксиому 2. Пусть a — ненулевой элемент множества; рассмотрим последовательность a, a^2, a^3, где $a^2 = a \cdot a$, $a^3 = a \cdot a \cdot a$ и т. д. Так как рассматриваемое множество конечно, то должны существовать два целых числа j и i, $j > i$, для которых

$$a^i = a^j = a^i \cdot a^{j-i}. \tag{6.4.6}$$

Покажем, что a^{j-i} является нейтральным элементом по операции умножения. Умножая обе части (6.4.6) на произвольный элемент b и сокращая на a^i (который является ненулевым элементом), получаем

$$b = b \cdot a^{j-i}.$$

Отсюда следует, что a^{j-i} является нейтральным элементом по операции умножения. Наконец, обратный по операции умножения элемент для a равен a^{j-i-1} (или a, если $j = i + 1$). Так как a — произвольный элемент, то любой ненулевой элемент имеет обратный и аксиома 2 справедлива. |

Чтобы пояснить на примере использование этой теоремы, выберем простое число p и рассмотрим множество целых чисел **0**, **1**, ..., **i**, где $i = p - 1$. Введем операции сложения и умножения по модулю p [это означает, что элемент поля $\mathbf{i} + \mathbf{j}$ задается остатком R_p $(i + j)$; от деления обычной суммы $i + j$ на p]. Обратный по сложению элемент для элемента i — это элемент, соответствующий $p - i$ Существование обратного по умножению элемента не совсем очевидно, но следует из теоремы 6.4.1. Поэтому для любого простого p неотрицательные целые числа, меньшие чем p, образуют поле в арифметике по модулю p. Если p не простое, эти элементы не могут образовывать поле в арифметике по модулю p, поскольку существуют два ненулевых элемента, обычное произведение которых равно p; отсюда следует, что их произведение по модулю p равно нулю. Ниже мы покажем, то существуют поля с p^n элементами, где p — простое, а n — большее 1 целое число, однако правила сложения и умножения в этих полях не являются сложением и умножением по модулю p^n. Как и в теории групп, *порядок* поля Галуа равен числу элементов в поле; в дальнейшем будем обозначать любое поле с q элементами через $GF(q)$.

Многочлены

Выражение вида $f_n D^n + f_{n-1} D^{n-1} + \ldots + f_0$, обозначаемое через $f(D)$, называется *многочленом* над $GF(q)$ степени n, если коэффициенты $f_n, \ldots, f_0$ являются элементами $GF(q)$ и если первый коэффициент f_n ненулевой. Удобно считать, что коэффициент f_i, связанный с многочленом n-го порядка, определен для всех $i \geqslant 0$, но удовлетворяет соотношению $f_i = 0$ для $i > n$. Поэтому можно рассматривать многочлен

над $GF(q)$ как способ представления бесконечной последовательности элементов поля $f_0, f_1, \ldots$, когда лишь конечное число членов последовательности отлично от нуля. Степенью многочлена в этом случае является наибольшее число n, для которого $f_n \neq 0$. В частном случае 0-многочлена, $f_n = 0$ для всех $n \geqslant 0$, условимся считать, что 0-многочлен имеет степень 0.

Символ D в многочлене $f(D)$ называется *неопределенным*. Его *нельзя* интерпретировать как переменную или неизвестный элемент поля, во-первых, потому, что в дальнейшем мы иногда будем подставлять в качестве D элемент, не принадлежащий первоначальному полю, и, во-вторых, потому, что чаще интерес будет представлять последовательность коэффициентов, определяемая многочленом, а не многочлен как функция. Во всяком случае, после того как будут определены правила обращения с многочленами, вопрос о том, что такое неопределенная, исчезнет сам по себе.

Назовем два многочлена равными, если каждый из них соответствует одной и той же последовательности коэффициентов. Например, пусть $f(D) = D^3 + D^2 + D$ и $g(D) = D$ — два многочлена над полем по модулю 2 (условимся при записи многочленов опускать слагаемые, коэффициенты при которых равны **0**, и записывать $\mathbf{1}D^i$ как D^i). В силу нашего определения, эти многочлены не равны. Вместе с тем, если подставить вместо D элементы поля **0** и **1**, то получим, что $f(\mathbf{0}) = \mathbf{0}$, $f(\mathbf{1}) = \mathbf{1}$ и $g(\mathbf{0}) = \mathbf{0}$, $g(\mathbf{1}) = \mathbf{1}$. Поэтому, если рассматривать $f(D)$ и $g(D)$ как функции переменной, определенной в поле по модулю 2, то они будут равны, хотя при рассмотрении их в виде многочленов они не равны.

Сумма двух многочленов над данным полем есть многочлен над тем же полем, определяемый по известному правилу

$$f(D) + g(D) = \sum_{i=0}^{\infty} (f_i + g_i) D^i. \tag{6.4.7}$$

Степень $f(D) + g(D)$ равна максимальному n, для которого $f_n + g_n \neq 0$ и, следовательно, не превосходит наибольшей среди степеней $f(D)$ и $g(D)$. Например, над полем по модулю 2

$$(D^2 + D + \mathbf{1}) + (D^2 + \mathbf{1}) = (\mathbf{1} \oplus \mathbf{1}) D^2 + D + (\mathbf{1} \oplus \mathbf{1}) = D.$$

Произведение двух многочленов над данным полем есть многочлен над тем же полем, определяемый соотношением

$$f(D) g(D) = \sum_i \left(\sum_{j=0}^{i} f_j g_{i-j} \right) D^i. \tag{6.4.8}$$

Непосредственно используя (6.4.8), можно показать, что для $g(D) = \mathbf{0}$

$$f(D)\mathbf{0} = \mathbf{0} \text{ для всех } f(D). \tag{6.4.9}$$

Кроме того,

$$f(D) \cdot g(D) \neq \mathbf{0} \text{ для } f(D) \neq \mathbf{0},\ g(D) \neq \mathbf{0}. \tag{6.4.10}$$

Чтобы убедиться в этом, предположим, что $f(D)$ имеет степень n, $f_n \neq \mathbf{0}$ и $g(D)$ имеет степень m, $g_m \neq \mathbf{0}$. Тогда из (6.4.8) следует, что

$f(D)g(D)$ имеет степень $n+m$ и первое слагаемое этого многочлена равно $f_n g_m D^{n+m}$.

Умножение многочлена $f(D)$ над некоторым полем на элемент α этого поля определяется соотношением

$$\alpha f(D) = \sum_i (\alpha f_i) D^i.$$

Аналогично, отрицательный многочлен для данного многочлена определяется как

$$-f(D) = \sum_i (-f_i) D^i.$$

Легко убедиться, что множество многочленов над полем образует абелеву группу по сложению. Также можно показать, что для умножения многочленов выполняются и ассоциативный и коммутативный законы и что справедлив дистрибутивный закон

$$[f(D) + g(D)]\, h(D) = f(D)h(D) + g(D)h(D).$$

Однако множество многочленов над полем не образует поле из-за отсутствия обратных по умножению элементов. Последнее служит примером того, что теорема 6.4.1 становится неверной без ограничения конечными множествами.

Сформулируем некоторые элементарные свойства многочленов, которые будут полезны в дальнейшем:

$$-[f(D)g(D)] = [-f(D)]\, g(D) = f(D)\, [-g(D)], \qquad (6.4.11)$$

$$f(D)g(D) = f(D)h(D) \Rightarrow g(D) = h(D) \text{ для } f(D) \neq \mathbf{0}. \qquad (6.4.12)$$

Доказательства (6.4.11) и (6.4.12) аналогичны доказательствам (6.4.4) и (6.4.5).

Хотя, вообще говоря, нельзя делить один многочлен на другой и получить многочлен — частное, деление все же может производиться, если при этом допускается остаток.

Теорема 6.4.2. (Алгоритм Евклида деления многочленов.) Пусть $f(D)$ и $g(D)$ — многочлены над $GF(q)$ и пусть $g(D)$ имеет степень, не меньшую 1. Тогда существуют единственные многочлены $h(D)$ и $r(D)$ над $GF(q)$, для которых

$$f(D) = g(D)h(D) + r(D), \qquad (6.4.13)$$

где степень $r(D)$ меньше, чем степень $g(D)$.

Доказательство. Покажем сначала, как найти $h(D)$ и $r(D)$, удовлетворяющие (6.4.13), а затем докажем, что это решение единственное. Пусть $f(D)$ имеет степень n и пусть $g(D)$ имеет степень m. Если $n < m$, то положим $h(D) = 0$, $r(D) = f(D)$. Если $n \geqslant m$, произведем деление $f(D)$ на $g(D)$ по тому же правилу, по которому производится деление многочленов в обычном поле действительных чисел, а затем положим $h(D)$ равным частному от деления, $r(D)$ — остатку. Это означает, что первое слагаемое $h(D)$ равно $f_n g_m^{-1} D^{n-m}$. Это первое слагаемое, умноженное на $g(D)$, вычитается из $f(D)$, и следующее слагаемое $h(D)$ находится путем деления на $g(D)$ этого остатка. Когда остаток

будет иметь степень, меньшую, чем степень $g(D)$, он выбирается в качестве $r(D)$. В качестве примера рассмотрим случай, когда $g(D) = D^2 + D + 1$ и $f(D) = D^3 + D + 1$ являются многочленами над полем по модулю 2:

$$
\begin{array}{r|l}
 & D+1 \\
\hline
D^2+D+1 & D^3+\quad D+1 \\
 & D^3+D^2+D \\
\hline
 & \quad D^2\quad +1 \\
 & \quad D^2+D+1 \\
\hline
 & \quad D
\end{array}
$$

В этом примере $h(D) = D + 1$ и $r(D) = D$.

Теперь предположим, что существуют два решения (6.4.13), задаваемые $h(D)$, $r(D)$ и $h'(D)$, $r'(D)$. Тогда

$$g(D)h(D) + r(D) = g(D)h'(D) + r'(D), \qquad (6.4.14)$$

$$g(D)[h(D) - h'(D)] = r'(D) - r(D). \qquad (6.4.15)$$

Теперь заметим, что $r'(D) - r(D)$ имеет степень, меньшую, чем степень $g(D)$, и поэтому $h(D) - h'(D) = 0$. Но отсюда следует, что $r'(D) - r(D) = 0$, что завершает доказательство единственности решения.

В дальнейшем при операциях над конечными полями нас часто будет интересовать не столько частное $h(D)$, сколько остаток $r(D)$ в (6.4.13). *По аналогии с арифметическими действиями по модулю простого числа назовем вычетом многочлена $f(D)$ по модулю многочлена $g(D)$ остаток от деления $f(D)$ на $g(D)$; будем обозначать его через* $R_{g(D)}[f(D)]$:

$$R_{g(D)}[f(D)] = r(D);\ r(D) \text{ определяется (6.4.13).} \qquad (6.4.16)$$

Алгоритм Евклида деления многочленов можно использовать для исследования существования делителей и корней многочлена над $GF(q)$. *Многочлен $g(D)$ является множителем (или делителем) другого многочлена $f(D)$, если существует многочлен $h(D)$ над рассматриваемым полем, удовлетворяющий соотношению*

$$f(D) = g(D)h(D). \qquad (6.4.17)$$

Другими словами, $f(D)$ делится на $g(D)$, если при использовании алгоритма Евклида (6.4.13) получим $r(D) = 0$. *Многочлен $f(D)$ называется приводимым, если существуют два многочлена $g(D)$ и $h(D)$ над рассматриваемым полем, каждый из которых имеет степень, не меньшую 1, удовлетворяющие* (6.4.17). *Многочлен называется неприводимым, если он не является приводимым.*

Часто бывает полезно разлагать многочлен на неприводимые множители. При этом всегда существует некоторая доля неопределенности, поскольку при разложении всегда можно умножить один из сомножителей на произвольный элемент поля, а другой сомножитель — на обратный ему элемент поля, не изменяя при этом произведение. *Нормi-*

рованным многочленом называется многочлен, у которого старший ненулевой коэффициент равен **1**; можно избежать указанной выше неопределенности, если при разложении многочлена всегда представлять его в виде произведения элемента поля и нормированных неприводимых множителей.

Теорема 6.4.3. (Единственность разложения.) Многочлен $f(D)$ над заданным полем единственным образом представляется в виде произведения элемента поля и нормированных неприводимых многочленов над данным полем, каждый из которых имеет степень, не меньшую 1.

Доказательство. Ясно, что входящий в разложение элемент поля единствен и равен f_n, где n — степень многочлена $f(D)$. Поэтому можно ограничиться рассмотрением нормированных многочленов. Предположим, что теорема неверна и существует некоторый нормированный многочлен $f(D)$ наименьшей степени, который может быть разложен двумя способами:

$$a_1(D)a_2(D)\ \ldots a_k(D) = b_1(D)\ldots\ b_j(D), \tag{6.4.18}$$

где многочлены $a_k(D)$ и $b_j(D)$ нормированные и неприводимые.

Каждый из многочленов $a_h(D)$ должен отличаться от каждого из $b_j(D)$; в противном случае должен существовать нормированный многочлен, степень которого меньше, чем степень $f(D)$, допускающий два различных разложения. Без ограничения общности допустим, что степень $b_1(D)$ не выше степени $a_1(D)$. Тогда получим

$$a_1(D) = b_1(D)h(D) + r(D), \tag{6.4.19}$$

где $r(D)$ имеет степень, меньшую,чем степень $b_1(D)$, и меньшую, чем $a_1(D)$. Подставляя (6.4.19) в (6.4.18), получаем

$$r(D)\,a_2(D)\ldots a_k(D) = b_1(D)\,[b_2(D)\ldots b_j(D) - h(D)\,a_2(D)\ldots a_k(D)].$$

Разложив $r(D)$ на множители и умножив произведение на элемент поля таким образом, чтобы нормировать множители, получим два разложения нормированного многочлена, имеющего степень, меньшую, чем степень $f(D)$, причем левое разложение на неприводимые множители не содержит в качестве множителя неприводимый многочлен $b_1(D)$. Полученное противоречие доказывает справедливость теоремы. |

Элемент α из поля называется *корнем* многочлена $f(D)$ над этим полем, если $f(\alpha) = 0$.

Теорема 6.4.4. Элемент α, принадлежащий полю, является корнем ненулевого многочлена $f(D)$ над этим полем тогда и только тогда, когда $(D - \alpha)$ является делителем $f(D)$. Кроме того, если $f(D)$ имеет степень n, то число элементов поля, являющихся корнями $f(D)$, не превосходит n.

Доказательство. Согласно алгоритму Евклида имеем

$$f(D) = (D - \alpha)h(D) + r(D). \tag{6.4.20}$$

Так как степень $(D-\alpha)$ равна 1, то степень $r(D)$ равна 0 и потому $r(D)$ является просто элементом поля r_0. Подставляя α вместо D, получаем $f(\alpha)=r_0$. Поэтому, если $f(\alpha)=0$, то $r_0=0$ и $(D-\alpha)$ является делителем $f(D)$. Обратно, если $(D-\alpha)$ является делителем $f(D)$, то имеем $f(D)=(D-\alpha)h(D)$ и $f(\alpha)=0$.

Теперь представим $f(D)$ в виде произведения элемента поля и неприводимых множителей степени не меньше 1. Так как степень $f(D)$ равна сумме степеней множителей, то существует не более n сомножителей. Это разложение единственно и, следовательно, $f(D)$ имеет не более n корней в рассматриваемом поле.

	0	1	t	$t+1$
0	0	1	t	$t+1$
1	1	0	$t+1$	t
t	t	$t+1$	0	1
$t+1$	$t+1$	t	1	0

Сложение

	0	1	t	$t+1$
0	0	0	0	0
1	0	1	t	$t+1$
t	0	t	$t+1$	1
$t+1$	0	$t+1$	1	t

* Умножение

Рис. 6.4.1. Поле многочленов в $GF(2)$ по модулю D^2+D+1.

Используем эти результаты, касающиеся многочленов, для конструирования нового примера конечного поля. Этот пример более важен, чем может показаться на первый взгляд; как будет показано в дальнейшем, он дает конкретное представление для *любого* конечного поля.

Пусть $f(D)$ — неприводимый многочлен степени n над конечным полем $GF(q)$; рассмотрим множество всех многочленов степени, не большей $(n-1)$ над $GF(q)$. Пусть $*$ означает специальную операцию над этими многочленами, которая дает вычет произведения многочленов по модулю $f(D)$,

$$g_1(D)*g_2(D)=R_{f(D)}[g_1(D)g_2(D)]. \tag{6.4.21}$$

Теперь покажем, что множество многочленов $g(t)$ над $GF(q)$, степени которых не превышают $n-1$, образует поле с операциями обычного сложения многочленов и умножения $*$ (неопределенная будет обозначаться в этом случае через t для напоминания о том, что рассматривается частное множество многочленов, степени которых не превышают $n-1$, и со специальной операцией $*$ умножения, в отличие от обычной операции умножения многочленов). Чтобы доказать, что получается поле, заметим, что аксиомы 1 и 3 непосредственно следуют из свойств сложения многочленов и умножения. Далее отметим, что при $g_1(t)*g_2(t)=0$ имеем

$$g_1(D)g_2(D)=f(D)h(D). \tag{6.4.22}$$

Однако, так как многочлен $f(D)$ неприводим и каждый из многочленов $g_1(D)$ и $g_2(D)$ имеет степень, меньшую чем $f(D)$, то из теоремы о единственности разложения следует, что либо $g_1(D)=0$, либо $g_2(D)=0$. Отсюда следует, что аксиома 2′в теореме 6.4.1 справедлива и мы имеем поле. Так как каждый из коэффициентов $g_0, g_1, \ldots, g_{n-1}$

может быть любым из q элементов первоначального поля, то вновь полученное поле содержит q^n элементов. *Назовем это новое поле полем многочленов над GF (q) по модулю f (D)*. В рассматриваемом случае необходимо, чтобы многочлен $f(D)$ был неприводимым, так как в противном случае $f(D) = g_1(D)g_2(D)$ для некоторых ненулевых g_1 и g_2 и, следовательно, $g_1(t) * g_2(t) = \mathbf{0}$, что противоречит (6.4.3).

Например, пусть $f(D) = D^2 + D + \mathbf{1}$ является многочленом над $GF(2)$. Тогда элементами поля по модулю $f(D)$ служат многочлены $\mathbf{0}, \mathbf{1}, t, t + \mathbf{1}$. На рис. 6.4.1 приведены таблицы сложения и $*$ умножения. Например, чтобы найти $t * t$, воспользуемся алгоритмом Евклида; получим $D^2 = (D^2 + D + \mathbf{1})\mathbf{1} + (D + \mathbf{1})$. Отсюда $R_{f(D)}[D^2] = D + \mathbf{1}$ и $t * t = t + \mathbf{1}$.

6.5. ЦИКЛИЧЕСКИЕ КОДЫ

В этом параграфе будет рассмотрен специальный класс кодов с проверкой на четность, известных под названием циклических кодов. Такие коды имеют два преимущества над обычными кодами с проверкой на четность: во-первых, операция кодирования для них легче в реализации и, во-вторых, математическая регулярность структуры кода делает возможным нахождение различных простых алгоритмов декодирования.

Прежде чем определить циклический код, обобщим коды с проверкой на четность на недвоичные алфавиты. Эти обобщенные коды будем называть *линейными* или *групповыми*, поскольку слово четность мало подходит для недвоичных алфавитов. Пусть $\mathbf{u} = (u_1, u_2, \ldots, u_L)$ — произвольная последовательность информационных символов, причем $\mathbf{u}_l$ является элементом некоторого конечного поля $GF(q)$. *Линейным (N, L)-кодом называется такой код, в котором кодовое слово* $\mathbf{x} = (x_1, \ldots, x_N)$, *соответствующее* $\mathbf{u}$, *является последовательностью* $N > L$ *букв из* $GF(q)$ *и образуется по правилу*

$$x_n = \sum_{l=1}^{L} u_l g_{l,n}; \quad 1 \leqslant n \leqslant N, \tag{6.5.1}$$

где элементы $g_{l,n}$ *произвольно выбираются среди элементов* $GF(q)$, *а сложение и умножение являются операциями в* $GF(q)$. Как и раньше, элементы $g_{l,n}$ будем представлять с помощью производящей матрицы G, как это показано на рис. 6.1.2, *а*, а образование кодовых слов будем описывать соотношением

$$\mathbf{x} = \mathbf{u}G. \tag{6.5.2}$$

Умножив обе части (6.5.2) на произвольный элемент α, принадлежащий $GF(q)$, можно убедиться, что если $\mathbf{x}$ — кодовое слово, соответствующее $\mathbf{u}$, то $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_N)$ — кодовое слово, соответствующее $\alpha\mathbf{u}$. Аналогично, если $\mathbf{x}_1$ и $\mathbf{x}_2$ — кодовые слова для $\mathbf{u}_1$ и $\mathbf{u}_2$ соответственно, то $\mathbf{x}_1 + \mathbf{x}_2$ — кодовое слово, соответствующее $\mathbf{u}_1 + \mathbf{u}_2$. На математическом языке это означает, что отображение информационных последовательностей в кодовые слова линейно и множество кодо-

вых слов образует линейное подпространство пространства N-мерных последовательностей над $GF(q)$.

Систематическим линейным кодом называется такой линейный код, у которого первые L компонент каждого кодового слова удовлетворяют соотношениям

$$x_l = u_l; \quad 1 \leqslant l \leqslant L. \tag{6.5.3}$$

Это достигается путем выбора $g_{l,n} = 1$ для $l = n$ и $g_{l,n} = 0$ для $n \leqslant L$, $n \neq l$. Порождающая матрица для систематического линейного кода представлена на рис. 6.1.2, *б*.

Проверочная матрица H, соответствующая систематическому линейному коду, должна быть некоторой модификацией матрицы, представленной на рис. 6.1.3, поскольку при переходе от (6.1.9) к (6.1.10) получаем

$$\mathbf{0} = \sum_{l=1}^{L} x_l g_{l,n} - x_n. \tag{6.5.4}$$

Поэтому матрица H принимает такой вид, как это представлено на рис. 6.5.1, и для всех кодовых слов выполняется соотношение

$$\mathbf{x}H = \mathbf{0}. \tag{6.5.5}$$

Несмотря на модификацию матрицы, результаты § 6.1 и 6.2 могут быть непосредственно перенесены на рассматриваемое обобщение кодов. Например, если кодовые слова передаются по каналу, у которого символы входного и выходного алфавитов являются элементами $GF(q)$, то можно представить принятую последовательность вектором $\mathbf{y}$, а шумовую последовательность — вектором $\mathbf{z}$, где $\mathbf{y} = \mathbf{x} + \mathbf{z}$. Определяя синдром $\mathbf{S}$ соотношением $\mathbf{S} = \mathbf{y}H$, получаем, как и ранее, что $\mathbf{S} = \mathbf{y}H = \mathbf{z}H$, и можно построить таблицу декодирования для нахождения $\mathbf{z}$ по синдрому $\mathbf{S}$.

Циклическим кодом над $GF(q)$ называется такой линейный код, у которого при любом циклическом сдвиге какого-либо кодового слова получается другое кодовое слово. Таким образом, если $(x_1, \ldots, x_N)$ — какое-либо кодовое слово, то $(x_2, x_3, \ldots, x_N, x_1)$ — другое кодовое слово. При изучении циклических кодов более удобно несколько изменить обозначения, последовательно перенумеровав элементы не с начала, а с конца, от $(N-1)$ до 0. Таким образом, запишем $\mathbf{x}$ в виде $(x_{N-1}, x_{N-2}, \ldots, x_0)$. Это эквивалентно следующему представлению последовательности $\mathbf{x}$ в виде многочлена над $GF(q)$:

$$x(D) = x_{N-1} D^{N-1} + x_{N-2} D^{N-2} + \ldots + x_0. \tag{6.5.6}$$

Если $x(D)$ — кодовое слово циклического кода (т. е. его коэффициенты являются буквами кодового слова), то вычет $Dx(D)$ по модулю $D^N - 1$ также является кодовым словом. Чтобы убедиться в этом, рассмотрим

$$Dx(D) = x_{N-1} D^N + \ldots + x_0 D, \tag{6.5.7}$$

$$Dx(D) = x_{N-1}(D^N - 1) + x_{N-2} D^{N-1} + \ldots + x_0 D + x_{N-1},$$

$$R_{D^N-1}[Dx(D)] = x_{N-2}D^{N-1} + \ldots + x_0 D + x_{N-1}. \qquad (6.5.8)$$

Допустим, что $g(D)$ — нормированный многочлен минимальной степени, являющийся кодовым словом циклического кода, и пусть степень $g(D)$ равна m. Из свойства линейности непосредственно следует, что если a_0 — произвольный элемент $GF(q)$, то $a_0 g(D)$ — также кодовое слово. Кроме того, из свойства цикличности непосредственно следует, что $a_1 Dg(D)$ является кодовым словом при всех a_1, принадлежащих $GF(q)$, и что $a_i D^i g(D)$ является кодовым словом для $i \leqslant N-1-m$. Складывая эти кодовые слова, в конце концов можно доказать, что для любого многочлена $a(D)$ над $GF(q)$ со степенью, не большей $N-1-m$, $a(D)\,g(D)$ является кодовым словом.

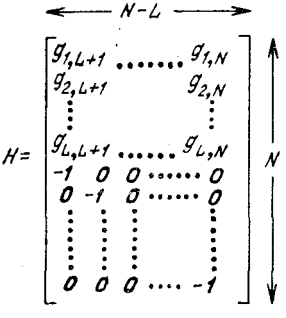


Рис. 6.5.1. Проверочная матрица систематического группового кода в $GF(q)$.

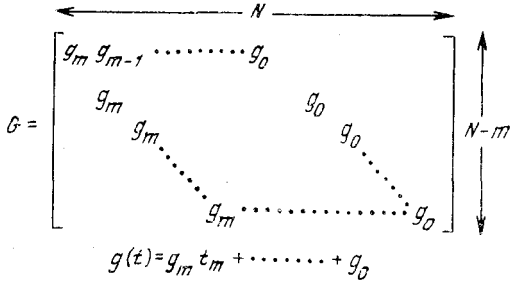


Рис. 6.5.2. Порождающая матрица циклического кода.

Покажем теперь, что *все* кодовые слова циклического кода представимы в таком виде. Согласно алгоритму Евклида деления многочленов, любой многочлен $x(D)$, степень которого не больше $N-1$, можно представить в виде

$$x(D) = a(D)g(D) + r(D). \qquad (6.5.9)$$

Так как $a(D)g(D)$ всегда является кодовым словом, то из свойства линейности вытекает, что если $x(D)$ — кодовое слово, то $r(D)$ — также кодовое слово. Но $r(D)$ имеет меньшую степень, чем $g(D)$; поэтому, если степень $r(D)$ отлична от нуля, то $r(D)$ всегда можно умножить на элемент $GF(q)$ таким образом, чтобы полученный в результате нормированный многочлен был бы кодовым словом, степень которого меньше степени $g(D)$. По определению $g(D)$ это невозможно и, следовательно, $r(D) = 0$. Таким образом, $x(D)$ является кодовым словом в том и только в том случае, если существует такой многочлен $a(D)$, степень которого не больше $N-m-1$, что

$$x(D) = a(D)g(D). \qquad (6.5.10)$$

Теперь покажем, что $g(D)$ должен быть делителем многочлена $D^N - 1$. Так как $g(D)$ — нормированный многочлен степени m, то

получаем

$$D^{N-m}g(D) = D^N - 1 + r(D). \qquad (6.5.11)$$

Поскольку $r(D) = R_{D^N-1}[D^{N-m}g(D)]$, то многочлен $r(D)$ получается циклическим сдвигом кодового слова и имеет $g(D)$ в качестве делителя. Тогда из соотношения (6.5.11) следует, что $g(D)$ является делителем $D^N - 1$.

Многочлен $g(D)$ называется порождающим многочленом циклического кода; смысл его состоит в том, что $g(D)$ является делителем всех кодовых слов. Множество кодовых слов — это множество линейных комбинаций $g(D)$ и его первых $N - m - 1$ циклических сдвигов. Порождающая матрица для циклического кода в несистематической форме представлена на рис. 6.5.2. В терминах числа информационных символов кода L получим, что $N - m = L$, а степень $g(D)$ равна $N - L$.

Теперь рассмотрим эту задачу несколько иначе; пусть $g(D)$ — любой нормированный многочлен в $GF(q)$ степени $N - L$, являющийся делителем $D^N - 1$. Покажем, что код, порождаемый $g(D)$ [т. е. код, у которого все кодовые слова являются линейными комбинациями $g(D)$ и его первых $L-1$ сдвигов], является циклическим кодом. Очевидно, что этот код линейный и имеет порождающую матрицу, представленную на рис. 6.5.2. Далее, если $x(D)$ — любое кодовое слово, то

$$Dx(D) = x_{N-1}(D^N - 1) + x_{N-2}D^{N-1} + \ldots + x_0 D + x_{N-1}. \qquad (6.5.12)$$

Так как $x(D)$ и $D^N - 1$ делятся на $g(D)$, то отсюда следует, что $x_{N-2}D^{N-1} + \ldots + x_0 D + x_{N-1}$ делится на $g(D)$ и что любой циклический сдвиг кодового слова является другим кодовым словом. Эти результаты можно сформулировать в виде следующей теоремы.

Теорема 6.5.1. Любой циклический код над $GF(q)$ с L информационными символами и с длиной блока N порождается нормированным многочленом $g(D)$ над $GF(q)$ степени $N - L$. Этот многочлен является делителем $D^N - 1$. Обратно, любой нормированный многочлен над $GF(q)$ степени $N - L$, который делит $D^N - 1$, порождает циклический код с L информационными символами и длиной блока N.

Точно так же, как порождающая матрица циклического кода может быть выражена через порождающий многочлен $g(D)$ $(N - L)$-й степени, проверочная матрица может быть выражена через многочлен L-й степени $h(D)$, называемый проверочным многочленом и определяемый соотношением

$$g(D)h(D) = D^N - 1. \qquad (6.5.13)$$

Если умножить любое кодовое слово $x(D) = a(D)g(D)$ на $h(D)$, то получим

$$f(D) = x(D)h(D) = a(D)g(D)h(D) = D^N a(D) - a(D). \qquad (6.5.14)$$

Так как степень $a(D)$ не превышает $L - 1$, то из рассмотрения

(6.5.14) следует, что $f(D) = \Sigma f_n D^n$ должен иметь нулевые слагаемые при $L \leqslant n \leqslant N-1$. Произведя умножение $x(D)\,h(D)$, получим

$$\sum_{i=0}^{L} h_i x_{n-i} = f_n = 0; \quad L \leqslant n \leqslant N-1. \tag{6.5.15}$$

Из (6.5.13) следует, что $h(D)$ — нормированный многочлен и что $h_L = 1$. Поэтому можно переписать (6.5.15) в виде

$$x_{n-L} = -\sum_{i=0}^{L-1} h_i x_{n-i}; \quad L \leqslant n \leqslant N-1. \tag{6.5.16}$$

Равенство (6.5.16) дает рекуррентное соотношение для вычисления в соответствующем порядке проверочных символов x_{N-L-1}, x_{N-L-2}, ..., x_0 по информационным символам x_{N-1}, ..., x_{N-L}. Поэтому любой $x(D)$, удовлетворяющий (6.5.15), является кодовым словом и любое кодовое слово удовлетворяет (6.5.15). На языке матриц это означает, что $\mathbf{x}H = \mathbf{0}$ тогда и только тогда, когда $\mathbf{x} = (x_{N-1}, \ldots, x_0)$ является кодовым словом, где H представлено на рис. 6.5.3.

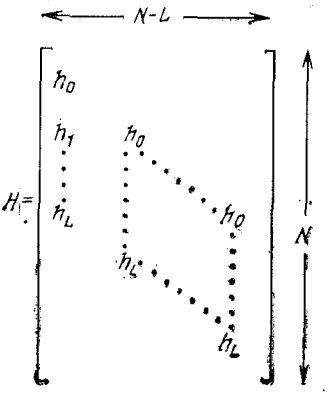


Рис. 6.5.3. Проверочная матрица циклического кода.

Так как многочлен $h(D)$ является делителем $D^N - 1$, то он также порождает циклический код, который называется *дуальным кодом* по отношению к коду, порождаемому $g(D)$. Поэтому все циклические сдвиги $h(D)$ являются линейными комбинациями $N-L$ сдвигов, изображенных на рис. 6.5.3, и часто удобно использовать эти дополнительные сдвиги в качестве проверочных соотношений, которым должны удовлетворять кодовые слова, порождаемые $g(D)$.

До настоящего момента рассмотрение носило в основном абстрактный характер, поэтому настало время обсудить вопрос о том, как в действительности можно реализовать кодирование циклического кода. Первый из способов кодирования основывается на (6.5.16) и иллюстрируется рис. 6.5.4.

Первоначально информационные символы хранятся в разрядах регистра сдвига, как это представлено на рис. 6.5.4. Содержимое в каждом разряде подается на выход разряда (т. е. на линию, выходящую из правой части разряда) и умножается на соответствующее h_i; эти произведения суммируются и умножаются на -1, что дает в итоге x_{N-L-1}, как это следует из (6.5.16). Затем регистр сдвига сдвигается на одну позицию вправо, символ x_{N-1} поступает в канал, а символ x_{N-L-1} подается в левый разряд регистра. После этого сдвига на выходе умножителя на (-1)появляется символ x_{N-L-2}. После N сдвигов в канал будет передан x_0, а кодер будет подготовлен к поступлению в него следующей информационной последовательности.

В случае двоичных циклических кодов техническая реализация этого устройства довольно проста; устройство включает в себя лишь L-разрядный двоичный регистр сдвига и сумматоры по модулю 2. Умножение на h_i производится просто путем размыкания цепи при $h_i = 0$ и замыкания — при $h_i = 1$. В случае циклических кодов в произвольном поле цепь немного более сложная, поскольку требуются устройства, выполняющие сложение и умножение элементов в этом поле.

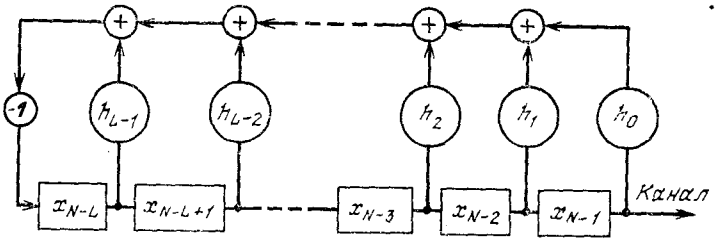


Рис. 6.5.4. L-разрядный кодер циклического кода.

Теперь обсудим другую реализацию циклического кодера, для которой требуется регистр сдвига на $N - L$ разрядов. Ясно, что выбор той или иной реализации зависит от того, велико или мало отношение числа информационных символов к длине блока. Информацион-

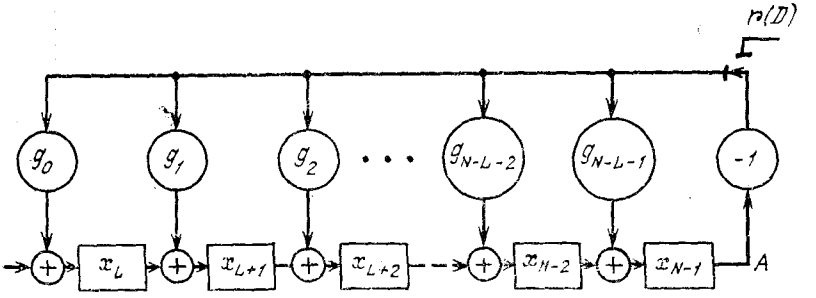


Рис. 6.5.5. $(N - L)$-разрядный кодер циклического кода.

ные символы в кодовом слове могут быть представлены в виде многочлена $x_{N-1}D^{N-1} + \ldots + x_{N-L}D^{N-L}$. Согласно алгоритму Евклида,

$$x_{N-1}D^{N-1} + \ldots + x_{N-L}D^{N-L} = a(D)\,g(D) + r(D), \qquad (6.5.17)$$

где степень $r(D)$ не превышает $N - L - 1$.

Из (6.5.17) следует, что $x_{N-1}D^{N-1} + \ldots + x_{N-L}D^{N-L} - r(D)$ является кодовым словом, а $-r(D)$ — многочлен, соответствующий проверочным символам. Любое устройство, совершающее деление $x_{N-1}D^{N-1} + \ldots + x_{N-L}D^{N-L}$ на $g(D)$, позволяет найти $r(D)$; такое устройство представлено на рис. 6.5.5.

Первоначально в регистре сдвига этого устройства хранятся первые $N - L$ информационных символов, причем правые разряды оста-

ются свободными, если число информационных символов меньше чем $N - L$. Первая операция состоит в вычитании из x_{L-1+j} произведений $g_j x_{N-1}$ при $0 \leqslant j \leqslant N - L - 1$, после чего регистр сдвигается на одну позицию вправо, и, если возможно, добавляется новый информационный символ. Символ x_{N-1} больше не требуется для вычисления $r(D)$ и поэтому сбрасывается. Можно показать, что хранящиеся в этот момент в разрядах регистра сдвига числа точно совпадают с теми числами, которые получились бы после первого вычитания, если бы деление многочленов производилось на бумаге. Этот цикл повторяется L раз; рассуждая, как в предыдущем случае, можно убедиться, что после этого в регистре останется многочлен $r(D)$; при этом слагаемые более высокого порядка хранятся в правых разрядах. После этого ключ в правом углу на рис. 6.5.5 переводится в вертикальное положение и $-r(D)$ считывается и передается в канал.

Из свойства линейности рассмотренной цепи следует, что аналогичный результат может быть получен и в том случае, когда первоначально регистр пуст; при этом информационные символы поступают в устройство в точке A через сумматор, который прибавляет поступающие информационные символы к выходным символам самого правого разряда регистра.

6.6. ПОЛЯ ГАЛУА

В настоящем параграфе будет изучена несколько подробнее, чем в § 6.4, структура полей Галуа. Полученные результаты потребуются нам в следующем параграфе, посвященном кодам Боуза — Чоудхури — Хоквингема (БЧХ); они играют главенствующую роль во многих проводящихся в настоящее время исследованиях по технике алгебраического кодирования.

Начнем с изучения мультипликативного порядка различных ненулевых элементов поля Галуа $GF(q)$, состоящего из q элементов. Так как ненулевые элементы поля, обозначенные, допустим, символами $\alpha_1, \alpha_2, \ldots, \alpha_{q-1}$, образуют абелеву группу по умножению в рассматриваемом поле, то из теоремы 6.3.2 следует, что каждый из этих элементов имеет мультипликативный порядок, на который делится число $q - 1$. Поэтому $\alpha_i^{q-1} - \mathbf{1} = \mathbf{0}$ при всех i, $1 \leqslant i \leqslant q - 1$. Отсюда следует, что каждый элемент α_i является корнем многочлена $D^{q-1} - \mathbf{1}$ [рассматриваемого здесь в качестве многочлена над $GF(q)$]. Так как степень этого многочлена равна $q - 1$ и известны ровно $q - 1$ его различных корней, т. е. $\alpha_1, \ldots, \alpha_{q-1}$, то получим

$$D^{q-1} - \mathbf{1} = \prod_{i=1}^{q-1} (D - \alpha_i). \tag{6.6.1}$$

Например, в поле целых чисел по модулю 3 это равенство будет

$$D^2 - \mathbf{1} = (D - \mathbf{1})\ \ (D - \mathbf{2}),$$

которое, очевидно, удовлетворяется, если производить сложение и умножение целых чисел по модулю 3.

Выше было показано, что все ненулевые элементы поля, состоящего из q элементов, имеют мультипликативный порядок, на который делится число $q-1$. *Примитивным элементом поля из q элементов называется элемент, у которого мультипликативный порядок точно равен* $q-1$. Если можно найти некоторый примитивный элемент поля, допустим α, то последовательность $\alpha, \alpha^2, \ldots, \alpha^{q-1}$ содержит все ненулевые элементы поля, а мультипликативная группа ненулевых элементов поля является циклической. При этом ясно, что не представляет труда нахождение мультипликативного порядка любого элемента, являющегося степенью элемента α (см. задачу 6.26).

Теорема 6.6.1. Любое поле Галуа содержит примитивный элемент.

Доказательство. Пусть максимальный мультипликативный порядок ненулевых элементов поля $GF(q)$ равен m. Из теоремы 6.3.3. следует, что каждый ненулевой элемент имеет мультипликативный порядок, который является делителем m, и поэтому этот элемент является корнем $D^m - 1$. Так как этот многочлен имеет $q-1$ корней, то имеем $m \geqslant q-1$. Поскольку согласно (6.6.1) каждый ненулевой элемент поля имеет мультипликативный порядок, на который делится число $q-1$, то $m \leqslant q-1$, что завершает доказательство. |

Доказанная выше теорема в принципе полностью характеризует мультипликативную группу ненулевых элементов поля Галуа, однако она ничего не говорит о структуре аддитивной группы или о связи между умножением и сложением. Ниже эта связь будет найдена и показано, что все поля Галуа могут быть представлены как поля многочленов по модулю неприводимого многочлена. Прежде всего *определим подполе некоторого поля как поле, элементы которого образуют подмножество элементов первоначального поля с операциями сложения и умножения, совпадающими с операциями первоначального поля. Если какое-либо поле F имеет подполе E, то говорят, что F является расширением E.*

Теорема 6.6.2. Любое поле Галуа содержит единственное подполе, число элементов в котором простое.

Доказательство. Любое подполе должно содержать элементы поля **0** и **1**. Оно должно также включать в себя $\mathbf{1}+\mathbf{1}$, $\mathbf{1}+\mathbf{1}+\mathbf{1}$, и т. д. Обозначим эти получающиеся при сложении элементы через **2**, **3**, **4**, ...; из теоремы 6.3.2 следует, что эти элементы, называемые *целыми* элементами поля, образуют циклическую подгруппу по операции сложения. Если эта подгруппа имеет p элементов, то сложение этих элементов является сложением по модулю p. Согласно дистрибутивному закону умножение этих элементов также является умножением по модулю p [т. е. $\mathbf{2}\cdot\mathbf{3} = (\mathbf{1}+\mathbf{1})\cdot\mathbf{3} = \mathbf{3}+\mathbf{3} = R_p$ (6)]. Отсюда следует, что число p простое, поскольку если бы p было произведением двух целых чисел $i > 1$ и $j > 1$, то для соответствующих им целых элементов поля удовлетворялось бы равенство $\mathbf{i}\cdot\mathbf{j} = \mathbf{0}$. Это невозможно, так как **i** и **j** — ненулевые элементы первоначального поля. Как было показано, целые по модулю простого числа элементы образуют поле, поэтому это поле как раз и является тем подполем, которое разыскивается. На-

конец, любое другое подполе должно содержать эти целые элементы поля, а аддитивная группа любого подполя должна содержать эти целые элементы как подгруппу. Поэтому число элементов в любом другом подполе делится на p и потому не является простым. |

Из этой теоремы непосредственно следует, что любое поле или подполе о простым числом элементов является при соответствующей нумерации элементов полем целых элементов по модулю простого числа. Характеристикой p поля Галуа называется число элементов в определенном выше простом подполе.

Если $P(D) = P_0 + P_1D + \ldots + P_nD^n$ — многочлен над полем E и если поле F является расширением E, то говорят, что элемент α из F является корнем многочлена $P(D)$, когда $P(\alpha) = 0$, т. е. когда

$$\sum_i P_i \alpha^i = 0.$$

В качестве часто используемого примера укажем, что привычно находить комплексные корни многочленов над полем действительных чисел. Если E является подполем поля Галуа F, то *минимальный многочлен $f_\alpha(D)$ над E элемента α, принадлежащего F, определяется как нормированный многочлен минимальной степени над E, у которого α является корнем*. В дальнейшем во всех случаях, когда подполе явно не определено, мы будем иметь в виду подполе с простым числом элементов, фигурирующее в теореме 6.6.2.

Теорема 6.6.3. Для любого подполя E поля Галуа $GF(q)$ каждому ненулевому элементу α, принадлежащему $GF(q)$, соответствует единственный минимальный многочлен $f_\alpha(D)$ над E, причем $f_\alpha(D)$ — неприводимый. Более того, для любого многочлена $P(D)$ над E многочлен $f_\alpha(D)$ является делителем $P(D)$ тогда и только тогда, когда α является корнем $P(D)$.

Доказательство. Было уже показано, что α является корнем $D^{q-1} - 1$. Так как $D^{q-1} - 1$ можно рассматривать как многочлен над E, то отсюда следует существование таких многочленов над E, для которых α является корнем, и, следовательно, существует некоторый нормированный многочлен минимальной степени, обозначаемый $f_\alpha(D)$, для которого α является корнем. Если бы $f_\alpha(D)$ был приводимым, то он мог бы быть представлен в виде произведения двух нормированных многочленов меньшей степени над полем E, т. е. $f_\alpha(D) = g(D)h(D)$. Отсюда следует, что $g(\alpha)h(\alpha) = 0$, и так как $g(\alpha)$ и $h(\alpha)$ — элементы, принадлежащие $GF(q)$, то один из них должен быть равным 0, что противоречит предположению о том, что $f_\alpha(D)$ имеет минимальную степень. Поэтому $f_\alpha(D)$ неприводим. Тогда для любого $P(D)$ над E

$$P(D) = f_\alpha(D)h(D) + r(D), \tag{6.6.2}$$

где $r(D)$ — многочлен над E, степень которого меньше степени $f_\alpha(D)$. Так как $f_\alpha(D) = 0$, то $P(\alpha) = r(\alpha)$. Поэтому $P(\alpha) = 0$ тогда и только тогда, когда $r(\alpha) = 0$. Но поскольку степень $r(D)$ меньше, чем степень $f_\alpha(D)$, то получаем, что $r(\alpha) = 0$, в свою очередь, тогда и толь-

ко тогда, когда $r(D)$ является нулевым многочленом. Поэтому $P(\alpha) = 0$ тогда и только тогда, когда $f_\alpha(D)$ является делителем $P(D)$. Наконец, если $P(D)$ является нормированным многочленом той же самой степени, что и $f_\alpha(D)$, то равенство $P(D) = f_\alpha(D)\, h(D)$ выполняется лишь при $h(D) = 1$, что доказывает единственность $f_\alpha(D)$. |

В доказанной выше теореме утверждение о неприводимости многочлена $f_\alpha(D)$ означает, что над полем E не существует многочленов меньшей степени, произведение которых было бы равно $f_\alpha(D)$. Если интерпретировать $f_\alpha(D)$ как многочлен над $GF(q)$, то $D - \alpha$ является очевидным делителем $f_\alpha(D)$.

Следствие. Пусть E — подполе поля Галуа $GF(q)$ и пусть $f_1(D), \ldots, f_L(D)$ — различные минимальные многочлены над E, соответствующие ненулевым элементам $GF(q)$. [т. е. последовательность $f_\alpha(D), \ldots, f_{\alpha_{q-1}}(D)$, из которой исключены одинаковые многочлены]. Тогда

$$D^{q-1} - 1 = \prod_{i=1}^{L} f_i(D). \tag{6.6.3}$$

Заметим, что соотношение (6.6.1) описывает разложение $D^{q-1} - 1$ на неприводимые многочлены над $GF(q)$, а (6.6.3) — на неприводимые многочлены над E.

Доказательство. Так как ненулевые элементы из $GF(q)$ являются корнями $D^{q-1} - 1$, то каждый минимальный многочлен является делителем $D^{q-1} - 1$. Поскольку минимальные многочлены неприводимы (над полем E), то произведение $\prod_{i=1}^{L} f_i(D)$ также является делителем $D^{q-1} - 1$. Наконец, все ненулевые элементы $GF(q)$ являются корнями многочлена $\prod_{i=1}^{L} f_i(D)$.

Следовательно, указанное выше произведение имеет степень, не меньшую $q - 1$, и (6.6.3) выполняется. |

Теорема 6.6.4. Пусть α — примитивный элемент поля Галуа $GF(q)$ характеристики p и пусть степень минимального многочлена $f(D)$ над $GF(p)$ элемента α равна n. Тогда число элементов в поле Галуа равно p^n и каждый элемент поля β может быть представлен в виде

$$\beta = \mathbf{i}_{n-1}\alpha^{n-1} + \mathbf{i}_{n-2}\alpha^{n-2} + \ldots + \mathbf{i}_1\alpha + \mathbf{i}_0 \tag{6.6.4}$$

при некотором наборе целых элементов поля $\mathbf{i}_0, \mathbf{i}_1, \ldots, \mathbf{i}_{n-1}$.

Доказательство. Так как α является корнем $f(D) = D^n + f_{n-1}D^{n-1} + \ldots + f_0$, то имеем $\alpha^n + f_{n-1}\alpha^{n-1} + \ldots + f_0 = 0$, или

$$\alpha^n = -\sum_{i=0}^{n-1} f_i \alpha^i. \tag{6.6.5}$$

Поскольку каждый из коэффициентов f_i является целым элементом

поля, то нетрудно видеть, что элемент α^n может быть представлен в виде (6.6.4). Умножая (6.6.5) на α, получаем

$$\alpha^{n+1} = -\sum_{i=0}^{n-1} f_i \alpha^{i+1}. \tag{6.6.6}$$

В силу того, что α^n можно представить в виде (6.6.4), α^{n+1} также можно представить в таком виде; последовательно умножая (6.6.6) на более высокие степени α, убеждаемся, что все степени α могут быть представлены в виде (6.6.4). Поэтому каждый элемент $GF(q)$ может быть представлен в виде (6.6.4). Теперь допустим, что имеются два различных множества целых элементов поля, входящих в (6.6.4), например $\mathbf{i}_{n-1}, \ldots, \mathbf{i}_0$ и $\mathbf{j}_{n-1}, \ldots, \mathbf{j}_0$, которые соответствуют одному и тому же элементу из $GF(q)$. Тогда

$$\mathbf{0} = (\mathbf{i}_{n-1} - \mathbf{j}_{n-1})\,\alpha^{n-1} + \ldots + (\mathbf{i}_1 - \mathbf{j}_1)\,\alpha + (\mathbf{i}_0 - \mathbf{j}_0). \tag{6.6.7}$$

Отсюда следует, что α является корнем многочлена

$$(\mathbf{i}_{n-1} - \mathbf{j}_{n-1})\,D^{n-1} + \ldots + (\mathbf{i}_0 - \mathbf{j}_0),$$

что невозможно, поскольку степень этого многочлена меньше n. Поэтому каждый из наборов $\mathbf{i}_{n-1}, \ldots, \mathbf{i}_0$ соответствует различным элементам поля и так как существует всего p^n таких наборов целых элементов поля, то $q = p^n$. |

Эта теорема имеет ряд интересных следствий. Во-первых, поскольку характеристикой всякого поля Галуа является некоторое простое число p, то число элементов во всяком поле Галуа должно представляться в виде $q = p^n$, где p — некоторое простое число, а n — некоторое целое число. Далее, если в соотношении (6.6.4) заменить α на не определенную t, то можно убедиться, что множество элементов поля может рассматриваться как множество многочленов над $GF(p)$ степени, не большей $n-1$ с умножением по модулю $f(t)$. Другими словами, это поле после введения в нем новых обозначений элементов будет совпадать с полем многочленов над $GF(p)$ по модулю $f(t)$ (т. е. будет иметь то же самое множество элементов и те же самые правила сложения и умножения). Два таких поля, отличающиеся лишь обозначениями элементов, называются *изоморфными*; из приведенных рассуждений следует, что *всякое* поле с p^n элементами изоморфно некоторому полю многочленов над $GF(p)$ по модулю некоторого неприводимого многочлена степени n. Наконец, согласно (6.6.3) из единственности разложения многочлена $D^{p^n-1} - 1$ над $GF(p)$ на неприводимые множители следует, что *всякое* поле с p^n элементами имеет одно и то же множество минимальных многочленов. Поэтому в любом поле с p^n элементами можно выбрать в качестве α корень фиксированного многочлена $f(D)$, использованного в теореме (6.6.4), и представить все элементы поля в виде (6.6.4). Таким образом, доказана следующая теорема.

Теорема 6.6.5. Все поля Галуа с p^n элементами изоморфны данному полю многочленов над $GF(p)$ по модулю неприводимого многочлена n-й степени.

Коды максимальной длины и коды Хэмминга

Допустим, что $h(D)$ — минимальный многочлен некоторого примитивного элемента α из $GF(p^m)$. Из теоремы 6.6.5 следует, что $h(D)$ остается минимальным многочленом примитивного элемента при любом представлении $GF(p^m)$; такой многочлен называется *примитивным многочленом* степени m. Циклический код с длиной блока $N = p^m - 1$, для которого примитивный многочлен $h(D)$ является проверочным многочленом, называется кодом максимальной длины. Кодовые слова такого кода можно порождать с помощью устройства, имеющего регистр сдвига и представленного на рис. 6.5.4 и вновь воспроизведенного на рис. 6.6.1.

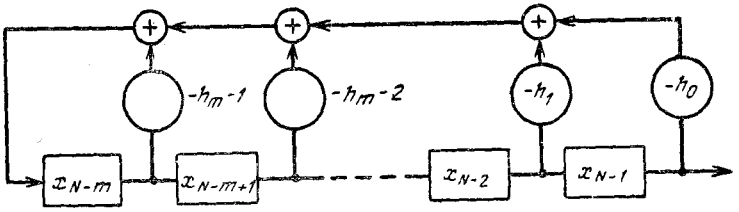


Рис. 6.6.1. Кодер для кода максимальной длины.

Одно из кодовых слов в этом коде соответствует порождающему многочлену $g(D) = [D^{p^m-1} - 1]/h(D)$. Покажем, что множество кодовых слов состоит из чисто нулевой последовательности и множества циклических сдвигов $g(D)$. Так как число информационных символов в коде равно m, то число кодовых слов равно p^m. Так как число циклических сдвигов многочлена $g(D)$ равно $p^m - 1$ [включая сам $g(D)$], то достаточно показать, что все циклические сдвиги $g(D)$ различны. Для этого предположим, что i-й и j-й циклические сдвиги дают одинаковые многочлены; $0 \leqslant i < j < p^m - 1$. Тогда получим

$$R_{D^{p^m-1}-1}[D^i g(D)] = R_{D^{p^m-1}-1}[D^j g(D)].$$

Это означает, что при некоторых $a(D)$ и $b(D)$

$$D^i g(D) - a(D)[D^{p^m-1} - 1] = D^j g(D) - b(D)[D^{p^m-1} - 1].$$

Разделив обе части равенства на $g(D)$ и преобразовав выражение, получим

$$[b(D) - a(D)]\, h(D) = D^j - D^i = D^i[D^{j-i} - 1].$$

Наконец, в силу того, что $j - i < p^m - 1$, получим, что примитивный элемент α не может быть корнем $D^{j-i} - 1$ и, следовательно [согласно теореме 6.6.3], $h(D)$ не является делителем $D^{j-i} - 1$. Полученное противоречие показывает, что все циклические сдвиги $g(D)$ различны.

Так как каждое кодовое слово определяется своими m информационными символами, из приведенных выше рассуждений следует, что

всем циклическим сдвигам $g(D)$ соответствуют различные начальные множества m символов и что они исчерпывают все $p^m - 1$ ненулевые комбинации m символов из $GF(q)$. Наглядно это интерпретируется на рис. 6.6.2, где представлено кодовое слово $g(D)$, соединенное своими концами в кольцо, так что циклические сдвиги $g(D)$ могут считываться с кольца, если начинать считывание с различных его точек.

Каждая последовательность без пропусков в кольце, состоящая из m символов, является информационной последовательностью одного-единственного циклического сдвига $g(D)$. Кроме того, можно показать, что информационная последовательность i-го циклического сдвига $g(D)$ совпадает с последовательностью, хранящейся в регистре сдвига на рис. 6.6.1 (причем наиболее «продвинутый» по ходу часовой стрелки символ кольца соответствует символу левого разряда регистра), после i сдвигов регистра при генерировании $g(D)$.

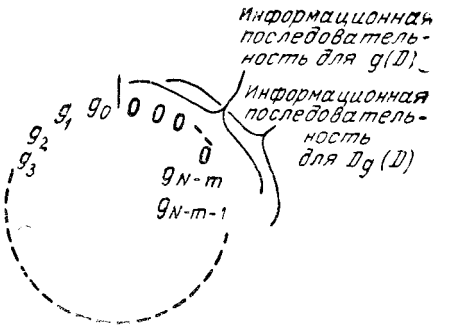


Рис. 6.6.2. Символы $g(D)$, расположенные по кругу.

Теперь заметим, что ровно p^{m-1} из $p^m - 1$ ненулевых последовательностей длины m начинаются каждым из $p - 1$ ненулевых элементов поля $GF(p)$ и ровно $p^{m-1} - 1$ начинаются нулевым элементом. Поскольку каждый символ кодового слова, соответствующего $g(D)$, является начальным символом для одной из указанных выше последовательностей длины m, то каждый ненулевой символ встречается в любом ненулевом кодовом слове точно p^{m-1} раз, а 0 встречается $p^{m-1} - 1$ раз. Поскольку разность двух кодовых слов является другим кодовым словом, то отсюда следует, что каждая пара различных кодовых слов совпадает в $p^{m-1} - 1$ позициях и отличается во всех остальных позициях. Можно показать (см. задачу 6.24), что рассматриваемый код обладает максимальным числом позиций, в которых могут отличаться все пары кодовых слов с данной длиной блока; поэтому такие коды весьма эффективны при исправлении ошибок.

Нетрудно убедиться, что если регистр сдвига на рис. 6.6.1 генерирует бесконечную последовательность символов, не останавливаясь после получения $p^m - 1$ символов, то получится периодическая последовательность с периодом $p^m - 1$; последовательные символы последовательности можно найти путем считывания по часовой стрелке символов с кольца, представленного на рис. 6.6.2. Так как будущий выход регистра сдвига с обратной связью зависит лишь от того, что хранится в нем в настоящий момент, то нетрудно убедиться, что после того как хранящиеся в нем символы совпадут с символами, хранящимися в какой-либо предыдущий момент времени, выходные символы регистра должны повторяться периодически. Если не считать нулевой последовательности, в m-разрядном регистре могут храниться

лишь $p^m - 1$ возможных сочетаний символов; поэтому выходные символы любого m-разрядного регистра с обратной связью и с элементами из $GF(p)$ являются периодическими, причем период не больше $p^m - 1$. Так как при использовании для обратной связи примитивного многочлена достигается именно этот максимальный период, то такие регистры называются регистрами *максимальной длины* с обратной связью, а коды называются кодами максимальной длины.

Периодические последовательности, выходящие из регистра сдвига максимальной длины с обратной связью, называются *псевдошумовыми* или $(p - n)$-последовательностями. Если выбрать последовательность i символов ($i \leqslant m$), случайно выбирая ее места в указанной выше последовательности, то вероятность ненулевой последовательности будет равна $p^{m-i}/(p^m - 1)$, а вероятность нулевой последовательности будет равна $(p^{m-i} - 1)/(p^m - 1)$. Поэтому при больших m для широкого круга задач можно рассматривать результирующую по-

Элементы $GF(2^4)$ как степени α	Элементы $GF(2^4)$ как многочлены $g(t)$				Минимальные многочлены
	g_3	g_2	g_1	g_0	
0	0	0	0	0	
1	0	0	0	1	$D+1$
α	0	0	1	0	D^4+D+1
α^2	0	1	0	0	D^4+D+1
α^3	1	0	0	0	$D^4+D^3+D^2+D+1$
α^4	0	0	1	1	D^4+D+1
α^5	0	1	1	0	D^2+D+1
α^6	1	1	0	0	$D^4+D^3+D^2+D+1$
α^7	1	0	1	1	D^4+D^3+1
α^8	0	1	0	1	D^4+D+1
α^9	1	0	1	0	$D^4+D^3+D^2+D+1$
α^{10}	0	1	1	1	D^2+D+1
α^{11}	1	1	1	0	D^4+D^3+1
α^{12}	1	1	1	1	$D^4+D^3+D^2+D+1$
α^{13}	1	1	0	1	D^4+D^3+1
α^{14}	1	0	0	1	D^4+D^3+1

Рис. 6.6.3. Элементы $GF(2^4)$ как многочлены по модулю D^4+D+1.

следовательность как случайную последовательность независимых равновероятных символов из $GF(p)$. Эти последовательности часто используются для определения дальности целей и синхронизации в радиолокации и связи. Например, путем использования m-разрядного регистра сдвига можно в двоичном случае добиться разрешения $2^m - 1$ различных значений дальности, или позиций синхронизации.

Теперь рассмотрим дуальный код для кода максимальной длины. В этом случае примитивный многочлен $h(D)$ является порождающим многочленом для кода, а $g(D)$ — проверочным. При этом столбцы проверочной матрицы можно трактовать как $g(D)$ и $(m - 1)$ его первые циклические сдвиги. Поэтому множество строк проверочной матрицы является множеством последовательностей m следующих друг за дру-

гом символов на кольце, представленном на рис. 6.6.2. Следовательно, все эти строки различны и получается циклическое представление кода Хэмминга.

Пример. Теперь обсудим более детально частный случай поля Галуа $GF\ (2^4)$ отчасти для того, чтобы сделать теорию чуть менее абстрактной, а отчасти для того, чтобы рассмотреть, как на практике производятся операции в поле Галуа. В качестве представления $GF\ (2^4)$ можно рассмотреть поле многочленов над $GF\ (2)$ по модулю многочлена $f\ (D) = D^4 + D + 1$. Разделив $f\ (D)$ на все многочлены первой и второй степени над $GF\ (2^4)$, можно убедиться, что $f\ (D)$ — неприводимый многочлен. На рис. 6.6.3 элементы $GF\ (2^4)$ представлены двумя способами, при одном из них — степенями элемента поля α, соответствующего многочлену t, а при другом — в виде многочленов $g(t) = g_3t^3 + g_2t^2 + g_1t + g_0$. В § 6.4 было указано, что сложение элементов поля производится путем сложения многочленов, а умножение — путем умножения многочленов по модулю $f\ (t)$. Умножение на элемент поля $\alpha = t$ выполняется особенно легко и может быть реализовано устройством, изображенным на рис. 6.6.4. После того как в регистр подан многочлен $g\ (t)$, регистр сдвигается на один разряд вправо, что соответствует умножению на t, а затем приводится по модулю $f(t)$ путем использования обратных связей. Читатель может проверить, что при последовательных сдвигах регистра, изображенного на рис. 6.6.4, последовательно генерируются многочлены, приведенные на рис. 6.6.3.

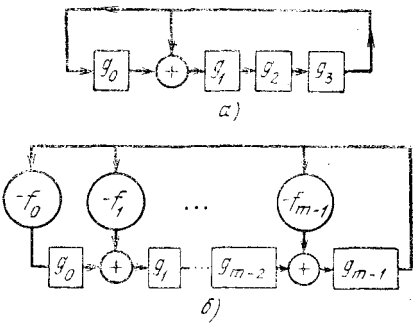


Рис. 6.6.4. Устройство для умножения произвольного элемента поля на $\alpha = t$ в поле многочленов по модулю $f(D)$:
а) $f(D) = D^4 + D + 1$; *б)* произвольный многочлен $f(D)$ степени m.

Покажем, что элемент поля $\alpha = t$ на рис. 6.6.3 имеет мультипликативный порядок 15 и потому примитивный . Для любого поля многочленов по модулю многочлена $f\ (D)$ минимальным многочленом для элемента поля $\alpha = t$ является $f\ (D)$ [так как вычет $f\ (t)$ по модулю $f\ (t)$ равен 0]; поэтому то обстоятельство, что α примитивный, означает, что нам повезло при выборе $f\ (D)$ в качестве примитивного многочлена.

Минимальные многочлены для всех элементов поля $GF\ (2^4)$ представлены на рис. 6.6.3. В задаче 6.29 разрабатывается способ вычисления этих минимальных многочленов, который основан на последующих результатах этого параграфа. Питерсон (1961) затабулировал эти минимальные многочлены для полей с характеристикой 2 вплоть до $GF\ (2^{17})$; поэтому, как правило, нет необходимости проделывать эти вычисления. В данном случае, однако, можно по крайней мере проверить, что приведенные на рис. 6.6.3 минимальные многочлены правильны. Например, минимальный многочлен для α^{10} записан в виде

$D^2 + D + 1$. Для того чтобы α^{10} было корнем $D^2 + D + 1$, должно выполняться $\alpha^{20} + \alpha^{10} + 1 = 0$. Так как $\alpha^{15} = 1$, то $\alpha^{20} = \alpha^5$, что приводит к соотношению $\alpha^{10} + \alpha^5 + 1 = 0$; складывая многочленные представления α^{10}, α^5 и 1, получаем, что утверждение действительно верно.

Для практического выполнения операций в произвольном поле Галуа $GF(p^m)$ обычно удобно представлять элементы поля как многочлены по модулю примитивного многочлена m-й степени над $GF(p)$; с точки зрения реализации это означает представление каждого элемента в виде последовательности m элементов в $GF(p)$. Как было показано, сложение в $GF(p^m)$ при этом соответствует сложению соответствующих последовательностей в $GF(p)$. Умножение на специально

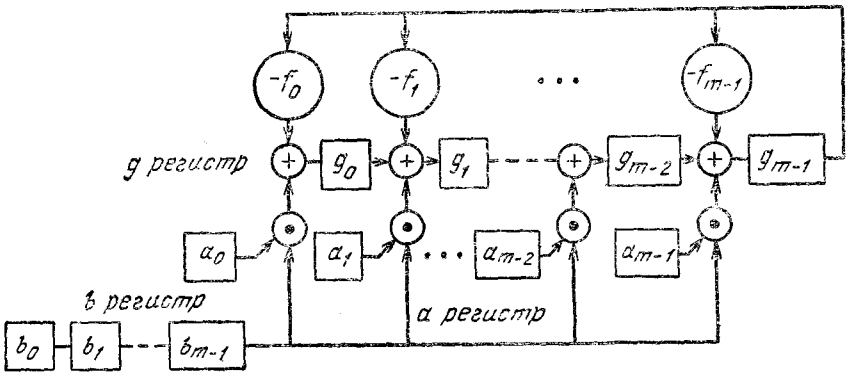


Рис. 6.6.5. Умножение элементов поля $a(t)$ и $b(t)$ в поле многочленов по модулю $f(t)$.

выбранный элемент $\alpha = t$ также выполняется устройством, представленным на рис. 6.6.4. И, наконец, умножение двух произвольных элементов поля может производиться устройством, изображенным на рис. 6.6.5.

Вначале при выполнении умножения этим устройством регистр g пуст, а $a(t)$ и $b(t)$ находятся в регистрах a и b соответственно. Затем регистры b и g сдвигаются на одну позицию вправо таким образом, что в регистре g будет многочлен $b_{m-1}a(t)$. Затем регистры b и g вновь сдвигаются вправо. После этого в регистре g будет храниться $R_{f(t)}[(b_{m-1}t + b_{m-2})a(t)]$. После m сдвигов регистров g и b в регистре g будет $R_{f(t)}[b(t)a(t)]$, или элемент поля $b(t) * a(t)$. Следует заметить, что для выполнения этим устройством умножения в $GF(p^m)$ необходимые время и сложность оборудования прямо пропорциональны m.

Существование полей Галуа

Как было показано, поля Галуа существуют только, когда число элементов равно степени простого числа, и существует (с точностью до нумерации элементов) с любым заданным числом элементов только одно поле. В последующих трех теоремах доказывается, что для любого

p^n поле Галуа $GF(p^n)$ действительно существует. Собственно, все, что нужно доказать — это существование неприводимых многочленов всех степеней над $GF(p)$ для всех простых чисел $p > 1$.

Теорема 6.6.6. Пусть α и β — элементы поля $GF(p^n)$ с характеристикой p. Тогда при всех целых $m \geqslant 1$

$$(\alpha+\beta)^{p^m} = \alpha^{p^m} + \beta^{p^m}. \tag{6.6.8}$$

Доказательство. При $m = 1$ из разложения бинома следует, что

$$(\alpha+\beta)^p = \alpha^p + p\alpha^{p-1}\beta + \ldots + \binom{p}{i}\alpha^{p-i}\beta^i + \ldots + \beta^p, \tag{6.6.9}$$

где $\binom{p}{i}\alpha^{p-i}\beta^i$ нужно рассматривать как сумму $\binom{p}{i}$ слагаемых, каждое из которых равно $\alpha^{p-i}\beta^i$. Вместе с тем при $1 \leqslant i \leqslant p-1$

$$\binom{p}{i} = \frac{p(p-1)!}{i!\,(p-i)!}. \tag{6.6.10}$$

Так как $\binom{p}{i}$ — целое число, а p — простое число, то знаменатель (6.6.10) является делителем $(p-1)!$; поэтому p является делителем $\binom{p}{i}$. Следовательно, при сложении **1** с самой собой $\binom{p}{i}$ раз получим **0** и все внутренние слагаемые в правой части (6.6.9) равны **0**; таким образом,

$$(\alpha+\beta)^p = \alpha^p + \beta^p. \tag{6.6.11}$$

Возводя (6.6.11) в степень p, получаем

$$(\alpha+\beta)^{p^2} = (\alpha^p+\beta^p)^p = \alpha^{p^2} + \beta^{p^2};$$

аналогично, возводя $(m-1)$ раз соотношение (6.6.11) в степень p, получаем (6.6.8). |

Следствие. Если

$$f(D) = \sum_{i=0}^{l} f_i D^i$$

— многочлен над $GF(p)$ и α — элемент поля $GF(p^n)$, то при любом $m \geqslant 0$

$$f(\alpha^{p^m}) = [f(\alpha)]^{p^m}. \tag{6.6.12}$$

В частности, если α — корень $f(D)$, то при любом $m \geqslant 0$

$$f(\alpha^{p^m}) = \mathbf{0}. \tag{6.6.13}$$

Доказательство. Имеем

$$[f(\alpha)]^{p^m} = \left[f_l\alpha^l + \sum_{i=0}^{l-1} f_i\alpha^i\right]^{p^m} = (f_l\alpha^l)^{p^m} + \left(\sum_{i=0}^{l-1} f_i\alpha^i\right)^{p^m} =$$

$$= \sum_{i=0}^{l} (f_i \alpha^i)^{p^m} = \sum_{i=0}^{l} f_i^{p^m} \alpha^{ip^m}. \qquad (6.6.14)$$

Так как f_i — элемент $GF(p)$, то $f_i^{p-1} = 1$, поэтому $f_i^p = f_i$. Последовательно возводя обе части равенства в степень p, получаем $f_i^{p^m} = f_i$. Поэтому (6.6.14) преобразовывается к виду

$$[f(\alpha)]^{p^m} = \sum_{i=0}^{l} f_i \alpha^{ip^m} = f(\alpha^{p^m}). \qquad (6.6.15)$$

Если $f(\alpha) = 0$, то получаем (6.6.13). |

Теорема 6.6.7. Пусть $f(D)$ — нормированный неприводимый многочлен n-й степени над $GF(p)$. Многочлен $D^{p^m} - 1$ делится на $f(D)$ тогда и только тогда, когда m делится на n.

Доказательство. Рассмотрим поле $GF(p^n)$ многочленов над $GF(p)$ по модулю $f(t)$. Многочлен t является элементом рассматриваемого поля; обозначим этот элемент через α. Так как $f(\alpha) = 0$, то α — корень $f(D)$. Любой другой элемент в этом поле можно представить в виде

$$\beta = \mathbf{i}_0 + \mathbf{i}_1 \alpha + \ldots + \mathbf{i}_{n-1} \alpha^{n-1}, \qquad (6.6.16)$$

где $\mathbf{i}_0, \ldots, \mathbf{i}_{n-1}$ — целые элементы поля. Выберем $\mathbf{i}_0, \ldots, \mathbf{i}_{n-1}$ таким образом, чтобы β было примитивным элементом; обозначим через $B(D)$ соответствующий ему многочлен $B(D) = \mathbf{i}_0 + \mathbf{i}_1 D + \ldots + \mathbf{i}_{n-1} D^{n-1}$, так что $\beta = B(\alpha)$. Теперь предположим, что $D^{p^m} - 1$ делится на $f(D)$. Тогда α является корнем $D^{p^m-1} - 1$ и

$$\alpha^{p^m} - \alpha = 0. \qquad (6.6.17)$$

Используя (6.6.17) и (6.6.12), получаем

$$\beta = B(\alpha) = B(\alpha^{p^m}) = [B(\alpha)]^{p^m} = \beta^{p^m}. \qquad (6.6.18)$$

Из (6.6.18) следует, что $p^m - 1$ делится на порядок элемента β. Вместе с тем, так как β — примитивный элемент, его порядок равен $p^n - 1$, и число $p^n - 1$ должно быть делителем $p^m - 1$. Выполнив это деление, получим

$$p^m - 1 = (p^n - 1)(p^{m-n} + p^{m-2n} + \ldots), \qquad (6.6.19)$$

откуда следует, что $p^m - 1$ делится на $p^n - 1$ тогда и только тогда, когда m делится на n. Таким образом, если $D^{p^m-1} - 1$ делится на $f(D)$, то m должно делиться на n. Обратно, если m делится на n, то $p^m - 1$ делится на порядок элемента α, поскольку $p^n - 1$ делится на порядок элемента α. Следовательно, α — корень $D^{p^m-1} - 1$ и $D^{p^m-1} - 1$ делится на минимальный многочлен $f(D)$ элемента α. |

Из этой теоремы следует, что при любых $m \geqslant 1$ все неприводимые делители $D^{p^m-1} - 1$ над $GF(p)$ имеют степени, равные m или делителям m.

Теорема 6.6.8. Для любого положительного целого числа m и любого простого числа p существуют неприводимые многочлены в $GF(p)$ степени m и, следовательно, поля с p^m элементами.

Доказательство этой теоремы основывается на том факте, что не существует достаточного числа многочленов степени $m/2$ или меньше, составляющих все делители многочлена $D^{p^m-1} - 1$. Прежде чем использовать это соображение, нужно доказать следующую лемму.

Лемма. Многочлен $D^{p^m-1} - 1$, рассматриваемый как многочлен над $GF(p)$, не имеет кратных нормированных неприводимых делителей положительной степени.

Доказательство. Предположим, что $f(D)$ — нормированный неприводимый делитель n-й степени в разложении $D^{p^m-1} - 1$. Так как m делится на n, то $p^m - 1$ делится на $p^n - 1$ [см. (6.6.19)]. Поэтому

$$D^{p^m-1} - 1 = (D^{p^n-1} - 1)\, A(D), \tag{6.6.20}$$

где

$$A(D) = D^{(p^m-1)-(p^n-1)} + D^{(p^m-1)-2(p^n-1)} + \ldots + 1. \tag{6.6.21}$$

Существование $f(D)$ предполагает существование поля Галуа $GF(p^n)$ и поэтому $f(D)$ не может быть кратным делителем $(D^{p^n-1} - 1)$. Кроме того, $f(D)$ — минимальный многочлен некоторого элемента α из $GF(p^n)$, поэтому если $A(D)$ делится на $f(D)$, то α — корень $A(D)$. Вместе с тем, так как $p^m - 1$ делится на порядок элемента α, то $\alpha^{(p^m-1)-i(p^n-1)} - 1 = 0$.

Следовательно,

$$A(\alpha) = 1 + 1 + \ldots + 1, \tag{6.6.22}$$

где число слагаемых в правой части (6.6.22) равно

$$(p^m - 1)/(p^n - 1) = p^{m-n} + p^{m-2n} + \ldots + 1.$$

Так как $A(\alpha)$ равно остатку от деления этого числа слагаемых на p, то $A(\alpha) = 1$. Поэтому α не может быть корнем $A(D)$ и $A(D)$ не может делиться на $f(D)$, что завершает доказательство. |

Доказательство теоремы. Пусть a_i равно числу нормированных неприводимых делителей многочлена $D^{p^m-1} - 1$, степень которых равна m/i. Так как сумма степеней этих делителей равна $p^m - 1$, то получим

$$p^m - 1 = a_1 m + a_2 \frac{m}{2} + a_3 \frac{m}{3} + \ldots + a_m. \tag{6.6.23}$$

Так как все неприводимые делители степени m/i являются делителями многочлена $D^{p^{m/i}-1} - 1$, то

$$a_i \leqslant \frac{p^{m/i} - 1}{m/i}, \tag{6.6.24}$$

$$p^m - 1 \leqslant a_1 m + \sum_{\substack{i=2 \\ i : \frac{m}{i} = \text{целое}}}^{m} [p^{m/i} - 1]. \qquad (6.6.25)$$

Заменяя в (6.6.25) m/i на j и строя границу сверху с помощью суммирования по всем j, $1 \leqslant j \leqslant m/2$, получаем

$$p^m - 1 \leqslant a_1 m + \frac{p^{\lfloor m/2 \rfloor + 1} - p}{p-1} - \left\lfloor \frac{m}{2} \right\rfloor.$$

Ясно, что это неравенство удовлетворяется при $a_1 > 0$, что завершает доказательство. |

6.7. БЧХ-КОДЫ

Коды Боуза, Чоудхури и Хоквингема (БЧХ), открытые Хоквингемом (1959) и независимо от него Боузе и Чоудхури (1960), представляют собой класс циклических кодов, которые обладают весьма мощной способностью исправлять ошибки и одновременно допускают простые алгоритмы декодирования. Наиболее простым примером БЧХ-кодов являются двоичные коды, но совершенно аналогично можно в качестве алфавита выбирать элементы произвольного поля Галуа $GF(q)$. Порождающий многочлен для этих кодов определяется в терминах некоторого расширения $GF(q^m)$ поля $GF(q)$. Пусть α — элемент $GF(q^m)$, мультипликативный порядок которого равен N; числа $r \geqslant 0$ и d, $2 \leqslant d \leqslant N$ — произвольные целые числа; пусть далее $f_r(D)$, $f_{r+1}(D)$, ..., $f_{r+d-2}(D)$ — минимальные многочлены для α^r, α^{r+1}, ..., α^{r+d-2}. *Для каждого набора определенных выше параметров (т. е. для q, m, α, r и d) существует БЧХ-код; его порождающий многочлен определяется равенством*

$$g(D) = \text{НОК}\,[f_r(D),\ f_{r+1}(D),\ \ldots,\ f_{r+d-2}(D)], \qquad (6.7.1)$$

где НОК означает наименьшее общее кратное.

Длина блока в коде по определению равна N, мультипликативному порядку элемента α. Так как каждый из элементов α^r, ..., α^{r+d-2} является корнем $D^N - 1$, то многочлен $D^N - 1$ делится на каждый из многочленов $f_r(D)$, ..., $f_{r+d-2}(D)$ (см. теорему 6.6.3) и, следовательно, на $g(D)$, что необходимо для циклических кодов. В неинтересном случае, когда $g(D) = D^N - 1$, последовательность, целиком состоящая из нулей, по определению, выбирается в качестве единого кодового слова.

Другое определение БЧХ-кода с параметрами, указанными выше, состоит в том, что последовательность x_{N-1}, ..., x_0 является кодовым словом тогда и только тогда, когда α^r, α^{r+1}, ..., α^{r+d-2} являются корнями $x_{N-1}D^{N-1} + x_{N-2}D^{N-2} + \ldots + x_0$. Чтобы убедиться в этом, заметим, что в силу определения (6.7.1) любой многочлен, соответствующий кодовому слову, делится на $g(D)$ и, следовательно, на все минимальные многочлены $f_r(D)$, ..., $f_{r+d-2}(D)$. Поэтому α^r, α^{r+1}, ..., α^{r+d-2} являются корнями многочлена, соответствую-

щего кодовому слову. Обратно, если многочлен $x_{N-1}D^{N-1} + \ldots + x_0$ не является кодовым словом, то он не делится на $g(D)$ и, следовательно, не делится, по крайней мере, на один из минимальных многочленов, допустим на $f_i(D)$. Но тогда α^i не является корнем $x_{N-1}D^{N-1} + \ldots + x_0$.

В большинстве приложений в качестве α выбирается примитивный элемент $GF(q^m)$, так что $N = q^m - 1$. В качестве r обычно выбирается 1. Позднее будет показано, что параметр d имеет смысл нижней границы для минимального расстояния в коде.

Пример. Так как введенные выше определения, без сомнения, кажутся несколько абстрактными, рассмотрим конкретный пример $q = 2$, $m = 4$, $r = 1$, $d = 5$. Пусть используется представление $GF(2^4)$, определенное таблицей на рис. 6.6.3. Выберем в качестве элемента α тот элемент на рис. 6.6.3, который представляется многочленом t. Как показано в § 6.6, минимальный многочлен для α над $GF(2)$ равен $f_1(D) = D^4 + D + 1$. Тогда из (6.6.13) следует, что элементам α^2, α^4 и α^8 соответствует тот же самый минимальный многочлен, что и элементу α. Как показано в § 6.6, минимальный многочлен для α^3 над $GF(2)$ равен $f_3(D) = D^4 + D^3 + D^2 + D + 1$. Тогда

$$g(D) = f_1(D)f_3(D) = D^8 + D^7 + D^6 + D^4 + 1. \qquad (6.7.2)$$

Интересная особенность, которую вскрывает данный пример, состоит в том, что при $q = 2$ и при любых i имеем $f_{2i}(D) = f_i(D)$. Поэтому при $r = 1$, $q = 2$ и нечетном d (6.7.1) можно переписать в виде

$$g(D) = \text{НОК}\,[f_1(D), f_3(D), \ldots, f_{d-2}(D)]. \qquad (6.7.3)$$

Так как степень $g(D)$ равна числу проверочных символов в циклическом коде, порождаемом $g(D)$, то отсюда следует, что при нечетных d число проверочных символов БЧХ-кода с $q = 2$, $r = 1$ и произвольными α и m не превышает $m[(d-1)/2]$. Допустим на время, что d равно нижней границе минимального расстояния в коде, тогда получим, что если число проверочных символов равно em, то исправляются все комбинации e ошибок. Если выбрать α примитивным, то длина блока в коде будет равна $2^m - 1$. Из (6.7.1) следует, что при $r \neq 1$ или $q \neq 2$ все комбинации e ошибок могут быть исправлены, если число проверочных символов не меньше $2em$.

Так как БЧХ-коды являются циклическими кодами, то из результатов § 6.5 следует, что проверочную матрицу для БЧХ-кода можно вычислить, используя многочлен $h(D) = [D^N - 1]/g(D)$. Проверочную матрицу можно привести к более удобной для наших целей форме, если вспомнить, что $x(D)$ является многочленом, соответствующим кодовому слову, тогда и только тогда, когда $x(\alpha^i) = 0$ при $r \leqslant i \leqslant r + d - 2$. Последнее соотношение можно переписать в виде

$$\sum_{n=0}^{N-1} x_n \alpha^{in} = 0; \quad r \leqslant i \leqslant r + d - 2. \qquad (6.7.4)$$

Определяя проверочную матрицу как

$$H = \begin{bmatrix} \alpha^{(N-1)r} & \alpha^{(N-1)(r+1)} & \dots & \alpha^{(N-1)(r+d-2)} \\ \alpha^{(N-2)r} & \dots & \dots & \dots \\ \vdots & \vdots & & \\ \alpha^{r} & \alpha^{r+1} & \dots & \alpha^{r+d-2} \\ \mathbf{1} & \mathbf{1} & \dots & \mathbf{1} \end{bmatrix}, \qquad (6.7.5)$$

можно переписать (6.7.4) в виде равенства

$$\mathbf{x}H = 0, \qquad (6.7.6)$$

справедливого тогда и только тогда, когда $\mathbf{x} = (x_{N-1}, \dots, x_0)$ является кодовым словом. Если элементам α^i и α^j при некоторых значениях i и j соответствуют одинаковые минимальные многочлены, то $x(\alpha^i) = 0$ тогда и только тогда, когда $x(\alpha^j) = 0$; при этом столбцы, соответствующие α^j, могут быть опущены в (6.7.5). Таким образом, для только что рассмотренного примера

$$H^T = \begin{bmatrix} \alpha^{N-1} & \alpha^{N-2} & \dots & \alpha & 1 \\ \alpha^{3(N-1)} & \alpha^{3(N-2)} & \dots & \alpha^3 & 1 \end{bmatrix}. \qquad (6.7.7а)$$

Напомним, что согласно § 6.6 элементы расширения $GF(q^m)$ поля $GF(q)$ могут рассматриваться как m-компонентные векторы над $GF(q)$. Умножение элемента α из $GF(q^m)$ на элемент x, принадлежащий $GF(q)$, соответствует скалярному умножению вектора α на скаляр x из $GF(q)$. Поэтому, если рассматривать элементы H в (6.7.5) как вектор-строки в поле $GF(q)$, то при матричном умножении $\mathbf{x}H$ производятся лишь операции в $GF(q)$. Например, для только что рассмотренного примера матрица H^T, определяемая (6.7.7а), может быть переписана следующим образом (см. рис. 6.6.3):

$$H^T = \begin{bmatrix} 1&1&1&1&0&1&0&1&1&0&0&1&0&0&0 \\ 0&1&1&1&1&0&1&0&1&1&0&0&1&0&0 \\ 0&0&1&1&1&1&0&1&0&1&1&0&0&1&0 \\ 1&1&1&0&1&0&1&1&0&0&1&0&0&0&1 \\ \multicolumn{15}{c}{\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots} \\ 1&1&1&1&0&1&1&1&1&0&1&1&1&1&0 \\ 1&0&1&0&0&1&0&1&0&0&1&0&1&0&0 \\ 1&1&0&0&0&1&1&0&0&0&1&1&0&0&0 \\ 1&0&0&0&1&1&0&0&0&1&1&0&0&0&1 \end{bmatrix} \qquad (6.7.7б)$$

Теорема 6.7.1. Минимальное расстояние для БЧХ-кода с определенным выше параметром d не меньше чем d.

Доказательство. Пусть $\mathbf{x} = (x_{N-1}, \dots, x_0)$ — последовательность символов из $GF(q)$; предположим, что все значения этих символов, кроме $d-1$, выбираются равными **0**, а остальные символы, например

$x_{n_1}, x_{n_2}, \ldots, x_{n_{d-1}}$, выбираются произвольно. Покажем, что при любых значениях целых чисел n_j (удовлетворяющих $N-1 \geqslant n_1 > n_2 > \ldots > n_{d-1} \geqslant 0$), x может быть кодовым словом тогда и только тогда, когда $x_{n_j} = 0$ при $1 \leqslant j \leqslant d-1$. Этим будет доказано, что все ненулевые кодовые слова отличаются от последовательности, целиком состоящей из нулей, не менее чем в d позициях и, следовательно, минимальное расстояние в коде не меньше d.

При заданном выборе целых чисел $\{n_j\}$ вектор x является кодовым словом тогда и только тогда [см. (6.7.4)], когда

$$\sum_{j=1}^{d-1} x_{n_j} \alpha^{in_j} = 0; \quad r \leqslant i \leqslant r+d-2. \tag{6.7.8}$$

Упростим обозначения, определив

$$\begin{aligned} V_j &= x_{n_j}, \\ U_j &= \alpha^{n_j}, \end{aligned} \quad 1 \leqslant j \leqslant d-1. \tag{6.7.9}$$

Тогда соотношение (6.7.8) переписывается в виде

$$\begin{array}{lllll} V_1 U_1^r & + V_2 U_2^r + & \ldots & + V_{d-1} U_{d-1}^r & = 0, \\ V_1 U_1^{r+1} & + & \ldots & + V_{d-1} U_{d-1}^{r+1} & = 0, \\ \vdots & & & \vdots & \vdots \\ V_1 U_1^{r+d-2} & + & \ldots & + V_{d-1} U_{d-1}^{r+d-2} & = 0. \end{array} \tag{6.7.10}$$

Эти соотношения представляют собой систему $d-1$ уравнений [над $GF(q^m)$] с $d-1$ неизвестным $V_1, \ldots, V_{d-1}$. Одно решение данной системы тривиально: $V_1 = \ldots = V_{d-1} = 0$. Для завершения доказательства необходимо показать, что это решение единственно [из того факта, что в $GF(q^m)$ не существует другого решения, в действительности следует, что не существует другого решения в $GF(q)$]. Вместе с тем это решение единственно, если уравнения линейно независимы (доказательство этого такое же, как в поле действительных чисел). Линейная зависимость уравнений означает, что существует такое множество не равных одновременно нулю элементов поля $\beta_0, \beta_1, \ldots, \beta_{d-2}$, для которых

$$\sum_{j=0}^{d-2} \beta_j U_i^{r+j} = 0 \quad \text{для} \quad 1 \leqslant i \leqslant d-1. \tag{6.7.11}$$

Предположим, что такое множество элементов поля существует, и пусть $\beta(D) = \beta_0 + \beta_1 D + \ldots + \beta_{d-2} D^{d-2}$. Соотношение (6.7.11) выражается через $\beta(D)$ следующим образом:

$$U_i^r \beta(U_i) = 0; \quad 1 \leqslant i \leqslant d-1. \tag{6.7.12}$$

Вместе с тем $\beta(D)$ является ненулевым многочленом, степень которого не больше $d-2$, а из (6.7.12) следует, что число корней $\beta(D)$ не мень-

ше $d-1$. Эти утверждения противоречивы и, следовательно, уравнения линейно не зависимы, что завершает доказательство. |

Так как минимальное расстояние в БЧХ-коде не меньше d, то, как известно, любая комбинация из $\lfloor (d-1)/2 \rfloor$ или менее ошибок может быть исправлена; ниже получим алгоритм исправления всех таких комбинаций ошибок. В действительности этот код способен исправлять также много комбинаций более чем $\lfloor (d-1)/2 \rfloor$ ошибок, однако до сих пор неизвестен простой алгоритм их исправления.

Обозначим через $x(D) = x_{N-1}D^{N-1} + \ldots + x_0$ многочлен, соответствующий переданному кодовому слову, через $y(D) = y_{N-1}D^{N-1} + \ldots + y_0$ — многочлен, соответствующий принятой последовательности, а через $z(D) = y(D) - x(D)$ — многочлен, соответствующий шумовой последовательности. Синдромом $\mathbf{S}$ назовем вектор с компонентами

$$S_i = y(\alpha^{r+i}), \quad 0 \leqslant i \leqslant d-2. \tag{6.7.13}$$

Отметим, что каждая из компонент S_i является элементом $GF(q^m)$; введя матрицу H согласно (6.7.5), получим $\mathbf{S} = \mathbf{y}H$. Так как α^{r+i} при $0 \leqslant i \leqslant d-2$ является корнем многочлена $x(D)$, то

$$S_i = x(\alpha^{r+i}) + z(\alpha^{r+i}) = z(\alpha^{r+i}); \quad 0 \leqslant i \leqslant d-2. \tag{6.7.14}$$

Теперь предположим, что при передаче произошло некоторое фиксированнсе число $e \leqslant \lfloor (d-1)/2 \rfloor$ ошибок, например на n_1-й, n_2-й, ..., n_e-й позициях. Тогда $z_n = 0$ при всех n, кроме $n_1, n_2, \ldots, n_e$, и (6.7.14) принимает вид

$$S_i = \sum_{j=1}^{e} z_{n_j} \alpha^{n_j (r+i)}; \quad 0 \leqslant i \leqslant d-2. \tag{6.7.15}$$

Чтобы упростить обозначения, введем понятия *значений ошибок* V_j и *локаторов ошибок* U_j, определяемых соотношениями:

$$\begin{aligned} V_j &= z_{n_j}, \\ U_j &= \alpha^{n_j}, \end{aligned} \qquad 1 \leqslant j \leqslant e \tag{6.7.16}$$

Соотношение (6.7.15) принимает вид

$$S_i = \sum_{j=1}^{e} V_j U_j^{r+i}; \quad 0 \leqslant i \leqslant d-2. \tag{6.7.17}$$

Таким образом, декодер может вычислить синдром $\mathbf{S}$ по принятой последовательности $\mathbf{y}$. Если декодер может решить (6.7.17) и определить значения ошибок и локаторы ошибок, то он может из (6.7.16) найти шумовую последовательность $\mathbf{z}$. (Напомним, что так как N равно мультипликативному порядку α, то все элементы $\alpha^0, \alpha^1, \alpha^2, \ldots, \alpha^{N-1}$ различны.) Перейдем теперь к главной задаче, которая согласно принятой здесь схеме состоит в нахождении решения (6.7.17).

Прежде всего определим многочлен*) бесконечной степени $S_\infty(D)$:

*) Обычно такие выражения не называют многочленами, но здесь это удобно.

$$S_\infty(D) = S_0 + S_1 D + S_2 D^2 + \ldots, \tag{6.7.18}$$

где S_i при всех $i \geqslant 0$ задается соотношением

$$S_i = \sum_{j=1}^{e} V_j U_j^{r+i}; \quad i \geqslant 0. \tag{6.7.19}$$

Напомним, что при $0 \leqslant i \leqslant d-2$ можно вычислять S_i непосредственно по $\mathbf{y}$; при $i > d-2$ величина S_i неизвестна, но по крайней мере определена для заданной шумовой последовательности $\mathbf{z}$. Можно переписать $S_\infty(D)$ в виде

$$S_\infty(D) = \sum_{i=0}^{\infty} \sum_{j=1}^{e} V_j U_j^{r+i} D^i =$$

$$= \sum_{j=1}^{e} V_j U_j^r \sum_{i=0}^{\infty} U_j^i D^i = \sum_{j=1}^{e} V_j U_j^r \frac{1}{1-U_j D}, \tag{6.7.20}$$

где через $1/(1-U_jD)$ обозначен многочлен бесконечной степени $1 + U_jD + U^2{}_jD^2 + \ldots$. Теперь введем многочлен $\sigma(D) = \sigma_0 + \sigma_1 D + \ldots + \sigma_e D^e$ следующим образом:

$$\sigma(D) = \prod_{j=1}^{e} (1 - U_j D). \tag{6.7.21}$$

Вычисляя произведение правых частей (6.7.20) и (6.7.21), получаем*)

$$\sigma(D) S_\infty(D) = \sum_{j=1}^{e} V_j U_j^r \prod_{\substack{l=1 \\ l \neq j}}^{e} (1 - U_l D) \stackrel{\Delta}{=} A(D). \tag{6.7.22}$$

Чтобы интерпретировать это равенство, полезно ввести некоторые дополнительные обозначения. Для произвольного многочлена $B(D)$ (конечного или бесконечного) пусть $[B(D)]_i^j$ определяется соотношением

$$[B(D)]_i^j = \begin{cases} \sum_{l=i}^{j} B_l D^l, & j \geqslant i, \\ 0, & j < i. \end{cases} \tag{6.7.23}$$

Если j превышает степень $B(D)$, равную L, условимся полагать $B_{L+1} = B_{L+2} = \ldots = B_j = 0$. Обозначим далее $S(D) = S_0 + S_1 D + \ldots + S_{d-2} D^{d-2}$.

---

*) Читатель, не привыкший к таким формальным преобразованиям, может непосредственно убедиться, что

$$[\prod_{l=1}^{e} (1 - U_l D)][1 + U_j D + U_j^2 D^2 + \ldots] = \prod_{l \neq j} (1 - U_l D),$$

и равенство выполняется для всех степеней D.

В введенных выше обозначениях

$$S(D) = [S_\infty(D)]_0^{d-2}. \tag{6.7.24}$$

Так как члены в $S_\infty(D)$, степень которых больше $d-2$, влияют лишь на те члены многочлена $S_\infty(D)\sigma(D)$, степень которых больше чем $d-2$, то из (6.7.22) получим

$$[\sigma(D)\,S(D)]_0^{d-2} = [A(D)]_0^{d-2}. \tag{6.7.25}$$

Наконец, заметим, что согласно равенству (6.7.22), определяющему $A(D)$, степень $A(D)$ не может превышать $e-1$. Поэтому скобки вокруг $A(D)$ в (6.7.25) могут быть опущены и, более того,

$$[\sigma(D)\,S(D)]_e^{d-2} = 0. \tag{6.7.26}$$

Из соотношения (6.7.26) следует, что коэффициент при D^l в произведении $\sigma(D)S(D)$ равен **0** для $e \leqslant l \leqslant d-2$. При более подробной записи (6.7.26) эквивалентно следующей системе равенств:

$$\begin{aligned}
&\sigma_0 S_e + \sigma_1 S_{e-1} + \ldots + \sigma_e S_0 = 0,\\
&\sigma_0 S_{e+1} + \sigma_1 S_e + \ldots + \sigma_e S_1 = 0,\\
&\quad\vdots \qquad\qquad\qquad\quad \vdots\\
&\sigma_0 S_{d-2} + \sigma_1 S_{d-3} + \ldots + \sigma_e S_{d-2-e} = 0.
\end{aligned} \tag{6.7.27}$$

Равенства (6.7.27) дают систему $d-1-e$ линейных уравнений, по которым декодер может найти e неизвестных $\sigma_1, \ldots, \sigma_e$ [согласно (6.7.21), $\sigma_0 = 1$]. Если эти уравнения разрешимы, то по $\sigma(D)$ могут быть найдены локаторы ошибок $U_1, \ldots, U_e$, поскольку $U_1^{-1}, \ldots, U_e^{-1}$ являются корнями $\sigma(D)$. Теперь общая картина процедуры декодирования ясна и можно сформулировать ее для последующих ссылок в виде четырех этапов.

Этап 1. Вычислить $S_0, \ldots, S_{d-2}$ по принятой последовательности $\mathbf{y}$.

Этап 2. Найти $\sigma(D)$ из (6.7.26) или (6.7.27).

Этап 3. Найти корни $\sigma(D)$ и, следовательно, локаторы ошибок.

Этап 4. Найти значения ошибок $V_1, \ldots, V_e$. Отметим, что в случае двоичных БЧХ-кодов все значения ошибок равны 1 (так как, по определению, они ненулевые) и, следовательно, этап 4 не нужен.

Теперь обсудим более подробно этап 2, а к рассмотрению реализации остальных этапов вернемся несколько позднее. Отметим, что при отыскании $\sigma(D)$ из (6.7.26) декодеру неизвестно число ошибок e. Следующая теорема утверждает, что $\sigma(D)$ может быть однозначно определен без предварительного знания e.

Теорема 6.7.2. Предположим, что произошло $e \leqslant \lfloor (d-1)/2 \rfloor$ ошибок и что $\sigma(D)$ определяется равенством (6.7.21). Пусть $\hat{e}$ — минимальное целое число, для которого существует многочлен $\hat{\sigma}(D)$ с $\hat{\sigma}_0 = 1$, степень которого не превышает $\hat{e}$ и который удовлетворяет соотношению $[\hat{\sigma}(D)S(D)]_{\hat{e}}^{d-2} = 0$. Тогда $e = \hat{e}$ и $\sigma(D) = \hat{\sigma}(D)$.

Доказательство. Соотношение $[\hat{\sigma}(D)\, S(D)]_{\hat{e}}^{d-2} = 0$ можно переписать в виде

$$\sum_{l=0}^{\hat{e}} \hat{\sigma}_l S_{i-l} = 0; \quad \hat{e} \leqslant i \leqslant d-2. \tag{6.7.28}$$

По определению S_i, это означает

$$\sum_{l=0}^{\hat{e}} \hat{\sigma}_l S_{i-l} = \sum_{l=0}^{\hat{e}} \hat{\sigma}_l \sum_{j=1}^{e} V_j U_j^{r+i-l} =$$

$$= \sum_{j=1}^{e} V_j U_j^{r+i} \sum_{l=0}^{\hat{e}} \hat{\sigma}_l U_j^{-l} = \sum_{j=1}^{e} V_j \hat{\sigma}(U_j^{-1}) U_j^{r+i} = 0;$$

$$\hat{e} \leqslant i \leqslant d-2. \tag{6.7.29}$$

Равенство (6.7.29) можно рассматривать как систему $d-1-\hat{e}$ линейных уравнений относительное e неизвестных $V_j\hat{\sigma}(U_j^{-1})$, где $1 \leqslant j \leqslant e$. Так как $\hat{e}$ равно минимальному целому числу, для которого выполняется (6.7.28), и так как $[\sigma(D)S(D)]_e^{d-2} = 0$, то должно выполняться $\hat{e} \leqslant e$. Аналогично, так как $e \leqslant \lfloor (d-1)/2 \rfloor$, то справедливо неравенство $e \leqslant d-1-\hat{e}$. Теперь рассмотрим лишь e первых уравнений в (6.7.29) при $\hat{e} \leqslant i \leqslant \hat{e}+e-1$. Эти e уравнений с e неизвестными $V_j\hat{\sigma}(U_j^{-1})$ являются линейно независимыми по тем же причинам, которые были указаны при установлении линейной независимости (6.7.10) в теореме 6.7.1. Поэтому единственное решение этих уравнений:

$$V_j \hat{\sigma}(U_j^{-1}) = 0; \quad 1 \leqslant j \leqslant e. \tag{6.7.30}$$

Поскольку $V_j \neq 0$ при $1 \leqslant j \leqslant e$, величины U_j^{-1} являются корнями многочлена $\hat{\sigma}(D)$ при $1 \leqslant j \leqslant e$. Так как, во-первых, степень $\hat{\sigma}(D)$ не превышает e, во-вторых, $\hat{\sigma}(D)$ имеет те же e корней, что и $\sigma(D)$, и, в-третьих, $\hat{\sigma}_0 = \sigma_0$, то выполняется равенство $\hat{\sigma}(D) = \sigma(D)$. Поскольку степень $\hat{\sigma}(D)$ равна e, мы получим $\hat{e} = e$, что завершает доказательство. |

Теперь опишем весьма простой итеративный алгоритм нахождения $\sigma(D)$.

Итеративный алгоритм*) для нахождения $\sigma(D)$

Согласно предыдущей теореме, если число ошибок не превышает $\lfloor (d-1)/2 \rfloor$, то определим $\sigma(D)$ через синдромные многочлены $S(D)$, разрешив уравнение $[\sigma(D)S(D)]_e^{d-2} = 0$ для минимального e и для

*) Этот алгоритм принадлежит Берлекэмпу (1967). Мы опишем здесь модификацию алгоритма Берлекэмпа, принадлежащую Месси (1968), и будем ближе придерживаться трактовки Месси.

многочлена $\sigma(D)$, $\sigma_0 = 1$, степени, не превышающей e. Эту задачу легче всего описать в терминах регистра сдвига с линейной обратной связью (РСЛОС), представленного на рис. 6.7.1.

Первоначально в регистре хранится последовательность элементов $S_0, S_1, \ldots, S_{l-1}$ из данного поля. Затем регистр вычисляет новый элемент S_l через величины обратных связей $-C_1, \ldots, -C_l$ согласно соотношению

$$S_l = -S_{l-1}C_1 - S_{l-2}C_2 - \ldots - S_0 C_l. \qquad (6.7.31)$$

Числа $C_1, \ldots, C_l$ являются элементами того же поля, что и S_i. Затем регистр сдвигается на одну позицию вправо, после чего слева в регистр

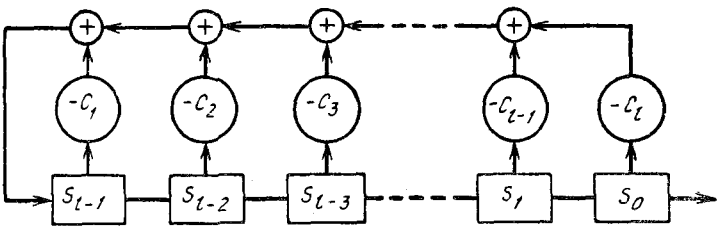


Рис. 6.7.1. Регистр сдвига с линейной обратной связью (РСЛОС).

вводится S_l. При каждом из последовательных сдвигов вправо вычисляется новый элемент, определяемый соотношением

$$S_i = -S_{i-1}C_1 - S_{i-2}C_2 - \ldots - S_{i-l}C_l; \; i \geqslant l. \qquad (6.7.32)$$

Назовем *длиной* РСЛОС число разрядов в регистре сдвига (l в обозначениях рис. 6.7.1). Назовем также *многочленом связей* РСЛОС многочлен $C(D) = 1 + C_1 D + C_2 D^2 + \ldots + C_l D^l$, где $C_1, C_2, \ldots, C_l$ — показанные на рис. 6.7.1 величины обратных связей, взятые со знаком минус. Поскольку РСЛОС полностью (если не считать первоначально хранящуюся в нем последовательность элементов) описывается заданием длины регистра и многочленом связей, то для обозначения РСЛОС с длиной регистра l и многочленом связей $C(D)$ будем использовать название $[C(D), l]$-регистр. Некоторые или все величины обратных связей $-C_1, -C_2, \ldots, -C_l$ могут равняться нулю; поэтому $C(D)$ может быть произвольным многочленом, степень которого не превышает l, причем $C_0 = 1$. Наконец, для любого многочлена $S(D)$ (конечного или бесконечного) будем говорить, что $[C(D), l]$-регистр *генерирует* $[S(D)]_0^L$ тогда и только тогда, когда регистр, первоначально хранивший $S_0, \ldots, S_{l-1}$, порождает оставшиеся элементы (если они есть) $S_l, \ldots, S_L$ методом, определяемым соотношением (6.7.32). Поскольку (6.7.32) эквивалентно утверждению, что коэффициент при D^i в произведении $C(D)S(D)$ равен 0, то $[C(D), l]$-регистр генерирует $[S(D)]_0^L$ тогда и только тогда, когда

$$[C(D)\,S(D)]_l^L = 0. \qquad (6.7.33)$$

Теперь вырисовывается связь между устройством регистра сдвига с линейной обратной связью и задачей нахождения $\sigma(D)$. Задача состоит в отыскании такого $[\sigma(D), e]$-регистра с минимальной длиной e, который генерировал бы $[S(D)]_0^{d-2}$. Заметим, что определение $[\sigma(D), e]$-регистра включает в себя ограничение степени $\sigma(D)$ максимальной величиной e и условие $\sigma_0 = 1$. Ниже дан алгоритм нахождения такого самого короткого регистра. Будет показано, что этот алгоритм работает в случае произвольного многочлена $S(D)$ над произвольным полем. Алгоритм состоит в нахождении последовательности регистров, первый из которых является самым коротким регистром, генерирующим S_0, второй — самым коротким регистром, генерирующим $S_0 + S_1 D$, и т. д. Регистр, воспроизводящий алгоритм генерирования $[S(D)]_0^{n-1}$, назовем $[C_n(D), l_n]$-регистром. Грубо говоря, алгоритм работает следующим образом: при заданном $[C_n(D), l_n]$-регистре, генерирующем $[S(D)]_0^{n-1}$, алгоритм проверяет, генерирует ли $[C_n(D), l_n]$-регистр также $[S(D)]_0^n$, т. е. выполняется ли $[C_n(D)S(D)]_{l_n}^n = 0$.

Так как по предположению $[C_n(D)S(D)]_{l_n}^{n-1} = 0$, то вопрос состоит в том, равен ли нулю коэффициент при D^n у многочлена $C_n(D)S(D)$. Эта сумма называется *следующей разностью*, d_n, алгоритма; выражая $C_n(D)$ через $1 + C_{n,1}D + C_{n,2}D^2 + \ldots$, имеем

$$d_n = S_n + \sum_{i=1}^{n} C_{n,i} S_{n-i}. \tag{6.7.34}$$

Обозначим через S'_n коэффициент при D^n, генерируемый регистром согласно (6.7.32); тогда получим $d_n = S_n - S'_n$, так что d_n равно разности между благоприятным значением следующего выходного символа S_n и действительным выходным символом регистра S'_n. Если $d_n = 0$, алгоритм увеличивает n на 1, сохраняя неизменным регистр. Если $d_n \neq 0$, к многочлену связи добавляется поправочный член, так, чтобы S_n генерировалось правильно.

Более подробное описание алгоритма состоит в следующем: при любом n параметры $[C_{n+1}(D), l_{n+1}]$-регистра определяются через параметры $[C_n(D), l_n]$-регистра и параметры априорного регистра в последовательности $[C_{k_n}, l_{k_n}]$, где $k_n < n$. При любом $n > 0$ целое число k_n определяется следующим рекуррентным соотношением:

$$k_n = \begin{cases} k_{n-1}, & \text{если } l_n = l_{n-1}, \\ n-1, & \text{если } l_n > l_{n-1}. \end{cases} \tag{6.7.35}$$

Многочлен $C_{n+1}(D)$ и l_{n+1} равны:

$$C_{n+1}(D) = C_n(D) - \frac{d_n}{d_{k_n}} D^{n-k_n} C_{k_n}(D), \tag{6.7.36}$$

$$l_{n+1} = \begin{cases} l_n; & d_n = 0, \\ \max\left[l_n,\ n-(k_n - l_{k_n})\right]; & d_n \neq 0, \end{cases} \tag{6.7.37}$$

где d_n и d_{k_n} определяются (6.7.34) или, точнее,

$$d_{k_n} = S_{k_n} + \sum_{i=1}^{k_n} C_{k_n,\,i}\, S_{k_n - i}.$$

Алгоритм начинает работу с $n = 0$ при начальных условиях

$$C_0(D) = C_{-1}(D) = 1,\quad l_0 = l_{-1} = 0,\quad k_0 = -1,\quad d_{-1} = 1.$$

Следующая ниже теорема утверждает, что $C_n(D)$ и l_n определяют $[C_n(D),\ l_n]$-регистр и что этот регистр генерирует $[S(D)]_0^{n-1}$. В последующей теореме будет показано, что $[C_n(D),\ l_n]$-регистр является самым коротким регистром, генерирующим $[S(D)]_0^{n-1}$.

Теорема 6.7.3. При любом $n \geqslant 0$:

(а) $k_n < n$, (6.7.38)

(б) $C_{n,0} = 1$, где $C_n(D) = C_{n,0} + C_{n,1} D + \ldots$, (6.7.39)

(в) степень $[C_n(D)] \leqslant l_n$, (6.7.40)

(г) $[C_n(D)\, S(D)]_{l_n}^{n-1} = 0$. (6.7.41)

Доказательство.

Утверждение а. Согласно начальным условиям при $n = 0$ имеем $k_0 < 0$. При $n > 0$ доказательство немедленно следует из (6.7.35) с помощью индукции по n.

Утверждение б. Согласно начальным условиям $C_{0,0} = 1$. Теперь предположим, что $C_{n,0} = 1$ при любом заданном n. Так как $n - k_n > 0$, то из (6.7.36) следует, что $C_{n+1,0} = 1$. Поэтому по индукции получим, что $C_{n,0} = 1$ при всех $n \geqslant 0$.

Утверждения в и г. Вновь воспользуемся индукцией по n. Согласно начальным условиям соотношения (6.7.40) и (6.7.41) выполняются при $n = -1, 0$. Пусть задано некоторое n; предположим, что при $-1 \leqslant i \leqslant n$

$$\text{степень}\quad [C_i(D)] \leqslant l_i, \tag{6.7.42}$$

$$[C_i(D)\, S(D)]_{l_i}^{i-1} = 0. \tag{6.7.43}$$

Доказательство будет завершено, если показать, что из этих соотношений следует справедливость (6.7.42) и (6.7.43) для $i = n + 1$. Рассмотрим отдельно случаи $d_n = 0$ и $d_n \neq 0$. При $d_n = 0$ имеем $C_{n+1}(D) = C_n(D)$ и $l_{n+1} = l_n$. Поэтому справедливость (6.7.42) при $i = n$ влечет за собой справедливость (6.7.42) при $i = n + 1$. Аналогично из (6.7.43) для $i = n$ следует, что

$$[C_{n+1}(D)\, S(D)]_{l_{n+1}}^{n-1} = 0.$$

Из (6.7.34) получим $[C_{n+1}(D)\, S(D)]_{l_n}^{n} = d_n = 0$. Поэтому $[C_{n+1}(D) \times S(D)]_{l_{n+1}}^{n} = 0$, откуда вытекает справедливость (6.7.43) для $i = n + 1$. Теперь предположим, что $d_n \neq 0$. Согласно (6.7.36) имеем:

$$\text{степень } [C_{n+1}(D)] \leqslant \max\{\text{степень } [C_n(D)],\ n-k_n+$$
$$+\text{степень } [C_{k_n}(D)]\} \leqslant \max\{l_n,\ n-k_n+l_{k_n}\} = l_{n+1},$$

где было использовано (6.7.42) при $i=n$ и $i=k_n$ и затем использовано (6.7.37).

Наконец, из (6.7.36) получим

$$[C_{n+1}(D)\,S(D)]_{l_{n+1}}^{n} = [C_n(D)\,S(D)]_{l_{n+1}}^{n} - \left[\frac{d_n}{d_{k_n}} D^{n-k_n} C_{k_n}(D)\,S(D)\right]_{l_{n+1}}^{n}. \tag{6.7.44}$$

n	S_n	l_n	$C_n(D)$	РСЛОС	d_n	k_n	l_{k_n}	$C_{k_n}(D)$	d_{k_n}
0	1	0	1		1	-1	0	1	1
1	1	1	$1+D$		0	0	0	1	1
2	1	1	$1+D$		0	0	0	1	1
3	0	1	$1+D$		1	0	0	1	1
4	1	3	$1+D+D^3$		0	3	1	$1+D$	1
5	1	3	$1+D+D^3$		1	3	1	$1+D$	1
6	0	3	$1+D+D^2$		0	3	1	$1+D$	1
7	1	3	$1+D+D^2$		0	3	1	$1+D$	1
8		3	$1+D+D^2$						

Рис. 6.7.2. Действия алгоритма в $GF(2)$ для $S(D)=1+D+D^2+D^4+D^5+D^7$ до $n=8$.

Так как $l_{n+1} \geqslant l_n$, то имеем

$$[C_n(D)\,S(D)]_{l_{n+1}}^{n} = d_n D^n. \tag{6.7.45}$$

Что касается последнего выражения в (6.7.44), то нетрудно убедиться, что D^{n-k_n} можно вынести за скобки, если уменьшить одновременно пределы на $n-k_n$. Таким образом,

$$\left[\frac{d_n}{d_{k_n}} D^{n-k_n} C_{k_n}(D)\,S(D)\right]_{l_{n+1}}^{n} - = \frac{d_n}{d_{k_n}} D^{n-k_n} [C_{k_n}(D)\,S(D)]_{(l_{n+1}-n+k_n)}^{k_n} = = \frac{d_n}{d_{k_n}} D^{n-k_n} d_{k_n} D^{k_n}, \tag{6.7.46}$$

где для доказательства того, что $l_{n+1} - n + k_n \geqslant l_{k_n}$, было использовано (6.7.37). Подставляя (6.7.45) и (6.7.46) в (6.7.44), получаем $[C_{n+1}(D)\,S(D)]_{l_{n+1}}^{n} = 0$, что завершает доказательство.

На рис. 6.7.2 представлен пример работы описанного алгоритма для простейшего случая многочленов над $GF(2)$. На рис. 6.7.3 изображены графики зависимости l_n и $n - l_n$ от n для этого примера. Имеются некоторые важные моменты во взаимосвязи между l_n и $n - l_n$, справедливые в более общих случаях. Заметим, во-первых, что l_n не убывает с ростом n; согласно (6.7.37) это выполняется в общем случае. Следующие три результата являются более тонкими.

Л е м м а 1. При любом $n \geqslant 0$

$$k_n - l_{k_n} > i - l_i; \quad -1 \leqslant i < k_n, \tag{6.7.47}$$

$$l_n = k_n - l_{k_n} + 1, \tag{6.7.48}$$

$l_{n+1} > l_n$ тогда и только тогда, когда $l_n \leqslant \frac{1}{2}$ и $d_n \neq 0$. (6.7.49)

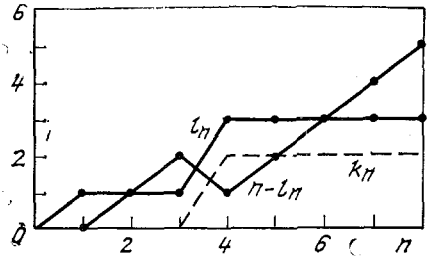


Рис. 6.7.3. Величины l_n и $n - l_n$ как функции n.

Доказательство. Сначала покажем, что из (6.7.48) следует (6.7.49). Согласно (6.7.37) $l_{n+1} > l_n$ тогда и только тогда, когда одновременно выполняется $d_n \neq 0$ и

$$n - (k_n - l_{k_n}) > l_n. \tag{6.7.50}$$

Согласно (6.7.48) соотношение (6.7.50) эквивалентно $n > 2l_n - 1$ или $l_n \leqslant n/2$, что доказывает (6.7.49). Совершенно очевидна справедливость (6.7.47) и (6.7.48) при $n = 0$; предположим, что они справедливы при некотором заданном n. Утверждение можно доказать по индукции, если показать справедливость этих соотношений при $n + 1$. Если $l_{n+1} = l_n$, то $k_{n+1} = k_n$, и (6.7.47), и (6.7.48) выполняются при $n + 1$. Если $l_{n+1} > l_n$, то из (6.7.35) следует, что $k_{n+1} = n$ и

$$k_{n+1} - l_{k_{n+1}} = n - l_n. \tag{6.7.51}$$

Величина k_n показывает, когда в последний раз до момента n произошло изменение длины регистра; поэтому $l_i = l_n$ при $k_n < i < n$ и, следовательно,

$$k_{n+1} - l_{k_{n+1}} > i - l_i; \quad k_n < i < n. \tag{6.7.52}$$

Кроме того, из (6.7.37) следует $n - l_n > k_n - l_{k_n}$; поэтому на основании (6.7.51) и (6.7.47) получим

$$k_{n+1} - l_{k_{n+1}} > i - l_i; \quad -1 \leqslant i \leqslant k_n, \tag{6.7.53}$$

что подтверждает справедливость (6.7.47) при $n + 1$.

Предполагая, как и ранее, что $l_{n+1} > l_n$, можно, используя соотношение (6.7.50), справедливое при n, и соотношение (6.7.51), получить

$$k_{n+1} - l_{k_{n+1}} = l_{n+1} + 1. \tag{6.7.54}$$

Последнее доказывает справедливость (6.7.48) при $n+1$, что завершает доказательство. |

Из этой леммы следует, что $n - l_n$ как функция n имеет вид возрастающей последовательности пиков; величина k_n для каждого n указывает на местонахождение предшествующего пика, который превышает любой из предыдущих пиков (мы считаем, что в точке n имеется пик, если $n - l_n \geqslant (n+1) - l_{n+1}$).

Прежде чем доказывать, что описанный алгоритм при любом n приводит к самому короткому возможному регистру, необходимо привести две леммы.

Л е м м а 2. Предположим, что $[A(D), l]$ и $[B(D), l]$ — это два регистра, удовлетворяющие соотношениям

$$[A(D)S(D)]_l^n = aD^n; \quad a \neq 0, \tag{6.7.55}$$

$$[B(D)S(D)]_l^n = 0; \tag{6.7.56}$$

тогда при некотором j, $0 \leqslant j \leqslant l$, существует $[F(D), l-j]$-регистр, удовлетворяющий соотношению

$$[F(D)S(D)]_{l-j}^{n-j} = fD^{n-j}; \quad f \neq 0. \tag{6.7.57}$$

Доказательство. Имеем

$$\{[A(D) - B(D)]S(D)\}_l^n = aD^n. \tag{6.7.58}$$

Пусть j — минимальное целое число, для которого $A_j \neq B_j$ и пусть $\gamma = A_j - B_j$. Пусть $F(D)$ определяется соотношением

$$A(D) - B(D) = \gamma D^j F(D). \tag{6.7.59}$$

Тогда $F_0 = \mathbf{1}$ и

$$\text{степень } F(D) < \\ < \min\ [\text{степень } A(D),\ \text{степень } B(D)] - j \leqslant l - j.$$

Поэтому $[F(D), l-j]$ является регистром. Подставляя (6.7.59) в (6.7.58) и замечая, что D^j можно вынести за скобки, если одновременно уменьшить пределы на j, получим

$$\gamma\,[F(D)S(D)]_{l-j}^{n-j} = a\ D^{n-j}.$$

Поскольку $a/\gamma \neq \mathbf{0}$, это завершает доказательство. |

Л е м м а 3. Предположим, что при заданных $S(D)$ и n, $[C_i(D), l_i]$-регистр является самым коротким регистром, генерирующим $[S(D)]_0^{i-1}$ при всех $i \leqslant n$. Тогда не существует такого $[A(D), l_A]$-регистра, для которого при некотором $n_A < n$ одновременно выполняются соотношения

$$n_A - l_A > k_n - l_{k_n} \tag{6.7.60}$$

и

$$[A(D)S(D)]_{l_A}^{n_A} = aD^{n_A}; \quad a \neq 0. \tag{6.7.61}$$

Доказательство. Покажем, что предположение о несправедливости леммы приводит к противоречию. Пусть $[A(D), l_A]$ — самый короткий регистр, для которого при $n_A < n$ выполняются (6.7.60) и (6.7.61).

Случай а. Допустим, что $n_A > n_k$. Известно, что $l_i = l_n$ при $n_k < i < n$ и потому $[C_n(D), l_n]$ является самым коротким регистром, генерирующим $[S(D)]_0^{i-1}$ при $n_k < i < n$. При $i = n_A$ отсюда следует, что $l_n \leqslant l_A$. Поэтому, поскольку $n_A < n$, то для $[C_n(D), l_A]$-регистра выполняется равенство

$$[C_n(D)\,S(D)]_{l_A}^{n_A} = 0. \tag{6.7.62}$$

Согласно предыдущей лемме из (6.7.61) и (6.7.62) следует существование $[F(D), l_A - j]$-регистра, для которого при некотором $j > 0$ выполняется равенство

$$[F(D)\,S(D)]_{l_A-j}^{n_A-j} = fD^{n_A-j};\ f \neq 0.$$

Этот регистр короче, чем $[A(D), l_A]$-регистр и удовлетворяет (6.7.60) и (6.7.61). Это противоречит сделанному предположению.

Случай б. Допустим, что $n_A \leqslant n_k$. По предложению, $[C_{n_A}(D), l_{n_A}]$-регистр является кратчайшим регистром, генерирующим $[S(D)]_0^{n_A-1}$ и поэтому $l_{n_A} \leqslant l_A$. Следовательно, в силу (6.7.47),

$$k_n - l_{kn} \geqslant n_A - l_{n_A} \geqslant n_A - l_A,$$

что противоречит (6.7.60). |

Теорема 6.7.4. При любом $S(D)$ и всех $n \geqslant 0$ не существует регистра, который генерирует $[S(D)]_0^{n-1}$ и имеет длину, меньшую, чем $[C_n(D), l_n]$-регистр, построенный с помощью описанного выше алгоритма.

Доказательство. Проведем индукцию по n. Очевидно, что теорема справедлива при $n = 0$. Допустим, что она выполняется для любых заданных $S(D)$ и некоторого данного n. Если $l_{n+1} = l_n$, то ясно, что $[C_{n+1}(D), l_{n+1}]$-регистр имеет минимальную длину среди регистров, генерирующих $[S(D)]_0^n$, поскольку он является самым коротким регистром, генерирующим $[S(D)]_0^{n-1}$, а любой регистр, генерирующий $[S(D)]_0^n$, генерирует также и $[S(D)]_0^{n-1}$. Теперь предположим, что $l_{n+1} > l_n$, так что

$$[C_n(D)\,S(D)]_{l_n}^{n} = d_n D^n;\ d_n \neq 0.$$

Рассмотрим какой-либо $[B(D), l_B]$-регистр, генерирующий $[S(D)]_0^n$. Тогда имеем $l_B \geqslant l_n$ и согласно лемме 2 из существования

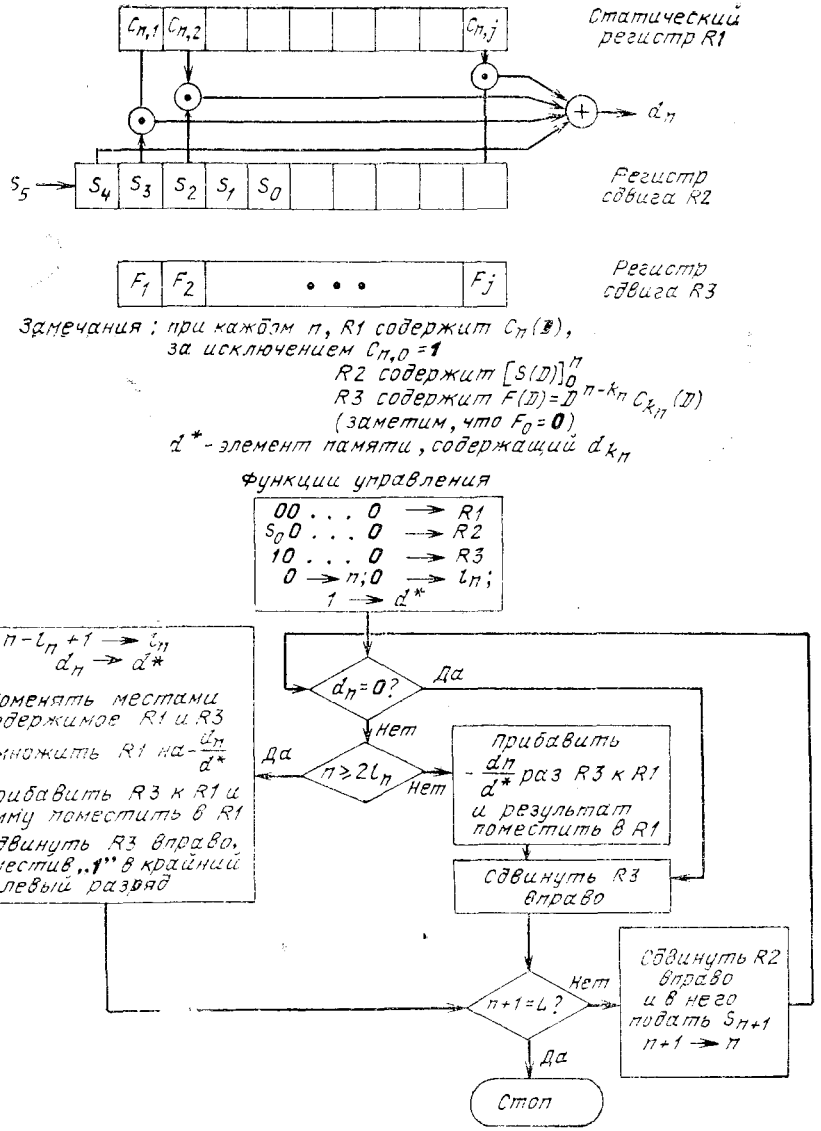


Рис. 6.7.4. Реализация РСЛОС-алгоритма для генерирования $[S(D)]_0^L$.

$[C_n(D),\ l_B]$- и $[B(D),\ l_B]$-регистров следует существование при некотором $j>0$ такого $[F(D),\ l_B-j]$-регистра, для которого

$$[F(D)S(D)]_{l_B-j}^{n-j}=fD^{n-i},\ f\neq 0.$$

Согласно лемме 3 $(n-j)-(l_B-j)\leqslant k_n-l_{kn}$. Поэтому $L_B\geqslant n-(k_n-l_{kn})=l_{n+1}$. Тогда $[B(D),\ l_B]$-регистр не короче, чем $[C_{n+1}(D),\ l_{n+1}]$-регистр, что завершает доказательство. |

Блок-схема, представленная на рис. 6.7.4, автором которой является Месси (1968), предлагает способ реализации алгоритма. Отметим, что соотношение (6.7.49) используется здесь в качестве критерия для указания изменения l_n и k_n. Длина регистров j, представленных на рис. 6.7.4, должна быть достаточной для хранения многочлена связей самого длинного ожидаемого регистра. Выберем для декодирования БЧХ-кодов $L=d-2$ и расположим $\sigma(D)$, кроме коэффициента $\sigma_0=1$, в $R1$ слева. Можно выбрать $j=\lfloor (d-1)/2 \rfloor$ и тем самым гарантировать исправление всех комбинаций не более чем $\lfloor (d-1)/2 \rfloor$ ошибок. В случае двоичных БЧХ-кодов элементы $\{S_i\}$ и $\{C_i\}$ являются элементами $GF(2^m)$ и каждый из регистров, представленных на рис. 6.7.4, может быть реализован при помощи m двоичных регистров. Устройство, выполняющее умножение в $GF(2^m)$, можно сделать так, как показано на рис. 6.6.5. Легко видеть, что сложность оборудования, необходимого для регистров и умножителей, пропорциональна md. Также нетрудно убедиться, что время, необходимое для нахождения $\sigma(D)$, пропорционально md [или чуть больше, что зависит от способа вычисления $(d^*)^{-1}$]. Безусловно, при построении такого устройства необходимо разрешить большое число технических вопросов; важно заметить, однако, что для такого устройства необходимо поразительно мало оборудования и поразительно мало вычислительного времени. Берлекэмп (1967) также доказал, что для двоичных кодов с $r=1$ и при нечетных n величина d_n всегда равна нулю. Использование этого обстоятельства, по существу, вдвое сокращает вычислительное время для нахождения $\sigma(D)$. На этом мы закончим обсуждение второго этапа процедуры декодирования БЧХ-кодов.

Теперь коротко рассмотрим реализации этапов 1, 3 и 4 для БЧХ-декодера. На этапе 1 можно вычислять элементы синдрома следующим способом:

$$S_i=\sum_{n=0}^{N-1} y_n\alpha^{(r+i)n}=(\ldots(y_{N-1}\alpha^{r+i}+y_{N-2})\alpha^{r+i}+y_{N-3})\alpha^{r+i}+\ldots+y_0).$$

Таким образом, возможный путь вычисления S_i состоит в сложении всех последовательно принятых символов в первоначально пустом регистре; сумма затем должна быть умножена на α^{r+i} и возвращена в регистр, ждущий следующего принятого символа.

Этап 3 легче всего реализовать при помощи процедуры Ченя (1964). Если число ошибок не превышает $\lfloor (d-1)/2 \rfloor$, то $\sigma(D)$, вычисленный на этапе 2, определяется соотношением (6.7.21), и ошибка

в n-й позиции (т. е. $z_n \neq 0$) произойдет тогда и только тогда, когда $\sigma(\alpha^{-n}) = 0$, или, что то же самое, тогда и только тогда, когда

$$\sum_{i=0}^{e} \sigma_i \alpha^{-ni} = 0. \tag{6.7.63}$$

Если обозначить

$$\sigma_{i,\,n} = \sigma_i \alpha^{-ni}, \tag{6.7.64}$$

то

$$\sigma_{i,\,N-1} = \sigma_i \alpha^{-(N-1)i} = \sigma_i \alpha^i \tag{6.7.65}$$

и при любом n

$$\sigma_{i,\,n-1} = \sigma_{i,\,n} \alpha^i. \tag{6.7.66}$$

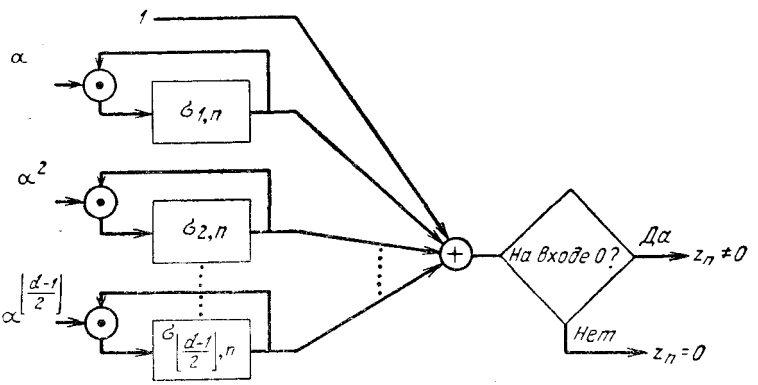


Рис. 6.7.5. Шаг 3 при декодировании БЧХ-кода: нахождение позиций ошибок. Первоначально в регистре хранятся $\sigma_1, \ldots, \sigma_{\lfloor (d-1)/2 \rfloor}$. Затем производится умножение и проверка гипотезы $z_{N-1} = 0$; после этого производится умножение и проверка гипотезы $z_{N-2} = 0$; и так далее до $z_0 = 0$.

Последнее предполагает использование схемы рис. 6.7.5 для реализации критерия (6.7.63) сначала для $n = N-1$, затем для $n = N-2$ и т. д.

Как уже было указано, в случае двоичных БЧХ-кодов этап 4 в процедуре декодирования не является необходимым. Поэтому принятые символы могут выходить из декодера синхронно с операциями, производимыми устройством на рис. 6.7.5, если на выходе устройства просто производится сложение y_n и z_n по модулю 2 при изменении n от $N-1$ до 0, что позволяет получить переданное кодовое слово, если произошло не более $\lfloor (d-1)/2 \rfloor$ ошибок.

Если при использовании двоичного кода произошло более чем $\lfloor (d-1)/2 \rfloor$ ошибок, то может произойти одно из следующих трех событий. Во-первых, длина [$\sigma(D)$, l]-регистра, который был на этапе 2, могла быть больше чем $\lfloor (d-1)/2 \rfloor$. Если в устройстве на рис. 6.7.4 $j = \lfloor (d-1)/2 \rfloor$, то в этом случае нельзя найти $\sigma(D)$, но довольно просто обнаружить это событие. Во-вторых, вполне возможно также, что l, найденное на этапе 2, не превышает $\lfloor (d-1)/2 \rfloor$, но $\sigma(D)$ не имеет l корней в $GF(2^m)$. В этом случае на этапе 3 будет

сделано меньше, чем l исправлений, но декодированная последовательность не будет представлять собой кодовое слово. Это вновь может быть легко обнаружено или путем подсчета числа исправлений, или путем проверки, является ли декодированная последовательность кодовым словом. Наконец, вполне возможно, что $[\sigma(D), l]$-регистр имеет длину $l \leqslant \lfloor (d-1)/2 \rfloor$ и что при декодировании будет найдено l ошибок. В этом случае декодированная последовательность будет кодовым словом, которое отличается от принятой последовательности более чем в $\lfloor (d-1)/2 \rfloor$ позициях. При этом нельзя обнаружить ошибку декодирования, но мы по крайней мере знаем, что декодер принял в двоичном симметричном канале решение по максимуму правдоподобия.

Обратимся теперь к отысканию значений ошибок (этап 4) в процессе декодирования недвоичных БЧХ-кодов. Определим многочлен $A(D)$ в (6.7.22) как

$$A(D) = \sum_{j=1}^{e} V_j U_j^r \prod_{l \neq j} (1 - U_l D). \tag{6.7.67}$$

Тогда согласно (6.7.25) $A(D)$ определяется следующим образом через $\sigma(D)$:

$$A(D) = [\sigma(D) S(D)]_0^{d-2}. \tag{6.7.68}$$

Многочлен $A(D)$ можно найти непосредственно из (6.7.68) или, что более красиво, его вычисление можно включить в виде составной части в итеративный алгоритм нахождения $\sigma(D)$. В частности, из начальных условий $A_{-1}(D) = -D^{-1}$ и $A_0(D) = 0$ вычислим при $n \geqslant 0$ многочлен $A_{n+1}(D)$ согласно формуле

$$A_{n+1}(D) = A_n(D) - \frac{d_n}{d_{k_n}} D^{n-k_n} A_{k_n}(D), \tag{6.7.69}$$

где d_n и k_n определяются (6.7.34) и (6.7.35). Для этого в блок-схему на рис. 6.7.4 необходимо включить два дополнительных регистра; никаких дополнительных логических элементов управления, по существу, не требуется, поскольку операции, производимые регистрами при вычислении $A_n(D)$ и $D^{n-k_n} A_{k_n}(D)$, аналогичны операциям, производимым регистрами при вычислении $C_n(D)$ и $D^{n-k_n} C_{k_n}(D)$. Доказательство того, что $[C_n(D) S(D)]_0^{n-1} = A_n(D)$ при любом $n \geqslant 0$ почти аналогично доказательству теоремы 6.7.3 и это приведено в виде задачи 6.35.

После нахождения $A(D)$ нетрудно убедиться, что согласно (6.7.67)

$$A(U_j^{-1}) = V_j U_j^r \prod_{l \neq j} (1 - U_l U_j^{-1}). \tag{6.7.70}$$

Можно упростить правую часть (6.7.70), если ввести производную $\sigma(D) = \sigma_0 + \sigma_1 D + \ldots + \sigma_e D^e$:

$$\sigma'(D) = \sigma_1 + 2\sigma_2 D + \ldots + e\sigma_e D^{e-1}. \tag{6.7.71}$$

Легко реализовать вычисление $\sigma'(D)$ по $\sigma(D)$; если q является степенью 2, то $\sigma'(D)$ равна сумме членов $\sigma(D)$ в нечетных степенях, деленной на D, так как

$$\sigma(D) = \prod_j (1 - U_j D),$$

то получим (см. задачу 6.36)

$$\sigma'(D) = -\sum_{j=1}^{e} U_j \prod_{l \neq i} [1 - U_l D],$$

$$\sigma'(U_j^{-1}) = -U_j \prod_{l \neq j} (1 - U_l U_j^{-1}). \tag{6.7.72}$$

Подставляя (6.7.72) в (6.7.70), получаем

$$V_j = -U_j^{1-r} \frac{A(U_j^{-1})}{\sigma'(U_j^{-1})}. \tag{6.7.73}$$

Используя определения U_j и V_j, данные в (6.7.16), получаем следующую формулу для нахождения каждого из ненулевых шумовых символов z_n:

$$z_n = -\frac{\alpha^{n(1-r)} A(\alpha^{-n})}{\sigma'(\alpha^{-n})}. \tag{6.7.74}$$

Каждая из трех функций, входящих в правую часть (6.7.74), может быть последовательно вычислена для всех n при изменении n от $N-1$ до 0 при помощи устройства того же типа, что и на рис. 6.7.5.

На этом заканчивается обсуждение декодирования БЧХ-кодов. Главный вывод состоит в том, что хотя по замыслу это декодирование является сложным, оно очень просто в смысле времени декодирования и требуемой сложности оборудования. Если не считать ячеек памяти, необходимых для запоминания принятого слова, и устройства вычисления обратных элементов в $GF(q^m)$, то количество требуемого оборудования пропорционально md. Время декодирования на этапе 2 пропорционально md, а на этапах 3 и 4 пропорционально mN.

Давайте посмотрим, что можно сказать о поведении двоичных БЧХ-кодов в пределе при $N \to \infty$. На время допустим, что $m(d-1)/2$ является хорошей оценкой для числа проверочных символов в коде, так что

$$\frac{m(d-1)}{2N} \approx 1 - R,$$

где R — скорость передачи в двоичных символах. Так как $m \geqslant \log_2(N+1)$, то при фиксированном R величина $(d-1)/2\ N$ должна стремиться к 0 при стремлении N к бесконечности. Поэтому число ошибок, которые можно исправить описанным алгоритмом декодирования, в конце концов становится меньше среднего числа ошибок в канале. Питерсон (1961) точно вычислил число проверочных символов в различных двоичных БЧХ-кодах; из его результатов следует, что при

фиксированном R величина $(d-1)/2\ N$ стремится к нулю с возрастанием N. Вместе с тем это уменьшение наступает при таких больших значениях N, что этот предел мало значит для практики.

Коды Рида-Соломона (1960) представляют собой интересный частный класс БЧХ-кодов с параметром m, равным 1; иначе говоря, в этом случае расширенное поле, в котором определено α, совпадает с полем символов для кодовых букв. Тогда минимальный многочлен для α^i равен $D-\alpha^i$, так что имеем

$$g(D)=\prod_{i=r}^{r+d-2}(D-\alpha^i).$$

Поэтому указанный код имеет $d-1$ проверочных символов и минимальное расстояние d. Нетрудно убедиться, что ни один групповой код с такими же объемом алфавита, длиной блока и числом проверочных символов не может иметь большего минимального расстояния, поскольку если выбрать все информационные символы, кроме одного, равными 0, то соответствующее кодовое слово будет иметь не более d ненулевых символов.

Так как в коде Рида-Соломона длина блока N равна $q-1$ или делителю числа $(q-1)$, то нетрудно заметить, что эти коды полезны лишь при больших объемах алфавита. Их можно эффективно использовать в непрерывных по времени каналах, где входной алфавит составляет огромное множество сигналов. Форни (1965) эффективно использовал их также в каскадной схеме, где символы кода Рида-Соломона являлись кодовыми словами в меньшем внутреннем коде. Форни показал, что такие коды можно использовать при скоростях передачи, как угодно близких к пропускной способности. Вероятность ошибки для них экспоненциально убывает с ростом длины блока, а сложность декодирования пропорциональна малой степени длины слова. Коды Рида-Соломона можно также непосредственно использовать в канале с малым входным алфавитом путем представления каждой буквы в кодовом слове в виде последовательности букв канала. Эта техника полезна для применения в каналах, где ошибки объединяются в группы, поскольку число операций декодирования зависит лишь от числа последовательностей выходных символов канала, содержащих ошибки.

6.8. СВЕРТОЧНЫЕ КОДЫ И ПОРОГОВОЕ ДЕКОДИРОВАНИЕ

Сверточные коды, которые будут определены в этом параграфе, отличаются от всех ранее рассмотренных кодов тем, что они являются неблоковыми кодами. Прежде чем определять эти коды, рассмотрим простой пример сверточного кода, иллюстрируемый рис. 6.8.1. В каждую единицу времени в кодер поступает новый двоичный символ источника u_n, а каждый из ранее поступивших символов источника сдвигается в регистре сдвига на один разряд вправо. Новый символ источника поступает непосредственно в канал, так что $x_n^{(1)}=u_n$. За каждым таким информационным символом следует проверочный символ

$$x_n^{(2)}=u_n\oplus u_{n-3}\oplus u_{n-4}\oplus u_{n-5}= \qquad (6.8.1)$$

$$= x_n^{(1)} \oplus x_{n-3}^{(1)} \oplus x_{n-4}^{(1)} \oplus x_{n-5}^{(1)}. \qquad (6.8.2)$$

Предполагая, что в начальном положении регистр сдвига заполнен нулями и что передача начинается с u_1, имеем $u_n = \mathbf{0}$, $x_n^{(1)} = \mathbf{0}$ и $x_n^{(2)} = \mathbf{0}$ при $n \leqslant 0$. Опишем теперь очень простой метод декодирования этого кода для случая передачи по двоичному каналу — пороговое декодирование.

Пусть

$\mathbf{x} = x_1^{(1)}\ x_1^{(2)}\ x_2^{(1)}\ x_2^{(2)}\ \ldots$ — передаваемая последовательность,
$\mathbf{y} = y_1^{(1)}\ y_1^{(2)}\ y_2^{(1)}\ y_2^{(2)}\ \ldots$ — принятая последовательность,
$\mathbf{z} = z_1^{(1)}\ z_1^{(2)}\ z_2^{(1)}\ z_2^{(2)}\ \ldots$ — шумовая последовательность,

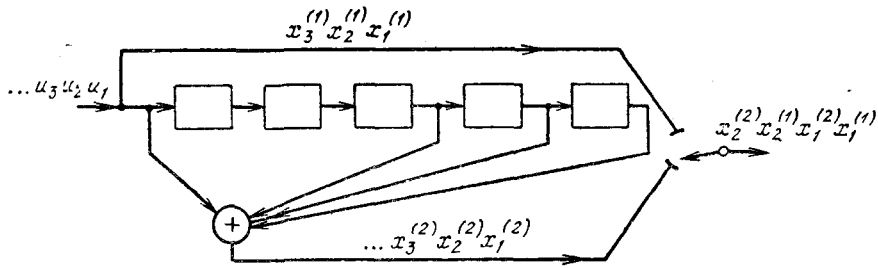


Рис. 6.8.1. Пример сверточного кода.

где $\mathbf{z} = \mathbf{y} \oplus \mathbf{x}$. Как и для кодов с проверкой на четность, определим синдром $\mathbf{S} = S_1\ S_2\ \ldots$ с помощью формулы

$$S_n = y_n^{(2)} \oplus y_n^{(1)} \oplus y_{n-3}^{(1)} \oplus y_{n-4}^{(1)} \oplus y_{n-5}^{(1)}. \qquad (6.8.3)$$

Следовательно, $S_n = \mathbf{0}$, если выполнена n-я проверка на четность, вычисленная в приемнике, и $S_n = \mathbf{1}$, в противном случае. Из (6.8.2) также имеем

$$S_n = z_n^{(2)} + z_n^{(1)} + z_{n-3}^{(1)} + z_{n-4}^{(1)} + z_{n-5}^{(1)}. \qquad (6.8.4)$$

Расписывая эти равенства при $1 \leqslant n \leqslant 6$, получаем:

$$\begin{aligned}
S_1 &= z_1^{(2)} \oplus z_1^{(1)}, \\
S_2 &= z_2^{(2)} \oplus z_2^{(1)}, \\
S_3 &= z_3^{(2)} \oplus z_3^{(1)}, \\
S_4 &= z_4^{(2)} \oplus z_4^{(1)} \oplus z_1^{(1)}, \\
S_5 &= z_5^{(2)} \oplus z_5^{(1)} \oplus z_2^{(1)} \oplus z_1^{(1)}, \\
S_6 &= z_6^{(2)} \oplus z_6^{(1)} \oplus z_3^{(1)} \oplus z_2^{(1)} \oplus z_1^{(1)}.
\end{aligned} \qquad (6.8.5)$$

Теперь остановимся на декодировании первого информационного символа u_1; декодирование эквивалентно определению, является ли $z_1^{(1)}$ единицей или нулем. Отметим, что величины S_1, S_4, S_5 и S_6 непосредственно содержат $z_1^{(1)}$. Например, если $z_1^{(1)} = \mathbf{1}$ и никаких других ошибок не произошло, все элементы S_1, S_4, S_5 и S_6 равны **1**,

тогда как, если ни одной ошибки не произошло, все они равны **0**. Это позволяет применить следующую стратегию декодирования: если большинство из элементов S_1, S_4, S_5 и S_6 равно **1**, то полагаем $z_1^{(1)} = 1$, в противном случае полагаем $z_1^{(1)} = 0$.

К описанной выше стратегии декодирования можно применить следующее простое усовершенствование. Легко видеть, что $z_2^{(1)}$ содержится в выражениях как для S_5, так и для S_6, поэтому если $z_1^{(1)} = 1$ и $z_2^{(1)} = 1$, то S_5 и S_6 будут равны **0**, S_1 и S_4 будут равны **1**, и в результате декодирование будет неправильным. Этой трудности можно избежать, суммируя S_2 и S_5, что приводит к соотношениям:

$$\begin{aligned} S_1 &= z_1^{(2)} \oplus z_1^{(1)}, \\ S_4 &= z_4^{(2)} \oplus z_4^{(1)} \oplus z_1^{(1)}, \\ S_5 \oplus S_2 &= z_5^{(2)} \oplus z_5^{(2)} \oplus z_2^{(2)} \oplus z_1^{(1)}, \\ S_6 &= z_6^{(2)} \oplus z_6^{(1)} \oplus z_3^{(1)} \oplus z_2^{(1)} \oplus z_1^{(1)}. \end{aligned} \tag{6.8.6}$$

Множество линейных комбинаций шумовых символов называется *ортогональным* к одному из шумовых символов, если этот символ входит (с ненулевыми коэффициентами) в каждую из линейных комбинаций того множества и ни один другой символ не входит (с ненулевыми коэффициентами) более чем в одну из этих линейных комбинаций. Таким образом, множество четырех линейных комбинаций в правой части (6.8.6) ортогонально к $z_1^{(1)}$. Заметим, что если $z_1^{(1)} \neq \mathbf{0}$ и все остальные $z_n^{(i)}$, входящие в линейные комбинации из множества, равны **0**, то значения всех четырех линейных комбинаций отличны от нуля. Если произойдет одна дополнительная ошибка (т. е. $z_n^{(i)} \neq \mathbf{0}$ для одного из других символов), то в силу ортогональности по меньшей мере значения трех линейных комбинаций будут отличны от нуля. Вместе с тем, если $z_1^{(1)} = \mathbf{0}$ и по крайней мере два других шумовых символа, входящих в совокупность, ненулевые, то значения не более чем двух линейных комбинаций отличны от нуля. Поэтому, если декодер после вычисления значений элементов, стоящих в левой части (6.8.6), положит $z_1^{(1)}$ равным **1**, когда большинство из этих символов равно **1**, и положит $z_1^{(1)} = \mathbf{0}$ в противном случае, то $z_1^{(1)}$ будет декодировано правильно, если число ненулевых шумовых символов в (6.8.6) не превышает два. Этот принцип непосредственно обобщает следующая теорема, утверждение которой справедливо для произвольного поля, хотя здесь будет рассматриваться $GF(2)$.

Теорема 6.8.1. Предположим, что при произвольном положительном целом e декодер может вычислить значения $2e$ линейных комбинаций шумовых символов, образующих множество, ортогональное к $z_1^{(1)}$. Тогда если число ненулевых шумовых символов, входящих в линейные комбинации, не превышает e, то к правильному декодированию $z_1^{(1)}$ приводит следующее правило. Если более половины линейных комбинаций имеют одно и то же значение α, то декодер полагает $z_1^{(1)} = \alpha$; в противном случае декодер полагает $z_1^{(1)} = 0$.

Устройство, изображенное на рис. 6.8.2, называется *пороговым декодером*; для кода, представленного на рис. 6.8.1, показано декодирование первого информационного символа с помощью описанного метода.

Пунктирные линии на рис. 6.8.2 относятся к декодированию последующих информационных символов, к которому мы теперь перейдем. Если все индексы в левой части (6.8.6) увеличить на 1, то получим:

$$\begin{aligned} S_2 &= z_2^{(2)} \oplus z_2^{(1)}, \\ S_5 &= z_5^{(2)} \oplus z_5^{(1)} \oplus z_2^{(1)} \oplus z_1^{(1)}, \\ S_6 \oplus S_3 &= z_6^{(2)} \oplus z_6^{(1)} \oplus z_3^{(2)} \oplus z_2^{(1)} \oplus z_1^{(1)}, \\ S_7 &= z_7^{(2)} \oplus z_7^{(1)} \oplus z_4^{(1)} \oplus z_3^{(1)} \oplus z_2^{(1)}. \end{aligned} \tag{6.8.7}$$

Следует заметить, что если исключить $z_1^{(1)}$, то множество этих соотношений будет ортогонально к $z_2^{(1)}$; поэтому исправление $z_2^{(1)}$ может

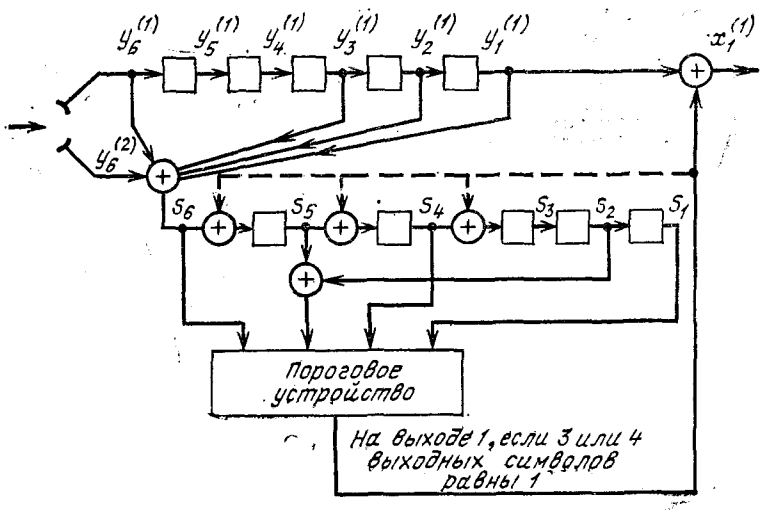


Рис. 6.8.2. Пороговый декодер.

быть выполнено представленным на рис. 6.8.2 устройством путем сдвига синдромных символов вправо после исправления $z_1^{(1)}$. Теперь объясним назначение пунктирных линий на рис. 6.8.2. Они изображают линии обратной связи, устраняющие влияние ошибки на синдром после того, как она исправлена. Поэтому, если $z_1^{(1)} = 1$ и декодирование проведено правильно, то декодер полагает $z_1^{(1)} = 0$ и изменяет S_4, S_5 и S_6 соответственно. Таким образом, если число ошибок в ортогональных соотношениях не превышает двух и при декодировании предшествующих символов не произошло ошибок, то все информационные символы последовательности будут декодированы правильно.

Значительно сложнее исследование вопроса о том, что случится с декодером на рис. 8.6.2 *после* того, как произойдет ошибка декодиро-

вания. Экспериментальные исследования показали, что, как правило, после того как декодер совершит ошибку декодирования, он сделает еще ошибки при декодировании последующих примерно пяти символов, а затем вновь будет декодировать правильно. Для рассмотренного частного кода Месси (1964) теоретически показал, что поступление в декодер достаточно длинной последовательности неискаженных символов после ошибки декодирования возвращает декодер к правильному декодированию. Это стремление к размножению ошибок характерно для схем декодирования сверточных кодов. Для очень простых кодов и декодеров типа рассмотренных выше это размножение ошибок не слишком серьезно и обычно приводит лишь к коротким пакетам ошибок декодирования. Однако если увеличивать длину кодового ограничения и усложнять схему декодирования, то это размножение ошибок приводит к более серьезным последствиям. Вместе с тем, чем серьезнее становится проблема размножения ошибок, тем легче декодеру распознать наличие ошибки декодирования. Если декодер имеет обратную связь с передатчиком, то передатчик может затем повторить передачу пропавших данных*). Кроме того, в кодер можно периодически подавать известную последовательность нулей, после чего декодер может начинать декодирование сначала.

Пример сверточного кодера, представленный на рис. 6.8.1, можно теперь обобщить в трех направлениях. Во-первых, можно произвольно выбрать длину регистра сдвига и его отводы, определяющие проверочные символы. Во-вторых, можно сделать так, чтобы символы источника и символы канала были бы элементами произвольного поля Галуа. Наконец, можно обобщить коды на скорости, отличные от скорости «один символ источника на два символа канала». Для того чтобы провести это последнее обобщение, разделим последовательность символов источника на блоки заданной длины λ. Каждому множеству символов источника кодер ставит в соответствие данное число ν символов канала, как показано на рис. 6.8.3. Если обозначить последовательность источника $u_1^{(1)}, u_1^{(2)}, \ldots, u_1^{(\lambda)}, u_2^{(1)}, u_2^{(2)}, \ldots, u_2^{(\lambda)}, \ldots$ и последовательность символов канала $x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(\nu)}, x_2^{(1)}, \ldots$, то правило образования символов канала по символам источника можно записать в виде

$$x_n^{(i)} = \sum_{l=0}^{L-1} \sum_{j=1}^{\lambda} g_{j,\,i}(l)\, u_{n-l}^{(j)}, \quad 1 \leqslant i \leqslant \nu, \tag{6.8.8}$$

где L (как показано на рис. 6.8.3) — длина регистра сдвига, а элементы $g_{j,\,i}(l)$ определяют связи между регистрами сдвига и сумматорами. При суммировании по модулю 2 коэффициент $g_{j,\,i}(l) = 1$, если l-й разряд j-го регистра сдвига связан с сумматором, образующим i-й поток символов канала. В общем случае все элементы $u_l^{(j)}$, $g_{j,\,i}(l)$ и $x_n^{(i)}$ принадлежат данному полю Галуа и элементы, поступающие на

*) Некоторые методы передачи, применяемые в случае, когда линия обратной связи зашумленная, рассмотрены Возенкрафтом и Хорстейном (1961) и Метцнером и Морганом (1960).

входы сумматоров (см. рис. 6.8.3), должны быть умножены на соответствующие элементы $g_{j,\,i}(l)$. В примере, приведенном на рис. 6.8.1, $\lambda = 1$, $\nu = 2$, $L = 6$ и $g_{1,1}(0) = g_{1,2}(0) = g_{1,3}(3) = g_{1,2}(4) = g_{1,2}(5) = \mathbf{1}$.

Сверточный код называется *систематическим*, если первые λ символов канала в каждом блоке длины ν совпадают с соответствующими λ символами источника. В этом случае, для $i \leqslant \lambda$ $g_{j,\,i}(l)$ равно **1** при $l = 0$, $i = j$ и **0** при других значениях l и j. Оба кодера, представленные на рис. 6.8.1 и 6.8.3, систематические.

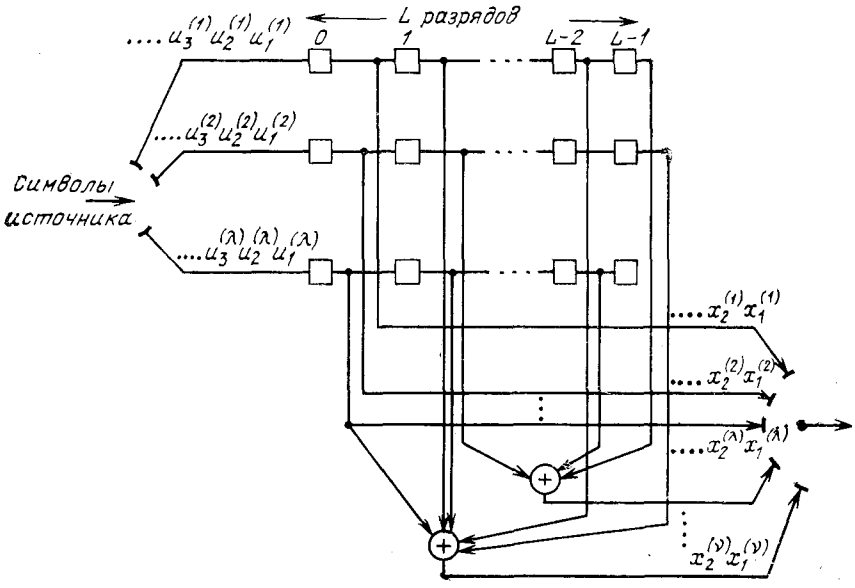


Рис. 6.8.3. Общий систематический сверточный кодер.

Длина кодового ограничения сверточного кода, определяемая как $N = \nu L$, равна числу символов канала, выходящих из декодера в течение времени между поступлением данного символа источника в кодер и выходом его из кодера. Длина кодового ограничения сверточного кода играет ту же роль, что и длина блока в блоковом коде.

Пороговое декодирование можно использовать для любого сверточного кода (и также, конечно, для блокового кода с проверкой на четность). Единственный вопрос состоит в том, сколько ортогональных линейных комбинаций можно найти для каждого символа. Обычно это решается методом проб и ошибок. Некоторые сверточные коды, соответствующие различным скоростям и длинам блоков, а также правила их ортогонализации табулированы Месси (1963). Значительно эффективнее применять пороговое декодирование для исправления ошибок при малых длинах ограничения; необходимо заметить также, что при возрастании длины кодового ограничения не может быть получена произвольно малая вероятность ошибки*). В § 6.9 мы изучим

*) При пороговом декодировании. (*Прим. ред.*).

другой метод декодирования сверточных кодов — последовательное декодирование, для которого вероятность ошибки может быть сделана произвольно малой с помощью увеличения длины кодового ограничения.

6.9. ПОСЛЕДОВАТЕЛЬНОЕ ДЕКОДИРОВАНИЕ

Идею последовательного декодирования можно лучше всего понять на следующем простом примере. Рассмотрим двоичный сверточный кодер, схема которого представлена на рис. 6.9.1. Как обычно, предполагается, что до поступления первого символа источника u_1 он содержит лишь нули. Первые три символа канала $x_1^{(1)}, x_1^{(2)}, x_1^{(3)}$ полностью определяются значением u_1. Следующие за ними три символа

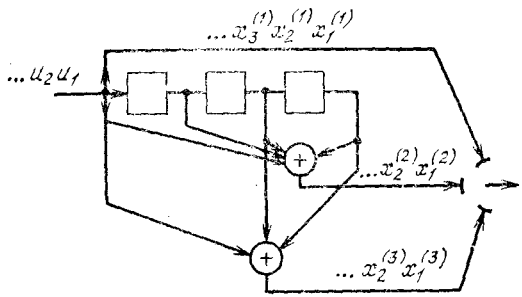


Рис. 6.9.1. Пример сверточного кода.

канала являются функциями обоих символов u_1 и u_2, а последующие три символа — функции u_1, u_2 и u_3 и т. д. Эта зависимость объясняется древовидной структурой множества кодовых последовательностей в канале, выявляемой из рассмотрения рис. 6.9.2.

Крайние слева тройки двоичных символов на рис. 6.9.2 соответствуют отклику кодера на поступление в него первого символа **1**, или **0**. Отходящие вправо от множества **111**, соответствующего отклику кодера на поступление в него первого символа **1**, тройки символов являются откликами на поступление вслед за первой **1** новой **1** или **0**, и т. д. Например, пунктиром на рис. 6.9.2 обозначен отклик кодера на поступление $u_1 = 1$, $u_2 = 1$, $u_3 = 0$, $u_4 = 0$.

Исследуем теперь качественно, как можно использовать при декодировании эту древовидность структуры. Позднее мы определим последовательное декодирование формально, развивая довольно простой по смыслу подход, который будет сейчас рассмотрен.

Допустим, что кодер, представленный на рис. 6.9.1, используется в двоичном симметричном канале и что первыми двенадцатью символами принятой последовательности являются **101 101 001 000**. Из рассмотрения дерева на рис. 6.9.2 видно, что первая тройка переданных символов может быть либо **111**, либо **000**. Поэтому, основываясь на первой тройке принятых символов, мы можем попробовать принять

гипотезу, что передано множество **111**, соответствующее верхнему разветвлению первого ребра дерева. Пусть принята эта гипотеза; тогда следующие три символа должны быть символами одной из двух верхних троек символов, обозначенных на рис. 6.9.2 как $x_2^{(1)}$, $x_2^{(2)}$, $x_2^{(3)}$. Очевидно, что предпочтительнее принятие гипотезы **101**, соответствующей верхнему разветвлению. Продолжая таким же образом, мы принимаем гипотезу, что третья тройка переданных символов равна **001**,

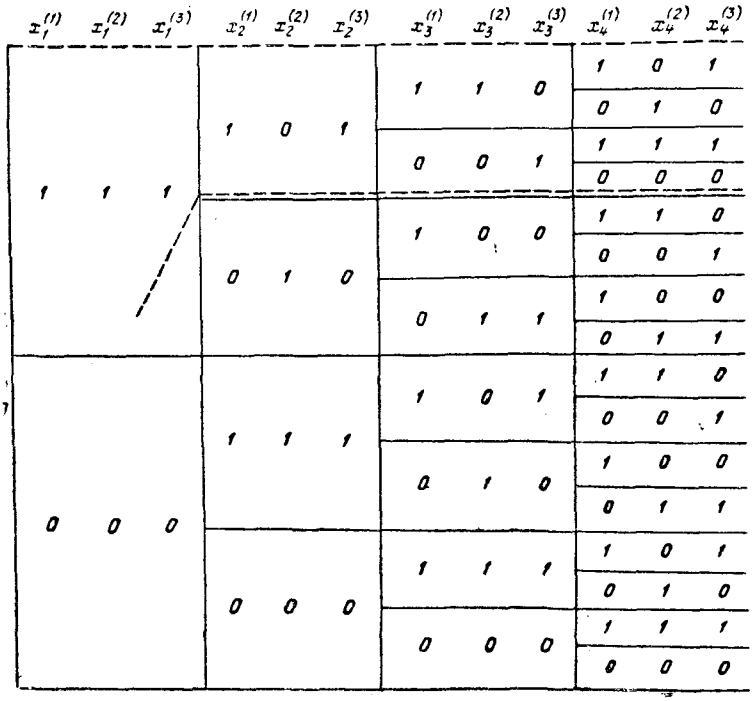


Рис. 6.9.2. Древовидная структура сверточного кода.

а четвертая **000**. Этим гипотезам соответствует пунктирная линия на рис. 6.9.2. Метод, который был продемонстрирован — это, по существу, декодирование символ за символом, при котором выбор гипотезы при декодировании тройки символов используется для уменьшения числа возможных выборов при декодировании последующей тройки. В рассмотренном примере совпадение всех, начиная с третьего, символов последовательности, которую мы на основании нашей гипотезы считаем переданной, с символами принятой последовательности подтверждает, что принятые ранее гипотезы правильны.

Теперь рассмотрим второй пример, который иллюстрирует, что случится, если была принята неправильная гипотеза. Предположим, что передана та же самая первоначальная последовательность **111 101**

001 000, но принята последовательность **010 101 001 000**... Декодер, которому не известно, что было передано, принимает гипотезу, что первые три переданные символа равны **000**, что соответствует нижнему ребру, выходящему из первой точки ветвления. Две следующие принятые гипотезы тогда соответствуют символам **111** и **101**. Из этих первых девяти символов последовательности, которую мы считаем переданной, три символа не совпадают с символами принятой последовательности. Случившееся является следствием того, что декодер, приняв однажды неправильную гипотезу, вынужден выбирать последующие гипотезы среди последовательностей, не имеющих отношения к принятой последовательности. Поэтому в конце концов декодер, как правило, способен распознать, что он, по-видимому, принял неправильную гипотезу. Тогда можно вернуться назад, проверить альтернативные гипотезы, после чего, вероятно, декодирование будет правильным. Таким образом, принятие каждой гипотезы упрощает дальнейший процесс проверки гипотез путем уменьшения числа возможных выборов и в то же время дает дополнительные данные для проверки правильности ранее принятых гипотез.

Последовательный декодер является декодером, предназначенным для кодов с древовидной структурой, который декодирует их путем принятия пробных гипотез относительно последовательных ребер дерева и изменяет эти гипотезы, если последующий выбор указывает на неправильность ранее принятых гипотез. К сожалению, чрезвычайная простота этой концепции исчезнет при четкой формулировке правила о том, должен ли декодер продолжать принимать новые гипотезы, или он должен вернуться для изменения старой гипотезы. Нашим целям отвечают три требования к стратегии поиска при последовательном декодировании. Во-первых, стратегия должна в конце концов с высокой вероятностью привести к правильному декодированию последовательности источника. Во-вторых, количество вычислений, требуемых декодером, не должно быть чрезмерным. Наконец (это требование не возникает для действующих систем), стратегия должна допускать математическое исследование. Первая такая стратегия (или алгоритм) была предложена Возенкрафтом (1957), который к тому же явился автором идеи последовательного декодирования. Мы ограничимся здесь рассмотрением усовершенствованного более позднего алгоритма, принадлежащего Фано (1963).

Предполагается, что канал является дискретным каналом без памяти с переходными вероятностями $P(j \mid k)$. Детальное рассмотрение кодера будет проведено позднее; коротко говоря, он состоит из трех блоков. Первый из них является двоичным сверточным кодером, аналогичным представленному на рис. 6.8.3 или описываемому соотношением (6.8.8). Каждую единицу времени в двоичный кодер поступают от источника λ двоичных символов, а он производит на выходе $a\nu$ двоичных символов, где a и ν целые. Второй блок осуществляет сложение фиксированной двоичной последовательности и этой двоичной последовательности. В-третьих, полученная сумма разбивается на подблоки из a символов каждый, и каждый подблок отображается во входную букву канала с помощью отображения, аналогичного представленному

на рис. 6.2.1*). Мы видим, что при таком кодировании каждые λ символов источника порождают ν символов канала и что скорость кода, вычисленная в натуральных единицах на символ канала, равна

$$R = \frac{\lambda}{\nu} \ln 2. \tag{6.9.1}$$

Графически можно представить такое кодирование в виде древовидной структуры, аналогичной изображенной на рис. 6.9.2, отличающейся, однако, тем, что из каждой точки ветвления выходит не 2 возможных ребра, как на рисунке, а 2^{λ}, и что каждое ребро состоит из ν символов канала.

Обозначим теперь через $\mathbf{x}_l = (x_1^{(1)}, \ldots, x_1^{(\nu)}, \ldots, x_2^{(1)}, \ldots, x_l^{(\nu)})$ первые νl символов кодовой последовательности, и через $\mathbf{y}_l = (y_1^{(1)}, \ldots, y_l^{(\nu)})$ первые νl символов принятой последовательности. Определим функцию $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$ следующим образом:

$$\Gamma(\mathbf{x}_l; \mathbf{y}_l) = \sum_{n=1}^{l} \sum_{i=1}^{\nu} \left[\ln \frac{P\left(y_n^{(i)} \mid x_n^{(i)}\right)}{\omega\left(y_n^{(i)}\right)} - B \right]. \tag{6.9.2}$$

В этом выражении B есть произвольное число, определяющее смещение, значение которого будет выбрано позднее, и $\omega(j)$ есть вероятность появления j-й буквы выходного алфавита канала,

$$\omega(j) = \sum_{k=0}^{K-1} Q(k) P(j \mid k), \tag{6.9.3}$$

где $Q(k)$ — относительная частота k-й буквы при отображении двоичных символов во входные символы канала [см. (6.2.6)]. Назовем $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$ *ценой* гипотезы $\mathbf{x}_l$ и будем использовать ее как меру согласования $\mathbf{x}_l$ с принятой последовательностью $\mathbf{y}_l$. Это разумная мера, так как при фиксированном $\mathbf{y}_l$ величина Γ является возрастающей функцией вероятности $P_l(\mathbf{y}_l \mid \mathbf{x}_l)$. При таком подходе довольно трудно понять назначение члена $\omega(y_n^{(i)})$ в соотношении (6.9.2); грубо говоря, он вводится для того, чтобы уменьшить зависимость источника от безусловных вероятностей символов выходной последовательности.

В случае двоичного симметричного канала с переходной вероятностью ε последние два блока описанного выше кодера могут быть отброшены и соотношение (6.8.8) полностью описывает кодирование. Тогда $\omega(y_n^{(i)})$ является постоянной и $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$ упрощается:

$$\Gamma(\mathbf{x}_l; \mathbf{y}_l) = -d(\mathbf{x}_l; \mathbf{y}_l) \ln \frac{1-\varepsilon}{\varepsilon} + \nu l \{\ln [2(1-\varepsilon)] - B\},$$

где $d(\mathbf{x}_l; \mathbf{y}_l)$ — расстояние Хэмминга между $\mathbf{x}_l$ и $\mathbf{y}_l$.

---

*) При таком отображении теряется некоторая информация относительно закодированной двоичной последовательности, но если код выбран соответствующим образом, то не теряется никакой информации относительно источника. Теорема кодирования § 6.2 убеждает нас в справедливости этого утверждения для блоковых кодов; ниже оно будет аналогичным образом доказано и для сверточных кодов.

Назначение декодера, выраженное в терминах цены Γ, состоит в проверке гипотез так чтобы величина $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$ возрастала с ростом l. Когда Γ начинает убывать с ростом l, по всей видимости, декодер попал на ложный путь и необходимо провести соответствующий поиск. Более наглядно эта идея иллюстрируется на рис. 6.9.3, на котором графическое представление величины $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$ отражает древовидную структуру возможных значений $\mathbf{x}_l$. Этот рисунок, который называется деревом принятых цен, соответствует кодеру рис. 6.9.1 и принятой последовательности **010 101 001 000**.

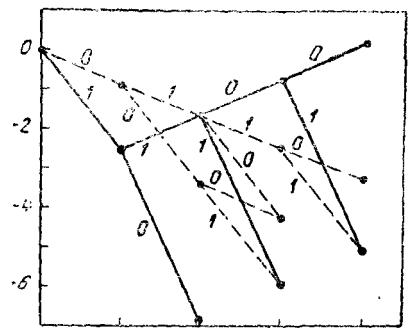


Рис. 6.9.3. Дерево принятых цен; ДСК; $\varepsilon=0{,}15$; кодер изображен на рис. 6.9.1; $\mathbf{y}_4 = \mathbf{010\ 101\ 001\ 000}$; $B = 1/3\ \ln 2$.

На рис. 6.9.3 узлы отмечены в соответствии с информационными последовательностями, ведущими в них. Иначе говоря, каждый узел на глубине l в дереве соответствует различным информационным последовательностям $\mathbf{u}_l$ из $l\lambda$ двоичных символов. Например, узел в верхнем правом углу рисунка соответствует информационной последовательности $\mathbf{u}_4 = \mathbf{1100}$, при передаваемой последовательности $\mathbf{x}_4 = \mathbf{111\ 101\ 001\ 000}$.

В принципе декодер может построить дерево принятых цен, аналогичное изображенному рис. 6.9.3, для произвольного сверточного кода и произвольной принятой последовательности. Однако *на практике* такое дерево декодер построить не может, так как число узлов в дереве растет экспоненциально с ростом l. То, что возможно — это включение в состав декодера точного аналога сверточного декодера.

Для любой заданной гипотетической последовательности информационных символов $\mathbf{u}_l$ декодер может, прогнав $\mathbf{u}_l$ через аналог кодера, найти соответствующую последовательность $\mathbf{x}_l$ и вычислить $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$, что является ценой узла, соответствующего $\mathbf{u}_l$.

Алгоритм декодирования, который будет описан, есть совокупность правил, определяющих движение от одного узла к другому. Мы будем допускать лишь три вида движений: вперед, вбок и назад. При *движении вперед* декодер продвигается в дереве полученных цен на одно ребро вправо от ранее проверенного узла. Технически это соответствует сдвигу вправо регистра сдвига в аналоге кодера и поступлению в него слева λ новых гипотетических информационных символов. Так как новая информационная последовательность $\mathbf{u}_{l+1}$ отличается от старой лишь наличием λ символов, прибавленных к ней, новая цена $\Gamma(\mathbf{x}_{l+1}; \mathbf{y}_{l+1})$ может быть найдена из $\Gamma(\mathbf{x}_l; \mathbf{y}_l)$ добавлением еще одного члена в сумме (6.9.2):

$$\Gamma(\mathbf{x}_{l+1}; \mathbf{y}_{l+1}) = \Gamma(\mathbf{x}_l; \mathbf{y}_l) + \sum_{i=1}^{\nu} \left[\ln \frac{P\left(y_{l+1}^{(i)} \mid x_{l+1}^{(i)}\right)}{\omega\left(y_{l+1}^{(i)}\right)} - B \right]. \qquad (6.9.4)$$

Символы, входящие в эту сумму, есть просто ν входных символов канала, порождаемых аналогом кодера. *Движение вбок* определяется как движение из одного узла в другой, отличающийся лишь в последнем ребре дерева. Технически это соответствует изменению крайне левых λ символов в регистре сдвига аналога кодера. Как и ранее, изменение в цене при переходе от одного узла к другому определяется изменением последних ν входных символов канала. *Движение назад* определяется как движение на одно ребро влево в дереве принятых цен. Технически это соответствует сдвигу регистра сдвига кодера влево и восстановлению последних λ символов, вышедших из регистра

Правило	Условия в узле		Действия, которые следует выполнить	
	Предыдущее движение	Сравнение Γ_{l-1} и Γ_l с первоначальным порогом	Окончательный порог	Движение
1	F или L	$\Gamma_{l-1}<T+\Delta$, $\Gamma_l \geqslant T$	Повышается*	F^+
2	F или L	$\Gamma_{l-1}\geqslant T+\Delta$, $\Gamma_l \geqslant T$	Не изменяется	F^+
3	F или L	любое Γ_{l-1}, $\Gamma_l<T$	Не изменяется	L или $B^{\ddagger}$
4	B	$\Gamma_{l-1}<T$, любое Γ_l	Понижается на Δ	F^+
5	B	$\Gamma_{l-1}\geqslant T$, любое Γ_l	Не изменяется	L или $B^{\ddagger}$

* Прибавить к порогу $j\Delta$, где j выбрано таким образом, что $\Gamma_l-\Delta<T+j\Delta\leqslant\Gamma_l$.

$^+$ Движение *вперед* (Forward) в первый из 2^λ узлов, исходящих из текущего узла (предполагается предварительное упорядочение 2^λ узлов).

‡ Движение *вбок* (Lateral) в следующий узел, отличающийся от текущего узла лишь в последнем ребре (предполагается то же предварительное упорядочение, что и выше); если текущий узел является последним среди этих 2^λ узлов, то совершается движение *назад* (Backward).

Рис. 6.9.4. Правила движения декодера.

сдвига вправо. Новое значение цены вычисляется из старого вычитанием последнего слагаемого из суммы по n в (6.9.2). Поэтому для каждого возможного движения изменение цены при переходе от старого узла к новому зависит лишь от последних ν гипотетических входных символов канала.

Алгоритм, использующий движение от одного узла к другому, является модификацией алгоритма Фано и представлен на рис. 6.9.4. В формулировку правила входит цена Γ_l текущего узла, который проверяется в данный момент, цена Γ_{l-1} узла, находящегося на одно ребро слева от текущего узла, и *порог* T. На порог T наложено ограничение, состоящее в том, что при увеличении он изменяется на некоторое фиксированное число Δ, величина которого определяется алгоритмом.

В начале декодирования устанавливается следующие начальные условия: проверяется исходный узел $\mathbf{u}_0$, что соответствует тому, что регистр сдвига содержит лишь нули, причем ни один информационный

u_l	Окончательный порог	Движение	v_l	Окончательный порог	Движение
—	0	*F*	010	—3	*L*
0	0	*L*	011	—3	*F*
1	0	*B*	0110	—3	*L*
—	—1,5	*F*	0111	—3	*B*
0	—1,5	*F*	011	—3	*B*
00	—1,5	*L*	01	—3	*B*
01	—1,5	*B*	0	—3	*L*
0	—1,5	*L*	1	—3	*F*
1	—1,5	*B*	10	—3	*L*
—	—3	*F*	11	—3	*F*
0	—3	*F*	110	—1,5	*F*
00	—3	*L*	1100	0	*F*
01	—3	*F*			

Рис. 6.9.5. Запись поиска декодера, изображенного на рис. 6.9.3 ($\Delta=1{,}5$).

символ еще не проверялся. Значение Γ_0 выбирается равным 0 и по определению $\Gamma_{-1} = -\infty$. Начальное значение порога не определено, но условимся считать, что декодер при начальной проверке следует правилу 1 с конечным порогом, равным 0.

Легко видеть, что движение вперед может быть предпринято в любой из 2^λ узлов, отстоящих на одно ребро вправо от текущего. Предполагается, что узлы предварительно упорядочены и движение вперед всегда совершается в первый по порядку узел, а движение вбок — в следующий по порядку после текущего. Эта упорядоченность несущественна для исследования, но здесь мы будем предполагать лексикографическую упорядоченность гипотетических информационных подпоследовательностей длины λ, считая **00...0** первой, **0...01** следующей и **11...1** последней последовательностью*).

В дальнейшем мы увидим, что правило 1 алгоритма декодирования является правилом, обычно используемым при проверке декодером узлов, соответствующих действительно переданной последовательности, когда действие шума не слишком велико. В этих условиях порог будет подниматься таким образом, чтобы разность между ценой узла и величиной порога не была бы больше Δ. После движения вперед цена этого прежнего узла будет появляться как Γ_{l-1}; то обстоятельство, что $\Gamma_{l-1} < T + \Delta$, вновь обеспечивает нам возможность применения правила 1 и нового повышения порога, если только новая цена Γ_l выше, чем старая цена Γ_{l-1}. Если первоначально проверялся ложный узел, то скорее всего цена узла будет меньше порога и декодер будет совершать движения вбок до тех пор, пока не перейдет к проверке правильного узла, а порог будет повышен вновь.

*) В первоначальном алгоритме Фано (1963) упорядочение узлов производилось в порядке убывания величины Γ. Вообще говоря, это уменьшает число узлов, которое должно быть проверено, но увеличивает количество вычислений, необходимых на проверку, из-за дополнительных вычислений, затрачиваемых на упорядочение узлов.

Если действие шумов сильнее обычного, поведение алгоритма существенно сложнее, что иллюстрируется схемой рис. 6.9.5. Грубо говоря, ситуация, представленная на рисунке, вызвана тем, что декодер пытается найти путь в дереве принятых цен, который находится выше текущего порога. Как мы позднее увидим, если такого пути не существует, декодер вынужден двигаться назад к узлу, где порог согласовывался бы с его текущей ценой. В этом узле порог понижается (по правилу 4) и декодер снова движется вперед, пытаясь найти узел, остающийся выше этого пониженного порога. Ограничение $\Gamma_{l-1} < T + \Delta$ в правиле 1 введено с целью предохранить порог от нового повышения в одном из тех узлов, которые были предварительно проверены. В самом деле, если бы это ограничение не было введено, декодер быстро вошел бы в цикл, проверяя снова и снова один и тот же узел при одном и том же пороге.

Прежде чем сформулировать эту идею точнее, необходимо ввести дополнительные определения. Назовем *F-проверкой* проверку по правилам 1, 2 или 4 (т. е. по тем правилам, для которых последующее движение совершается вперед). *Предшественниками* узла $\mathbf{u}_l$ являются узлы, связывающие $\mathbf{u}_l$ с началом $\mathbf{u}_0$, т. е. узлы, соответствующие префиксам информационной последовательности $\mathbf{u}_l$. *Путь цен* $\Gamma_0, \ldots, \Gamma_l$, относящийся к узлу $\mathbf{u}_l$, образован ценами предшественников узла $\mathbf{u}_l$ и ценой самого $\mathbf{u}_l$. *Путь порогов* $T_0, \ldots, T_l$, относящийся к проверке $\mathbf{u}_l$, образован окончательными порогами при последних проверках гипотез относительно их предшественников $\mathbf{u}_l$ (T_l — это окончательный порог для текущей проверки). *Потомками* узла $\mathbf{u}_l$ называются узлы, для которых $\mathbf{u}_l$ является предшествующим узлом, т. е. узлы, расположенные направо от $\mathbf{u}_l$. *Непосредственными потомками* $\mathbf{u}_l$ называются 2^λ потомков, отстоящих от $\mathbf{u}_l$ на одно ребро. *Путь* из узла $\mathbf{u}_i$ к потомку $\mathbf{u}_l$ лежит *выше порога* T, если цены в узлах пути удовлетворяют неравенству $\Gamma_j \geqslant T$ при $i \leqslant j \leqslant l$.

Следующая теорема, которая доказывается в приложении 6А, описывает работу алгоритма. С точки зрения последующего анализа наиболее важными пунктами теоремы являются пункты (*б*) и (*в*); эти пункты лучше всего понять, связав чтение теоремы с последующим обсуждением.

Теорема 6.9.1.

а) Путь порогов $T_0, \ldots, T_l$ и путь цен $\Gamma_0, \ldots, \Gamma_l$, связанные с проверкой узла $\mathbf{u}_l$, удовлетворяют следующим соотношениям при $0 \leqslant i \leqslant l-1$:

$$T_i \leqslant \Gamma_i, \tag{6.9.5}$$

$$T_{i+1} \geqslant T_i, \tag{6.9.6}$$

$$T_{i+1} \geqslant T_i + \Delta \Rightarrow T_i + \Delta > \Gamma_i, \tag{6.9.7}$$

$$T_{i+1} \geqslant T_i + \Delta \Rightarrow T_{i+1} > \Gamma_i. \tag{6.9.8}$$

б) Для каждого узла, относительно которого когда-либо проводилась F-проверка, окончательный порог T при первой F-проверке связан с ценой Γ этого узла соотношением

$$T \leqslant \Gamma < T + \Delta. \tag{6.9.9}$$

Окончательный порог при каждой последующей F-проверке этого узла лежит на Δ ниже, чем при предшествующей F-проверке этого узла.

в) Пусть относительно узла $\mathbf{u}$ проведена F-проверка с окончательным значением порога, равным T. Тогда, прежде чем узел $\mathbf{u}$ сможет быть проверен вновь, должны быть проведены F-проверки с окончательным порогом T в каждом из потомков узла $\mathbf{u}$, для которых путь, ведущий из $\mathbf{u}$, лежит выше T. Кроме того, в течение времени между данной проверкой узла $\mathbf{u}$ и его первой перепроверкой порог не может быть сделан ниже T.

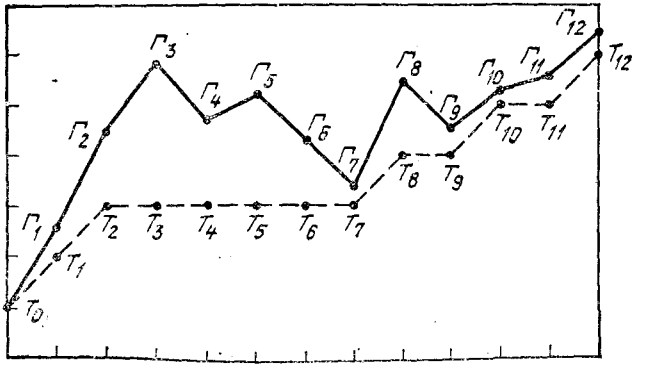


Рис. 6.9.6. Траектория порогов, связанная с проверкой последовательности $\mathbf{u}_{12}$.

Обсуждение. Пусть заданы бесконечные*) передаваемая и принятая последовательности символов; назовем *правильным путем* последовательность цен $\Gamma_0, \Gamma_1, \ldots$ в узлах, соответствующих переданной информационной последовательности. Предположим, что этот правильный путь флуктуирует с положительным смещением, а каждый другой путь в дереве принятых цен флуктуирует с отрицательным смещением. Обозначим через T_l окончательный порог при первой F-проверке узла $\mathbf{u}_l$, где $\mathbf{u}_l$ — произвольный узел правильного пути. Согласно пункту б) теоремы $T_l \leqslant \Gamma_l < T_l + \Delta$. Назовем узел $\mathbf{u}_l$ правильного пути *прорвавшимся*, если $\Gamma_i \geqslant T_l$ для всех $i > l$. Из пункта в) теоремы вытекает, что если узел $\mathbf{u}_l$ прорвался, то он не будет перепроверяться до тех пор, пока не будет проверен каждый последующий узел правильного пути, т. е., иными словами, вообще не будет перепроверяться. Так как все неправильные пути флуктуируют вниз, то декодер в результате движется вперед от одного прорвавшегося узла к следующему; поэтому использование алгоритма приведет в конце концов к декодированию каждого символа правильного пути. Отметим, однако, что декодер никогда не может быть вполне уверен, что он «декоди-

*) Если последовательности имеют конечную длину, но декодирование заканчивается, когда относительно конечного символа последовательности произведена F-проверка, то можно использовать те же рассуждения.

ровал» узел, так как всегда не исключена возможность, что последующие узлы упадут ниже данного порога. В дальнейшем мы подробнее обсудим этот момент.

На рис. 6.9.6 представлена типичная последовательность узлов вдоль правильного пути. Предположим, что $\Gamma_l \geqslant \Gamma_{12}$ для всех $l > 12$; тогда прорвавшимися узлами будут $\mathbf{u}_0$, $\mathbf{u}_1$, $\mathbf{u}_7$, $\mathbf{u}_9$, $\mathbf{u}_{10}$, $\mathbf{u}_{11}$ и $\mathbf{u}_{12}$. На рис. 6.9.6 также изображен путь порогов до прорыва в $\mathbf{u}_{12}$. Эти пороги полностью определяются пунктом (а) теоремы 6.9.1 через путь цен и окончательный порог в последовательности. Чтобы убедиться в этом, начнем с рассмотрения окончательного порога (T_{12} на рис. 6.9.6) и продвинемся по горизонтали влево. Согласно (6.9.8) пороги понижаются только, если необходимо избежать пересечения с траекторией пути; согласно (6.9.7) понижение не больше чем это необходимо для того, чтобы избежать пересечения.

Сложность последовательного декодирования

Число проверок, которое последовательный декодер должен произвести при декодировании последовательности принятых символов, является случайной величиной, зависящей от принятой последовательности, последовательности на выходе источника и сверточного кодера. Если принятые символы поступают в декодер с фиксированной по времени скоростью, и проверки проводятся с фиксированной скоростью, на входе декодера образуется очередь из принятых символов. Ясно, что изучение поведения этой очереди имеет первостепенное значение при рассмотрении работы последовательного декодера.

Чтобы лучше понять поведение этой очереди, удобно представить себе дерево принятых цен (соответствующее данной последовательности источника, данному кодеру и данной принятой последовательности) состоящим из правильного пути, соответствующего передаваемой последовательности источника, и множества деревьев неправильных путей, причем из каждого узла правильного пути исходят различные деревья. *Обозначим* через W_n число F-проверок, проведенных в n-м узле правильного пути и во всех узлах дерева неправильных путей, исходящих из n-го узла. Сумма по n величин W_n равна общему числу F-проверок, которые необходимо провести для декодирования последовательности; W_n можно интерпретировать как число F-проверок, которые необходимо провести для декодирования n-го подблока. Если n-й узел правильного дерева — прорвавшийся, то все F-проверки, входящие в W_0, ..., W_{n-1}, будут выполнены раньше любой F-проверки, входящей в W_n, W_{n-1}, ...

В дальнейшем нас будет интересовать исследование распределения W_n. Это позволит нам лучше понять статистическое поведение очереди; если вероятность того, что W_n примет очень большое значение не мала, то будет существовать тенденция к созданию больших очередей. Оправданием для рассмотрения только F-проверок и пренебрежения движениями вбок и назад могут служить следующие соображения. Каждый узел имеет 2^λ непосредственных потомков и на каждую F-проверку данного узла из каждого из этих потомков приходится

лишь одно движение вбок или назад. Поэтому при любой фиксированной длине l дерева общее число боковых и обратных движений на длине l не более чем в 2^{λ} раз превышает число движений вперед на длине $l-1$. Суммируя по l, получаем, что общее число проверок за некоторое фиксированное время не более чем в $2^{\lambda}+1$ раз превышает число F-проверок.

В свете доказательства теоремы кодирования не должно показаться неожиданным, что поведение случайной величины W_n легче исследовать для ансамбля кодов, чем для некоторого частного кода. В каждом из кодов ансамбля, который будет рассматриваться, кодовые последовательности получаются следующим образом: информационная последовательность подается в двоичный несистематический сверточный кодер, выходная последовательность этого кодера суммируется с произвольной двоичной последовательностью, а сумма затем преобразуется во входные символы канала. Точнее, двоичная последовательность сообщения разбивается на подблоки по λ символов в каждом, $\mathbf{u}_1^{(1)}$, ..., $\mathbf{u}_1^{(\lambda)}$, $\mathbf{u}_2^{(1)}$, ..., $\mathbf{u}_2^{(\lambda)}$, ... Каждым λ входным символам сверточного кодера соответствуют $a\nu$ выходных символов (для соответствующим образом выбранных a и ν), образующих двоичную последовательность $s_1^{(1)}$, ..., $s_1^{(a\nu)}$, $s_2^{(1)}$, ... Как и в соотношении (6.8.8) (и на рис. 6.8.3), выходные символы образуются из входных по правилу

$$s_n^{(i)} = \sum_{l=0}^{L-1} \sum_{j=1}^{\lambda} g_{j,\,i}(l)\, u_{n-l}^{(j)};\ 1 \leqslant i \leqslant a\nu, \tag{6.9.10}$$

где $g_{j,\,i}(l) = 0$ при $l < 0$.

Для простоты исследования будем считать L (длину кодового ограничения в блоке) бесконечной. Позднее, при изучении вероятности ошибки мы, естественно, будем рассматривать кодовые ограничения конечной длины. Выходная последовательность сверточного кодера $\mathbf{s}$ суммируется далее по модулю 2 с произвольной двоичной последовательностью $\mathbf{v}$, при этом образуется последовательность

$$s_1^{(1)} \oplus v_1^{(1)}, \ldots, s_1^{(a\nu)} \oplus v_1^{(a\nu)},\ s_2^{(1)} \oplus v_2^{(1)}, \ldots .$$

Эта последовательность затем разбивается на подблоки по a символов в каждом, и каждая двоичная последовательность из a символов отображается во входную букву канала с помощью отображения, аналогичного представленному на рис. 6.2.1; при этом k-я буква входного алфавита канала используется с относительной частотой $Q(k)$, определенной как и в (6.2.6). Так же как и в (6.2.6), используемое значение a определяется по значениям $Q(k)$. Можно заметить, что после выполнения этих этапов кодирования на каждые λ на двоичных символов источника будет приходиться ν символов канала, поэтому скорость кода равна $R = (\lambda/\nu)\ln 2$ нат на символ канала.

Пусть заданы значения λ, ν и a и отображение a-символьной двоичной последовательности в символы канала; рассмотрим ансамбль кодов, в котором значения всех элементов $g_{j,\,i}(l)$ и всех элементов последовательности $\mathbf{v}$ являются значениями независимых равновероятных двоичных символов.

Л е м м а 6.9.1. Для рассмотренного выше ансамбля кодов кодовая последовательность $x_1^{(1)}, \ldots, x_1^{(\nu)}, x_2^{(1)}, \ldots$, соответствующая какой-либо данной информационной последовательности, является случайной последовательностью с символами, статистически независящими один от другого и выбираемыми с вероятностями $Q(k)$. Более того, если две информационные последовательности $\mathbf{u}$ и $\mathbf{u}'$ совпадают в первых $b-1$ подблоках и отличаются в b-м подблоке, то соответствующие кодовые последовательности совпадают в первых $b-1$ подблоках и статистически независимы во всех последующих подблоках.

Доказательство. Если задана некоторая произвольная выходная последовательность двоичного сверточного кодера $\mathbf{s}$, то из независимости и равновероятности двоичных символов последовательности $\mathbf{v}$ вытекает, что $\mathbf{s} \oplus \mathbf{v}$ состоит из независимых равновероятных двоичных символов, поэтому любая кодовая последовательность состоит из независимых букв с вероятностями появления $Q(k)$. Предположим, что $\mathbf{u}$ и $\mathbf{u}'$ совпадают в первых $b-1$ подблоках и отличаются в b-м подблоке, и что $\mathbf{s}$ и $\mathbf{s}'$ — соответствующие им выходные последовательности двоичного сверточного кодера. Тогда $\mathbf{s}'' = \mathbf{s} \oplus \mathbf{s}'$ является выходной последовательностью сверточного кодера при входной последовательности $\mathbf{u}'' = \mathbf{u} \oplus \mathbf{u}'$; поскольку первые $b-1$ подблоков последовательности $\mathbf{u}''$ целиком состоят из нулей, то первые $b-1$ подблоков $\mathbf{s}''$ — также нулевые. Пусть теперь, по крайней мере при одном значении j, например j', величина $u_b''^{(j')} = 1$ и при любых $n \geqslant b$

$$s_n''^{(i)} = g_{j',\,i}(n-b) \oplus \sum g_{j,\,i}(n-b)\, u_b''^{(j)} \oplus \sum_{l=b+1}^{n} \sum_{f=1}^{\lambda} g_{j,\,i}(n-l)\, u_l''^{(j)}.$$

Так как $g_{j',\,i}(n-b)$ является двоичной случайной величиной с равновероятными значениями, не зависящей от всех остальных $g_{j,\,i}(l)$, то $s_n''^{(i)}$ — также двоичная случайная величина с равновероятными значениями. Точно так же величина $s_n''^{(i)}$ статистически не зависит от всех $s_m''^{(i)}$ при $m < n$ и всех $s_n''^{(i')}$ при $i' \neq i$, так как они не зависят от $g_{j',i}(n-b)$. Поэтому все подблоки последовательности $\mathbf{s}''$, кроме первых $b-1$, образуются независимыми и равновероятными двоичными символами. Следовательно, $\mathbf{s} \oplus \mathbf{v}$ и $\mathbf{s}' \oplus \mathbf{v}$ статистически независимы во всех подблоках, кроме первых $b-1$, и поэтому $\mathbf{x}$ и $\mathbf{x}'$ статистически независимы во всех подблоках, кроме первых $b-1$. |

Найдем теперь верхнюю границу $\overline{W}_0$ среднего числа F-проверок, выполняемых последовательным декодером в начальном узле и в дереве неправильных путей, исходящем из начального узла. Усреднение будет проводиться по ансамблю кодов, шумам канала и последовательностям сообщения. После этого найдем верхнюю границу для ρ-го момента $\overline{W}_0^{\rho}$ величины W_0 при всех $\rho \geqslant 1$. Те же границы, как мы увидим, справедливы для W_n в каждом узле правильного пути.

Случайные величины W_0 зависят как от цен узлов вдоль правильного пути, так и от цен узлов в неправильном поддереве, исходящем из начального узла. Число непосредственных потомков начального

узла, принадлежащих неправильному поддереву, равно $2^{\lambda}-1$ (оставшийся непосредственный потомок принадлежит правильному пути). Число узлов в данном неправильном поддереве, лежащих на глубине 2, равно $2^{\lambda}(2^{\lambda}-1)$, и вообще число узлов на глубине l в данном неправильном поддереве равно $2^{\lambda(l-1)}(2^{\lambda}-1)$. Обозначим через $m(l)$ m-й узел среди этих узлов, лежащих на глубине l, при некотором произвольном упорядочении; пусть $\Gamma'_{m(l)}$ является ценой этого узла. Обозначим через Γ_n цену n-го узла на правильном пути и через Γ_{min} — нижнюю грань Γ_n при $0 \leqslant n < \infty$.

Л е м м а 6.9.2. Узел $m(l)$ может проверяться в i-й раз, $i \geqslant 1$, лишь в том случае, когда $\Gamma'_{m(l)} \geqslant \Gamma_{min} - 2\Delta + i\Delta$.

Доказательство. Согласно пункту б) теоремы 6.9.1, при первой F-проверке узла $m(l)$ окончательный порог не может превысить $\Gamma'_{m(l)}$. Аналогично, окончательный порог при i-й F-проверке узла $m(l)$ не превышает $\Gamma'_{m(l)} - i\Delta + \Delta$. Лемма будет доказана, если показать, что порог никогда не может быть сделан ниже чем $\Gamma_{min} - \Delta$. Согласно пункту б) теоремы 6.9.1 порог при последовательных F-проверках в начальном узле уменьшается от 0 скачками, равными Δ. Согласно пункту в) теоремы в любом другом узле порог не может быть сделан ниже финального порога при последней F-проверке в начальном узле. Следовательно, впервые порог принимает значение, равное Γ_{min} или меньше, при F-проверке в начальном узле и этот порог больше чем $\Gamma_{min} - \Delta$. Так как траектория правильного пути проходит всюду не ниже этого порога, то начальный узел в дальнейшем никогда не будет проверен вновь и порог никогда не будет сделан ниже $\Gamma_{min} - \Delta$.

Определим теперь двоичную случайную величину $w[m(l), i]$ с помощью соотношения

$$w[m(l), i] = \begin{cases} 1, & \text{если } \Gamma'_{m(l)} \geqslant \Gamma_{min} - 2\Delta + i\Delta, \\ 0 & \text{в противном случае.} \end{cases} \tag{6.9.11}$$

Так как i-й раз F-проверка в $m(l)$-м узле может быть выполнена только в том случае, если $w[m(l), i] = 1$, то легко видеть, что число F-проверок в $m(l)$-м узле ограничено сверху суммой

$$\sum_{i=1}^{\infty} w[m(l), i].$$

Поэтому общее число F-проверок в исходном узле и всех узлах, которые принадлежат неправильному поддереву, исходящему из начального узла, ограничено неравенством

$$W_0 \leqslant \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} w[m(l), i]. \tag{6.9.12}$$

По условию, в сумму по $m(l)$ при $l = 0$ войдет только одно слагаемое, соответствующее исходному узлу. При $l \geqslant 1$ в сумму по $m(l)$ войдут слагаемые, соответствующие $(2^{\lambda}-1)\,2^{\lambda(l-1)}$ узлам неправильного поддерева, лежащим на глубине l. Оценим теперь сверху математи-

ческое ожидание величины W_0. Так как математическое ожидание суммы равно сумме математических ожиданий, то

$$\overline{W}_0 \leqslant \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} \overline{w[m(l), i]}. \tag{6.9.13}$$

С другой стороны, согласно (6.9.11)

$$\overline{w[m(l), i]} = \Pr[\Gamma'_{m(l)} \geqslant \Gamma_{min} - 2\Delta + i\Delta]. \tag{6.9.14}$$

В силу статистической независимости передаваемой кодовой последовательности и кодовой последовательности, соответствующей $m(l)$, задача нахождения вероятности в правой части (6.9.14) аналогична задаче нахождения вероятности ошибки при передаче двух случайно выбираемых кодовых слов. Единственное различие заключается в том, что должны рассматриваться последовательности вдоль правильного пути различной длины. Принимая во внимание эту аналогию с задачами передачи двух кодовых слов, не следует удивляться, что в ответе появляется значение функции $E_0(\rho, \mathbf{Q})$ при $\rho = 1$:

$$E_0(1, \mathbf{Q}) = -\ln \sum_{j=0}^{J-1} \left[\sum_{k=0}^{K-1} Q(k) \sqrt{P(j|k)} \right]^2. \tag{6.9.15}$$

Следующая лемма доказана в приложении 6Б.

Лемма 6.9.3. Рассмотрим ансамбль кодов, в котором кодовая последовательность, соответствующая каждой последовательности сообщения, представляет собой последовательность статистически независимых символов, принимающих значения с вероятностями $Q(k)$. Обозначим через Γ_{min} нижнюю грань цен вдоль правильного пути в дереве принятых цен и через Γ'_l — цену в узле $\mathbf{u}'_l$, где $\mathbf{u}'_l$ — последовательность сообщения из l подблоков, такая, что первые l подблоков кодовой последовательности, соответствующей $\mathbf{u}'_l$, статистически независимы от переданной кодовой последовательности. Тогда, если смещение B удовлетворяет неравенству $B \leqslant E_0(1, \mathbf{Q})$, то

$$\Pr[\Gamma'_l \geqslant \Gamma_{min} + (i-2)\Delta] \leqslant (l+1) \times \\ \times \exp\left[-\frac{(i-2)\Delta}{2} - \nu l \frac{E_0(1, \mathbf{Q}) + B}{2} \right]. \tag{6.9.16}$$

Сравнивая (6.9.13), (6.9.14) и (6.9.16), получаем

$$\overline{W}_0 \leqslant \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} (l+1) \exp\left[-\frac{(i-2)\Delta}{2} - \nu l \frac{E_0(1, \mathbf{Q}) + B}{2} \right]. \tag{6.9.17}$$

Из (6.9.1) видно, что $2^{\lambda} = e^{\nu R}$ и поэтому приведенная выше сумма по $m(l)$ содержит менее чем $e^{\nu l R}$ одинаковых членов; следовательно,

$$\overline{W}_0 \leqslant \sum_{l=0}^{\infty} \sum_{i=1}^{\infty} (l+1) \exp\left\{ -\frac{(i-2)\Delta}{2} - \nu l \left[\frac{E_0(1, \mathbf{Q}) + B}{2} - R \right] \right\}. \tag{6.9.18}$$

Эти ряды могут быть без труда просуммированы, если использовать разложение

$$\frac{z}{1-z}=\sum_{i=1}^{\infty} z^i; \qquad z<1, \tag{6.9.19}$$

и его производную

$$\frac{1}{(1-z)^2}=\sum_{l=0}^{\infty}(l+1)z^l; \qquad z<1. \tag{6.9.20}$$

Эти ряды сходятся, если

$$R<\frac{E_0(1, \mathbf{Q})+B}{2}; \tag{6.9.21}$$

при этом получим

$$\overline{W}_0 \leqslant \frac{e^{\Delta/2}}{1-e^{-\Delta/2}}\left\{1-\exp\left[-\nu\frac{E_0(1, \mathbf{Q})+B}{2}+\nu B\right]\right\}^{-2}. \tag{6.9.22}$$

Приведенная граница дает некоторое представление о том, какие значения должны выбираться для смещения B и порогового интервала Δ. Граница убывает с возрастанием B, но справедлива лишь при $B \leqslant (1, \mathbf{Q})$. Поэтому граница имеет минимум при $B = E_0(1, \mathbf{Q})$. Аналогично, минимизируя по Δ, находим, что минимум достигается при $e^{\Delta/2} = 2$ или $\Delta = 2 \ln 2$. Используя эти значения, убеждаемся, что при $R < E_0(1, \mathbf{Q})$

$$\overline{W}_0 \leqslant 4\{-\exp[-\nu E_0(1, \mathbf{Q})+\nu R]\}^{-2}. \tag{6.9.23}$$

Эти рассуждения оставляют некоторое сомнение в том, что минимум $\overline{W}_0$ достигается при этих значениях B и Δ, поскольку они минимизируют лишь границу для $\overline{W}_0$. Для слагаемых, соответствующих малым значениям i и l, которые доминируют в сумме (6.9.13), использованная нами экспоненциальная граница является довольно грубой. Блюстейн и Жордан (1963) экспериментально показали, что, как правило, величина $\overline{W}_0$ на два порядка меньше, чем приведенная здесь граница, и что она мало зависит от смещения и от порогового интервала. Важное значение неравенства (6.9.23) состоит не в том, что оно задает явную границу, а в доказательстве того, что граница конечна при $R < E_0(1, \mathbf{Q})$.

Рассмотрим теперь произвольный узел правильного пути, например n-й узел. Заметим, что статистическое описание неправильного поддерева и правильного пути, исходящих из этого узла, совершенно аналогично статистическому описанию неправильного поддерева и правильного пути, исходящих из начального узла. Единственная разница состоит в том, что к множеству цен узлов добавлена цена Γ_n. Лемма применяется здесь точно так же, как и раньше, и (6.9.23) позволяет получить верхнюю границу для $\overline{W}_n$. Поэтому (6.9.23) оценивает сверху среднее число F-проверок на декодированный подблок.

Из этой границы можно также увидеть довольно ясно, почему «работает» метод последовательного декодирования. Число узлов в неправильном поддереве растет экспоненциально с глубиной дерева, а вероятность проверки узла является экспоненциально убывающей функцией глубины дерева. До тех пор пока скорость остается меньше, чем $E_0(1, \mathbf{Q})$, убывание вероятности более чем компенсирует возрастание числа узлов. Ясно, что именно древовидная структура сверточных кодов делает возможным такой обмен.

Теперь найдем верхнюю границу $\overline{W_0^\rho}$ при $\rho \geqslant 1$. Из (6.9.12) имеем

$$\overline{W_0^\rho} \leqslant \overline{\left\{ \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} w[m(l), i] \right\}^\rho}. \tag{6.9.24}$$

Применяя в правой части (6.9.24) неравенство Минковского (см. задачу 4.15 (*e*)), получаем

$$\left[\overline{W_0^\rho}\right]^{1/\rho} \leqslant \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} \left[\overline{w[m(l), i]^\rho}\right]^{1/\rho}. \tag{6.9.25}$$

Так как $w[m(l), i]$ принимает лишь значения 0 и 1, то $w[m(l), i]^\rho = w[m(l), i]$ и из (6.9.14) и (6.9.16) имеем

$$\overline{w[m(l), i]^\rho} \leqslant (l+1) \exp\left[-\frac{(i-2)\Delta}{2} - \nu l \frac{E_0(1, \mathbf{Q}) + B}{2} \right],$$

$$\left[\overline{W_n^\rho}\right]^{1/\rho} \leqslant \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} (l+1)^{1/\rho} \exp\left[-\frac{(i-2)\Delta}{2\rho} - \nu l \frac{E_0(1, \mathbf{Q}) + B}{2\rho} \right]. \tag{6.9.26}$$

Оценивая сверху $(l+1)^{1/\rho}$ величиной $l+1$ и суммируя эти ряды таким же образом, как и ряды в (6.9.17), убеждаемся, что суммы сходятся, если

$$R < \frac{E_0(1, \mathbf{Q}) + B}{2\rho} \tag{6.9.27}$$

и

$$\left[\overline{W_0^\rho}\right]^{1/\rho} \leqslant \frac{\exp[\Delta/(2\rho)]}{1 - \exp[-\Delta/(2\rho)]} \left\{ 1 - \exp\left[-\nu \frac{E_0(1, \mathbf{Q}) + B}{2\rho} + \nu R \right] \right\}^{-2} \tag{6.9.28}$$

К $[\overline{W_n^\rho}]^{1/\rho}$ вновь применима та же самая граница. Если смещение B равно $E_0(1, \mathbf{Q})$, то $\overline{W_n^\rho}$, как нетрудно убедиться, конечно при R, меньших, чем $E_0(1, \mathbf{Q})/\rho$. Границу (6.9.28) можно использовать для оценки функции распределения W_n. Применяя обобщенное неравенство Чебышева, имеем

$$\Pr[W_n \geqslant i] \leqslant i^{-\rho} \overline{W_n^\rho}. \tag{6.9.29}$$

Наши результаты можно сформулировать в виде следующей теоремы.

Теорема 6.9.2. При применении последовательного декодирования в дискретном канале без памяти среднее число F-проверок, приходя-

щихся на декодированный подблок, удовлетворяет неравенству

$$\overline{W}_n \leqslant 4\{1-\exp[-\nu E_0(1, \mathbf{Q})+\nu R]\}^{-2} \qquad (6.9.30)$$

при $R < E_0(1, \mathbf{Q})$.

Здесь ν означает число символов канала, приходящихся на подблок, $R = (\lambda/\nu)\ln 2$ является скоростью кода в натах на символ канала, $E_0(1, \mathbf{Q})$ определяется соотношением (6.9.15). Смещение B выбирается равным $E_0(1, \mathbf{Q})$, а пороговый интервал — равным $2 \ln 2$. Усреднение проводится по ансамблю кодов бесконечной длины, определенных на стр. 292. Кроме того, при любом $\rho \geqslant 1$, для которого $R < E_0(1, \mathbf{Q})/\rho$, величина $\overline{W_n^\rho}$ также конечна.

Севэдж (1966) получил более сильные оценки для $\overline{W_n^\rho}$ с помощью рассмотрения большего ансамбля кодов. Он рассмотрел ансамбль случайно выбираемых древовидных кодов, в котором кодовые последовательности имеют такую же древовидную структуру, как и на рис. 6.9.2, но где каждый символ в дереве выбирается независимо с вероятностями $Q(k)$, $0 \leqslant k \leqslant K-1$. Севэдж показал, что в этом ансамбле при целых ρ величина $\overline{W_0^\rho}$ конечна, если $R < E_0(\rho, \mathbf{Q})/\rho$. Отсюда не обязательно следует, что для этих скоростей существуют сверточные коды, для которых $\overline{W_0^\rho}$ конечно. Однако довольно правдоподобно, что при всех положительных ρ и всех $R < E_0(\rho, \mathbf{Q})/\rho$ величина $\overline{W_n^\rho}$ конечна, где усреднение производится по ансамблю сверточных кодов с бесконечной длиной кодового ограничения и скоростью R. Это предположение было доказано Фэлконером (1966) для случая $0 < \rho \leqslant 1$.

Вопрос о том, что произойдет с $\overline{W_n^\rho}$ при $R > E_0(\rho, \mathbf{Q})/\rho$, детально исследован Джекобсом и Берлекэмпом (1967). Их результаты применимы к произвольным древовидным кодам (т. е. кодам, структура которых описывается рис. 6.9.2) и к классу последовательных алгоритмов, включающему описанный здесь алгоритм Фано. Пусть $E_0(\rho) = \max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})$ и пусть $\hat{E}_0(\rho)$ выпуклая $\frown$ оболочка $E_0(\rho)$, т. е. $\hat{E}_0(\rho)$ — минимальная выпуклая $\frown$ функция, равная или бо́льшая чем $E_0(\rho)$. Она образуется заменой всех невыпуклых $\frown$ участков $E_0(\rho)$ прямолинейными сегментами. Джекобс и Берлекэмп показали, что в наших обозначениях для любого древовидного кода со скоростью $R > \hat{E}_0(\rho)/\rho$

$$\lim_{N\to\infty} \overline{\left(\frac{1}{N}\sum_{n=0}^{N-1} W_n\right)^\rho} = \infty. \qquad (6.9.31)$$

Бесконечность моментов W_n высокого порядка указывает на то, что длинные очереди являются трудной проблемой в последовательном декодировании. Джекобс и Берлекэмп показали, что вероятность того, что либо очередь при декодировании символа превысит данную длину i, либо будет сделана ошибка, не может убывать с ростом i быстрее, чем $i^{-\rho}$, для скорости кода, равной $\hat{E}_0(\rho)/\rho$.

Вероятность ошибки при последовательном декодировании

Чтобы оценить вероятность ошибки при последовательном декодировании, следует рассмотреть коды с конечной длиной кодового ограничения. Если для кодов с бесконечной длиной кодового ограничения $\overline{W}_n$ конечно при любом n, то из (6.9.18) следует, что вероятность проверки узла на глубине l дерева неправильных путей стремится к нулю с возрастанием l и вероятность ошибки, очевидно, равна нулю.

Предположим, что длина ограничения кодера равна L подблокам. Предположим также, что в кодер вместо бесконечной последовательности подблоков поступает от источника конечное число подблоков L_T, где L_T, как правило, гораздо больше L. После того как в кодер поступит L_T подблоков сообщения, в него подается известная последовательность $L - 1$ подблоков, состоящих из нулей. В такой ситуации каждый узел дерева принятых цен, лежащий на глубине не больше $L_T - 1$, имеет 2^λ непосредственных потомков, а каждый узел на глубине, не меньшей L_T, только одного непосредственного потомка. Декодирование проводится согласно правилам, указанным на рис. 6.9.4, за исключением того, что декодер распознает, что узел порядка L_T или больше имеет лишь одного непосредственного потомка и поэтому правила 3 и 5 всегда приводят к движению назад для узлов порядка, большего чем L_T. Декодирование заканчивается при первой проверке в узле порядка $L_T + L - 1$ и гипотетическая последовательность источника при этой проверке выбирается в качестве декодированной последовательности. При конечных L и L_T декодирование в конце концов должно закончиться.

Из рассмотрения ансамбля кодов, введенных после соотношения (6.9.10), с конечным L, следует, что две кодовые последовательности, соответствующие двум заданным последовательностям источника, статистически независимы в первых L подблоках, исходящих из первого подблока, в котором они отличаются, но зависимы в остальных. Вероятность ошибки декодирования для этого ансамбля может быть ограничена сверху величиной, по существу, эквивалентной границе случайного кодирования для блоковых кодов (см. задачу 6.42). Однако интереснее здесь слегка модифицировать ансамбль и затем вывести гораздо меньшую верхнюю границу для вероятности ошибки. Для этого рассмотрим двоичный сверточный кодер, изменяющийся во времени; двоичная выходная последовательность кодера образуется по правилу [сравнить с (6.9.10)]:

$$s_n^{(i)} = \sum_{l=n-L+1}^{n} \sum_{j=1}^{\lambda} g_{j,i}(n, l)\, u_l^{(j)}; \quad 1 \leqslant i \leqslant a\nu. \tag{6.9.32}$$

В данном ансамбле каждый элемент $g_{j,i}(n, l)$ выбирается так, что он является статистически независимой от других величин двоичной случайной величиной, принимающей оба значения с одинаковыми вероятностями. В остальном ансамбль аналогичен прежнему; $\mathbf{s}$ суммируется со случайной двоичной последовательностью $\mathbf{v}$, после чего сумма отображается во входные буквы канала. В обозначениях

рис. 6.8.3 это соответствует случайному перемешиванию связей между λ регистрами сдвига и ν сумматорами по модулю 2, порождающими двоичную выходную последовательность; перемешивание производится после поступления в кодер каждого из подблоков символов сообщения.

Л е м м а 6.9.4. *В рассмотренном выше ансамбле кодов кодовая последовательность, соответствующая какой-либо данной информационной последовательности, представляет собой последовательность статистически независимых букв, выбираемых с вероятностями, равными $Q(k)$, где $Q(k)$ — относительная частота появления k-й входной буквы при отображении двоичной последовательности длины a во входную букву канала. Пусть далее кодовые последовательности* $\mathbf{x}$ *и* $\mathbf{x}'$ *соответствуют последовательности сообщений* $\mathbf{u}$ *и* $\mathbf{u}'$; *тогда* $\mathbf{x}$ *и* $\mathbf{x}'$ *совпадают в некотором подблоке, если* $\mathbf{u}$ *и* $\mathbf{u}'$ *совпадают в этом подблоке и $L-1$ предыдущих. Во всех остальных подблоках последовательности* $\mathbf{x}$ *и* $\mathbf{x}'$ *статистически независимы.*

Доказательство этой леммы почти идентично доказательству леммы 6.9.1 и потому опускается.

Утверждение леммы в обозначениях рис. 6.8.3 означает, что $\mathbf{x}$ и $\mathbf{x}'$ совпадают в тех подблоках, которые соответствуют одним и тем же символам в регистре сдвига. В других случаях $\mathbf{x}$ и $\mathbf{x}'$ независимы.

Пусть задан код из рассмотренного выше ансамбля. Обозначим через $\mathbf{u}$ переданную последовательность сообщения и через $\mathbf{u}'$ — декодированную методом последовательного декодирования последовательность сообщения. Как правило, ошибки декодирования, если они происходят, имеют тенденцию группироваться в пакеты. Точнее, *пакетом ошибок декодирования называется последовательность одного и более подблоков, обладающих следующими свойствами: первый и последний подблоки последовательности содержат ошибки (т. е.* $\mathbf{u}$ *и* $\mathbf{u}'$ *отличаются в этих подблоках); никакие $L-1$ последовательных подблоков внутри этой последовательности не являются одновременно безошибочными; по $L-1$ подблоков с каждой стороны последовательности не содержат ошибки.* Это определение однозначно определяет множество одного или более пакетов ошибок для любого $\mathbf{u}' \neq \mathbf{u}$.

Обозначим через $P_e(b, c)$ вероятность (в ансамбле кодов, сообщений и шумов в канале) того, что с началом в b-м подблоке и концом в c-м подблоке имел место пакет ошибок. Обозначим через $P_{e,n}$ вероятность ошибочного декодирования n-го подблока. Согласно нашему определению пакета любой подблок, ошибочно декодированный, должен входить в некоторый пакет ошибок декодирования, поэтому

$$P_{e,n} \leqslant \sum_{b=1}^{n} \sum_{c=n}^{L_T} P_e(b, c). \tag{6.9.33}$$

Чтобы найти верхнюю границу $P_e(b, c)$, рассмотрим сначала простейший случай, т. е. случай $b = 1$. Пусть $\mathbf{u}$ — переданная информационная последовательность и пусть $\mathbf{u}'_{c+L-1}$ — первые $c + L - 1$ подблоков декодированной последовательности. Для того чтобы в промежутке от 1-го до c-го подблока появился пакет ошибок декодиро-

вания, должно быть $u_1' \neq u_1$, $u_c' \neq u_c$ и $u_{c+i}' = u_{c+i}$, $1 \leqslant i \leqslant L-1$. Так как каждый внутренний подблок может быть выбран не более чем 2^{λ} способами, то последовательность $\mathbf{u}'_{c+L-1}$, содержащая пакет ошибок декодирования в промежутке от 1-го до c-го подблока, может быть выбрана менее чем $2^{\lambda c}$, способами.

Обозначим через Γ_l, $0 \leqslant l \leqslant L_T + L - 1$ путь цен вдоль правильного пути в дереве полученных цен, и пусть

$$\Gamma_{min} = \min_{0 \leqslant l \leqslant L_T + L - 1} \Gamma_l.$$

Пусть далее Γ'_{c+L-1} является ценой узла $\mathbf{u}'_{c+L-1}$, где $\mathbf{u}'_{c+L-1}$ выбрано таким образом, что он соответствует пакету ошибок декодирования в промежутке от 1-го до c-го подблока. Совершенно ясно, что проведение в узле $\mathbf{u}'_{c+L-1}$ F-проверки необходимо для того, чтобы $\mathbf{u}'_{c+L-1}$ было декодировано. Вместе с тем согласно лемме 6.9.2 F-проверка $\mathbf{u}'_{c+L-1}$ может наступить только тогда, когда $\Gamma'_{c+L-1} \geqslant \geqslant \Gamma_{\min} - \Delta$. Поэтому

$$\Pr\,[\text{декодирования } \mathbf{u}'_{c+L+1}] \leqslant \Pr\,[\Gamma'_{c+L-1} \geqslant \Gamma_{min} - \Delta]. \quad (6.9.34)$$

Согласно лемме 6.9.4 $c + L - 1$ подблоков кодовой последовательности, соответствующей последовательности $\mathbf{u}'_{c+L-1}$, статистически не зависят от переданной последовательности и поэтому из леммы 6.9.3 вытекает

$$\Pr\,[\text{декодирования } \mathbf{u}'_{c+L-1}] \leqslant (c+L) \times \times \exp\left[\frac{\Delta}{2} - \nu(c+L-1)\,\frac{E_0(1, \mathbf{Q}) + B}{2}\right] \quad (6.9.35)$$

при $B \leqslant E_0(1, \mathbf{Q})$. Так как число способов выбора $\mathbf{u}'_{c+L-1}$, соответствующих пакету ошибок в промежутке от 1-го до c-го подблока, меньше чем $2^{\lambda c}$, то получим

$$P_e(1, c) \leqslant 2^{\lambda c}(c+L) \exp\left[\frac{\Delta}{2} - \nu(c+L-1)\,\frac{E_0(1, \mathbf{Q}) + B}{2}\right]. \quad (6.9.36)$$

Теперь найдем эквивалентную границу для $P_e(b, c)$ при произвольных $b \geqslant 1$. Рассмотрим некоторый конкретный код ансамбля; обозначим через $\mathbf{u}$ переданное сообщение, а через $\mathbf{x}$ — переданную кодовую последовательность. Пусть $\mathbf{u}'_{b-1}$ первые $b-1$ подблоков произвольной последовательности сообщения, удовлетворяющей условию $u'_{b-i} = u_{b-i}$ при $1 \leqslant i \leqslant \min(b-1, L-1)$. Рассмотрим потомков $\mathbf{u}'_{b-1}$ в дереве принятых цен. Один из путей потомков, выходящий из $\mathbf{u}'_{b-1}$, соответствует подблокам переданного сообщения u_b, u_{b+1}, ... Кодовая последовательность, соответствующая $\mathbf{u}'_{b-1}$, u_b, u_{b+1}, ..., совпадает в b-м, $(b+1)$-м, ... подблоках с кодовой последовательностью, соответствующей переданной последовательности сообщения $\mathbf{u}$, так как в регистре сдвига, порождающем кодовые последовательности, в обоих случаях будут храниться одни и те же символы. Поэтому изменение вдоль пути цен от узла $\mathbf{u}'_{b-1}$ до потомка $\mathbf{u}'_{b-1}$, u_b, ..., u_l совпадают с $\Gamma_l - \Gamma_{b-1}$ (изменением вдоль пути цен от $\mathbf{u}'_{b-1}$ до $\mathbf{u}_l$).

Пусть $\mathbf{u}'_{c+L-1}$ — какой-либо потомок $\mathbf{u}'_{b-1}$, лежащий на глубине в $c+L-1$ подблоков, и пусть $\Gamma'_{c+L-1}-\Gamma'_{b-1}$ — изменение цены при переходе в дереве принятых цен от узла $\mathbf{u}'_{b-1}$ к $\mathbf{u}'_{c+L-1}$. Чтобы декодировать узел $\mathbf{u}'_{c+L-1}$, необходимо произвести в нем, по крайней мере, F-проверку, а для того чтобы провести в этом узле F-проверку, необходимо, чтобы

$$\Gamma'_{c+L-1}-\Gamma'_{b-1}\geqslant \min_{b-1\leqslant l\leqslant L_T+L-1}(\Gamma_l-\Gamma_{b-1})-\Delta. \qquad (6.9.37)$$

Справедливость этого следует из того, что в противном случае последовательность $\mathbf{u}'_{b-1}$, u_b, u_{b+1}, ... не даст возможности понизить порог настолько, чтобы была возможна F-проверка $\mathbf{u}'_{c+L-1}$. Если декодирование $\mathbf{u}'_{c+L-1}$ приводит к пакету ошибок в промежутке от b-го до c-го подблока, то необходимо, чтобы $u'_c\neq u_c$ и $u'_{c+i}=u_{c+i}$ при $1\leqslant i\leqslant L-1$. Поэтому последовательность u'_b, ..., u'_{c+L-1}, содержащая пакет ошибок декодирования в промежутке от b-го до c-го подблока, может быть выбрана менее чем $2^{\lambda(c-b+1)}$ способами. Наконец, выполнение неравенства (6.9.37) для некоторой выбранной подпоследовательности зависит лишь от кодовых последовательностей, начинающихся с b-го подблока, и не зависит от выбора $\mathbf{u}'_{b-1}$ [удовлетворяющих ограничению, состоящему в том, что $u'_{b-i}=u_{b-i}$ для $1\leqslant i\leqslant \min(b-1, L-1)$]. Поэтому вероятность появления пакета ошибок декодирования в промежутке от b-го до c-го подблока ограничена сверху вероятностью того, что неравенство (6.9.37) не выполняется при всех рассмотренных выше способах выбора u'_b, ..., u'_{c+L-1}; число этих способов меньше чем $2^{\lambda(c-b+1)}$. Так как эта вероятность не зависит от $\mathbf{u}'_{b-1}$, удобно выбрать $\mathbf{u}'_{b-1}=\mathbf{u}_{b-1}$.

В ансамбле кодов подблоки от b-го до $(c+L-1)$-го кодовой последовательности, соответствующей какому-либо из рассмотренных выше узлов $\mathbf{u}'_{c+L-1}$, статистически не зависят от подблоков от b-го до (L_T+L-1)-го кодовой последовательности, которая соответствует последовательности $\mathbf{u}$ (см. лемму 6.9.4). Поэтому согласно лемме 6.9.3 имеем при $B\leqslant E_0(1, \mathbf{Q})$

$$\Pr\{[\Gamma'_{c+L-1}-\Gamma'_{b-1}]\leqslant \min_{b-1\leqslant l\leqslant L_T-L-1}[\Gamma_l-\Gamma_{b-1}]-\Delta\}\leqslant (c+L-b+1)\times$$

$$\times\exp\left[\frac{\Delta}{2}-\nu(c+L-b)\frac{E_0(1, \mathbf{Q})+B}{2}\right], \qquad (6.9.38)$$

$$P_e(b, c)\leqslant 2^{\lambda(c-b+1)}(c+L-b+1)\times$$

$$\times\exp\left[\frac{\Delta}{2}-\nu(c+L-b)\frac{E_0(1, \mathbf{Q})+B}{2}\right]. \qquad (6.9.39)$$

Подставляя (6.9.39) в (6.9.33) и суммируя по b и c, получаем границу вероятности ошибки на подблок. Далее можно получить верхнюю границу, суммируя по b от $-\infty$ до n и по c от n до ∞; эти суммы можно вычислить согласно формуле суммы геометрической прогрессии

(6.9.19) и ее производной (6.9.20). В результате при $z < 1$ [см. (6.9.42)] имеем

$$P_{e,n} \leqslant A \exp\left[-\nu L \frac{E_0(1, \mathbf{Q}) + B}{2}\right], \tag{6.9.40}$$

$$A = 2^{\lambda} e^{\Delta/2}\left[\frac{L}{(1-z)^2} + \frac{1+z}{(1-z)^3}\right], \tag{6.9.41}$$

$$z = 2^{\lambda} \exp\left[-\nu \frac{E_0(1, \mathbf{Q}) + B}{2}\right]. \tag{6.9.42}$$

Заметим, что минимум границы (6.9.40) по $B \leqslant E_0(1, \mathbf{Q})$ достигается при $B = E_0(1, \mathbf{Q})$. Обозначим также через $R = (\lambda/\nu) \ln 2$ скорость источника в натуральных единицах на символ канала для первых

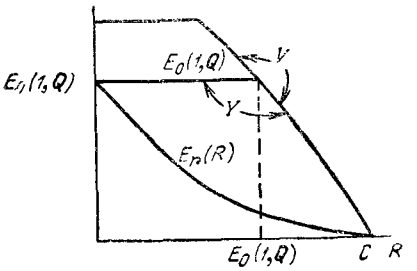


Рис. 6.9.7. Сравнение показателя экспоненты последовательного декодирования $E_0(1, \mathbf{Q})$ с показателем экспоненты блокового случайного кодирования; кривая Y является графиком показателя экспоненты в верхней границе Юдкина для P_e при последовательном кодировании; кривая V — график показателя экспоненты в нижней границе Витерби для P_e сверточных кодов.

L_T подблоков. Фактическая скорость несколько меньше и отличается множителем $L_T/(L_T+L-1)$ в силу наличия концевой последовательности в полном блоке. Подставляя $R = (\lambda/\nu) \ln 2$ и смещение $B = E_0 \times (1, \mathbf{Q})$ в (6.9.40) — (6.9.42), получаем

$$P_{e,n} \leqslant A \exp\left[-\nu L E_0(1, \mathbf{Q})\right], \tag{6.9.43}$$

$$A = \left[\frac{L}{(1-z)^2} + \frac{1+z}{(1-z)^3}\right] \exp\left[\nu R + \frac{\Delta}{2}\right], \tag{6.9.44}$$

$$z = \exp\left\{-\nu\left[E_0(1, \mathbf{Q}) - R\right]\right\}. \tag{6.9.45}$$

Эта граница справедлива при $R < E_0(1, \mathbf{Q})$, что совпадает с ограничением, необходимым для того, чтобы $\overline{W}_n$ было бы конечно.

Заметим, что вероятность ошибочного декодирования подблока в (6.9.43) убывает экспоненциально с ростом длины кодового ограничения νL, причем показатель экспоненты равен $E_0(1, \mathbf{Q})$. То что этот показатель не зависит от скорости при $R < E_0(1, \mathbf{Q})$, не является свидетельством слабости границы. Согласно (6.9.39) он равен показателю в экспоненте вероятности появления пакета ошибок декодирования длины l. Так как число последовательностей, отличающихся от $\mathbf{u}$ только в данном подблоке, равно $2^{\lambda} - 1$, то эта экспонента не зависит от R.

На рис. 6.9.7 показатель экспоненты $E_0(1, \mathbf{Q})$ сравнивается с показателем экспоненты случайного кодирования $E_r(R)$ для блоковых кодов. Представляется удивительным, что показатель экспоненты для последовательного декодирования больше чем $E_r(R)$. Отчасти это

объясняется тем, что ограничения на блок данных в сверточном коде простираются вне данного блока, а отчасти тем, что декодеру разрешается проводить поиск во всем блоке $(L + L_T - 1)\nu$ принятых символов, прежде чем он примет какое-либо решение.

Юдкин (1964) с помощью более тонких методов, чем применявшиеся при выводе (6.9.43), получил верхнюю границу вероятности ошибки для последовательного декодирования при всех скоростях вплоть до пропускной способности, и его экспонента также представлена на рис. 6.9.7. Витерби (1967) недавно нашел нижнюю границу минимальной вероятности ошибки, которая может быть достигнута для сверточных кодов, и график соответствующего показателя экспоненты также изображен на рис. 6.9.7. Заметим, что последовательное декодирование имеет вероятность ошибки с наилучшей возможной экспонентой при $R \geqslant E_0(1, \mathbf{Q})$ и фактически наилучшую экспоненту при R, несколько меньших $E_0(1, \mathbf{Q})$, когда $\overline{W_n}$ также конечно.

Граница вероятности ошибки (6.9.40) не зависит от общей длины L_T и поэтому нет необходимости периодически вставлять концевую последовательность при последовательном декодировании. Вместе с тем при практическом применении, особенно если L больше или равно 20, желательно выбирать L_T конечным; это позволяет декодеру в тех редких случаях, когда поиск очень затягивается, считать принятый блок стертым и переходить к следующему блоку. Вероятность такого стирания, конечно, тесно связана с функцией распределения W_n.

6.10. КОДИРОВАНИЕ В КАНАЛАХ С ПАКЕТАМИ ОШИБОК

В предыдущих параграфах в основном рассматривались методы кодирования в каналах без памяти. В этом параграфе рассматриваются каналы с двоичными входными и выходным алфавитами, в которых ошибки передачи имеют тенденцию к объединению в пакеты. Большинство двоичных систем связи (за исключением космических каналов) ведет себя именно таким образом. Трудно найти вероятностные модели этих каналов, которые подходили бы для целей изучения в них кодирования. Недостаточно найти модель, которая описывала бы типичное поведение канала, поскольку при отклонении поведения канала от типичного любой разумный метод кодирования приведет к ошибке декодирования. Причинами нетипичного поведения канала служат различные редкие события, которые по своей природе затрудняют вероятностное моделирование. По этой причине не будем исследовать вероятность ошибки для различных методов кодирования в этих каналах, а найдем другие меры их характеристики.

Наиболее естественным методом кодирования, используемым в каналах с пакетами ошибок, является обнаружение ошибок и переспрос. В этом случае данные на выходе источника кодируются кодом с проверкой на четность (лучше всего, в силу простоты реализации, использовать циклический код). Приемник вычисляет для принятой последовательности синдром $\mathbf{S}$ и если $\mathbf{S} = 0$, то считается, что информационные символы приняты без ошибок. Если $\mathbf{S} \neq 0$, приемник запрашивает, чтобы передатчик повторил данное кодовое слово. При этом ошиб-

ка декодирования произойдет только в том случае, если в передаче произойдут ошибки и последовательность ошибок совпадает с одним из кодовых слов. Для (N, L)-кода с проверкой на четность кодовыми словами являются лишь 2^L последовательностей из общего числа 2^N последовательностей длины N или лишь одна из каждых 2^{N-L} последовательностей. Отсюда видно, что ошибками декодирования можно пренебречь уже при не слишком больших значениях $N - L$ и что число ошибок декодирования слабо зависит от конкретной статистики шумов. Кодер и декодер для такого рода циклического кода могут быть реализованы при помощи регистров сдвига с $(N - L)$ разрядами (см. рис. 6.5.5).

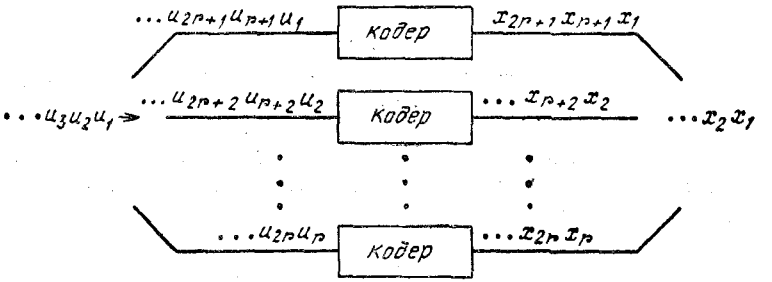


Рис. 6.10.1. Кодирование с перемежением.

Наряду с простотой этот метод имеет несколько недостатков. Во-первых, необходимо иметь надежную линию обратной связи от приемника к передатчику для передачи запросов о повторении передачи. Во-вторых, возникают проблемы хранения данных как в передатчике, так и в приемнике в моменты, когда необходимо повторение передачи. В-третьих, если канал сильно зашумлен, принимается лишь малое число кодовых слов. Существенны или нет первые два недостатка — это зависит от наличия линии обратной связи и от природы источника (т. е. от того, порождает ли источник данные по запросу передатчика, или он порождает их с постоянной по времени скоростью). Если существенным является третий недостаток, то описанная выше схема повторной передачи может быть дополнена каким-либо методом исправления ошибок, рассмотренным ниже. Следует отметить здесь, что если модулятор и демодулятор цифровых данных предназначены для эффективной передачи большого числа двоичных символов в секунду, то в передаче будут происходить частые ошибки, исправление которых совершенно необходимо вне зависимости от того, используется повторение передачи или нет.

Другой простой метод достижения надежной передачи в канале с пакетами ошибок состоит в перемежении или перемешивании символов. В принципе этот метод сводится к тому, что поступающий поток двоичных данных разбивается на фиксированное число, допустим на r, потоков данных, как показано на рис. 6.10.1. После этого каждый из r потоков данных кодируется отдельно, а закодированные последо-

вательности объединяются для передачи по каналу. На выходе канала принятый поток данных вновь разделяется на r потоков, каждый поток декодируется отдельно и, наконец, декодированные данные вновь объединяются.

Идея рассмотренного метода состоит в том, что следующие друг за другом буквы в любом кодовом слове отделены в канале промежутком в r единиц времени. Поэтому, если влияние памяти в канале исчезает с увеличением времени разделения символов, шумы в канале, действующие на следующие друг за другом буквы кодового слова, практически становятся независимыми при достаточно больших r. Таким образом, в канале с пакетами ошибок можно использовать в сочетании с перемежением символов любой из методов кодирования, ранее рассмотренных применительно к каналам без памяти.

Эти рассуждения показывают, что наличие памяти в канале не уменьшает его пропускной способности. Чтобы сделать это утверждение более точным, предположим, что в данном дискретном канале с памятью можно задать переходные вероятности $P(y \mid x)$ отдельных букв*). Тогда, если память в канале уменьшается достаточно быстро по времени, то любой метод кодирования, который приводит к данной вероятности ошибки в дискретном канале без памяти с переходными вероятностями $P(y \mid x)$, будет иметь практически ту же самую вероятность ошибки в данном канале с памятью, если только использовать перемежение с достаточно большим r. Поэтому канал с памятью имеет пропускную способность, по меньшей мере равную пропускной способности соответствующего канала без памяти. Мы не формулируем этот результат в виде теоремы, поскольку точно указать, как быстро память должна убывать со временем, довольно трудно.

Для реализации перемежения символов не всегда необходимо использовать r отдельных кодеров и декодеров. Например, если кодеры, представленные на рис. 6.10.1, являются циклическими (N, L) кодерами, каждый с производящим многочленом $g(D)$, то изображенные на рис. 6.10.1 перемежение и кодирование могут быть произведены с помощью (Nr, Lr) циклического кодера с производящим многочленом $g(D^r)$. Аналогично, если r нечетно и кодирование производится одинаковыми сверточными кодерами, каждый из которых рассчитан на скорость $1/2$ (в битах), как и на рис. 6.8.1, то перемежение и кодирование могут быть выполнены единственным сверточным кодером, у которого между каждыми двумя соседними разрядами регистра сдвига, представленного на рис 6.8.1, добавлено $r - 1$ разрядов и каждый проверочный символ проходит через $(r - 1)/2$-разрядный регистр сдвига.

С практической точки зрения основное преимущество перемежения состоит в том, что оно мало чувствительно по отношению к статистике памяти в канале и требует лишь настолько большого r, чтобы почти все действие памяти в канале исчезло. Недостаток перемежения (или по крайней мере декодера, использующего перемежения) состоит в том,

*) Как указано в § 4.6, это возможно не всегда; в частности, невозможно для каналов с междусимвольной интерференцией.

что при принятии решения о декодировании не принимается во внимание память канала.

Прежде чем перейти к дальнейшему изложению, необходимо условиться о критерии оценки методов кодирования в каналах с пакетами ошибок. Для простоты будем в дальнейшем предполагать, что входные **x**- и выходные **y**-последовательности канала являются двоичными. Шумовая последовательность $\mathbf{z} = \ldots, z_{-1}, z_0, z_1, \ldots$ равна $\mathbf{x} \oplus \mathbf{y}$. При попытке определить, что такое пакет, заметим, что две ошибки (т. е. две единицы) в последовательности **z**, отделенные несколькими нулями, могут рассматриваться либо как две изолированные ошибки, либо как пакет, состоящий из двух ошибок. Чтобы разрешить такого вида неопределенность, *условимся считать совокупность последователь-*

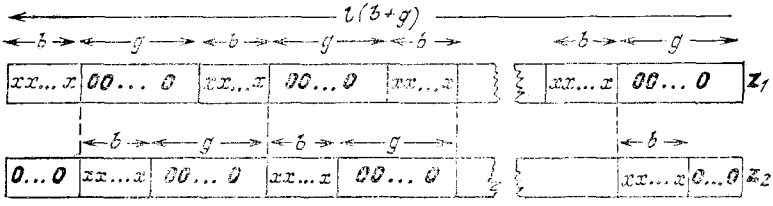


Рис. 6.10.2. Два типа шумовых последовательностей; x — произвольные двоичные символы.

ных символов шумовой последовательности $z_n, z_{n+1}, \ldots, z_{n+b-1}$ *пакетом ошибок относительно защитного интервала* g, *если, во-первых,* $z_n = z_{n+b-1} = 1$, *во-вторых, если по* g *символов с каждой стороны множества* $z_n \ldots z_{n+b-1}$ *равны* 0, *и, в-третьих, если в множестве* $z_n, \ldots, z_{n+b-1}$ *нет* g *последовательных символов, одновременно равных* 0. Длиной пакета b является мощность рассматриваемого множества. Заметим, что все символы шумовой последовательности в пакете не обязательно являются ошибками. В частном случае $b = 1$ пакетом является изолированная ошибка, с каждой стороны которой расположены по g безошибочных символов. Данное выше определение однозначно разбивает любую последовательность $\mathbf{z} = \ldots, z_{-1}, z_0, z_1, \ldots$ на множество пакетов, отделенных друг от друга интервалами по крайней мере из g безошибочных символов.

Условимся говорить, что кодер и декодер имеют при защитном интервале g корректирующую пакеты способность b, если b равно максимальному целому числу, при котором любая шумовая последовательность **z**, содержащая лишь пакеты длины b или меньше при защитном интервале g, декодируется правильно. Аналогично, корректирующей пакеты способностью кодера (или кода) при защитном интервале g называется наибольшее целое b, для которого существует декодер, такой, что корректирующая пакеты способность этих кодера и декодера при защитном интервале g равна b. Будем использовать величину корректирующей пакеты способности кода при данном защитном интервале в качестве критерия эффективности кода при исправлении пакетов ошибок. Следует заметить, что это лишь грубый

критерий. Например, в канале, в котором длинные пакеты, содержащие относительно мало ошибок, гораздо более вероятны, чем более короткие пакеты, содержащие большое число ошибок, предпочтительнее использовать код, способный к исправлению высоко вероятных длинных пакетов за счет менее вероятных коротких пакетов.

Теперь исследуем верхнюю границу корректирующей пакеты способности кода в зависимости от его скорости R и защитного интервала g. Эта граница справедлива для блоковых кодов, сверточных кодов и любых других классов кодов. Предположим, однако, что существует произвольная, но конечная задержка декодирования в N символов источника, т. е. в тот момент, когда n-й ($n > N$) символ источника поступает в кодер, по крайней мере $n - N$ символов источника должны быть декодированы. Вспоминая, что скорость R (измеряемая числом двоичных символов) определяется как отношение числа символов источника к числу символов канала, можно трактовать это условие как требование, чтобы за время, в течение которого принято L символов канала, декодировалось бы не менее $RL - N$ символов источника. Теперь предположим, что используется код, имеющий корректирующую пакеты способность b при защитном интервале g, и рассмотрим случай, когда число принятых символов L кратно $b + g$, т. е. $L = l(b + g)$.

Рассмотрим далее два типа шумовых последовательностей, представленных на рис. 6.10.2. В каждом из этих двух случаев шумовые последовательности таковы, что имеют нули на указанных на рисунке позициях и могут иметь произвольные значения в позициях, обозначенных символами x (мы предполагаем здесь, что $b \leqslant g$). Пусть $\mathbf{x}_1$ и $\mathbf{x}_2$ — кодовые последовательности, соответствующие двум различным наборам первых $[RL - N]$ символов источника и $\mathbf{z}_1$ и $\mathbf{z}_2$ — шумовые последовательности первого и второго типа соответственно Так как по предположению эти шумовые последовательности не могут привести к ошибке декодирования, то имеем

$$\mathbf{x}_1 \oplus \mathbf{z}_1 \neq \mathbf{x}_2 \oplus \mathbf{z}_2. \tag{6.10.1}$$

Более обще, если $\mathbf{z}_1$ и $\mathbf{z}_1'$ — шумовые последовательности первого типа и $\mathbf{z}_2$ и $\mathbf{z}_2'$ — шумовые последовательности второго типа, то должно выполняться соотношение

$$\mathbf{x}_1 \oplus \mathbf{z}_1 \oplus \mathbf{z}_2 \neq \mathbf{x}_2 \oplus \mathbf{z}_1' \oplus \mathbf{z}_2'. \tag{6.10.2}$$

Чтобы доказать (6.10.2), предположим, что (6.10.2) не выполняется при каком-либо выборе последовательностей, и убедимся, что такое предположение приводит к противоречию. Если в (6.10.2) имеет место знак равенства, то, прибавив к обеим сторонам этого равенства $\mathbf{z}_1'$ и $\mathbf{z}_2'$, получим в результате

$$\mathbf{x}_1 \oplus \mathbf{z}_1 \oplus \mathbf{z}_1' = \mathbf{x}_2 \oplus \mathbf{z}_2 \oplus \mathbf{z}_2'. \tag{6.10.3}$$

Поскольку $(\mathbf{z}_1 \oplus \mathbf{z}_1')$ — шумовая последовательность первого типа, а $\mathbf{z}_2 \oplus \mathbf{z}_2'$ — шумовая последовательность второго типа, то равенство (6.10.3) противоречит (6.10.1) и, следовательно, соотношение (6.10.2) верно. Наконец, заметим, что если $\mathbf{x}_1$ и $\mathbf{x}_2$ одинаковы и соответствуют

одним и тем же $(RL - N)$ первым символам источника, но или $\mathbf{z}_1 \neq \mathbf{z}_1'$, или $\mathbf{z}_2 \neq \mathbf{z}_2'$, то соотношение (6.10.2) вновь сохраняет силу.

Теперь легко понять, что $\mathbf{x}_1$ можно выбрать по крайней мере $2^{RL-N} = 2^{Rl(b+g)-N}$ различными способами, каждый из которых соответствует различным выборам первых $(RL - N)$ символов источника. Аналогично имеется 2^{lb} различных способов выбора $\mathbf{z}_1$ и 2^{lb} различных способов выбора $\mathbf{z}_2$. Согласно (6.10.2) при различных выборах тройки $(\mathbf{x}_1, \mathbf{z}_1, \mathbf{z}_2)$ получаем различные последовательности $\mathbf{x}_1 \oplus \mathbf{z}_1 \oplus \mathbf{z}_2$. Поскольку двоичную последовательность длины $l(b+g)$ можно выбрать $2^{l(b+g)}$ различными способами, то имеют место неравенства:

$$2^{Rl(b+g)-N} \cdot 2^{lb} \cdot 2^{lb} \leqslant 2^{l(b+g)}, \tag{6.10.4}$$

$$R(b+g) - \frac{N}{l} + 2b \leqslant b + g. \tag{6.10.5}$$

Так как N фиксировано, а неравенство (6.10.5) должно выполняться при всех $l \geqslant 1$, то можно перейти к пределу при $l \to \infty$; при этом получим

$$R(b+g) + 2b \leqslant b + g. \tag{6.10.6}$$

Аналогичные расчеты в предположении, что шумовые последовательности второго типа (рис. 6.10.2) имеют пакеты длины g, могут быть проведены при $b \geqslant g$. В результате получим, что $R = 0$ (см. задачу 6.46).

Преобразовывая (6.10.6) в выражение (6.10.7), приведенное ниже, получаем следующую теорему.

Теорема 6.10.1. Для того чтобы двоичные кодер и декодер с ограниченной задержкой декодирования и скоростью $R > 0$ (в двоичных символах на символ канала) имели корректирующую пакеты способность b относительно защитного интервала g, необходимо, чтобы

$$\frac{g}{b} \geqslant \frac{1+R}{1-R}. \tag{6.10.7}$$

В дальнейшем будут рассмотрены сначала циклический, а затем сверточный коды, исправляющие пакеты ошибок. Будет показано, что (по крайней мере для сверточных кодов) в пределе при больших g можно приблизиться к границе (6.10.7) сколь угодно близко. Мы также увидим, что большинство пакетов, имеющих длины, гораздо большие, чем значение b, определяемое (6.10.7), может быть исправлено при больших g.

Циклические коды

Для исправления пакетов ошибок рассмотрим использование двоичного циклического (N, L)-кода с порождающим многочленом $g(D)$ степени $N - L$. Пусть $x(D)$ — многочлен, соответствующий передаваемому кодовому слову, $z(D)$ — многочлен, соответствующий,

шумовой последовательности, и $y(D) = x(D) + z(D)$ — многочлен, соответствующий принятой последовательности.

В случае циклического кода пакет ошибок удобно определить несколько иначе, чем это сделано выше. Найдем в шумовой последовательности $\mathbf{z} = (z_{N-1} \ldots, z_0)$ самый длинный интервал циклически следующих друг за другом нулей и рассмотрим оставшуюся часть последовательности как пакет. Например, если

$$\mathbf{z} = (1, 0, 0, \ldots, 0, 0, 1), \tag{6.10.8}$$

то согласно этому определению $\mathbf{z}$ состоит не из одиночных пакетов, а из пакета длины 2, состоящего из z_{N-1}, и из z_0. Чтобы хотя бы частично убедиться в разумности этого довольно странного определения, предположим, что циклический код способен исправлять все пакеты длины b или меньше. Тогда согласно нашему старому определению корректирующая пакеты способность кода при защитном интервале $N - b$ не меньше b. Однако если код не исправляет пакеты в циклическом смысле, то он не будет исправлять и два пакета в канале, отделенные участком из $N - b$ неискаженных символов в середине блока.

В случае циклического кода оптимальным декодером для пакетов ошибок называется такой декодер, который при заданном $y(D)$ выбирает кодовое слово $x(D)$, при котором $y(D) - x(D)$ состоит из самых коротких пакетов. Такой декодер минимизирует вероятность ошибки декодирования в канале, для которого любой пакет данной длины менее вероятен, чем любой пакет меньшей длины. Покажем теперь, что задача построения оптимального декодера для пакетов оказывается удивительно простой.

При любом i, $0 \leqslant i \leqslant N - 1$, назовем i-м циклическим сдвигом многочлена $y(D)$, соответствующего принятой последовательности, многочлен

$$y^{(i)}(D) = R_{D^N-1}[D^{N-i} y(D)]. \tag{6.10.9}$$

Кроме того, введем многочлен $B_i(D)$:

$$B_i(D) = R_{g(D)}[y^{(i)}(D)]. \tag{6.10.10}$$

Так как многочлен $y^{(i)}(D) - B_i(D)$ делится на $g(D)$, то он соответствует кодовому слову. Поскольку при каждом циклическом сдвиге кодового слова также получается кодовое слово, то можно, сдвигая полином $y^{(i)}(D)$ назад к $y(D)$ и одновременно полином $B_i(D)$ на i позиций в том же направлении, убедиться, что многочлен $y(D) - R_{D^N-1}[D^i B_i(D)]$ также соответствует кодовому слову. Иначе говоря, многочлен $R_{D^N-1}[D^i B_i(D)]$ при заданном $y(D)$ и при любом i определяет возможную шумовую последовательность, которая могла бы наложиться на переданное слово, давая в итоге $y(D)$. Из (6.10.10) следует, что каждый из полиномов $B_i(D)$ имеет степень не больше $N - L - 1$ и потому $R_{D^N-1}[D^i B_i(D)]$ соответствует последовательности, у которой все ненулевые символы циклически следуют друг за другом в ряду, состоящем не более чем из $N - L$ символов.

Теперь допустим, что передано $x(D)$ и на него наложился пакет из $b \leqslant N - L$ ошибок в промежутке от z_i до z_{i-b-1} (или до $z_{i+b-1-N}$ при модифицированном определении пакета). Этот пакет можно представить с помощью многочлена $R_{D^N-1}[D^i \beta(D)]$, где степень $\beta(D)$ равна $b-1$; при этом получим $y(D) = xD) + R_{D^N-1}[D^i \beta(D)]$. Тогда имеем $y^{(i)}(D) = x^{(i)}(D) + \beta(D)$, где $x^{(i)}(D)$ — циклический сдвиг $x(D)$. В силу определения (6.10.10) это означает, что $\beta(D) = B_i(D)$.

Из проведенных рассуждений вытекает следующая последовательность действий оптимального декодера: нахождение многочленов $B_i(D)$ для всех i; выбор значения i, при котором $B_i(D)$ соответствует

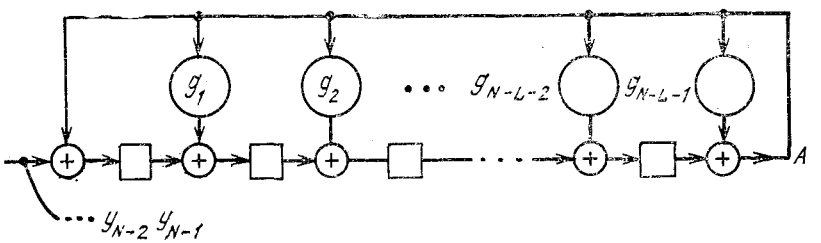


Рис. 6.10.3. Устройство для вычисления $B_i(D)$ при каждом i.

самому короткому пакету, и затем сложение $R_{D^N-1}[D^i B_i(D)]$ при найденном значении i и $y(D)$. Схема устройства, вычисляющего $B_i(D)$ при любом i, представлена на рис 6.10.3. Можно заметить, что это — просто устройство, производящее деление $y(D)$ на $g(D)$, и что после поступления y_0 в левый разряд регистра сдвига, в его разрядах будет храниться как раз полином $B_0(D)$ (члены более высокого порядка находятся справа). Однако если после этого вновь сдвинуть регистр, не подавая при этом на его вход символы принятой последовательности, то в разрядах регистра будет храниться многочлен $R_{g(D)}[Dy(D)]$, который, как нетрудно убедиться, как раз совпадает с $B_{N-1}(D)$. Путем подобных последовательных сдвигов порождаются $B_{N-2}(D)$, $B_{N-3}(D)$, ..., $B_1(D)$. Можно также убедиться, что самый простой способ нахождения $B_i(D)$, соответствующего наименьшему пакету, состоит в нахождении самой длинной последовательности нулей в точке A после поступления слева в регистр сдвига y_0. Наконец, для выполнения декодирования необходимо, чтобы регистр сдвига совершил цикл более чем N раз; после того как в A появится самый длинный отрезок нулей, в регистре будет содержаться пакет ошибок.

Иногда желательно несколько видоизменить такой декодер. Например, можно не рассматривать модифицированные пакеты, так как они обычно гораздо менее вероятны, чем обычные пакеты. Также любой пакет, имеющий длину больше заданной, можно относить к обнаруживаемым ошибкам. Наконец, при определении, какой из многочленов $B_i(D)$ следует использовать, декодер может принимать во внимание как число ошибок в пакете, так и длину пакета.

Теперь исследуем корректирующую пакеты способность b цик-

лических кодах, где b — наибольшее целое число, такое, что все пакеты длины b или меньше могут быть исправлены оптимальным декодером. Если корректирующая пакеты способность циклического кода равна b, то существует некоторый пакет длины $b+1$, который может быть декодирован как некоторый другой пакет длины $b+1$ или меньше; следовательно, сумма этих пакетов является кодовым словом. Произведя циклический сдвиг этого кодового слова так, чтобы пакет неисправимых ошибок начинался с нулевой позиции, представим это сдвинутое кодовое слово в виде

$$x(D) = B(D) + D^m B'(D), \tag{6.10.11}$$

где степень $B(D)$ равна b и степень B'_D не больше b.

Так как $x(D)$ — кодовое слово, то его можно представить в виде $A(D)\,g(D)$, где степень $A(D)$, равная, например, a, представляется в виде разности степени члена самого высокого порядка в $D^m B'(D)$ и $N-L$. Поэтому, для того чтобы степень $B'(D)$ не превосходила $b+1$, необходимо, чтобы коэффициенты при D^i в многочлене $A(D)\,g(D)$ были бы равны нулю для

$$b+1 \leqslant i \leqslant N-L+a-b-1. \tag{6.10.12}$$

Другими словами, для того чтобы существовал неисправимый пакет длины $b+1$, необходимо, чтобы существовали некоторое целое число a, $0 \leqslant a \leqslant L$, и некоторый ненулевой многочлен $A(D)$ степени a, такой, что коэффициенты членов многочлена $A(D)\,g(D)$ равны нулю при всех i, удовлетворяющих (6.10.12). Длиной исправимого пакета циклического кода называется такое наименьшее b, при котором такое решение существует. При любых заданных a и b нахождение такого решения эквивалентно нахождению совокупности $(a-1)$ коэффициентов $A_1, \ldots, A_{a-1}$, которые удовлетворяли бы $N-L+a-2b-1$ линейным уравнениям, указанным в (6.10.12). Легко видеть, что $b \leqslant (N-L)/2$, поскольку пакет $B(D)$, образуемый с помощью усечения $g(D)$ до первых $\lfloor (N-L+1)/2 \rfloor$ членов, всегда может быть воспринят как $g(D) - B(D)$. Также легко увидеть, что граница $b \leqslant \lfloor (N-L)/2 \rfloor$, справедливая для циклических кодов, эквивалентна более общей границе, приведенной в теореме 6.10.1.

К сожалению, довольно мало изучена проблема нахождения при заданных N и L многочлена $g(D)$, максимизирующего b. Файр (1959) нашел большой класс циклических кодов со сравнительно большими значениями b, а Элспас и Шорт (1962) опубликовали небольшую таблицу циклических кодов с оптимальными значениями b. Казами (1963, 1964) также указал упрощенную процедуру решения данных выше уравнений и также привел таблицу укороченных циклических кодов с оптимальными значениями b. Для любого циклического (N, L) кода укороченный циклический код с длиной блока N', $N-L < N' < N$ образуется с помощью отбрасывания первых $N-N'$ информационных символов циклического кода и заменой этих отброшенных символов нулями при вычислении проверочных символов.

Теперь исследуем долю исправимых пакетов ошибок, имеющих длины, большие, чем корректирующая пакеты способность.

Теорема 6.10.2. Пусть циклический $(N,\ L)$-код имеет корректирующую пакеты способность, равную b. Тогда при $b < b' \leqslant N - L$ доля $f(b')$ пакетов длины b', неправильно декодированных оптимальным для пакетов декодером, ограничена неравенствами

$$f(b') \geqslant \begin{cases} 2^{-e(b')}; & b' = b+1, \\ {}^1/_2\, 2^{-e(b')}; & b+1 < b' \leqslant N-L; \end{cases} \qquad (6.10.13)$$

$$f(b') \leqslant \begin{cases} 2(N-1)\,2^{-e(b')}; & b' = b+1 \text{ и } N-L-b+1 \leqslant b', \\ (N-1)\,2^{-e(b')}; & b+1 < b' < N-L-b+1, \end{cases} \qquad (6.10.14)$$

где

$$e(b') = \begin{cases} b; & b+1 \leqslant b' \leqslant N-L-b, \\ N-L-b'+1; & N-L-b+1 < b' < N-L. \end{cases}$$

Обсуждение. Наибольший интерес эта теорема представляет для больших b и $(N-L)$, когда величина коэффициентов в (6.10.14) не имеет большого значения. Весьма интересно, что при $b' < N-L-\log_2 N$ исправляется большинство пакетов. При фиксированном значении скорости L/N и возрастании N отношение этой оценки к верхней границе корректирующей пакеты способности стремится к 2. График функции $e(b')$ приведен на рис. 6.10.4. Можно заметить, что при возрастании корректирующей пакеты способности b горизонтальный участок функции $e(b')$ поднимается, уменьшая долю неисправимых пакетов в окрестности $b' = (N-1)/2$.

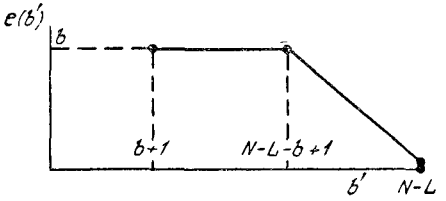


Рис. 6.10.4. Показатель экспоненты доли неисправных пакетов длины b'.

Доказательство. Назовем *синдромом* $S(D)$, соответствующим шумовой последовательности $z(D)$, многочлен

$$S(D) = R_{D^N-1}[z(D)\,h(D)], \qquad (6.10.15)$$

где $h(D) = (D^N - 1)/g(D)$. Заметим, что две шумовые последовательности имеют один и тот же синдром, если их сумма является кодовым словом. Отметим также, что множество всех синдромов, соответствующих всем различным $z(D)$, просто совпадает с множеством кодовых слов в дуальном коде, порождаемом многочленом $h(D)$. Пакет ошибок называется путающим, если его синдром совпадает с синдромом другого пакета меньшей или равной длины. Отметим, что любой неисправимый пакет будет путающим и по крайней мере половина путающих пакетов любой длины декодируется оптимальным декодером неправильно. Наконец, доля путающих пакетов длины b', расположенных на позициях от 0 до $b'-1$, совпадает с долей путающих пакетов длины b', начинающихся в любой другой позиции.

При любых целых значениях m, $1 \leqslant m \leqslant N-1$, и любых b', $0 \leqslant b' \leqslant N-L$, обозначим через $A_{m,\,b'}(u,\ v)$ (рис. 6.10.5) множество

синдромов $S(D)$, для которых $S_{m-1-(N-L-b')} = u$, $S_{L+b'-1} = v$ и $S_i = 0$ при $m-(N-L-b') \leqslant i \leqslant m-1$ и $L+b' \leqslant i \leqslant N-1$. Коэффициенты здесь выбираются как вычеты по модулю N. Из рассмотрения положений интервалов нулей на рис. 6.10.5 видно, что каждый синдром в- $A_{m,b'}$ (1,1) может быть представлен как в виде $S(D) = B_1(D)\ h(D)$, так и в виде $S(D) = R_{D^N-1}[D^m B_2(D) \times \times h(D)]$, где каждый из полиномов $B_1(D)$ и $B_2(D)$ имеет степень b'. Найдем сначала границу числа элементов в этих множествах синдромов, а затем укажем ее связь с числом неисправимых пакетов длины b'. Заметим, что $A_{m,b'}$ (0, 0) всегда содержит все нулевые синдромы и образует подгруппу множества всех синдромов относительно сложе-

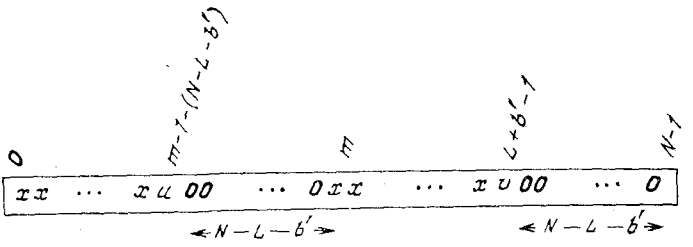


Рис. 6.10.5. Множество синдромов $A_{m,b}(u, v)$; x — произвольные двоичные символы.

ния многочленов по модулю 2. Заметим также, что при любых u, v множество $A_{m,b'}(u, v)$ или пусто или образует смежный класс для $A_{m,b'}$ (0, 0). Так как общее число синдромов является степенью 2, то из теоремы Лагранжа (теорема 6.3.1) следует, что каждая из этих подгрупп и каждый из этих смежных классов имеют число элементов, равное степени 2. Определим множество $A_{m,b'}$ следующим образом:

$$A_{m,b'} = \bigcup_{u=0,1} \bigcup_{v=0,1} A_{m,b'}(u, v). \qquad (6.10.16)$$

Из рассмотрения рис. 6.10.3 непосредственно проверяется, что

$$A_{m,b'-1} = A_{m,b'}(0, 0). \qquad (6.10.17)$$

Отсюда следует, что

$$\|A_{m,b'}\| = \|A_{m,b'-1}\|\,\alpha(m, b'), \qquad (6.10.18$$

где $\alpha(m,b')$ равно числу комбинаций u и v, для которых $A_{m,b'}(u, v)$ не пусто. Так как $\|A_{m,b'}\|$ и $\|A_{m,b'-1}\|$ являются степенью 2, то $\alpha(m, b')$ равно 1, 2 или 4.

Далее заметим, что согласно рис. 6.10.5, если $S(D) \in A_{m,b'-1}(u, v)$, то $DS(D) \in A_{m,b'}(u, v)$. Поэтому, если множество $A_{m,b'}(u, v)$ не пусто для одного из значений b', то оно не пусто и для всех больших значений b'. Следовательно, $\alpha(m, b')$ не убывает с ростом b'.

Согласно определению величины b (корректирующей пакеты способности кода) существует некоторый пакет длины $b+1$, про-

стирающийся от 0-й до b-й позиций, который имеет такой же синдром, что и некоторый пакет с длиной, не превышающей $b+1$, начинающийся, например, на позиции m'. Этот синдром принадлежит $A_{m',\,b+1}$ $(1,1)$. Более того,

$$\|A_{b+1,\,m'}(1,1)\|=1. \tag{6.10.19}$$

Доказательство этого опирается на тот факт, что бо́льшее число синдромов содержат ненулевой синдром из $A_{m',\,b+1}$, $(0,0)$. Но такой синдром содержал бы два интервала, каждый из которых целиком состоит не менее чем из $N-L-b$ нулей, либо сливающихся, либо нет. Если интервалы сливаются в один, то длина получающегося интервала должна быть не менее $N-L$, что влечет за собой требование, чтобы синдром был равен $\mathbf{0}$. Если интервалы не сливаются, то существуют два пакета, длина которых не превышает b, с таким синдромом, что противоречит определению b.

Учитывая, что эти множества являются смежными классами и используя (6.10.16) и (6.10.17), получаем

$$\|A_{m',\,b+1}\|\geqslant 2;\quad \|A_{m',\,b}\|=1. \tag{6.10.20}$$

Наконец, отметим, что при $b'=N-L$ все синдромы принадлежат $A_{m',\,b'}$ и

$$\|A_{m'\,N-L}\|=2^{N-L}. \tag{6.10.21}$$

Так как величина $\alpha\,(m',\,b')$, входящая в соотношение (6.10.18), не убывает с ростом b' и принимает при $b+1\leqslant b'\leqslant N-L$ лишь значения 2 или 4, соотношения (6.10.20) и (6.10.21) полностью определяют $\|A_{m',\,b'}\|$ при всех промежуточных значениях величины b':

$$\|A_{m'\,b'}\|=\begin{cases}2^{b'-b}; & b\leqslant b'\leqslant N-L-b,\\ 2^{2b'-(N-L)}; & N-L-b\leqslant b'\leqslant N-L.\end{cases} \tag{6.10.22}$$

Используя равенство (6.10.17) и равенство $\|A_{m'\,b'}(0,0)\|=\|A_{m'\,b'}(1,1)\|$, получаем

$$\|A_{m',\,b}(1,1)\|=\begin{cases}2^{b'-b-1}; & b+1\leqslant b'\leqslant N-L-b+1,\\ 2^{2b'-2-(N-L)}; & N-L-b+1\leqslant b'\leqslant N-L.\end{cases} \tag{6.10.23}$$

Каждый синдром $S(D)\in A_{m',\,b'}(1,1)$, для которого $S_0=1$, соответствует путающему пакету длины b', расположенному на позициях от 0 до $b'-1$. При $b'=b+1$ единственный синдром из $A_{m',\,b+1}(1,1)$ имеет $S_0=1$, а при бо́льших b' ровно половина синдромов из $A_{m',\,b'}(1,1)$ имеет $S_0=1$. Чтобы показать это, заметим, что если $S(D)\in A_{m',\,b+1}(1,1)$, то $S'(D)=D^{b'-b-1}S(D)$ принадлежит $A_{m',\,b'}(1,1)$ с $S_0'=0$ и $S''(D)=(1+D^{b'-b-1})\,S(D)$ принадлежит $A_{m',\,b'}(1,1)$ с $S_0''=1$. Так как множество $S(D)\in A_{m',\,b'}(1,1)$ с $S_0=1$ и аналогичное множество с $S_0=0$ являются оба смежными классами подгруппы $S(D)\in A_{m',\,b'}(0,0)$ с $S_0=0$, то эти множества имеют одинаковую мощность. Далее, так как существует всего $2^{b'-2}$ различных пакетов, занимающих позиции от 0 до $b'-1$, то доля путаю-

щих пакетов, занимающих позиции от 0 до $b'-1$, ограничена снизу величинами

$$\|A_{m', b'}(1, 1)\| 2^{-b'+2}; \quad b'=b+1,$$
$$\|A_{m', b'}(1, 1)\| 2^{-b'+1}; \quad b'>b+1.$$

Так как по крайней мере половина путающих пакетов декодируется неправильно, последнее соотношение в сочетании с (6.10.23) приводит к нижней границе $f(b')$ вида (6.10.13).

Чтобы получить верхнюю границу $f(b')$, следует прежде всего оценить $\|A_{m, b'}(1, 1)\|$ при каждом значении m. Пусть b_m при любом заданном m — наибольшее целое число, для которого $\|A_{m, b_m}\| = 1$. Тогда $\|A_{m, b_m+1}(u, v)\| = 1$ при некоторой ненулевой паре значений u, v. Если $\|A_{m, b_m+1}(1, 1)\| = 1$, то, повторяя предыдущие рассуждения, можно получить

$$\|A_{m, b'}(1, 1)\| \leqslant \|A_{m', b'}(1, 1)\|; \quad b+1 \leqslant b' \leqslant N-L. \quad (6.10.24)$$

Вместе с тем, если $\|A_{m, b_m+1}(1, 1)\| = 0$, то $\|A_{m, b'}(1, 1)\| = 0$ при всех b', таких, что $\alpha(m, b') = 2$, и $\alpha(m, b') = 4$ для $\|A_{m, b'}(1, 1)\| > 0$. Следовательно,

$$\|A_{m, b'}(1, 1)\| = \begin{cases} 2^{2b'-2-(N-L)} \\ \text{или} \\ 0 \end{cases} \quad (6.10.25)$$

и поэтому (6.10.24) справедливо при всех m, $1 \leqslant m \leqslant N-L$.

Любой неисправимый пакет длины b', занимающий позиции от 0 до $b'-1$, соответствует синдрому $S(D)$ в $A_{m, b'}(1, 1)$ при некотором m, $1 \leqslant m \leqslant N-L$, с $S_0 = 1$. Поэтому общее число неисправимых пакетов длины b', занимающих позиции от 0 до $b'-1$, ограничено сверху величиной $^1/_2 (N-1)\|A_{m', b'}(1, 1)\|$ для $b+1 < b' \leqslant N-L-b+1$ и величиной $(N-1)\|A_{m', b'}(1, 1)\|$ для других значений b'. Учитывая (6.10.23), отсюда получаем верхнюю оценку $f(b')$ вида (6.10.14).

Другой метод блокового кодирования для каналов с пакетами ошибок основан на использовании кодов Рида-Соломона (1961). Используемые символы принадлежат полю $GF(2^m)$ при некотором m, длина блока равна $N = 2^m - 1$. Для произвольно выбранного минимального расстояния кода, заданного нечетным числом d, число информационных символов равно $L = N - d + 1$ и любая конфигурация $(d-1)/2 = (N-L)/2$ ошибок может быть исправлена. Если представить каждую букву кодового слова m двоичными символами, то получим двоичный код с Lm информационными символами и длиной блока Nm. Любая шумовая последовательность, которая искажает не более чем $(N-L)/2$ из этих последовательностей длины m, исправляется; поэтому корректирующая пакеты способность этого кода равна $m\,[(N-L)/2-1]+1$, при этом код исправляет большое число

конфигураций из более коротких пакетов. Отсюда видно, что, увеличивая m при фиксированном L/N, мы подходим сколь угодно близко к теоретическому пределу корректирующей пакеты способности, определяемому теоремой 6.10.1.

Сверточные коды

В этом параграфе рассматриваются некоторые частные методы исправления пакетов ошибок при использовании сверточных кодов. Все эти методы могут быть различным образом модифицированы и обобщены. Однако в настоящее время разработка сверточных кодов для исправления пакетов ошибок еще не вышла из той стадии, когда лучше всего обучить этому искусству на примере. Первый из обсуждаемых методов был независимо развит Ивадари (1967) и Месси*). Он применим

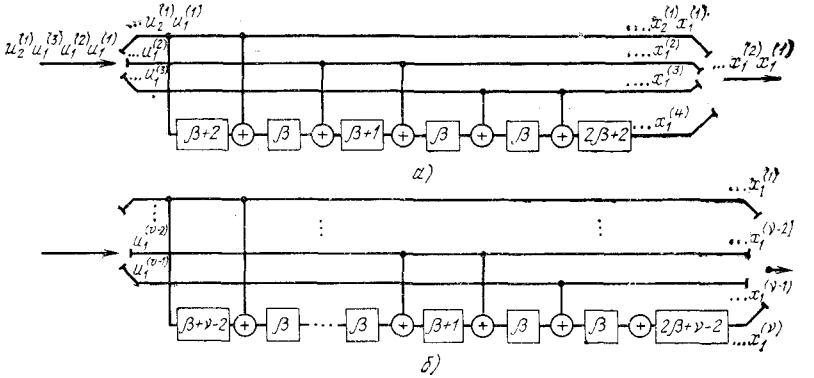


Рис. 6.10.6. Сверточный кодер для $R=3/4$ (бит) ($а$); β означает регистр сдвига на β разрядов. Сверточный кодер для произвольного $R=(\nu-1)/\nu$ ($б$).

к сверточным кодам со скоростями передачи (в двоичных единицах) вида $(\nu-1)/\nu$, где ν — произвольное положительное целое число. Данный метод позволяет для любого положительного целого числа β достичь исправляющей пакеты способности $\beta\nu$ при защитном интервале $\beta\nu(2\nu-1)+\frac{1}{2}\nu^2(\nu-1)-1$. В пределе при больших β отношение длины защитного интервала к корректирующей пакеты способности стремится к $2\nu-1$, что соответствует верхней границе отношения длины защитного интервала к корректирующей пакеты способности, определяемой теоремой 6.10.1.

Блок-схема кодера для метода декодирования Ивадари — Месси изображена на рис. 6.10.6, вначале для частного случая $\nu=4$ ($R=3/4$ бит), а затем — в общем случае. На рис. 6.10.7 представлен декодер для $\nu=4$, из приведенных ниже объяснений станет довольно очевидно, как модифицировать его применительно к произволь-

*) Неопубликованная заметка, июль 1967 г.

ному ν. В случае $\nu = 4$ кодирование производится согласно следующему правилу: при любом n,

$$x_n^{(i)} = u_n^{(i)}; \quad i = 1, 2, 3,$$

$$x_n^{(4)} = u_{n-2\beta-2}^{(3)} \oplus u_{n-3\beta-2}^{(3)} \oplus u_{n-4\beta-2}^{(2)} \oplus u_{n-5\beta-3}^{(2)} \oplus u_{n-6\beta-3}^{(1)} \oplus u_{n-7\beta-5}^{(2)}. \tag{6.10.26}$$

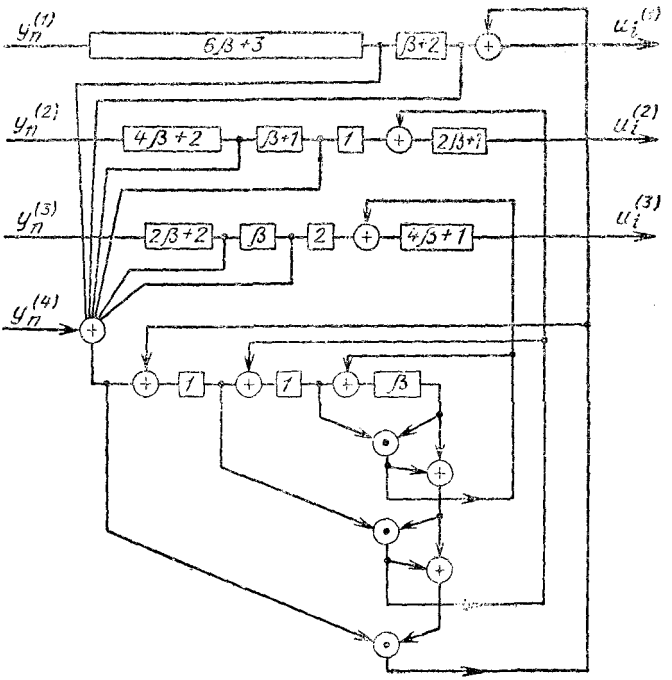


Рис. 6.10.7. Декодер для исправления пакетов (метод Ивадари — Месси); $R=3/4$.

Символы синдрома S_n определяются через принятые по каналу символы с помощью соотношения

$$S_n = y_n^{(4)} \oplus y_{n-2\beta-2}^{(3)} \oplus y_{n-3\beta-2}^{(3)} \oplus y_{n-4\beta-2}^{(2)} \oplus y_{n-5\beta-3}^{(2)} \oplus y_{n-6\beta-3}^{(1)} \oplus y_{n-7\beta-5}^{(1)}. \tag{6.10.27}$$

Из (6.10.26) непосредственно следует, что значения принятых символов в (6.10.27) можно заменить шумовыми символами. Также, если шумовая последовательность содержит лишь пакеты не больше чем из $\nu\beta$ символов при защитном интервале $\beta\nu(2\nu-1) + {}^1/_2\nu^2(\nu-1) - 1$, то, как нетрудно убедиться, каждый синдром S_n может содержать не более одного неисправленного шумового символа. Действительно, защитный интервал был выбран так, что в случае, когда $z_{n-7\beta-5}^{(1)}$ является последней ошибкой какого-либо пакета, $z_n^{(4)}$ не может быть первой ошибкой последующего пакета.

Чтобы понять работу декодера, полезно рассмотреть прохождение через декодер пакета из $\nu\beta$ ошибок. Такой пакет порождает четыре последовательных пакета из единиц в регистре синдрома. Первый из этих пакетов порождается ошибками в четвертом потоке поступающих символов, второй — ошибками в третьем потоке символов, следующий — ошибками во втором потоке и последний — ошибками в первом потоке. Так как длина первого пакета единиц в синдроме не превышает β, то этот пакет не может быть заблокирован каким-либо из элементов «и», расположенным в нижней части рис. 6.10.7; поэтому пакет исправляется. Последовательность синдрома должна содержать не менее $(\beta + 2)$ нулей между первым и вторым пакетом из единиц; следовательно, эти пакеты не могут перекрыться в регистре синдрома.

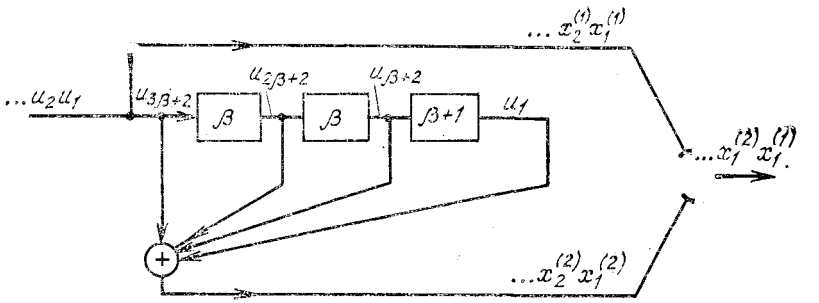


Рис. 6.10.8. Кодер для диффузного порогового декодирования.

Появление каждой ошибки в третьем потоке принятых символов увеличивает на 2 число единиц в синдроме и эти единицы отстоят друг от друга на β символов, что приводит к исправлению в соответствующий момент символов в третьем потоке. Сумматоры по модулю 2 в нижней части рисунка не позволяют проводить какие-либо исправления в первом и во втором потоках в течение этого времени. Каждое исправление также приводит к изменению соответствующего символа синдрома, так что после исправления последней ошибки в третьем потоке принятых символов регистр синдрома будет целиком заполнен нулями, за исключением, быть может, двух самых левых разрядов. Исправление второго, а затем первого потока принятых символов производится аналогично.

Весьма близкий класс кодов был исследован в совместной работе Вайнера и Эша (1963), Берлекэмпом (1964) и Месси (1965). Рассмотренный в этих работах метод приводит к несколько меньшему защитному интервалу при заданной корректирующей пакеты способности, чем метод Ивадари — Месси, однако увеличение сложности оборудования делает его менее полезным в практических применениях.

Другой метод, развитый с несколько других позиций, диффузное пороговое декодирование, принадлежит Месси и Коленбергу (1964). При заданной корректирующей пакеты способности этот метод требует несколько большего защитного интервала и несколько более сложного оборудования, чем предыдущие методы, но он несколько более

приспособлен к исправлению конфигураций пакетов и независимых ошибок. Лучше всего описать этот метод на примере, представленном на рис. 6.10.8 и 6.10.9. Параметр β на этих рисунках произволен и определяет корректирующую пакеты способность кода. Как мы скоро увидим, корректирующая пакеты способность этого кода равна 2β символов при защитном интервале, содержащем $6\beta + 2$ символов.

Исследование декодера, представленного на рис. 6.10.9, почти аналогично исследованию порогового декодера на рис. 6.8.2. В част-

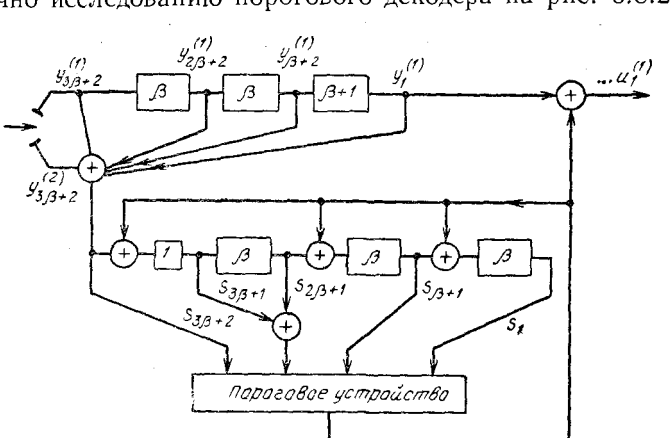


Рис. 6.10.9. Диффузный пороговый декодер.

ности, предполагая, что $y_1^{(1)}$ занимает самый правый разряд регистра сдвига, и что никаких ошибок декодирования ранее сделано не было, получаем, что в пороговое устройство поступают символы:

$$
\begin{aligned}
S_1 &= z_1^{(1)} \oplus z_1^{(2)}, \\
S_{\beta+1} &= z_1^{(1)} \oplus z_{\beta+1}^{(1)} \oplus z_{\beta+1}^{(2)}, \\
S_{2\beta+1} + S_{3\beta+1} &= z_1^{(1)} \oplus z_{2\beta+1}^{(2)} \oplus z_{3\beta+1}^{(1)} \oplus z_{3\beta+1}^{(2)}, \\
S_{3\beta+2} &= z_1^{(1)} \oplus z_{\beta+2}^{(1)} \oplus z_{2\beta+2}^{(1)} \oplus z_{3\beta+2}^{(1)} \oplus z_{3\beta+2}^{(2)}.
\end{aligned}
\tag{6.10.28}
$$

Можно заметить, что линейные комбинации символов шума в правой части (6.10.28) ортогональны к $z_1^{(1)}$ и поэтому $y_1^{(1)}$ будет декодирован правильно, если только в этих соотношениях появится не более двух ошибок. Что касается пакетов ошибок, то, как легко проверить, используя (6.10.28), любой пакет, состоящий не более чем из 2β символов, может влиять максимум на два символа в (6.10.28). Заметим, что защитный интервал выбран таким образом, что если $z_1^{(2)}$ — последний символ какого-либо пакета, то $z_{3\beta+2}^{(2)}$ должен предшествовать началу следующего пакета, но если $z_1^{(1)}$—последний символ какого-либо пакета, то $z_{3\beta+2}^{(2)}$ может быть первым символом следующего пакета.

Можно заметить, что этот метод и метод перемежения символов основаны на близких идеях. В обоих методах символы, рассматриваемые кодером совместно, в канале разнесены так, что пакеты, по существу, преобразуются в независимые ошибки. Однако можно убедиться, что отношение длины защитного интервала к длине пакета для декодера, представленного на рис. 6.10.9, гораздо меньше, чем то, которое может быть получено с помощью перемежения символов в коде, изображенном на рис. 6.8.1.

Существует также тесная связь между методом Ивадари — Месси и диффузным пороговым декодированием. В терминах независимых ошибок декодер, представленный на рис. 6.10.7, может рассматриваться как пороговый декодер, исправляющий одиночную ошибку, а декодер на рис. 6.10.9 — как пороговый декодер, исправляющий двойные ошиб-

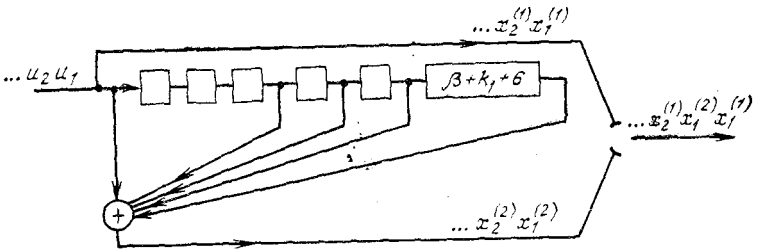


Рис. 6.10.10. Кодер для разнесения пакетов по времени.

ки. Поэтому не удивительно, что диффузный пороговый декодер, исправляющий двойные ошибки, имеет бо́льшую сложность, будучи при этом более приспособлен к исправлению различных типов ошибок, чем декодер Ивадари — Месси.

В качестве последнего примера исправления пакетов ошибок с помощью сверточных кодов рассмотрим следующий метод, принадлежащий Галлагеру (1965б). Сначала рассмотрим частный пример, представленный на рис. 6.10.10 и 6.10.11. Изображенные здесь кодер и декодер предназначены для исправления большинства пакетов с длиной не более 2β при защитном интервале длины $2(\beta + 10 + k_1 + k_2) - 1$. Параметры β_1, k_1 и k_2 произвольны, но можно считать, что обычно β имеет порядок величины 1000, а k_1 и k_2 меньше чем 10. Заметим, что поскольку мы пытаемся исправлять пакеты с длиной, почти в три раза большей верхней оценки корректирующей пакеты способности, определяемой теоремой 6.10.1, то не *все* пакеты длины 2β могут быть исправлены.

Чтобы понять качественно работу декодера, представленного на рис. 6.10.11, рассмотрим, что произойдет, если пакет длины 2β поступит в декодер. Поступление пакета ошибок в декодер приводит к тому, что многие из символов синдрома примут значение 1; фактически ошибки, проходя через первые пять разрядов верхнего регистра сдвига, порождают пакеты единиц в последовательностях синдрома, длина которых не превышает $\beta + 6$. Однако в тот момент, когда начало этого пакета единиц в последовательности синдрома поступает в первый

разряд регистра синдрома, что ведет к обнаружителю ошибок, все ошибки в информационном потоке y_i^1 будут находиться в не имеющем отводов верхнем регистре сдвига, длина которого $\beta + k_1 + b$. Поэтому каждое последующее появление 1 в последовательности синдрома является следствием ошибки в информационном потоке символов в правой части верхнего регистра сдвига. Эта ошибка исправляется, если та часть синдромного регистра, которая связана с обнаружением пакетов, содержит единицу.

Эффективность этого метода можно грубо оценить, предположив, что символы шума внутри пакета являются независимыми равновероятными двоичными символами. Если первая ошибка в пакете пора-

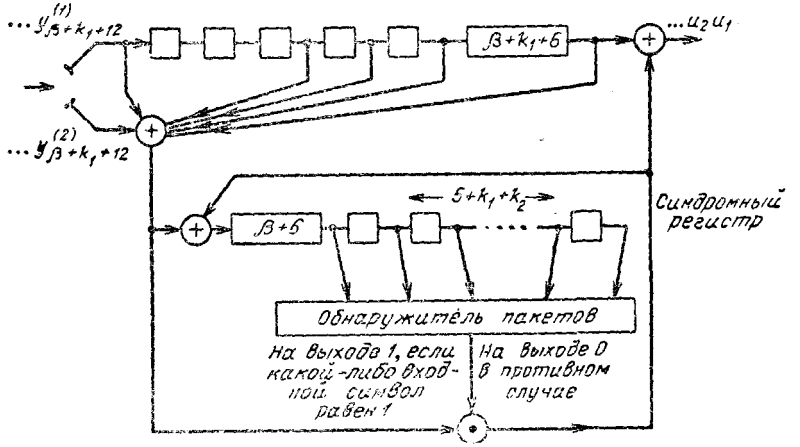


Рис. 6.10.11. Декодер для разнесения пакетов по времени.

жает символ информационного потока, то в тот момент, когда она исправляется, $6 + k_1$ крайне левых символов синдрома, связанных с устройством, обнаруживающим пакеты, будут независимыми равновероятными двоичными символами и вероятность того, что ошибка не будет исправлена, равна 2^{-6-k_1}. Аналогично, вероятность того, что ошибка в середине пакета не будет исправлена, равна $2^{-6-k_1-k_2}$, а вероятность того, что не будет исправлена последняя ошибка в пакете, равна, грубо говоря, 2^{-6-k_2}. Отсюда видно, что с возрастанием k_1 и k_2 вероятность неудачи при исправлении пакета убывает, но возрастают сложность оборудования и требуемая длина защитного интервала.

В действительности этот метод является разновидностью передачи с разнесением по времени. Если данный интервал информационного потока искажен пакетом ошибок, то последующие символы в потоке проверочных символов могут использоваться для преобразования потока. Основное различие состоит в том, что при обычной передаче с разнесением по времени в канале производятся измерения, чтобы определить, какой поток символов наиболее надежен, в то время как

здесь для этого используются кодовые связи. Во многих физических каналах существует столь много причин, приводящих к пакетам ошибок, что трудно надежно определить по измерениям в канале место пакета, поэтому использование для этого кодовых связей более надежно и более просто.

В описанном методе могут быть сделаны некоторые изменения. Ясно, что можно изменить данное отдельное множество разрядов регистра в левой части рис. 6.10.10, выходные символы которых суммируются при вычислении проверочных символов. Более существенная модификация состоит в использовании устройства (рис. 6.10.11), обнаруживающего пакеты как для исправления относительно изолированных ошибок, так и для обнаружения более крупных пакетов. Наконец, можно использовать тот же метод при произвольных скоростях, представимых в виде $R = (\nu - 1)/\nu$, для исправления большинства пакетов длины $\nu\beta$ с длиной защитного интервала, чуть большей $\nu(\nu - 1)\beta$.

ИТОГИ И ВЫВОДЫ

Предыдущие главы были посвящены изучению теоретических ограничений при передаче данных по каналам с шумами; в настоящей главе были рассмотрены методы достижения характеристик, близких к этим ограничениям. Для практики важно установить обменные соотношения между скоростью передачи данных, стоимостью системы и вероятностью ошибки для какого-либо данного метода кодирования и данного канала. К сожалению, стоимость системы является в условиях быстро меняющейся технологии производства весьма нечеткой характеристикой, а в реальных каналах действует большое число факторов, трудно поддающихся математическому описанию. Поэтому, все что можно сделать здесь — это описать некоторые методы кодирования и декодирования, показать, как можно их реализовать и дать некоторое представление об их характеристиках при использовании в различных простых математических моделях каналов. Мы не пытались указать, какой метод следует использовать в данной ситуации, но была сделана попытка подготовить читателя к осмысленному решению конкретных технических проблем в этой области.

Быть может, главное значение методов кодирования состоит в том, что они практически полезны для широкого класса систем связи. Это можно легко упустить из-за устойчивой человеческой тенденции путать привычное с полезным. Кроме того, недооценка методов кодирования связана с большим числом вычислительных операций, производимых кодером и декодером. Однако эти операции становятся в настоящее время высоко надежными и легкими в реализации, хотя остаются трудными для понимания.

Первые шесть параграфов главы посвящены изложению основ, необходимых для исследований и чтения литературы по алгебраическому кодированию. В § 6.7 изложены БЧХ коды, которые являются наиболее интересными и практически перспективными среди алгебраических кодов. В § 6.8 введены сверточное кодирование и пороговое

декодирование. Параграф 6.9 служит довольно полным введением в последовательное декодирование. Для каналов, не подверженных длительным замираниям, последовательное декодирование является наиболее перспективным методом кодирования среди существующих в настоящее время. Наконец, в § 6.10 дано краткое введение в теорию методов кодирования в каналах, пораженных пакетами шумов.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Питерсону (1961) принадлежит классическая книга по методам алгебраического кодирования, в которой обсуждаются различные вопросы, не освещенные здесь. Берлекэмп (1968) внес большой вклад в теорию алгебраического кодирования и изложил ее на современном уровне. Для всех тех, кто может получить лекции Месси (1967), мы рекомендуем их как превосходное изложение предмета. В книге Возенкрафта и Джекобса (1965) дано введение в последовательное декодирование и связанные с ним вопросы, рассмотренные в § 6.9. Книга Месси (1963) является классическим пособием по пороговому декодированию.

Невозможно перечислить здесь все значительные статьи по методам кодирования. Довольно полная библиография работ в данной области приведена Питерсоном (1961), Питерсоном и Месси (1963) и и Коутцом (1967). В историческом плане следует, однако, отметить следующее. Работа Хэмминга (1950) предопределила большинство работ в теории алгебраического кодирования. Слепян (1956) первый изложил коды с проверкой на четность на четкой математической основе. Элайсу (1955) принадлежат результаты § 6.2, а также первое описание сверточных кодов. Прейндж (1957) первый изучил циклические коды. Методы кодирования для кодов Боуза — Чоудхури (1960) и Хоквингема (1959) были впервые развиты Питерсоном (1960) и Цирлером (1960). Последовательное декодирование было изобретено Возенкрафтом (1957), а рассмотренный здесь алгоритм принадлежит Фано (1963). Первые коды, эффективно исправляющие пакеты ошибок, были построены Хейгельбергером (1959) и Файром (1959).

ПРИЛОЖЕНИЕ 6А

Прежде чем приступить к доказательству теоремы 6.9.1, полезно переписать правила алгоритма декодирования так, как представлено на рис. 6А.1. Заметим, что условия, указанные на рис. 6А.1, содержат все условия, представленные на рис. 6.9.4, а также некоторые дополнительные условия. Поэтому, если применяется *некоторое* правило из описания алгоритма на рис. 6А.1, то оно должно совпадать с правилом, указанным на рис. 6.9.4. Чтобы показать, что некоторое правило применимо к каждой проверке на рис. 6А.1, можно воспользоваться индукцией по последовательным проверкам. Рассмотрим каждое из правил на рис. 6А.1; предположим, что данное правило применимо, и рассмотрим условия, которые возникают при следующей проверке. Например, если применяется правило 1, то для следующих проверок $\Gamma_{l-1} \geqslant T$ и это является новым условием, добавляемым к правилам 1, 2 и 3.

Правило	Условия в узле		Действия, которые следует выполнить	
	Предыдущее движение	Сравнение Γ_{l-1} и Γ_l с первоначальным порогом T	Окончательный порог	Движение
1	F или L	$T \leqslant \Gamma_{l-1} < T+\Delta$, $\Gamma_l \geqslant T$	Повышается	F
2	F или L	$\Gamma_{l-1} \geqslant T+\Delta$, $\Gamma_l \geqslant T$	Не изменяется	F
3	F или L	$\Gamma_{l-1} \geqslant T$, $\Gamma_l < T$	Не изменяется	L или B
4	B	$\Gamma_{l-1} < T$, $\Gamma_l \geqslant T$	Понижается на Δ	F
5	B	$\Gamma_{l-1} \geqslant T$, $\Gamma_l \geqslant T$	Не изменяется	L или B

Рис. 6А.1. Множество правил, эквивалентное множеству правил, указанному на рис. 6.9.4.

Доказательство теоремы 6.9.1 (*пункт а*). Мы хотим показать, что пути порогов и цен, связанных с проверками узла $\mathbf{u}_l$, удовлетворяют соотношениям (для $0 \leqslant i \leqslant l-1$):

$$T_i \leqslant \Gamma_i \tag{6А.1}$$

$$T_{i+1} \geqslant T_i, \tag{6А.2}$$

$$T_{i+1} \geqslant T_i + \Delta \Rightarrow T_i + \Delta > \Gamma_i, \tag{6А.3}$$

$$T_{i+1} \geqslant T_i + \Delta \Rightarrow T_{i+1} > \Gamma_i. \tag{6А.4}$$

Последней проверкой для любого узла $\mathbf{u}_i$, предшественника текущего узла $\mathbf{u}_l$, должна быть F-проверка, поскольку $\mathbf{u}_l$ может быть достигнуто лишь при движении вперед из $\mathbf{u}_i$. Поэтому к $\mathbf{u}_l$ должно быть применено одно из правил 1, 2 или 4, и в каждом случае должно выполняться (6А.1). Далее заметим, что (6А.4) является прямым следствием (6А.3); соотношение (6А.4) является соединением неравенств в (6А.3). Теперь докажем (6А.2) и (6А. 3),используя индукцию по последовательности узлов, просмотренных декодером. Для первой проверки в начальном узле справедливость (6А.2) и (6А.3) тривиальна, поскольку множество i пусто при $l = 0$. Теперь предположим, что (6А.2) и (6А.3) выполняются при проверках в произвольном узле $\mathbf{u}_l$ с последовательностью порогов $T_0, \ldots, T_l$. Рассматривая сначала движение вперед, затем вбок и затем назад из $\mathbf{u}_l$, непосредственными, но утомительными рассуждениями показывается, что соотношения (6А.2) и (6А.3) всегда выполняются в следующем проверяемом узле.

Движение вперед. После движения из $\mathbf{u}_l$ вперед в узел $\mathbf{u}_{l+1}$ пороги $T_0, \ldots, T_l$ не изменяются, но к этой последовательности добавляется новый порог T_{l+1}. Так как движение в $\mathbf{u}_{l+1}$ было движением вперед, к $l+1$ применимы правила 1, 2 или 3 и окончательный порог T_{l+1} больше или равен первоначальному порогу T_l, что доказывает (6А.2) для $\mathbf{u}_{l+1}$. Если $T_{l+1} \geqslant T_l + \Delta$, то порог был повышен и должно применяться правило 1. Так как первоначальный порог равен T_l и предшествующая цена равна Γ_l, то крайне левое условие в правиле 1 принимает вид $\Gamma_l < T_l + \Delta$, что устанавливает справедливость (6А.3).

Движение вбок. Пусть $T_0, \ldots, T_l$ — путь порогов при проверках узла $\mathbf{u}_l$ и предположим, что из $\mathbf{u}_l$ производится движение вбок или назад. Тогда к $\mathbf{u}_l$ должно быть применимо правило 3 или 5 и в обоих случаях согласно этим правилам $T_l \leqslant \Gamma_{l-1}$. Мы видим, что движения вбок или назад могут быть произведены лишь при $l > 0$. В силу предположений индукции к $\mathbf{u}_l$ применимо (6А.3) и, следовательно, (6А.4). Из (6А.4) вытекает, что $T_l \geqslant T_{l-1} + \Delta \Rightarrow \Leftarrow T_l > \Gamma_{l-1}$ при $i = l-1$. Поскольку, как было уже показано, $T_l \leqslant \Gamma_{l-1}$, то должно выполняться неравенство $T_l < T_{l-1} + \Delta$. Так как согласно (6А.2) $T_l \geqslant T_{l-1}$ и так как T может изменяться, лишь принимая приращения Δ, то выводим, что при L и B движениях из $\mathbf{u}_l$ имеет место равенство

$$T_l = T_{l-1}. \tag{6А.5}$$

При движении вбок из узла $\mathbf{u}_l$ в узел $\mathbf{u}_l'$ путь порогов $T_0, \ldots, T_{l-1}$ не изменяется. Первоначальный порог при проверках в $\mathbf{u}_l'$ равен T_l' и согласно (6А.5) $T_l' = T_{l-1}$. Пусть T_l' — конечный порог для проверок в $\mathbf{u}_l'$. Так как в узле $\mathbf{u}_l'$ должны применяться правила 1, 2 или 3, то порог не может понижаться и потому $T_l' \geqslant T_{l-1}$, что доказывает (6А.2). Если $T_l' \geqslant T_{l-1} + \Delta$, то порог в $\mathbf{u}_l'$ должен быть повышен, применяется правило 1 и $\Gamma_{l-1} < T_{l-1} + \Delta$, что доказывает (6А.3).

Движение назад. Вновь допустим, что $T_0, \ldots, T_l$ является путем порогов при проверках $\mathbf{u}_l$, и предположим, что было совершено движение назад к $\mathbf{u}_{l-1}$. Согласно (6А.5) первоначальный порог при проверке $\mathbf{u}_{l-1}$ равен $T_l = T_{l-1}$. Так как из $\mathbf{u}_l$ было совершено движение назад, то в $\mathbf{u}_{l-1}$ применяется либо правило 5, либо правило 4. Если применяется правило 5 ,то конечный порог T_{l-1}' при новых проверках $\mathbf{u}_{l-1}$ равен первоначальному порогу, который, как указано выше, равен T_{l-1}. Поэтому, так как для $\mathbf{u}_l$ выполняются соотношения (6А.2) и (6А.3), то они также выполняются и для новой проверки $\mathbf{u}_{l-1}$. Если применяется правило 4, то согласно рис. 6А.1 (учитывая, что T_{l-1}—первоначальный порог), $\Gamma_{l-2} < T_{l-1}$. Согласно (6А.1) $T_{l-2} \leqslant \Gamma_{l-2}$ и в силу того, что пороги могут изменяться лишь на приращения Δ, то $T_{l-2} \leqslant T_{l-1} - \Delta$. Окончательный порог T_{l-1}' при новой проверке $\mathbf{u}_{l-1}$ равен $T_{l-1} - \Delta$, так что $T_{l-2} \leqslant T_{l-1}'$, что доказывает (6А.2) для $\mathbf{u}_{l-1}$. Наконец, если $T_{l-1}' \geqslant T_{l-2} + \Delta$, то $T_{l-1} \geqslant T_{l-2} + \Delta$, и из справедливости соотношения (6А.3) для $\mathbf{u}_l$ следует его справедливость для $\mathbf{u}_{l-1}$, что завершает доказательство пункта (а).

Следующее следствие из пункта (а) теоремы будет необходимо при доказательстве пунктов (б) и (в).

С л е д с т в и е. Если в узле $\mathbf{u}$ проведена F-проверка с окончательным порогом T, то T является начальным порогом при первых последующих проверках каждого из непосредственных потомков узла $\mathbf{u}$ и при первой последующей проверке узла $\mathbf{u}$.

Доказательство. Для проверки первого непосредственного потомка узла $\mathbf{u}$ это утверждение очевидно. Первая проверка каждого из других непосредственных потомков должна производиться при движении вбок из ранее просмотренных непосредственных потомков узла $\mathbf{u}$. Но согласно (6А.5) такое движение вбок может выполняться лишь тогда, когда окончательный порог, установленный до выполнения движения, равен T. Аналогично, первое движение назад к $\mathbf{u}$ совершается из последнего среди непосредственных потомков и порог вновь равен T. |

Доказательство теоремы 6.9.1 (пункт б). Мы хотим показать, что окончательный порог T при первой F-проверке каждого узла связан с ценой узла Γ соотношением

$$T \leqslant \Gamma < T + \Delta, \tag{6А.6}$$

а также что каждая из последующих F-проверок этого узла проводится с окончательным порогом на Δ ниже, чем предыдущее значение.

Воспользуемся индукцией вдоль пути узлов, сначала доказав справедливость теоремы для начального узла, а затем показав, что если теорема верна для какого-либо данного узла, то она справедлива для любого из непосредственных потомков этого узла. Начальные проверки $\mathbf{u}_0$ являются F-проверками и удовлетворяют (6А.6) в силу начальных условий, накладываемых на декодер. Согласно следствию из пункта (а) первоначальный порог для каждой последующей проверки начального узла равен окончательному порогу при предыдущей проверке. Так как $\Gamma_{-1} = -\infty$, то к каждой такой проверке применимо правило 4; причем эта F-проверка производится с окончательным порогом, сниженным на Δ. Теперь предположим, что утверждение пункта (б) теоремы справедливо для узла $\mathbf{u}_{l-1}$ с ценой, равной Γ_{l-1}, и пусть $\mathbf{u}_l$ — его непосредственный потомок с ценой Γ_l. По предположению окончательный порог T при первой F-проверке узла $\mathbf{u}_{l-1}$ удовлетворяет неравенствам:

$$T \leqslant \Gamma_{l-1} < T + \Delta. \tag{6А.7}$$

Согласно следствию, T есть первоначальный порог при первой проверке $\mathbf{u}_l$. Теперь рассмотрим отдельно случаи $\Gamma_l \geqslant T$ и $\Gamma_l < T$.

Если $\Gamma_l \geqslant T$, первая проверка $\mathbf{u}_l$ удовлетворяет условиям правила 1, и окончательный порог T установлен таким образом, что он удовлетворяет условию (6А.6) для $\mathbf{u}_l$. Согласно следствию первоначальный порог при следующем возвращении к $\mathbf{u}_l$ равен T_l. Если $T_l \geqslant T + \Delta$ (т. е. если порог повышен при первоначальной проверке $\mathbf{u}_l$), то согласно (6А.7) $T_l > \Gamma_{l-1}$, применяется правило 4 и при возвращении совершается F-проверка с окончательным порогом $T_l - \Delta$. Рассуждая таким же образом, находим, что окончательный порог уменьшается на Δ при каждом последующем возвращении к $\mathbf{u}_l$ до тех пор, пока окончательный порог не будет равен T. При следующем возвращении к $\mathbf{u}_l$ первоначальный порог меньше или равен Γ_{l-1}, применяется правило 5 и совершается движение вбок или назад. Прежде чем можно будет совершить следующую F-проверку $\mathbf{u}_l$, по предположению должна быть совершена другая F-проверка $\mathbf{u}_{l-1}$ при окончательном пороге $T - \Delta$. Поэтому первоначальный и окончательный пороги при следующих F-проверках $\mathbf{u}_l$ также равны $T - \Delta$. Аналогично, последовательные F-проверки $\mathbf{u}_l$ чередуются с F-проверками $\mathbf{u}_{l-1}$, каждый раз с величиной порога на Δ ниже, чем при предыдущей проверке.

Наконец, рассмотрим случай $\Gamma_l < T$. Тогда при первой проверке $\mathbf{u}_l$ применяется правило 3, совершается движение вбок или назад и $\mathbf{u}_l$ перепроверяется лишь после следующей F-проверки $\mathbf{u}_{l-1}$, на этот раз при окончательном пороге $T - \Delta$. При каждой из последовательных проверок $\mathbf{u}_l$ первоначальный порог понижается на Δ до тех пор, пока порог не станет меньше или равен Γ_l; совершаемая в этот момент проверка — это F-проверка и для $\mathbf{u}_l$ выполняется (6А.6). Как и ранее, окончательные пороги при последовательных проверках $\mathbf{u}_l$ уменьшаются по сравнению с предыдущим значением на величину Δ.

Доказательство теоремы 6.9.1 (пункт в). Пусть в узле $\mathbf{u}_l$ производится F-проверка с окончательным порогом T_l. Мы хотим показать, что прежде чем $\mathbf{u}_l$ можно будет проверить вновь, каждый из потомков $\mathbf{u}_l$, для которого путь из $\mathbf{u}_l$ лежит выше T_l, должен будет пройти F-проверку с окончательным порогом T_l. В процессе доказательства пункта (б) для каждого из непосредственных потомков $\mathbf{u}_l$, например для $\mathbf{u}_{l+1}$, было показано, что если путь из $\mathbf{u}_l$ лежит выше T_l (т. е. если $\Gamma_{l+1} \geqslant T_l$), то прежде чем произвести первую перепроверку узла $\mathbf{u}_l$, будет произведена F-проверка узла $\mathbf{u}_{l+1}$ с окончательным порогом T. Теорема доказывается индукцией по длине пути потомков узла $\mathbf{u}_l$, т. е. если потомок $\mathbf{u}_{l+j}$, находящийся на глубине j от узла $\mathbf{u}_l$, проходит F-проверку с окончательным порогом T_l, то каждый из непосредственных потомков узла $\mathbf{u}_{l+j}$, для которого $\Gamma_{l+j+1} \geqslant T_l$, проходит F-проверку с окончательным порогом T_l, прежде чем $\mathbf{u}_{l+j}$ и, следовательно, $\mathbf{u}_l$ пройдет перепроверку. Из (6.9.3) следует, что этот порог не может быть сделан ниже T_l до тех пор, пока $\mathbf{u}_l$ не будет перепроверен. Это завершает доказательство. |

ПРИЛОЖЕНИЕ 6Б

В этом приложении находится верхняя граница $\Pr[\Gamma'_{m(l)} > \Gamma_{min} + \alpha]$, где α — произвольная постоянная. Случайная переменная Γ_{min} равна $\inf_{n \geqslant 0} \Gamma_n$, где $\Gamma_0, \Gamma_1, \Gamma_2, \ldots$ — цены узлов правильного пути в дереве принятых цеп. В обозначениях переданной и принятой по каналу последовательности $\Gamma_0 = 0$ и при $n > 0$

$$\Gamma_n = \sum_{i=1}^{n} \gamma_i, \tag{6Б.1}$$

$$\gamma_i = \sum_{a=1}^{\nu} \left[\ln \frac{P\left(y_i^{(a)} \mid x_i^{(a)}\right)}{\omega\left(y_i^{(a)}\right)} - B \right]. \tag{6Б.2}$$

В ансамбле кодов величины $x_n^{(a)}$ являются независимыми выборками букв входного алфавита с вероятностями $Q(k)$. Величины $y_n^{(a)}$ через переходные вероятности канала $P(j|k)$ статистически связаны с $x_n^{(a)}$; легко видеть, что γ_i в (6Б.2) — статистически независимые случайные величины. Значение случайной величины $\Gamma'_{m(l)}$ является ценой некоторого данного узла на глубине l в дереве принятых цен, определяемой с помощью равенства

$$\Gamma'_{m(l)} = \sum_{i=1}^{l} \gamma'_i, \tag{6Б.3}$$

где

$$\gamma'_i = \sum_{a=1}^{\nu} \left[\ln \frac{P\left(y_i^{(a)} \mid x_i'^{(a)}\right)}{\omega\left(y_i^{(a)}\right)} - B \right], \tag{6Б.4}$$

а $\mathbf{x}' = x'^{(1)}_1, \ldots, x'^{(\nu)}_1, x'^{(1)}_2, \ldots$ — кодовая последовательность, соответствующая узлу $m(l)$. По предположению путь $m(l)$ соответствует последовательности источника, отличающейся от переданной последовательности в первом подблоке. В ансамбле кодов величины $x'^{(a)}_n$ статистически независимы от величин $x_n^{(a)}$ из (6Б.2), а также статистически независимы друг от друга. Поэтому γ'_i из (6Б.4) — также статистически независимые случайные величины. Следует заметить, что так как $y_n^{(a)}$ из (6Б.4) — те же самые, что $y_n^{(a)}$ из (6Б.2), то γ_i и γ'_i, вообще говоря, не являются статистически независимыми случайными величинами (см. задачу 6.43).

Положим при $n > l$

$$\Gamma_{n,\,l} = \sum_{i=l+1}^{n} \gamma_i$$

и при $n = l$ положим $\Gamma_{n,\,l} = 0$. Отсюда получим $\Gamma_n = \Gamma_l + \Gamma_{n,\,l}$ при $n \geqslant l$. Пусть

$$\min \Gamma_{n,\,l} = \inf_{n \geqslant l} \Gamma_{n,\,l}.$$

Событие, состоящее в том, что $\Gamma'_{m(l)} \geqslant \Gamma_{min} + \alpha$, является объединением двух событий, первое из которых состоит в том, что при $0 \leqslant n \leqslant l-1$ выполняется $\Gamma'_{m(l)} \geqslant \Gamma_n + \alpha$, а второе — в том, что $\Gamma'_{m(l)} \geqslant \Gamma_l + \min \Gamma_{n,\,l} + \alpha$. Следовательно,

$$\Pr\left[\Gamma'_{m(l)} \geqslant \Gamma_{min} + \alpha\right] \leqslant \sum_{n=0}^{l-1} \Pr\left[\Gamma'_{m(l)} \geqslant \Gamma_n + \alpha\right] + \\ + \Pr\left[\Gamma'_{m(l)} \geqslant \Gamma_l + \min \Gamma_{n,l} + \alpha\right]. \tag{6Б.5}$$

Теперь найдем границу сверху для каждого из слагаемых суммы в правой части (6Б.5), применяя границу Чернова к $\Gamma'_{m(l)}$ и Γ_n в (6Б.1) и (6Б.3):

$$\Pr\left[\Gamma'_{m(l)} \geqslant \Gamma_n + \alpha\right] \leqslant \overline{\exp\left\{ s \left[\sum_{i=1}^{l} \gamma'_i - \sum_{i=1}^{n} \gamma_i - \alpha \right] \right\}} \tag{6Б.6}$$

при всех $s \geqslant 0$. Используя статистическую независимость пар γ'_i и γ_i для различных значений i, это неравенство можно переписать в виде

$$\Pr\left[\Gamma'_{m(l)} \geqslant \Gamma_n + \alpha\right] \leqslant e^{-s\alpha} \prod_{i=1}^{n} \overline{\exp\left[s\left(\gamma'_i - \gamma_i\right)\right]} \prod_{i=n+1}^{l} \exp\left(s\gamma'_i\right). \tag{6Б.7}$$

Удобно выбрать $s = {}^1/_2$; используя (6Б.2) и (6Б.4), получаем

$$\overline{\exp\left[{}^1/_2(\gamma_i' - \gamma_i)\right]} = \sum \prod_{a=1}^{\nu} Q\left(x_i^{(a)}\right) P\left(y_i^{(a)} \mid x_i^{(a)}\right) Q\left(x_i'^{(a)}\right) \times$$
$$\times \left[\frac{P\left(y_i^{(a)} \mid x_i'^{(a)}\right)}{\omega\left(y_i^{(a)}\right)}\right]^{1/2} \left[\frac{P\left(y_i^{(a)} \mid x_i^{(a)}\right)}{\omega\left(y_i^{(a)}\right)}\right]^{-1/2}, \qquad (6Б.8)$$

где суммирование производится по всем возможным значениям $x_i^{(a)}$, $x_i'^{(a)}$, $y_i^{(a)}$. Произведя суммирование отдельно для каждого a, $1 \leqslant a \leqslant \nu$ [как в (5.5.7) — (5.5.10)], получим

$$\overline{\exp\left[{}^1/_2(\gamma_i' - \gamma_i)\right]} = \left\{\sum_{j=0}^{J-1}\left[\sum_{k=0}^{K-1} Q(k)\sqrt{P(j \mid k)}\right] \times \right.$$
$$\left. \times \left[\sum_{k'=0}^{K-1} Q(k')\sqrt{P(j \mid k')}\right]\right\}^{\nu} = \left\{\sum_{j=0}^{J-1}\left[\sum_{k=0}^{K-1} Q(k)\sqrt{P(j \mid k)}\right]^2\right\}^{\nu} =$$
$$= \exp\left[-\nu E_0(1, \mathbf{Q})\right]. \qquad (6Б.9)$$

Аналогично имеем

$$\overline{\exp\left({}^1/_2\gamma_i'\right)} = \sum \prod_{a=1}^{\nu} \left\{Q\left(x_i^{(a)}\right) P\left(y_i^{(a)} \mid x_i^{(a)}\right) Q\left(x_i'^{(a)}\right) \times \right.$$
$$\left. \times \left[\frac{P\left(y_i^{(a)} \mid x_i'^{(a)}\right)}{\omega\left(y_i^{(a)}\right)}\right]^{1/2} e^{-B/2}\right\}. \qquad (6Б.10)$$

Вспоминая что $\omega\left(y_i^{(a)}\right) = \sum Q\left(x_i^{(a)}\right) P\left(y_i^{(a)} \mid x_i^{(a)}\right)$, это выражение приводим к виду

$$\overline{\exp\left({}^1/_2\gamma_i'\right)} = \left[\sum_{j=0}^{J-1}\sum_{k=0}^{K-1} Q(k)\,\omega(j)^{1/2} P(j \mid k)^{1/2} e^{-B/2}\right]^{\nu}. \qquad (6Б.11)$$

Применяя к сумме по j неравенство Коши, получаем

$$\overline{\exp\left({}^1/_2\gamma_i\right)} \leqslant \left[\sqrt{\sum_j \omega(j)} \cdot \sqrt{\sum_j \left[\sum_k Q(k)\sqrt{P(j \mid k)}\right]^2}\, e^{-B/2}\right]^{\nu} =$$
$$= \exp\left\{-\frac{\nu}{2}\left[E_0(1, \mathbf{Q}) + B\right]\right\}. \qquad (6Б.12)$$

При ограничении $B \leqslant E_0(1, Q)$ можно также оценить сверху величину $\overline{\exp[{}^1/_2(\gamma_i' - \gamma_i)]}$, входящую в (6Б.9), величиной $\exp\{-(\nu/2)\,[E_0(1, \mathbf{Q}) + B]\}$. Подставляя эту границу и (6Б.12) в (6Б.7) с $s = {}^1/_2$, получаем, что при $n < l$

$$\Pr\left[\Gamma'_{m(l)} \geqslant \Gamma_n + \alpha\right] \leqslant \exp\left\{-\frac{\alpha}{2} - \frac{\nu l}{2}\left[E_0(1, \mathbf{Q}) + B\right]\right\}. \qquad (6Б.13)$$

Далее следует найти верхнюю границу второго члена в (6Б.5), $\Pr\left[\Gamma'_{m(l)} \geqslant \right.$ $\left. \geqslant \Gamma_l + \min \Gamma_{n,l} + \alpha\right]$. Напомним, что

$$\Gamma_{n,l} = \sum_{i=l+1}^{n} \gamma_i$$

при $n > l$ и что γ_i — независимые одинаково распределенные случайные величины. Последовательность $\Gamma_{n,l}$ при $n = l, l+1, \ldots$ называется *случайным*

блужданием. Приведенная ниже лемма является хорошо известным результатом теории случайных блужданий*) и часто используется в теории информации. Мы сформулируем ее и затем воспользуемся результатом здесь, а докажем ее в следующем разделе настоящего приложения.

Лемма 6Б.1. Пусть $z_1, z_2, \ldots$ — последовательность независимых одинаково распределенных дискретных случайных величин. Пусть $S_0 = 0$ и при любом $n > 0$

$$S_n = \sum_{i=1}^{n} z_i.$$

Положим $$S_{min} = \inf_{n \geqslant 0} S_n.$$

Тогда при любом $r \leqslant 0$, таком, что $\overline{\exp(rz_i)} \leqslant 1$, и при любом u

$$\Pr[S_{min} \leqslant u] \leqslant e^{-ru}. \tag{6Б.14}$$

Применительно к нашему случаю $z_i = \gamma_{i+l}$, $i \geqslant 1$, так, что $\overline{S_{min} = \min \Gamma_{n,l}}$. Теперь покажем, что $\overline{\exp[-(1/2)\gamma_i]} \leqslant 1$, откуда следует согласно (6Б.14), что

$$\Pr[\min \Gamma_{n,l} \leqslant u] \leqslant e^{u/2}. \tag{6Б.15}$$

Аналогично (6Б.10)—(6Б.12) имеем

$$\overline{\exp\left(-\frac{1}{2}\gamma_i\right)} = \sum \prod_{a=1}^{v} \left\{ Q\left(x_i^{(a)}\right) P\left(y_i^{(a)} \mid x_i^{(a)}\right) \left[\frac{P\left(y_i^{(a)} \mid x_i^{(a)}\right)}{\omega\left(y_i^{(a)}\right)} \right]^{-1/2} e^{B/2} \right\} =$$

$$= \left[\sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) \sqrt{\omega(j) P(j \mid k)}\, e^{B/2} \right]^{v} \leqslant \exp\left\{ -\frac{v}{2} [E_0(1, \mathbf{Q}) - B] \right\}. \tag{6Б.16}$$

На последнем шаге мы воспользовались неравенством Коши аналогично тому, как это сделано в (6Б.11) и (6Б.12). Так как по предположению $B \leqslant E_0(1, \mathbf{Q})$, то $\overline{\exp(-{}^1\!/_2\gamma_i)} \leqslant 1$ и (6Б.15) справедливо.

Случайная величина $\Gamma'_l - \Gamma_l - \alpha$ статистически независима от случайной величины $\min \Gamma_{n,l}$ и, следовательно,

$$\Pr[\Gamma'_l - \Gamma_l - \alpha - \min \Gamma_{n,l} \geqslant 0] = \sum_u \Pr[\Gamma'_l - \Gamma_l - \alpha = u] \Pr[\min \Gamma_{n,l} \leqslant u], \tag{6Б.17}$$

где суммирование по u проводится по всем дискретным значениям, принимаемым случайной величиной $\Gamma'_l - \Gamma_l - \alpha$. Ограничивая сверху (6Б.17) с помощью (6Б.15), имеем

$$\Pr[\Gamma'_l - \Gamma_l - \alpha - \min \Gamma_{n,l} \geqslant 0] \leqslant \sum_u \Pr[\Gamma'_l - \Gamma_l - \alpha = u]\, e^{u/2} =$$

$$= \overline{\exp[{}^1\!/_2(\Gamma'_l - \Gamma_l - \alpha)]} = \tag{6Б.18}$$

$$= \overline{\exp\left\{ \frac{1}{2} \left[\sum_{i=1}^{l} (\gamma'_i - \gamma_i) - \alpha \right] \right\}} = \tag{6Б.19}$$

$$= \exp\left(-\frac{\alpha}{2}\right) \overline{\left[\exp\left(\frac{\gamma'_l - \gamma_l}{2} \right) \right]}^{l} = \tag{6Б.20}$$

*) См., например, книгу Кокса и Миллера (1965).

$$= \exp\left[-\frac{\alpha}{2} - \frac{\nu l}{2} E_0(1, \mathbf{Q})\right] \leqslant \tag{6Б.21}$$

$$\leqslant \exp\left\{-\frac{\alpha}{2} - \frac{\nu l}{2}[E_0(1, \mathbf{Q}) + B]\right\}. \tag{6Б.22}$$

При выводе (6Б.20) использована статистическая независимость γ_i' — γ_i при различных значениях i, в (6Б.21) использовалось соотношение (6Б.9), а в (6Б.22) — предположение, что $B \leqslant E_0(1, \mathbf{Q})$. Наконец, подставляя (6Б.22) и (6Б.13) в (6Б.5), имеем

$$\Pr[\Gamma'_{m(l)} \geqslant \Gamma_{min} + \alpha] \leqslant (l+1) \exp\left\{-\frac{\alpha}{2} - \frac{\nu l}{2}[E_0(1, \mathbf{Q}) + B]\right\}, \tag{6Б.23}$$

что завершает доказательство леммы 6.9.3. |

Случайные блуждания и доказательство леммы 6Б.1

Последовательность случайных величин $S_0 = 0$, S_1, S_2, ... называется *случайным блужданием*, если при любом целом $n > 0$ можно представить S_n в виде

$$\sum_{i=1}^{n} z_i,$$

где z_1, z_2, ... — последовательность независимых одинаково распределенных случайных величин. Пусть $g(r) = \overline{\exp(r, z_i)}$ — производящая функция моментов каждой из величин z_i и пусть $P(z)$ — распределение вероятностей этих случайных величин. Для простоты записи будем предполагать, что z_i — дискретные случайные величины, но результаты можно легко распространить на произвольные случайные величины, для которых $g(r)$ существует в окрестности $r = 0$.

При любом заданном r определим перекошенные распределения вероятностей $Q_r(z)$ с помощью равенств

$$Q_r(z) = \frac{P(z)\, e^{rz}}{g(r)}. \tag{6Б.24}$$

Пусть $z_{i,r}$; $i = 1, 2, \ldots$ — последовательность статистически независимых случайных величин, каждая из которых принимает значения с вероятностями $Q_r(z)$, и рассмотрим «перекошенное» случайное блуждание, для которого $S_{0,r} = 0$ и

$$S_{n,r} = \sum_{i=1}^{n} z_{i,r}$$

при $n > 0$.

Нас интересует вероятность того, что при таком перекошенном случайном блуждании первый переход ниже некоторой точки $u < 0$ произойдет при данном целом n, и определим вероятность $f_{r,n}(u, v)$ при $v \leqslant u$ с помощью равенства

$$f_{r,n}(u, v) = \Pr[S_{l,r} > u;\ 1 \leqslant l \leqslant n-1;\ S_{n,r} = v]. \tag{6Б.25}$$

Заметим, что $\sum_{v \leqslant u}^{\infty} f_{r,n}(u, v)$ является вероятностью того, что при перекошенном блуждании первое достижение значения, меньшего или равного u, произойдет в момент n. Следовательно, $\sum_{n=1}^{\infty} \sum_{v \leqslant u} f_{r,n}(u, v)$ является вероятностью того, что при перекошенном случайном блуждании когда-нибудь будет достигнуто значение,

меньшее или равное u. Наконец, при $r = 0$ эти вероятности относятся к первоначальному случайному блужданию.

Теперь предположим, что $a_1, \ldots, a_n$ являются последовательностью значений, которые могут принимать случайные величины $z_1, \ldots, z_n$. Из (6Б.24) получим

$$\Pr[z_{1,r}=a_1, \ \ldots \ , z_{n,r}=a_n] = \frac{\Pr[z_1=a_1, \ldots, z_n=a_n] \exp\left(r \sum_{i=1}^{n} a_i\right)}{[g(r)]^n}. \qquad (6Б.26)$$

Следовательно,

$$\Pr\left[S_{1,r}=a_1, \ldots, S_{n,r}=\sum_{i=1}^{n} a_i\right] =$$

$$= \frac{\Pr\left[S_1=a_1, \ldots, S_n=\sum_{i=1}^{n} a_i\right] \exp\left(r \sum_{i=1}^{n} a_i\right)}{[g(r)]^n}. \qquad (6Б.27)$$

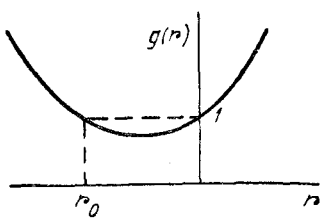


Рис. 6Б.1. График функции $g(r)=\overline{\exp(rz_i)}$ для $\overline{z_i}>0$, где z_i принимает как положительные, так и отрицательные значения.

Если просуммировать (6Б.27) по всем последовательностям $a_1, \ldots, a_n$, для которых $\sum_{i=1}^{l} a_i > u$ при $1<l<n-1$ и $\sum_{i=1}^{n} a_i = v$, то, получим, что

$$f_{r,n}(u, v) = \frac{f_{0,n}(u, v)\, e^{rv}}{[g(r)]^n}. \qquad (6Б.28)$$

Предположим теперь, что r выбрано так, что $\overline{z_{i,r}} < 0$. Тогда, используя закон больших чисел, получаем

$$\lim_{n\to\infty} \Pr[S_{n,r} \leqslant u] = 1$$

и с вероятностью 1 перекошенное случайное блуждание в конце концов достигает значений, меньших или равных любому $u < 0$. Поэтому, суммируя обе части (6Б.28) по $v \leqslant u$ и по n, получаем

$$1 = \sum_{n=1}^{\infty} \sum_{v \leqslant u} f_{0,n}(u, v)\, e^{rv} [g(r)]^{-n}. \qquad (6Б.29)$$

Этот результат известен как тождество Вальда для блужданий с одним барьером u (величина u называется барьером, так как в теории случайных блужданий часто заканчивают блуждание после первого пересечения данного значения).

Здесь мы хотим использовать этот результат для вывода границы сверху величины

$$\Pr[S_{min} \leqslant u] = \sum_{n=1}^{\infty} \sum_{v \leqslant u} f_{0,n}(u, v).$$

Предположим, что $\overline{z_i} > 0$, в противном случае $\Pr[S_{min} \leqslant u] = 1$. Также предположим, что z_i принимает по меньшей мере одно отрицательное значение с отличной от нуля вероятностью, в противном случае $\Pr[S_{min} < u) = 0$. График функции $g(r)$ представлен на рис. 6Б.1.В частности, функция $g(r)$ выпукла $\cup$ и стремится к бесконечности как при $r \to \infty$, так и при $r \to -\infty$. Поэтому уравнение $g(r) = 1$ имеет два решения: одно при $r = 0$ и одно при $r = r_0$, где $r_0 < 0$. Так как

$$\overline{z_{i,r}} = \frac{1}{g(r)} \frac{dg(r)}{dr},$$

то отсюда видно, что $\overline{z_{i,r}} < 0$ при $r = r_0$; применяя (6Б.29), получаем

$$1 = \sum_{n=1}^{\infty} \sum_{v \leqslant u} f_{0,n}(u, v) e^{r_0 v} \geqslant \tag{6Б.30}$$

$$\geqslant e^{r_0 u} \sum_{n=1}^{\infty} \sum_{v \leqslant u} f_{0,n}(u, v), \tag{6Б.31}$$

$$\Pr[S_{min} \leqslant u] \leqslant e^{-r_0 u}. \tag{6Б.32}$$

Так как $g(r) \leqslant 1$ лишь при $r_0 \leqslant r \leqslant 0$, то можно далее получить границу сверху в (6Б.32) вида e^{-ru} для любых $r \leqslant 0$, таких, что $g(r) \leqslant 1$. Наконец, заметив, что $\Pr[S_{min} \leqslant u] \leqslant \exp(-ru)$ также при $u \geqslant 0$, завершим доказательство леммы 6Б.1. Граница (6Б.32) является довольно точной. Чтобы увидеть это, предположим, что существует минимальное значение, например z_{min}, которое может принимать случайная величина z. Тогда v в (6Б.30) должна удовлетворять неравенствам $u - z_{min} < v \leqslant u$ и поэтому $\Pr[S_{min} \leqslant u]$ можно ограничить снизу с помощью неравенства

$$\Pr[S_{min} \leqslant u] \geqslant \exp[-r_0(u - z_{min})]. \tag{6Б.33}$$

Можно также показать (см. Феллер (1966), т. 2, гл. XII, § 5), что асимптотически при больших u $\Pr[S_{min} \leqslant u] \sim Ce^{-r_0 u}$, где C не зависит от u.

7

ДИСКРЕТНЫЕ ПО ВРЕМЕНИ КАНАЛЫ БЕЗ ПАМЯТИ

7.1. ВВЕДЕНИЕ

В гл. 4 и 5 для дискретных каналов без памяти были доказаны теорема кодирования и ее обращение. Под дискретными понимались каналы, в которых вход и выход были временными последовательностями букв, выбранных из конечного алфавита. В этой главе результаты гл. 4 и 5 будут обобщены на случай, когда входные и выходные алфавиты бесконечны. Простейшим и наиболее важным примером такого обобщения является случай, когда входной и выходной алфавиты образованы множествами действительных чисел и канал статистически описывается условной плотностью вероятности $p_{Y|X}(y|x)$. Предполагается, что канал является каналом без памяти, т. е. что если $\mathbf{x} = (x_1, \ldots, x_N)$ — последовательность N входных символов, то для соответствующей выходной последовательности $\mathbf{y} = (y_1, \ldots, y_N)$ условная плотность вероятности задается равенством

$$p_N(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} p_{Y|X}(y_n|x_n). \tag{7.1.1}$$

Другими словами, каждая выходная буква статистически зависит только от соответствующей входной буквы и эта статистическая зависимость остается неизменной во времени (т. е. не меняется от положения в последовательности).

В общем случае дискретный по времени канал без памяти задается произвольным входным пространством X, произвольным выходным пространством Y, и для каждого элемента x входного пространства условной вероятностной мерой*) на выходе $P_{Y|X}$. Входом канала является последовательность букв входного пространства, выходом — последовательность букв выходного пространства и каждая выходная буква зависит вероятностно только от соответствующей входной буквы; эта зависимость задается вероятностной мерой $P_{Y|X}$ (т. е. так же как и в гл. 4 для любого заданного n величины x_n и y_n условно не зависят от всех других входов и выходов).

Развиваемый здесь общий подход к изучению таких каналов состоит в том, чтобы ограничиться использованием конечного множества

*) Для полной корректности следует также определить события на выходе, замкнутые относительно операции дополнения и относительно конечных или счетных объединений и пересечений. Для каждого x входного пространства $P_{|Y}$ должно быть вероятностной мерой на этом классе выходных событий.

букв входного алфавита, скажем $a_1, a_2, \ldots, a_K$, и разбиением выходного пространства на конечное множество непересекающихся событий, скажем $B_1, \ldots, B_J$, объединение которых образует все выходное пространство. Тогда в принципе можно построить квантующее устройство, для которого входом в каждый момент времени является выход канала y, а выходом — событие B_j, содержащее y. Канал и квантующее устройство в совокупности образуют дискретный канал без памяти с переходными вероятностями $P_{Y|X}(B_j|a_k)$. Изучение первоначального канала будет основано на рассмотрении поведения всех таким образом полученных дискретных каналов без памяти. Такой подход имеет преимущество в том, что он тесно связан со способами физического использования канала и в легкости аналитического исследования.

При изучении таких каналов возникает новая проблема, связанная с ограничениями на входы канала. Рассмотрим канал примера 4 гл. 2, в котором выход канала образован суммой входа и независимой гауссовой случайной величины

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-x)^2}{2\sigma^2}\right]. \tag{7.1.2}$$

Согласно (2.4.36), если вход—гауссовская случайная величина с дисперсией $\mathscr{E}$, то

$$I(X;Y) = \frac{1}{2}\log\left(1+\frac{\mathscr{E}}{\sigma^2}\right). \tag{7.1.3}$$

При $\mathscr{E}$ произвольно большом количество информации $I(X;Y)$ становится сколь угодно большим и, выбирая сколь угодно большое множество входов, разнесенных сколь угодно далеко по амплитуде, видим, что, по существу, без кодирования может быть достигнута произвольно высокая скорость передачи при произвольно малой вероятности ошибки. Однако если рассмотреть эти входы как выборки переданных непрерывных сигналов, то можно заметить, что этот результат получается при использовании сколь угодно большой мощности. Для этого канала и для обширного класса каналов, связанных с этим примером, мы можем получить физически важные и математически интересные результаты, если зададим ограничения на входы канала.

Простейшим при нашем подходе типом ограничений, накладываемых на вход канала, является ограничение на амплитуду: входной алфавит просто ограничен значениями x, меньшими или равными некоторому фиксированному числу A. Если входное пространство определено как интервал от $-A$ до $+A$, то это ограничение можно не учитывать. Более общий и важный тип ограничения — ограничение на энергию. Этот тип ограничения будет точно описан позднее, однако сущность его состоит в том, что вход канала должен иметь среднеквадратическое значение, не большее некоторого фиксированного числа $\mathscr{E}$. Этот тип ограничений касается не входного пространства, а относительных частот, с которыми различные входы могут быть использованы. Как будет показано в следующей главе, ограничение на энергию является естественным при представлении непрерывного по времени

канала в виде параллельного соединения каналов с дискретным временем.

В следующих параграфах сначала будет проведено исследование каналов без ограничений (или при наличии амплитудных ограничений) на входы, затем каналов с ограничениями на входы и затем изучено несколько примеров, в том числе важный пример канала с аддитивным гауссовым шумом.

7.2. ОТСУТСТВИЕ ОГРАНИЧЕНИЙ НА ВХОДЕ

Как было показано, если ограничиться конечным множеством букв на входе и произвести разбиения на выходе, то общий дискретный по времени канал без памяти можно использовать как дискретный канал без памяти. Таким образом, любая вероятность ошибки, которая может быть достигнута с помощью кодирования в любом таком дискретном канале без памяти, может быть достигнута и в общем канале при использовании его как соответствующего дискретного канала.

Для заданного дискретного по времени канала без памяти пусть X_d — конечное множество входных букв канала $(a_1, \ldots, a_K)$ с вероятностями $Q(a_1), \ldots, Q(a_K)$. Пусть Y_p — разбиение выходов канала на события $B_1, \ldots, B_J$. Совместный ансамбль $X_d Y_p$ имеет совместное распределение вероятностей $Q(a_K)\, P_{Y|X}(B_j|a_k)$ и среднюю взаимную информацию (в натуральных единицах)

$$I(X_d; Y_p) = \sum_{k=1}^{K} \sum_{j=1}^{J} Q(a_k)\, P_{Y|X}(B_j|a_k) \ln \frac{P_{Y|X}(B_j|a_k)}{\sum_{i=1}^{K} Q(a_i)\, P_{Y|X}(B_j|a_i)}. \tag{7.2.1}$$

Определим функцию $E_0(\rho, X_d, Y_p)$ с помощью равенства

$$E_0(\rho, X_d, Y_p) = -\ln \sum_{j=1}^{J} \left[\sum_{k=1}^{K} Q(a_k)\, P_{Y|X}(B_j|a_k)^{1/(1+\rho)} \right]^{1+\rho}. \tag{7.2.2}$$

Согласно теореме (5.6.2) существует блоковый код длины N с $M = e^{NR}$ кодовыми словами, для которого при переходных вероятностях канала $P_{Y|X}(B_j|a_k)$ вероятность ошибки удовлетворяет неравенству $P_e \leqslant \exp\{-N[E_0(\rho, X_d, Y_p) - \rho R]\}$ для всех ρ, $0 \leqslant \rho \leqslant 1$. *Это приводит нас к определению показателя экспоненты случайного кодирования для данного дискретного по времени канала без памяти*

$$E_r(R) = \sup [E_0(\rho, X_d, Y_p) - \rho R]. \tag{7.2.3}$$

Верхняя грань берется по всем конечным выборам входных букв, всем распределениям вероятностей входных букв, всем разбиениям выходного пространства и всем ρ, $0 \leqslant \rho \leqslant 1$. *Аналогично определяется пропускная способность канала (в натуральных единицах)*:

$$C = \sup I(X_d; Y_p), \tag{7.2.4}$$

где верхняя грань определяется как и выше.

Теорема 7.2.1. (Теорема кодирования.) Пусть для дискретного по времени канала без памяти $E_r(R)$ и C определены равенствами (7.2.3) и (7.2.4). Для любых $R \geqslant 0$, $N \geqslant 1$ и $E < E_r(R)$ существует блоковый код длины N с $M = \lceil e^{NR} \rceil$ кодовыми словами, для которого

$$P_e \leqslant \exp(-NE). \tag{7.2.5}$$

Здесь

$$P_e = \sum_{m=1}^{M} \Pr(m) P_{e,m}$$

— средняя вероятность ошибки; $\Pr(m)$ — вероятность передачи m-го кодового слова; $P_{e,m}$ — вероятность ошибки для m-го кодового слова. Кроме того,

$$E_r(R) > 0 \text{ для всех } R,\ 0 \leqslant R < C. \tag{7.2.6}$$

(*Замечания.* Неравенство (7.2.5) утверждает, что для заданных N, R можно либо найти коды, для которых $P_e \leqslant \exp[-NE_r(R)]$, либо, по крайней мере, найти коды, для которых P_e сколь угодно близко к $\exp[-NE_r(R)]$. Альтернативное утверждение состоит в том, что для заданных N, R имеем $\inf P_e \leqslant \exp[-NE_r(R)]$, где нижняя грань берется по всем кодам с заданными N и R.)

Доказательство. Для заданных N, R и $E < E_r(R)$ выберем ρ, X_d и Y_p так, что

$$E_0(\rho, X_d, Y_p) - \rho R \geqslant E. \tag{7.2.7}$$

Это всегда возможно, так как E строго меньше, чем верхняя грань левой части (7.2.7) по ρ, X_d, Y_p. Из теоремы 5.6.2 следует, что для дискретного канала, соответствующего X_d, Y_p, существует код с заданными N и R, для которого

$$P_e \leqslant \exp\{-N[E_0(\rho, X_d, Y_p) - \rho R]\}. \tag{7.2.8}$$

Так как эта вероятность ошибки также может быть достигнута для общего дискретного по времени канала, то из (7.2.7) и (7.2.8) получаем (7.2.5). Для того чтобы установить справедливость (7.2.6), примем $R < C$, выберем число R_1, $R < R_1 < C$, и выберем X_d, Y_p, удовлетворяющие неравенству

$$I(X_d; Y_p) \geqslant R_1. \tag{7.2.9}$$

Из теоремы 5.6.4 следует, что показатель экспоненты случайного кодирования для канала, соответствующего X_d, Y_p, положителен при $R < R_1$, следовательно, $E_r(R)$ также положительно для данного R. |

Использование $E < E_r(R)$ в (7.2.5) вместо $E_r(R)$ вызывает некоторое раздражение, однако следующий пример (рис. 7.2.1) показывает, что этого нельзя избежать. Легко видеть, что при использовании такого канала любой код будет иметь ненулевую вероятность ошибочного декодирования. Вместе с тем при любых N и R, если использовать только входы $L < k \leqslant L + K$, то при достаточно больших L и K вероятность ошибки можно уменьшить до сколь угодно малого

положительного значения. Наконец, с помощью разбиения выходного пространства на множества, сопоставления одного множества разбиения каждому j, $L < j \leqslant L + K$ и одного множества всем остальным выходам, после устремления L и K к ∞ видим, что $C = \infty$, $E_r(R) = \infty$. Таким образом, для этого примера нельзя достичь $P_e \leqslant \exp[-NE_r(R)]$, хотя можно достичь (7.2.5) для любого конечного E.

Теорема 7.2.2. (Обращение теоремы кодирования.) Пусть дискретный стационарный источник с алфавитом объема M имеет энтропию $H_\infty(U)$ и порождает одну букву каждые τ_s секунд. Пусть дискретный по времени канал без памяти имеет пропускную способность C и используется один раз каждые τ_c секунд. Пусть последовательность символов источника длины L связана с адресатом каналом, по которо-

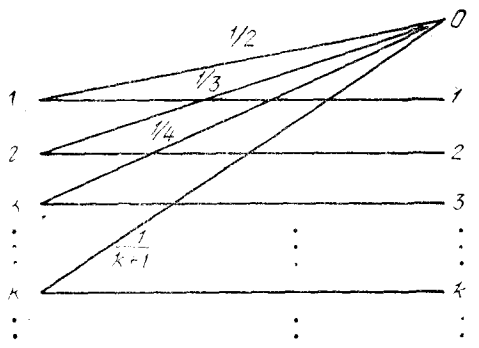


Рис. 7.2.1. Канал со стиранием и с бесконечным алфавитом.

му передается последовательность N символов, где N равно $\lfloor L\tau_s/\tau_c \rfloor$. Тогда в пределе при $L \to \infty$ вероятность ошибки на букву источника $\langle P_e \rangle$ удовлетворяет соотношению

$$\langle P_e \rangle \log(M-1) + \mathcal{H}(\langle P_e \rangle) \geqslant H_\infty(U) - \frac{\tau_s}{\tau_c} C. \qquad (7.2.10)$$

Доказательство. Формулировка этой теоремы совпадает с формулировкой теоремы 4.3.4. Однако она применима к более широкому классу каналов. В доказательстве теоремы 4.3.4 канал рассматривается лишь при установлении следующих двух соотношений:

$$I(\mathbf{U}^L; \mathbf{V}^L) \leqslant I(\mathbf{X}^N; \mathbf{Y}^N), \qquad (7.2.11)$$

$$I(\mathbf{X}^N; \mathbf{Y}^N) \leqslant NC. \qquad (7.2.12)$$

Таким образом, здесь достаточно установить справедливость этих двух соотношений для дискретного по времени канала без памяти.

Доказательство соотношения (7.2.11). Для любого заданного значения L и для любых заданных источника и кодера число кодовых слов конечно. Следовательно, только конечное множество входных букв канала может быть использовано в соответствующем ансамбле

$\mathbf{X}^N$ и, таким образом, $\mathbf{X}^N$ — дискретный ансамбль. Аналогично, любой заданный декодер разбивает пространство $\mathbf{Y}^N$ не более чем на M^L областей декодирования, соответствующих элементам пространства $\mathbf{V}^L$. Обозначим разбиение выходного пространства через $\mathbf{Y}_p^N$. Теорема 4.3.3 устанавливает, что

$$I(\mathbf{U}^L;\ \mathbf{V}^L) \leqslant I(\mathbf{X}^N;\ \mathbf{Y}_p^N). \tag{7.2.13}$$

Далее, по определению (2.5.1),

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) = \sup I(\mathbf{X}^N; \mathbf{Y}_p^N), \tag{7.2.14}$$

где верхняя грань берется по всем разбиениям Y^N. Сравнивая (7.2.13) и (7.2.14), получаем (7.2.11).

Доказательство соотношения (7.2.12). Выражение $I(\mathbf{X}^N; \mathbf{Y}^N)$ может быть переписано следующим образом*):

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) = I(X_1 \ldots X_N;\quad Y_1 \ldots Y_N). \tag{7.2.15}$$

Повторно применяя соотношение (2.2.29) к правой части (7.2.15) и замечая, что все члены конечны, получаем

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) = \sum_{n=1}^{N} I(\mathbf{X}^N;\ Y_n \,|\, Y_1 \ldots Y_{n-1}). \tag{7.2.16}$$

Используя (2.5.4) и (2.5.5), имеем

$$I(\mathbf{X}^N;\ Y_n \,|\, Y_1 \ldots Y_{n-1}) = I(Y_n;\ \mathbf{X}^N Y_1 \ldots Y_{n-1}) - \\ - I(Y_n;\quad Y_1 \ldots Y_{n-1}) \leqslant I(Y_n;\ \mathbf{X}^N Y_1 \ldots Y_{n-1}), \tag{7.2.17}$$

$$I(Y_n;\quad \mathbf{X}^N Y_1 \ldots Y_{n-1}) = I(Y_n;\ X_n) + \\ + I(Y_n;\ Y_1 \ldots Y_{n-1} X_1 \ldots X_{n-1} X_{n+1} \ldots X_N \,|\, X_n). \tag{7.2.18}$$

Согласно теореме 2.3.3 последнее слагаемое в (7.2.18) равно нулю; сопоставляя эти соотношения, получаем

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n;\quad Y_n). \tag{7.2.19}$$

Так как X_n дискретно, то $I(X_n;\ Y_n)$ определяется как верхняя грань по всем разбиениям Y_n и в соответствии с определением C в (7.2.4) имеем $I(X_n;\ Y_n) \leqslant C$, а отсюда непосредственно следует (7.2.12). |

Устанавливая справедливость теоремы кодирования и ее обращения при большой степени общности, представленные выше результаты дают слабое указание на то, как вычислять $E_r(R)$ или C. В § 2.5 было показано, что $I(X_d;\ Y_p)$ не убывает при измельчании разбиения пространства Y. Теперь установим тот же самый результат для $E_0(\rho, X_d, Y_p)$. Пусть Y_p — разбиение с событиями $B_1, \ldots, B_J$ и пусть $Y_{p'}$ — подразбиение Y_p с событиями B_{ij}, где $\bigcup_i B_{ij} = B_j$.

*) Это нетривиальная подстановка, см. обсуждение, следующее за формулой (2.5.3).

По неравенству Минковского (см. задачу 4.15.з) для $\rho \geqslant 0$ имеем

$$\sum_i \left[\sum_k Q(a_k) P_{Y|X}(B_{ij}|a_k)^{1/(1+\rho)}\right]^{1+\rho} \leqslant$$
$$\leqslant \left\{\sum_k Q(a_k) \left[\sum_i P_{Y|X}(B_{ij}|a_k)\right]^{1/(1+\rho)}\right\}^{1+\rho} =$$
$$= \left[\sum_k Q(a_k) P_{Y|X}(B_j|a_k)^{1/(1+\rho)}\right]^{1+\rho}. \qquad (7.2.20)$$

Суммируя обе части (7.2.20) по j и беря минус логарифм от результата, получаем

$$E_0(\rho, X_d, Y_{p'}) \geqslant E_0(\rho, X_d, Y_p). \qquad (7.2.21)$$

С физической точки зрения этот результат не удивителен. При более тонком разбиении выхода канала декодер имеет большую информацию о принятой последовательности и вероятность ошибочного декодирования будет меньше.

Если выходное пространство канала — действительная прямая и канал описывается плотностью вероятности $p_{Y|X}(y|x)$, то выходное пространство можно разбить на интервалы, подразбить интервалы на все более мелкие интервалы и в пределе получить

$$E_0(\rho, X_d, Y) = -\ln \int_{-\infty}^{\infty} dy \left[\sum_k Q(a_k) P_{Y|X}(y|a_k)^{1/(1+\rho)}\right]^{1+\rho}. \qquad (7.2.22)$$

То что правая часть (7.2.22) действительно является верхней гранью $E_0(\rho, X_d, Y_p)$ по *всем* разбиениям Y, следует из тех же соображений, которые были использованы при рассмотрении (7.2.20); при этом применяется интегральная форма неравенства Минковского*)

$$\int_{y \in B_j} dy \left[\sum_k Q(a_k) p_{Y|X}(y|a_k)^{1/(1+\rho)}\right]^{1+\rho} \leqslant$$
$$\leqslant \left\{\sum_k Q(a_k) \left[\int_{y \in B_j} p_{Y|X}(y|a_k)\, dy\right]^{1/(1+\rho)}\right\}^{1+\rho} =$$
$$= \left\{\sum_k Q(a_k) P_{Y|X}(B_j|a_k)^{1/(1+\rho)}\right\}^{1+\rho}. \qquad (7.2.23)$$

Суммируя обе части по j, получаем желаемый результат.

Это сводит проблему нахождения $E_r(R)$ для канала с переходной плотностью вероятности к проблеме вычисления выражения

$$E_r(R) = \sup_{0 \leqslant \rho \leqslant 1} \sup_{X_d} [E_0(\rho, X_d, Y) - \rho R], \qquad (7.2.24)$$

где $E_0(\rho, X_d, Y)$ определено (7.2.22). Очень мало можно сказать в общем случае о нахождении верхней грани $E_0(\rho, X_d, Y)$ по всем дискретным входам. В некоторых случаях (см. задачу 7.5) $E_0(\rho, X_d, Y)$ достигает максимума на дискретных входах и в этом случае условия

*) См. Харди, Литльвуд и Полиа (1934), теорема 201.

(5.6.37) и (5.6.38), обобщенные на непрерывный выход, указывают, что максимум достигается. В других случаях (см. задачу 7.4) верхняя грань по $E_0(\rho, X_d, Y)$ достигается в пределе, когда X_d стремится к плотности вероятности на входном пространстве, и, используя (5.6.37) и (5.6.38), эту оптимальную плотность вероятности можно иногда найти*). В задаче 7.3 обращено внимание на одну особенность функции $E_0(\rho, X_d, Y)$, где показано, что

$$\sup_{X_d} E_0(\rho, X_d, Y)$$

может быть разрывна по ρ в точке $\rho = 0$. Это указывает на существование каналов с бесконечной пропускной способностью, но с конечным показателем экспоненты случайного кодирования.

Рассмотренная в § 5.7 граница случайного кодирования для процедуры с выбрасыванием обобщается на общие дискретные по времени каналы без памяти таким же образом, как и граница случайного кодирования.

Определим следующие величины:

$$E_x(\rho, X_d, Y_p) = \\ = -\rho \ln \sum_{k,\,i} Q(a_k) Q(a_i) \left[\sum_j \sqrt{P_{Y|X}(B_j|a_k) P_{Y|X}(B_j|a_i)} \right]^{1/\rho}, \quad (7.2.25)$$

$$E_{ex}(R') = \sup [-\rho R' + E_x(\rho, X_d, Y_p)], \quad (7.2.26)$$

где верхняя грань берется по $\rho \geqslant 1$, X_d и Y_p. Тогда для любых $R' \geqslant 0$, $N \geqslant 1$ и $E < E_{ex}(R')$ существует блоковый код длины N с $M = \lceil {}^1/_4 \, e^{NR} \rceil$ кодовыми словами, для которого

$$P_{e,\,m} \leqslant \exp(-NE) \text{ для всех } m, \ 1 \leqslant m \leqslant M. \quad (7.2.27)$$

Доказательство этого неравенства аналогично доказательству теоремы 7.2.1.

7.3. ОГРАНИЧЕНИЯ НА ВХОДЕ

В § 7.1 мы определили ограничение на энергию, как ограничение на среднеквадратическое значение входных сигналов. Здесь рассматривается несколько более общая задача. Пусть X — входное пространство дискретного по времени канала без памяти и пусть $f(x)$ — действительная функция, определенная на входных буквах. Ограничение при использовании канала сводится к тому, что математическое ожидание $f(x)$ меньше или равно некоторому фиксированному значению $\mathscr{E}$. Если X — множество действительных чисел и $f(x) = x^2$, то получается описанное выше ограничение на энергию. В более общем случае, например, X может быть классом функций $x(t)$, а $f(x)$ может быть $\int x^2(t)\,dt$ или любым другим функционалом от $x(t)$.

С точки зрения теории кодирования следует более точно объяснить, что означает ограничение на математическое ожидание $f(x)$. Одна

*) В этом случае достаточность условий (5.6.37) и (5.6.38) все еще имеет место, хотя доказательства гл. 4 и 5 не проходят.

из разумных интерпретаций этого ограничения состоит в том, что каждое кодовое слово удовлетворяет этому ограничению, т. е. для каждого кодового слова $\mathbf{x}_m = (x_{m,1}, \ldots, x_{m,N})$, требуется, чтобы

$$\sum_{n=1}^{N} f(x_{n,m}) \leqslant N\mathscr{E}.$$

Другая разумная интерпретация состоит в задании вероятностной меры на сообщениях $\Pr(m)$ и требовании, чтобы

$$\sum_{m=1}^{M} \Pr(m) \sum_{n=1}^{N} f(x_{n,m}) \leqslant N\mathscr{E}.$$

Заметим, что класс кодов, для которого каждое кодовое слово удовлетворяет ограничению, содержится в классе кодов, для которых удовлетворяется ограничение при усреднении по кодовым словам. Таким образом, любая вероятность ошибочного декодирования, которая может быть достигнута на некотором коде первого класса, может быть также достигнута на коде (в частности, на том же коде) последнего класса. Обратно, любая нижняя граница вероятности ошибки последнего класса также будет нижней границей первого класса. Поэтому теорема кодирования будет доказываться при ограничении на *каждое* кодовое слово, а ее обращение — когда удовлетворяется ограничение только при усреднении по множеству кодовых слов. Таким образом, каждая теорема будет применима к обоим случаям, и будет показано, что нет существенной разницы в том, какой из двух случаев рассматривается. Начнем с изучения обращения теоремы кодирования, так как она почти не отличается от соответствующей теоремы для случая без ограничений.

Используя обозначения последнего параграфа, пропускную способность дискретного по времени канала без памяти с ограничением на входе $\overline{f(x)} \leqslant \mathscr{E}$, определим следующим образом:

$$C = \sup I(X_d;\ Y_p), \tag{7.3.1}$$

где верхняя грань берется по всем разбиениям выходного пространства, всем дискретным множествам входов $(a_1, \ldots, a_K)$ и всем распределениям вероятностей $Q(a_1), \ldots, Q(a_K)$, удовлетворяющим ограничению

$$\sum_{k=1}^{K} Q(a_k) f(a_k) \leqslant \mathscr{E}. \tag{7.3.2}$$

Заметим, что при этом определении функция $f(x)$ и присутствующее в ограничении значение $\mathscr{E}$ рассматриваются как неотъемлемые части описания канала.

Теорема 7.3.1. (Обращение теоремы кодирования.) Пусть дискретный стационарный источник с алфавитом объема M имеет энтропию $H_\infty(U)$ и порождает одну букву каждые τ_s секунд. Пусть дискретный по времени канал без памяти с ограничением на входе $\overline{f(x)} \leqslant \mathscr{E}$ имеет пропускную способность C, определенную (7.3.1). Пусть последовательность символов источника длины L связана с адресатом каналом, по

которому передается последовательность N символов $x_1, \ldots, x_N$, где N равно $\lfloor L\tau_s/\tau_c \rfloor$. Пусть ансамбль $\mathbf{X}^N$, образовавшийся на входе канала, удовлетворяет ограничению

$$\sum_{n=1}^{N} \overline{f(x_n)} \leqslant N\mathscr{E}. \tag{7.3.3}$$

Тогда, в пределе при $L \to \infty$, вероятность ошибки на букву источника $\langle P_e \rangle$ удовлетворяет соотношению

$$\langle P_e \rangle \log(M-1) + \mathscr{H}(\langle P_e \rangle) \geqslant H_\infty(U) - \frac{\tau_s}{\tau_c} C. \tag{7.3.4}$$

(*Замечания.* Заметим, что условие (7.3.3) означает, что ограничение удовлетворяется при усреднении по кодовым словам. Допускается нарушение ограничения для отдельных кодовых слов, а также для отдельных букв в последовательности N посылок по каналу. Другими словами, при ограничении на энергию допускается распределение энергии, приходящейся на блок, любым желаемым образом между N посылками по каналу.)

Доказательство. Доказательство теоремы 7.2.2 применимо здесь в той части, где оно касается установления справедливости неравенств

$$I(\mathbf{U}^L;\ \mathbf{V}^L) \leqslant I(\mathbf{X}^N;\ \mathbf{Y}^N) \text{ и } I(\mathbf{X}^N;\ \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n;\ Y_n).$$

В конце доказательства теоремы 7.2.2 было показано, что $I(X_n;\ Y_n) \leqslant C$ для каждого n. Этот результат не сохраняется здесь, поскольку для каждой отдельной посылки по каналу не требуется, чтобы удовлетворялись ограничения. Для того чтобы завершить доказательство, надо показать, что

$$\sum_{n=1}^{N} I(X_n;\ Y_n) \leqslant NC. \tag{7.3.5}$$

Пусть $a_1, \ldots, a_K$ — множество входных букв канала, используемых для какого-либо кода, $Q_n(a_k)$ — вероятность появления входной буквы a_k при n-м использовании канала. Согласно (7.3.3) имеем

$$\sum_{n=1}^{N} \sum_{k=1}^{K} Q_n(a_k) f(a_k) \leqslant N\mathscr{E}. \tag{7.3.6}$$

Введем вероятности $Q(a_k)$

$$Q(a_k) = \frac{1}{N} \sum_{n=1}^{N} Q_n(a_k). \tag{7.3.7}$$

Подставляя (7.3.7) в (7.3.6), получаем

$$\sum_{k=1}^{K} Q(a_k) f(a_k) \leqslant \mathscr{E}. \tag{7.3.8}$$

Пусть $I(X;\ Y)$ — средняя взаимная информация между входом и выходом канала при использовании букв $a_1, \ldots, a_K$ с вероятностями $Q(a_1), \ldots, Q(a_K)$. Так как (7.3.8) эквивалентно (7.3.2), то

$$I(X;\ Y) \leqslant C. \tag{7.3.9}$$

В теореме 4.4.2 было показано, что средняя взаимная информация в канале с дискретным входом является выпуклой $\frown$ функцией входных вероятностей*). Из (4.4.5) (если положить θ_n равным $1/N$) следует

$$\sum_n \frac{1}{N} I(X_n;\ Y_n) \leqslant I(X;\ Y). \tag{7.3.10}$$

Сочетая (7.3.9) и (7.3.10), получаем (7.3.5), что и доказывает теорему. |

Перед тем как приступить к формулировке и доказательству теоремы кодирования для дискретного по времени канала без памяти с ограничением на входе, следует рассмотреть влияние ограничения на входе на дискретный канал без памяти. Так же как и в гл. 5, обозначим через $P(j\,|\,k)$ переходные вероятности для дискретного канала без памяти с входным алфавитом $0, \ldots, K-1$ и выходным алфавитом $0, \ldots, J-1$. Пусть $f(k)$ — функция, определенная на входных буквах; рассмотрим класс кодов, для которых каждое кодовое слово $\mathbf{x} = (x_1, \ldots, x_N)$ удовлетворяет ограничению

$$\sum_{n=1}^{N} f(x_n) \leqslant N\mathscr{E}, \tag{7.3.11}$$

где $\mathscr{E}$ — заданная постоянная.

Построим теперь ансамбль кодов, в котором каждое кодовое слово удовлетворяет (7.3.11). Пусть $Q(k)$ — вероятности входных букв, удовлетворяющие неравенству

$$\sum_k Q(k) f(k) \leqslant \mathscr{E}. \tag{7.3.12}$$

Пусть $Q_N(\mathbf{x})$ — распределение вероятностей на последовательностях N входов канала, задаваемое соотношением

$$Q_N(\mathbf{x}) = \mu^{-1} \varphi(\mathbf{x}) \prod_{n=1}^{N} Q(x_n), \tag{7.3.13}$$

где

$$\varphi = (\mathbf{x}) = \begin{cases} 1 \quad \text{для } N\mathscr{E} - \delta < \sum_n f(x_n) \leqslant N\mathscr{E}, \\ 0 \text{ в остальных случаях;} \end{cases} \tag{7.3.14}$$

$$\mu = \sum_{\mathbf{x}} \varphi(\mathbf{x}) \prod_{n=1}^{N} Q(x_n) \tag{7.3.15}$$

и δ — произвольное положительное число, определенное ниже.

*) Утверждение теоремы 4.4.2 относится к каналам с дискретным выходом, однако легко видеть, что доказательство также применимо к случаю произвольного выхода.

Можно заметить, что $Q_N(\mathbf{x})$ является условной вероятностью последовательности $\mathbf{x}$ при условии $N\mathscr{E}-\delta<\Sigma f(x_n)\leqslant N\mathscr{E}$, если буквы слова $\mathbf{x}$ выбраны независимо с вероятностями $Q(k)$. Величина μ — это вероятность того, что при таком независимом выборе последовательностей удовлетворяется условие $N\mathscr{E}-\delta<\Sigma f(x_n)\leqslant N\mathscr{E}$.

Рассмотрим ансамбль кодов с M кодовыми словами блоковой длины N, в котором кодовые слова выбраны независимо с вероятностями $Q_N(\mathbf{x})$. Из теоремы 5.6.1 следует, что для каждого сообщения, $1\leqslant m\leqslant M$, вероятность ошибки, усредненная по ансамблю кодов, ограничена сверху при всех ρ, $0\leqslant\rho\leqslant 1$, выражением

$$P_{e,m}\leqslant(M-1)^{\rho}\sum_{\mathbf{y}}\left[\sum_{\mathbf{x}}Q_N(\mathbf{x})\,P_N(\mathbf{y}\mid\mathbf{x})^{1/(1+\rho)}\right]^{1+\rho}. \qquad (7.3.16)$$

Неравенство (7.3.16) не очень удобно по форме, так как при больших N трудно иметь дело с входящими в него суммами. Если оценить сверху $Q_N(\mathbf{x})$ и упростить получившееся выражение, то можно получить более удобный результат. Для любого $r\geqslant 0$ можно построить верхнюю границу для $\varphi(\mathbf{x})$ [см. (7.3.14)]:

$$\varphi(\mathbf{x})\leqslant\exp\{r[\Sigma f(x_n)-N\mathscr{E}+\delta]\}. \qquad (7.3.17)$$

Неравенство (7.3.17), очевидно, справедливо, когда $\varphi(\mathbf{x})=0$. Когда $\varphi(\mathbf{x})=1$, выражение, стоящее в скобках в (7.3.17), неотрицательно и правая часть (7.3.17) больше или равна 1. Сочетая (7.3.13) и (7.3.17), имеем для любого $r\geqslant 0$

$$Q_N(\mathbf{x})\leqslant\mu^{-1}e^{r\delta}\prod_{n=1}^{N}Q(x_n)\,e^{r[f(x_n)-\mathscr{E}]}. \qquad (7.3.18)$$

Мажорируя $Q_N(\mathbf{x})$ в (7.3.16) выражением (7.3.18) и повторяя выкладки, проведенные при переходе от (5.6.11) к (5.6.13), получаем

$$P_{e,m}\leqslant\left[\frac{e^{r\delta}}{\mu}\right]^{1+\rho}\exp\{-N[E_0(\rho,\ \mathbf{Q},\ r)-\rho R]\}, \qquad (7.3.19)$$

$$E_0(\rho,\ \mathbf{Q},\ r)=-\ln\sum_j\left\{\sum_k Q(k)\,e^{r[f(k)-\mathscr{E}]}P(j\mid k)^{1/(1+\rho)}\right\}^{1+\rho}, \qquad (7.3.20)$$

где M и R связаны равенством $M=\lceil e^{NR}\rceil$. Используя соображения следствия 2 теоремы 5.6.2, можно заметить, что существует также код, для которого при любом m, $1\leqslant m\leqslant M$, и любом ρ, $0\leqslant\rho\leqslant 1$,

$$P_{e,m}\leqslant\left(\frac{2e^{r\delta}}{\mu}\right)^{2}\exp\{-N[E_0(\rho,\ \mathbf{Q},\ r)-\rho R]\}. \qquad (7.3.21)$$

Приведенная выше граница записана с помощью некоторых произвольных параметров, а именно $0\leqslant\rho\leqslant 1$, $r\geqslant 0$, $\mathbf{Q}$ и $\delta\geqslant 0$. Прежде чем попытаться оптимизировать выражение по этим параметрам, полезно исследовать поведение границы при $r=0$ и сколь угодно большом δ. В этом случае (7.3.19) упрощается следующим образом:

$$P_{e,m}\leqslant\left(\frac{1}{\mu}\right)^{1+\rho}\exp\{-N[E_0(\rho,\ \mathbf{Q})-\rho R]\}, \qquad (7.3.22)$$

где $E_0(\rho, \mathbf{Q}) - \rho R$ — показатель экспоненты, знакомый по теореме 5.6.2, за исключением множителя $(1/\mu)^{1+\rho}$; (7.3.22) эквивалентно теореме 5.6.2.

Так как теперь μ — вероятность того, что при независимом выборе букв с распределением $Q(k)$ последовательность x удовлетворяет неравенству

$$\sum_{n=1}^{N} f(x_n) \leqslant N\mathscr{E},$$

то из центральной предельной теоремы следует, что μ стремится к $^1/_2$ при возрастании N, если $\sum_k Q(k) f(k) = \mathscr{E}$, и μ стремится к 1, если $\sum_k Q(k) f(k) < \mathscr{E}$. Таким образом, множитель $(1/\mu)^{1+\rho}$ не отражается на экспоненциальной зависимости границы от N.

Пропускная способность этого канала в натах задается как частный случай формулы (7.3.1):

$$C = \max \sum_k \sum_j Q(k) P(j \mid k) \ln \frac{P(j \mid k)}{\sum_i Q(i) P(j \mid i)}, \tag{7.3.23}$$

где максимум берется по всем заданиям вероятностей $Q(k)$, удовлетворяющих (7.3.12). Используя рассуждения теоремы 5.6.4, можно показать, что для $\mathbf{Q}$, максимизирующем (7.3.23) при условии (7.3.12), выражение

$$\max_{0 \leqslant \rho \leqslant 1} E_0(\rho, \mathbf{Q}) - \rho R$$

положительно при $R < C$. В итоге получаем, что при $r = 0$ и сколь угодно большом δ и при подходящим образом выбранных ρ и $\mathbf{Q}$ вероятность $P_{e,m}$, как следует из (7.3.21), экспоненциально убывает с N для любого $R < C$.

Как побочный результат приведенного выше рассмотрения находим, что если максимум $E_0(\rho, \mathbf{Q})$ по всем векторам вероятностей $\mathbf{Q}$ достигается на векторе $\mathbf{Q}$, который удовлетворяет ограничению $\Sigma Q(k) f(k) \leqslant \mathscr{E}$, то для канала с ограничением можно достигнуть того же показателя экспоненты, что и для канала без ограничения.

Обратимся теперь к более интересной ситуации, в которой при заданном ρ, максимум $E_0(\rho, \mathbf{Q})$ достигается на $\mathbf{Q}$, не удовлетворяющем ограничению. В этом случае оказывается, что максимум $E_0(\rho, \mathbf{Q}, r)$ при заданных ограничениях достигается на $\mathbf{Q}$, удовлетворяющем $\Sigma f(k) Q(k) = \mathscr{E}$, и на $r > 0$. Наглядно причину этого можно пояснить следующим образом. Предположим, что используется ансамбль кодов, в котором все буквы кодовых слов выбираются независимо и так, что $\Sigma Q(k) f(k) = \mathscr{E}$. Сумма $\sum_{n=1}^{N} f(x_n)$ для большинства кодовых слов будет близка к $N\mathscr{E}$. Однако небольшое число кодовых слов, для которых $\Sigma f(x_n)$ существенно меньше, чем $N\mathscr{E}$, будут представителями ансамбля с меньшим значением $E_0(\rho, \mathbf{Q})$, а следовательно, с много большей ве-

роятностью ошибки. Вероятность ошибки, обусловленная этими немногими словами, определяет границу вероятности ошибки при $r = 0$ [в действительности эти слова не принадлежат ансамблю, однако они появляются в границе, поскольку граница для $Q_N(\mathbf{x})$ имеет вид (7.3.18)]. Смысл выбора $r > 0$ состоит в том, чтобы уменьшить влияние этих немногих плохих слов на границу.

Простейшим способом максимизации $E_0(\rho, \mathbf{Q}, r)$ по r и $\mathbf{Q}$ является нахождение стационарной точки относительно r и $\mathbf{Q}$ при ограничениях

$$\sum_k Q(k) = 1 \quad \text{и} \quad \sum_k Q(k)[f(k) - \mathscr{E}] = 0.$$

Используя λ и γ как множители Лагранжа, находим стационарную точку функции

$$\sum_j \left\{ \sum_k Q(k)\, e^{r[f(k)-\mathscr{E}]} P(j|k)^{1/(1+\rho)} \right\}^{1+\rho} + \\ + \lambda \sum_k Q(k) + \gamma \sum_k Q(k)[f(k) - \mathscr{E}]. \tag{7.3.24}$$

Взяв частные производные по всем $Q(k)$, получаем условия

$$(1+\rho) \sum_j \alpha_j^\rho\, e^{r[f(k)-\mathscr{E}]} P(j|k)^{1/(1+\rho)} + \lambda + \gamma[f(k) - \mathscr{E}] \geqslant 0 \tag{7.3.25}$$

с равенством при $Q(k) > 0$, где α_j определяется формулой

$$\alpha_j = \sum_k Q(k)\, e^{r[f(k)-\mathscr{E}]} P(j|k)^{1/(1+\rho)}. \tag{7.3.26}$$

Неравенство в (7.3.25) соответствует тому, что максимум $E_0(\rho, \mathbf{Q}, r)$ может достигаться на границе, когда некоторые $Q(k) = 0$, точно так же как и в теореме 4.4.1. Взяв частные производные функции (7.3.24) по r, получаем условие

$$(1+\rho) \sum_j \alpha_j^\rho \sum_k Q(k)[f(k) - \mathscr{E}]\, e^{r[f(k)-\mathscr{E}]} P(j|k)^{1/(1+\rho)} = 0. \tag{7.3.27}$$

Умножая (7.3.25) на $Q(k)$ и суммируя по k, находим, что

$$\lambda = -(1+\rho) \sum_j \alpha_j^{1+\rho}.$$

Аналогично, умножив (7.3.25) на $Q(k)[f(k) - \mathscr{E}]$, просуммировав по k и сравнив с (7.3.27), находим, что $\gamma = 0$. Таким образом, (7.3.25) преобразуется в

$$\sum_j \alpha_j^\rho\, e^{r[f(k)-\mathscr{E}]} P(j|k)^{1/(1+\rho)} \geqslant \sum_j \alpha_j^{1+\rho} \tag{7.3.28}$$

для всех k с равенством, если $Q(k) > 0$. Хотя это выше и не было доказано, однако можно показать, что соотношение (7.3.28) определяет множество необходимых и достаточных условий на r и на удовлетворяющий указанным ограничениям вектор $\mathbf{Q}$, которые максимизируют $E_0(\rho, \mathbf{Q}, r)$. Однако более важно то, что можно показать, что получающийся показатель экспоненты дает точное выражение для функции

надежности канала с ограничением при скоростях, лежащих между R_{cr} и C, где R_{cr} определена в § 5.8*).

Для величины μ из (7.3.19) трудно найти хорошую границу, однако она может быть точно оценена для больших N. Предположим, что $\mathbf{Q}$ удовлетворяет соотношению $\Sigma Q(k) f(k) = \mathscr{E}$, и рассмотрим

$$\sum_{n=1}^{N} f(x_n)$$

как сумму независимых случайных величин, выбранных в соответствии с вероятностной мерой $\mathbf{Q}$. Тогда μ — вероятность того, что эта сумма расположена между средним значением и величиной, на δ меньшей среднего значения. Из центральной предельной теоремы следует**), что для фиксированного δ

$$\lim_{N\to\infty} \sqrt{N}\,\mu = \frac{\delta}{\sqrt{2\pi}\sigma_f}, \tag{7.3.29}$$

$$\sigma_f^2 = \sum Q(k)\,[f(k) - \mathscr{E}]^2. \tag{7.3.30}$$

Из (7.3.29) можно увидеть, что для фиксированных r и δ $[e^{r\delta}/\mu]^{1+\rho}$ возрастает с N, как $N^{(1+\rho)/2}$, и, таким образом, этот коэффициент не влияет на экспоненциальную зависимость границы от N.

Граница вероятности ошибки для процедуры с выбрасыванием § 5.7 может быть распространена на случай, когда имеются ограничения на входе, точно так же как и граница случайного кодирования. Рассмотрим (5.7.7), которое справедливо для любого ансамбля кодов, и верхние границы (7.3.18) для $Q_N(\mathbf{x})$ и $Q_N(\mathbf{x}')$. Раскрывая произведения, находим, что для всех $M \geqslant 2$ и $N \geqslant 1$ существует блоковый код длины N с M кодовыми словами, каждое кодовое слово которого $\mathbf{x}_m$ удовлетворяет ограничению

$$\sum_{n=1}^{N} f(x_{n,m}) \leqslant N\mathscr{E},$$

а также удовлетворяет границе

$$P_{e,\,m} \leqslant -N\,[E_x(\rho, \mathbf{Q}, r) - \rho R'], \tag{7.3.31}$$

где

$$E_x(\rho, \mathbf{Q}, r) = -\rho \ln \left\{ \sum_{k=0}^{K-1} \sum_{i=0}^{K-1} Q(k)\,Q(i)\,e^{r\,[f(k)+f(i)-2\mathscr{E}]} \times \right.$$

*) Чтобы доказать это, нужно начать с нижней границы вероятности — ошибки для кодов с фиксированной композицией, полученной Шенноном, Галлагером, Берлекэмпом (1967), см. неравенство (4.16). После оптимизации по композициям, удовлетворяющим ограничению, следует продолжить доказательство, как и в теореме 6 этой работы.

**) Для нерешетчатых распределений (7.3.29) следует из теоремы 1, гл. XVI, § 4, Феллер (1966), т. 2. Для решетчатых распределений равенство (7.3.29) справедливо только лишь, когда δ кратно шагу; это следует из теоремы 3, гл. XV, § 5, Феллер (решетчатая случайная величина f — это случайная величина, которая может быть сведена к целочисленной случайной величине z с помощью преобразования $f = zh + a$; шаг определяется как наибольшее h, для которого это сведение может быть проведено).

$$\times \left(\sum_{j=0}^{J-i} \sqrt{P(j \mid k) P(j \mid i)} \right)^{1/\rho} \Bigg\}, \tag{7.3.32}$$

$$R' = \frac{\ln M}{N} + \frac{2}{N} \ln \frac{2e^{r\delta}}{\mu}. \tag{7.3.33}$$

В приведенных выше выражениях $\rho \geqslant 1$, $r \geqslant 0$ и $\delta > 0$ — произвольны, а μ определено в (7.3.15) и оценено в (7.3.29).

Теперь можно применить эти результаты к произвольному дискретному по времени каналу без памяти с ограничением на входе $f(x) \leqslant \mathscr{E}$. Как в § 7.2, пусть X_d — конечное множество входных букв канала $a_1, \ldots, a_K$ с заданными вероятностями $Q(a_1), \ldots, Q(a_K)$, удовлетворяющими ограничению $\Sigma Q(a_k) f(a_k) \leqslant \mathscr{E}$. Пусть Y_p — разбиение выходного пространства на события $B_1, \ldots, B_J$. Пусть $E_0(\rho, X_d, Y_p, r)$ и $E_x(\rho, X_d, Y_p, r)$ задаются соотношениями

$$E_0(\rho, X_d, Y_p, r) =$$

$$= -\ln \sum_{j=1}^{J} \left[\sum_{k=1}^{K} Q(a_k) e^{r[f(a_k) - \mathscr{E}]} P_{Y \mid X}(B_j \mid a_k)^{1/(1+\rho)} \right]^{1+\rho}, \tag{7.3.34}$$

$$E_x(\rho, X_d, Y_p, r) = -\rho \ln \left\{ \sum_{k=1}^{K} \sum_{i=1}^{K} Q(a_k) Q(a_i) \times \right.$$

$$\left. \times e^{r[f(a_k) + f(a_i) - 2\mathscr{E}]} \left(\sum_{j=1}^{J} \sqrt{P_{Y \mid X}(B_j \mid a_k) P_{Y \mid X}(B_j \mid a_i)} \right)^{1/\rho} \right\}. \tag{7.3.35}$$

Определим для рассматриваемого канала показатель экспоненты случайного кодирования и показатель экспоненты для процедуры с выбрасыванием равенствами

$$E_r(R) = \sup [E_0(\rho, X_d, Y_p, r) - \rho R], \tag{7.3.36}$$

$$E_{ex}(R) = \sup [E_x(\rho, X_d, Y_p, r) - \rho R]. \tag{7.3.37}$$

Верхняя грань в обоих приведенных выше равенствах берется по всем конечным наборам входных букв; всем вероятностям, удовлетворяющим ограничению; всем разбиениям на выходе; всем $r \geqslant 0$ и всем ρ, удовлетворяющим $0 \leqslant \rho \leqslant 1$ для (7.3.36) и $\rho \geqslant 1$ для (7.3.37).

Теорема 7.3.2. (Теорема кодирования.) Для произвольного дискретного по времени канала без памяти с ограничением на входе $\overline{f(x)} \leqslant \mathscr{E}$ пусть $E_r(R)$, $E_{ex}(R)$ и C определены в (7.3.36), (7.3.37) и (7.3.1). Пусть $R \geqslant 0$ и $E < \max [E_r(R), E_{ex}(R)]$ произвольны. Тогда для всех достаточно больших N существует блоковый код длины N с $M = \lceil e^{NR} \rceil$ кодовыми словами $\mathbf{x}_1, \ldots, \mathbf{x}_M$, каждое из которых удовлетворяет ограничению

$$\sum_{n} f(x_{m,n}) \leqslant N\mathscr{E}$$

и для каждого из которых удовлетворяется неравенство

$$P_{e,m} \leqslant \exp(-NE). \tag{7.3.38}$$

Кроме того, $E_r(R) > 0$ для всех R, $0 \leqslant R < C$.

Доказательство. Выберем E_1, удовлетворяющее условию $E < E_1 < \max[E_r(R), E_{ex}(R)]$. Если $E_r(R) \geqslant E_{ex}(R)$, выберем X_d, Y_p, $r \geqslant 0$ и $0 \leqslant \rho \leqslant 1$ так, что

$$E_1 \leqslant E_0(\rho, X_d, Y_p, r) - \rho R.$$

Из (7.3.21) следует, что для любого $\delta > 0$ и каждого $N \geqslant 1$ существует код для этих X_d, Y_p с $M = \lceil e^{NR} \rceil$ кодовыми словами, каждое из которых удовлетворяет ограничению и удовлетворяет неравенству

$$P_{e,m} \leqslant [e^{r\delta}/\mu]^{1+\rho} \exp(-NE_1); \quad 1 \leqslant m \leqslant M. \tag{7.3.39}$$

Так как для достаточно больших N выражение $[e^{r\delta}/\mu]^{1+\rho}$ возрастает как $N^{(1+\rho)/2}$ и так как $E < E_1$, то для достаточно больших N имеем

$$P_{e,m} \leqslant [e^{r\delta}/\mu]^{1+\rho} \exp(-NE_1) \leqslant \exp(-NE). \tag{7.3.40}$$

Эти же рассуждения применимы и в случае, когда $E_{ex}(R) > E_r(R)$, за исключением того, что E_x должно стоять вместо E_0. Далее примем, что $R < C$ и выберем X_d, Y_p так, чтобы $R < I(X_d; Y_p) < C$. Из соображений, следующих за (7.3.23), видно, что для $r = 0$

$$\max_{0 \leqslant \rho \leqslant 1} [E_0(\rho, X_d, Y_p, r) - \rho R] > 0$$

и, следовательно, $E_r(R) > 0$. |

Теорема 7.3.2 обладает большой общностью, однако часто ее трудно применить ввиду трудности вычисления верхних граней, введенных при определении $E_r(R)$ и $E_{ex}(R)$. Для каналов, задаваемых переходной плотностью вероятности, обе величины $E_0(\rho, X_d, Y_p, r)$ и $E_x(\rho, X_d, Y_p, r)$ не убывают при измельчении разбиения пространства Y. Этот результат доказывается точно так же, как и для каналов без ограничений на входе. Верхняя грань по Y_p достигается и имеет вид

$$E_0(\rho, X_d, Y, r) = -\ln \int \left[\sum_k Q(a_k) e^{r[f(a_k) - \mathscr{E}]} \times \right.$$

$$\left. \times p(y \mid a_k)^{1/(1+\rho)} \right]^{1+\rho} dy, \tag{7.3.41}$$

$$E_x(\rho, X_d, Y, r) = -\rho \ln \left\{ \sum_k \sum_i Q(a_k) Q(a_i) \times \right.$$

$$\left. \times e^{r[f(a_k) + f(a_i) - 2\mathscr{E}]} \left[\int \sqrt{p(y \mid a_k) p(y \mid a_i)}\, dy \right]^{1/\rho} \right\}. \tag{7.3.42}$$

Если верхняя грань E_0 и E_x по X_d достигается в пределе на распределениях, сходящихся к плотности вероятности $q(x)$, то можно

положить по определению

$$E_0(\rho, X, Y, r) = -\ln \int \left[\int q(x)\, e^{r[f(x)-\mathscr{E}]}\, p(y|x)^{1/(1+\rho)}\, dx \right]^{1+\rho} dy, \tag{7.3.43}$$

$$E_x(\rho, X, Y, r) = -\rho \ln \int\int q(x)\, q(x')\, e^{r[f(x)+f(x')-2\mathscr{E}]} \times$$

$$\times \left[\int \sqrt{p(y|x)\, p(y|x')}\, dy \right]^{1/\rho} dx\, dx'. \tag{7.3.44}$$

Показатели экспонент $E_r(R)$ и $E_{ex}(R)$ могут теперь быть найдены прямо из (7.3.43) и (7.3.44) после максимизации лишь по ρ и r. В этом случае можно получить в некотором смысле более точные результаты, повторяя выводы теорем 5.6.1 и 5.7.1 и используя при этом плотности вместо вероятностей и интегралы вместо сумм. Упрощая результаты, так же как при переходе от (7.3.16) к (7.3.2), находим, что существует код, для которого каждое кодовое слово удовлетворяет как ограничению, так и следующим границам вероятности ошибки:

$$P_{e,m} \leqslant [2e^{r\delta}/\mu]^2 \exp\{-N[E_0(\rho, X, Y, r) - \rho R]\}; \quad 0 \leqslant \rho \leqslant 1, \tag{7.3.45}$$

$$P_{e,m} \leqslant \exp\{-N[E_x(\rho, X, Y, r) - \rho R']\}; \quad \rho \geqslant 1, \tag{7.3.46}$$

где R' определено в (7.3.33), $\delta > 0$ — произвольно и μ задается приближенно равенством (7.3.29), в котором $\sigma_f^2 = \int q(x)\,[f(x) - \mathscr{E}]^2\, dx$. Ограничения при выводе этих соотношений состоят в том, что плотности и интегралы существуют; $\int q(x)\, f(x)\, dx = \mathscr{E}$ и $\int_{-\infty}^{\infty} q(x)\,[f(x)]^2\, dx$ конечен. Ясно, что (7.3.45) и (7.3.46) справедливы, независимо от того, максимизирует ли входная плотность $q(x)$ величины E_0 и E_x или нет.

7.4. АДДИТИВНЫЙ ШУМ И АДДИТИВНЫЙ ГАУССОВ ШУМ

В этом параграфе результаты, полученные в § 7.2 и 7.3, применяются в важном и простом частном случае каналов с аддитивным шумом. Канал с аддитивным шумом определяется как канал, для которого входное пространство — множество действительных чисел (или действительных векторов) и выход представляется как сумма входа и статистически независимой случайной величины (или вектора), называемой шумом*). Для простоты примем, что шум z имеет плотность вероятности $p_Z(z)$. Для данного входа x выход принимает значение y, тогда и только тогда, когда $z = y - x$, и так как z не зависит от x, то переходная плотность вероятности канала задается равенством

$$p_{Y|X}(y|x) = p_Z(y - x). \tag{7.4.1}$$

*) Для пуританина, у которого вызывает беспокойство определение статистической независимости при отсутствии какой-либо вероятностной меры на входном пространстве, канал с аддитивным шумом может быть определен как канал, удовлетворяющий соотношению (7.4.1), т. е. канал, для которого переходная вероятностная мера является функцией только разности $y - x$.

Вычисление средней взаимной информации и пропускной способности для канала с аддитивным шумом сильно упрощается в силу того, что условная энтропия выхода при заданном входе $H(Y \mid X)$ равна энтропии шума $H(Z)$ и, следовательно, не зависит от входного распределения. Для того чтобы убедиться в этом, будем считать, что $p_X(x)$ — плотность вероятности на входе. Используя (7.4.1) в выражении, определяющем условную энтропию (2.4.25), имеем

$$H(Y \mid X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_X(x)\, p_Z(y-x) \log p_Z(y-x)\, dy\, dx =$$

$$= -\int_{-\infty}^{\infty} p_X(x) \int_{-\infty}^{\infty} p_Z(z) \log p_Z(z)\, dz\, dx =$$

$$= \int p_X(x)\, H(Z)\, dx = H(Z). \qquad (7.4.2)$$

Те же самые соображения, очевидно, применимы для дискретного распределения на входе. Следовательно, средняя взаимная информация между выходом и входом канала задается равенствами

$$I(X;\ Y) = H(Y) - H\ (Y \mid X) = H(Y) - H(Z). \qquad (7.4.3)$$

В этом выражении $H(Y)$ зависит от входного распределения, а $H(Z)$ не зависит. Таким образом, проблема нахождения пропускной способности для канала с аддитивным шумом сводится к максимизации $H(Y)$ при заданных ограничениях на входе. Следующие два примера показывают, как эта задача может быть иногда решена.

Пример. Сначала рассмотрим шум с плотностью вероятности $p_Z(z) = 1/2$ для $-1 \leqslant z \leqslant 1$ и $p_Z(z) = 0$ во всех других точках. Предположим, что амплитуда на входе принимает значения из интервала $-1 \leqslant x \leqslant 1$. Так как выход y равен $x+z$, то значения сигнала на выходе лежат в интервале $-2 \leqslant y \leqslant 2$. Наша задача заключается в том, чтобы найти в этом интервале $p_Y(y)$, максимизирующее $H(Y)$, и затем попытаться найти входное распределение, которое приводит к этой максимизирующей плотности $p_Y(y)$. Нетрудно догадаться, что $H(Y)$ достигает максимума на плотности вероятности равномерного распределения и это будет подтверждено применением вариационного исчисления. Пусть

$$F[p(y)] = -\int_{-2}^{2} p(y) \log p(y)\, dy + \lambda \int_{-2}^{2} p(y)\, dy, \qquad (7.4.3a)$$

где λ — множитель Лагранжа для ограничения $\int p(y)\, dy = 1$. Функция $p(y)$ является стационарной точкой, если выражение

$$\partial F[p(y) + \varepsilon h(y)]/\partial \varepsilon\,|_{\varepsilon=0}$$

равно нулю для всех $h(y)$,

$$\left.\frac{\partial F[p(y)+\varepsilon\, h(y)]}{\partial \varepsilon}\right|_{\varepsilon=0} = -\int_{-2}^{2} h(y)\,[\log p(y) + \log e - \lambda]\, dy. \qquad (7.4.4)$$

Следовательно, стационарная точка достигается, если

$$\log p(y) + \log e - \lambda = 0; \quad -2 \leqslant y \leqslant 2. \tag{7.4.5}$$

Это означает, что $p(y)$ постоянна в рассматриваемом интервале, или $p(y) = 1/4$.

Заметим, что дискретное распределение на входе $P_X(-1) = P_X(+1) = 1/2$ дает $P_Y(y) = 1/4$ при $-2 < y \leqslant 2$ и поэтому на нем, по-видимому, достигается пропускная способность. Для того чтобы строго доказать, что на этом распределении достигается пропускная способность, выберем произвольное конечное множество входных букв $a_1, \ldots, a_K$ между -1 и $+1$, включая -1 и $+1$. Для рассматриваемого входного распределения

$$I(x = a_k; Y) = \int_{-2}^{2} p_Z(y - a_k) \log \frac{p_Z(y - a_k)}{1/4}\, dy = 1 \text{ бит}.$$

Следовательно, это распределение удовлетворяет необходимым и достаточным условиям теоремы 4.5.1, при выполнении которых достигается пропускная способность. Аналогично можно проверить, что это распределение удовлетворяет необходимым и достаточным условиям, при выполнении которых максимизируется $E_r(R)$. Это не удивительно, так как использование на входе лишь $x = -1$ и $x = +1$ превращает рассматриваемый канал в двоичный канал без шума; $y > 0$ означает, что $x = +1$, а $y < 0$ означает, что $x = -1$. В задаче 7.5, в которой продолжается изучение этого примера, рассматриваются произвольные амплитудные ограничения на x. Оказывается, что плотность равномерного распределения y в общем случае не получается, хотя пропускная способность всегда достигается на дискретном входном распределении.

Аддитивный гауссов шум и ограничение на энергию входного сигнала

Рассматривая второй пример, предположим, что в канале с аддитивным шумом задано ограничение на энергию на входе

$$\overline{x^2} \leqslant \mathscr{E}. \tag{7.4.6}$$

Пусть шум имеет плотность $p_Z(z)$ с нулевым средним и дисперсией σ^2. Предположение $\bar{z} = 0$ не приводит к потере общности, так как можно всегда добиться его выполнения, сдвигая нуль на осях z и y. В дальнейшем мы примем, что $p_Z(x)$ — плотность гауссовского распределения, однако в течение некоторого времени будем считать ее произвольной. Среднеквадратическое значение выхода канала ограничено следующим образом:

$$\overline{y^2} = \overline{(x+z)^2} = \overline{x^2} + \overline{z^2} \leqslant \mathscr{E} + \sigma^2. \tag{7.4.7}$$

Найдем теперь максимум $H(Y)$ при ограничении на $\overline{y^2}$, а затем попытаемся найти соответствующее входное распределение, приводящее

к плотности $p_Y(y)$, на которой достигается максимум. Можно опять использовать методы вариационного исчисления, распространяя предел в (7.4.3) до ∞ и добавляя второе ограничение, $\gamma \int y^2 p(y)\,dy$. Это приводит к условиям, аналогичным (7.4.5),

$$\log p(y) + \log e - \lambda - \gamma y^2 = 0 \text{ для всех } y.$$

Решение этого уравнения, удовлетворяющее ограничению (7.4.7), имеет вид

$$p(y) = \frac{1}{\sqrt{2\pi(\mathscr{E} + \sigma^2)}} \exp\left[-\frac{y^2}{2(\mathscr{E} + \sigma^2)}\right]. \tag{7.4.8}$$

Теорема 7.4.1. Максимальное значение энтропии

$$H(Y) = -\int_{-\infty}^{\infty} p(y) \log p(y)\,dy,$$

взятое по всем плотностям вероятностей, удовлетворяющим ограничению

$$\int_{-\infty}^{\infty} y^2 p(y)\,dy = A, \tag{7.4.9}$$

достигается только на плотности гауссовского распределения

$$\varphi_A(y) = \frac{1}{\sqrt{2\pi A}} \exp\left(-\frac{y^2}{2A}\right) \tag{7.4.10}$$

и равно

$$H(Y) = {}^1/_2 \log(2\pi e A). \tag{7.4.11}$$

Доказательство. Эту теорему можно было бы доказать, развивая соображения гл. 4 о свойствах выпуклых функций и доказывая, что решение (7.4.8) вариационного уравнения дает единственный максимум. Однако следующее доказательство в некотором смысле проще. Пусть $p(y)$ — произвольная плотность вероятности, удовлетворяющая (7.4.9), и пусть $\varphi_A(y)$ — плотность гауссовского распределения. Тогда

$$\int p(y) \log \frac{1}{\varphi_A(y)}\,dy = \int p(y)\left[\log\sqrt{2\pi A} + \frac{y^2}{2A}\log e\right]dy =$$
$$= \log\sqrt{2\pi A} + \frac{A}{2A}\log e = {}^1/_2 \log(2\pi e A). \tag{7.4.12}$$

Применяя (7.4.12), получаем

$$H(Y) - {}^1/_2 \log(2\pi e A) = \int p(y) \log \frac{\varphi_A(y)}{p(y)}\,dy \leqslant$$
$$\leqslant \log e \int p(y)\left[\frac{\varphi_A(y)}{p(y)} - 1\right]dy = 0,$$

где было использовано неравенство $\log z \leqslant (z-1)\log e$. Равенство здесь достигается тогда и только тогда, когда $\varphi_A(y)/p(y) = 1$ для всех y.

Поскольку $H(Y)$ возрастает с A, то ясно, что изменение ограничения в теореме на $\int y^2 p(y)\,dy \leqslant A$ не может изменить результат. Для получения пропускной способности остается найти плотность вероятности входного сигнала, приводящую к плотности гауссовского распределения на выходе. Один из печальных фактов нашей жизни состоит в том, что если сумма двух независимых случайных величин является гауссовской случайной величиной, то каждая из случайных величин должна быть гауссовской*). Однако, к счастью, наибольший интерес представляет ситуация, когда аддитивный шум гауссов. В этом случае максимум $H(Y)$ и пропускная способность достигаются на гауссовском x. Для этого случая выражение $I(X; Y)$ уже было вычислено в (2.4.36). Таким образом, доказана следующая теорема.

Теорема 7.4.2. Пусть задан дискретный по времени канал без памяти с аддитивным гауссовым шумом, дисперсия которого σ^2, и пусть ограничение на входе имеет вид $\overline{x^2} \leqslant \mathscr{E}$. Тогда пропускная способность равна

$$C = {}^1/_2 \log\left(1 + \frac{\mathscr{E}}{\sigma^2}\right). \tag{7.4.13}$$

Вычисление пропускной способности канала с негауссовым аддитивным шумом — задача утомительная и неблагодарная. Ограничимся здесь границами для пропускной способности, даваемыми следующей теоремой; эта теорема фактически показывает, что при заданной дисперсии шума гауссов шум является наихудшим с точки зрения пропускной способности аддитивным шумом.

Теорема 7.4.3. Пусть задан дискретный по времени канал без памяти с аддитивным шумом (дисперсия которого σ^2) и с ограничением на входе $\overline{x^2} \leqslant \mathscr{E}$. Тогда

$${}^1/_2 \log[2\pi e(\mathscr{E} + \sigma^2)] - H(Z) \geqslant C \geqslant \frac{1}{2}\log\left(1 + \frac{\mathscr{E}}{\sigma^2}\right). \tag{7.4.14}$$

Доказательство. Левая часть неравенства следует из соотношения $I(X; Y) = H(Y) - H(Z)$ и того, что выражение ${}^1/_2 \log[2\pi e(\mathscr{E} + \sigma^2)]$ является верхней границей $H(Y)$. Для того чтобы установить справедливость правой части неравенства, положим

$$p_X(x) = \frac{1}{\sqrt{2\pi\mathscr{E}}} \exp\left(-\frac{x^2}{2\mathscr{E}}\right)$$

*) Г. Крамер. Случайные величины и распределения вероятностей. М., ИЛ, 1937.

и покажем, что получающаяся средняя взаимная информация удовлетворяет неравенству

$$I(X;Y) \geqslant {}^1/_2 \log\left(1+\frac{\mathscr{E}}{\sigma^2}\right). \tag{7.4.15}$$

Это полностью докажет теорему, так как C — верхняя грань $I(X;Y)$ по всем допустимым входным распределениям. Пусть $p_Z(z)$ — плотность вероятности шума и пусть $\varphi_{\sigma^2}(z)$ — плотность вероятности гауссовского распределения с дисперсией σ^2. Пусть $p_Y(y)$ — выходная плотность и пусть $\varphi_A(y)$ — плотность гауссовского распределения с дисперсией $A = \mathscr{E} + \sigma^2$. Тогда, так же как в (7.4.12), имеем

$$\begin{aligned}\iint p_X(x)\, p_Z(y-x) \log \frac{\varphi_{\sigma^2}(y-x)}{\varphi_A(y)}\, dy\, dx &= \int p_Z(z) \log \varphi_{\sigma^2}(z)\, dz - \\ - \int p_Y(y) \log \varphi_A(y)\, dy &= -{}^1/_2 \log(2\pi e\sigma^2) + \\ + {}^1/_2 \log(2\pi e A) &= {}^1/_2 \log\left(1+\frac{\mathscr{E}}{\sigma^2}\right).\end{aligned} \tag{7.4.16}$$

Далее, используя (7.4.16),

$$\begin{aligned}&-I(X;Y) + {}^1/_2 \log\left(1+\frac{\mathscr{E}}{\sigma^2}\right) = \\ &= \iint p_X(x)\, p_Z(y-x) \log \frac{p_Y(y)\, \varphi_{\sigma^2}(y-x)}{p_Z(y-x)\, \varphi_A(y)}\, dx\, dy \leqslant \\ &\leqslant \log e \left\{ \iint \frac{p_X(x)\, p_Y(y)\, \varphi_{\sigma^2}(y-x)}{\varphi_A(y)}\, dy\, dx - 1 \right\}.\end{aligned} \tag{7.4.17}$$

Однако, так как $p_X(x)$ — плотность гауссовского распределения, то $\int p_X(x)\, \varphi_{\sigma^2}(y-x)\, dx = \varphi_A(y)$. Следовательно, двойной интеграл в (7.4.17) сводится к $\int p_Y(y)\, dy = 1$ и правая часть (7.4.17) равна нулю, что завершает доказательство. |

Далее используем границы вероятности ошибки (7.3.45) и (7.3.46) для канала с аддитивным гауссовым шумом, описываемым плотностью

$$p(y \mid x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-x)^2}{2\sigma^2}\right]. \tag{7.4.18}$$

Каждое кодовое слово $\mathbf{x}_m = (x_{m,1}, \ldots, x_{m,N})$ удовлетворяет ограничению

$$\sum_{n=1}^{N} x_{m,n}^2 \leqslant N\mathscr{E}. \tag{7.4.19}$$

Выберем для ансамбля кодов плотность вероятности на входе

$$q(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\mathscr{E}}\right). \tag{7.4.20}$$

Имеется целый ряд причин для выбора здесь гауссовской плотности. Первая состоит в том, что она легко интегрируема, вторая причина — получающая совместная плотность имеет сферическую симметрию и третья причина в том, что эта плотность приводит к экспоненте случайного кодирования, показатель которой в нелинейной части совпадает с показателем экспоненты нижней границы вероятности ошибки (см. Шеннон, 1959).

Подставляя (7.4.18) и (7.4.20) в выражение для $E_0(\rho, X, Y, r)$ в (7.3.43) и заменяя $f(x)$ на x^2, мы можем дополнить показатель экспоненты до полного квадрата и полученное выражение проинтегрировать. После замены r на переменную s результат представляется в виде

$$E_0(\rho, X, Y, s) = \\ = s(1+\rho)\mathscr{E} + {}^1/_2 \ln(1-2s\mathscr{E}) + \frac{\rho}{2}\ln\left[1-2s\mathscr{E} + \frac{\mathscr{E}}{(1+\rho)\sigma^2}\right]. \quad (7.4.21)$$

Этот результат справедлив при $0 \leqslant s < 1/(2\mathscr{E})$. Для больших s, $E_0 = -\infty$, что бесполезно. Вместо максимизации этого выражения по s удобно произвести следующие замены:

$$A = \mathscr{E}/\sigma^2, \quad (7.4.22)$$

$$\beta = 1 - 2s\mathscr{E} + A/(1+\rho). \quad (7.4.23)$$

Величина A — отношение сигнал/шум в канале и можно ожидать, что в определенном масштабе получающаяся граница будет зависеть лишь от A, а не отдельно от $\mathscr{E}$ и σ^2. Используя (7.4.22) и (7.4.23), исключим $\mathscr{E}$, σ^2 и s в (7.4.21) и получим выражение для E_0 в виде функции от A, β и ρ:

$$\widetilde{E}_0(A, \beta, \rho) = \frac{1}{2}\left[(1-\beta)(1+\rho) + A + \ln\left(\beta - \frac{A}{1+\rho}\right) + \rho\ln\beta\right]. \quad (7.4.24)$$

Ограничение $0 \leqslant s < 1/(2\mathscr{E})$ появляется здесь как

$$\frac{A}{1+\rho} < \beta \leqslant 1 + \frac{A}{1+\rho}. \quad (7.4.25)$$

Функция $\widetilde{E}_0$ имеет стационарную точку относительно β, определяемую из равенства

$$\frac{\partial \widetilde{E}_0}{\partial \beta} = \frac{1}{2}\left[-(1+\rho) + \frac{1+\rho}{\beta(1+\rho) - A} + \frac{\rho}{\beta}\right] = 0. \quad (7.4.26)$$

Левая часть (7.4.26) в области, задаваемой (7.4.25), убывает по β от $+\infty$ до некоторого отрицательного значения. Следовательно, $\widetilde{E}_0$ максимизируется единственным значением β из этой области, которое удовлетворяет (7.4.26). Преобразуя (7.4.26) и решая получившееся квадратное уравнение

$$\beta^2 - \beta\left(1 + \frac{A}{1+\rho}\right) + \frac{A\rho}{(1+\rho)^2} = 0, \quad (7.4.27)$$

находим это значение

$$\beta=\frac{1}{2}\left(1+\frac{A}{1+\rho}\right)\left[1+\sqrt{1-\frac{4A\rho}{(1+\rho+A)^2}}\right]. \quad (7.4.28)$$

Далее $\widetilde{E}_0-\rho R$ имеет стационарную точку относительно ρ, определяемую из равенства

$$\frac{\partial[\widetilde{E}_0-\rho R]}{\partial\rho}=\frac{1}{2}\left[1-\beta+\frac{\beta}{(1+\rho)\beta-A}-\frac{1}{(1+\rho)}+\ln\beta\right]-R=0. \quad (7.4.29)$$

Для β, удовлетворяющих (7.4.26), оно сводится к

$$R=\tfrac{1}{2}\ln\beta. \quad (7.4.30)$$

Теперь для β и ρ, удовлетворяющих (7.4.26) и (7.4.29), имеем

$$E_r(R)=\widetilde{E}_0-\rho R=\frac{1}{2}\left[(1-\beta)(1+\rho)+A+\ln\left(\beta-\frac{A}{1+\rho}\right)\right]. \quad (7.4.31)$$

Теперь можно получить явное выражение для $E_r(R)$, разрешая (7.4.26) относительно $(1+\rho)$ через β и A. Это приводит к

$$1+\rho=\frac{A}{2\beta}\left[1+\sqrt{1+\frac{4\beta}{A(\beta-1)}}\right]. \quad (7.4.32)$$

Подставляя (7.4.32) в (7.4.31) и несколько упрощая получающееся выражение, находим

$$E_r(R)=\frac{A}{4\beta}\left[(\beta+1)-(\beta-1)\sqrt{1+\frac{4\beta}{A(\beta-1)}}\right]+$$
$$+\frac{1}{2}\ln\left\{\beta-\frac{A(\beta-1)}{2}\left[\sqrt{1+\frac{4\beta}{A(\beta-1)}}-1\right]\right\}, \quad (7.4.33)$$

где согласно (7.4.30), $\beta=e^{2R}$.

Равенство (7.4.33) справедливо для $0\leqslant\rho\leqslant 1$. Взяв значения β из (7.4.28) для $\rho=0$ и $\rho=1$ и подставляя их в (7.4.30), находим, что (7.4.33) справедливо для

$$\tfrac{1}{2}\ln\left[\frac{1}{2}+\frac{A}{4}+\frac{1}{2}\sqrt{1+\frac{A^2}{4}}\right]\leqslant R\leqslant\tfrac{1}{2}\ln(1+A). \quad (7.4.34)$$

Для R, меньших чем левая часть (7.4.34), следует выбрать $\rho=1$, что дает

$$E_r(R)=1-\beta+\frac{A}{2}+\frac{1}{2}\ln\left(\beta-\frac{A}{2}\right)+\frac{1}{2}\ln\beta-R, \quad (7.4.35)$$

где

$$\beta=\frac{1}{2}\left[1+\frac{A}{2}+\sqrt{1+\frac{A^2}{4}}\right], \quad (7.4.36)$$

Экспонента случайного кодирования $E_r(R)$ для некоторых значений A изображена на рис. 7.4.1.

Следует еще рассмотреть коэффициент $[2e^{s\delta}/\mu]^2$ в (7.3.45). Решая уравнение (7.4.23) относительно s, получаем

$$2s\mathscr{E} = 1 - \beta + A/(1+\rho). \tag{7.4.37}$$

Умножая числитель и знаменатель правой части на β и сравнивая с (7.4.27), получаем полезное в дальнейшем выражение для s

$$2s\mathscr{E} = \frac{\rho A}{(1+\rho)^2\beta}. \tag{7.4.38}$$

Из (7.4.29) имеем для больших N

$$\mu \approx \frac{\delta}{\sqrt{2\pi N}\,\sigma_f}. \tag{7.4.39}$$

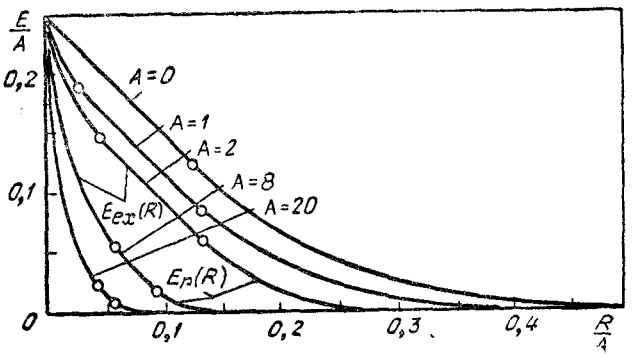


Рис. 7.4.1. $E_r(R)$ и $E_{ex}(R)$ для дискретного по времени канала с аддитивным гауссовым шумом при различных отношениях сигнал/шум A.

При использовании этого приближения для μ величина $e^{s\delta}/\mu$ достигает минимума при $\delta = 1/s$, давая

$$\left[\frac{2e^{s\delta}}{\mu}\right] \approx 2s\mathscr{E}e\sqrt{4\pi N} = \frac{\rho A e\sqrt{4\pi N}}{(1+\rho)^2\beta}. \tag{7.4.40}$$

Для того чтобы получить точное выражение для μ, заметим, что μ равно вероятности, с которой случайная величина с распределением χ^2 с N степенями свободы принимает значение из интервала между ее средним и числом на $\delta/\mathscr{E}$ меньше этого среднего. Таким образом, μ может быть найдено из таблиц распределения χ^2.

В случае границы случайного кодирования для процедуры с выбрасыванием подставим (7.4.18) и (7.4.20) в (7.3.44). Интегрируя, получаем

$$E_x(\rho, X, Y, s) = 2\rho s\mathscr{E} + \frac{\rho}{2}\ln\left(1 - 2s\mathscr{E} + \frac{\mathscr{E}}{2\rho\sigma^2}\right) + \\ + \frac{\rho}{2}\ln(1 - 2s\mathscr{E}). \tag{7.4.41}$$

После подстановок $A = \mathscr{E}/\sigma^2$ и

$$\beta = 1 - 2s\mathscr{E} + \frac{A}{2\rho} \tag{7.4.42}$$

равенство (7.4.41) принимает вид

$$\widetilde{E}_x(A, \beta, \rho) = (1-\beta)\rho + \frac{A}{2} + \frac{\rho}{2} \ln\left[\beta\left(\beta - \frac{A}{2\rho}\right)\right], \tag{7.4.43}$$

где β ограничено интервалом $A/(2\rho) < \beta < 1 + A/(2\rho)$. Взяв $\partial\widetilde{E}_x/\partial\beta$, находим, что максимум по β существует и достигается при

$$\beta^2 - \beta\left(1 + \frac{A}{2\rho}\right) + \frac{A}{4\rho} = 0 \tag{7.4.44}$$

или

$$\beta = \frac{1}{2} + \frac{A}{4\rho} + \frac{1}{2}\sqrt{1 + \frac{A^2}{4\rho^2}}. \tag{7.4.45}$$

Равенство (7.4.44) может быть переписано в виде

$$\rho = \frac{A(2\beta-1)}{4\beta(\beta-1)}. \tag{7.4.46}$$

Далее, $\widetilde{E}_x(A, \beta, \rho) - \rho R'$ имеет максимум по ρ, который достигается при

$$R' = 1 - \beta + \frac{A}{2(2\rho\beta - A)} + \frac{1}{2} \ln\left[\beta\left(\beta - \frac{A}{2\rho}\right)\right]. \tag{7.4.47}$$

Для β и ρ, которые удовлетворяют обоим равенствам (7.4.44) и (7.4.47), можно подставить (7.4.46) вместо ρ в третье слагаемое правой части (7.4.47) и получить

$$R' = {}^1/_2 \ln\left[\beta\left(\beta - \frac{A}{2\rho}\right)\right], \tag{7.4.48}$$

$$E_{ex}(R') = (1-\beta)\rho + \frac{A}{2}. \tag{7.4.49}$$

При использовании выражения (7.4.46) для ρ эти равенства упрощаются:

$$R' = {}^1/_2 \ln\frac{\beta^2}{2\beta - 1}, \tag{7.4.50}$$

$$E_{ex}(R') = \frac{A}{4\beta}. \tag{7.4.51}$$

Решая (7.4.50) относительно β, получаем явное выражение

$$E_{ex}(R') = \frac{A}{4}\left(1 - \sqrt{1 - e^{-2R'}}\right). \tag{7.4.52}$$

Это справедливо для $\rho \geqslant 1$, или, комбинируя (7.4.45) и (7.4.50), для

$$R' \leqslant {}^1/_2 \ln\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 + \frac{A^2}{4}}\right). \tag{7.4.53}$$

Из (7.3.33) следует, что R' связана со скоростью R соотношением

$$R' = R + \frac{2}{N} \ln \frac{2e^{s\delta}}{\mu}. \tag{7.4.54}$$

Параметр s задается соотношением (7.4.42) следующим образом:

$$2s\mathscr{E}\rho = (1-\beta)\rho + \frac{A}{2} = \frac{A}{4\beta}, \tag{7.4.55}$$

где было использовано соотношение (7.4.46). Подставляя это значение s в (7.4.40), получаем

$$\frac{2e^{s\delta}}{\mu} \approx 2s\mathscr{E}e\sqrt{4\pi N} = \frac{Ae\sqrt{4\pi N}}{4\rho\beta}. \tag{7.4.56}$$

Эти результаты подытоживаются в следующей теореме.

Теорема 7.4.4. Пусть для дискретного по времени канала с аддитивным гауссовым шумом переходная плотность вероятности имеет вид

$$p(y \mid x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-(y-x)^2/(2\sigma^2)\right],$$

а ограничение имеет вид $\overline{x^2} \leqslant \mathscr{E} = \sigma^2 A$. Тогда для любой длины блока N и любой скорости $0 \leqslant R < C = 1/2 \ln(1+A)$ существует код с $M = \lceil e^{NR} \rceil$ кодовыми словами, каждое из которых удовлетворяет ограничению (7.4.19) и границе вероятности ошибки

$$P_{e,m} \leqslant \left[\frac{2e^{s\delta}}{\mu}\right]^2 e^{-NE_r(R)}, \tag{7.4.57}$$

где $E_r(R)$ задается соотношениями (7.4.33) и (7.4.36), а $2e^{s\delta}/\mu$ задается приближенно (7.4.40). Кроме того, если R', задаваемое (7.4.54) и (7.4.56), удовлетворяет (7.4.53), то также для всех кодовых слов

$$P_{e,m} \leqslant \exp\left\{-\frac{NA}{4}\left[1 - \sqrt{1-e^{-2R'}}\right]\right\}. \tag{7.4.58}$$

Хотя это непосредственно не очевидно из выражения для $E_r(R)$, однако из § 7.3 следует, что $\max[E_r(R), E_{ex}(R)]$ является выпуклой $\smile$ невозрастающей положительной функцией для $0 \leqslant R < {}^1/_2 \ln(1+A)$.

7.5. ПАРАЛЛЕЛЬНЫЕ КАНАЛЫ С АДДИТИВНЫМ ГАУССОВЫМ ШУМОМ

В следующей главе канал с непрерывным временем и аддитивным гауссовым шумом будет сведен к множеству параллельных дискретных по времени каналов с аддитивным гауссовым шумом. Следующая теорема позволяет найти пропускную способность такого параллельного соединения каналов.

Теорема 7.5.1. Рассмотрим множество из N параллельных дискретных по времени каналов с аддитивными гауссовыми шумами и дис-

персиями шумов $\sigma_1^2, \ldots, \sigma_N^2$. Пусть входы каналов удовлетворяют ограничению

$$\sum_{n=1}^{N} \overline{x_n^2} = \mathscr{E}. \tag{7.5.1}$$

Тогда пропускная способность достигается на входах, представляющих собой статистически независимые гауссовские случайные ве-

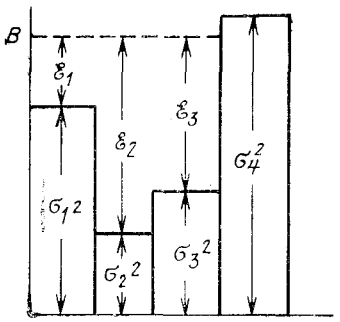


Рис. 7.5.1. Интерпретация, связанная с «наполнением водой» для параллельных дискретных по времени каналов с аддитивным гауссовым шумом.

личины с нулевыми средними и дисперсиями

$$\overline{x_n^2} = \mathscr{E}_n, \tag{7.5.2}$$

где $\mathscr{E}_n$ удовлетворяют соотношениям

$$\sigma_n^2 + \mathscr{E}_n = B \quad \text{для} \quad \sigma_n^2 < B, \tag{7.5.3}$$

$$\mathscr{E}_n = 0 \quad \text{для} \quad \sigma_n^2 \geqslant B \tag{7.5.4}$$

и где B выбрано так, что $\Sigma \mathscr{E}_n = \mathscr{E}$. Пропускная способность параллельного соединения каналов равна

$$C = \sum_{n=1}^{N} {}^1/_2 \ln\left(1 + \frac{\mathscr{E}_n}{\sigma_n^2}\right) \text{ нат} = \tag{7.5.5}$$

$$= \sum_{n:\, \sigma_n^2 \leqslant B} {}^1/_2 \ln (B/\sigma_n^2) \text{ нат}. \tag{7.5.6}$$

Обсуждение. Графическая интерпретация распределения для входных энергий, используемых в различных каналах, приведена на рис. 7.5.1. Можно представить себе общую энергию $\mathscr{E}$ как объем воды, которая помещается в резервуар с неровной формой дна, определяемой дисперсиями шума. Уровень, до которого поднимается вода, равен B, а $\mathscr{E}_n$ определяет глубину воды в различных частях резервуара.

Доказательство. Пусть $\mathbf{X}^N = (X_1 X_2 \ldots X_N)$ и $\mathbf{Y}^N = (Y_1 Y_2 \ldots Y_N)$ — совместные входные и выходные ансамбли. Так же как и (7.2.19), доказывается, что

$$I(\mathbf{X}^N; \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n; Y_n)$$

с равенством, когда входы независимы. Пусть $\mathscr{E}_n$ — среднеквадратическое значение n-го входа для любого заданного совместного ансамбля. Тогда, используя теорему 7.4.2, имеем

$$\sum I(X_n; Y_n) \leqslant \sum {}^1/_2 \ln\left(1 + \frac{\mathscr{E}_n}{\sigma_n^2}\right) \tag{7.5.7}$$

с равенством, когда входы — гауссовские случайные величины с нулевым средним. Правая часть (7.5.7) — выпуклая $\frown$ функция вектора $(\mathscr{E}_1, \ldots, \mathscr{E}_N)$. Теперь осталось провести максимизацию этой функции в выпуклой области, где $\mathscr{E}_n \geqslant 0$, $1 \leqslant n \leqslant N$ и $\Sigma\mathscr{E}_n \leqslant \mathscr{E}$. Очевидно, максимум имеет место, когда $\Sigma\mathscr{E}_n/\mathscr{E} = 1$ и, следовательно, рассматриваемая задача тождественна максимизации выпуклой функции вектора вероятностей. Из теоремы 4.4.1 вытекает, что необходимые и достаточные условия максимума будут

$$\frac{\partial \sum {}^1/_2 \ln(1 + \mathscr{E}_n/\sigma_n^2)}{\partial \mathscr{E}_n} \leqslant \alpha \text{ для всех } n$$

с равенством, когда $\mathscr{E}_n > 0$, и α, выбранным так, что удовлетворяется равенство $\Sigma\mathscr{E}_n = \mathscr{E}$. Дифференцируя, выводим

$$\frac{1}{2(\sigma_n^2 + \mathscr{E}_n)} \leqslant \alpha.$$

Выбирая $B = 1/(2\alpha)$, получаем (7.5.3) и (7.5.4). То что получается максимум, задаваемый (7.5.5) и (7.5.6), следует из соотношений (7.5.7), (7.5.4) и (7.5.5). |

Далее распространим границу случайного кодирования и границу для процедуры с выбрасыванием на параллельные каналы с аддитивным гауссовым шумом. Для простоты обозначения примем, что длина блока равна 1. Результаты могут быть применены к произвольной длине блока N', если рассмотреть множество параллельных каналов с N' повторениями каждого из первоначальных каналов.

Для N параллельных каналов с дисперсиями шумов $\sigma_1^2, \ldots, \sigma_N^2$ соответственно совместная переходная плотность вероятности множества каналов равна

$$p_N(\mathbf{y} \mid \mathbf{x}) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left[-\frac{(y_n - x_n)^2}{2\sigma_n^2}\right]. \tag{7.5.8}$$

Примем, что каждое кодовое слово $\mathbf{x}_m$ должно удовлетворять соотношению

$$\sum_n (x_{m,n})^2 \leqslant \mathscr{E}.$$

В качестве входной плотности для ансамбля кодов выберем

$$q_N(\mathbf{x}) = \frac{\varphi(\mathbf{x})}{\mu} \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\mathscr{E}_n}} \exp\left(-\frac{x_n^2}{2\mathscr{E}_n}\right), \tag{7.5.9}$$

где $\mathscr{E}_n$ будут выбраны ниже, но так, чтобы они удовлетворяли соотношению

$$\sum_n \mathscr{E}_n = \mathscr{E}. \tag{7.5.10}$$

Так же как и в § 7.3,

$$\varphi(\mathbf{x}) = \begin{cases} 1, & \mathscr{E} - \delta < \sum_n x_n^2 \leqslant \mathscr{E}, \\ 0, & \text{в других случаях} \end{cases} \tag{7.5.11}$$

и μ выбрано так, чтобы интеграл от $q_N(\mathbf{x})$ был равен 1. Используя теорему 5.6.1 в интегральной форме, находим, что вероятность ошибки по ансамблю кодов с M кодовыми словами удовлетворяет неравенству

$$\overline{P_{e,m}} \leqslant (M-1)^\rho \int_{y_1} \dots \int_{y_N} \left[\int_{x_1} \dots \int_{x_N} q_N(\mathbf{x})\, p_N(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}\, d\mathbf{x} \right]^{1+\rho} d\mathbf{y}. \tag{7.5.12}$$

для любых ρ, $0 \leqslant \rho \leqslant 1$. Ограничивая сверху $\varphi(\mathbf{x})$ для любого $s \geqslant 0$ неравенством

$$\varphi(\mathbf{x}) \leqslant \exp\left[s\delta + s\sum_n (x_n^2 - \mathscr{E}_n)\right], \tag{7.5.13}$$

оценку (7.5.12) можно упростить и привести к виду

$$\overline{P_{e,m}} \leqslant \left[\frac{e^{s\delta}}{\mu}\right]^{1+\rho} (M-1)^\rho \exp\left[-\sum_{n=1}^{N} E_0(\rho, X_n, Y_n, s)\right] \tag{7.5.14}$$

$$E_0(\rho, X_n, Y_n, s) = -\ln \int \left\{ \int \frac{1}{\sqrt{2\pi\mathscr{E}_n}} \exp\left[-\frac{x^2}{2\mathscr{E}_n} + s(x^2 - \mathscr{E}_n)\right] \times \right.$$
$$\left. \times \left[\frac{1}{\sqrt{2\pi\sigma_n^2}} \exp -\frac{(y-x)^2}{2\sigma_n^2}\right]^{1/(1+\rho)} dx \right\}^{1+\rho} dy. \tag{7.5.15}$$

Выражение в (7.5.15) совпадает с интегралом, который вычислен в (7.4.2) так, что

$$E_0(\rho, X_n, Y_n, s) = s(1+\rho)\mathscr{E}_n +$$
$$+ {}^1/_2 \ln(1 - 2s\mathscr{E}_n) + \frac{\rho}{2} \ln\left(1 - 2s\mathscr{E}_n + \frac{\mathscr{E}_n}{(1+\rho)\sigma_n^2}\right). \tag{7.5.16}$$

Как и в (7.3.21) для любого $R \geqslant 0$, теперь можно установить существование кода с $M = \lceil e^R \rceil$ кодовыми словами, каждое из которых удовлетворяет неравенству

$$P_{e,m} \leqslant \left[\frac{2e^{s\delta}}{\mu}\right]^2 \exp\left[\rho R - \sum_{n=1}^{N} E_0(\rho, X_n, Y_n, s)\right]. \tag{7.5.17}$$

Максимизируем теперь экспоненту в (7.5.17) по $0 \leqslant \rho \leqslant 1$, $s \geqslant 0$, с помощью той же самой процедуры, которая использовалась для одного канала с аддитивным гауссовым шумом. Имеется дополнительная задача максимизации по энергии отдельных входов при ограничениях

$$\mathscr{E}_n \geqslant 0, \quad \sum_n \mathscr{E}_n = \mathscr{E},$$

Положим

$$A_n = \mathscr{E}_n/\sigma_n^2, \tag{7.5.18}$$

$$\beta_n = 1 - 2s\mathscr{E}_n + A_n/(1+\rho). \tag{7.5.19}$$

Тогда $E_0(\rho, X_n, Y_n, s)$ в (7.5.16) можно заменить выражением

$$\widetilde{E}_0(A_n, \beta_n, \rho) = \frac{1}{2}\left[(1-\beta_n)(1+\rho) + A_n + \ln\left(\beta_n - \frac{A_n}{1+\rho}\right) + \rho \ln \beta_n\right]. \tag{7.5.20}$$

Нужно максимизировать

$$E = -\rho R + \sum_{n=1}^{N} \widetilde{E}_0(A_n, \beta_n, \rho) \tag{7.5.21}$$

по A_n, β_n и ρ. На величины A_n наложены ограничения $A_n \geqslant 0$, $\Sigma A_n \sigma_n^2 = \mathscr{E}$. Так как E — выпуклая $\cap$ по A_n функция, то необходимое и достаточное условие максимума E по A_n состоит в том, чтобы для некоторого λ и всех n

$$\frac{\partial E}{\partial A_n} = \frac{1}{2}\left[1 - \frac{1}{\beta_n(1+\rho) - A_n}\right] \leqslant \lambda\sigma_n^2 \tag{7.5.22}$$

с равенством, когда $A_n > 0$. Временно не будем рассматривать связь между β_n, накладываемую равенствами (7.5.19), а в дальнейшем покажем, что полученное решение будет удовлетворять (7.5.19) при некотором $s \geqslant 0$. Таким образом, так же как и в (7.4.26) и (7.4.28), имеем

$$\frac{\partial E}{\partial \beta_n} = \frac{1}{2}\left[-(1+\rho) + \frac{(1+\rho)}{\beta_n(1+\rho) - A_n} + \frac{\rho}{\beta_n}\right] = 0, \tag{7.5.23}$$

$$\beta_n = \frac{1}{2}\left(1 + \frac{A_n}{1+\rho}\right)\left[1 + \sqrt{1 - \frac{4A_n\rho}{(1+\rho+A_n)^2}}\right]. \tag{7.5.24}$$

Сочетая (7.5.22) и (7.5.23), получаем

$$\frac{\rho}{2(1+\rho)\beta_n} \leqslant \lambda\sigma_n^2. \tag{7.5.25}$$

Если ввести величину B, равную $\rho/[2(1+\rho)\lambda]$, то (7.5.25) принимает вид

$$\beta_n \sigma_n^2 \geqslant B \text{ с равенством, если } A_n > 0. \tag{7.5.26}$$

Далее из (7.5.24) видно, что если $A_n = 0$, то $\beta_n = 1$. Следовательно, если $A_n = 0$, то (7.5.26) означает, что $\sigma_n^2 \geqslant B$. Кроме того, из (7.5.23) видно, что $\partial E/\partial \beta_n$ убывает по β_n и возрастает по A_n. Таким образом, решение для β_n в (7.5.24) является возрастающей функцией A_n, и $\beta_n > 1$ для $A_n > 0$. Поскольку (7.5.26) удовлетворяется с равенством для $A_n > 0$, то это означает, что $\sigma_n^2 < B$, если $A_n > 0$. Другими словами, по всем каналам с дисперсией шума, меньшей чем B, посылаются сигналы с положительной энергией, а все каналы с дисперсией шума, превосходящей B, не используются (т. е. их входы всегда

равны 0). Для каналов, удовлетворяющих условию $\sigma_n^2 < B$, значения A_n с помощью (7.5.23) можно выразить через β_n и получить

$$A_n = \frac{(1+\rho)^2(\beta_n - 1)\beta_n}{(1+\rho)\beta_n - \rho}.$$

Так как $\beta_n = B/\sigma_n^2$ и $A_n = \mathscr{E}_n/\sigma_n^2$, то это равенство можно записать в виде

$$\mathscr{E}_n = \frac{(1+\rho)^2(B-\sigma_n^2)B}{(1+\rho)B - \rho\sigma_n^2}; \quad \sigma_n^2 < B. \tag{7.5.27}$$

Просуммировав по n, получим неявное выражение для параметра B

$$\mathscr{E} = \sum_{n:\sigma_n^2 < B} \frac{(1+\rho)^2(\beta - \sigma_n^2)B}{(1+\rho)B - \rho\sigma_n^2}. \tag{7.5.28}$$

Правая часть (7.5.28) — непрерывная возрастающая функция B. Это следует из того, что каждое слагаемое является непрерывной функцией и когда новое слагаемое входит в сумму, оно возрастает, начиная с нулевого значения. Следовательно, (7.5.28) имеет единственное решение для B, которое в свою очередь определяет β_n по (7.5.26) и $\mathscr{E}_n$ по (7.5.27). Если $\mathscr{E} = 0$, то любое s согласуется с (7.5.19). Если $\mathscr{E}_n > 0$ и β_n удовлетворяет (7.5.23), то значение s, удовлетворяющее (7.5.19), вычислено в (7.4.38) и равно

$$s = \frac{\rho A_n}{2(1+\rho)^2 \beta_n \mathscr{E}_n} = \frac{\rho}{2(1+\rho)^2 B}. \tag{7.5.29}$$

Таким образом, одно и то же значение s удовлетворяет (7.5.19) для всех n и наше решение является совместным.

Далее выполним максимизацию по ρ. Имеем

$$\frac{\partial E}{\partial \rho} = -R + \frac{1}{2}\sum_n \left[1 - \beta_n + \frac{\beta_n}{(1+\rho)\beta_n - A_n} - \right.$$
$$\left. - \frac{1}{1+\rho} + \ln \beta_n\right] = 0. \tag{7.5.30}$$

Так же как (7.4.29), это выражение упрощается к

$$R = \sum_n {}^1/_2 \ln \beta_n, \tag{7.5.31}$$

$$R = \sum_{n:\sigma_n^2 < B} {}^1/_2 \ln \frac{B}{\sigma_n^2}. \tag{7.5.32}$$

Наконец, $E_r(R)$ совпадает с E, вычисленным при заданных ρ, β_n и A_n,

$$E_r(R) = \frac{1}{2}\sum_n \left[(1-\beta_n)(1+\rho) + A_n + \ln\left(\beta_n - \frac{A_n}{1+\rho}\right)\right]. \tag{7.5.33}$$

В (7.5.33) только слагаемые с $\sigma_n^2 < B$ дают вклад в сумму. Используя соотношение (7.5.19) для слагаемых, не содержащих знак логарифма, и (7.5.23) для слагаемых со знаком логарифма, получаем

$$E_r(R) = \frac{1}{2} \sum_{n:\, \sigma_n^2 < B} \left[2s\mathscr{E}_n (1+\rho) - \ln\left(1+\rho-\frac{\rho}{\beta_n}\right)\right].$$

Суммируя по n и используя выражение (7.5.29) для s, получаем

$$E_r(R) = \frac{\rho\mathscr{E}}{2B(1+\rho)} - \sum_{n:\, \sigma_n^2 < B} {}^1/_2 \ln\left(1+\rho-\frac{\rho\sigma_n^2}{B}\right). \qquad (7.5.34)$$

Равенства (7.5.28), (7.5.32) и (7.5.34) — параметрические соотношения с параметрами B и ρ ($0 \leqslant \rho \leqslant 1$), связывающие энергию $\mathscr{E}(B, \rho)$, скорость $R(B)$ и показатель экспоненты $E_r(B, \rho)$. Легко видеть, что $\mathscr{E}(B, \rho)$ является непрерывной и строго возрастающей функцией B и ρ (при $B > \min \sigma_n^2$). Следовательно, уравнение $\mathscr{E} = \mathscr{E}(B, \rho)$ при фиксированной энергии $\mathscr{E}$ определяет B как функцию ρ и это неявно определяет $R(B)$ как функцию ρ (или ρ как функцию R и $\mathscr{E}$). Для $\rho = 0$, как это можно увидеть, сравнив (7.5.28) и (7.5.32) с теоремой 7.5.1, получающаяся скорость совпадает просто с пропускной способностью канала при заданном $\mathscr{E}$. Показатель экспоненты $E(B, \rho)$ равен нулю при $\rho = 0$. При фиксированной энергии $\mathscr{E}$ с возрастанием ρ величина B убывает, $R(B)$ также убывает, а $E(B, \rho)$ возрастает. Как и ранее, наклон E как функции R равен $-\rho$, что легче всего можно увидеть из графика рис. 5.6.4. Когда при фиксированном $\mathscr{E}$ значение ρ возрастает до 1, B убывает до критического значения B_{cr}, задаваемого соотношением

$$\mathscr{E} = \mathscr{E}(B_{cr}, 1) = \sum_{n:\, \sigma_n^2 \leqslant B_{cr}} \frac{4(B_{cr}-\sigma_n^2) B_{cr}}{2B_{cr}-\sigma_n^2}. \qquad (7.5.35)$$

Значению B_{cr} соответствует критическое значение скорости, определяемое по формуле

$$R_{cr} = \sum_{n:\, \sigma_n^2 \leqslant B_{cr}} {}^1/_2 \ln \frac{B_{cr}}{\sigma_n^2}. \qquad (7.5.36)$$

Параметрические уравнения (7.5.28), (7.5.32) и (7.5.34) применимы только при фиксированном $\mathscr{E}$ и для R из области $R_{cr} \leqslant R \leqslant C$. Для $R < R_{cr}$ показатель экспоненты E максимизируется на значении $\rho = 1$ с решениями для A_n и β_n, указанными выше. Это приводит к показателю экспоненты $E_r(R)$, задаваемому равенством

$$E_r(R) = E(B_{cr}, 1) + R_{cr} - R. \qquad (7.5.37)$$

Следует отметить, что во всем интервале скоростей распределение энергии по множеству каналов дается формулой (7.5.27) при соответствующим образом выбранных B и ρ. С изменением R при фиксированном $\mathscr{E}$ обе величины B и ρ изменяются и изменяется распределение

энергии. Другими словами, использование теоремы 7. 5.1 на практике для распределения энергии между параллельными каналами с гауссовыми шумами не всегда разумно, даже тогда, когда в системах применяются сложные системы кодирования.

Коэффициент $[2e^{s\delta}/\mu]^2$ в выражении вероятности ошибки может быть оценен как и ранее с помощью центральной предельной теоремы. Имеем*)

$$\mu \approx \frac{\delta}{\sqrt{4\pi \sum_n \mathscr{E}_n^2}}. \tag{7.5.38}$$

Выбирая $\delta = 1/s$ и применяя (7.5.29), получаем

$$\frac{2e^{s\delta}}{\mu} \approx \frac{e\rho \sqrt{4\pi \sum_n \mathscr{E}_n^2}}{(1+\rho)^2 B}. \tag{7.5.39}$$

Граница случайного кодирования для процедуры с выбрасыванием для параллельных каналов с аддитивными гауссовыми шумами получается как обобщение результата для одного канала с гауссовым шумом и доказывается тем же методом, что и граница случайного кодирования. Используя ансамбль кодов, описанный соотношением (7.5.9), найдем, что для любого $R \geqslant 0$ существует код с $M = \lceil e^{R'} \rceil$ кодовыми словами, каждое из которых удовлетворяет ограничению на энергию и

$$P_{e,\,m} \leqslant \exp\{-\rho R' + \sum_n E_x(\rho,\, X_n,\, Y_n,\, s)\}, \tag{7.5.40}$$

где

$$R' = R + 2 \ln \frac{2e^{s\delta}}{\mu}, \tag{7.5.41}$$

$$E_x(\rho,\, X_n,\, Y_n,\, s) = 2\rho s \mathscr{E}_n + \frac{\rho}{2} \ln\left(1 - 2s\mathscr{E}_n + \frac{\mathscr{E}_n}{2\rho\sigma_n^2}\right) + \frac{\rho}{2} \ln(1 - 2s\mathscr{E}_n).$$

Подставляя $A_n = \mathscr{E}_n/\sigma_n^2$ и

$$\beta_n = 1 - 2s\mathscr{E}_n + A_n/(2\rho), \tag{7.5.42}$$

выражение $E_x(\rho, X_n, Y_n, s)$ можно переписать следующим образом:

$$\widetilde{E}_x(A_n,\, \beta_n,\, \rho) = (1 - \beta_n)\rho + \frac{A_n}{2} + \frac{\rho}{2} \ln\left[\beta_n \left(\beta_n - \frac{A_n}{2\rho}\right)\right]. \tag{7.5.43}$$

Теперь найдем максимум величины

$$E = -\rho R + \sum_{n=1}^{N} \widetilde{E}_x(A_n,\, \beta_n,\, \rho) \tag{7.5.44}$$

*) Эта аппроксимация хороша лишь тогда, когда каждое $\mathscr{E}_n$ мало сравнительно с $\mathscr{E}$ (см. Феллер (1966), т. 2, задача 19, гл. 15). Также предполагается, что δ мало относительно стандартного отклонения $\sqrt{2\Sigma\mathscr{E}_n^2}$, таким образом, принимается, что плотность $\sum_n x_n^2$ постоянна между $\mathscr{E} - \delta$ и $\mathscr{E}$. Другими словами, приближение (7.5.38) точно лишь в том случае, если одновременно $\mu \ll 1$ и $\mathscr{E}_n \ll \mathscr{E}$ для всех n.

по A_n, β_n и ρ. На A_n опять наложены ограничения $A_n \geqslant 0$, $\sum \sigma_n^2 A_n = \mathscr{E}$. Так как E — выпуклая $\smile$ по A_n функция, то необходимые и достаточные условия для максимума E по A_n состоят в том, что для некоторого λ и всех n

$$\frac{\partial E}{\partial A_n} = \frac{1}{2}\left[1 - \frac{\rho}{2\rho\beta_n - A_n}\right] \leqslant \lambda\sigma_n^2 \tag{7.5.45}$$

с равенством, если $A_n > 0$. Опять временно не будем учитывать связи между β_n, задаваемыми (7.5.42), и максимизируем E по каждому β_n отдельно. Получаем

$$\frac{\partial E}{\partial \beta_n} = -\rho + \frac{\rho}{2\beta_n} + \frac{\rho^2}{2\rho\beta_n - A_n} = 0, \tag{7.5.46}$$

$$\beta_n = \frac{1}{2} + \frac{A_n}{4\rho} + \frac{1}{2}\sqrt{1 + \frac{A_n^2}{4\rho^2}}. \tag{7.5.47}$$

Для A_n и β_n, удовлетворяющих обоим равенствам (7.5.46) и (7.5.45), можно упростить (7.5.45), приведя его к виду

$$\frac{1}{4\beta_n} \leqslant \lambda\sigma_n^2. \tag{7.5.48}$$

Следовательно, если B положить равным $1/(4\lambda)$, то (7.5.48) примет вид

$$\beta_n \sigma_n^2 \geqslant B \tag{7.5.49}$$

с равенством, если $A_n > 0$.

Из (7.5.47) видно, что $\beta_n = 1$ при $A_n = 0$ и поэтому из (7.5.49) следует, что $\sigma_n^2 \geqslant B$. Имеем $\beta_n > 1$ при $A_n > 0$ и, следовательно, так как $\beta_n \sigma_n^2 = B$, то $\sigma_n^2 < B$. Итак, с положительной энергией используются только те каналы, для которых $\sigma_n^2 < B$.

Далее выведем выражение для $\mathscr{E}_n$. Из (7.5.46) находим

$$A_n = \frac{4\rho\beta_n(\beta_n - 1)}{2\beta_n - 1}. \tag{7.5.50}$$

Для $\sigma_n^2 < B$ имеем $\beta_n = B/\sigma_n^2$, и (7.5.50) принимает вид

$$\mathscr{E}_n = \frac{4\rho B(B - \sigma_n^2)}{2B - \sigma_n^2}, \tag{7.5.51}$$

$$\mathscr{E} = \sum_{n:\,\sigma_n^2 < B} \frac{4\rho B(B - \sigma_n^2)}{2B - \sigma_n^2}. \tag{7.5.52}$$

Равенство (7.5.52) дает неявное решение для B через $\mathscr{E}$ и оно, в свою очередь, определяет $\mathscr{E}_n$ из (7.5.51) и β_n из (7.5.49).

Для $A_n > 0$ и β_n, удовлетворяющих (7.5.46), значение s, удовлетворяющее (7.5.42), найдено в (7.4.55). Имеем

$$s = \frac{A_n}{8\beta_n \mathscr{E}\rho} = \frac{1}{8B\rho}. \tag{7.5.53}$$

Следовательно, одно и то же значение s удовлетворяет (7.5.42) для всех n и решения для β_n и $\mathscr{E}_n$ совместны.

Далее, E максимизируется по ρ, когда $\partial E/\partial\rho = 0$ или когда

$$R' = \sum_n (1-\beta_n) + \frac{A_n}{2(2\rho\beta_n - A_n)} + {}^1/_2 \ln\left[\beta_n\left(\beta_n - \frac{A_n}{2\rho}\right)\right]. \quad (7.5.54)$$

Так же как при переходе от (7.4.47) к (7.4.51), из полученного выражения будем иметь

$$R' = \sum_n {}^1/_2 \ln \frac{\beta_n^2}{2\beta_n - 1}, \quad (7.5.55)$$

$$E_{ex}(R') = \sum_n \frac{A_n}{4\beta_n}. \quad (7.5.56)$$

С помощью (7.5.49) эти равенства приводятся к виду

$$R' = \sum_{n:\,\sigma_n^2 < B} {}^1/_2 \ln \frac{B^2}{\sigma_n^2(2B - \sigma_n^2)}, \quad (7.5.57)$$

$$E_{ex}(R') = \frac{\mathscr{E}}{4B}. \quad (7.5.58)$$

Равенства (7.5.52), (7.5.57) и (7.5.58) связывают $\mathscr{E}$, R' и $E_x(R')$ параметрически через ρ и B. Они справедливы только для $\rho \geqslant 1$, которое для заданного $\mathscr{E}$ определяет верхнюю границу для B из (7.5.52). Сравнивая (7.5.52) при $\rho = 1$ с (7.5.36) видим, что этим предельным значением B будет как раз B_{cr}. Следовательно, при данном $\mathscr{E}$ граница для процедуры с выбрасыванием справедлива, когда

$$0 \leqslant R' \leqslant \sum_{n:\,\sigma_n^2 < B_{cr}} {}^1/_2 \ln \frac{B_{cr}^2}{\sigma_n^2(2B_{cr} - \sigma_n^2)}. \quad (7.5.59)$$

Наконец, применяя приближение (7.5.38) для μ, выбирая $\delta = 1/s$ и применяя равенства (7.5.53) для s, можно связать R' со скоростью R из (7.5.41) с помощью соотношения

$$R' \approx R + 2\ln\left[\frac{e}{4B\rho}\sqrt{4\pi\sum_n \mathscr{E}_n^2}\right]. \quad (7.5.60)$$

Результаты этого параграфа суммируются в следующей теореме, принадлежащей Эберту (1965).

Теорема 7.5.2. Пусть задано множество N параллельных дискретных по времени каналов с аддитивными гауссовыми шумами и дисперсии шумов равны $\sigma_1^2, \ldots, \sigma_N^2$. Для любого *) $B > 0$ и любого ρ, $0 \leqslant \rho \leqslant 1$, существует такой код с $M = \lceil e^{R(B)} \rceil$ кодовыми словами,

*) В тривиальном случае, когда $B \leqslant \min \sigma_n^2$, все значения $\mathscr{E}$, R и E равны нулю.

что каждое кодовое слово $\mathbf{x}_m$ удовлетворяет ограничению

$$\sum_{n=1}^{N} x_{m,n}^2 \leqslant \mathscr{E}(B, \rho)$$

и для каждого кодового слова вероятность ошибки ограничена неравенством

$$P_{e,m} \leqslant \left[\frac{2e^{s\delta}}{\mu}\right]^2 \exp\{-E(B, \rho)\}, \tag{7.5.61}$$

где $R(B)$, $\mathscr{E}(B, \rho)$ и $E(B, \rho)$ задаются равенствами (7.5.32), (7.5.28), и (7.5.34). Для фиксированного $\mathscr{E} = \mathscr{E}(B, \rho)$ при изменении ρ от 0 до 1 $R(B)$ строго и непрерывно убывает от C до R_{cr}, а $E(B, \rho)$ строго и непрерывно возрастает от 0 до $E(B_{cr}, 1)$. Для $R < R_{cr}$ существуют коды с $M = \lceil e^R \rceil$ кодовыми словами, энергия каждого из которых не более $\mathscr{E}$, и для каждого слова

$$P_{e,m} \leqslant \left[\frac{2e^{s\delta}}{\mu}\right]^2 \exp\{-[E(B_{cr}, 1) + R_{cr} - R]\}. \tag{7.5.62}$$

Кроме того, для любого

$$B \geqslant \min \sigma_n^2$$

и любого $\rho \geqslant 1$ существуют коды с $M = \lceil \exp\{R_x'(B) - 2\ln(2e^{s\mu}/\delta)\} \rceil$ словами, энергия каждого из которых не более $\mathscr{E}_x(B, \rho)$, и для каждого слова

$$P_{e,m} \leqslant \exp[-E_{ex}(B, \rho)], \tag{7.5.63}$$

где R_x', $\mathscr{E}_x$ и E_{ex} задаются формулами (7.5.57), (7.5.52) и (7.5.58) соответственно. Для фиксированного $\mathscr{E} = \mathscr{E}_x(B, \rho)$ при возрастании ρ от 1 до ∞, функция $R_x'(B)$ строго и непрерывно убывает от $R_{x\,cr}$ до 0, а $E_x(B, \rho)$ строго возрастает.

В следующей главе будет показано, что при применении этих результатов для параллельных каналов к каналам с непрерывным временем коэффициент в (7.5.61) и разность между R и R' в (7.5.60) будут несущественны и основное внимание сконцентрируется на показателях экспонент $E_r(R)$ и $E_{ex}(R')$. Эберт (1965) вывел также нижние границы вероятности ошибки для параллельных каналов с аддитивными гауссовыми шумами; показатель экспоненты его нижней границы совпадает с $E_r(R)$ для $R_{cr} \leqslant R \leqslant C$.

ИТОГИ И ВЫВОДЫ

В этой главе результаты гл. 4 и 5 были распространены на дискретные по времени каналы без памяти с произвольными входными и выходными пространствами. Основу развиваемого здесь подхода составляло использование лишь конечного множества букв входного алфавита и разбиение выходного алфавита для сведения канала к дискрет-

ному каналу без памяти. Главное отличие этой главы от гл. 4 и 5 состоит во введении ограничений на входной алфавит. В §§ 7.4 и 7.5 общий результат первых трех параграфов был применен к важным случаям дискретных по времени каналов с аддитивным гауссовым шумом и параллельных дискретных по времени каналов с аддитивными гауссовыми шумами. Важность этих каналов станет ясной в гл. 8, когда непрерывный по времени канал с аддитивным гауссовым шумом будет сведен к множеству параллельных дискретных по времени каналов с аддитивными гауссовыми шумами.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Большинство результатов §§ 7.1 и 7.3 новые, за исключением тех, которые ранее появились в работе, изучавшей ограничения на входе (Галлагер, 1965). Однако некоторые результаты были получены независимо Вагнером (1968). Граница вероятности ошибки при случайном кодировании для дискретного по времени канала была получена Шенноном (1959), который также вывел нижнюю границу сферической упаковки для вероятности ошибки. Слепян (1963) вычислил значение этих верхних и нижних границ для некоторого интервала кодовых длин и числа кодовых слов. Результаты § 7.5 о параллельных дискретных по времени каналах с аддитивными шумами принадлежат Эберту (1966), который также вывел нижние границы вероятности ошибки для этих каналов.

8

НЕПРЕРЫВНЫЕ КАНАЛЫ

8.1. ОРТОНОРМАЛЬНЫЕ РАЗЛОЖЕНИЯ СИГНАЛОВ И БЕЛЫЙ ГАУССОВ ШУМ

В этой главе рассматриваются каналы, в которых сигналы на входе и выходе являются функциями времени и время здесь определяется на континууме, а не в дискретных точках. Это сразу же приводит к необходимости рассмотрения понятия «вероятности функции». Вероятность одного события из дискретного множества событий — понятие довольно простое, и при введении функции распределения вероятностей случайной величины с непрерывным множеством значений не нужно слишком сильно пересматривать основные понятия. Конечно, можно было бы описать случайные функции (или случайные процессы, как их обычно называют) в некоторый момент времени распределением вероятности, однако в общем случае даже совместное распределение вероятностей для большого числа моментов времени было бы недостаточным для полного статистического описания процесса. В принципе случайный процесс считается полностью заданным, если имеется правило, по которому может быть вычислено совместное распределение вероятностей для любого конечного множества моментов времени. Мы не будем следовать этому подходу в настоящей главе, а рассмотрим вместо этого подход, основанный на представлении любой заданной действительной или комплексной функции с помощью разложения в ряд по ортонормальным функциям. При этом случайная функция будет описываться с помощью совместного распределения вероятностей коэффициентов такого разложения в ряд. Хотя вначале этот подход может показаться довольно абстрактным и громоздким, он окажется впоследствии более полезным как для развития интуиции, так и при доказательстве теорем по сравнению с подходом, основанным на описании поведения функций в различные моменты времени.

На протяжении этой главы мы будем иметь дело с некоторыми математическими понятиями, такими, как сходимость рядов, изменение порядка суммирования и интегрирования, и существование пределов. Связанные с ними вопросы нельзя разрешить на основе физических соображений, поскольку они относятся только к математическим моделям физических задач. В действительности, часто можно глубже проникнуть в суть физических явлений, исследуя случаи, когда математические пределы перестают существовать. Вместе с тем, если непрестанно беспокоиться о сходимости и пределах, то мы затем-

ним более важные вопросы и потеряем читателей, которые не имеют либо подготовки, либо склонности следовать сложным обоснованиям предельного перехода. Многие из этих вопросов, касающихся сходимости, можно обойти или по крайней мере упростить, если ограничиться рассмотрением функций конечной энергии, т. е. функций $x(t)$, для которых*) $\int |x(t)|^2 dt < \infty$. Эти функции часто называют функциями из L_2. Хотя они образуют достаточно общий класс функций, тем не менее отметим, что как импульсные функции, так и синусоиды бесконечной длительности не будут функциями с конечной энергией.

Две функции $\varphi_1(t)$ и $\varphi_2(t)$ называются ортогональными, если $\int \varphi_1(t)\,\varphi_2^*(t)\,dt = 0$, где $\varphi_2^*(t)$ — функция, комплексно-сопряженная с $\varphi_2(t)$. Функция называется нормированной, если ее энергия $\int |\varphi_1(t)|^2 dt$ равна 1. *Ортонормальное множество определяется как множество функций* $\varphi_1(t)$, $\varphi_2(t)$, ..., *каждая из которых нормирована и каждая пара которых ортогональна*; таким образом, для всех φ_i, φ_j из множества имеем

$$\int \varphi_i(t)\,\varphi_j^*(t)\,dt = \delta_{ij}, \tag{8.1.1}$$

где $\delta_{ij} = 1$ для $i = j$ и $\delta_{ij} = 0$ в других случаях.

Предположим теперь, что функция $x(t)$ может быть представлена через ортонормальное множество функций в виде

$$x(t) = \sum_{i=1}^{k} x_i\,\varphi_i(t). \tag{8.1.2}$$

В этих условиях коэффициенты x_i удовлетворяют соотношениям

$$x_i = \int x(t)\,\varphi_i^*(t)\,dt. \tag{8.1.3}$$

Чтобы увидеть это, нужно подставить (8.1.2) в (8.1.3) и проинтегрировать полученное выражение, учитывая (8.1.1)

Пусть теперь $x(t)$ — произвольная функция из L_2 и пусть x_i определяется (8.1.3). Следует исследовать, может ли разложение

$$\sum_{i=1}^{\infty} x_i\,\varphi_i(t)$$

по-прежнему быть использовано для представления $x(t)$. Пусть $x_{r,k}(t)$ — остаток, получающийся, когда k членов разложения использованы для представления $x(t)$:

$$x_{r,k}(t) = x(t) - \sum_{i=1}^{k} x_i\,\varphi_i(t). \tag{8.1.4}$$

Если умножить обе части (8.1.4) на $\varphi_i^*(t)$ и проинтегрировать, то можно сразу же заметить, что $x_{r,k}(t)$ ортогональна $\varphi_i(t)$ при всех $i \leqslant k$.

---

*) На протяжении этой главы любой интеграл, в котором не указаны пределы, берется в пределах от $-\infty$ до $+\infty$.

Энергия $x_{r,\,k}(t)$ задается выражением

$$\int |x_{r,k}(t)|^2\,dt = \int \left[x(t) - \sum_{i=1}^{k} x_i \varphi_i(t)\right]\left[x^*(t) - \sum_{i=1}^{k} x_i^* \varphi_i^*(t)\right] dt =$$

$$= \int |x(t)|^2\,dt - \sum_i x_i^* \int x(t)\,\varphi_i^*(t)\,dt - \sum_i x_i \int x^*(t)\,\varphi_i(t)\,dt +$$

$$+ \sum_{i,j} x_i x_j^* \int \varphi_i(t)\,\varphi_j^*(t)\,dt = \int |x(t)|^2\,dt - \sum_{i=1}^{k} |x_i|^2. \qquad (8.1.5)$$

При выводе (8.1.5) была использована формула (8.1.3) и ее комплексное сопряжение $x_i^* = \int x^*(t)\varphi_i(t)\,dt$.

Так как энергия $x_{r,\,k}$ неотрицательна для всех k, то

$$\sum_{i=1}^{k} |x_i|^2 \leqslant \int |x(t)|^2\,dt, \qquad (8.1.6)$$

$$\sum_{i=1}^{\infty} |x_i|^2 \leqslant \int |x(t)|^2\,dt. \qquad (8.1.7)$$

Соотношение (8.1.7) известно как неравенство Бесселя и оно будет часто использоваться в последующем изложении.

Теперь можно исследовать предельное поведение $x_{r,k}(t)$, рассматривая разность между $x_{r,k}(t)$ и $x_{r,l}(t)$ при $l > k$. Из (8.1.4) следует, что разность равна $\sum_{i=k+}^{l} x_i \varphi_i(t)$ и имеет энергию $\sum_{i=k+1}^{l} |x_i|^2$.

Неравенство Бесселя показывает, что $\sum_{i=1}^{\infty} |x_i|^2$ ограничена и поэтому $\sum_{i=k+1}^{l} |x_i|^2$ должна стремиться к 0, при неограниченном возрастании k и l. Таким образом, $x_{r,k}(t)$ стремится к пределу*) $x_r(t)$, который ортогонален ко всем $\varphi_i(t)$:

$$x_r(t) = \text{l. i. m.}_{k\to\infty} x_{r,k}(t). \qquad (8.1.8)$$

Обозначение l. i. m. в (8.1.8) используется для предела в среднем**). Под этим понимается, что $\lim_{k\to\infty} \int |x_r(t) - x_{r,k}(t)|^2\,dt = 0$ и это равенство не обязательно означает, что $x_{r,\,k}(t)$ сходится к пределу для каждого значения t. Теперь можно представить $x(t)$ как

$$x(t) = \sum_{i=1}^{\infty} x_i \varphi_i(t) + x_r(t), \qquad (8.1.9)$$

*) Существование этой предельной функции вытекает из теоремы Рисса — Фишера [см., например, Рисс и Надь (1955)].

**) Чаще такой предел называют пределом в среднем квадратичном (*Прим. ред.*).

$$x_i = \int x(t)\,\varphi_i^*(t)\,dt; \quad \int x_r(t)\,\varphi_i^*(t)\,dt = 0 \text{ для всех } i.$$

Под бесконечной суммой в (8.1.9), а также на протяжении этой главы понимается предел

$$\lim_{k\to\infty} \sum_{i=1}^{k} x_i\,\varphi_i(t).$$

На геометрическом языке (8.1.9) утверждает, что функция $x(t)$ может быть разложена на две составляющие: одну в подпространстве, порожденном $\varphi_i(t)$, и другую — ортогональную подпространству $\varphi_i(t)$.

Обычно при использовании (8.1.9) с бесконечным рядом можно обращаться как с конечным. Как правило, это можно оправдать с помощью неравенства Шварца, которое устанавливает, что для двух функций $x(t)$ и $y(t)$ из L_2

$$\left|\int x(t)\,y^*(t)\,dt\right|^2 \leqslant \int |x(t)|^2\,dt \int |y(t)|^2\,dt. \tag{8.1.10}$$

Для того чтобы проверить справедливость (8.1.10), положим, что

$$\varphi_1(t) = y(t)/\sqrt{\int |y(t)|^2\,dt} \tag{8.1.11}$$

будет ортонормальным множеством, содержащим лишь один элемент. Тогда неравенство Бесселя, примененное к $x(t)$, имеет вид

$$\left|\int x(t)\,\varphi_1^*(t)\,dt\right|^2 \leqslant \int |x(t)|^2\,dt. \tag{8.1.12}$$

Подставляя (8.1.11) в (8.1.12), получаем (8.1.10).

Для того чтобы построить пример, показывающий, как неравенство Шварца может применяться, когда имеют дело с разложением функции $x(t)$ в бесконечный ряд (8.1.9), рассмотрим интеграл

$$\int \sum_{i=1}^{\infty} x_i\,\varphi_i(t)\,y^*(t)\,dt = \sum_{i=1}^{k} x_i \int \varphi_i(t)\,y^*(t)\,dt + \int \left[\sum_{i=k+1}^{\infty} x_i\,\varphi_i(t)\right] y^*(t)\,dt. \tag{8.1.13}$$

Применяя неравенство Шварца к последнему интегралу в (8.1.13), получаем

$$\left|\int \sum_{i=k+1}^{\infty} x_i\,\varphi_i(t)\,y^*(t)\,dt\right|^2 \leqslant \int \left|\sum_{i=k+1}^{\infty} x_i\,\varphi_i(t)\right|^2 dt \times \int |y(t)|^2\,dt = \sum_{i=k+1}^{\infty} |x_i|^2 \int |y(t)|^2\,dt. \tag{8.1.14}$$

Вспоминая, что рассматривается только функция с интегрируемым квадратом, находим, что предел правой части (8.1.14) при $k \to \infty$

равен 0. А тогда можно перейти к пределу в (8.1.13) и, следовательно, изменить порядок интегрирования и суммирования:

$$\int \sum_{i=1}^{\infty} x_i \varphi_i(t)\, y^*(t)\, dt = \sum_{i=1}^{\infty} x_i \int \varphi_i(t)\, y^*(t)\, dt. \qquad (8.1.15)$$

Говорят, что множество ортонормированных функций полно в классе функций, если все функции этого класса содержатся в подпространстве, порожденном $\varphi_i(t)$, или иначе, если $x_r(t)$ равно нулю для всех функций класса. В этом случае ясно, что член в левой части (8.1.5) стремится к нулю с возрастанием k и, следовательно,

$$\int |x(t)|^2 dt = \sum_{i=1}^{\infty} |x_i|^2, \qquad (8.1.16)$$

если $x(t)$ принадлежит подпространству, порожденному множеством $\varphi_i(t)$. Равенство (8.1.16) называется энергетическим уравнением и часто будет использоваться в последующем изложении.

Если $x(t)$ принадлежит подпространству, порожденному ортонормальным множеством $\varphi_i(t)$ и если $y(t)$ — любая другая функция из L_2, то имеет место также равенство Парсеваля

$$\int x(t)\, y^*(t)\, dt = \sum x_i\, y_i^*, \qquad (8.1.17)$$

где $y_i = \int y(t)\, \varphi_i^*(t)\, dt$. Для того чтобы вывести (8.1.17), заметим, что $x(t) = \sum x_i \varphi_i(t)$ и, следовательно, (8.1.17) эквивалентно (8.1.15).

В качестве довольно частого примера множества ортонормальных функций рассмотрим

$$\varphi_i(t) = \begin{cases} \sqrt{\frac{1}{T}} \exp\left(\frac{j2\pi it}{T}\right); & |t| \leqslant T/2, \\ 0 & \text{в других точках,} \end{cases} \qquad (8.1.18)$$

где $j = \sqrt{-1}$ и i — какое-либо целое число. Тогда равенство (8.1.9) принимает вид

$$x(t) = \begin{cases} \sqrt{\frac{1}{T}} \sum_{i=-\infty}^{\infty} x_i \exp\left(\frac{j2\pi it}{T}\right); & |t| \leqslant T/2, \\ 0 & \text{в других точках,} \end{cases} \qquad (8.1.19)$$

где

$$x_i = \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} x(t) \exp\left(\frac{-j2\pi it}{T}\right) dt.$$

Это просто разложение функции на интервале $(-T/2, T/2)$ в ряд Фурье. Это множество функций полно в классе функций с конечной энергией, определенных на интервале $(-T/2, T/2)$*) и, следовательно, остаточный член $x_r(t)$ отвечает $x(t)$ вне этого интервала.

*) См., например, Ахиезер и Глазман, стр. 41 (1950).

Предположим, что мы хотим построить множество сигналов, которые ограничены во времени интервалом $(-T/2, T/2)$, и также приближенно ограничены по частоте частотами, меньшими некоторого максимального значения W. Можно сделать некоторое продвижение в решении этой задачи, если рассмотреть $\varphi_i(t)$ из (8.1.18) как комплексную синусоиду частоты i/T. Так как $\varphi_i(t)$ усечена, то она не имеет ограниченную полосу частот; ее преобразование Фурье имеет вид

$$\Phi_i(f) = \int \varphi_i(t)\, e^{-j2\pi f t}\, dt = \sqrt{T}\, \frac{\sin \pi T (f - i/T)}{\pi T (f - i/T)}. \qquad (8.1.20)$$

График функции $\Phi_i(f)$ изображен на рис. 8.1.1 и из него ясно, что $\varphi_i(t)$ имеет наибольшую энергию на частотах в окрестности i/T. Если рассматриваются функции, являющиеся линейными комбинациями $\varphi_i(t)$ с $-WT \leqslant i \leqslant WT$, где WT — целое, то

$$x(t) = \sum_{i=-WT}^{WT} x_i \varphi_i(t). \qquad (8.1.21)$$

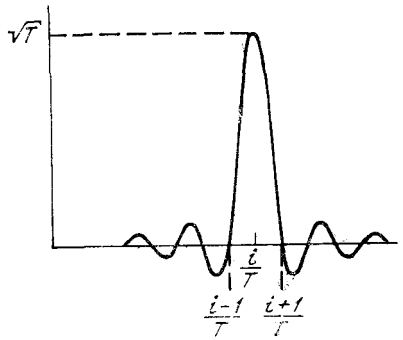


Рис. 8.1.1. Вид функции $\Phi_i(f) = \sqrt{T}\, \frac{\sin \pi T [f - (i/T)]}{\pi T [f - (i/T)]}$.

Этот класс функций строго ограничен во времени интервалом $(-T/2, T/2)$ и в некотором смысле ограничен по частоте полосой $(-W, W)$. Следовательно, произвольная комплексная функция в этом классе задается с помощью $2WT + 1$ комплексных чисел. Если потребовать, чтобы $x(t)$ была действительной, то $x_{-i} = x_i^*$ и $x(t)$ определяется $2WT + 1$ действительными числами, а именно числом x_0, которое действительно, и действительными и мнимыми частями x_i с $i > 0$. О классе функций, в котором любая функция может быть определена n действительными числами, говорят, что он имеет n степеней свободы и, следовательно, класс действительных функций $x(t)$, удовлетворяющих (8.1.21), имеет $2WT + 1$ степеней свободы. Заметим, что не имеет никакого смысла говорить о числе степеней свободы функций без указания вначале класса функций, к которому она принадлежит.

Если попытаться уточнить смысл, в котором функции, удовлетворяющие (8.1.21), имеют ограниченную полосу частот, то можно столкнуться с целым рядом проблем и в действительности может случиться так, что при достаточно специальном выборе x_i наибольшая часть энергии $x(t)$ окажется вне полосы частот $(-W, W)$. В § 8.4 этот вопрос о сигнале, ограниченном во времени и по частоте, получит более удовлетворительное математическое рассмотрение на основе использова-

ния другого множества ортонормальных функций. Однако рассматриваемое здесь приближение, использующее ограниченные во времени синусоиды, весьма полезно для более глубокого проникновения в различные проблемы техники связи; этого не следует избегать в силу недостаточной точности понятия ограниченной полосы частот.

Второе множество ортонормальных функций, которое весьма полезно в приобретении навыков понимания существа многих задач, составляют отсчетные функции

$$\theta_i(t) = \sqrt{2W}\,\frac{\sin 2\pi W\,[t - i/(2W)]}{2\pi W\,[t - i/(2W)]}. \tag{8.1.22}$$

Для того чтобы увидеть, что эти функции ортонормальны, надо прежде всего установить равенство Парсеваля, связывающее преобразования Фурье. Пусть $X(f)$ и $Y(f)$ — преобразования Фурье функций $x(t)$ и $y(t)$. Тогда*)

$$\int x(t)\,y^*(t)\,dt = \int X(f)\,Y^*(f)\,df. \tag{8.1.23}$$

Следовательно, если $\{\varphi_i(t)\}$ — какое-либо множество ортонормальных функций и $\Phi_i(f)$ — преобразование Фурье $\varphi_i(t)$ для каждого i, то $\{\Phi_i(f)\}$ — также множество ортонормальных функций, удовлетворяющих

$$\delta_{i,l} = \int \varphi_i(t)\,\varphi_l^*(t)\,dt = \int \Phi_i(f)\,\Phi_l^*(f)\,df. \tag{8.1.24}$$

Полагая, что $\Phi_i(f)$ задается (8.1.20) и подставляя t вместо f и $2W$ вместо T, видим, что $\theta_i(t)$ в (8.1.22) являются ортонормальными.

Используя соотношение (8.1.20), можно найти преобразование Фурье $\theta_i(t)$

$$\Theta_i(f) = \begin{cases} \sqrt{\dfrac{1}{2W}}\exp\left(-\dfrac{j2\pi i f}{2W}\right); & |f| \leqslant W, \\ 0 \qquad\qquad\qquad\quad ; & |f| > W. \end{cases} \tag{8.1.25}$$

Таким образом, мы видим, что все $\theta_i(t)$ имеют полосу частот, ограниченную W, в том смысле, что их преобразования Фурье равны 0 при $|f| > W$.

Функции $\theta_i(t)$ называются отсчетными функциями, поскольку любая функция $x(t)$ с полосой частот, ограниченной $|f| \leqslant W$, может быть представлена через эти функции и значения $x(t)$ в точках, отделенных интервалами $1/2W$:

$$x(t) = \sum_{i=-\infty}^{\infty} \sqrt{\frac{1}{2W}}\,x\left(\frac{i}{2W}\right)\theta_i(t). \tag{8.1.26}$$

*) Для общих функций с конечной энергией эти преобразования Фурье существуют в том смысле, что $X(f) = \lim\limits_{T\to\infty} \int\limits_{-T}^{T} x(t)\,e^{-j2\pi ft}\,dt$; $x(t) = \lim\limits_{F\to\infty} \int\limits_{-F}^{F} X(f) e^{j2\pi ft}\,df$ и (8.1.23) всегда справедливо. См. Титчмарш (1948), теоремы 48 и 49.

Для того чтобы вывести (8.1.26), обозначим через $X(f)$ преобразование Фурье $x(t)$. Так как $X(f) = 0$ для $|f| > W$, то она может быть разложена в ряд Фурье

$$X(f) = \sum_{i=-\infty}^{\infty} x_i \Theta_i(f), \tag{8.1.27}$$

$$x_i = \int X(f) \Theta_i^*(f)\, df.$$

Используя (8.1.25) и тот факт, что $X(f) = 0$, для $|f| > W$, имеем

$$x_i = \int_{-W}^{W} X(f) \sqrt{\frac{1}{2W}} \exp\left(\frac{j2\pi if}{2W}\right) df =$$

$$= \sqrt{\frac{1}{2W}} \int_{-\infty}^{\infty} X(f) \exp\left(\frac{j2\pi if}{2W}\right) df = \sqrt{\frac{1}{2W}}\, x\left(\frac{i}{2W}\right).$$

Подставляя это выражение для x_i в (8.1.27) и беря преобразование Фурье, получаем (8.1.26). В приведенном выше выводе предполагалось, что $x(t)$ обладает достаточно хорошим поведением, так что обратное преобразование $X(f)$ всюду сходится к $x(t)$.

Теперь аналогично тому как разложение Фурье использовалось для образования сигналов, ограниченных во времени интервалом $(-T/2, T/2)$ и приближенно ограниченных по полосе частот $|f| \leqslant W$, можно использовать разложение по выборочным функциям для образования сигналов с точно ограниченной полосой частот и приближенно ограниченных во времени. В этом случае следует использовать $\theta_i(t)$, для которых $|i| \leqslant WT$. Опять получаем $2WT + 1$ степеней свободы, и функция $x(t)$ из рассматриваемого класса равна нулю во всех точках отсчета с $|i/(2W)| > T/2$. Это представление рассматривается при тех же самых ограничениях, что и представление рядом Фурье, и на самом деле это то же самое представление, в котором, однако, роль времени играет частота, и наоборот.

Гауссовские случайные процессы

В этом параграфе дается краткое описание гауссовских случайных процессов и показывается, почему они так часто являются разумной моделью для реальных шумов. Случайный процесс*) $z(t)$ можно представлять себе как множество функций с вероятностной мерой, заданной на этом множестве. Точнее, процесс можно задать как совокупность случайных величин $z(t)$: каждая отдельная случайная величина соответствует каждому действительному значению параметра t. Одним из методов задания случайного процесса служит правило, которое

*) Более точно, здесь рассматриваются случайные процессы с непрерывным параметром, в отличие от случайных процессов с дискретным параметром, рассмотренных в § 3.5.

каждому конечному множеству моментов времени $t_1, \ldots, t_n$ относит совместную функцию распределения $F_{t_1, t_2, \ldots, t_n}(z_1, z_2, \ldots, z_n)$ случайных величин $z(t_1), z(t_2), \ldots, z(t_n)$. При каждом выборе действительных чисел $z_1, z_2, \ldots, z_n$ это распределение является вероятностью того, что $z(t_1) \leqslant z_1, z(t_2) \leqslant z_2, \ldots, z(t_n) \leqslant z_n$. Случайный процесс называется стационарным, если вероятностная мера инвариантна относительно сдвига во времени или, точнее, если для каждого конечного множества моментов времени $t_1, \ldots, t_n$, для каждого интервала времени T и для каждого выбора действительных чисел $z_1, \ldots, z_n$ имеем

$$F_{t_1, \ldots, t_n}(z_1, \ldots, z_n) = F_{t_1+T, \ldots, t_n+T}(z_1, \ldots, z_n). \qquad (8.1.28)$$

Говорят, что среднее значение случайного процесса равно нулю, если для каждого t математическое ожидание $z(t)$ равно нулю. В дальнейшем будут рассматриваться лишь процессы с нулевым средним. В действительности это не приводит здесь к потере общности, так как произвольный случайный процесс может быть разбит на две части: $\overline{z(t)}$ и $z(t) - \overline{z(t)}$, где $\overline{z(t)}$ — детерминированная функция (хотя не обязательно из L_2), а $z(t) - \overline{z(t)}$ — случайный процесс с нулевым средним.

Автокорреляционной функцией случайного процесса $z(t)$ называется функция двух действительных переменных, определяемая равенством

$$\mathcal{R}(t_1, t_2) = \overline{z(t_1)\, z(t_2)}. \qquad (8.1.29)$$

Автокорреляционная функция случайного процесса, очевидно, не дает полного описания процесса, однако, как сейчас будет показано, описание с ее помощью достаточно для ответа на вопросы, касающиеся линейной фильтрации случайных процессов.

Предположим, что линейный инвариантный во времени фильтр имеет импульсный отклик $h(t)$. Под этим понимается, что функция на выходе фильтра равна свертке функции на входе с $h(t)$ и, следовательно, если вход — случайный процесс $z(t)$, то выход — другой случайный процесс $y(t)$, определяемый равенством*)

$$y(t) = \int h(t - \tau)\, z(\tau)\, d\tau. \qquad (8.1.30)$$

Тогда автокорреляционная функция $y(t)$ задается равенством

$$\mathcal{R}_y(t_1, t_2) = \overline{y(t_1)\, y(t_2)} =$$

$$= \overline{\iint h(t_1 - \tau_1)\, h(t_2 - \tau_2)\, z(\tau_1)\, z(\tau_2)\, d\tau_1\, d\tau_2}. \qquad (8.1.31)$$

*) Рассуждения, которые начинаются здесь и оканчиваются формулой (8.1.34), нужны лишь для введения и наглядного разъяснения и поэтому здесь опускается точное определение того, как понимать интеграл в (8.1.30), а также опускается любое обоснование изменения порядка интегрирования и математического ожидания в (8.1.32). Более точное рассмотрение этого можно найти, например, у Давенпорта и Рута (1958), гл. 4, и у Яглома (1952), гл. 1 и 2.

Изменяя порядок интегрирования и математического ожидания, получаем

$$\mathcal{R}_y(t_1, t_2) = \iint h(t_1 - \varepsilon_1)\, h(t_2 - \tau_2)\, \mathcal{R}_z(\tau_1, \tau_2)\, d\tau_1\, d\tau_2. \quad (8.1.32)$$

Следовательно, корреляционная функция случайного процесса на выходе линейного инвариантного во времени фильтра определяется по корреляционной функции случайного процесса на входе. Если случайный процесс $z(t)$ стационарный, то $\mathcal{R}_z(t_1, t_2)$ — функция только разности $t = t_1 - t_2$ и $\mathcal{R}_z$ выражается как функция только одного переменного $\mathcal{R}_z(t)$. Случайный процесс называется *стационарным в широком смысле*, если его автокорреляционная функция $\mathcal{R}(t_1, t_2)$ является функцией только $t = t_1 - t_2$. Из (8.1.32) легко заметить, что если $z(t)$ стационарен в широком смысле, то $y(t)$ также стационарен в широком смысле. Если $z(t)$ стационарен в широком смысле, то (8.1.32) можно интерпретировать как свертку $\mathcal{R}_z(t)$ с $h(t)$ и полученного результата с $h(-t)$. Поэтому если определить *спектральную плотность мощности* $S_z(f)$ стационарного в широком смысле случайного процесса $z(t)$ как преобразование Фурье $\mathcal{R}_z(t)$, $S_z(f) = \int \mathcal{R}_z(t)\, e^{-j2\pi ft}\, dt$ и ввести $H(f) = \int h(t)\, e^{-j2\pi ft}\, dt$ — частотную характеристику фильтра, то получим

$$S_y(f) = S_z(f)\, |H(f)|^2. \quad (8.1.33)$$

Для истолкования смысла спектральной плотности мощности определим *мощность* стационарного в широком смысле случайного процесса $y(t)$, как $\overline{y^2(t)} = \mathcal{R}_y(0)$. Так как $S_y(f)$ — преобразование Фурье $\mathcal{R}_y(t)$, то имеем

$$\overline{y^2(t)} = \int_{-\infty}^{\infty} S_y(f)\, df = \int_{-\infty}^{\infty} S_z(f)\, |H(f)|^2\, df. \quad (8.1.34)$$

Если $|H(f)|^2$ равно единице в узкой полосе частот и нулю в других точках, то мощность на выходе фильтра равна интегралу от $S_z(f)$ по этой узкой полосе, и $S_z(f)$ физически можно интерпретировать как плотность мощности на единицу полосы частот на частоте f.

Рассмотрим теперь кратко реальные физические шумы и выясним, почему часто они адекватно могут моделироваться с помощью одного частного класса случайных процессов, называемых гауссовскими случайными процессами. Для многих шумов, по существу, равна нулю физическая связь между значениями шума в любые два момента времени, отделяемые более чем очень малым интервалом Δ, который называется *временем когерентности* шума*). При создании моделирующего шум случайного процесса целесообразно принять, что шум приближенно статистически независим в два момента времени, отделенных более чем интервалом Δ. Следовательно, если такой шум подается на вход фильтра с импульсным откликом $h(t)$ и если $h(t)$ существенно отличен от нуля на интервале, много большем чем Δ, то на основании

*) Для процессов, содержащих последовательность импульсов, Δ должно быть также большим, чем среднее время действия импульсов. (*Прим. ред.*)

центральной предельной теоремы следует ожидать, что отклик фильтра в любой заданный момент времени вполне можно моделировать гауссовской случайной величиной. Если Δ столь мало, что это предположение приемлемо для всех интересующих нас фильтров, то модель шума как случайного процесса можно упростить, приняв, что выход любого линейного фильтра в любой данный момент времени является гауссовской случайной величиной. Такой случайный процесс известен как гауссовский. Более точно, *случайный гауссовский процесс**) $z(t)$ с нулевым средним определяется как процесс, обладающий тем свойством, что для любой функции $x(t)$ из L_2 значение $\int x(t)\, z(t)\, dt$ является гауссовской случайной величиной с нулевым средним и конечной дисперсией.* Для ранее рассмотренных случайных процессов этот интеграл можно понимать обычным образом. Однако здесь желательно рассмотреть несколько более широкий класс случайных процессов, включающий *белый гауссовский случайный процесс (или белый гауссов шум), который определяется на основе того свойства, что для любой функции $x(t)$ из L_2 $x = \int x(t)\, z(t)\, dt$ — гауссовская случайная величина с нулевым средним и дисперсией*

$$\overline{x^2} = \frac{N_0}{2} \int x^2(t)\, dt, \tag{8.1.35}$$

где $N_0/2$ — постоянная, не зависящая от $x(t)$. Как мы увидим далее, этот процесс столь «дикий», что его нельзя определить как совокупность случайных величин, каждая из которых соответствует одному из значений параметра t. Вместе с тем, все что будет нужно в последующем изложении — это случайные величины $\int x(t)\, z(t)\, dt$ и, следовательно, случайный процесс будем считать определенным, если имеется правило, задающее случайную величину $\int x(t)\, z(t)\, dt$ для всех функций $x(t)$ из L_2. При таком подходе нет нужды беспокоиться о том, что означает указанный выше интеграл или каким образом определить $z(t)$ как совокупность случайных величин параметра t. Указанный выше подход подобен тому, который используется при рассмотрении обобщенных функций, когда обобщенные функции не определяются через их значения для каждого значения аргумента, а вместо этого они определяются через интеграл от их произведения на каждую функцию из подходящим образом определенного класса функций. По этой причине белый гауссов шум обычно называют обобщенным случайным процессом. Следующее условие линейности должно быть наложено на случайные величины $\int \vartheta(t)\, z(t)\, dt$: для любого множества постоянных $a_1, \ldots, a_k$ и любого множества функций $\vartheta_1(t), \ldots, \vartheta_k(t)$ из L_2 требуется, чтобы

$$\int \left[\sum_{i=1}^{k} a_i\, \vartheta_i(t) \right] z(t)\, dt = \sum_{i=1}^{k} a_i \int \vartheta_i(t)\, z(t)\, dt. \tag{8.1.36}$$

*) Это определение не является наиболее общим, которое может быть дано. Можно показать, что для стационарных процессов указанное определение приводит к тому, что спектральная плотность процесса ограничена (см. задачу 8.2).

Для гауссовских случайных процессов с нулевым средним значением случайные величины $z_i = \int \vartheta_i(t)\, z(t)\, dt$ и случайная величина $y = \int [\Sigma a_i \vartheta_i(T)] z(t)\, dt$ имеют нулевое среднее и являются гауссовскими. Следовательно, любая линейная комбинация гауссовских случайных величин z_i также является гауссовской случайной величиной. Множество случайных величин, для которых каждая конечная линейная комбинация гауссовская, называется множеством *совместно гауссовских* случайных величин так, что множество z_i, рассмотренное выше, является множеством совместно гауссовских случайных величин.

Можно легко найти совместную характеристическую функцию и совместную плотность вероятности множества совместно гауссовских случайных величин $z_1, \ldots, z_k$. Совместная характеристическая функция $z_1, \ldots, z_k$ равна по определению

$$M_{z_1, \ldots, z_k}(v_1, \ldots, v_k) = \overline{\exp j \sum_{i=1}^{k} v_i z_i}. \tag{8.1.37}$$

Пусть случайная величина y равна

$$\sum_{i=1}^{k} v_i z_i.$$

Так как y — гауссовская величина с нулевым средним, то ее характеристическая функция равна

$$M_y(u) = \overline{e^{juy}} = \int \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{y^2}{2\sigma_y^2} + juy\right) dy = \exp(-\sigma_y^2 u^2), \tag{8.1.38}$$

где σ_y^2 — дисперсия y,

$$\sigma_y^2 = \sum_{i=1}^{k} \sum_{l=1}^{k} v_i v_l \overline{z_i z_l}.$$

Заметив, что $\overline{e^{jy}}$ составляет правую часть (8.1.37), можно подставить $u = 1$ в (8.1.38) и получить

$$M_{z_1, \ldots, z_k}(v_1, \ldots, v_k) = \exp\left(-\sum_{i=1}^{k} \sum_{l=1}^{k} v_i v_l \overline{z_i z_l}\right). \tag{8.1.39}$$

Так как совместная характеристическая функция $z_1, \ldots, z_k$ является многомерным преобразованием Фурье совместной плотности вероятности случайных величин $z_1, \ldots, z_k$, то совместную плотность вероятности можно найти преобразованием, обратным преобразованию Фурье (8.1.39). Получаем

$$p(z_1, \ldots, z_k) = \frac{\exp\left(-\frac{1}{2|\Lambda|} \sum_{i=1}^{k} \sum_{l=1}^{k} \Lambda_{i,l} z_i z_l\right)}{(2\pi)^{n/2} |\Lambda|^{1/2}}, \tag{8.1.40}$$

где Λ — матрица $k \times k$ с элементами $\overline{z_i z_l}$; $|\Lambda|$ — определитель матрицы Λ, а $\Lambda_{i,l}$ — алгебраическое дополнение элемента i, l матрицы Λ. Если $|\Lambda| = 0$, то некоторая линейная комбинация z_i имеет нулевую

дисперсию и совместная плотность вероятности существует только в смысле δ-функций. Важно заметить, что совместная плотность определяется только через множество коэффициентов корреляции $\overline{z_i z_l}$. Если $\overline{z_i z_l} = 0$ для всех $i \neq l$, то можно увидеть, что $p(z_1, \ldots z_k)$, представляется произведением $p_{z_1}(z_1) \ldots p_{z_k}(z_k)$. Таким образом, получаем важный результат, что, если совместно гауссовские случайные величины с нулевым средним некоррелированы (т. е. $\overline{z_i z_l} = 0$ для $i \neq l$), то они статистически независимы.

Для белого гауссова шума этот результат имеет интересное следствие, состоящее в том, что если $\vartheta_1(t), \ldots, \vartheta_k(t)$ ортогональны, то $z_i = \int \vartheta_i(t)\, z(t)\, dt$, $1 \leqslant i \leqslant k$, образуют множество статистически независимых случайных величин. Для того чтобы увидеть это, следует заметить, что

$$\overline{(z_i + z_l)^2} = \frac{N_0}{2} \int [\vartheta_i(t) + \vartheta_l(t)]^2\, dt.$$

Раскрывая скобки и опуская члены в квадрате, получаем

$$\overline{z_i z_l} = \frac{N_0}{2} \int \vartheta_i(t)\, \vartheta_l(t)\, dt = 0. \tag{8.1.41}$$

Теперь пусть $\{\vartheta_i(t)\}$ — полное множество ортонормальных функций и пусть для гауссовского процесса с нулевым средним $\{z_i\}$ — множество совместно гауссовских случайных величин, задаваемых равенством $z_i = \int \vartheta_i(t)\, z(t)\, dt$. Тогда для произвольной функции из L_2

$$x(t) = \sum_{i=1}^{\infty} x_i\, \vartheta_i(t)$$

имеем*)

$$\int x(t)\, z(t)\, dt = \operatorname*{l.i.m.}_{k \to \infty} \sum_{i=1}^{k} x_i z_i, \tag{8.1.42}$$

где предел понимается в том смысле, что

$$\lim_{k \to \infty} \overline{\left[\int x(t)\, z(t)\, dt - \sum_{i=1}^{k} x_i z_i \right]^2} = 0.$$

Из этого следует, что гауссовский процесс с нулевым средним полностью определяется коэффициентами корреляции $\overline{z_i z_l}$ приведенного выше разложения. Для белого гауссова шума $\overline{z_i z_l} = (N_0/2)\delta_{il}$, где δ_{il} равно 1 для $i = l$ и равно 0 для $i \neq l$.

Теперь предположим, что $z(t)$ — гауссовский процесс с нулевым средним и что $z(t)$ также определяется в виде совокупности слу-

*) Для доказательства равенства (8.1.42) и существования предела в общем случае см. задачу 8.1.

чайных величин параметра t. Тогда для полного множества ортонормальных функций $\{\vartheta_i(t)\}$ с $z_i = \int \vartheta_i(t) z(t) dt$ коэффициент корреляции $\overline{z_i z_l}$ задается равенством

$$\overline{z_i z_l} = \int\int \vartheta_i(t)\, \vartheta_l(\tau)\, \overline{z(t) z(\tau)}\, dt\, d\tau. \tag{8.1.43}$$

Отсюда видно, что рассматриваемый процесс полностью определяется автокорреляционной функцией $\mathcal{R}(t, \tau) = \overline{z(t)z(\tau)}$.

Далее покажем, что если $\mathcal{R}(t, \tau)$ — непрерывная функция, то для каждого t, $z(t)$ — гауссовская случайная величина с нулевым средним. Для любого данного t определим $u_n(\tau)$ по формуле

$$u_n(\tau) = \begin{cases} n; & t - \frac{1}{2n} \leqslant \tau \leqslant t + \frac{1}{2n}, \\ 0; & \text{в других точках,} \end{cases} \tag{8.1.44}$$

и положим $y_n = \int u_n(\tau) z(\tau) d\tau$. Для каждого n имеем: y_n — гауссовская случайная величина с нулевым средним, $y_n - z(t) = \int u_n(\tau) [z(\tau) - z(t)]\, d\tau$ и, следовательно,

$$\overline{[y_n - z(t)]^2} = \int\int u_n(\tau_1)\, u_n(\tau_2)\, \overline{[z(\tau_1) - z(t)]\, [z(\tau_2) - z(t)]}\, d\tau_1\, d\tau_2. \tag{8.1.45}$$

Так как $\mathcal{R}(t, \tau)$ непрерывна, то $\overline{[z(\tau_1) - z(t)]\, [z(\tau_2) - z(t)]}$ стремится к нулю при τ_1 и τ_2, стремящимся к t и, следовательно,

$$\lim_{n \to \infty} \overline{[y_n - z(t)]^2} = 0. \tag{8.1.46}$$

Отсюда видно, что $z(t)$ — гауссовская случайная величина с нулевым средним. Слегка обобщая это доказательство и рассматривая линейные комбинации $z(t_1)$, ..., $z(t_k)$, можно увидеть, что $z(t_1)$, ..., $z(t_k)$ является множеством совместно гауссовских случайных величин. Обратно, можно показать [Давенпорт и Рут (1958)], что если случайный процесс имеет непрерывную автокорреляционную функцию и если для любого множества моментов времени t_1, ..., t_k средние значения случайных величин $z(t_1)$, ..., $z(t_k)$ равны нулю и эти величины являются совместно гауссовскими, то $z(t)$ — гауссовский случайный процесс с нулевым средним (в смысле первоначального определения).

Теперь рассмотрим прохождение гауссовского случайного процесса с нулевым средним через линейный инвариантный во времени фильтр с импульсным откликом $h(t)$ из L_2. Получающийся случайный процесс $y(t) = \int h(t - \tau) z(\tau) d\tau$ также является гауссовским случайным процессом с нулевым средним, так как для любого множества моментов времени t_1, ..., t_k случайные величины $y(t_1)$, ..., $y(t_k)$ имеют нулевые средние и являются совместно гауссовскими (для доказательства того, что $y(t)$ имеет непрерывную автокорреляционную функцию, см. задачу 8.5).

В частном случае, когда $z(t)$ — белый гауссов шум, процесс $y(t) = \int h(t-\tau)z(\tau)\,d\tau$ весьма просто характеризуется. Используя те же самые рассуждения, как и при выводе (8.1.41), имеем

$$\overline{y(\tau_1)\,y(\tau_2)} = \frac{N_0}{2}\int h(t-\tau_1)\,h(t-\tau_2)\,dt.$$

Следовательно, $y(t)$ является стационарным процессом с автокорреляционной функцией

$$\mathcal{R}_y(\tau) = \frac{N_0}{2}\int h(t)\,h(t+\tau)\,dt. \qquad (8.1.47)$$

Полагая $\tau = 0$, обозначая через $H(f)$ преобразование Фурье $h(t)$, и вспоминая, что $\int |H(f)|^2 df = \int h^2(t)dt$ [см. (8.1.23)], имеем

$$\overline{y^2(t)} = \frac{N_0}{2}\int |H(f)|^2\,df. \qquad (8.1.48)$$

Сравнивая это равенство с (8.1.34), можно заметить, что $N_0/2$ имеет смысл спектральной плотности мощности белого гауссова шума. Из этого видно, что случайный процесс, называемый белым гауссовым шумом, является разумной моделью для шума, который имеет почти постоянную спектральную плотность в интересующей нас области частот. Предположение, что спектральная плотность постоянна на всех частотах, сильно упрощает вычисление для этой модели. Формально автокорреляционная функция белого гауссова шума является обратным преобразованием Фурье от $N_0/2$ и равна δ — импульсу со значением $N_0/2$ в $\tau = 0$. Это означает, что если возникает желание истолковать $z(t)$ как совокупность случайных величин параметра t, то приходится признать, что дисперсия $z(t)$ при каждом t должна быть бесконечной. К тому же заключению можно прийти, основываясь на соотношениях (8.1.44) и (8.1.45) для $z(t)$. Это явление способствует пониманию, почему мы сосредоточиваем внимание на линейных операциях $\int x(t)\,z(t)\,dt$ случайного процесса, а не на случайных величинах, зависящих от времени. На случайные величины $z(t)$ параметра t часто сильно влияет спектральная плотность мощности той области частот, которая не представляет физического интереса, и поэтому часто имеется весьма малая связь между реализациями шума и выборочными функциями (во времени) случайного процесса, являющегося моделью шума.

Взаимная информация для каналов с непрерывным временем

Пусть $x(t)$ из L_2 является сигналом на входе, а $y(t)$ — сигналом на выходе канала с непрерывным временем. Пусть $\varphi_1(t)$, $\varphi_2(t)$, ... — полное множество действительных ортонормальных функций, определенных на интервале $(0, T)$. Тогда $x(t)$ можно представить на интервале $(0, T)$ в виде

$$x(t) = \sum x_n\,\varphi_n(t),\ 0 \leqslant t \leqslant T, \quad x_n = \int x(t)\,\varphi_n(t)\,dt. \qquad (8.1.49)$$

Аналогично, если $\theta_1(t)$, $\theta_2(t)$, ... — другое полное множество ортонормальных функций на интервале $(0, T)$, то $y(t)$ можно представить с помощью случайных величин

$$y_n = \int y(t)\,\theta_n(t)\,dt. \tag{8.1.50}$$

Множество $\{\theta_n(t)\}$ может совпадать с множеством $\{\varphi_n(t)\}$ и такой выбор часто оказывается удобным.

Положим $\mathbf{x}^N$ и $\mathbf{y}^N$ — последовательности $\mathbf{x}^N = (x_1, \ldots, x_N)$, $\mathbf{y}^N = (y_1, \ldots, y_N)$. Канал может быть описан статистически через совместные условные плотности вероятности $p_N(\mathbf{y}^N \mid \mathbf{x}^N)$, заданные для всех N. Для того чтобы избежать влияние $x(t)$ при $t < 0$, примем, что $x(t) = 0$ для $t < 0$. Для простоты примем также, что для всех N входной ансамбль может быть описан совместной плотностью вероятности $q_N(\mathbf{x}^N)$. Взаимная информация*) между $x(t)$ и $y(t)$ для $0 \leqslant t \leqslant T$, если она существует, определяется равенством**)

$$I_T[x(t); y(t)] = \lim_{N\to\infty} I(\mathbf{x}^N; \mathbf{y}^N), \tag{8.1.51}$$

где ***)

$$I(\mathbf{x}^N; \mathbf{y}^N) = \log \frac{p_N(\mathbf{y}^N \mid \mathbf{x}^N)}{\int q_N(\mathbf{x}_1^N)\, p_N(\mathbf{y}^N \mid \mathbf{x}_1^N)\, d\mathbf{x}_1^N}. \tag{8.1.52}$$

Средняя взаимная информация между входом и выходом на интервале $(0,T)$, если она существует, определяется равенствами

$$I_T[X(t); Y(t)] = \lim_{N\to\infty} I(\mathbf{X}^N; \mathbf{Y}^N), \tag{8.1.53}$$

$$I(\mathbf{X}^N; \mathbf{Y}^N) = \int q_N(\mathbf{x}^N)\, p_N(\mathbf{y}^N \mid \mathbf{x}^N)\, I(\mathbf{x}^N; \mathbf{y}^N)\, d\mathbf{x}^N\, d\mathbf{y}^N. \tag{8.1.54}$$

Заметим, что $I(\mathbf{X}^N; \mathbf{Y}^N)$ является неявной функцией T и $q_N(\mathbf{x}^N)$. *Пропускная способность канала на единицу времени определяется равенством*

$$C = \lim_{T\to\infty} C_T, \tag{8.1.55}$$

---

*) В русской терминологии — информационная плотность. (*Прим. ред.*)

**) Не следует удивляться, что это определение равносильно общему определению взаимной информации, введенному в гл. 2. Однако имеются некоторые математические тонкости, возникающие при доказательстве этого, и заинтересованному читателю следует обратиться к статье Гельфанда и Яглома (1957). То, что это определение равносильно общему определению гл. 2, означает также, что оно не зависит от используемого множества ортонормальных функций.

***) Как указано в гл. 2, $I(\mathbf{x}^N; \mathbf{y}^N)$ часто существует, даже если не существуют плотности вероятностей.

где

$$C_T = \frac{1}{T}[\sup I(\mathbf{X}^N; \mathbf{Y}^N)], \tag{8.1.56}$$

а верхняя грань берется по всем входным распределениям вероятностей, согласующимся с ограничениями на входе канала. Величина, стоящая в скобках в приведенном выше выражении, представляет собой максимум взаимной информации, которая может быть передана за время T. Для произвольного непрерывного канала при $T \to \infty$ указанный выше предел не обязательно существует, и пропускная способность определена лишь в том случае, когда этот предел существует. Если C существует, то обращение теоремы кодирования, очевидно, остается в силе, однако прямая теорема кодирования не обязательно имеет место, т. е. можно построить примеры каналов с пропускной способностью, определяемой (8.1.55), но таких, что при скоростях, меньших пропускной способности, данные не могут быть переданы с произвольно малой вероятностью ошибки.

8.2. БЕЛЫЙ ГАУССОВ ШУМ И ОРТОГОНАЛЬНЫЕ СИГНАЛЫ

Здесь ортонормальное разложение, рассмотренное в последнем параграфе, применяется к нахождению пропускной способности канала с аддитивным гауссовым шумом с мощностным ограничением на входе. Затем исследуется вероятность ошибки, достигаемая при ортогональных сигналах на входе такого канала.

Пусть $\varphi_1(t)$, $\varphi_2(t)$, ... — полное множество действительных ортонормальных на интервале $(0, T)$ функций. Вход $x(t)$ для $0 \leqslant t \leqslant T$ может быть представлен в виде (8.1.49). Аналогично шум $z(t)$ может быть представлен в виде

$$z_n = \int z(t)\,\varphi_n(t)\,dt. \tag{8.2.1}$$

Для белого гауссова шума со спектральной плотностью $N_0/2$ компоненты шума $z_1, z_2, \ldots$, по определению, статистически независимые гауссовские случайные величины со средним 0 и дисперсией $N_0/2$. Предполагается также, что они статистически не зависят от $x(t)$. Принятая функция $y(t)$ равна сумме $x(t)$ и $z(t)$ и допускает представление

$$y_n = \int y(t)\,\varphi_n(t)\,dt = x_n + z_n. \tag{8.2.2}$$

Равенство (8.2.2), как показано на рис. 8.2.1, сводит канал непрерывного времени к бесконечному множеству параллельных дискретных по времени каналов с аддитивным гауссовым шумом.

Предположим, что мощность на входе канала ограничена величиной S, так что

$$\int_0^T x^2(t)\,dt \leqslant ST.$$

Из соотношения между энергиями находим, что это неравенство равносильно неравенству

$$\sum_n \overline{x_n^2} \leqslant ST. \tag{8.2.3}$$

Из теоремы 7.4.2 следует, что средняя взаимная информация в n-м канале ограничена сверху следующим образом:

$$I(X_n; Y_n) \leqslant \frac{1}{2} \log\left(1 + \frac{2\overline{x_n^2}}{N_0}\right)$$

с равенством, если x_n — гауссовская случайная величина с нулевыми средними значениями. Используя неравенство $\log(1 + z) \leqslant z \log e$, получаем

$$I(X_n; Y_n) \leqslant \frac{\overline{x_n^2}}{N_0} \log e. \tag{8.2.4}$$

Так же как и при доказательстве (7.2.19), можно показать, что средняя взаимная информация в первых N каналах ограничена сверху следующим образом:

$$I(\mathbf{X}^N; \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n; Y_n) \leqslant \sum_{n=1}^{N} \frac{1}{2} \log\left(1 + \frac{2\overline{x_n^2}}{N_0}\right) \tag{8.2.5}$$

с равенством, если $x_1, \ldots, x_N$ — статистически независимые гауссовские случайные величины с нулевыми средними. Из (8.2.3) и (8.2.4) видно, что можно перейти к пределу при $N \to \infty$ и получить

$$I_T[X(t); Y(t)] \leqslant \sum_{n=1}^{\infty} \frac{1}{2} \log\left(1 + \frac{2\overline{x_n^2}}{N_0}\right) \leqslant \tag{8.2.6}$$

$$\leqslant \frac{ST}{N_0} \log e, \tag{8.2.7}$$

где при выводе (8.2.7) использовались неравенства (8.2.3) и (8.2.4). Равенство в (8.2.7) имеет место, если x_n — статистически независимые гауссовские случайные величины с нулевыми средними.

Предположим теперь, что мы ограничиваемся функциями $x(t)$, являющимися линейными комбинациями только первых N ортонормальных функций. Из теоремы 7.5.1 следует, что при ограничениях

$$\sum_{n=1}^{N} \overline{x_n^2} \leqslant ST,$$

значение $I(\mathbf{X}^N; \mathbf{Y}^N)$ достигает максимума на независимых гауссовских случайных величинах $x_1, \ldots, x_N$ с нулевыми средними и с дисперсиями ST/N и

$$\max I(\mathbf{X}^N; \mathbf{Y}^N) = \frac{N}{2} \log\left(1 + \frac{2ST}{N_0 N}\right).$$

При этом ограничении $\overline{x_n^2} = 0$ для $n > N$ и, следовательно, $I(\mathbf{X}^N; \mathbf{Y}^N) = I_T[X(t); Y(t)]$. Пропускная способность в секунду при этом ограничении задается формулой

$$C = \frac{N}{2T} \log\left(1 + \frac{2ST}{N_0 N}\right). \tag{8.2.8}$$

Сформулируем этот результат как теорему, используя обозначение $2WT$ вместо N.

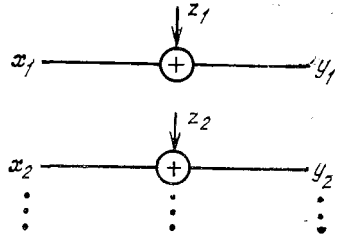


Рис. 8.2.1. Параллельные дискретные по времени каналы, соответствующие непрерывному по времени каналу.

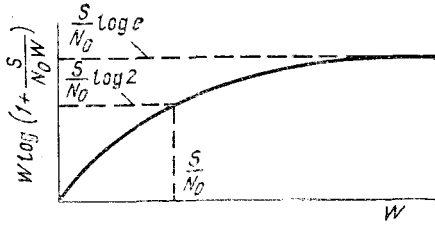


Рис. 8.2.2. Пропускная способность канала с белым гауссовым шумом и с $2WT$ степенями свободы.

Теорема 8.2.1. Пусть выход непрерывного по времени канала представляется как сумма входа и белого гауссова шума со спектральной плотностью $N_0/2$. Пусть вход ограничен по мощности величиной S и представляется на интервале времени длины T как линейная комбинация $2WT$ ортонормальных функций. Тогда пропускная способность канала на единицу времени задается равенством

$$C = W \log\left(1 + \frac{S}{WN_0}\right). \tag{8.2.9}$$

Равенство (8.2.9) — знаменитая формула Шеннона для пропускной способности канала с белым гауссовым шумом и сигналом на входе, ограниченным по полосе и мощности. Как было показано, для больших WT имеется около $2WT$ степеней свободы у множества входных сигналов, ограниченных во времени интервалом длины T, а по частоте, приближенно, полосой W. В § 8.5 эта связь между числом степеней свободы и шириной полосы частот станет яснее и результат (8.2.9) будет установлен для каналов с ограниченной полосой частот, а не для каналов с ограниченным числом степеней свободы.

На рис. 8.2.2 приведен график C как функции W в соответствии с (8.2.9). Из него видно, что C быстро возрастает с ростом W до тех пор, пока W не становится приближенно равной S/N_0. Затем C возрастает более медленно, приближаясь к пределу $(S/N_0) \log e$ при $W \to \infty$. Это в сочетании с (8.2.7) приводит к следующему следствию из теоремы 8.2.1.

Следствие. Пропускная способность на единицу времени канала с белым гауссовым шумом, с входной мощностью, ограниченной S, и с неограниченным числом степеней свободы задается формулой

$$C_\infty = (S/N_0)\log e. \tag{8.2.10}$$

Для того чтобы достичь пропускной способности при ограничении $2WT$ на число степеней свободы, следует энергию сигнала на одну степень задать равенством $\overline{x_n^2} = S/2W$. Следовательно, точка $W = S/N_0$ рис. 8.2.2 соответствует отношению энергий сигнала и шума, равному единице на одну степень свободы. При $W > S/N_0$ энергия сигнала на степень свободы меньше, чем энергия шума на степень свободы и при $W \to \infty$ энергия сигнала на степень свободы стремится к нулю. Этот результат сначала кажется странным и противоречащим интуиции, так как, когда W возрастает, мощность сигнала распространяется все более тонким слоем на все большее число степеней свободы и, следовательно, кажется утопающей в шуме. Однако, как будет показано далее, различимость любого заданного кодового слова в шуме никак не связана с числом ортонормальных функций, используемых для задания кодового слова. Только наличие большого числа степеней свободы позволяет произвести хорошее разделение различных кодовых слов.

Вероятность ошибки для двух кодовых слов

Для непрерывных по времени каналов кодовые слова в коде будут функциями времени (или в более общем случае векторными функциями времени). При заданном множестве ортонормальных функций $\varphi_1(t), \varphi_2(t), \ldots$ эти кодовые слова могут быть представлены как векторы. Таким образом, m-е кодовое слово $x_m(t)$ может быть представлено в виде

$$\begin{aligned} x_m(t) &= \sum_n x_{m,n}\,\varphi_n(t), \\ x_{m,n} &= \int x_m(t)\,\varphi_n(t)\,dt. \end{aligned} \tag{8.2.11}$$

Если кодовые слова имеют не более чем N степеней свободы, то эти кодовые слова можно рассматривать как блоки длины N, и поэтому здесь могут быть непосредственно применены результаты гл. 7. Однако будет более поучительно начать здесь сначала и вывести вновь те результаты, которые относятся к частному случаю, когда допустимое число степеней свободы неограниченно. Начнем со случая двух кодовых слов $x_1(t)$ и $x_2(t)$ и предположим, что $x_1(t)$ и $x_2(t)$ линейные комбинации первых N функций из множества ортонормальных функций

$$\begin{aligned} x_1(t) &= \sum_{n=1}^{N} x_{1,n}\,\varphi_n(t), \\ x_2(t) &= \sum_{n=1}^{N} x_{2,n}\,\varphi_n(t). \end{aligned} \tag{8.2.12}$$

Белый гауссов шум со спектральной плотностью $N_0/2$ складывается с передаваемым сигналом и принятая функция $y(t)$ имеет компоненты

$$y_n = \begin{cases} x_{1,n} + z_n, & n \leqslant N, \quad \text{если послано } x_1(t), \\ x_{2,n} + z_n, & n \leqslant N, \quad \text{если послано } x_2(t), \\ z_n, & n > N, \end{cases} \tag{8.2.13}$$

где z_n — независимые гауссовские случайные величины со средними, равными нулю и дисперсиями $N_0/2$. Пусть масштаб измерения амплитуды для $x(t)$ и $y(t)$ выбран так, что $N_0/2 = 1$. Тогда, если положить

$$\mathbf{x}_1 = (x_{1,1}, \ldots, x_{1,N}), \quad \mathbf{x}_2 = (x_{2,1}, \ldots, x_{2,N}) \text{ и } \mathbf{y} = (y_1, \ldots, y_N),$$

то совместную условную плотность вероятности $\mathbf{y}$ при условии, что задано $\mathbf{x}_1$ или $\mathbf{x}_2$, можно записать в виде

$$\begin{aligned} p_N(\mathbf{y} \mid \mathbf{x}_1) &= \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} \sum_{n=1}^{N} (y_n - x_{1,n})^2 \right], \\ p_N(\mathbf{y} \mid \mathbf{x}_2) &= \frac{1}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2} \sum_{n=1}^{N} (y_n - x_{2,n})^2 \right]. \end{aligned} \tag{8.2.14}$$

Определим логарифм отношения правдоподобия $r_{1,2}(\mathbf{y})$:

$$r_{1,2}(\mathbf{y}) = \ln \frac{p_N(\mathbf{y} \mid \mathbf{x}_1)}{p_N(\mathbf{y} \mid \mathbf{x}_2)} = \tag{8.2.15}$$

$$= -\frac{1}{2} \sum_{n=1}^{N} (y_n - x_{1,n})^2 + \frac{1}{2} \sum_{n=1}^{N} (y_n - x_{2,n})^2 = \tag{8.2.16}$$

$$= \sum_{n=1}^{N} y_n x_{1,n} - \sum_{n=1}^{N} y_n x_{2,n} - \frac{1}{2} \sum_{n=1}^{N} x_{1,n}^2 + \frac{1}{2} \sum_{n=1}^{N} x_{2,n}^2. \tag{8.2.17}$$

Здесь логарифм отношения правдоподобия играет во многом ту же самую роль, как в гл. 5. При декодировании по максимуму правдоподобия сообщение 1 декодируется, когда $r_{1,2}(\mathbf{y}) > 0$, а сообщение 2 — в противном случае. При декодировании по минимуму вероятности ошибки с априорными вероятностями q_1 и q_2 сообщение 1 декодируется, когда $r_{1,2}(\mathbf{y}) > \ln(q_2/q_1)$. Заметим, что при $n > N$ величины y_n опускаются из рассмотрения, так как эти величины не зависят от посланного сообщения и не влияют на значение логарифма отношения правдоподобия. Даже если N бесконечно в (8.2.12), $r_{1,2}(\mathbf{y})$ вполне определено, хотя предел условных плотностей вероятностей в (8.2.14) не существует.

Логарифм отношения правдоподобия может быть вычислен довольно легко по принятой функции, если заметить, что (8.2.17) может быть переписано на основе равенства Парсеваля следующим образом:

$$r_{1,2}(\mathbf{y}) = \int y(t)\, x_1(t)\, dt - \int y(t)\, x_2(t)\, dt - \frac{1}{2} \int x_1^2(t)\, dt + \frac{1}{2} \int x_2^2(t)\, dt. \tag{8.2.18}$$

Следовательно, единственные операции, которые следует произвести над $y(t)$, состоят в умножении $y(t)$ на $x_1(t)$ и $x_2(t)$ и интегрировании; такая процедура называется корреляционным декодированием. Другой способ приема, эквивалентный указанному, состоит в построении фильтров с импульсными откликами $h_1(t) = x_1(T - t)$ и $h_2(t) = x_2(T - t)$. Эти фильтры называются *согласованными* фильтрами и когда $y(t)$ проходит через них, то выход в момент $t = T$ совпадает с приведенным выше; такое декодирование называется декодированием с согласованными фильтрами.

Из (8.2.17) видно, что $r_{1,2}(\mathbf{y})$ линейно связана с y и равна постоянной, сложенной с проекцией y на $\mathrm{x}_1 - \mathrm{x}_2$. Следовательно, $r_{1,2}(\mathbf{y})$

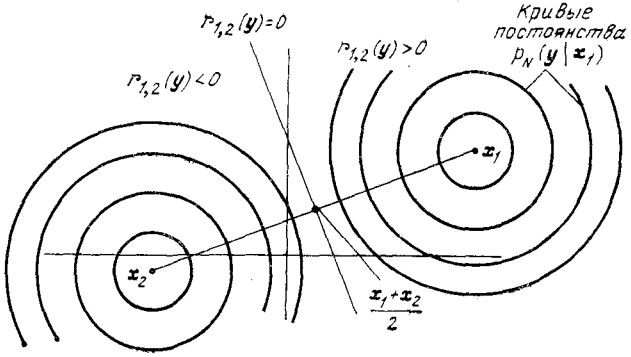


Рис. 8.2.3. Геометрическая интерпретация двух кодовых слов в канале с белым гауссовым шумом.

постоянно на любой гиперплоскости, перпендикулярной к прямой, соединяющей x_1, и x_2, и $r_{1,2}(\mathbf{y}) = 0$ для $\mathbf{y} = 1/2(\mathrm{x}_1 + \mathrm{x}_2)$, что можно легко проверить с помощью (8.2.16) (см. рис. 8.2.3).

Теперь может быть вычислена вероятность ошибки при использовании декодирования по максимуму правдоподобия. Если послано сообщение 1, то, используя выражение $x_{1,n} + z_n$ вместо y_n, получаем

$$r_{1,2}(\mathbf{y}) = \sum_n z_n (x_{1,n} - x_{2,n}) + \frac{1}{2} \sum_n (x_{1,n} - x_{2,n})^2. \qquad (8.2.19)$$

Следовательно, $r_{1,2}(\mathbf{y})$ — гауссовская случайная величина со средним $^1/_2 \sum_n (x_{1,n} - x_{2,n})^2$ и дисперсией $\sum_n (x_{1,n} - x_{2,n})^2$. Вероятность ошибки совпадает с вероятностью того, что $r_{1,2}(\mathbf{y}) < 0$ (или вероятностью того, что значение этой величины более чем на $^1/_2 \sqrt{\sum (x_{1,n} - x_{2,n})^2}$ стандартных отклонений меньше среднего значения)

$$P_{e,1} = \Phi\left[-\frac{1}{2} \sqrt{\sum_n (x_{1,n} - x_{2,n})^2} \right], \qquad (8.2.20)$$

где Φ — функция распределения нормированной гауссовской случайной величины

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp(-v^2/2)\, dv. \tag{8.2.21}$$

Так как в случае, когда послано сообщение 2, вероятность ошибки, очевидно, является той же самой, то общая вероятность ошибки задается равенством

$$P_e = \Phi\left\{-\frac{1}{2}\sqrt{\int [x_1(t) - x_2(t)]^2\, dt}\right\}, \tag{8.2.22}$$

где было использовано соотношение (8.1.16) для энергии. Следует подчеркнуть, что P_e зависит только от энергии разности $x_1(t) - x_2(t)$, а не от особенностей выбираемых функций.

Далее рассмотрим важный случай, когда оба кодовых слова имеют одинаковую энергию

$$\int x_1^2(t)\, dt = \int x_2^2(t)\, dt = \frac{2E}{N_0}. \tag{8.2.23}$$

Здесь используется масштаб для амплитуды, при котором $N_0/2 = 1$, однако E в (8.2.23) можно истолковать как энергию кодового слова при некотором произвольном масштабе для амплитуды и $N_0/2$ как значение спектральной плотности в том же самом масштабе. Полагая, что

$$\lambda = \frac{N_0}{2E} \int x_1(t)\, x_2(t)\, dt$$

— нормированная корреляция между кодовыми словами, перепишем равенство (8.2.22) следующим образом:

$$P_e = \Phi\left[-\sqrt{\frac{(1-\lambda)\, E}{N_0}}\right]. \tag{8.2.24}$$

Вероятность ошибки минимизируется по λ при выборе $x_2(t) = -x_1(t)$, в этом случае $\lambda = -1$. Любое убывание вероятности ошибки ниже этого минимального значения требует увеличения E, которое при фиксированной мощности сигнала требует увеличения времени передачи одного бита. Другая альтернатива, которая будет теперь исследована, состоит в увеличении числа кодовых слов M. Это позволяет увеличить длину кодовых слов (а следовательно, и E) без уменьшения скорости передачи R.

Когда M велико, возникает проблема выбора кодовых слов. Из анализа случая двух кодовых слов ясно, что энергия разности $\int [x_m(t) - x_{m'}(t)]^2 dt$ должна быть большой для всех $m \neq m'$. Можно получить некоторое представление относительно возможных значений энергии этих разностей, оценивая среднюю по m и m' энергию разности при условии, что

$$\frac{1}{M} \sum_{m} \int x_m^2(t)\, dt \leqslant \frac{2E}{N_0}.$$

Тогда средняя энергия разности удовлетворяет соотношениям

$$\frac{1}{M(M-1)}\sum_{m}\sum_{m'\neq m}\int [x_m(t)-x_{m'}(t)]^2\,dt=$$

$$=\frac{1}{M(M-1)}\sum_{m}\sum_{m'}\int [x_m(t)-x_{m'}(t)]^2\,dt= \tag{8.2.25}$$

$$=\frac{1}{M(M-1)}\sum_{m,m'}\int [x_m^2(t)+x_{m'}^2(t)]\,dt-$$

$$-\frac{2}{M(M-1)}\int\sum_{m}x_m(t)\sum_{m'}x_{m'}(t)\,dt\leqslant \tag{8.2.26}$$

$$\leqslant\frac{M}{M-1}\frac{4E}{N_0}. \tag{8.2.27}$$

Соотношение (8.2.27) следует из (8.2.26), если пренебречь вторым членом в (8.2.26), который всегда отрицателен, и использовать ограничение на энергию $x_m(t)$. Когда $M=2$, эта граница равна $8E/N_0$ и равна энергии разности для $x_{m'}(t)=-x_m(t)$. При $M\to\infty$ граница сходится к значению $4E/N_0$, которое, как можно легко увидеть, является энергией разности для ортогональных кодовых слов энергии $2E/N_0$. Поэтому из (8.2.27) следует, что для больших M существует большое число пар кодовых слов, для которых энергия разности немного больше, чем для ортогональных кодовых слов.

В оставшейся части этого параграфа будут найдены верхние и нижние границы вероятности ошибки для множества ортогональных кодовых слов равной энергии. Однако прежде всего свяжем ортогональные коды с другим хорошо известным классом кодов — симплексными кодами. Пусть $x_1(t),\ \ldots,\ x_M(t)$ множество ортогональных функций равной энергии; определим кодовые слова ассоциированного симплексного кода с помощью равенств

$$\xi_m(t)=x_m(t)-\frac{1}{M}\sum_{m'}x_{m'}(t),\quad 1\leqslant m\leqslant M. \tag{8.2.28}$$

Геометрически $\xi_m(t)$ могут быть истолкованы как вершины $(M-1)$-мерного равностороннего симплекса с центром в начале координат. Так как кодовые слова симплексного кода получаются в результате простого смещения связанного с ним ортогонального кода, то видно, что эти коды имеют одинаковую вероятность ошибочного декодирования. Однако энергия симплексных кодовых слов меньше, чем энергия ортогональных слов, в $(M-1)/M$ раз. Интуитивно правдоподобно, что при заданных M и E симплексный код дает минимум возможной вероятности ошибки в канале с белым гауссовым шумом. Однако строгое доказательство этого еще не найдено.

Вероятность ошибки для ортогональных кодовых слов

Пусть $x_1(t),\ \ldots,\ x_M(t)$ — ортогональные сигналы, имеющие энергию $A^2=2E/N_0$; здесь, так же как и для случая двух кодовых слов, выбирается масштаб амплитуды, нормирующий шум. Шум предпо-

лагается аддитивным гауссовым и белым. Определим ортонормальное множество $\varphi_1(t), \ldots, \varphi_M(t)$ следующим образом:

$$\varphi_m(t) = \frac{x_m(t)}{A}. \tag{8.2.29}$$

Тогда $x_m(t)$ представляется в виде $\mathbf{x}_m = (0, \ldots, 0, A, 0, \ldots, 0)$, где A стоит на m-й позиции. Пусть $y(t)$ — принятая функция и пусть $y_m = \int y(t)\ \varphi_m(t)\, dt$. Если передано сообщение m, то $y_m = A + z_m$, а для $m' \neq m$ имеем $y_{m'} = z_{m'}$, где z_m — независимые нормированные гауссовские случайные величины. Полагая $\mathbf{y} = (y_1, \ldots, y_M)$, имеем условные совместные плотности вероятностей

$$p(\mathbf{y} \mid \mathbf{x}_m) = \left(\frac{1}{2\pi}\right)^{M/2} \exp\left[-\frac{(y_m - A)^2}{2} - \sum_{m' \neq m} \frac{y_{m'}^2}{2}\right], \tag{8.2.30}$$

$$p(\mathbf{y} \mid \mathbf{x}_m) = \left(\frac{1}{2\pi}\right)^{M/2} \exp\left[y_m A - \frac{A^2}{2} - \sum_{m'} \frac{y_{m'}^2}{2}\right]. \tag{8.2.31}$$

Будем считать, что используется декодирование по максимуму правдоподобия. Из (8.2.31) видно, что правило декодирования состоит в том, что выбирается m, для которого y_m наибольшее. При этом, если послано сообщение m, то вероятность ошибки равна вероятности того, что $y_{m'} \geqslant y_m$ для некоторого $m' \neq m$, $1 \leqslant m' \leqslant M$. Эту вероятность можно записать в виде

$$P_{e,m} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_m - A)^2}{2}\right] Q(y_m)\, dy_m, \tag{8.2.32}$$

где, для заданного y_m, $Q(y_m)$ — вероятность того, что, для некоторого m', $m' \neq m$, $y_{m'} \geqslant y_m$. Эта вероятность равна единице минус вероятность того, что $y_{m'} < y_m$ для всех m', и так как все $y_{m'}$ — независимые нормированные гауссовские случайные величины, то имеем

$$Q(y_m) = 1 - [\Phi(y_m)]^{M-1} = 1 - [1 - \Phi(-y_m)]^{M-1}. \tag{8.2.33}$$

Равенства (8.2.32) и (8.2.33) дают точное выражение для $P_{e,m}$; оно было табулировано Витерби (1961) для различных значений A и M. Однако здесь мы хотим найти простые и допускающие наглядное толкование границы для $P_{e,m}$.

Для того чтобы получить весьма простую первоначальную границу, заметим, что для заданных m и m' вероятность, что $y_{m'} \geqslant y_m$, когда передано m, равна вероятности ошибки для двух кодовых слов, определенной в (8.2.24) при $\lambda = 0$. Следовательно, используя аддитивную границу для $M - 1$ возможных выборов m', имеем

$$P_{e,m} \leqslant (M-1)\, \Phi(-A/\sqrt{2}). \tag{8.2.34}$$

При другом методе, который оказывается лучшим, когда M велико, можно использовать аддитивную границу для $Q(y_m)$ и получить

$$Q(y_m) \leqslant (M-1)\, \Phi(-y_m). \tag{8.2.35}$$

Для малых y_m правая часть (8.2.35) больше 1, хотя $Q(y_m)$ всегда не больше чем 1. Определим y_0 из равенства

$$M \exp\left(-\frac{y_0^2}{2}\right) = 1. \tag{8.2.36}$$

Для больших M значение y_0 является приближенным значением y_m, для которого $(M-1)\,\Phi(-y_m) = 1$. Поэтому мы используем (8.2.35) для $y_m > y_0$ и $Q(y_m) \leqslant 1$ для $y_m \leqslant y_0$. После этого (8.2.32) принимает вид

$$P_{e,m} \leqslant \int_{-\infty}^{y_0} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_m-A)^2}{2}\right] dy_m +$$

$$+(M-1)\int_{y_0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_m-A)^2}{2}\right] \Phi(-y_m)\, dy_m. \tag{8.2.37}$$

Оценим теперь $\Phi(-y_m)$, используя следующее известное неравенство*) для гауссовской функции распределения при $y_m > 0$,

$$\left(\frac{1}{y_m} - \frac{1}{y_m^3}\right) \frac{\exp(-y_m^2/2)}{\sqrt{2\pi}} < \Phi(-y_m) < \frac{1}{y_m\sqrt{2\pi}} \exp(-y_m^2/2). \tag{8.2.38}$$

Подставляя правую часть (8.2.38) в (8.2.37) и оценивая сверху $1/y_m$ в интеграле величиной $1/y_0$, получаем

$$P_{e,m} \leqslant \Phi(y_0-A) + \frac{M-1}{y_0\sqrt{2\pi}} \int_{y_0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_m-A)^2}{2} - \frac{y_m^2}{2}\right] dy_m.$$

Раскрывая квадрат в подынтегральном выражении и оценивая сверху $M-1$ величиной $\exp y_0^2/2$, будем иметь

$$P_{e,m} \leqslant \Phi(y_0-A) + \frac{\exp[y_0^2/2 - A^2/4]}{\sqrt{4\pi}\, y_0} \Phi\left(\sqrt{2}\left(\frac{A}{2} - y_0\right)\right). \tag{8.2.39}$$

Если $A/2 < y_0 < A$, то можно применить (8.2.38) для обоих слагаемых в (8.2.39). Экспонента в каждом слагаемом будет одной и той же, что дает

$$P_{e,m} \leqslant \frac{\exp[-(A-y_0)^2/2]}{\sqrt{2\pi}} \left[\frac{1}{A-y_0} + \frac{1}{\sqrt{2\pi}\, y_0 (2y_0 - A)}\right]. \tag{8.2.40}$$

Пусть теперь сигналы имеют продолжительность T и мощность $S = E/T$. Тогда

$$A = \sqrt{2TS/N_0} = \sqrt{2TC_\infty}, \tag{8.2.41}$$

где C_∞ — пропускная способность канала, измеренная в натуральных единицах в секунду. Из (8.2.36) также имеем

$$y_0 = \sqrt{2 \ln M} = \sqrt{2RT}, \tag{8.2.42}$$

*) См. Феллер (1950), т. 1, гл. VII, § 1.

где R — скорость, измеренная в натуральных единицах в секунду. Следовательно, (8.2.40) справедливо при $C_\infty/4 < R < C_\infty$ и после подстановки приведенных выше выражений (8.2.40) принимает вид

$$P_{e,m} \leqslant \frac{\exp\left[-T\left(\sqrt{C_\infty}-\sqrt{R}\right)^2\right]}{\sqrt{4\pi T}}\left[\frac{1}{\sqrt{C_\infty}-\sqrt{R}}+\frac{1}{\sqrt{4\pi TR}\left(2\sqrt{R}-\sqrt{C_\infty}\right)}\right]. \tag{8.2.43}$$

Для $R \leqslant C_\infty/4$ используем (8.2.34). Оценивая сверху $M-1$ величиной e^{RT} и применяя (8.2.38), получаем из (8.2.34)

$$P_{e,m} \leqslant \frac{\exp\left[-T\left(\frac{C_\infty}{2}-R\right)\right]}{\sqrt{2\pi TC_\infty}}. \tag{8.2.44}$$

Неравенства (8.2.43) и (8.2.44) показывают, что для любого заданного $R < C_\infty$ вероятность ошибки стремится к нулю по меньшей мере экспоненциально с ростом T. Показатель экспоненты (равный $\left(\sqrt{C_\infty}-\sqrt{R}\right)^2$ для $C_\infty/4 < R < C_\infty$ и равный $C_\infty/2 - R$ для $R \leqslant C_\infty/4$) изображен на рис. 8.2.4. Из рисунка видно, что рассматриваемый показатель экспоненты имеет такой же вид, как кривая зависимости показателя экспоненты от скорости для каналов с очень большим шумом. Это неудивительно, поскольку для больших T число дискретных по времени каналов, требуемых для представления кодовых слов, растет экспоненциально и отношение средней энергии сигнала к шуму на одну степень свободы стремится к 0.

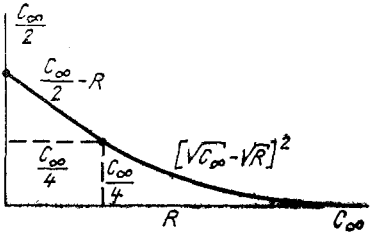


Рис. 8.2.4. Показатель экспоненты в зависимости от скорости для канала с белым гауссовым шумом.

Эти результаты можно сравнить с результатами гл. 7, соответствующими тому же каналу, но с ограничением на число степеней свободы в секунду. Из сравнения видно, что потеря в показателе экспоненты, вызываемая ограничением, мала, если только число степеней свободы достаточно для того, чтобы сделать энергетическое отношение сигнал/шум на степень свободы меньше чем 1.

Оценим теперь снизу $P_{e,m}$. Пусть y — некоторое число, определяемое ниже. Используя (8.2.32) и учитывая, что $Q(y_m)$ убывает с y_m, получаем

$$P_{e,m} \geqslant \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_m-A)^2}{2}\right] Q(y_m)\, dy_m \geqslant \tag{8.2.45}$$

$$\geqslant Q(y)\,\Phi(y-A). \tag{8.2.46}$$

Другими словами, $P_{e,m}$ оценивается снизу с помощью подсчета только тех ошибок, для которых $y_m < y$ и $y_{m'} \geqslant y$ для некоторого m'.

Используя биномиальное разложение для $Q(y)$, данное в (8.2.33), получаем

$$Q(y) = (M-1)\,\Phi(-y) - \binom{M-1}{2}\Phi^2(-y) + \cdots$$

Это знакопеременный ряд и первые два члена дают нижнюю границу для $Q(y)$. Это следует из того, что либо члены убывают по величине, либо оценка отрицательна. Для $y \geqslant y_0$, $Q(y)$ можно оценить дальше следующим образом:

$$\begin{aligned} Q(y) &\geqslant (M-1)\,\Phi(-y)\left[1 - \frac{M-2}{2}\,\Phi(-y)\right] \geqslant \\ &\geqslant (M-1)\,\Phi(-y)\left[1 - \frac{M-2}{2\sqrt{2\pi}y}\,e^{-y^2/2}\right] \geqslant \\ &\geqslant (M-1)\,\Phi(-y)\left[1 - \frac{1}{2\sqrt{2\pi}y}\right]; \quad y \geqslant y_0. \end{aligned} \tag{8.2.47}$$

Здесь было использовано (8.2.38) для оценки снизу $-\Phi(-y)$, а затем то, что $M = \exp(y_0^2/2)$ и $y \geqslant y_0$. Подставляя (8.2.47) в (8.2.46), используя соотношение $M = \exp(y_0^2/2)$ еще раз и оценивая Φ снизу с помощью (8.2.38), получаем

$$\begin{aligned} P_{e,m} \geqslant &\left(1 - \frac{1}{M}\right)\left(1 - \frac{1}{2\sqrt{2\pi}y}\right)\left(\frac{1}{y} - \frac{1}{y^3}\right) \times \\ \times &\left[\frac{1}{A-y} - \frac{1}{(A-y)^3}\right]\frac{1}{2\pi}\exp\left[\frac{y_0^2}{2} - \frac{y^2}{2} - \frac{(A-y)^2}{2}\right]. \end{aligned} \tag{8.2.48}$$

Граница (8.2.48) справедлива для любых y, находящихся между y_0 и A. Экспоненциальный множитель максимизируется для $A/2 < y_0 < A$ при $y = y_0$ и для $y_0 \leqslant A/2$ при $y = A/2$. Отсюда, используя $A = \sqrt{2TC_\infty}$ и $y_0 = \sqrt{2TR}$, (8.2.48) можно привести к виду

$$\begin{aligned} P_{e,m} \geqslant &\left(1 - \frac{1}{M}\right)\left(1 - \frac{1}{4\sqrt{\pi RT}}\right)\left(1 - \frac{1}{2RT}\right)\left[1 - \frac{1}{2T\left(\sqrt{C_\infty} - \sqrt{R}\right)^2}\right] \times \\ &\times \frac{\exp\left[-T\left(\sqrt{C_\infty} - \sqrt{R}\right)^2\right]}{4\pi T\left[\sqrt{RC_\infty} - R\right]} \quad \text{для} \quad \frac{C_\infty}{4} < R < C_\infty, \end{aligned} \tag{8.2.49}$$

$$P_{e\,m} \geqslant \left(1 - \frac{1}{M}\right)\left(1 - \frac{1}{2\sqrt{\pi TC_\infty}}\right)\left(1 - \frac{2}{TC_\infty}\right)^2 \frac{\exp\left[-T\left(\frac{C_\infty}{2} - R\right)\right]}{\pi C_\infty T}$$

$$\text{для} \quad R \leqslant \frac{C_\infty}{4}. \tag{8.2.50}$$

Отсюда видно, что экспоненты в этих нижних границах совпадают с экспонентами верхних границ для всех $R < C_\infty$.

Так как симплексный код имеет ту же вероятность ошибки, что и ортогональный код с энергией, большей в $M/(M-1)$ раз, то можно применить границы (8.2.43), (8.2.44), (8.2.49) и (8.2.50) для симплексного кода, заменив лишь C_∞ во всех равенствах на $C_\infty[M/(M-1)]^{1/2}$.

8.3. ЭВРИСТИЧЕСКОЕ ИЗУЧЕНИЕ ПРОПУСКНОЙ СПОСОБНОСТИ КАНАЛА С АДДИТИВНЫМ ГАУССОВЫМ ШУМОМ И ОГРАНИЧЕНИЯМИ НА ПОЛОСУ ЧАСТОТ

В предыдущих параграфах было показано, как представить сигнал и шум с помощью ортонормальных разложений, и это было использовано для нахождения пропускной способности канала с аддитивным гауссовым шумом в случае, когда имелись ограничения на мощность. Затем рассматривалась вероятность ошибочного декодирования, когда в качестве множества кодовых слов бралось множество ортогональных функций.

В этом анализе имеются два неприятных момента. Во-первых, чтобы сделать вероятность ошибки малой для скорости, близкой к пропускной способности, надо использовать огромное число ортогональных функций и это требует весьма большую полосу частот. Во-вторых, оправдание использования белого гауссова шума как модели реального шума было основано на том, что он дает приемлемое и простое приближение в интересующей нас области частот. Вместе с тем при использовании для кодирования все большего и большего числа ортогональных функций в конце концов должна превыситься область частот, в которой имеет какой-либо смысл предположение о белом гауссовом шуме. Фактически, если принять точку зрения, что полная мощность принятого шума конечна, то спектральная мощность шума должна стремиться к нулю при возрастании частоты и среднюю взаимную информацию в канале можно сделать сколь угодно большой, помещая входной сигнал на произвольно больших частотах.

Физически, приведенная выше аргументация не совсем верна, так как некоторые из аддитивных шумов возникают в приемнике и, увеличивая частоту входных сигналов, надо модифицировать приемник так, чтобы он принимал эти высокие частоты; это, в свою очередь, порождает аддитивный шум на этих частотах.

Однако физические доводы, подобные этому, не дают полностью удовлетворительного выхода из этого затруднения. Действительные трудности состоят в том, что модель гауссова белого шума и модель сигнала, не ограниченного по частоте, являются весьма неустойчивыми. Получаемые результаты очень сильно зависят от того, что происходит на бесконечно больших частотах.

Один из распространенных способов избежать эти трудности состоит в допущении, что сигнал не содержит частот, больших, чем некоторая максимальная частота W. В этом случае для представления входа может быть использована теорема отсчетов и, так как имеются $2W$ отсчетов в секунду, то из (8.2.9) можно заключить, что пропускная способность равна $W\log[1 + S/(N_0W)]$. Однако при этом подходе возникают некоторые чисто математические трудности, состоящие, в частности, в том, что определение пропускной способности, данное в гл. 4, неприменимо здесь, так как ограниченный по полосе частот сигнал одновременно не может быть строго ограничен по времени. Ниже мы вернемся к разрешению этих трудностей и сделаем точным приведенный выше результат.

Другой распространенный метод обхода неприятностей, связанных с произвольно большими частотами, сводится к тому, что принимается, что спектральная плотность шума возрастает с частотой при $f \to \infty$. Этот подход неудовлетворителен как с физической, так и с математической точки зрения.

Предлагаемый здесь подход состоит в предположении, что сигнал, во-первых, ограничивается по мощности и, во-вторых, ограничивается по частоте с помощью пропускания его перед передачей через линейный инвариантный во времени фильтр. Следовательно, тогда, когда используются сигналы с очень высокими частотами, фильтр ослабляет эти высокие частоты до величины, много меньшей спектральной плотности шума.

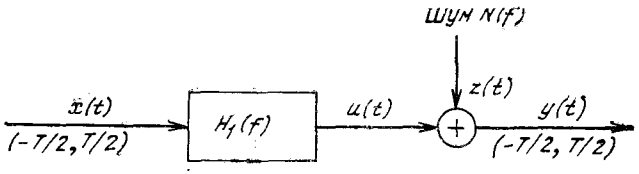


Рис. 8.3.1. Сумма профильтрованного сигнала и шума.

Математически этот подход имеет то преимущество, что он абсолютно ясно определен и допускает точный анализ. Физически его преимущество заключается в том, что он намного ближе, чем другие подходы, отражает виды ограничений по частоте, возникающие в реальных системах связи. Последнее преимущество этого подхода состоит в том, что, варьируя отклик фильтра и спектральную плотность шума, можно понять довольно много об устойчивости пропускной способности канала и экспоненты вероятности ошибки к относительно малым изменениям модели. Действительно, приняв этот подход, мы сможем дать точный вывод пропускной способности при строго ограниченном по частоте входе и увидеть, насколько устойчив этот результат.

В этом параграфе приводится эвристический вывод пропускной способности канала с отфильтрованным входом и аддитивным небелым гауссовым шумом. Вывод крайне прост и крайне убедителен. Вместе с тем он не строг и содержит ряд пробелов, которые не могут быть удовлетворительно заполнены. Следующие два параграфа будут посвящены строгому выводу того же самого результата другим методом.

Рассматриваемая ситуация изображена на рис. 8.3.1. Вход канала $x(t)$ равен нулю вне интервала $(-T/2, T/2)$ и ограничен по мощности величиной S в том смысле, что $x(t)$ выбирается из ансамбля, для которого

$$\int_{-T/2}^{T/2} x^2(t)\,dt \leqslant ST. \tag{8.3.1}$$

Вход пропускается через фильтр с частотной характеристикой $H_1(f)$ и выход фильтра обозначается через $u(t)$. Шум образован выборочной функцией $z(t)$ стационарного гауссова шума со спектральной плотностью $N(f)$, который добавляется к $u(t)$. Выход канала $y(t)$ равен сумме $u(t) + z(t)$, рассматриваемой на интервале $(-T/2, T/2)$. Рассмотрим, сколь большой может быть сделана средняя взаимная информация в секунду между $x(t)$ и $y(t)$.

Представим $x(t)$ рядом Фурье:

$$x(t) = \sum_{i=-\infty}^{\infty} x_i \vartheta_i(t),$$

$$\vartheta_i(t) = \begin{cases} \sqrt{2/T}\cos\dfrac{2\pi i t}{T}; & i > 0 \\ \sqrt{1/T}\qquad\quad ; & i = 0 \\ \sqrt{2/T}\sin\dfrac{2\pi i t}{T}; & i < 0 \end{cases}; \quad |t| \leqslant T/2, \tag{8.3.2}$$

$$\vartheta_i(t) = 0; \quad |t| > T/2.$$

Отклик фильтра на $\vartheta_i(t)$ равен $\int \vartheta_i(\tau)\, h_1(t-\tau)\, d\tau$, где $h_1(t)$ — импульсный отклик фильтра и обратное преобразование Фурье $H_1(f)$. Если T намного больше, чем эффективная продолжительность импульсного отклика, то следует ожидать, что отклик фильтра на $\vartheta_i(\tau)$ должен быть приближенно синусоидой частоты $|i|/T$ и продолжительности $(-T/2, T/2)$. Также следует ожидать, что этот отклик относительно $\vartheta_i(\tau)$ ослабится примерно в $|H_i(i/T)|$ раз. Это побуждает определить множество функций $\theta_i(t)$:

$$\theta_i(t) = \frac{1}{|H_1(i/T)|} \int \vartheta_i(\tau)\, h_1(t-\tau)\, d\tau. \tag{8.3.3}$$

Из приведенного выше рассмотрения следует ожидать, что $\theta_i(t)$ приближенно, с точностью до сдвига фазы, равна $\vartheta_i(t)$

$$\theta_i(t) \approx \alpha\vartheta_i(t) + \beta\vartheta_{-i}(t); \quad \alpha^2 + \beta^2 = 1. \tag{8.3.4}$$

Более того, из (8.3.3) видно, что фазовый сдвиг между $\theta_i(t)$ и $\vartheta_i(t)$ примерно такой же, как и фазовый сдвиг между $\theta_{-i}(t)$ и $\vartheta_{-i}(t)$ и поэтому $\theta_i(t)$ и $\theta_{-i}(t)$ приближенно ортогональны. Из (8.3.4) также вытекает, что $\theta_i(t)$ приближенно ортогональны к $\theta_j(t)$ при $j \neq i$. Следовательно, при аппроксимации, которая улучшается с возрастанием T, $\theta_i(t)$ можно приближенно рассматривать как множество ортонормальных функций.

Далее вычислим $u(t)$ через $\theta_i(t)$:

$$u(t) = \int x(\tau)\, h_1(t-\tau)\, d\tau = \int \sum_i x_i \vartheta_i(\tau)\, h_1(t-\tau)\, d\tau.$$

Применяя (8.3.3), получаем равенства

$$u(t) = \sum u_i \theta_i(t),$$
$$u_i = x_i |H_1(i/T)|. \tag{8.3.5}$$

Шум $z(t)$ также можно разложить по $\theta_i(t)$:

$$z_i = \int z(t)\,\theta_i(t)\,dt. \tag{8.3.6}$$

В течение некоторого времени шум будет рассматриваться как результат прохождения белого гауссова шума с единичной спектральной плотностью через физически нереализуемый фильтр с частотной характеристикой, равной $\sqrt{N(f)}$.

Следовательно, если положить, что $n(\tau)$ — белый гауссов шум, то

$$z_i = \int z(t)\,\theta_i(t)\,dt = \iint n(\tau)\,g(t-\tau)\,\theta_i(t)\,dt d\tau,$$

где

$$g(t) = \int \sqrt{N(f)}\,e^{i2\pi ft}\,df.$$

Полагая

$$\psi_i(\tau) = \int g(t-\tau)\,\theta_i(t)\,dt, \tag{8.3.7}$$

получаем

$$z_i = \int n(\tau)\,\psi_i(\tau)\,d\tau. \tag{8.3.8}$$

Из рассуждений, аналогичных приведенным выше, следует, что все функции $\psi_i(\tau)$ приближенно являются синусоидами; функции множества $\{\psi_i(\tau)\}$ приближенно ортогональны и

$$\int \psi_i^2(\tau)\,d\tau \approx N(i/T). \tag{8.3.9}$$

Отсюда, используя (8.1.41), имеем

$$\overline{z_i z_j} \approx N(i/T)\,\delta_{ij}. \tag{8.3.10}$$

Суммируя $u(t)$ и $z(t)$, находим, что принятый сигнал на интервале $(-T/2, T/2)$ приближенно задается соотношениями

$$y(t) \approx \sum y_i\,\theta_i(t), \tag{8.3.11}$$

$$y_i = x_i\,|H_1(i/T)| + z_i. \tag{8.3.12}$$

Величины y_i можно рассматривать как выходы множества параллельных дискретных по времени каналов с аддитивными гауссовыми шумами. Существенно то, что канал был разбит на узкие полосы частот, каждая из которых имеет ширину $1/T$. Каждая полоса частот имеет две степени свободы, соответствующие синусу и косинусу.

Все приведенные выше утверждения можно было бы сформулировать чуть более тщательно, однако во всем этом подходе имеется существенный недостаток, который, по-видимому, весьма трудно преодолеть. Когда T становится большим, число параллельных каналов на единицу полосы частот возрастает. При этом, хотя шум в любых двух каналах становится статистически независимым при $T \to \infty$, неясно, становится ли любой канал статистически независимым от

множества всех других каналов. Для того чтобы сделать рассуждения более точными, проще всего отказаться от подхода, основанного на рядах Фурье, и использовать другое множество ортонормальных функций; это будет сделано в следующих двух параграфах.

В остающейся части этого параграфа, тем не менее, будем считать, что (8.3.10) выполняется со строгим равенством и что (8.3.12) задает множество параллельных (независимых) каналов. Теперь можно использовать результаты гл. 7 для нахождения пропускной способности этого параллельного соединения каналов.

Используя неравенство Бесселя, ограничение на мощность (8.3.1) можно записать в виде

$$\sum_i \overline{x_i^2} \leqslant ST. \tag{8.3.13}$$

Если рассмотреть $y_i/|H_1(i/T)|$ как выход i-го канала, то этот выход будет равен x_i, сложенной с независимой гауссовской случайной величиной дисперсии $N(i/T)/|H_1(i/T)|^2$. Из теоремы 7.5.1 следует, что пропускная способность этого параллельного соединения (нормированная на единицу времени) равна

$$C_T = \sum_{i \in I_B} \frac{1}{2T} \log \frac{|H_1(i/T)|^2 B}{N(i/T)}, \tag{8.3.14}$$

где I_B — множество i, для которых $N(i/T)/|H(i/T)|^2 \leqslant B$, а B является решением уравнения

$$S = \frac{1}{T} \sum_{i \in I_B} \left[B - \frac{N(i/T)}{|H_1(i/T)|^2} \right]. \tag{8.3.15}$$

Для того чтобы достичь пропускной способности, количество энергии, которое должно быть использовано в каждом канале, следует задать равенством

$$\overline{x_i^2} = \begin{cases} B - \dfrac{N(i/T)}{|H_1(i/T)|^2}; & i \in I_B, \\ 0 \quad ; & i \notin I_B. \end{cases} \tag{8.3.16}$$

Этот результат допускает то же толкование, что и результат теоремы 7.5.1 (см. рис. 7.5.1).

Пусть теперь $T \to \infty$; в пределе (8.3.14) и (8.3.15) становятся интегралами Римана:

$$C = \lim_{T \to \infty} C_T = \int_{f \in F_B} 1/2 \log \left[\frac{|H_1(f)|^2 B}{N(f)} \right] df, \tag{8.3.17}$$

где F_B — область f, для которой $N(f)/|H_1(f)|^2 \leqslant B$ и B является решением уравнения

$$S = \int_{f \in F_B} \left[B - \frac{N(f)}{|H_1(f)|^2} \right] df. \tag{8.3.18}$$

Спектральная плотность мощности входного ансамбля, на котором достигается пропускная способность, задается равенством, которое следует из (8.3.16),

$$S_x(f) = \begin{cases} B - \dfrac{N(f)}{|H_1(f)|^2}; & f \in F_B, \\ 0 \quad ; & f \notin F_B. \end{cases} \tag{8.3.19}$$

В § 8.5 будет доказано, что при некоторых небольших ограничениях на $N(f)$ и $H_1(f)$ переход от равенства (8.3.17) к (8.3.19) действительно верен.

Интерпретация этих равенств почти тождественна интерпретации теоремы 7.5.1 и дана на рис. 8.3.2.

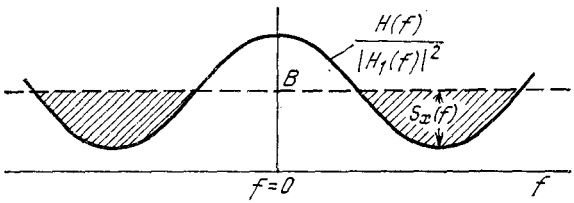


Рис. 8.3.2. Распределение входной мощности для достижения пропускной способности.

Сравнивая рис. 8.3.2 с равенствами (8.3.17) — (8.3.19), можно заметить, что мощность S равна всей площади заштрихованной области на рис. 8.3.2 и что соответствующая спектральная плотность мощности при любом заданном f равна высоте заштрихованной области на этой частоте. Это толкование обычно называется интерпретацией с наполнением водой, так как можно представлять себе, что $N(f)/|H_1(f)|^2$ описывает дно резервуара, а S — количество налитой воды. Предположив, что области соединены, видим, что вода (мощность) распределяется таким же образом, как и при достижении пропускной способности. Пропускная способность пропорциональна интегралу (по заштрихованной части f) от логарифма отношения уровня воды B и дна резервуара $N(f)/|H_1(f)|^2$.

Одной интересной особенностью этого результата является то, что пропускная способность не зависит от значения $N(f)/|H_1(f)|^2$ для частот вне заштрихованной области, т. е. от значений $N(f)/|H_1(f)|^2$, которые больше, чем B. Другими словами, C не зависит от особенностей поведения $N(f)$ и $|H_1(f)|^2$ при $f \to \infty$. Исключение из этого правила возникает, когда $N(f)$ стремится к 0 при возрастании f быстрее, чем $|H_1(f)|^2$. В этом случае пропускная способность бесконечна и любое количество информации может быть получено с помощью передачи на достаточно высоких частотах. Это, конечно, указывает, что математическая модель не отражает основных черт физической ситуации.

Теперь можно применить эти результаты к сигналу с ограниченной полосой частот на фоне белого гауссова шума со спектральной плотностью $N(f) = N_0/2$. Если вход имеет полосу частот W вокруг не-

которой центральной частоты f_c, то ограничения можно представить в виде

$$H_1(f) = \begin{cases} 1; & 0 \leqslant f_c - \frac{W}{2} \leqslant |f| \leqslant f_c + \frac{W}{2}, \\ 0 & \text{на других частотах.} \end{cases} \tag{8.3.20}$$

В этом случае $N(f)/|H_1(f)|^2$ равно $N_0/2$ или ∞ в зависимости от того, находится f внутри полосы частот или нет. Подынтегральные выражения в (8.3.17) и (8.3.18) не зависят от f внутри полосы частот, и F_B должно совпадать со всей областью частот внутри полосы. Таким образом, интегрируя, получаем

$$C = W \log \frac{2B}{N_0}, \tag{8.3.21}$$

$$S = 2W\left[B - \frac{N_0}{2}\right]. \tag{8.3.22}$$

Решая (8.3.22) относительно B и подставляя решение в (8.3.21), имеем

$$C = W \log \left(1 + \frac{S}{N_0 W}\right). \tag{8.3.23}$$

Это является известной теоремой Шеннона о пропускной способности канала, ограниченного по полосе. Эта формула часто употребляется неправильно, главным образом, из-за непонимания того, что она применима только к аддитивному гауссову шуму. Заметим также, что для частоты вне полосы, где $H_1(f) = 1$, несущественно, каково значение спектральной плотности, так как $N(f)/|H_1(f)|^2 = \infty$ для любого ненулевого значения $N(f)$.

Имеется ряд опубликованных в литературе математических парадоксов, касающихся формулы (8.3.23); в них считается, что $N(f) = 0$ вне полосы. В этом случае значение $N(f)/|H_1(f)|^2$ не определено вне полосы, и пропускную способность можно сделать сколь угодно большой, если приписать сколь угодно малые значения отношению $N(f)/|H_1(f)|^2$ вне полосы. Физически, конечно, проблема может быть очень легко разрешена — стоит только заметить, что $N(f)$ не может быть точно равной нулю. Другими словами, когда анализ математической модели физической задачи приходит к неопределенности, что означает, что модель слишком идеализирована и должна быть изменена.

8.4. ПРЕДСТАВЛЕНИЕ ЛИНЕЙНЫХ ФИЛЬТРОВ И НЕБЕЛЫЙ ШУМ

Приступая к изложению точной трактовки сигналов, ограниченных как по мощности, так и по частоте, начнем с анализа ситуации, изображенной на рис. 8.4.1.

На рис. 8.4.1 сигнал $x(t)$ на входе канала имеет произвольную продолжительность T, ограничен по мощности значением S и пропущен через линейный инвариантный во времени фильтр с импульсным откликом $h_1(t)$. Выход фильтра $u(t)$ задается равенством

$$u(t) = \int x(\tau) h_1(t - \tau)\, d\tau.$$

Принимаемый сигнал $y(t)$ равен сумме $u(t)$ и белого гауссова шума. Рассмотрим два случая, когда выход канала является неограниченным по продолжительности сигналом $y(t)$ и когда рассматриваемый выход представляет собой часть сигнала $y(t)$, заданную на конечном временно́м интервале. Как указано выше, фильтр*) $h_1(t)$ можно рассматривать как часть канала или как некоторое ограничение, вводимое для регулирования спектральной плотности $u(t)$.

Первая проблема при анализе ситуации, изображенной на рис. 8.4.1, заключается в нахождении подходящего представления для $x(t)$, $u(t)$ и $y(t)$. Естественно, было бы удобным представить каждую из этих функций ортонормальным разложением и вопрос состоит в том, какое ортонормальное разложение следует выбрать. Было бы особенно хорошо, если бы можно было найти два множества ортонормальных функций, скажем $\{\varphi_i(t)\}$ и $\{\theta_i(t)\}$, таких, что множество $\{\varphi_i(t)\}$ могло быть использовано для представления входа фильтра; множество $\{\theta_i(t)\}$ могло быть использовано для представления выхода фильтра и для каждого i нормированный отклик на $\varphi_i(t)$ был равен $\theta_i(t)$. Цель этого параграфа — показать, что такие множества функций существуют и что они обладают рядом других свойств, которые делают их весьма естественными множествами функций для представления фильтра, его входа и выхода. Имеется одна неприятная особенность у этих функций, которая часто на первых порах раздражает. В лучшем случае, нахождение их не упорядочено, в худшем — фактически невозможно. Здесь нас это не будет касаться, так как в последующем изложении никогда не придется фактически вычислять эти функции. Они будут использованы исключительно как умозрительный инструмент со знанием того, что они могут быть вычислены, но что выигрыш от более глубокого проникновения не оправдал бы усилий по их отысканию.

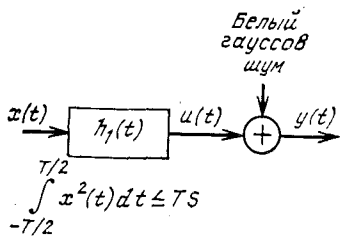


Рис. 8.4.1. Простой канал, ограниченный по мощности и полосе частот.

Что важно для дальнейших обобщений, построение этих множеств функций не зависит от того, инвариантен ли фильтр во времени, а поэтому построение будет проводиться для произвольных линейных, изменяющихся во времени фильтров. Пусть $h(t, \tau)$ — выход в момент времени t на δ-импульс на входе в момент времени τ, т. е. для заданного входа $x(t)$ выход задается равенством

$$u(t) = \int_{-\infty}^{\infty} h(t, \tau)x(\tau)\, d\tau. \qquad (8.4.1)$$

*) Не предполагается, что $h_1(t)$ определяет физически реализуемый фильтр, т. е. что $h_1(t) = 0$ для $t < 0$. Если фильтр играет роль математического условия, то нет причины для такого ограничения.

Примем в дальнейшем изложении, что вход фильтра равен нулю вне интервала $(-T/2,\ T/2)$ и что нас интересует представление выхода только на некотором интервале $(-T_0/2,\ T_0/2)$. Для того чтобы заложить это в нашу модель и избежать постоянного беспокойства относительно пределов интегрирования, положим $h(t,\ \tau)$ равным нулю вне интересующей нас области, т. е.

$$h(t,\ \tau) = 0;\ |\tau| \geqslant T/2 \text{ или } |t| \geqslant T_0/2. \tag{8.4.2}$$

Следовательно, если линейный инвариантный во времени фильтр имеет импульсный отклик $h_1(t)$, то фильтр будет представляться с помощью

$$h(t,\ \tau) = \begin{cases} h_1(t-\tau); & |\tau| \leqslant T/2,\ |t| \leqslant T_0/2, \\ 0 & \text{в других точках.} \end{cases} \tag{8.4.3}$$

Если на функцию $x(\tau)$ наложено ограничение, что она может быть отлична от нуля только на интервале $(-T/2,\ T/2)$, то выход $u(t)$, задаваемый (8.4.1), является выходом линейного инвариантного во времени фильтра $h_1(t)$ на интервале $(-T_0/2,\ T_0/2)$. Вместе с тем функция $u(t)$, определяемая (8.4.1), равна нулю вне этого интервала, в то время как выход действительного фильтра может быть там отличен от нуля.

Продолжительность сигнала на входе T или на выходе T_0, или и того и другого может быть бесконечной, однако всегда будет предполагаться, что

$$\int\int h^2(t,\ \tau)\, dt d\tau < \infty. \tag{8.4.4}$$

Позволяя избежать паталогических ситуаций, условие (8.4.4) в то же время исключает две весьма общие в инженерной практике ситуации. Первая включает случай, когда $h(t,\ \tau)$ содержит δ-импульсы, другая включает случай, когда T и T_0 бесконечны и фильтр инвариантен во времени. Таким образом, в нашем подходе оба этих случая должны рассматриваться как предельные и не всегда имеется гарантия, что эти пределы существуют.

Найдем теперь функции $\varphi_i(\tau)$ и $\theta_i(t)$, которые связаны равенством

$$\theta_i(t) = \alpha_i \int h(t,\ \tau)\, \varphi_i(\tau)\, d\tau. \tag{8.4.5}$$

Требование, чтобы $\theta_i(t)$ были бы ортонормальны, приводит к соотношениям

$$\delta_{ij} = \int \theta_i(t)\, \theta_j(t)\, dt = \alpha_i\, \alpha_j \int\int\int h(t,\ \tau_1)\, h(t,\ \tau_2) \times$$
$$\times\, \varphi_i(\tau_1)\, \varphi_j(\tau_2)\, d\tau_2\, d\tau_1\, dt = \tag{8.4.6}$$
$$= \alpha_i\, \alpha_j \int\int \mathcal{R}(\tau_1,\ \tau_2)\, \varphi_i(\tau_1)\, \varphi_j(\tau_2)\, d\tau_1\, d\tau_2, \tag{8.4.7}$$

где

$$\mathcal{R}(\tau_1,\ \tau_2) = \int h(t,\ \tau_1)\, h(t,\ \tau_2)\, dt. \tag{8.4.8}$$

Задача нахождения множества ортонормальных функций, удовлетворяющих (8.4.7), очень часто встречается в математике и физике. Как

будет видно из дальнейшего, она равносильна нахождению множества чисел $\lambda_1, \lambda_2, \ldots$ и множества функций $\varphi_1(\tau), \varphi_2(\tau) \ldots$, которые удовлетворяют интегральному уравнению

$$\int \mathscr{R}(\tau_1, \tau_2)\, \varphi_i(\tau_2)\, d\tau_2 = \lambda \varphi_i(\tau_1). \tag{8.4.9}$$

В следующей теореме суммируются свойства, которыми обладают λ_i, $\varphi_i(\tau)$ и $\theta_i(t)$. Эти свойства делают более легким обращение с сигналами на входе и выходе фильтра $h(t, \tau)$, однако, как отмечено выше, они не дают указаний, как в действительности найти решение (8.4.9); к счастью, нет нужды в явных решениях для получения большинства последующих результатов.

Теорема 8.4.1. Пусть $h(t, \tau)$ отлична от нуля и интегрируема в квадрате [т. е. удовлетворяет (8.4.4)]. Тогда существуют последовательность (конечная или бесконечная) невозрастающих положительных чисел $\lambda_1 \geqslant \ldots \geqslant \lambda_i \geqslant \ldots > 0$ и взаимно однозначно соответствующие этим числам*) два множества ортонормальных функций $\varphi_i(\tau)$ и $\theta_i(t)$, которые обладают следующими свойствами.

а) $\varphi_i(\tau)$ и λ_i удовлетворяют интегральному уравнению

$$\int \mathscr{R}(\tau_1, \tau_2)\, \varphi_i(\tau_2)\, d\tau_2 = \lambda_i \varphi_i(\tau_1), \tag{8.4.10}$$

где $\mathscr{R}(\tau_1, \tau_2)$ задается (8.4.8).

б) $\varphi_i(\tau)$ и $\theta_i(t)$ связаны соотношениями

$$\sqrt{\lambda_i}\, \theta_i(t) = \int h(t, \tau)\, \varphi_i(\tau)\, d\tau, \tag{8.4.11}$$

$$\sqrt{\lambda_i}\, \varphi_i(\tau) = \int h(t, \tau)\, \theta_i(t)\, dt. \tag{8.4.12}$$

в) $\theta_i(t)$ и λ_i удовлетворяют интегральному уравнению

$$\int \mathscr{R}_0(t_1, t_2)\, \theta_i(t_2)\, dt_2 = \lambda_i \theta_i(t_1), \tag{8.4.13}$$

где

$$\mathscr{R}_0(t_1, t_2) = \int h(t_1, \tau)\, h(t_2, \tau)\, d\tau. \tag{8.4.14}$$

г) Пусть $x(\tau)$ — произвольная функция из L_2 и пусть $(x, \varphi_i) = \int x(\tau)\, \varphi_i(\tau)\, d\tau$. Тогда следующие три утверждения вытекают одно из другого:

$$(x, \varphi_i) = 0 \text{ для всех } i \Longleftrightarrow \int \mathscr{R}(\tau_1, \tau_2)\, x(\tau_2)\, d\tau_2 = 0 \Longleftrightarrow \int h(t, \tau)\, x(\tau)\, d\tau = 0. \tag{8.4.15}$$

К тому же, если $x(\tau)$ разлагается в ряд

$$x(\tau) = \sum_i x_i \varphi_i(\tau) + x_r(\tau); \quad x_i = (x, \varphi_i), \tag{8.4.16}$$

то

$$\int h(t, \tau)\, x(\tau)\, d\tau = \sum x_i \sqrt{\lambda_i}\, \theta_i(t), \tag{8.4.17}$$

*) Заметим, что любые решения $\varphi(\tau)$ (8.4.10) с $\lambda = 0$ не рассматриваются в ортонормальном множестве $\{\varphi_i(\tau)\}$.

$$\int \mathscr{R}(\tau_1,\ \tau_2)\, x(\tau_2)\, d\tau_2 = \sum x_i\, \lambda_i\, \varphi_i(\tau_1). \tag{8.4.18}$$

д) Пусть $u(t)$ — произвольная функция из L_2. Тогда следующие утверждения вытекают одно из другого:

$$(u,\ \theta_i) = 0 \text{ для всех } i \Longleftrightarrow \int \mathscr{R}_0(t_1,\ t_2)\ u(t_2)\, dt_2 = 0 \Longleftrightarrow \int h(t,\ \tau)\, u(t)dt = 0. \tag{8.4.19}$$

Если $u(t)$ разлагается в ряд

$$u(t) = \sum u_i\, \theta_i(t) + u_r(t);\quad u_i = (u,\ \theta_i), \tag{8.4.20}$$

то

$$\int h(t,\ \tau)\, u(t)\, dt = \sum u_i \sqrt{\lambda_i}\, \varphi_i(\tau), \tag{8.4.21}$$

$$\int \mathscr{R}_0(t_1,\ t_2)\, u(t_2)\, dt_2 = \sum u_i\, \lambda_i\, \theta_i(t_1). \tag{8.4.22}$$

е)
$$h(t,\tau) = \sum \sqrt{\lambda_i}\, \varphi_i(\tau)\, \theta_i(t), \tag{8.4.23}$$

$$\iint h^2(t,\ \tau)\, dt\, d\tau = \sum \lambda_i, \tag{8.4.24}$$

$$\mathscr{R}(\tau_1,\ \tau_2) = \sum \lambda_i\, \varphi_i(\tau_1)\, \varphi_i(\tau_2), \tag{8.4.25}$$

$$\mathscr{R}_0(t_1,\ t_2) = \sum \lambda_i\, \theta_i(t_1)\, \theta_i(t_2), \tag{8.4.26}$$

$$\iint \mathscr{R}^2(\tau_1,\ \tau_2)\, d\tau_1\, d\tau_2 = \iint \mathscr{R}_0^2(t_1,\ t_2)\, dt_1\, dt_2 = \sum \lambda_i^2. \tag{8.4.27}$$

ж) λ_i и $\varphi_i(\tau)$ являются решениями следующих задач по отысканию максимума

$$\lambda_i = \max \left\| \int h(t,\ \tau)\ x(\tau)\, d\tau \right\|^2, \tag{8.4.28}$$

$$\lambda_i = \max \left\| \int \mathscr{R}(\tau_1,\ \tau_2)\, x\ (\tau_2)\, d\tau_2 \right\|. \tag{8.4.29}$$

Эти максимизации производятся при ограничениях $\|x\| = 1$ и $(x, \varphi_j) = 0$ для $1 \leqslant i < i$, где $\|x\|$ определяется как $\sqrt{\int x^2(\tau)\, d\tau}$. В каждом случае в качестве $\varphi_i(\tau)$ можно взять функцию $x(\tau)$, которая максимизирует приведенные выше выражения.

Доказательство. Пусть $\lambda_i \neq 0$, $\varphi_i(\tau)$ и $\lambda_j \neq 0$, $\varphi_j(\tau)$ — два нормированных решения интегрального уравнения (8.4.10). Пусть $\theta_i(t)$ и $\theta_j(t)$ задаются равенством $\sqrt{|\lambda_i|}\, \theta_i(t) = \int h(t,\tau)\, \varphi_i(\tau)\, d\tau$.

Тогда так же, как и в (8.4.7), выводим

$$\int \sqrt{|\lambda_i \lambda_j|}\, \theta_i(t)\, \theta_j(t)\, dt = \iint \mathscr{R}(\tau_1,\ \tau_2)\, \varphi_i(\tau_1)\, \varphi_j(\tau_2)\, d\tau_1\, d\tau_2 =$$

$$= \lambda_j \int \varphi_i(\tau_1)\, \varphi_j(\tau_1)\, d\tau_1 = \tag{8.4.30}$$

$$= \lambda_i \int \varphi_i(\tau_2)\, \varphi_j(\tau_2)\, d\tau_2. \tag{8.4.31}$$

При выводе (8.4.30) в интегрировании по τ_2 использовалось равенство (8.4.10). В (8.4.31) был использован тот факт, что $\mathscr{R}(\tau_1, \tau_2) = \mathscr{R}(\tau_2, \tau_1)$

и интегрирование сначала было произведено по τ_1. Из (8.4.30) и (8.4.31) видно, что если $\lambda_i \neq \lambda_j$, то φ_i и φ_j должны быть ортогональны, и отсюда следует, что θ_i и θ_j ортогональны. Если $i = j$, то из (8.4.30) и нормированности $\varphi_i(\tau)$ следует также нормированность $\theta_i(t)$; кроме того, так как в этом случае оба интеграла в (8.4.30) положительны, то $\lambda_i > 0$. Наконец, если $\lambda_i = \lambda_j$, а $\varphi_i(\tau)$ и $\varphi_j(\tau)$ линейно независимы, то любая линейная комбинация $\varphi_i(\tau)$ и $\varphi_j(\tau)$ также удовлетворяет (8.4.10). В последующем изложении, если более чем одна линейно независимая функция $\varphi(\tau)$ удовлетворяет (8.4.10) для одного и того же значения λ, то в множество $\varphi_i(\tau)$ будем включать только ортонормальный базис множества решений (8.4.10) с этим значением λ и повторять это значение λ соответствующее число раз в последовательности λ_i. Было показано, что при таком условии ненулевые λ, удовлетворяющие (8.4.10), положительны, что соответствующие $\varphi_i(\tau)$ — ортонормальны и что $\theta_i(t)$, задаваемые (8.4.11), ортонормальны. В дальнейшем будет показано, что ненулевые λ_i, удовлетворяющие (8.4.10), могут быть упорядочены в убывающую последовательность, а сейчас будем считать, что они расставлены произвольным образом.

Далее убедимся в справедливости (8.4.12), умножая обе части (8.4.11) на $h(t, \tau_1)$ и интегрируя по t. Получим

$$\sqrt{\lambda_i} \int \theta_i(t)\, h(t,\ \tau_1)\, dt = \iint h(t,\ \tau)\, h(t,\ \tau_1)\, \varphi_i(\tau)\, d\tau dt =$$
$$= \int \mathcal{R}(\tau_1,\ \tau_2)\, \varphi_i(\tau_2)\, d\tau_2 = \lambda_i\, \varphi_i(\tau_1).$$

Это соотношение равносильно (8.4.12). Равенство (8.4.13) получается таким же образом; в этом случае умножаем обе части (8.4.12) на $h(t_1, \tau)$, интегрируем по τ и используем (8.4.11).

До этого места были указаны некоторые свойства, которыми должны обладать решения интегрального уравнения (8.4.10), но еще не было показано, что (8.4.10) вообще имеет какое-либо решение.

С одной стороны, доказательство существования решения (8.4.10) весьма громоздко и, с другой стороны, это существование является центральным фактом теории линейных интегральных уравнений и функционального анализа; поэтому мы просто сформулируем результат*). Если ядро $\mathcal{R}(\tau_1, \tau_2)$ отлично от нуля и интегрируемо в квадрате и если $\mathcal{R}(\tau_1, \tau_2) = \mathcal{R}(\tau_2, \tau_1)$, то $\mathcal{R}(\tau_1, \tau_2)$ имеет по крайней мере одно ненулевое собственное значение λ_i и собственную функцию $\varphi_i(\tau)$ [т. е. решение (8.4.10)]; в действительности имеется достаточное число собственных функций и собственных значений, так что если функция $x(\tau)$ ортогональна всем $\varphi_i(\tau)$, то она должна удовлетворять равенству

$$\int \mathcal{R}(\tau_1,\ \tau_2)\ x(\tau_2) d\tau_2 = 0.$$

Для того чтобы использовать здесь этот результат, надо показать, что $\mathcal{R}(\tau_1, \tau_2)$ интегрируема в квадрате. Применяя неравенство Шварца к $\mathcal{R}(\tau_1, \tau_2) = \int h(t, \tau_1)\, h(t, \tau_2) dt$, получаем

*) См., например, Рисс и Надь (1955), Ахиезер и Глазман (1950) или Курант и Гильберт (1951). Однако Курант и Гильберт не рассматривают случая, когда интервал T бесконечен.

$$\mathscr{R}^2(\tau_1,\tau_2) \leqslant \int h^2(t,\tau_1)\,dt \int h^2(t,\tau_2)\,dt,$$

$$\iint \mathscr{R}^2(\tau_1,\tau_2)\,d\tau_1\,d\tau_2 \leqslant \left[\int h^2(t,\tau)\,dt\,d\tau\right]^2 < \infty.$$

Следовательно, сформулированный выше результат может быть применен и это показывает, что первое утверждение (8.4.15) влечет за собой второе. Второе утверждение также влечет за собой третье, так как, умножая второе выражение (8.4.15) на $x(\tau_1)$ и интегрируя, получаем

$$0 = \iint \mathscr{R}(\tau_1,\tau_2)\,x(\tau_2)\,x(\tau_1)\,d\tau_2 d\tau_1.$$

Применяя (8.4.8) к $\mathscr{R}(\tau_1, \tau_2)$, преобразуем это выражение к виду

$$0 = \iiint h(t, \tau_2)\ x(\tau_2)\,h(t, \tau_1)\ x(\tau_1) d\tau_1 d\tau_2 dt,$$

$$0 = \int\left[\int h(t,\ \tau)\ x(\tau)\ d\tau\right]^2 dt.$$

Следовательно,

$$\int h(t,\tau)\ x(\tau) d\tau = 0.$$

Ниже будет показано, что третье утверждение влечет за собой первое, и будет установлена справедливость (8.4.17).

Используя разложение $x(\tau)$ (8.4.16), имеем

$$\int h(t,\ \tau)\,x(\tau)\,d\tau = \int h(t,\tau) \sum_{i=1}^{\infty} x_i\,\varphi_i(\tau)\,d\tau + \\ + \int h(t,\tau)\,x_r(\tau)\,d\tau. \qquad (8.4.32)$$

В силу (8.4.15) последнее слагаемое в (8.4.32) равно нулю, так как $x_r(\tau)$ ортогонально по всем $\varphi_i(\tau)$. Из (8.1.15) следует, что можно изменить порядок суммирования и интегрирования в первом слагаемом правой части (8.4.32) и таким образом получить

$$\int h(t,\tau)\,x(\tau)\,d\tau = \sum_{i=1}^{\infty} \int h(t,\tau)\,x_i\,\varphi_i(\tau)\,d\tau. \qquad (8.4.33)$$

Используя (8.4.11) при интегрировании в правой части, получаем (8.4.17). Равенство (8.4.18) получается точно таким образом, за исключением того, что (8.4.10) используется при интегрировании каждого слагаемого. Покажем теперь, что если $u(t) = \int h(t, \tau)\,x(\tau)d\tau$ и если $u(t)$ равно 0, то $x(\tau)$ ортогонально ко всем φ_i. Из (8.4.17) и неравенства Бесселя выводим

$$0 = \int u^2(t)\,dt \geqslant \sum x_i^2\,\lambda_i.$$

Так как все λ_i положительны, то все x_i должны равняться нулю и третье утверждение (8.4.15) влечет за собой первое.

Часть д) теоремы доказывается таким же образом, как и часть г), и она фактически двойственна утверждению части г). Перейдем к выводу части е) и разложения $h(t, \tau)$. Так как $h(t, \tau)$ — функция двух

переменных, то ее можно разложить в ряд по функциям $\varphi_i(\tau)$ и $\theta_i(t)$. Коэффициенты равны

$$h_{ij} = \iint h(t,\tau)\,\varphi_i(\tau)\,\theta_j(t)\,d\tau dt =$$
$$= \int_t \sqrt{\lambda_i}\,\theta_i(t)\,\theta_j(t)\,dt = \sqrt{\lambda_i}\,\delta_{ij},$$

$$h(t,\tau) = \sum_i \sqrt{\lambda_i}\,\varphi_i(\tau)\,\theta_i(t) + h_r(t,\tau),$$

где $h_r(t, \tau)$ ортогональна ко всем $\varphi_i(\tau)\theta_j(t)$. Покажем теперь, что для любого $x(\tau)$ имеем $\int h_r(t,\tau)\,x(\tau)d\tau = 0$.

$$\int h_r(t,\tau)\,x(\tau)\,d\tau = \int h(t,\tau)\,x(\tau)\,d\tau -$$
$$-\int \sum_i \sqrt{\lambda_i}\,\theta_i(t)\,\varphi_i(\tau)\,x(\tau)\,d\tau.$$

Применяя (8.4.17) к первому слагаемому правой части приведенного выше выражения и используя соотношения (8.1.15) для обоснования изменения порядка суммирования и интегрирования во втором слагаемом, получаем

$$\int h_r(t,\tau)\,x(\tau)\,d\tau = \sum x_i \sqrt{\lambda_i}\,\theta_i(t) - \sum x_i \sqrt{\lambda_i}\,\theta_i(t) = 0.$$

Так как $\int h_r(t, \tau)x(\tau)d\tau = 0$ для всех $x(\tau)$, то $h_r(t, \tau)$ должно равняться нулю и (8.4.23) доказано. Для того чтобы показать, что это приводит к $h_r(t, \tau) = 0$, следует разложить h_r по функциям полного множества ортонормальных функций и прийти немедленно к противоречию, если предположить, что какой-либо член отличен от нуля. Равенство (8.4.24) следует из (8.4.23). Для того чтобы показать это, надо в соотношении (8.1.5) заменить $x(t)$ на $h(t, \tau)$ и проинтегрировать по t и τ. Так как сумма $\sum_i \lambda_i$ конечна, то λ_i не могут иметь предельных точек, кроме нуля, и могут быть расположены в порядке невозрастания.

Равенства (8.4.25) и (8.4.26) доказываются так же, как (8.4.23), равенство (8.4.27) доказывается так же, как (8.4.24).

Наконец, обратимся к задаче нахождения максимума в (8.4.28). Используя (8.4.17), имеем

$$\left\| \int h(t,\tau)\,x(\tau)\,d\tau \right\|^2 = \left\| \sum x_j \sqrt{\lambda_j}\,\theta_j(t) \right\|^2 = \sum_j x_j^2\,\lambda_j.$$

Поскольку при максимизации $x_1, \ldots, x_{i-1}$, по условию, равны 0 и λ_j убывает с j, то имеем

$$\left\| \int h(t,\tau)\,x(\tau)\,d\tau \right\|^2 \leqslant \lambda_i \sum_{j=i}^{\infty} x_j^2.$$

Так как $\|x\|$ должно быть равно 1, то λ_i является верхней границей для правой части (8.4.28). Однако если $x(\tau) = \varphi_i(\tau)$, то $\| \int h(t,\tau)\,x(\tau)d\tau \|^2 = \lambda_i$. Следовательно, (8.4.28) справедливо и $\varphi_i(\tau)$ — максимизирующая функция. Равенство (8.4.29) получается точно таким же образом.

Профильтрованный шум и разложение Карунена — Лоэва

В качестве примера использования предыдущей теоремы рассмотрим прохождение белого гауссова шума с единичной спектральной плотностью через фильтр с импульсным откликом $g(t, \tau)$. Будем интересоваться главным образом случаем, когда $g(t, \tau)$ — инвариатный во времени фильтр с ограниченным по времени выходом

$$g(t, \tau) = \begin{cases} g_1(t-\tau); & |t| \leqslant T_0/2, \\ 0 \quad ; & |t| > T_0/2. \end{cases} \tag{8.4.34}$$

Независимо от того, удовлетворяется (8.4.34) или нет, будем предполагать, что $g(t, \tau)$ интегрируема в квадрате. Пусть $\varphi_i(\tau)$, $\theta_i(t)$ и λ_i — соответственно собственные функции на входе, собственные функции на выходе и собственные значения фильтра $g(t, \tau)$ в смысле теоремы 8.4.1. Пусть $n(\tau)$ представляет собой выборочную функцию белого гауссова шума с единичной спектральной плотностью. Напомним, что, как это следует из предыдущего рассмотрения белого гауссова шума, эта выборочная функция $n(\tau)$ не является вполне определенной функцией времени, однако по предположению действие $n(\tau)$ на линейный фильтр вполне определено. Таким образом, положим, что $z(t)$ — выход фильтра $g(t, \tau)$, соответствующий входу $n(\tau)$:

$$z(t) = \int g(t, \tau)\ n(\tau) d\tau. \tag{8.4.35}$$

Определим также n_i и z_i равенствами

$$n_i = \int n(\tau)\varphi_i(\tau) d\tau, \tag{8.4.36}$$

$$z_i = \int z(t)\theta_i(t) dt. \tag{8.4.37}$$

Из (8.1.41) видно, что n_i — статистически независимые гауссовские случайные величины с нулевыми средними и единичными дисперсиями. Величины z_i и n_i могут быть связаны с помощью подстановки в (8.4.35) разложения для $g(t, \tau)$, задаваемого (8.4.23)

$$z(t) = \int \sum_i \sqrt{\lambda_i}\, \varphi_i(\tau)\, \theta_i(t)\, n(\tau)\, d\tau = \sum_{i=1}^{L} \sqrt{\lambda_i}\, n_i\, \theta_i(t) + \\ + \int \sum_{i=L+1}^{\infty} \sqrt{\lambda_i}\, \varphi_i(\tau)\, \theta_i(t)\, n(\tau)\, d\tau. \tag{8.4.38}$$

Теперь обозначим остаточный член в (8.4.38) через $z_L(t)$ и положим

$$g_L(t, \tau) = \sum_{i=L+1}^{\infty} \sqrt{\lambda_i}\, \varphi_i(\tau)\, \theta_i(t),$$

$$z_L(t) = \int g_L(t, \tau)\, n(\tau)\, d\tau. \tag{8.4.39}$$

Из (8.1.47), подставляя $g_L(t, \tau)$ вместо $h(t - \tau)$, имеем

$$\overline{z_L^2(t)} = \int g_L^2(t, \tau)\, d\tau.$$

Так как $g_L(t, \tau)$ стремится к 0 с возрастанием L в смысле предела в среднем, то

$$\lim_{L\to\infty} \int \overline{z_L^2(t)}\, dt = 0. \tag{8.4.40}$$

В силу того, что $z_L^2(t)$ неотрицательна, из (8.4.40) следует, что $\lim z_L^2(t) = 0$ почти всюду с вероятностью 1. Следовательно, $z(t)$ можно представить в виде

$$z(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\, n_i\, \theta_i(t) = \sum_{i=1}^{\infty} z_i\, \theta_i(t), \tag{8.4.41}$$

$$z_i = \sqrt{\lambda_i}\, n_i. \tag{8.4.42}$$

Так как n_i — независимые случайные величины с дисперсией 1, то z_i — независимые гауссовские случайные величины с нулевыми средними, удовлетворяющие соотношению

$$\overline{z_i z_j} = \delta_{ij} \lambda_i. \tag{8.4.43}$$

Разложение $z(t)$ в (8.4.41) известно как разложение Карунена — Лоэва. Мы видим, что функции $\theta_i(t)$ особенно удобны для представления $z(t)$ по двум причинам. Во-первых, они ортонормальны, и, во-вторых, случайные величины z_i статистически независимы.

Заметим, что мы не показали, что множество функций $\theta_i(t)$ полно, а показали лишь, что оно достаточно полно для того, чтобы представить выход фильтра $g(t, \tau)$ после прохождения через него белого гауссова шума. Покажем теперь, что если $g(t, \tau)$ ограничено во времени согласно (8.4.34), то множество $\theta_i(t)$ является полным на интервале $(-T_0/2, T_0/2)$.

Предположим, что функция $v(t)$ интегрируема в квадрате, отлична от нуля только в интервале $(-T_0/2, T_0/2)$ и ортогональна ко всем $\theta_i(t)$. Тогда из (8.4.19) получаем

$$\begin{aligned} \int g(t,\tau) v(t) dt &= 0, \\ \int g_1(t-\tau) v(t) dt &= 0. \end{aligned} \tag{8.4.44}$$

Полагая, что $G_1(f)$ и $V(f)$ — преобразования Фурье g_1 и v, и применяя теорему о свертке, получаем из (8.4.44)

$$\begin{aligned} G_1^*(f) V(f) &= 0, \\ V(f) &= 0; \quad f : G(f) \neq 0. \end{aligned}$$

Далее покажем, что $V(f)$ — аналитическая всюду функция f и, следовательно, равенство $V(f) = 0$ на любом интервале, где $G_1(f) \neq 0$ означает, что равны нулю все члены в разложении $V(f)$ в ряд Тейлора в окрестности какой-либо точки этого интервала; это, в свою очередь, означает, что $V(f) = 0$ всюду. Имеем

$$V(f) = \int_{-T_0/2}^{T_0/2} v(t)\, e^{-j2\pi ft}\, dt, \quad \frac{dV(f)}{df} = \int_{-T_0/2}^{T_0/2} -j2\pi t v(t)\, e^{-j2\pi ft}\, dt.$$

Так как $v(t)$ интегрируема в квадрате, то $dV(f)/df$ существует и конечна для всех комплексных f. Следовательно, $V(f)$ аналитическая функция и $v(t) = 0$ почти всюду.

Автокорреляция выходного процесса, т. е. функция $\overline{z(t_1)z(t_2)}$, может теперь быть найдена с помощью (8.4.41)

$$\overline{z(t_1)z(t_2)} = \overline{\sum_{i,j} z_i z_j \theta_i(t_1)\theta_j(t_2)}. \tag{8.4.45}$$

Изменяя порядок суммирования и усреднения и используя (8.4.43), получаем

$$\overline{z(t_1)z(t_2)} = \sum_i \lambda_i \theta_i(t_1)\theta_i(t_2) = \tag{8.4.46}$$

$$= \mathscr{R}_0(t_1, t_2), \tag{8.4.47}$$

где $\mathscr{R}_0$ задается равенством

$$\mathscr{R}_0(t_1,t_2) = \int g(t_1,\tau)g(t_2,\tau)\,d\tau. \tag{8.4.48}$$

Эти равенства справедливы в обычном среднеквадратическом смысле.

В этом месте можно вновь рассмотреть предыдущее изложение, начиная с предположения, что $z(t)$ — выборочная функция или усеченная выборочная функция произвольного гауссовского случайного процесса с нулевым средним и автокорреляционной функцией $\mathscr{R}_0(t_1, t_2)$. В предположении, что $\mathscr{R}_0(t_1, t_2)$ интегрируема в квадрате, можно положить, что $\theta_i(t)$ и λ_i — собственные функции и собственные значения (8.4.13). Можно опять представить $z(t)$ в виде (8.4.37) и коэффициенты z_i будут совместно гауссовскими с нулевыми средними. Покажем теперь, что коэффициенты будут некоррелированы и, следовательно, статистически независимыми. Имеем

$$\overline{z_i z_j} = \iint \overline{z(t_1)z(t_2)}\theta_i(t_1)\theta_j(t_2)dt_1dt_2 = \tag{8.4.49}$$

$$= \iint \mathscr{R}_0(t_1, t_2)\theta_i(t_1)\theta_j(t_2)dt_1dt_2 = \tag{8.4.50}$$

$$= \lambda_i\delta_{ij}. \tag{8.4.51}$$

Далее нужно исследовать, будет ли равен нулю остаточный член

$$z_r(t) = z(t) - \sum_i z_i\theta_i(t).$$

Применяя неравенство Бесселя в виде (8.1.5), имеем

$$\int z_r^2(t)\,dt = \int z^2(t)\,dt - \sum_{i=1}^{\infty} z_i^2,$$

$$\overline{\int z_r^2(t)\,dt} = \int \mathscr{R}_0(t, t)\,dt - \sum_i \lambda_i. \tag{8.4.52}$$

Если имеется некоторая интегрируемая в квадрате функция $g(t, \tau)$, для которой $\mathscr{R}_0(t_1, t_2)$ — корреляция на выходе, задаваемая (8.4.48), то $\int \mathscr{R}_0(t, t)dt = \iint g^2(t, \tau)d\tau dt$. Из (8.4.24) следует, что это выражение равно $\Sigma\lambda_i$, и $z_r(t)$ равно нулю почти всюду с вероятностью 1.

Если также T_0 конечно и $\mathcal{R}_0(t_1, t_2)$ непрерывна, то теорема Мерсера*) утверждает, что (8.4.26) сходится равномерно для всех t_1, t_2. Подставляя (8.4.26) в (8.4.52) и интегрируя, находим, что правая часть опять равна нулю. Из математических соображений следует, что $z_r(t)$ не равна нулю для полностью произвольных автокорреляционных функций. Например, если $\mathcal{R}_0$ равна сумме непрерывной функции и функции, которая равна 1 для $t_1 = t_2$ и равна 0 в других точках, то добавляемая функция не отражается на $\theta_i(t)$ или λ_i, однако она изменяет $\int \mathcal{R}_0(t, t)\, dt$. В последующем изложении мы будем игнорировать такие паталогии, поскольку они не соответствуют случаям, представляющим какой-либо физический интерес. Таким образом, для всех случаев, представляющих интерес, опять приходим к представлению Карунена — Лоэва (8.4.41)**)

Подытоживая изложенные выше результаты, получаем, что профильтрованный гауссов белый шум можно представить в виде (8.4.41) и гауссовский случайный процесс с автокорреляционной функцией $\mathcal{R}_0(t_1, t_2)$ можно представить в виде (8.4.41). Следовательно, если для заданной автокорреляционной функции случайного процесса можно найти функцию $g(t, \tau)$, которая удовлетворяет (8.4.48), то этот случайный процесс можно рассматривать как результат прохождения белого шума через фильтр $g(t, \tau)$. Рассмотрение небелого гауссова шума как профильтрованного белого шума является весьма полезным в различных задачах.

В случае стационарного гауссовского процесса с интегрируемой спектральной плотностью $N(f)$ довольно легко найти $g(t, \tau)$, удовлетворяющую (8.4.48). Определим

$$g_1(t) = \int \sqrt{N(f)}\, e^{j2\pi ft}\, df. \tag{8.4.53}$$

Полагая, что $g(t, \tau)$ является усеченным вариантом $g_1(t)$ и задается (8.4.34), имеем

$$\int g(t_1, \tau)\, g(t_2, \tau)\, d\tau = \int g_1(t_1 - \tau)\, g_1(t_2 - \tau)\, d\tau = \tag{8.4.54}$$

$$= \int N(f)\, e^{j2\pi f(t_1 - t_2)}\, df = \tag{8.4.55}$$

$$= \mathcal{R}(t_1 - t_2); \quad |t_1| \leqslant T_0/2, \ |t_2| \leqslant T_0/2. \tag{8.4.56}$$

Здесь $\mathcal{R}(\tau)$ — автокорреляционная функция процесса, задаваемая обратным преобразованием Фурье $N(f)$. Полагая $\mathcal{R}_0(t_1, t_2)$ равной $\mathcal{R}(t_1 - t_2)$ для $|t_1|$ и $|t_2|$, меньших или равных $T_0/2$, получаем (8.4.48). Для проверки заметим, что если белый шум подать на фильтр, например, с частотным откликом $\sqrt{N(f)}$, то на выходе будет процесс со спектральной плотностью $N(f)$.

---

*) Например, см. Рисс и Надь (1955).

**) Теорема Карунена—Лоэва [см. Лоэв (1955)] фактически применима к случаю, когда R_0 непрерывна и T_0 конечно. Она утверждает также, что (8.4.41) сходится в среднеквадратическом для каждого значения t, тогда как мы голословно утверждали наличие среднеквадратической сходимости по ансамблю и t.

Идеальные фильтры нижних частот

Важным частным случаем фильтра теоремы 8.4.1 является идеальный фильтр нижних частот, отсекающий частоты выше некоторой заданной частоты W. Такой фильтр описывается частотной характеристикой

$$H_1(f) = \begin{cases} 1; & |f| \leqslant W, \\ 0; & |f| > W. \end{cases} \tag{8.4.57}$$

Импульсный отклик $h_1(t)$ является обратным преобразованием Фурье $H_1(f)$,

$$h_1(t) = \frac{\sin 2\pi W t}{\pi t}. \tag{8.4.58}$$

Рассмотрим теперь вход лишь на некотором интервале $(-T/2, T/2)$ и определим

$$h(t, \tau) = \begin{cases} h_1(t-\tau); & |\tau| \leqslant T/2, \\ 0 \quad ; & |\tau| > T/2. \end{cases} \tag{8.4.59}$$

Множество входных собственных функций $\{\varphi_i(\tau)\}$ для фильтра $h(t, \tau)$ задаются (8.4.10) как решения интегрального уравнения

$$\int \mathscr{R}(\tau_1, \tau_2)\, \varphi_i(\tau_2) d\tau_2 = \lambda_i \varphi_i(\tau_1), \tag{8.4.60}$$

где

$$\mathscr{R}(\tau_1, \tau_2) = \begin{cases} \mathscr{R}_1(\tau_2 - \tau_1); & |\tau_1| \leqslant T/2, \quad |\tau_2| \leqslant T/2, \\ 0 \quad ; & \text{во всех других точках,} \end{cases}$$

$$\mathscr{R}_1(\tau_2 - \tau_1) = \int h_1(t-\tau_1) h_1(t-\tau_2) dt. \tag{8.4.61}$$

Взяв преобразование Фурье от обеих частей (8.4.61), найдем, что преобразованием Фурье от $\mathscr{R}_1$ является функция $|H_1(f)|^2$, которая численно равна $H_1(f)$. Следовательно,

$$\mathscr{R}(\tau_1, \tau_2) = \begin{cases} \dfrac{\sin 2\pi(\tau_1 - \tau_2) W}{\pi(\tau_1 - \tau_2)}; & |\tau_1| \leqslant T/2, \quad |\tau_2| \leqslant T/2, \\ 0 \quad ; & \text{во всех других точках.} \end{cases} \tag{8.4.62}$$

Входные собственные функции $\varphi_i(\tau)$ [которые ограничены интервалом $(-T/2, T/2)$] связаны с выходными собственными функциями $\theta_i(t)$ соотношениями

$$\theta_i(t) = \frac{1}{\sqrt{\lambda_i}} \int h_1(t-\tau)\, \varphi_i(\tau)\, d\tau, \tag{8.4.63}$$

$$\varphi_i(\tau) = \begin{cases} \dfrac{1}{\sqrt{\lambda_i}} \int h_1(t-\tau)\, \theta_i(t)\, dt; & |\tau| \leqslant T/2, \\ 0 \quad ; & |\tau| > T/2. \end{cases} \tag{8.4.64}$$

Из (8.4.63) видно, что $\theta_i(t)$ равна умноженной на $1/\sqrt{\lambda_i}$ функции $\varphi_i(t)$, усеченной к полосе частот $|f| \leqslant W$. Другими словами, преобразо-

вания Фурье $\theta_i(t)$ и $\varphi_i(t)$ связаны соотношением

$$\Theta_i(f) = \begin{cases} \frac{1}{\sqrt{\lambda_i}} \Phi_i(f); & |f| \leqslant W, \\ 0 \quad ; & |f| > W. \end{cases} \tag{8.4.65}$$

Кроме того, так как функция $\theta_i(\tau)$ имеет ограниченную полосу частот, то она проходит без изменений через фильтр $h_1(t)$, т. е. $\int h_1(\tau - t)\theta_i(t)dt$ равно $\theta_i(\tau)$. Так как $h_1(t)$ — четная функция, то вместе с (8.4.64) это дает

$$\varphi_i(\tau) = \begin{cases} \frac{1}{\sqrt{\lambda_i}} \theta_i(\tau); & |\tau| \leqslant T/2, \\ 0 \quad ; & |\tau| > T/2. \end{cases} \tag{8.4.66}$$

Следовательно, функции $\theta_i(t)$ имеют специфическое свойство: они ортонормальны на бесконечном интервале, а также ортогональны на интервале $(-T/2, T/2)$. Функции $\theta_i(t)$ известны как волновые функции вытянутого сфероида и они часто встречаются в разных задачах физики и математики. Имеется большая литература по этим функциям, и читатель, в частности, может обратиться к работам Слепяна, Поллака и Ландау (1961), (1962) и (1964). Некоторые весьма полезные свойства этих функций состоят в том, что все их собственные значения λ_i различны и каждое соответствует единственной нормированной собственной функции (с точностью до знака). Функция $\theta_i(t)$ имеет точно $i - 1$ нулей внутри интервала $(-T/2, T/2)$ и является четной функцией для нечетного i, и наоборот.

Как следует из обсуждения теоремы 8.4.1, $\varphi_1(\tau)$ является нормированной функцией на интервале $(-T/2, T/2)$, содержащей наибольшую энергию λ_1 в полосе $-W \leqslant f \leqslant W$. Аналогично $\varphi_i(\tau)$ — нормированная функция на $(-T/2, T/2)$, которая имеет наибольшую энергию λ_i в интервале $-W \leqslant f \leqslant W$ при условии ортогональности к $\varphi_1(\tau), \ldots, \varphi_{i-1}(\tau)$. Из тех же соображений следует, что $\theta_i(t)$ является нормированной функцией с ограниченной полосой частот $(-W, W)$, которая имеет наибольшую энергию λ_i на временно́м интервале $(-T/2, T/2)$ при условии ортогональности к $\theta_1(t), \ldots, \theta_{i-1}(t)$.

Теперь можно возвратиться к тому, чтобы дать более точное толкование утверждения, что класс сигналов, который приближенно ограничен во времени и по частоте, имеет около $2WT$ степеней свободы. Рассмотрим множество функций, которые являются линейными комбинациями первых n собственных функций (8.4.60) $\varphi_1(\tau), \ldots, \varphi_n(\tau)$. Пусть

$$x_n(\tau) = \sum_{i=1}^{n} x_i \varphi_i(\tau)$$

— произвольная функция этого класса и пусть

$$u_n(t) = \sum_{i=1}^{n} x_i \sqrt{\lambda_i}\, \theta_i(t)$$

— часть $x_n(\tau)$, лежащая в полосе частот $-W \leqslant f \leqslant W$. Доля энергии $x_n(\tau)$, которая содержится в полосе $-W \leqslant f \leqslant W$, задается соотношением

$$1 \geqslant \frac{\int u_n^2(t)\,dt}{\int x_n^2(\tau)\,d\tau} = \frac{\sum_{i=1}^{n} \lambda_i x_i^2}{\sum_{i=1}^{n} x_i^2} \geqslant \lambda_n. \tag{8.4.67}$$

Последнее приведенное выше неравенство следует из того, что все λ_i в сумме, стоящей в числителе, ограничены снизу λ_n. Таким образом, рассматривается класс ограниченных по времени функций, имеющих n степеней свободы, а у всех функций доля энергии, содержащаяся в полосе частот $-W \leqslant f \leqslant W$, заключена между λ_n и 1. Заметим, кроме того, что одна из функций $\varphi_n(\tau)$ имеет в полосе $-W \leqslant f \leqslant W$ энергию, в точности равную λ_n.

Прежде чем перейти к изучению поведения λ_n в зависимости от n, установим, что любое другое множество функций на интервале $(-T/2, T/2)$, имеющих n степеней свободы, также содержит функцию с λ_n или меньшей долей ее энергии в полосе $-W \leqslant f \leqslant W$. Под множеством функций с n степенями свободы понимается множество линейных комбинаций n линейно независимых функций. Могут быть два случая: или эти функции образуют то же пространство, что и $\varphi_1(\tau), \ldots, \varphi_n(\tau)$, или имеется линейная комбинация этих функций, ортогональная к $\varphi_1(\tau), \ldots, \varphi_n(\tau)$. В первом случае $\varphi_n(\tau)$ содержится в этом множестве и имеет в полосе $-W \leqslant f \leqslant W$ энергию, равную λ_n. Во втором случае функция, ортогональная к $\varphi_1(\tau) \ldots, \varphi_n(\tau)$, является линейной комбинацией $\varphi_i(\tau)$ с $i > n$. Следовательно, так как λ_i при $i > n$ ограничены сверху λ_n, то доля энергии этой функции в полосе $-W \leqslant f \leqslant W$ не больше, чем λ_n.

Последний вопрос, на который теперь следует ответить, состоит в том, как λ_n зависит от n, W и T. Изменяя масштаб времени в (8.4.10) и (8.4.62), видим, что λ_n зависит только от n и от произведения WT, и мы будем писать $\lambda_n(WT)$ для того, чтобы подчеркнуть эту зависимость. Слепян (1965) показал, что если для каждых n и WT ввести число α, определяемое равенством

$$n = 2WT + 1 + \frac{\alpha}{\pi^2} \ln(4\pi WT), \tag{8.4.68}$$

то

$$\lim_{WT \to \infty} \left| \lambda_n(WT) - \frac{1}{1 + e^{\alpha}} \right| = 0, \tag{8.4.69}$$

где n в (8.4.69) для каждого WT таковы, что α лежит в ограниченной области, не зависящей от WT.

Графическое пояснение этой зависимости приведено на рис. 8.4.2. Основное, что здесь следует отметить, это то, что для $n \ll 2WT + 1$ имеем $\lambda_n \approx 1$ и для $n \gg 2WT + 1$ имеем $\lambda_n = 0$. Переходная область между этими экстремальными значениями имеет ширину, пропорцио-

нальную $\ln(4\pi WT)$. В частности, для любого фиксированного $\varepsilon > 0$ эти равенства означают, что

$$\lim_{WT\to\infty} \lambda_{2WT(1+\varepsilon)}(WT) = 0, \tag{8.4.70}$$

$$\lim_{WT\to\infty} \lambda_{2WT(1-\varepsilon)}(WT) = 1. \tag{8.4.71}$$

Следовательно, если используется множество функций на $(-T/2, T/2)$ с $2WT(1+\varepsilon)$ степенями свободы, то для произвольно больших WT некоторые из этих функций будут иметь в полосе $-W \leqslant f \leqslant W$ исчезающе малую долю энергии. Обратно, при $2WT(1-\varepsilon)$ степенях свободы минимум доли энергии в полосе $-W \leqslant f \leqslant W$, взятый по всем функциям в классе, будет сходиться к 1 при неограниченном возрастании WT.

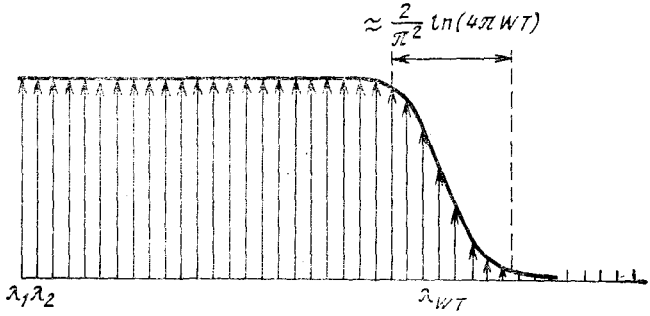


Рис. 8.4.2. Поведение собственных значений (8.4.60) при больших WT.

Последняя особенность функции вытянутого сфероида, которая сейчас будет показана, состоит в том, что $\varphi_i(\tau)$ и $\theta_i(t)$ с точностью до масштаба являются преобразованиями Фурье друг друга. Используя $H_1(f)$ из (8.4.57), для того чтобы получить ограничение по времени, можно переписать (8.4.66) следующим образом:

$$\varphi_i(\tau) = \frac{1}{\sqrt{\lambda_i}}\,\theta_i(\tau)\,H_1\left(\frac{2W\tau}{T}\right). \tag{8.4.72}$$

Взяв преобразование Фурье от обеих частей (8.4.72), получим

$$\Phi_i(f_2) = \frac{T}{2W\sqrt{\lambda_i}} \int \Theta_i(f_1)\,h_1\left[\frac{T(f_2-f_1)}{2W}\right] df_1. \tag{8.4.73}$$

Подставляя (8.4.65) в (8.4.73) и используя то, что $\mathcal{R}(\tau_1, \tau_2) = h_1(\tau_2 - \tau_1)$ для $|\tau_1|, |\tau_2| \leqslant T/2$, это равенство приводим к виду

$$\Theta_i(f_2) = \frac{T}{2W\lambda_i} \int_{-W}^{W} \Theta_i(f_1)\,\mathcal{R}\left(\frac{Tf_2}{2W}, \frac{Tf_1}{2W}\right) df_1. \tag{8.4.74}$$

Наконец, изменяя масштаб f_1 и f_2 в $2W/T$ раз, видим, что $\Theta_i(2Wf/T)$

удовлетворяет тому же самому интегральному уравнению (8.4.10), что и $\varphi_i(\tau)$. Так как его решения единственны с точностью до множителя, то, используя нормировку и то, что нумерация функций $\theta_i(t)$ несущественна, получаем

$$\Theta_i\left(\frac{2Wf}{T}\right) = \pm \sqrt{\frac{T}{2W}}\,\varphi_i(f)\,(\sqrt{-1})^{i-1}. \qquad (8.4.75)$$

Из (8.4.75) видно, что $\theta_i(t)$ не стремится ни к синусоидам, ни к отсчетным функциям на интервале $(-T/2,\ T/2)$, когда T становится большим. Это, в свою очередь, объясняет, почему эвристические соображения, приведенные в § 8.3, не могут быть без большого труда сделаны точными.

8.5. КАНАЛЫ С АДДИТИВНЫМ ГАУССОВЫМ ШУМОМ И СИГНАЛАМИ НА ВХОДЕ, ОГРАНИЧЕННЫМИ ПО МОЩНОСТИ И ПО ЧАСТОТЕ

В этом параграфе результаты, полученные в предыдущем параграфе, используются для того, чтобы получить строгое решение задач, рассмотренных в § 8.3. Канал изображен на рис. 8.5.1. Вход $x(t)$ рассматривается на временном интервале $(-T/2,\ T/2)$ и проходит через линейный инвариантный во времени фильтр с импульсным откликом $h_1(\tau)$ и частотной характеристикой $H_1(f) = \int h_1(\tau)\,e^{-j2\pi f\tau}\,d\tau$. Стационарный гауссов шум с нулевым средним и спектральной плотностью $N(f)$ добавляется к выходу фильтра и результат наблюдается на интервале $(-T_0/2,\ T_0/2)$. Вход $x(t)$ ограничен по мощности величиной S.

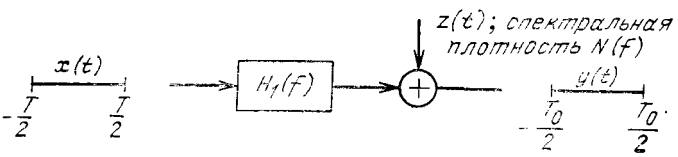


Рис. 8.5.1.

В обращении теоремы кодирования под этим будет пониматься, что математическое ожидание величины

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

не более ST. Для теоремы кодирования будет использоваться более сильное ограничение

$$\int_{-T/2}^{T/2} x_m^2(t)\,dt \leqslant ST$$

для каждого кодового слова.

Покажем сначала, как свести непрерывный по времени канал, изображенный на рис. 8.5.1, к множеству параллельных дискретных по времени каналов с аддитивными гауссовыми шумами. Тогда могут быть непосредственно применены результаты § 7.5. Затем рассмотрим более трудную задачу о переходе к пределу при $T \to \infty$ и условии $T = T_0$. Окончательные результаты будут те же самые, как и в § 8.3. В дальнейшем будем предполагать, что

$$\int_{-\infty}^{\infty} \frac{|H_1(f)|^2}{N(f)}\, df < \infty, \tag{8.5.1}$$

а также, что или $\int N(f)df < \infty$, или что шум белый (т. е. $N(f)$ не зависит от f).

Одним из недостатков развиваемого здесь подхода является предположение о том, что вход равен нулю вне интервала $(-T/2, T/2)$. Другими словами, когда здесь будет использоваться кодовое ограничение длины T, то не будет учитываться межсимвольная интерференция между последовательными кодовыми словами. Это не приводит к каким-либо трудностям при рассмотрении обращения теоремы кодирования, так как легко показать, что межсимвольная интерференция не может уменьшить вероятность ошибки. Также не возникают трудности при доказательстве того, что сколь угодно малая вероятность ошибки может быть достигнута при любой скорости, меньшей пропускной способности, так как в принципе можно передавать только одно кодовое слово сколь угодно большой длины. Кажется также, что межсимвольная интерференция не уменьшает показатель экспоненты вероятности ошибки в пределе, когда T становится большим, однако до сих пор это строго не доказано. Интуитивные соображения состоят в следующем. Если при некотором малом $\varepsilon > 0$ отделить кодовые слова длительности T в канале защитным интервалом $T^{1-\varepsilon}$, то в пределе при $T \to \infty$ отношение защитного интервала к длине кодового слова стремится к 0. Вместе с тем вклад в энергию на любом наблюдаемом интервале, вносимый кодовыми словами из других интервалов, стремится к нулю при $T \to \infty$ и, следовательно, этот вклад не должен влиять на вероятность ошибки*).

В § 8.4 было показано, что гауссов шум с интегрируемой и конечной спектральной плотностью $N(f)$ может рассматриваться как результат прохождения белого гауссова шума единичной спектральной плотности через (нереализуемый) фильтр с частотной характеристикой $\sqrt{N(f)}$ и импульсным откликом

$$g_1(t) = \int \sqrt{N(f)}\, e^{j2\pi ft}\, df. \tag{8.5.2}$$

Так как $N(f)$ — четная функция f, то $g_1(t)$ — действительная функция, и так как $\sqrt{N(f)}$ интегрируема в квадрате, то $g_1(t)$ также интегрируема

*) Это утверждение справедливо при выполнении некоторых условий на скорость затухания межсимвольной интерференции. (*Прим. ред.*).

в квадрате. Если шум белый со спектральной плотностью $N_0/2$, то фильтр $g_1(t)$ можно рассматривать как умножитель, усиливающий вход в $\sqrt{N_0/2}$ раз.

Разобьем умозрительно фильтр $H_1(f)$ на две части, одну с частотной характеристикой $H_1(f)/\sqrt{N(f)}$ и другую с частотной характеристикой $\sqrt{N(f)}$ (рис. 8.5.2). Опять, если $N(f) = N_0/2$, второй фильтр следует рассматривать как умножитель. Если $H_1(f)$ и $N(f)$ равны нулю для какого-либо f, то положим, по определению, $H_1(f)/\sqrt{N(f)}$ равным нулю в этой точке f. Читателю со слабым радиотехническим образованием было бы полезно проверить, что частотная характеристика двух после-

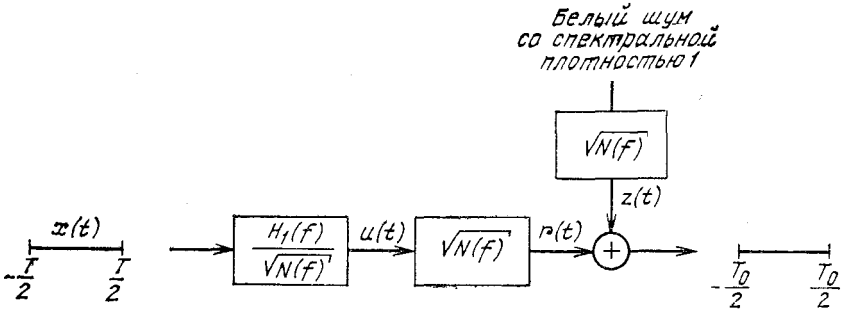


Рис. 8.5.2. Эквивалентное представление рис. 8.5.1.

довательных линейных инвариантных во времени фильтров действительно равна произведению частотных характеристик. Обозначим импульсный отклик фильтра $H_1(f)/\sqrt{N(f)}$ через

$$K_1(t) = \int \frac{H_1(f)}{\sqrt{N(f)}} e^{j2\pi ft}\, df. \tag{8.5.3}$$

Функция $K_1(t)$, подобно $g_1(t)$, действительная и интегрируема в квадрате.

Вход $x(\tau)$ и выход $r(t)$ первоначального фильтра связаны в терминах новых фильтров $K_1(t)$ и $g_1(t)$ соотношениями

$$r(t_1) = \int g_1(t_1 - t)\, u(t) dt, \tag{8.5.4}$$

$$u(t) = \int K_1(t - \tau)\, x(\tau) d\tau. \tag{8.5.5}$$

Так как $u(t)$ и белый шум (рис. 8.5.2) фильтруются одним и тем же фильтром и затем складываются, то систему можно представить в виде, указанном на рис. 8.5.3, где белый шум непосредственно добавляется к $u(t)$ и затем результат фильтруется с помощью $g_1(t)$. Хотя канал, изображенный на рис. 8.5.3, выглядит совсем отличным от канала на рис. 8.5.1, они тождественны в том смысле, что выходная функция $y(t)$ в обоих случаях равна сумме $r(t)$ и гауссова шума со спектральной плотностью $N(f)$; пока мы основываемся на том, что приемник наблю-

дает только $y(t)$, эти рисунки можно использовать на равных основаниях.

Для того чтобы использовать разложения последнего параграфа, удобно заменить инвариантные во времени фильтры рис. 8.5.3 на ме-

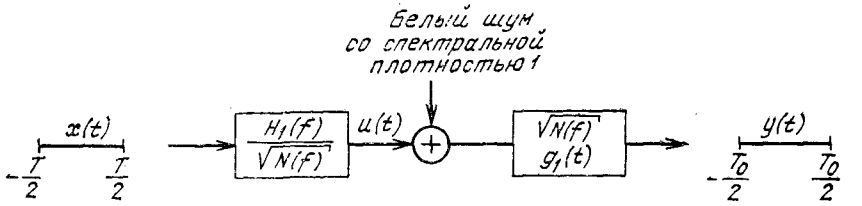


Рис. 8.5.3. Эквивалентное представление рис. 8.5.1.

няющиеся во времени фильтры, которые автоматически ограничивают вход интервалом $(-T/2,\ T/2)$ и выход интервалом $(-T_0/2,\ T_0/2)$. Таким образом, определим

$$K(t,\tau)=\begin{cases}K_1(t-\tau); & |\tau|\leqslant T/2,\\ 0 \quad ; & |\tau|>T/2,\end{cases} \tag{8 5.6}$$

$$g(t,\tau)=\begin{cases}g_1(t-\tau); & |t|\leqslant T_0/2,\\ 0 \quad ; & |t|>T_0/2.\end{cases} \tag{8.5.7}$$

Этот канал изображен на рис. 8.5.4, однако ограничения на длительность входа и на интервал наблюдения опущены, так как эти операции выполняются в канале.

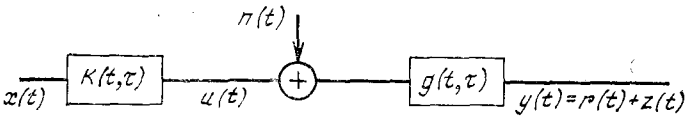


Рис. 8.5.4. Эквивалентное представление рис. 8.5.1.

Пусть $\xi_i(\tau)$, $\eta_i(t)$ и μ_i, $1\leqslant i\leqslant\infty$, соответственно собственные функции на входе, собственные функции на выходе и собственные значения фильтра $K(t,\tau)$ в смысле теоремы 8.4.1. Тогда $x(\tau)$ и $u(t)$ можно представить равенствами

$$x(\tau)=\sum_i x_i\,\xi_i(\tau), \tag{8.5.8}$$

$$u(t)=\sum_i x_i\sqrt{\mu_i}\,\eta_i(t). \tag{8.5.9}$$

Если можно было бы забыть о фильтре $g(t,\tau)$ на рис. 8.5.4, то тогда белый шум можно было бы представлять через ортонормальные функции $\eta_i(t)$ и для каждого i приемник мог бы вычислять $x_i\sqrt{\mu_i}+n_i$, где n_i — независимые нормированные гауссовские случайные величины.

К сожалению, приемник не может даже в принципе вычислить эти величины. Трудность состоит в том, что выход фильтра $g(t,\tau)$ не

определяет однозначно вход фильтра. Другими словами, в общем случае существует ненулевой вход фильтра $g(t, \tau)$, для которого выход равен нулю. Аналитически, пусть $\varphi_{i,g}(\tau)$, $\theta_{i,g}(t)$ и $\lambda_{i,g}$ для $1 \leqslant i < \infty$ соответственно входные и выходные собственные функции и собственные значения $g(t, \tau)$. Тогда из (8.4.15) следует, что вход фильтра $g(t, \tau)$ приводит к нулевому выходу тогда и только тогда, когда вход ортогонален к $\varphi_{i,g}(\tau)$ при всех i, $1 \leqslant i < \infty$.

Теперь предположим, что $u(t)$ (сигнал на входе фильтра $g(t, \tau)$) разделен на две компоненты: одна — линейная комбинация $\varphi_{i,g}(\tau)$ и другая — ортогональная ко всем $\varphi_{i,g}(\tau)$. Если фильтр $K(t, \tau)$ изменить так, чтобы он подавил компоненту, ортогональную ко всем $\varphi_{i,g}(\tau)$, то это не изменит сигнал на выходе $g(t, \tau)$. Вместе с тем ниже будет показано, что, когда $K(t, \tau)$ изменяется таким образом, фильтр $g(t, \tau)$ не разрушает информацию о $x(\tau)$.

Для того чтобы точно описать, как следует изменить фильтр $K(t, \tau)$, заметим, что выход фильтра $K(t, \tau)$ задается равенством

$$u(t) = \int x(\tau) K(t, \tau) d\tau. \tag{8.5.10}$$

Требуется заменить фильтр $K(t, \tau)$ на новый фильтр $K_0(t, \tau)$ с выходом

$$u_0(t) = \sum_i \varphi_{i,g}(t) \int_{-\infty}^{\infty} u(t_1) \varphi_{i,g}(t_1) dt_1. \tag{8.5.11}$$

Подставляя (8.5.10) в (8.5.11) и изменяя порядок интегрирования, получаем

$$u_0(t) = \int x(\tau) \left[\sum_i \varphi_{i,g}(t) \int K(t_1, \tau) \varphi_{i,g}(t_1) dt_1 \right] d\tau. \tag{8.5.12}$$

Следовательно, модифицированный фильтр должен иметь отклик

$$K_0(t, \tau) = \sum_i \varphi_{i,g}(t) \int K(t_1, \tau) \varphi_{i,g}(t_1) dt_1. \tag{8.5.13}$$

Теперь можно заменить фильтр $K(t, \tau)$ на рис. 8.5.4 на фильтр $K_0(t, \tau)$, что показано на рис. 8.5.5. Таким образом, $u_0(t)$ на рис. 8.5.5

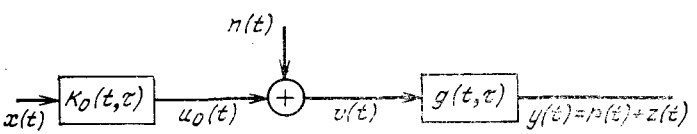


Рис. 8.5.5. Эквивалентное представление рис. 8.5.1.

отличается от $u(t)$ на рис. 8.5.4 только членом, ортогональным ко всем собственным функциям $\varphi_{i,g}$, и, следовательно, отклик $r(t)$, изображенный на рис. 8.5.5, в точности равен отклику на рис. 8.5.4. В частном случае, когда $N(f) = N_0/2$, действие $g(t, \tau)$ сводится к тому, что вход умножается на $N_0/2$ при $|t| \leqslant T_0/2$, и $K_0(t, \tau)$ определяется анало-

гично с помощью равенства

$$K_0(t,\tau)=\begin{cases} K(t,\tau); & |t|\leqslant T_0/2,\\ 0 \quad ; & |t|>T_0/2. \end{cases} \tag{8.5.14}$$

Далее покажем, что $K_0(t, \tau)$ интегрируема в квадрате, что позволяет использовать разложения § 8.4. Заметим, что для K_0, задаваемого (8.5.13), и для любого данного τ функция $K_0(t, \tau)$ является разложением $K(t, \tau)$ по функциям $\varphi_{i,g}(t)$. Следовательно, по неравенству Бесселя

$$\int_{-\infty}^{\infty} K_0^2(t,\tau)\,dt \leqslant \int_{-\infty}^{\infty} K^2(t,\tau)\,dt, \text{ для всех } \tau, \tag{8.5.15}$$

$$\iint K_0^2(t,\tau)\,dt\,d\tau \leqslant \int_{-T/2}^{T/2}\left[\int K_1^2(t-\tau)dt\right]d\tau = T\int_{-\infty}^{\infty} K_1^2(t)\,dt < \infty. \tag{8.5.16}$$

Для $K_0(t, \tau)$, задаваемого (8.5.14), интегрируемость в квадрате K_0 непосредственно следует из интегрируемости в квадрате $K(t, \tau)$.

Теперь пусть $\varphi_i(\tau)$, $\theta_i(t)$ и λ_i (для $1 \leqslant i < \infty$) — соответственно входные и выходные собственные функции и собственные значения фильтра $K(t, \tau)$ в смысле теоремы 8.4.1. Как было указано выше, нам не нужно заниматься нахождением собственных функций, а нужно лишь исследовать их предельное поведение при $T \to \infty$. Таким образом, сложное выражение (8.5.13) для K_0 в действительности не представляет особого интереса. Функции $x(\tau)$, $u_0(t)$ и $v(t) = u_0(t) + n(t)$ можно разложить в виде

$$x(\tau)=\sum_i x_i\varphi_i(\tau)+x_r(\tau), \tag{8.5.17}$$

$$u_0(t)=\sum_i x_i\sqrt{\lambda_i}\,\theta_i(t), \tag{8.5.18}$$

$$v(t)=\sum_i\left[x_i\sqrt{\lambda_i}+n_i\right]\theta_i(t)+n_r(t). \tag{8.5.19}$$

Как отмечено выше, $v(t)$ не вполне определена, так как содержит белый гауссов шум, однако в соответствии с определением белого гауссового шума коэффициенты $v_i = x_i\sqrt{\lambda_i} + n_i$ четко определены и случайные величины n_i — независимые нормированные гауссовские случайные величины. Остаточный член $n_r(t)$ в (8.5.19) представляет собой компоненту шума, ортогональную ко всем функциям $\theta_i(t)$. Точнее, $\int n_r(t)\theta_i(t)dt = 0$ для всех i. Точно так же для любой функции $\psi(t)$ с единичной энергией, ортогональной к $\theta_i(t)$ при всех i, величина $\int n_r(t)\,\psi(t)dt$ является нормированной гауссовской случайной величиной, не зависимой от всех x_i и n_i. Если образовать множество ортонормальных функций $\{\psi_j(t)\}$, каждая из которых ортонормирована с функциями множества $\{\theta_i(t)\}$, то канал может быть представлен как бесконечное множество параллельных каналов, соединенных последовательно с фильтром $g(t, \tau)$ (рис. 8.5.6). В последующем изложении

будет показано, что коэффициенты $v_1, v_2, \ldots$ могут быть определены по выходу канала $y(t)$.

Теперь предположим, что функция $v(t)$ известна приемнику (т. е. известны последовательность $\{v_i\}$ и последовательность $\{w_j\}$). Для любой вероятностной меры, заданной на $x(\tau)$, средняя взаимная информация между $x(\tau)$ и $v(t)$ равна, очевидно, средней взаимной информации между последовательностью $\{x_i\}$ и последовательностью $\{v_i\}$ (так как последовательность $\{w_i\}$ на рис. 8.5.6 не зависит от пары $\{x_j\}$ и $\{v_j\}$). Аналогично для любого множества кодовых слов декодирование по максимуму правдоподобия и декодирование по максимуму апостериорной вероятности зависит только от последователь-

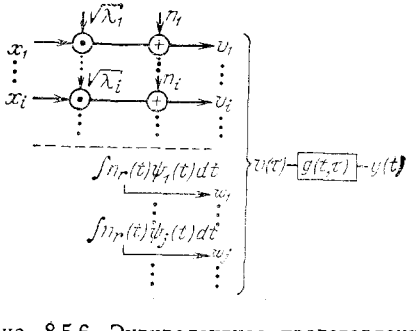


Рис. 8.5.6. Эквивалентное представление рис. 8.5.1.

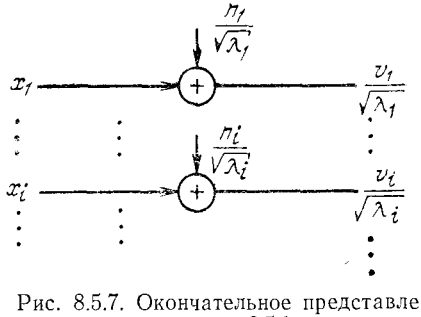


Рис. 8.5.7. Окончательное представление рис. 8.5.1.

ности $\{v_i\}$. Следовательно, только последовательность $\{v_i\}$ является существенной в сигнале $v(t)$. Покажем теперь, что последовательность $\{v_i\}$ может быть определена по окончательному выходу канала $y(t)$. Так как $y(t)$ определяется по $v(t)$, то отсюда следует, что дополнительные шумы $\{w_j\}$ можно не рассматривать (это и так почти очевидно) и канал можно представить в виде, изображенном на рис. 8.5.7. На рис. 8.5.7 аддитивный шум помещен слева от умножителя, для того чтобы сделать окончательное представление эквивалентным модели параллельного канала из § 7.5. Шум в i-м канале является теперь гауссовской случайной величиной с нулевым средним и дисперсией $1/\lambda_i$.

Л е м м а 8. 5. 1. Определенная выше последовательность случайных величин $\{v_i\}$ может быть однозначно найдена по выходу канала $y(t)$.

Доказательство. Если шум белый, то $y(t) = \sqrt{N_0/2}v(t)$ и доказательство тривиально. Для небелого шума величины v_i определяются через выходные собственные функции фильтра $K_0(t, \tau)$ с помощью равенства

$$v_i = \int v(t)\,\theta_i(t)\,dt. \tag{8.5.20}$$

Из (8.5.11) следует, что выход фильтра $K_0(t, \tau)$ при любом входе является линейной комбинацией входных собственных функций $\varphi_{i,g}(t)$

фильтра $g(t, \tau)$. Следовательно, для каждого i функция $\theta_i(t)$ является линейной комбинацией функций множества $\{\varphi_{i,g}(t)\}$ и может быть представлена в виде

$$\theta_i(t) = \sum_{j=1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t); \qquad \alpha_{ij} = \int \theta_i(t)\, \varphi_{j,\,g}(t)\, dt. \tag{8.5.21}$$

(Следует отметить, что в общем случае это разложение не может быть получено для выходных собственных функций фильтра $K(t, \tau)$ и фактически это вызвало введение фильтра $K_0(t, \tau)$.) Подставляя (8.5.21) в (8.5.20), изменяя порядок суммирования и интегрирования для произвольного числа m слагаемых, получаем

$$v_i = \sum_{j=1}^{m} \alpha_{ij} \int v(t)\, \varphi_{j,g}(t)\, dt + \int v(t) \sum_{j=m+1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t)\, dt. \tag{8.5.22}$$

В терминах выходных собственных функций $\theta_{j,\,g}(t)$ фильтра $g(t, \tau)$ имеем

$$\int y(t)\, \theta_{j,\,g}(t)\, dt = \int r(t)\, \theta_{j,\,g}(t)\, dt + \int z(t)\, \theta_{j,\,g}(t)\, dt. \tag{8.5.23}$$

Однако

$$\int r(t)\, \theta_{j,\,g}(t)\, dt = \iint u_0(\tau)\, g(t, \tau)\, \theta_{j,g}(t)\, d\tau\, dt = \\ = \sqrt{\lambda_{j,\,g}} \int u_0(\tau)\, \varphi_{j,\,g}(\tau)\, d\tau. \tag{8.5.24}$$

Из (8.4.42) имеем

$$\int z(t)\, \theta_{j,\,g}(t)\, dt = z_{j,\,g} = \sqrt{\lambda_{j,\,g}} \int n(\tau)\, \varphi_{j,g}(\tau)\, d\tau. \tag{8.5.25}$$

Сочетая (8.5.24) и (8.5.25), имеем

$$\int y(t)\, \theta_{j,\,g}(t)\, dt = \sqrt{\lambda_{j,\,g}} \int v(\tau)\, \varphi_{j,\,g}(\tau)\, d\tau. \tag{8.5.26}$$

Подставляя это выражение в (8.5.22), получаем

$$v_i = \sum_{j=1}^{m} \frac{\alpha_{ij}}{\sqrt{\lambda_{j,g}}} \int y(t)\, \theta_{j,g}(t)\, dt + \int v(t) \sum_{j=m+1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t)\, dt. \tag{8.5.27}$$

Покажем теперь, что остаточный член в (8.5.27) стремится к нулю при $m \to \infty$. Имеем

$$\int v(t) \sum_{j=m+1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t)\, dt = \int u_0(t) \sum_{j=m+1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t)\, dt + \\ + \int n(t) \sum_{j=m+1}^{\infty} \alpha_{j,g} \varphi_{j,\,g}(t)\, dt. \tag{8.5.28}$$

Первое выражение в правой части (8.5.28) может быть оценено с помощью неравенства Шварца

$$\left[\int u_0(t) \sum_{j=m+1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t)\, dt\right]^2 \leqslant \left[\int u_0^2(t)\, dt\right] \int \left[\sum_{j=m+1}^{\infty} \alpha_{ij} \varphi_{j,\,g}(t)\right]^2 dt =$$

$$= \left[\int u_0^2(t)\,dt\right] \sum_{j=m+1}^{\infty} \alpha_{ij}^2. \qquad (8.5.29)$$

В силу ортонормальности функций множества $\{\theta_i(t)\}$ из (8.5.21) следует, что $\sum_{j=1}^{\infty} \alpha_{ij}^2 = 1$, и поэтому $\lim_{m\to\infty} \sum_{j=m+1}^{\infty} \alpha_{ij}^2 = 0$. Так как $u_0(t)$ имеет конечную энергию, то из (8.5.29) вытекает, что

$$\lim_{m\to\infty} \left[\int u_0(t) \sum_{j=m+1}^{\infty} \alpha_{ij}\, \varphi_{j,\,g}(t)\,dt\right]^2 = 0. \qquad (8.5.30)$$

Аналогично, последнее выражение в правой части (8.5.28) является гауссовской случайной величиной с нулевым средним и дисперсией

$$\sum_{j=m+1}^{\infty} \alpha_{ij}^2.$$

Таким образом, это выражение сходится в среднем квадратическом к нулю при $m \to \infty$ и

$$v_i = \lim_{m\to\infty} \sum_{j=1}^{m} \frac{\alpha_{ij}}{\sqrt{\lambda_{j,\,g}}} \int y(t)\,\theta_{j,\,g}(t)\,dt, \qquad (8.5.31)$$

что завершает доказательство. |

Заметим, что в (8.5.30) не утверждается существование

$$\lim_{m\to\infty} \sum_{j=1}^{\infty} \frac{\alpha_{ij}}{\sqrt{\lambda_{i,\,g}}}\, \theta_{j,\,g}(t)$$

и в действительности эта функция не существует в общем случае. Это означает, что в общем случае v_i не могут быть вычислены точно по $y(t)$ с помощью лишь корреляционной операции, хотя они могут быть аппроксимированы с помощью этой операции сколь угодно точно.

Как было показано, модель параллельного канала на рис. 8.5.7 эквивалентна модели канала на рис. 8.5.1. Теперь можно применить теорему 7.5.1 для определения максимума средней взаимной информации и теорему 7.5.2 для определения верхних границ минимальной достижимой вероятности ошибки при любой заданной длительности T на входе и интервала наблюдения T_0. Ограничение на энергию $\mathscr{E}$ в этих теоремах должно быть заменено на ST, а дисперсии шума σ_i^2 — на $1/\lambda_i$.

Далее исследуем поведение собственных значений λ_i в пределе, когда T и T_0 стремятся к ∞. Проведем это с помощью ряда лемм, сначала исследуя предельное поведение собственных значений μ_i фильтра $K(t, \tau)$ и затем устанавливая связь между множеством $\{\mu_i\}$ и собственными значениями λ_i фильтра $K_0(t, \tau)$. Читатель при первом чтении может опустить доказательства. Для заданного интервала T соб-

ственные значения μ_i (обозначаемые в дальнейшем, как $\mu_i(T)$) являются решениями интегрального уравнения [см. (8.4.10) и (8.4.56)]

$$\int_{-T/2}^{T/2} \mathscr{R}(\tau_1-\tau_2)\,\vartheta_i(\tau_2)\,d\tau_2 = \mu_i(T)\,\vartheta_i(\tau_1);\quad -T/2 \leqslant \tau_1 \leqslant T/2, \tag{8.5.32}$$

где

$$\mathscr{R}(\tau) = \int \frac{|H_1(f)|^2}{N(f)}\, e^{i2\pi f\tau}\, df. \tag{8.5.33}$$

Поведение множества собственных значений $\{\mu_i(T)\}$ в пределе при $T \to \infty$ было исследовано Кацем, Мардоком и Сеге (1953). Полученный ими результат формулируется в нижеследующей лемме и читатель, интересующийся доказательством, должен обратиться к книге Гренандера и Сеге (1958).

Лемма 8.5.2. Пусть $\mathscr{R}(\tau)$ — действительная функция, преобразование Фурье которой действительная интегрируемая ограниченная функция $F(f)$. Для любого $T > 0$ обозначим через $N_T(a, b)$ — число собственных значений уравнения (8.5.32), удовлетворяющих неравенствам $a \leqslant \mu_i(T) < b$. Тогда, если a и b одновременно положительны (или отрицательны), и если множество точек f, для которых $F(f) = a$ или $F(f) = b$ имеет меру нуль (т. е. если $F(f)$ не равна a или b на некотором ненулевом интервале), то

$$\lim_{T\to\infty} \frac{1}{T} N_T(a, b) = \int_{f:\, a \leqslant F(f) < b} df. \tag{8.5.34}$$

Заметим, что (8.5.34) представляет собой как раз тот результат, который нужен, когда собственные функции (8.5.32) аппроксимируют рассмотренное в § 8.3 множество синусов и косинусов, разделенных по частоте $1/T$.

Для приложений интересно не число собственных значений в данной области, а асимптотическое поведение суммы функций от собственных значений. Следующая лемма относится к этой задаче.

Лемма 8.5.3. Пусть $\mathscr{R}(\tau)$ — корреляционная функция с собственными значениями $\mu_i(T)$, определяемыми уравнением (8.5.32), и пусть $F(f) \geqslant 0$—преобразование Фурье $\mathscr{R}(\tau)$. Предположим, что $F(f)$ интегрируема и ограничена. Пусть $g(x)$ — неубывающая функция x, определенная для $x \geqslant 0$, и $g(0) = 0$. Пусть кроме того $g(x)$ — функция с ограниченным наклоном, удовлетворяющая для некоторого фиксированного числа B и всех $x_1 > 0$, $x_2 > 0$ неравенству $|g(x_1) - g(x_2)| \leqslant B\,|x_1 - x_2|$. Тогда

$$\lim_{T\to\infty} \frac{1}{T} \sum_i g(\mu_i(T)) = \int g[F(f)]\, df. \tag{8.5.35}$$

Доказательство. Так как $F(f)$ ограничена, то можно найти достаточно большое число A, такое, что

$$F(f) < A \text{ для всех } f. \tag{8.5.36}$$

Умножая (8.5.32) на $\varphi_i(\tau_1)$ и интегрируя обе части по τ_1, получаем

$$\mu_i(T) = \int\int \varphi_i(\tau_1)\,\varphi_i(\tau_2)\,\mathcal{R}(\tau_1-\tau_2)\,d\tau_1\,d\tau_2 =$$
$$= \int |\Phi_i(f)|^2 F(f)\,df < A\int |\Phi_i(f)|^2\,df = A.$$

Следовательно, при всех T все собственные значения меньше, чем A. Теперь пусть ε — произвольно малое положительное число. Поскольку $F(f)$ интегрируема, то f_1 можно выбрать достаточно большим так, что

$$\int\limits_{|f|\geqslant f_1} F(f)\,df \leqslant \varepsilon.$$

Пусть a_0 — положительное число, $a_0 \leqslant \varepsilon/f_1$. Тогда

$$\int\limits_{f:F(f)<a_0} F(f)\,df \leqslant \int\limits_{\substack{f:F(f)<a_0\\ |f|\leqslant f_1}} F(f)\,df + \int\limits_{f:|f|\geqslant f_1} F(f)\,df \leqslant 2f_1 a_0 + \varepsilon \leqslant 3\varepsilon. \tag{8.5.37}$$

Теперь пусть $a_1 < \ldots < a_n$ — множество чисел, для которого $a_1 > a_0$, $a_n = A$ и $a_j - a_{j-1} \leqslant \varepsilon a_0/B$; $1 \leqslant j \leqslant n$. Кроме того, пусть эти числа таковы, что множество f, для которых $F(f) = a_j$, $0 \leqslant j \leqslant n$, имеет меру нуль. Для любого $T > 0$ имеем

$$\sum_{j=1}^{n} g(a_{j-1})\,N_T(a_{j-1}, a_j) \leqslant \sum_{j=1}^{n} \sum_{i:a_{j-1}\leqslant\mu_i(T)<a_j} g[\mu_i(T)] \leqslant$$
$$\leqslant \sum_{j=1}^{n} g(a_j)\,N_T(a_{j-1}, a_j), \tag{8.5.38}$$

где $N_T(a_{j-1}, a_j)$ — число целых i, для которых $a_{j-1} \leqslant \mu_i(T) < a_j$. Заметим, что центральная часть выражения (8.5.38) равна как раз $\Sigma g[\mu_i(T)]$, где сумма берется по тем i, для которых $a_0 \leqslant \mu_i(T)$. Разделив все члены (8.5.38) на T, переходя к пределу при $T \to \infty$ и используя (8.5.34), получим

$$\sum_{j=1}^{n} [g(a_{j-1}) \int\limits_{f:a_{j-1}\leqslant F(f)<a_j} df] \leqslant \lim_{T\to\infty} \frac{1}{T} \sum_{i:a_0\leqslant\mu_i(T)} g[\mu_i(T)] \leqslant$$
$$\leqslant \sum_{j=1}^{n} [g(a_j) \int\limits_{f:a_{j-1}\leqslant F(f)<a_j} df]. \tag{8.5.39}$$

Заметим, что внешние части (8.5.39) также ограничивают $\int g[F(f)]df$, где интеграл берется по множеству f, для которого $a_0 \leqslant F(f)$. Следовательно, разность внешних частей ограничивает разность между центральной частью в (8.5.39) и этим интегралом. Таким образом,

$$\left|\lim_{T\to\infty} \frac{1}{T} \sum_{a_0\leqslant\mu_i(T)} g[\mu_i(T)] - \int\limits_{f:a_0\leqslant F(f)} g[F(f)]\,df\right| \leqslant$$

$$\leqslant \sum_{j=1}^{n} [g(a_j) - g(a_{j-1})] \int\limits_{f:a_{j-1} \leqslant F(f) < a_j} df \leqslant a_0 \varepsilon \sum_{j=1}^{n} \int\limits_{f:a_{j-1} \leqslant F(f) < a_j} df =$$

$$= a_0 \varepsilon \int\limits_{f:F(f) \geqslant a_0} df \leqslant a_0 \varepsilon \int \frac{F(f)}{a_0} df = \varepsilon \int F(f)\, df. \qquad (8.5.40)$$

Далее рассмотрим значения i, для которых $\mu_i(T) < a_0$. Заметим, что поскольку величины a_0 и A в (8.5.40) были выбраны независимо от функции g, то (8.5.40) также справедливо для частного случая $g(x) = x$, поэтому

$$\left| \lim_{T \to \infty} \frac{1}{T} \sum_{a_0 \leqslant \mu_i(T)} \mu_i(T) - \int\limits_{f:a_0 \leqslant F(f)} F(f)\, df \right| \leqslant \varepsilon \int F(f)\, df. \qquad (8.5.41)$$

Теперь пусть $K_1(t) = \int \sqrt{F(f)}\, e^{j2\pi ft}\, df$. Из (8.4.24) следует, что для любого T

$$\sum_i \mu_i(T) = \int_{-T/2}^{T/2} \left[\int_{-\infty}^{\infty} K_1^2(t-\tau)\, dt \right] d\tau = T \int_{-\infty}^{\infty} K_1^2(t)\, dt = T \int F(f)\, df, \qquad (8.5.42)$$

где в соотношении (8.5.42) использовано равенство Парсеваля для преобразований Фурье (8.1.23). Подставляя (8.5.42) в (8.5.41), получаем

$$\left| \lim_{T \to \infty} \frac{1}{T} \sum_{\mu_i(T) < a_0} \mu_i(T) - \int\limits_{f:F(f) < a_0} F(f)\, df \right| \leqslant \varepsilon \int F(f)\, df. \qquad (8.5.43)$$

Тогда из (8.5.37) имеем

$$\lim_{T \to \infty} \frac{1}{T} \sum_{\mu_i(T) < a_0} \mu_i(T) \leqslant \varepsilon \int F(f)\, df + 3\varepsilon. \qquad (8.5.44)$$

Так как $g[\mu_i(T)] \leqslant B\mu_i(T)$, то

$$\lim_{T \to \infty} \frac{1}{T} \sum_{\mu_i(T) < a_0} g[\mu_i(T)] \leqslant B\varepsilon \left[\int F(f)\, df + 3 \right] \qquad (8.5.45)$$

Аналогично

$$\int\limits_{f:F(f) < a_0} g[F(f)]\, df \leqslant 3\varepsilon B. \qquad (8.5.46)$$

Сочетая (8.5.45) и (8.5.46) с (8.5.40), получаем

$$\left| \lim_{T \to \infty} \frac{1}{T} \sum_i g[\mu_i(T)] - \int_{-\infty}^{\infty} g[F(f)]\, df \right| \leqslant \varepsilon [(B+1) \int F(f)\, df + 6B]. \qquad (8.5.47)$$

Так как $\varepsilon > 0$ произвольно, то этим завершается доказательство.

Цель следующих трех лемм показать, что при стремлении T и T_0 к ∞, результат (8.5.34) применим также к собственным значениям λ_i фильтра $K_0(t, \tau)$.

Лемма 8.5.4. Пусть T и T_0 — любые заданные положительные числа и пусть $\{\mu_i\}$ — множество собственных чисел для $K(t, \tau)$ и $\{\lambda_i\}$ — множество собственных чисел для $K_0(t, \tau)$, упорядоченных, как указано в теореме 8.4.1. Тогда для всех $i \geqslant 1$

$$\lambda_i \leqslant \mu_i. \tag{8.5.48}$$

Доказательство. Первые i входных собственных функций $\vartheta_1(\tau), \ldots, \vartheta_i(\tau)$, связанных с $K_0(t, \tau)$, линейно независимы и, следовательно, все они не могут быть линейными комбинациями первых $i-1$ собственных функций $\xi_1(\tau), \ldots, \xi_{i-1}(\tau)$, связанных с $K(t, \tau)$. Поэтому можно выбрать функцию

$$x(\tau) = \sum_{j=1}^{i} x_j \vartheta_j(\tau), \tag{8.5.49}$$

которая является линейной комбинацией $\vartheta_1(\tau), \ldots, \vartheta_i(\tau)$ и ортогональна к $\xi_1(\tau), \ldots, \xi_{i-1}(\tau)$. Тогда $x(\tau)$ может быть также представлена с помощью множества функций $\{\xi_i(\tau)\}$ в виде

$$x(\tau) = \sum_{j=i}^{\infty} y_j \xi_j(\tau); \quad y_j = \int \xi_j(\tau)\, x(\tau)\, d\tau. \tag{8.5.50}$$

Функцию $x(\tau)$ можно нормировать к единичной энергии, так что

$$\sum_{j=1}^{i} x_j^2 = 1 = \sum_{j=i}^{\infty} y_j^2.$$

Отклики фильтров $K_0(t, \tau)$ и $K(t, \tau)$ на $x(\tau)$ задаются соотношениями

$$u_0(t) = \int x(\tau) K_0(t, \tau)\, d\tau = \sum_{j=1}^{i} x_j \sqrt{\lambda_j}\, \theta_j(t), \tag{8.5.51}$$

$$u(t) = \int x(\tau) K(t, \tau)\, d\tau = \sum_{j=i}^{\infty} y_j \sqrt{\mu_j}\, \eta_j(t), \tag{8.5.52}$$

$$\int u_0^2(t)\, dt = \sum_{j=1}^{i} x_j^2 \lambda_j \geqslant \lambda_i \sum_{j=1}^{i} x_j^2 = \lambda_i, \tag{8.5.53}$$

$$\int u^2(t)\, dt = \sum_{j=i}^{\infty} y_j^2 \mu_j \leqslant \mu_i \sum_{j=i}^{\infty} y_j^2 = \mu_i. \tag{8.5.54}$$

Для K_0, задаваемого равенством (8.5.13), $u_0(t)$ является проекцией $u(t)$ на пространство, порожденное $\{\vartheta_{i,g}(t)\}$ и, следовательно, $\int u_0^2(t)\, dt \leqslant \int u^2(t)\, dt$. Для K_0, задаваемого равенством (8.5.14), $u_0(t)$ равна $u(t)$, усеченной вне интервала $(-T_0/2, T_0/2)$, и справедлив тот же самый результат. Сочетая этот результат с (8.5.53) и (8.5.54), имеем $\lambda_i \leqslant \mu_i$. |

Для простоты предположим теперь, чть выходной интервал T_0 равен входному интервалу T. Обозначая через $\lambda_i(T)$ и $\mu_i(T)$ собственные значения фильтров $K(t, \tau)$ и $K_0(t, \tau)$ соответственно, мы хотим по-

казать, что $(1/T)\ \Sigma\ \lambda_i(T)$ стремится к $(1/T)\Sigma\mu_i(T)$ в пределе при $T \to \infty$. Ключом к этому результату служит следующая лемма, которая является небольшим обобщением результата Келли, Рида и Рута*).

Л е м м а 8.5.5. Пусть для функции $g_1(t)$ из L_2

$$g_T(t, \tau) = \begin{cases} g_1(t-\tau); & t \leqslant T/2, \\ 0 \quad\ ; & t > T/2, \end{cases}$$

т. е. $g_T(t, \tau)$ представляет фильтр $g_1(t)$ с выходом на временном интервале $(-T/2, T/2)$. Пусть $\{\varphi_{i,T}(\tau)\}$ — входные собственные функции фильтра $g_T(t, \tau)$ в смысле теоремы 8.4.1. Пусть $T_1 > 0$ и τ_1 удовлетворяют неравенству

$$T \geqslant T_1 + 2|\tau_1|. \tag{8.5.55}$$

Тогда, если функция $v(\tau)$ ортогональна к $\vartheta_{i,T}(\tau + \tau_1)$ при всех i, то $v(\tau)$ также ортогональна к $\vartheta_{i,T_1}(\tau)$ при всех i. Кроме того, если $v(\tau)$ разлагается в виде

$$v(\tau) = \sum_i v_{i,T}\,\vartheta_{i,T}(\tau) + v_{r,T}(\tau); \quad v_{i,T} = (v, \vartheta_{i,T}) \tag{8.5.56}$$

и если преобразование Фурье $v(\tau)$ равно нулю всюду, где преобразование Фурье $g_1(t)$ равно нулю, то

$$\lim_{T\to\infty} v_{r,T}(\tau) = 0. \tag{8.5.57}$$

В сущности, лемма устанавливает, что когда T неограниченно возрастает, произвольную функцию можно представить все лучше и лучше с помощью множества функций $\{\vartheta_{i,T}\}$ или временных сдвигов $\vartheta_{i,T}$.

Доказательство. Если $v(\tau)$ ортогонально к $\vartheta_{i,T}(\tau + \tau_1)$ при всех i, то

$$\int v(\tau - \tau_1)\,\vartheta_{i,T}(\tau)d\tau = 0 \text{ при всех } i.$$

В силу (8.4.15) это означает, что

$$\int g_1(t-\tau)v(\tau-\tau_1)d\tau = 0; \quad |t| \leqslant T/2.$$

Подставляя $\tau_2 = \tau - \tau_1$ и $t_2 = t + \tau_1$, получаем

$$\int g_1(t_2-\tau_2)v(\tau_2)d\tau_2 = 0; \quad |t_2 - \tau_1| \leqslant T/2. \tag{8.5.58}$$

Из (8.5.55) вытекает, что (8.5.58) должно удовлетвориться для $|t_2| \leqslant T_1/2$, и, используя утверждения (8.4.15) в обратном направлении, находим, что $v(\tau)$ ортогональна к $\vartheta_{i,T_1}(\tau)$ при всех τ.

Далее покажем, что $v_{r,T}(\tau) \to 0$ при $T \to \infty$. Так как по определению $v_{r,T}(\tau)$ ортогональна к $\vartheta_{i,T}(\tau)$ при всех i, то, используя (8.5.56), получаем

$$\int v_{r,T}(\tau)\,v(\tau)\,d\tau = \int v_{r,T}^2(\tau)\,d\tau = \|v_{r,T}\|^2. \tag{8.5.59}$$

---

*) Kelly, Reed, Root. «The detection of Radar Echoes in Noise, I» Jour. Siam, 8. (2), June 1960 (см. приложение А).

Для $T > T_1$ также получаем, что $v_{r,T}(\tau)$ ортогональна к $\vartheta_{i,T_1}(\tau)$ при всех i. Используя T_1 вместо T в (8.5.56) и подставляя это выражение в (8.5.59), получаем

$$\int v_{r,T}(\tau)\, v_{r,T_1}(\tau)\, d\tau = \| v_{r,T} \|^2. \tag{8.5.60}$$

Применяя неравенство Шварца к левой части (8.5.60), видим, что

$$\| v_{r,T} \| \leqslant \| v_{r,T_1} \|; \quad T_1 \leqslant T.$$

Следовательно, $\| v_{r,T} \|$ убывает с T и должна стремиться к пределу. Наконец, используя (8.5.59) и (8.5.60), имеем

$$\int [v_{r,T_1}(\tau) - v_{r,T}(\tau)]^2\, d\tau = \| v_{r,T_1} \|^2 - \| v_{r,T} \|^2. \tag{8.5.61}$$

Так как $\| v_{r,T} \|$ стремится к пределу, то предел правой части (8.5.60) при T и T_1, стремящимся к ∞, равен 0. Поэтому $v_{r,T}(\tau)$ сходится к предельной функции *) $v_{r,\infty}(\tau)$ и эта предельная функция ортогональна к $\vartheta_{i,T}(\tau)$ при всех i и всех T. Однако, в силу (8.4.15), это означает, что

$$\int g_1(t-\tau)\, v_{r,\infty}(\tau)\, d\tau = 0 \quad \text{при всех } t. \tag{8.5.62}$$

Взяв преобразование Фурье от (8.5.62), находим

$$\sqrt{N(f)}\, V_{r,\infty}(f) = 0 \text{ при всех } f.$$

По предположению $V_{r,\infty}(f) = 0$, когда $\sqrt{N(f)}$ равно 0; следовательно, $V_{r,\infty}(f) = 0$ и соотношение (8.5.57) установлено. |

Л е м м а 8.5.6. Пусть $\mu_i(T)$ и $\lambda_i(T)$ — собственные значения соответственно фильтров $K_T(t, \tau)$ и $K_{0,T}(t, \tau)$, где $K_T(t, \tau)$ равно $K_1(t-\tau)$, усеченному вне $|\tau| \leqslant T/2$, и $K_{0,T}(t, \tau)$ задается равенством (8.5.13) или (8.5.14) при $T_0 = T$. Тогда

$$\lim_{T\to\infty} \frac{1}{T} \sum_i \lambda_i(T) = \lim_{T\to\infty} \frac{1}{T} \sum_i \mu_i(T). \tag{8.5.63}$$

Доказательство. Используя (8.4.24), имеем

$$\frac{1}{T} \sum_i \mu_i(T) = \frac{1}{T} \iint K_T^2(t, \tau)\, dt\, d\tau = \int K_1^2(t)\, dt. \tag{8.5.64}$$

Так как $\lambda_i(T) \leqslant \mu_i(T)$, то для доказательства леммы достаточно показать, что для каждого $\varepsilon > 0$

$$\lim_{T\to\infty} \frac{1}{T} \iint K_{0,T}^2(t, \tau)\, dt\, d\tau \geqslant \int K_1^2(t)\, dt - \varepsilon. \tag{8.5.65}$$

Для $K_{0,T}$, задаваемого (8.5.13), положим $K_1(\tau)$ равным $v(\tau)$ из предыдущей леммы, и из (8.5.13) видим, что $K_1(t) - K_{0,T}(t, 0)$ играет роль $v_{r,T}(\tau)$. Так как $H_1(f)/\sqrt{N(f)} = 0$, когда $\sqrt{N(f)} = 0$, то предыдущая

---

*) Это следует из теоремы Рисса — Фишера (см. Рисс и Надь (1955)).

лемма утверждает, что $K_{0,T}(t, 0)$ стремится к $K_1(t)$ при T, стремящемся к бесконечности, и для любого $\varepsilon > 0$ можно выбрать T_ε так, чтобы

$$\int K_{0,T}^2(t, 0)\, dt \geqslant \int K_1^2(t)\, dt - \varepsilon;\ T \geqslant T_\varepsilon. \tag{8.5.66}$$

Далее выпишем

$$K_{0,T}(t+\tau, \tau) = \sum_i \vartheta_{i,T}(t+\tau) \int \vartheta_{i,T}(t_1+\tau) K_1(t_1)\, dt_1. \tag{8.5.67}$$

Следовательно, $K_{0,T}(t+\tau, \tau)$ является проекцией $K_1(t)$ на множество функций $\vartheta_{i,T}(t+\tau)$. Остаток [т. е. часть $K_1(t)$, ортогональная ко всем $\vartheta_{i,T}(t+\tau)$] согласно предыдущей лемме, также ортогонален ко всем $\vartheta_{i,T_\varepsilon}(t)$, если $T \geqslant T_\varepsilon + 2|\tau|$. Таким образом, энергия остатка ограничена сверху энергией остатка разложения $K_1(t)$ по функциям множества $\{\vartheta_{i,T_\varepsilon}(t)\}$. Из (8.5.66) вытекает, что эта энергия ограничена сверху ε и

$$\int K_{0,T}^2(t+\tau, \tau)\, dt \geqslant \int K_1^2(t)\, dt - \varepsilon;\ T \geqslant T_\varepsilon + 2|\tau|, \tag{8.5.68}$$

$$\frac{1}{T} \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} K_{0,T}^2(t, \tau)\, dt\, d\tau \geqslant \frac{1}{T} \int\limits_{|\tau| \leqslant \frac{T-T_\varepsilon}{2}} \int K_{0,T}^2(t, \tau)\, dt\, d\tau. \tag{8.5.69}$$

Неравенство (8.5.68) справедливо в области значений τ, указанной в правой части (8.5.69). Теперь имеем

$$\frac{1}{T} \int\int K_{0,T}^2(t, \tau)\, dt\, d\tau \geqslant \left(\frac{T-T_\varepsilon}{T}\right) \left[\int K_1^2(t)\, dt - \varepsilon\right]. \tag{8.5.70}$$

Переходя к пределу в (8.5.70) при $T \to \infty$, получаем (8.5.65).

Если шум в канале белый, то $K_{0,T}(t, \tau)$ задается (8.5.14) и непосредственно видно, что (8.5.68) опять удовлетворяется для любого $\varepsilon > 0$ и достаточно большего T_ε. Таким образом, как и ранее, получаем (8.5.63). |

Теперь следующая лемма связывает полученные результаты вместе.

Лемма 8.5.7. Пусть для канала рис. 8.5.1 $T_0 = T$ и предположим, что $F(f) = |H_1(f)|^2/N(f)$ — ограниченная и интегрируемая функция. Пусть $g(x)$ — неубывающая функция ограниченного наклона, определенная при $x > 0$, и $g(0) = 0$. Тогда собственные числа $\lambda_i(T)$ фильтра $K_0(t, \tau)$ удовлетворяют соотношению

$$\lim_{T \to \infty} \frac{1}{T} \sum_i g[\lambda_i(T)] = \int g\left[\frac{|H_1(f)|^2}{N(f)}\right] df. \tag{8.5.71}$$

Доказательство. Пусть B — верхняя граница наклона g. Тогда, так как $\lambda_i(T) \leqslant \mu_i(T)$, то

$$0 \leqslant g[\mu_i(T)] - g[\lambda_i(T)] \leqslant B[\mu_i(T) - \lambda_i(T)],$$

$$0 \leqslant \frac{1}{T} \sum_i g\,[\mu_i(T)] - \frac{1}{T} \sum_i g\,[\lambda_i(T)] \leqslant$$
$$\leqslant \frac{B}{T} \left[\sum_i \mu_i(T) - \sum_i \lambda_i(T)\right].$$

Из леммы 8.5.6 следует, что

$$\lim_{T\to\infty} \left\{\frac{1}{T} \sum_i g\,[\mu_i(T)] - \frac{1}{T} \sum_i g\,[\lambda_i(T)]\right\} = 0.$$

Это равенство в сочетании с леммой 8.5.3 завершает доказательство. |

Теперь можно использовать эту лемму для нахождения пропускной способности канала рис. 8.5.1 при ограничении на мощность и найти экспоненциальные границы для вероятности ошибки. Для того чтобы определить пропускную способность канала при ограниченной мощности S на входе, определим сначала C_T как умноженный на $1/T$ максимум средней взаимной информации между $x(t)$ и $y(t)$, когда $x(t)$ и $y(t)$ рассматриваются на временном интервале $(-T/2, T/2)$ и максимизация проводится по всем распределениям вероятностей на входе при условии, что математическое ожидание $\int x^2(t)dt$ не больше, чем ST. Фактически C_T уже было найдено на основе изображенного на рис. 8.5.7 представления в виде параллельных каналов и на основе теоремы 7.5.1. Пропускная способность определяется по формуле

$$C = \lim_{T\to\infty} C_T.$$

Существование этого предела доказывается в следующей теореме.

Теорема 8.5.1. Предположим, что для канала на рис. 8.5.1 с ограничением на мощность S и с $T_0 = T$ функция $|H_1(f)|^2/N(f)$ ограничена и интегрируема и что или $\int N(f)df < \infty$ или $N(f)$ — плотность белого шума. Тогда пропускная способность канала C задается параметрически равенствами

$$C = \int_{f:\,\frac{N(f)}{|H_1(f)|^2} \leqslant B} \frac{1}{2} \log\left[\frac{|H_1(f)|^2 B}{N(f)}\right] df, \tag{8.5.72}$$

$$S = \int_{f:\,\frac{N(f)}{|H_1(f)|^2} \leqslant B} \left[B - \frac{N(f)}{|H_1(f)|^2}\right] df. \tag{8.5.73}$$

Доказательство. Из теоремы 7.5.1 следует, что для множества параллельных каналов на рис. 8.5.7 максимум средней взаимной информации на единицу времени связан с ограничением на мощность S параметрическими равенствами

$$\tilde{C}_T(B) = \frac{1}{T} \sum_{i:\lambda_i(T) \geqslant (1/B)} \frac{1}{2} \log\left[\lambda_i(T)\, B\right], \tag{8.5.74}$$

$$\widetilde{S}_T(B) = \frac{1}{T} \sum_{i:\lambda_i(T) \geqslant (1/B)} \left[B - \frac{1}{\lambda_i(T)} \right], \tag{8.5.75}$$

т. е. при ограничении на мощность S находится значение B, которое удовлетворяет равенству $S = \widetilde{S}_T(B)$, и для этого значения B имеем $C_T = \widetilde{C}_T(B)$. Если для заданного B определить функцию

$$g(x) = \begin{cases} 0 & ; \ x < 1/B, \\ 1/2 \ \log(xB); & x \geqslant 1/B, \end{cases} \tag{8.5.76}$$

то (8.5.74) можно переписать в виде

$$\widetilde{C}_T(B) = \frac{1}{T} \sum_i g\,[\lambda_i(T)]. \tag{8.5.77}$$

Отсюда, используя лемму 8.5.7, имеем

$$\widetilde{C}_\infty(B) = \lim_{T \to \infty} \widetilde{C}_T(B) = \int g\left[\frac{|H_1(f)|^2}{N(f)} \right] df = \\ = \int\limits_{f:\ \frac{|H_1(f)^2|}{N(f)} \geqslant \frac{1}{B}} 1/2 \log\left[\frac{|H_1(f)|^2}{N(f)} B \right] df. \tag{8.5.78}$$

Переопределив функцию $g(x)$, полагая ее равной 0 для $x \leqslant 1/B$ и равной $B - 1/x$ для $x \geqslant 1/B$, ту же самую лемму можно применить к (8.5.73). Таким образом,

$$\widetilde{C}_\infty(B) = \lim_{T \to \infty} \widetilde{S}_T(B) = \int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B}} \left[B - \frac{N(f)}{|H_1(f)|^2} \right] df. \tag{8.5.79}$$

Пусть при заданном ограничении на мощность S значение B удовлетворяет равенству $S = \widetilde{S}_\infty(B)$. Пусть $\varepsilon > 0$ — произвольно мало. Так как $\widetilde{C}_\infty(B)$ — непрерывная функция, то существует $\delta > 0$, такое, что $\widetilde{C}_\infty(B + \delta) \leqslant \widetilde{C}_\infty(B) + \varepsilon$. Для этого δ существует некоторое T_1, такое, что для $T \geqslant T_1$

$$\widetilde{C}_T(B + \delta) \leqslant \widetilde{C}_\infty(B + \delta) + \varepsilon \leqslant \widetilde{C}_\infty(B) + 2\varepsilon. \tag{8.5.80}$$

Вместе с тем, так как $\widetilde{S}_\infty(B)$ строго монотонно возрастает с B (для $B > \inf N(f)/|H_1(f)|^2$), то имеем $\widetilde{S}_\infty(B + \delta) > \widetilde{S}_\infty(B)$ и существует T_2 такое, что для $T \geqslant T_2$

$$\widetilde{S}_T(B + \delta) \geqslant \widetilde{S}_\infty(B) = S. \tag{8.5.81}$$

Наконец, для любого T, C_T — монотонно возрастающая функция мощностного ограничения S, так что, используя (8.5.81) при $T \geqslant T_2$, имеем $C_T \leqslant \widetilde{C}_T(B + \delta)$. Следовательно, при $T \geqslant T_1$, $T \geqslant T_2$ имеем $C_T \leqslant \widetilde{C}_\infty(B) + 2\varepsilon$. Обращая неравенства, участвующие в приведен-

ных выше рассуждениях, имеем также $C_T \geqslant \widetilde{C}_\infty(B) - 2\varepsilon$ для всех достаточно больших T. Так как ε произвольно, то отсюда следует, что

$$\lim_{T\to\infty} C_T = \widetilde{C}_\infty(B),$$

что завершает доказательство. |

Обращение теоремы кодирования применимо здесь, так же как и для дискретных каналов. Точнее, имеет место следующая теорема. Ее доказательство опускается, так как оно, в сущности, повторяет доказательства, приведенные в гл. 4, 6 и 7.

Теорема 8.5.2. Пусть дискретный стационарный источник с алфавитом объема M имеет энтропию $H_\infty(U)$ и производит одну букву каждые τ_s секунд. Пусть последовательность букв источника произвольной длины L связана с адресатом посредством непрерывного по времени канала, используемого $T = L\tau_s$ секунд. Пусть C_T — умноженная на $1/T$ верхняя грань средней взаимной информации между входом и выходом канала на этом интервале, взятая по всем распределениям вероятностей на входе. Предположим, что $\lim_{T\to\infty} C_T$ существует и определим

$$C = \lim_{T\to\infty} C_T.$$

Тогда для любого $\varepsilon > 0$ и для всех достаточно больших L вероятность ошибки на символ $\langle P_e \rangle$ в последовательности из L букв источника удовлетворяет неравенству

$$\langle P_e \rangle \log(M-1) + \mathscr{H}(\langle P_e \rangle) \geqslant H_\infty(U) - \tau_s C - \varepsilon.$$

Можно применить те же соображения к границе случайного кодирования и границе для процедуры с выбрасыванием теоремы 7.5.2. Определим

$$\widetilde{S}_T(B,\rho) = \frac{1}{T} \sum_{i:\lambda_i(T)\geqslant\frac{1}{B}} \frac{(1+\rho)^2\,[B\lambda_i(T)-1]\,B}{(1+\rho)\,B\lambda_i(T)-\rho}, \tag{8.5.82}$$

$$\widetilde{R}_T(B) = \frac{1}{T} \sum_{i:\lambda_i(T)\geqslant\frac{1}{B}} \frac{1}{2}\ln[B\lambda_i(T)], \tag{8.5.83}$$

$$\widetilde{E}_T(B,\rho) = \frac{\rho\widetilde{S}_T(B,\rho)}{2B(1+\rho)} - \frac{1}{T} \sum_{i:\lambda_i(T)\geqslant\frac{1}{B}} \frac{1}{2}\ln\left[1+\rho-\frac{\rho}{B\lambda_i(T)}\right]. \tag{8.5.84}$$

Из теоремы 7.5.2 следует, что для любого выбранного ρ, $0 < \rho \leqslant 1$, и любого $B > 0$ существует код с $M = \lceil \exp\,[T\,\widetilde{R}_T(B)] \rceil$ кодовыми словами, каждое из которых ограничено временным интервалом

$(-T/2,\ T/2)$, имеет энергию не более $T\widetilde{S}_T(B, \rho)$ и вероятность ошибки, ограниченную неравенством

$$P_{e.m} \leqslant \left[\frac{2e^{s\delta}}{\mu}\right]^2 \exp\left[-T\widetilde{E}_T(B, \rho)\right]. \tag{8.5.85}$$

Для фиксированных B и ρ можно применить лемму 8.5.7 к (8.5.82), (8.5.83) и (8.5.84) точно так же, как в доказательстве теоремы 8.5.1, и получить

$$\widetilde{S}_\infty(B, \rho) = \int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B}} \frac{(1+\rho)^2 [B|H_1(f)|^2 - N(f)]B}{(1+\rho)B|H_1(f)|^2 - \rho N(f)}\, df, \tag{8.5.86}$$

$$\widetilde{R}_\infty(B) = \int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B}} \frac{1}{2} \ln\left[B\frac{|H_1(f)|^2}{N(f)}\right] df, \tag{8.5.87}$$

$$\widetilde{E}_\infty(B, \rho) = \frac{\rho\,\widetilde{S}_\infty(B, \rho)}{2B(1+\rho)} - \int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B}} \frac{1}{2} \ln\left[1+\rho - \frac{\rho N(f)}{B|H_1(f)|^2}\right] df. \tag{8.5.88}$$

С помощью того же типа (ε, δ)-рассуждений, как и в предыдущей теореме, найдем, что для любого $\varepsilon > 0$ существует T_1, такое, что для всех $T \geqslant T_1$ существуют коды с $M = \lceil \exp[T\widetilde{R}_\infty(B)] \rceil$ кодовыми словами, каждое из которых ограничено во времени интервалом $(-T/2,\ T/2)$, имеет энергию, не большую чем $T\widetilde{S}_\infty(B, \rho)$, и вероятность ошибки, ограниченную неравенством

$$P_{e,\,m} \leqslant \left[\frac{2\,e^{s\delta}}{\mu}\right]^2 \exp\{-T[\widetilde{E}_\infty(B, \rho) - \varepsilon]\}. \tag{8.5.89}$$

Как показано в (7.5.39), коэффициент задается приближенным равенством

$$\frac{2\,e^{s\delta}}{\mu} \approx \frac{e\rho\sqrt{4\pi\sum\limits_i \mathscr{E}_i^2}}{(1+\rho)^2 B}, \tag{8.5.90}$$

где

$$\mathscr{E}_i = \begin{cases} \dfrac{(1+\rho)^2[B\lambda_i(T) - 1]B}{(1+\rho)B\lambda_i(T) - \rho}; & \lambda_i(T) \geqslant 1/B, \\ 0 \quad ; & \lambda_i(T) < 1/B. \end{cases} \tag{8.5.91}$$

Из (8.5.91) легко заметить, что $\mathscr{E}_i \leqslant (1+\rho)B$ и, следовательно,

$$\sum_i \mathscr{E}_i^2 \leqslant (1+\rho)B\sum_i \mathscr{E}_i = (1+\rho)B\widetilde{S}_T(B, \rho)T. \tag{8.5.92}$$

Из (8.5.92) видно, что при фиксированных B и ρ для больших T приближенное выражение в (8.5.90) пропорционально $\sqrt{T}$, а также видно,

что приближение становится лучше с возрастанием T. Отсюда следует, что для достаточно больших T коэффициент в (8.5.89) может быть включен в ε. Таким образом, для любого ε существует T_2 (зависящее от ε, B и ρ), такое, что для $T \geqslant T_2$

$$P_{e,\,m} \leqslant \exp\{-T\,[\widetilde{E}_\infty(B,\,\rho) - \varepsilon]\}. \tag{8.5.93}$$

Так как значения ρ ограничены интервалом $0 < \rho \leqslant 1$, то, так же как в § 7.5, эта граница вероятности ошибки справедлива только в некотором диапазоне скоростей $[R = \widetilde{R}_\infty(B)]$ и мощностей $[S = \widetilde{S}_\infty(B,\,\rho)]$. Поскольку функция $\widetilde{S}_\infty(B,\,\rho)$ строго возрастающая и непрерывная по B и ρ (для $B \geqslant \inf N(f)/|H_1(f)|^2$), то при фиксированном S уравнение $S = \widetilde{S}_\infty(B,\,\rho)$ определяет B как функцию ρ и это неявно определяет $\widetilde{R}_\infty(B)$ как функцию ρ. При $\rho = 0$ значение $\widetilde{R}_\infty(B)$ равно пропускной способности при заданном S, что можно увидеть, сравнивая (8.5.86) и (8.5.87) с (8.5.72) и (8.5.73). Аналогично при $\rho = 0$ имеем $\widetilde{E}_\infty(B,\,\rho) = 0$. С возрастанием ρ при фиксированном S (а следовательно, при изменении B) $\widetilde{R}_\infty(B)$ убывает, а $\widetilde{E}_\infty(B,\,\rho)$ возрастает. Как и раньше, наклон E как функция R при фиксированном S равен $-\rho$. Когда ρ возрастает до 1, B убывает до критического значения B_{cr}, задаваемого формулой

$$S = \widetilde{S}_\infty(B_{cr},\,1) = \int\limits_{f\,:\,\frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B_{cr}}} \frac{4B_{cr}\,[B_{cr}\,|H_1(f)|^2 - N(f)]}{2B_{cr}\,|H_1(f)|^2 - N(f)}\,df. \tag{8.5.94}$$

Соответствующее B_{cr} значение критической скорости задается формулой

$$R_{cr} = \widetilde{R}_\infty(B_{cr}) = \int\limits_{f\,:\,\frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B_{cr}}} {}^1\!/_2 \ln\left[B_{cr}\,\frac{|H_1(f)|^2}{N(f)}\right] df. \tag{8.5.95}$$

Следовательно, при фиксированном S граница вероятности ошибки в (8.5.93) справедлива для скоростей из интервала $R_{cr} \leqslant R < C$.

Для скоростей $R < R_{cr}$ и любого заданного T граница вероятности ошибки задается (7.5.60). Как и выше, переходя к пределу, находим, что для любого $\varepsilon > 0$ существует достаточно большое T_1, такое, что при всех $T \geqslant T_1$ существует код с $M = \lceil \exp(RT) \rceil$ кодовыми словами, каждое из которых ограничено временным интервалом $(-T/2,\,T/2)$, имеет энергию, не большую ST, и

$$P_{em} \leqslant \exp\{-T\,[\widetilde{E}_\infty(B_{cr},\,1) + R_{cr} - R - \varepsilon]\}. \tag{8.5.96}$$

Наконец, для границы при процедуре с выбрасыванием, определим величины

$$\widetilde{S}_{x,\,T}(B,\,\rho) = \frac{1}{T} \sum_{i\,:\,\lambda_i(T) \geqslant \frac{1}{B}} \frac{4\rho B\,[B\lambda_i(T) - 1]}{2B\lambda_i(T) - 1}, \tag{8.5.97}$$

$$\widetilde{R}'_{x,T}(B)=\frac{1}{T}\sum_{i:\,\lambda_i(T)>\frac{1}{B}}{}^{1}/_{2}\ln\frac{B^2\lambda_i^2(T)}{2B\lambda_i(T)-1}, \tag{8.5.98}$$

$$\widetilde{E}_{x,T}(B,\rho)=\frac{\widetilde{S}_{x,T}(B,\rho)}{4B}. \tag{8.5.99}$$

Из теоремы 7.5.2 следует, что для любого $\rho>1$ и любого $B>0$ существует код с $M=\lceil\exp[T\widetilde{R}'_{x,T}(B)-2\ln(2e^{s\delta}/\mu)]\rceil$ кодовыми словами, каждое из которых ограничено во времени интервалом $(-T/2, T/2)$, имеет энергию, не большую $T\widetilde{S}_x(B,\rho)$, и вероятность ошибки, удовлетворяющую неравенству

$$P_{e,m}\leqslant\exp[-T\widetilde{E}_x(B,\rho)]. \tag{8.5.100}$$

Для фиксированных B и ρ можно применить лемму 8.5.7 к (8.5.97) и затем к (8.5.99), чтобы получить

$$\widetilde{S}_{x,\infty}(B,\rho)=\lim_{T\to\infty}\widetilde{S}_{x,T}(B,\rho)=$$

$$=\int\limits_{f:\,\frac{|H_1(f)|^2}{N(f)}\geqslant\frac{1}{B}}\frac{4\rho B[B|H_1(f)|^2-N(f)]}{2B|H_1(f)|^2-N(f)}\,df, \tag{8.5.101}$$

$$\widetilde{R}'_{x,\infty}(B)=\lim_{T\to\infty}\widetilde{R}'_{x,T}(B)=$$

$$=\int\limits_{f:\,\frac{|H_1(f)|^2}{N(f)}\geqslant\frac{1}{B}}\frac{1}{2}\ln\frac{B^2|H_1(f)|^4/N(f)}{2B|H_1(f)|^2-N(f)}\,df, \tag{8.5.102}$$

$$\widetilde{E}_{x,\infty}(B,\rho)=\lim_{T\to\infty}\widetilde{E}_{x,T}(B,\rho)=\frac{\widetilde{S}_{x,\infty}(B,\rho)}{4B}. \tag{8.5.103}$$

Если определить

$$\widetilde{R}_{x,T}(B,\rho)=\widetilde{R}'_{x,T}(B)-\frac{2}{T}\ln\frac{2e^{s\delta}}{\mu}, \tag{8.5.104}$$

то из тех же соображений, которые были использованы при переходе от (8.5.90) к (8.5.92), ясно, что

$$\widetilde{R}_{x,\infty}(B,\rho)=\lim_{T\to\infty}\widetilde{R}_{x,T}(B,\rho)=\widetilde{R}'_{x,\infty}(B). \tag{8.5.105}$$

Наконец, применяя (ε,δ)-технику, использованную в теореме 8.5.1, получаем, что для любого $B>0$, любого $\rho>1$ и любого произвольно малого $\varepsilon>0$ найдется T_1, такое, что при $T\geqslant T_1$ существует код с $M=\lceil\exp[T\widetilde{R}_{x,\infty}(B,\rho)]\rceil$ кодовыми словами, каждое из которых ограничено во времени интервалом $(-T/2, T/2)$, имеет энергию, не большую $\widetilde{S}_{x,\infty}(B,\rho)$, и вероятность ошибки, удовлетворяющую неравенству

$$P_{e,m}\leqslant\exp\{-T[\widetilde{E}_{x,\infty}(B,\rho)-\varepsilon]\}. \tag{8.5.106}$$

Для фиксированной мощности $S = \tilde{S}_{x,\infty}(B, \rho)$ при возрастании ρ от 1 значение B убывает от B_{cr} [определенного равенством (8.5.94)] и стремится к

$$\min_f \frac{N(f)}{|H_1(f)|^2}$$

при $\rho \to \infty$. Соответственно, когда ρ возрастает от 1, значение $\tilde{R}_{x,\infty}(B, \rho)$ убывает от $R_{x,cr}$ до 0, где

$$R_{x,cr} = \int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)} > \frac{1}{B_{cr}}} {}^1/_2 \ln \frac{B_{cr}^2 |H_1(f)|^4 / N(f)}{2B_{cr}|H_1(f)|^2 - N(f)}\, df. \qquad (8.5.107)$$

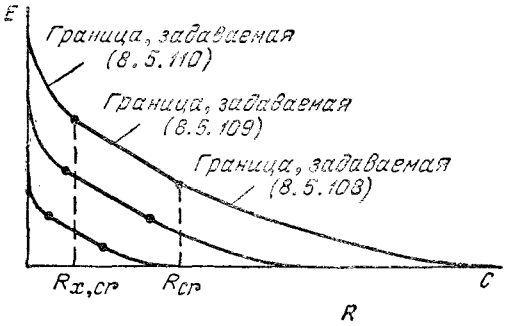


Рис. 8.5.8. Кривые показателя экспоненты как функции скорости при фиксированной мощности S (различные кривые соответствуют различным S).

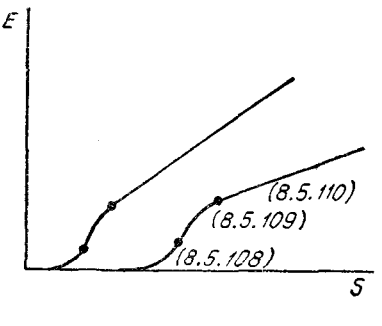


Рис. 8.5.9. Кривые показателя экспоненты как функции мощности при фиксированной скорости R (кривая, расположенная выше, соответствует меньшей скорости).

Аналогично при $\rho \to \infty$ и $R \to 0$ значение $\tilde{E}_{x,\infty}(B, \rho)$ возрастает, стремясь к

$$\frac{S}{4} \max_f \frac{|H_1(f)|^2}{N(f)}.$$

На рис. 8.5.8 и 8.5.9 изображено поведение этих показателей экспонент, как функций скорости R и мощности S. Полученные результаты суммируются в следующей теореме.

Теорема 8.5.3. Предположим, что для канала, изображенного на рис. 8.5.1 с $T_0 = T$, выполнены те же условия, что и в теореме 8.5.1. Тогда для любого $B > 0$, любого ρ, $0 \leqslant \rho \leqslant 1$, и любого произвольно малого $\varepsilon > 0$ найдется $T_1(\varepsilon, B, \rho)$, такое, что для любого $T \geqslant T_1(\varepsilon, B, \rho)$ существует код с $M = \lceil \exp[T\tilde{R}_\infty(B, \rho)] \rceil$ кодовыми словами, каждое из которых ограничено во времени интервалом $(-T/2, T/2)$, имеет энергию, не большую $T\tilde{S}_\infty(B, \rho)$, и вероятность ошибки, удовлетворяющую неравенству

$$P_{e,m} \leqslant \exp\{-T[\tilde{E}_\infty(B, \rho) - \varepsilon]\}. \qquad (8.5.108)$$

Для фиксированного $S = \tilde{S}_\infty(B,\rho)$ функция $\tilde{R}_\infty(B)$ строго и непрерывно убывает от C до R_{cr} при возрастании ρ от 0 до 1, и $E_\infty(B, \rho)$ строго и непрерывно возрастает при возрастании ρ от 0 до 1 (т. е. $\tilde{E}_\infty(B, \rho) > 0$ для всех $\tilde{R}_\infty(B) < C$). При фиксированном S и $T \geqslant T_1(\varepsilon, B_{cr}, 1)$ и любой $R \leqslant R_{cr}$ существует код с $M = \lceil \exp(TR) \rceil$ кодовыми словами, каждое из которых ограничено во времени интервалом $(-T/2, T/2)$, имеет энергию, не большую TS, и вероятность ошибки, удовлетворяющую неравенству

$$P_{e,m} \leqslant \exp\{-T[\tilde{E}_\infty(B_{cr}, 1) + R_{cr} - R - \varepsilon]\}. \qquad (8.5.109)$$

Наконец, для любого $B > 0$, любого $\rho > 1$ и любого $\varepsilon > 0$ найдется $T_1(\varepsilon, B, \rho)$, такое, что для $T \geqslant T_1(\varepsilon, B, \rho)$ существует код с $M = \lceil \exp[T\tilde{R}_{x,\infty}(B, \rho)] \rceil$ кодовыми словами, каждое из которых ограничено во времени интервалом $(-T/2, T/2)$, имеет энергию, не большую $T\tilde{S}_{x,\infty}(B, \rho)$, и вероятность ошибки, удовлетворяющую неравенству

$$P_{e,m} \leqslant \exp\{-T[\tilde{E}_{x,\infty}(B, \rho) - \varepsilon]\}. \qquad (8.5.110)$$

При фиксированном $S = \tilde{S}_{x,\infty}(B, \rho)$ значение $\tilde{R}_{x,\infty}(B, \rho)$ убывает от $R_{x,cr}$ до 0 с возрастанием ρ, а $\tilde{E}_{x,\infty}(B, \rho)$ возрастает с возрастанием ρ. Для фиксированного S показатель экспоненты как функция скорости, определяемая этими тремя границами (вторая граница применяется для скоростей $R_{x,cr} < R < R_{cr}$), непрерывен и имеет непрерывные производные.

8.6. ДИСПЕРГИРУЮЩИЕ КАНАЛЫ С ЗАМИРАНИЯМИ

В предыдущих параграфах изучались модели каналов, в которых принятый сигнал был суммой переданного сигнала (быть может профильтрованного и ослабленного известным образом) и гауссова шума. Такая модель обычно подходит для космических каналов связи и, быть может менее точно, для проводных и кабельных каналов связи. Однако во многих системах связи путь, по которому проходит сигнал, изменяется со временем и может быть заранее предсказан только статистически. Изменение пути приводит к изменениям энергии принятого сигнала во времени (которые называются замираниями), а также к дисперсионным изменениям принятого сигнала во времени. Эти эффекты особенно характерны для систем связи, использующих тропосферное рассеивание, орбитальное дипольное рассеивание и высокочастотную радиосвязь. В последующем изложении сначала будет построена математическая модель связи по таким каналам и затем будет доказана теорема кодирования для этой модели. Построение модели будет проведено без детального обоснования и заинтересованному читателю следует обратиться к книге Кеннеди (1969), в которой проведено подробное рассмотрение таких моделей.

Наиболее просто поведение таких систем можно представить себе, рассматривая рассеяние электромагнитных волн облаком рассеивающих частиц, как показано на рис. 8.6.1. Принятый сигнал (отделенный от всякого аддитивного шума) может рассматриваться как взвешенная сумма задержанных во времени переданных сигналов, где каждый задержанный сигнал соответствует рассеиванию от одного из слоев, показанных на рис. 8.6. 1. В пределах любого слоя рассеивающие частицы будут типично для этого слоя двигаться и вращаться так, что каждая рассеивающая частица будет вносить некоторый допплеровский сдвиг в принятый сигнал. Следовательно, если косинусоида $\cos 2\pi f_c t$ была передана, то сигнал, приходящий от какого-либо слоя, будет размазан по частоте около f_c. *Функция рассеивания* $\sigma(\tau, f)$ для

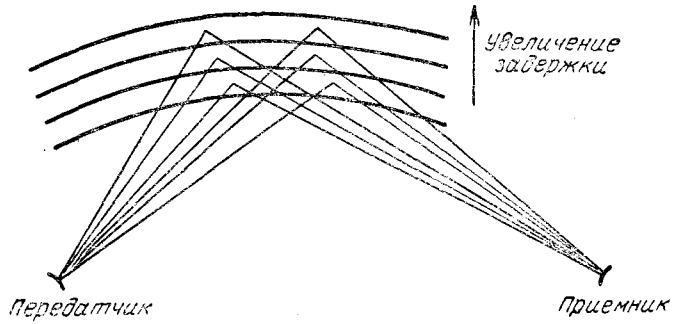


Рис. 8.6.1. Диспергирующий канал с замираниями.

такого канала определяется (с точностью до нормирующего множителя) как средняя по времени мощность, принятая в частотном интервале, расположенном около частоты $f_c + f$, и относящаяся к слою, вносящему задержку τ. Функция рассеивания обычно нормируется так, чтобы $\int\int \sigma(\tau, f) d\tau df = 1$. Здесь молчаливо принимается, что $\sigma(\tau, f)$ не зависит от f_c; обычно это предположение достаточно хорошо соблюдается для широкого диапазона f_c.

Если имеется большое число рассеивающих слоев, двигающихся более или менее случайно относительно друг друга, и если можно пренебречь многократным рассеиванием от одной частицы к другой, то можно рассматривать принятую функцию, порожденную фиксированным входом, как сумму весьма большого числа более или менее независимых функций, дающих малые приращения. Разумно поэтому предположить, что для каждого фиксированного входа выход является гауссовским случайным процессом. Если при этом предположении передано $A\cos 2\pi f_c t$, то принятая функция (в отсутствии аддитивного шума) может быть представлена в виде

$$r(t) = v_1(t) \cos 2\pi f_c t + v_2(t) \sin 2\pi f_c t, \qquad (8.6.1)$$

где $v_1(t)$ и $v_2(t)$ — гауссовские случайные процессы. Если предположить далее, что среда статистически стационарна и что фаза сигнала, приходящего от каждой отражающей частицы, равномерно распре-

делена между 0 и 2π, то можно показать, что $v_1(t)$ и $v_2(t)$ имеют одну и ту же спектральную плотность. Кроме того, если функция рассеивания удовлетворяет условию $\sigma(\tau, f) = \sigma(\tau, -f)$, то $v_1(t)$ и $v_2(t)$ статистически независимы. Наконец (предполагая, что f_c много больше, чем любой допплеровский сдвиг), приемник может наблюдать $v_1(t)$ и $v_2(t)$ раздельно.

Пусть $S\sigma(f)$ — спектральная плотность $v_1(t)$ (и $v_2(t)$) и $\sigma(f)$ нормирована, $\int \sigma(f)dt = 1$. Можно заметить, что S — средняя мощность принятой функции $r(t)$ и что $\sigma(f) = \int \sigma(\tau, f)\, d\tau$, где $\sigma(\tau, f)$ — функция рассеивания, определенная ранее. Средняя мощность S зависит от канала и, конечно, она также прямо пропорциональна мощности передатчика $A^2/2$.

Исследуем теперь вероятность ошибки, которую можно достичь при кодировании в указанном выше канале. Для того чтобы сделать анализ по возможности более простым, рассмотрим случай, когда кодовые слова представляют собой множество разделенных по частоте синусоид на фиксированном интервале времени $(-T/2, T/2)$, т. е.

$$x_m(t) = \begin{cases} A\cos 2\pi (f_c + \Delta m)t; & -T/2 \leqslant t \leqslant T/2, \\ 0 \qquad\qquad\quad ; & |t| > T/2. \end{cases} \tag{8.6.2}$$

Предположим, что Δ выбрано достаточно большим, так что $\sigma(f) = 0$ для $|f| \geqslant \Delta/2$. Будем считать, что в приемнике имеется набор параллельно соединенных фильтров с единичным усилением и полосой частот каждого фильтра Δ и m-й фильтр настроен на частоту $f_c + \Delta m$. Если теперь рассмотреть принятую функцию как сумму сигнала и белого гауссова шума со спектральной плотностью $N_0/2$, то выход каждого фильтра на посланное сообщение m можно представить в виде

$$y_m(t) = [v_{1,m}(t) + n_{1,m}(t)]\cos 2\pi (f_c + \Delta m)\, t + \\ + [v_{2,m}(t) + n_{2,m}(t)]\sin 2\pi (f_c + \Delta m)\, t, \tag{8.6.3}$$

$$y_{m'}(t) = n_{1,m'}(t)\cos 2\pi (f_c + \Delta m')\, t + \\ + n_{2,m'}(t)\sin 2\pi (f_c + \Delta m')\, t \quad \text{для всех} \quad m' \neq m. \tag{8.6.4}$$

В этих выражениях $n_{1;m}(t)$, $n_{2,m}(t)$, $n_{1,m'}(t)$ и $n_{2,m'}(t)$ — независимые стационарные гауссовские процессы с нулевыми средними, имеющие спектральную плотность N_0 при $|f| \leqslant \Delta/2$ и 0 при $|f| > \Delta/2$. Так как нас интересует главным образом нахождение верхней границы вероятности ошибки для системы и так как шум при $|f| > \Delta/2$ всегда несуществен, то модель можно упростить, предположив, что все указанные выше шумы являются белыми со спектральной плотностью N_0.

Процессы $v_{1,m}(t)$ и $v_{2,m}(t)$ из (8.6.3) не стационарны, так как вход $s_m(t)$ ограничен во времени интервалом $(-T/2, T/2)$. Вместе с тем, если ввести параметр L, как разброс в задержке, вызванной как средой, так и фильтром приемника, и если подходящим образом сдвинуть начало отсчета времени приемника, то можно утверждать, что выход на интервале $[-(T-L)/2, (T-L)/2]$ не будет зависеть от того, является ли вход усеченным вне интервала $(-T, T)$ или нет. Другими словами, $v_{1,m}(t)$ и $v_{2,m}(t)$ на интервале $[-(T-L)/2, (T-L)/2]$ можно рас-

сматривать как выборочные функции независимых стационарных случайных гауссовских процессов со спектральными плотностями $S\sigma(f)$.

До сих пор наши рассуждения были эвристическими и приближенными, что было естественно, так как мы имели дело с классом недостаточно точно определенных физических каналов. Теперь, однако, имеется математическая модель, с которой можно работать, и начиная с этого места рассуждения будут точными. Вновь дадим краткое описание модели; имеем множество M кодовых слов длины T, задаваемых (8.6.2). Приемник наблюдает функции $y_{1,m}(t)$ и $y_{2,m}(t)$ для $1 \leqslant m \leqslant M$, где если передано сообщение m, то

$$y_{i,m}(t) = v_{i,m}(t) + n_{i,m}(t); \quad i = 1,\ 2, \tag{8.6.5}$$

а для всех $m' \neq m$

$$y_{1,m'}(t) = n_{i,m'}(t); \quad i = 1,\ 2. \tag{8.6.6}$$

Будем считать, что все функции $y_{i,m}(t)$ при $i = 1, 2$ и $1 \leqslant m \leqslant M$ наблюдаются только в интервале $[-(T - L)/2, (T - L)/2]$. В этом интервале все функции $v_{i,m}(t)$, $n_{i,m}(t)$ и $n_{i,m'}(t)$ для $m' \neq m$ являются выборочными функциями независимых стационарных гауссовских процессов с нулевыми средними; $v_{i,m}(t)$ (при $i=1,2$) имеют спектральную плотность $S\sigma(f)$, а $n_{i,m}(t)$ и $n_{i,m'}(t)$ (при всех $m' \neq m$ и $i = 1,2$) имеют спектральную плотность N_0. Наконец, примем, что модель справедлива для всех значений M и T.

Пусть $\mathscr{R}(\tau) = \int S\sigma(f) e^{j2\pi f\tau} df$ — автокорреляционная функция неусеченных процессов $v_{i,m}(t)$ и пусть $\{\vartheta_j(\tau)\}$ и $\{\lambda_j\}$ — собственные функции и собственные значения $\mathscr{R}(\tau)$ на интервале $(-T_1/2, T_2/2)$, где $T_1 = T - L$,

$$\int_{-T_1/2}^{T_1/2} \mathscr{R}(\tau_2 - \tau_1)\,\vartheta_j(\tau_2)\,d\tau_2 = \lambda_j \vartheta_j(\tau_1); \quad |\tau_1| < T_1/2. \tag{8.6.7}$$

Определим случайные величины $v_{i,m,j}$, $n_{i,m,j}$ и $y_{i,m',j}$ равенствами

$$v_{i,m,j} = \int_{-T_1/2}^{T_1/2} v_{i,m}(t)\,\vartheta_j(t)\,dt, \tag{8.6.8}$$

$$n_{i,m,j} = \int_{-T_1/2}^{T_1/2} n_{i,m}(t)\,\vartheta_j(t)\,dt, \tag{8.6.9}$$

$$y_{i,m',j} = \int_{-T_1/2}^{T_1/2} y_{i,m'}(t)\,\vartheta_j(t)\,dt = \begin{cases} v_{i,m',j} + n_{i,m',j}; & m' = m, \\ n_{i,m',j}; & m' \neq m. \end{cases} \tag{8.6.10}$$

При принятых предположениях $y_{i,m',j}$ может быть вычислено в приемнике для всех i, m', j. Оно представляет собой выборочное значение гауссовской случайной величины с нулевым средним и дисперсией, задаваемой равенством

$$\overline{y^2_{i,m',j}} = \begin{cases} \lambda_j + N_0; & m' = m, \\ N_0; & m' \neq m, \end{cases} \tag{8.6.11}$$

где m — переданное сообщение.

Пусть $\mathbf{y}_{m'}$ — последовательность выходных случайных величин $(y_{1,m',1}, \ldots, y_{1,m',J}, y_{2,m',1}, \ldots, y_{2,m',J})$. Сначала будем считать, что J произвольно, но фиксированно, а затем рассмотрим предел при $J \to \infty$. При $m \neq m'$ совместная плотность вероятности $\mathbf{y}_{m'}$ задается равенством

$$p_0(\mathbf{y}_{m'}) = \prod_{i=1}^{2} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi N_0}} \exp\left(-\frac{y_{i,m',j}^2}{2N_0}\right). \tag{8.6.12}$$

Если $m' = m$, где m — переданное сообщение, то совместная плотность вероятности $\mathbf{y}_{m'}$ задается равенством

$$p_1(\mathbf{y}_{m'}) = \prod_{i=1}^{2} \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi(N_0+\lambda_j)}} \exp\left[-\frac{y_{i,m',j}^2}{2(N_0+\lambda_j)}\right]. \tag{8.6.13}$$

Следовательно, при условии, что сообщение m передано, совместная плотность вероятности всего множества принятых случайных величин $y_{i,m',j}$, где $1 \leqslant j \leqslant J$, задается равенствами

$$p(\mathbf{y}_1, \ldots, \mathbf{y}_M \mid \mathbf{x}_m) = p_1(\mathbf{y}_m) \prod_{m' \neq m} p_0(\mathbf{y}_{m'}) = \tag{8.6.14}$$

$$= \frac{p_1(\mathbf{y}_m)}{p_0(\mathbf{y}_m)} \prod_{m'=1}^{M} p_0(\mathbf{y}_{m'}). \tag{8.6.15}$$

Декодер по максимуму правдоподобия, принимающий решение по этому множеству случайных величин (с индексами $1 \leqslant j \leqslant J$), будет декодировать такое m, которое максимизирует $p_1(\mathbf{y}_1, \ldots, \mathbf{y}_M \mid \mathbf{x}_m)$ или, что эквивалентно, такое m, которое максимизирует $p_1(\mathbf{y}_m)/p_0(\mathbf{y}_m)$. Верхнюю границу вероятности ошибки для этого декодера по максимуму правдоподобия можно получить с помощью той же последовательности рассуждений, как в доказательстве теоремы 5.6.1. В частности,

$$P_{e,m} = \int p_1(\mathbf{y}_m) \Pr(\text{ошибка} \mid m, \mathbf{y}_m)\, d\mathbf{y}_m, \tag{8.6.16}$$

где $\Pr(\text{ошибка} \mid m, \mathbf{y}_m)$ — вероятность ошибки при условии, что передано сообщение m и принята некоторая последовательность $\mathbf{y}_m$. Для заданного $\mathbf{y}_m$ пусть $A_{m'}$ — событие, состоящее в том, что

$$\frac{p_1(\mathbf{y}_{m'})}{p_0(\mathbf{y}_{m'})} \geqslant \frac{p_1(\mathbf{y}_m)}{p_0(\mathbf{y}_m)}. \tag{8.6.17}$$

Ошибка происходит тогда и только тогда, когда событие $A_{m'}$ произойдет при некотором m' и, следовательно,

$$\Pr(\text{ошибка} \mid m, \mathbf{y}_m) = P\left(\bigcup_{m' \neq m} A_{m'}\right) \leqslant$$

$$\leqslant \left[\sum_{m' \neq m} P(A_{m'})\right]^{\rho} \text{ для любого } \rho, \tag{8.6.18}$$

$$0 \leqslant \rho \leqslant 1,$$

где было использовано (5.6.2) и все вероятности в правой части являются условными при заданных m и $\mathbf{y}_m$. Строя границы тем же методом, как и при выводе (5.6.8), получаем

$$P(A_{m'}) = \int\limits_{\mathbf{y}_{m'}:\ \frac{p_1(\mathbf{y}_{m'})}{p_0(\mathbf{y}_{m'})} \geqslant \frac{p_1(\mathbf{y}_m)}{p_0(\mathbf{y}_m)}} p_0(\mathbf{y}_{m'})\, d\mathbf{y}_{m'} \leqslant$$

$$\leqslant \int p_0(\mathbf{y}_{m'}) \left[\frac{p_1(\mathbf{y}_{m'})}{p_0(\mathbf{y}_{m'})} \frac{p_0(\mathbf{y}_m)}{p_1(\mathbf{y}_m)}\right]^{1/(1+\rho)} d\mathbf{y}_{m'}. \qquad (8.6.19)$$

Подставляя (8.6.19) в (8.6.18) и (8.6.18) в (8.6.16) и замечая, что $\mathbf{y}_{m'}$ здесь является переменной интегрирования, получаем

$$P_{e,m} \leqslant \int p_1(\mathbf{y}_m)(M-1)^\rho \left\{\int p_0(\mathbf{y}_{m'}) \left[\frac{p_1(\mathbf{y}_{m'})\, p_0(\mathbf{y}_m)}{p_0(\mathbf{y}_{m'})\, p_1(\mathbf{y}_m)}\right]^{1/(1+\rho)} d\mathbf{y}_{m'}\right\}^\rho d\mathbf{y}_m =$$

$$= (M-1)^\rho \int p_1(\mathbf{y}_m)^{1/(1+\rho)}\, p_0(\mathbf{y}_m)^{\rho/(1+\rho)}\, d\mathbf{y}_m \times$$

$$\times \left[\int p_1(\mathbf{y}_{m'})^{1/(1+\rho)}\, p_0(\mathbf{y}_{m'})^{\rho/(1+\rho)}\, d\mathbf{y}_{m'}\right]^\rho =$$

$$= (M-1)^\rho \left[\int p_1(\mathbf{y}_m)^{1/(1+\rho)}\, p_0(\mathbf{y}_m)^{\rho/(1+\rho)}\, d\mathbf{y}_m\right]^{1+\rho}. \qquad (8.6.20)$$

Подставляя (8.6.12) и (8.6.13) в (8.6.20), заметим, что интеграл по $\mathbf{y}_m$ распадается на произведение интегралов по компонентам $y_{i,m,j}$. Они представляют собой интегралы от гауссовских функций, зависящих от j, но не от i. Имеем

$$\left[\int p_1(\mathbf{y}_m)^{1/(1+\rho)}\, p_0(\mathbf{y}_m)^{\rho/(1+\rho)}\, d\mathbf{y}_m\right]^{1+\rho} =$$

$$= \prod_{j=1}^{J} \frac{[1+(\lambda_j/N_0)]^\rho}{\{1+\rho\lambda_j/[(1+\rho)\, N_0]\}^{1+\rho}}. \qquad (8.6.21)$$

Так как этот результат справедлив для всех J, то можно перейти к пределу при $J \to \infty$ и получить результат в виде

$$P_{e,m} \leqslant (M-1)^\rho \exp[-TE_0(\rho, T)], \qquad (8.6.22)$$

где

$$E_0(\rho, T) = \frac{1}{T} \sum_{j=1}^{\infty} \left[(1+\rho) \ln\left(1 + \frac{\rho\lambda_j}{(1+\rho)\, N_0}\right) - \rho \ln\left(1 + \frac{\lambda_j}{N_0}\right)\right]. \qquad (8.6.23)$$

Как обычно в задачах, связанных с собственными значениями, результаты можно упростить, полагая интервал времени большим. Для того чтобы указать на зависимость собственных значений от интервала приема $T_1 = T - L$, будем теперь писать $\lambda_j(T_1)$ вместо λ_j. Взяв производную от $E_0(\rho, T)$ по $\lambda_j(T_1)$, можно заметить, что каждое слагаемое возрастает по λ_j с ограниченным наклоном. Поэтому, если $S\sigma(f)$ ограничена и интегрируема, то можно применить лемму 8.5.3 и получить*)

*) Из доказательства леммы 8.5.3 следует, что сходимость в (8.6.24) равномерна по ρ при $0 \leqslant \rho \leqslant 1$, однако этот результат здесь не потребуется.

$$\lim_{T_1\to\infty}\frac{1}{T_1}\sum_{j=1}^{\infty}(1+\rho)\ln\left[1+\frac{\rho\lambda_j(T_1)}{(1+\rho)N_0}\right]-\frac{1}{T_1}\sum_{j=1}^{\infty}\rho\ln\left[1+\frac{\lambda_j(T_1)}{N_0}\right]=$$

$$=\int_{-\infty}^{\infty}\left\{(1+\rho)\ln\left[1+\frac{\rho S\sigma(f)}{(1+\rho)N_0}\right]-\rho\ln\left(1+\frac{S\sigma(f)}{N_0}\right)\right\}df. \quad (8.6.24)$$

Так как L фиксировано, то T_1/T стремится к 1 при $T_1\to\infty$ и поэтому

$$E_0(\rho)=\lim_{T\to\infty}E_0(\rho,T)=$$

$$=\int\left\{(1+\rho)\ln\left[1+\frac{\rho S\sigma(f)}{(1+\rho)N_0}\right]-\rho\ln\left[1+\frac{S\sigma(f)}{N_0}\right]\right\}df. \quad (8.6.25)$$

Отсюда следует, что для любого ρ, $0\leqslant\rho\leqslant 1$, и любого $\varepsilon>0$ можно выбрать T достаточно большим, так что для любого M и всех m, $1\leqslant m\leqslant M$,

$$P_{e,m}\leqslant(M-1)^{\rho}\exp\{-T[E_0(\rho)-\varepsilon]\}. \quad (8.6.26)$$

Прежде чем интерпретировать этот результат, полезно ввести еще одну степень свободы в эту ситуацию. Величина S задает полную мощность сигнала, имеющуюся на приемнике, и так как S пропорциональна мощности передатчика, то ее можно понимать как мощность передатчика, нормированную к приемнику. Если передатчик не имеет ограничений на пиковую мощность, но имеет ограничение на среднюю мощность S (при нормировке на приемнике), то одна из возможностей, существующая для передатчика, состоит в использовании на передачу кодовых слов доли времени θ с мощностью передачи S/θ в течение этого времени. Назовем θ *коэффициентом занятости* передачи. Можно также сделать θ большим, чем 1, перекрывая время передачи последовательных кодовых слов путем использования разных полос частот для последовательных слов. Пусть $T'=T/\theta$ — время (в секундах) между началами последовательных кодовых слов. Тогда можно получить произвольную скорость передачи R в натуральных единицах в секунду, используя M кодовых слов, где

$$M=\lceil e^{T'R}\rceil. \quad (8.6.27)$$

Оценивая сверху (8.6.26) с помощью неравенства $M-1<\exp(T'R)$, считая, что S/θ — средняя принимаемая мощность в течение передачи, и подставляя $T'\theta$ вместо T, находим, что для любого $\varepsilon>0$ существует такое достаточно большое T', что

$$P_{e,m}\leqslant\exp\{-T'[-\rho R+\tilde{E}_0(\rho,S,\theta)-\varepsilon]\}, \quad (8.6.28)$$

$$\tilde{E}_0(\rho,S,\theta)=\theta\int\left\{(1+\rho)\ln\left[1+\frac{\rho A(f)}{1+\rho}\right]-\rho\ln[1+A(f)]\right\}df, \quad (8.6.29)$$

где

$$A(f)=\frac{S\sigma(f)}{\theta N_0}. \quad (8.6.30)$$

В этом случае, требуемая величина T' при заданном $\varepsilon > 0$ зависит от θ и становится неограниченной при $0 \to 0$.

Свойства показателя экспоненты в (8.6.28) при фиксировнных θ и S аналогичны свойствам показателя экспоненты — $\rho R + E_0(\rho, \mathbf{Q})$ при фиксированном $\mathbf{Q}$, которые были рассмотрены в § 5.6. В частности, $E_0(0, S, \theta) = 0$ и

$$\frac{\partial E_0(\rho, S, \theta)}{\partial \rho} = \theta \int \left\{ \frac{A(f)}{1+\rho+\rho A(f)} - \ln\left[1 + \frac{A(f)}{1+\rho+\rho A(f)}\right] \right\} df > 0, \quad (8.6.31)$$

$$\frac{\partial^2 E_0(\rho, S, \theta)}{\partial \rho^2} = \theta \int \frac{-A^2(f)}{[1+\rho+\rho A(f)]^2 (1+\rho)} df < 0. \quad (8.6.32)$$

Определяя $E(R, S, \theta)$ как

$$\max_{0 \leqslant \rho \leqslant 1} [-\rho R + \tilde{E}_0(\rho, S, \theta)],$$

можно получить обычные параметрические уравнения, связывающие R и $E(R, S, \theta)$.

$$R = \theta \int \left\{ \frac{A(f)}{1+\rho+\rho A(f)} - \ln\left[1 + \frac{A(f)}{1+\rho+\rho A(f)}\right] \right\} df, \quad (8.6.33)$$

$$E(R, S, \theta) = \theta \int \left\{ \ln\left[1 + \frac{\rho A(f)}{1+\rho}\right] - \frac{\rho A(f)}{1+\rho+\rho A(f)} \right\} df. \quad (8.6.34)$$

Эти уравнения справедливы при $0 \leqslant \rho \leqslant 1$, а для R, меньших, чем значение, определяемое (8.6.33) при $\rho = 1$, имеем

$$E(R, S, \theta) = \theta \int \left\{ 2 \ln\left[1 + \frac{A(f)}{2}\right] - \ln[1 + A(f)] \right\} df - R. \quad (8.6.35)$$

Пропускная способность $C(S, \theta)$ при заданных коэффициенте занятости и мощности определена как R, задаваемая (8.6.33) при $\rho = 0$,

$$C(S, \theta) = \theta \int \{A(f) - \ln[1 + A(f)]\} df =$$
$$= \theta \int \left\{ \frac{S\sigma(f)}{\theta N_0} - \ln\left[1 + \frac{S\sigma(f)}{\theta N_0}\right] \right\} df. \quad (8.6.36)$$

Используя те же рассуждения, как и в § 5.6 [или непосредственно рассматривая (8.6.34) и (8.6.35)], нетрудно заметить, что $E(R, S, \theta) > 0$ при всех $R < C(S, \theta)$.

Наконец, определим пропускную способность канала $C(S)$ при заданной мощности S как верхнюю грань $C(S, \theta)$ по $\theta > 0$. Легко видеть, что эта верхняя грань задается формулой

$$C(S) = \lim_{\theta \to 0} C(S, \theta) = \int \frac{S\sigma(f)}{N_0} df = S/N_0, \quad (8.6.37)$$

где $\sigma(f)$ нормирована равенством $\int \sigma(f) df = 1$.

Этот результат является интересным и неожиданным, так как он утверждает, что пропускная способность такого канала совпадает с пропускной способностью неограниченного по полосе частот гауссовского канала, в котором имеется ограничение на мощность S и в ко-

тором действует аддитивный шум со спектральной плотностью $N_0/2$. Для любой $R < C(S)$ значение θ можно выбрать столь малым, что $R < C(S,\theta)$. Для этого коэффициента занятости вероятность ошибки убывает экспоненциально с T'.

Для того чтобы установить обратный результат, т. е. то, что надежная передача невозможна, если $R > C(S)$, представим канал как последовательное соединение двух каналов, первый из которых будет диспергирующим каналом с замираниями, а второй будет каналом с аддитивным белым шумом*). Так как средняя мощность на входе второго канала с аддитивным шумом ограничена S, то средняя взаимная информация в секунду, передаваемая по нему, не превышает S/N_0. Из (2.3.19, б) следует, что средняя взаимная информация в секунду, передаваемая по всему каналу, не превышает S/N_0. Следовательно, теорема 8.5.2 применима также к этому каналу. Отсюда можно увидеть, что рассуждения, приведенные выше, не зависят от особенностей анализируемой модели канала, и для них существенны только мощность белого шума, ограничение на мощность принимаемого сигнала и независимость белого шума от остальной части канала и сигнала.

Эти результаты можно представить в виде следующей теоремы.

Теорема 8.6.1. Для модели канала, определенной в тексте, примыкающем к равенствам (8.6.5) и (8.6.6), при дополнительном условии, что $\sigma(f)$ ограничена и интегрируема, пропускная способность канала при ограничении на мощность S задается формулой $C(S) = S/N_0$. При любой $R < C(S)$ можно получить сколь угодно малую вероятность ошибки, взяв время передачи T' достаточно большим и коэффициент занятости достаточно малым. При $R > C(S)$ применима теорема 8.5.2. Для любого заданного коэффициента занятости θ, для любого $\varepsilon > 0$, для всех достаточно больших T' и для любой $R < C(S, \theta)$ код с $M = \lceil e^{T'R} \rceil$ кодовыми словами имеет вероятность ошибки, удовлетворяющую при всех m, $1 \leqslant m \leqslant M$, неравенству

$$P_{e,\,m} \leqslant \exp\{-T'[E(R, S, \theta) - \varepsilon]\}, \qquad (8.6.38)$$

где $E(R, S, \theta)$ задается соотношениями (8.6.33) и (8.6.35).

Сформулированные результаты не означают, что код, образованный разнесенными по частоте синусоидами, в каком-либо смысле минимизирует вероятность ошибки, которая может быть достигнута в рассматриваемом канале. Однако если не ограничиться синусоидами, то можно получить способ воздействия на выбор собственных чисел $\{\lambda_j\}$ в (8.6.23). В случае, когда допустимо полное управление выбором λ_j, при условии, что $\Sigma\lambda_j = S$, Кеннеди (1964) показал, что имеется оптимальное значение λ_j, скажем λ_{opt}, зависящее только от ρ, и что $E_0(\rho, T)$ максимизируется на таких λ_j, среди которых S/λ_{opt} равны λ_{opt} и остальные равны нулю.

---

*) В действительности для заданной модели второй канал следует рассматривать как $2M$ параллельных каналов с белыми гауссовыми шумами, имеющими спектральные плотности N_0 и общую мощность на входе $2S$ (т. е. S для $v_{1,\,m}(t)$ и S для $v_{2,\,m}(t)$). При этом получается тот же результат, что и выше.

ИТОГИ И ВЫВОДЫ

В этой главе были рассмотрены два частных класса непрерывных по времени каналов. Первый класс составляли каналы, в которых переданный сигнал сначала фильтровался, а затем складывался со стационарным гауссовым шумом. Фильтр можно рассматривать либо как частотное ограничение на входе, либо как часть канала. Второй класс составляли каналы, в которых передающая среда была диспергирующей и изменяющейся во времени.

В первом параграфе было показано, как представить функции времени и гауссовские случайные процессы с помощью ортонормальных разложений. Наше представление гауссовских случайных процессов было там не совсем обычным, так как мы определяли процесс через линейные операции над процессом, а не через совместные распределения процесса для всех конечных множеств моментов времени. Этот подход имеет те преимущества, что позволяет удовлетворительно с точки зрения физики описывать белый шум и избежать все математические тонкости и трудности, которые возникают при переходе от точечного описания к описанию с помощью линейных операций. В § 8.2 исследовалась оптимальная вероятность ошибки и оптимальный приемник множества ортогональных сигналов в белом шуме. Было показано также, что эти результаты могут быть непосредственно преобразованы в результаты для симплексного множества сигналов.

В § 8.3 дан эвристический вывод выражения для пропускной способности канала с фильтром и аддитивным стационарным гауссовым шумом. В § 8.4 и 8.5 это исследование продолжено со строгим анализом пропускной способности и верхних границ для минимума достижимой вероятности ошибки. Однако проведенный анализ не был полным; в нем не учитывалась интерференция между последовательными кодовыми словами. Последняя задача остается открытой для исследования.

В § 8.6 сначала была разработана математическая модель для передачи сообщений с помощью кода, образованного разнесенными по частоте синусоидами, по диспергирующему каналу с замираниями и аддитивным стационарным белым гуссовым шумом. Затем были выведены экспоненциальные границы вероятности ошибки для этой модели канала при использовании указанного класса кодов, и было показано, что пропускная способность канала равна S/N_0, где S — ограничение на мощность принятого сигнала, а $N_0/2$ — спектральная плотность шума.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Рассмотренные здесь ортонормальные разложения и интегральные уравнения хорошо освещены в литературе; можно рекомендовать, например, Куранта и Гильберта (1959) и Рисса и Надь (1955). По гауссовским случайным процессам Возенкрафт и Джекобс (1965) и Давенпорт и Рут (1958) написали превосходные книги для инженеров, а Дуб (1953) и Лоев (1955) написали превосходные математические

книги. Другой подход к обнаружению сигнала в небелом гауссовом шуме, не использующий ортонормальные разложения, можно найти у Кайлата (1967). Верхние и нижние границы вероятности ошибки для ортогональных сигналов на фоне белого гауссового шума были найдены Фано (1961) и Зеттербергом (1961) соответственно. Пропускная способность каналов с аддитивным не белым гауссовым шумом была найдена Шенноном (1948) и со строгим выводом Пинскером (1957). Пропускная способность и теорема кодирования для рассмотренных здесь каналов с фильтром были получены Холзингером (1964). Изложение § 8.4 и 8.5 весьма близко следует Холзингеру, за исключением доказательств, предложенных здесь для некоторых преобразований, которые ранее проводились формально. Вайнер (1966) провел исследование ряда различных математических моделей для частного случая строго ограниченного по полосе частот сигнала в белом гауссовом шуме и вывел для них теоремы кодирования. Его выводы создают дополнительную уверенность в том, что результаты являются нечувствительными к малым изменениям модели. Рут и Варея (1967) рассмотрели обобщение предложенной здесь модели, когда фильтр и шум недостаточно известны.

Кеннеди (1969) предложил хорошо написанную и значительно более полную разработку надежной передачи по диспергирующим каналам с замираниями и вывел верхнюю и нижнюю границы вероятности ошибки для более широкого класса систем связи. Результаты, указанные здесь, принадлежат главным образом Кеннеди. Верхняя граница вероятности ошибки, задаваемая (8.6.22) и (8.6.23), была выведена независимо Юдкиным (1964) и Кеннеди (1964), а Пирс (1961) ранее нашел выражение для $P_{e,\,m}$ при $M = 2$ для эквивалентной задачи разнесения, когда все λ_i были равны. Результат, что $C(S) = S/N_0$, впервые был получен (без какой-либо сопутствующей теоремы кодирования) Джекобсом (1963). Витерби (1967) также рассматривал случай, разнесенных по частоте синусоид, изложенный здесь, и получил верхнюю и нижнюю границы для $P_{e,\,m}$, в которых показатели экспонент совпадают в областях, где справедливы формулы (8.6.33) и (8.6.34). Ричтерс (1967) распространил результаты Кеннеди на случай, когда вход канала имеет ограниченную полосу частот. Он показал, что полоса частот, требуемая для надежной передачи, быстро возрастает при приближении скорости к $C(S)$, однако при малых скоростях и умеренных полосах частот экспонента близка к результату, соответствующему бесконечной полосе. Важность этого результата очевидна, так как изученные здесь синусоиды, разнесенные по частоте, требуют полосу частот, растущую экспоненциально с T_1; это ситуация, которая быстро становится физически нереальной.

9

КОДИРОВАНИЕ ИСТОЧНИКА С ЗАДАННЫМ КРИТЕРИЕМ ВЕРНОСТИ

9.1. ВВЕДЕНИЕ

В гл. 3 рассматривалось кодирование выхода источника для минимизации среднего числа кодовых букв на букву источника, при условии, что буквы источника можно было восстановить по кодовой последовательности. Если выход источника — последовательность непрерывно-значных случайных величин или случайный процесс, то, очевидно, невозможно закодировать выход источника в последовательность дискретных кодовых букв, по которой выход источника может быть точно восстановлен. В таких случаях можно только потребовать, чтобы восстановленное сообщение аппроксимировало источник с заданным критерием верности. Например, если выход источника порождает последовательность действительных чисел $u_1, u_2, \ldots$, и эта последовательность восстанавливается как последовательность $v_1, v_2, \ldots$, то можно потребовать, чтобы математическое ожидание $(u_l - v_l)^2$, взятое по ансамблю источника и усредненное по времени, было меньше некоторого предписанного значения. В более общем случае можно определить меру искажения [$(u_l - v_l)^2$ — в рассмотренном выше примере] как произвольную действительную функцию выхода источника и восстановленной последовательности. Тогда критерий верности определяется как максимум допустимого значения среднего искажения.

Основная цель этой главы — найти минимальное число двоичных символов на букву источника или на единицу времени, требуемых для кодирования источника так, чтобы по кодовой последовательности можно было восстановить сообщение, удовлетворяющее заданному критерию верности. Естественно, что этот предел зависит от статистики источника, меры искажения и критерия верности.

Смысл представления выхода источника последовательностью двоичных символов заключается в том, чтобы отделить задачу представления источника от задачи передачи информации. Мы уже знаем из теоремы кодирования, если скорость двоичной последовательности (в битах в секунду) меньше пропускной способности канала (в битах в секунду), по которому последовательность должна быть передана, то последовательность может быть воспроизведена на выходе канала с произвольно малой вероятностью ошибки. Так как при стремлении к 0 вероятности ошибки влияние этих ошибок на полное искажение обычно становится малым, то можно в действительности отделить задачу кодирования для канала от задачи кодирования для источника.

В дальнейшем будет показано, что в некотором смысле ничего не теряется при таком промежуточном представлении двоичными символами. Точнее, будет показано, что если пропускная способность канала слишком мала для надежной передачи с минимальной скоростью, требуемой для представления источника с заданным критерием верности, то этот критерий не может быть удовлетворен, независимо от того, какого вида обработка используется между источником и каналом.

В принципе теория, развиваемая здесь, применима к задачам, обычно классифицируемым как квантование, преобразование аналог—цифра, сжатие полосы частот и редукция данных. На практику эта теория еще не оказывает большого влияния. Частично это объясняется отсутствием множества эффективных методов для осуществления кодирования источника при условии, что удовлетворяется критерий верности. Более существенной причиной является трудность нахождения приемлемых математических моделей для практически важных источников. Например, трудно дать статистическое описание речевого сигнала и еще труднее найти разумную меру искажения в этом случае. Однако даже в такой сложной проблеме могут быть весьма полезными те результаты и то понимание, которые возникают, из развиваемого здесь теоретического подхода.

9.2. ДИСКРЕТНЫЕ ИСТОЧНИКИ БЕЗ ПАМЯТИ И МЕРЫ ИСКАЖЕНИЯ ОТДЕЛЬНОЙ БУКВЫ

В этом и последующих трех параграфах будет изучено кодирование источника с критерием верности для следующего случая. Будет рассматриваться дискретный источник U без памяти с алфавитом $(0, \ldots, K-1)$ и вероятностями букв $Q(0), \ldots, Q(K-1)$. Всюду в дальнейшем предполагается, что K конечно и что $Q(k) > 0$ для всех k, $0 \leqslant k \leqslant K-1$ (если $Q(k)$ равно 0, то можно просто исключить эту букву из рассмотрения). Выход источника представляет собой последовательность $u_1, u_2, \ldots$ значений, выбранных независимо из заданного алфавита с заданными вероятностями букв. Последовательность источника должна быть представлена при поступлении ее адресату последовательностью букв $v_1, v_2, \ldots$, каждая из которых выбрана из алфавита $(0, 1, \ldots, J-1)$, $J < \infty$. Наконец, имеется мера искажения $d(k; j)$, определенная для $0 \leqslant k \leqslant K-1$, $0 \leqslant j \leqslant J-1$ и задающая численное значение искажения в случае, если буква k источника воспроизводится у адресата как буква j. В дальнейшем изложении всюду будем предполагать, что $d(k; j) \geqslant 0$ и что для каждого k имеется по крайней мере одно j, для которого $d(k; j) = 0$. Легко видеть, что это предположение в действительности не приводит к потере общности, так как если $d'(k; j)$ — произвольная функция k, j и по определению

$$m(k) = \min_j d'(k; j),$$

то $d(k; j) = d'(k; j) - m(k)$ будет иметь требуемый вид. Так как разность между $d(k; j)$ и $d'(k; j)$ является функцией только буквы k источника, то она не зависит от преобразований, проводимых между источни-

ком и адресатом. Иногда мы будем ограничиваться рассмотрением *конечных* мер искажения (при этом считается, что $d(k; j)$ конечна для всех k, j), но в бо́льшей части изложения будут допускаться и бесконечные значения $d(k; j)$. Бесконечное искажение между заданными k и j равносильно абсолютному запрещению того, чтобы буква k источника была представлена у адресата буквой j. Полное искажение между последовательностью $u_1, \ldots, u_L$ букв источника и последовательностью $v_1, \ldots, v_L$ букв адресата определяется как

$$\sum_{l=1}^{L} d(u_l; v_l).$$

Примером меры искажения является $d(k; j) = 0$ при $j = k$ и $d(k; j) = 1$ при $j \neq k$ в случае, когда $J = K$. Такая мера искажения букв будет приемлемой, если требуется точное воспроизведение букв источника и все ошибки считаются одинаково серьезными. Во втором примере положим $J = K + 1$, и путь $d(k; j) = 0$ при $j = k$, $d(k; j) = 1$ при $j \neq k$, $j < J - 1$, и $d(k; J - 1) = {}^1/_2$ для всех k. В этом примере выход $J - 1$ можно интерпретировать как стертый символ. Такая мера искажения является разумной, если все ошибки одинаково опасны, а стирание лишь наполовину опасно по сравнению с ошибкой. В третьем примере положим $J = K$, и путь $d(k; j) = (j - k)^2$ для всех j, k. Такая мера искажения приемлема, если буквы источника представляют собой значения амплитуды и большие ошибки в амплитуде более опасны, чем малые. В последнем примере положим $J = 2$, и пусть $d(k; j) = 0$, когда $k + j$ четно, и $d(k; j) = 1$, когда $k + j$ нечетно. Такая мера искажения является подходящей, если существенно только то, четны или нечетны буквы источника. Этот последний пример несколько искусствен, однако он показывает, что мера искажения может быть использована для указания степени важности воспроизведения отдельных сторон выхода источника.

Когда источник связан с адресатом каналом и некоторыми преобразованиями, статистика источника, статистика канала и операции преобразующих устройств определяют совместную вероятностную меру на входной последовательности **u** и выходной последовательности **v**. Вероятностная мера в свою очередь определяет среднее искажение на букву источника. Нас интересует нахождение минимума среднего искажения, которое может быть достигнуто при заданном канале. Будет показано, что этот минимум зависит лишь от пропускной способности канала и к нему можно подойти сколь угодно близко с помощью преобразователя, который отображает сначала выход источника в двоичный поток данных со скоростью, сколь угодно близкой к пропускной способности канала, а затем кодирует двоичный поток данных для передачи по каналу.

Начнем с изложения фундаментального определения, которое на первый взгляд кажется несколько произвольным. *Скорость как функция искажения* для заданных дискретного источника без памяти и меры искажения определяется следующим образом. Рассмотрим произвольное множество переходных вероятностей $P(j|k)$, где $P(j|k)$ —

условная вероятность буквы j у адресата при условии, что k была буквой источника. Эти вероятности вместе с вероятностями букв источника определяют среднюю взаимную информацию

$$\mathcal{I}(\mathbf{Q};\mathbf{P}) = \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} Q(k)\,P(j|k)\ln\frac{P(j|k)}{\sum\limits_{i} Q(i)\,P(j|i)}, \tag{9.2.1}$$

и среднее искажение

$$\overline{d} = \sum_{k}\sum_{j} Q(k)\,P(j|k)\,d(k;j). \tag{9.2.2}$$

Тогда скорость как функция искажения для источника относительно заданной меры искажения определяется равенством

$$R(d^*) = \min_{\mathbf{P}:\,\overline{d}\leqslant d^*} \mathcal{I}(\mathbf{Q};\mathbf{P}). \tag{9.2.3}$$

Минимизация в (9.2.3) проводится по всем переходным вероятностям при ограничении, что среднее искажение меньше или равно d^*. Ниже будет показано, что если для заданного d^* пропускная способность канала, связывающего источник и адресат, меньше чем $R(d^*)$ натуральных единиц на символ источника, то независимо от того, какие преобразования проводятся до и после передачи по каналу, среднее искажение должно быть больше d^*. Обратно, будет показано, что источник может быть закодирован в $R(d^*)/\ln 2$ двоичных символов на букву источника и что двоичные символы могут быть декодированы в буквы адресата таким образом, что среднее искажение на букву не будет превышать d^* более чем на произвольно малую величину. В свете этих результатов разумно рассматривать $R(d^*)$ как скорость источника в натуральных единицах на символ относительно критерия верности d^*.

Отложим вопрос о том, как вычислять $R(d^*)$ до вывода указанных выше результатов и вывода некоторых общих свойств этой функции. Покажем сначала, что $R(d^*)$ неотрицательна, не возрастает с d^* и выпукла $\smile$ по d^*. Неотрицательность очевидна, так как $\mathcal{I}(\mathbf{Q};\mathbf{P}) \geqslant 0$. Заметим далее, что минимум в (9.2.3) берется по множеству, удовлетворяющему некоторым ограничениям и расширяющемуся при увеличении d^*. Следовательно, получающийся минимум $R(d^*)$ не возрастает с d^*. Для того чтобы показать, что $R(d^*)$ выпукло $\smile$, положим, что для d_1^* минимум в (9.2.3) достигается на $P_1(j|k)$, а для d_2^* минимум достигается на $P_2(j|k)$. Тогда, так как d_1^* и d_2^* соответственно больше или равны искажениям, которые относятся к $\mathbf{P}_1$ и $\mathbf{P}_2$, то имеем для любого θ, $0<\theta<1$,

$$\theta d_1^* + (1-\theta)\,d_2^* \geqslant \sum_{k,\,j} Q(k)\,[\theta P_1(j|k) + (1-\theta)\,P_2(j|k)]\,d(j;k). \tag{9.2.4}$$

Это означает, что $\theta\mathbf{P}_1 + (1-\theta)\,\mathbf{P}_2$ принадлежит множеству, удовлетворяющему заданным ограничениям, по которому $\mathcal{I}(\mathbf{Q},\mathbf{P})$ минимизируется для нахождения $R[\theta d_1^* + (1-\theta)\,d_2^*]$. Следовательно,

$$R[\theta d_1^* + (1-\theta)\,d_2^*] \leqslant \mathcal{I}[\mathbf{Q};\ \theta\mathbf{P}_1 + (1-\theta)\,\mathbf{P}_2]. \tag{9.2.5}$$

Так как теорема 4.4.3 утверждает, что $\mathcal{J}(\mathbf{Q};\ \mathbf{P})$ выпуклая $\smile$ по $\mathbf{P}$, то (9.2.5) можно далее ограничить сверху неравенством

$$R[\theta d_1^* + (1-\theta) d_2^*] \leqslant \theta \mathcal{J}(\mathbf{Q};\ \mathbf{P}_1) + (1-\theta)\mathcal{J}(\mathbf{Q};\ \mathbf{P}_2) = \\ = \theta R(d_1^*) + (1-\theta) R(d_2^*), \tag{9.2.6}$$

которое представляет требуемый результат.

На рис. 9.2.1 изображен вид функции $R(d^*)$. Очевидно, что наименьшее возможное значение для среднего искажения равно нулю и достигается при отображении каждой буквы k алфавита источника в букву адресата j, для которой $d(k; j)=0$. При $d^*<0$ функция $R(d^*)$ не определена, так как по определению $d(k; j) \geqslant 0$.

Далее определим d_{max}^* как наименьшее d^*, для которого $R(d^*) = 0$ (рис. 9.2.1). Значение d_{max}^* можно вычислить по формуле

$$d_{max}^* = \min_j \sum_k Q(k)\, d(k; j). \tag{9.2.7}$$

Для того чтобы показать это, заметим, что если $\mathcal{J}(\mathbf{Q};\ \mathbf{P}) = 0$, то буква адресата должна быть статистически независима от буквы источника и, следовательно, среднее искажение при условии, что выбрана буква адресата j, равно

$$\sum_k Q(k)\, d(k; j).$$

Следовательно, минимум $\bar{d}$ при условии $\mathcal{J}(\mathbf{Q};\ \mathbf{P}) = 0$ достигается всегда при выборе j, определяемом из (9.2.7). Если для каждого j имеется некоторое k, для которого $d(k; j) = \infty$, то, очевидно, $d_{max}^* = \infty$.

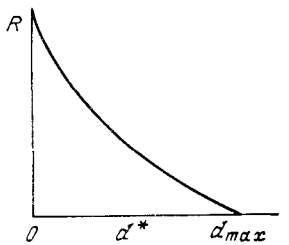


Рис. 9.2.1. Вид типичной функции $R(d^*)$.

Перейдем теперь к выводу того, что если выход источника с заданным $R(d^*)$ передается по каналу с пропускной способностью C натуральных единиц на символ источника, то среднее искажение на букву $\bar{d}$ должно удовлетворить неравенству $C \geqslant R(\bar{d})$.

Теорема 9.2.1. Пусть $R(d^*)$ — скорость как функция искажения для заданных дискретного источника без памяти и меры искажения. Пусть $\mathbf{U}^L\mathbf{V}^L$ — совместный ансамбль, $\mathbf{U}^L$ — ансамбль последовательностей источника $\mathbf{u} = (u_1, \ldots, u_L)$ длины L и $\mathbf{V}^L$ — ансамбль последовательностей алфавита адресата $\mathbf{v} = (v_1, \ldots, v_L)$. Пусть

$$\bar{d}_L = \overline{\frac{1}{L} \sum_{l=1}^{L} d(u_l; v_l)} \tag{9.2.8}$$

является средним искажением на букву для этого совместного ансамбля. Тогда

$$\frac{1}{L} I(\mathbf{U}^L;\ \mathbf{V}^L) \geqslant R(\bar{d}_L). \tag{9.2.9}$$

Доказательство. Имеем

$$\frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) = \frac{1}{L} (H(\mathbf{U}^L) - H(\mathbf{U}^L | \mathbf{V}^L)), \tag{9.2.10}$$

$$H(\mathbf{U}^L) = \sum_{l=1}^{L} H(U_l), \tag{9.2.11}$$

$$H(\mathbf{U}^L | \mathbf{V}^L) = \sum_{l=1}^{L} H(U_l | \mathbf{V}^L U_1 \dots U_{l-1}) \leqslant \sum_{l=1}^{L} H(U_l | V_l). \tag{9.2.12}$$

В равенстве (9.2.11) использовано то, что источник без памяти, а в (9.2.12) использованы соотношения (2.2.30) и (2.3.13). Подставляя (9.2.11) и (9.2.12) в (9.2.10), получаем

$$\frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) \geqslant \frac{1}{L} \sum_{l=1}^{L} I(U_l; V_l). \tag{9.2.13}$$

Полагая, что $\overline{d(l)}$ — среднее искажение l-й буквы, в соответствии с определением $R(d^*)$ в (9.2.3) находим, что $I(U_l; V_l) \geqslant R[\overline{d(l)}]$. Используя выпуклость $\smile$ функции $R(d^*)$, теперь имеем

$$\frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) \geqslant \sum_{l=1}^{L} \frac{1}{L} R[\overline{d(l)}] \geqslant$$

$$\geqslant R\left[\sum_{l=1}^{L} \frac{1}{L} \overline{d(l)}\right] = R(\overline{d}_L). \quad |$$

Приведем интересное истолкование этой теоремы. Пусть $P(j|k)$ — переходные вероятности, на которых достигается $R(d^*)$ из (9.2.3) при заданном d^*, и рассмотрим передачу L букв последовательности источника по дискретному каналу без памяти с этими переходными вероятностями. Тогда

$$\frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) = R(d^*)$$

и среднее искажение на букву между $\mathbf{U}^L$ и $\mathbf{V}^L$ не превышает d^*. Вместе с тем для любых других $\mathbf{U}^L\mathbf{V}^L$ со средним искажением на букву, не большим d^*, эта теорема утверждает, что

$$\frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) \geqslant R(d^*).$$

Следовательно, $R(d^*)$ может быть эквивалентно определена для любого $L \geqslant 1$ равенством

$$R(d^*) = \min_{\overline{d}_L \leqslant d^*} \frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L), \tag{9.2.14}$$

где минимизация проводится по $P_L(\mathbf{v}|\mathbf{u})$ при условии, что $\overline{d}_L \leqslant d^*$.

Предположим теперь, что при заданном L выбирается множество, скажем M, последовательностей адресата $\mathbf{v}_1, \dots, \mathbf{v}_M$, каждая из которых

имеет длину L, и множество последовательностей источника длины L отображается в это множество последовательностей адресата. Будем называть такое множество последовательностей и такое отображение (L, M)-кодом источника, а сами последовательности $\mathbf{v}_1, \ldots, \mathbf{v}_M$ — кодовыми словами. Каждое кодовое слово при таком кодировании можно представить последовательностью из $\lceil \log_2 M \rceil$ двоичных символов, и если эти двоичные символы передаются надежно по каналу с шумом, то искажение между последовательностью адресата (одним из кодовых слов) и последовательностью источника равно просто искажению, введенному при первоначальном кодировании.

Вероятности последовательностей из L букв источника и кодирование, определяемое заданным (L, M)-кодом источника, определяют совместный ансамбль $\mathbf{U}^L\mathbf{V}^L$ последовательностей источника и последовательностей адресата. Точнее, вероятность последовательности источника $\mathbf{u} = (u_1, \ldots, u_L)$ задается равенством

$$Q_L(\mathbf{u}) = \prod_{l=1}^{L} Q(u_l),$$

а условная вероятность последовательности адресата $\mathbf{v}$ при заданной последовательности источника $\mathbf{u}$ определяется равенством

$$P(\mathbf{v} \mid \mathbf{u}) = \begin{cases} 1, & \text{если } \mathbf{v} \text{ кодовое слово, в которое отображается } \mathbf{u}, \\ 0 & \text{в других случаях.} \end{cases}$$

Из теоремы 9.2.1 следует, что если $\bar{d}_L$ — среднее искажение на букву между последовательностями источника и последовательностями адресата для этого кода, то

$$\frac{1}{L} I(\mathbf{U}^L; \mathbf{V}^L) \geqslant R(\bar{d}_L).$$

Вместе с тем, так как имеется не более M последовательностей адресата (т. е. кодовых слов) с отличной от нуля вероятностью, то

$$\log M \geqslant H(\mathbf{V}^L) \geqslant I(\mathbf{U}^L; \mathbf{V}^L). \tag{9.2.15}$$

Из этих соотношений получаем простое следствие теоремы 9.2.1.

Следствие. Для того чтобы (L, M)-код источника имел среднее искажение на букву $\bar{d}_L$, необходимо, чтобы энтропия множества кодовых слов удовлетворяла неравенству

$$\frac{1}{L} H(\mathbf{V}^L) \geqslant R(\bar{d}_L), \tag{9.2.16}$$

а M — неравенству

$$\frac{\log M}{L} \geqslant R(\bar{d}_L). \tag{9.2.17}$$

Переходим теперь к изложению основного результата, который фактически содержит это следствие как частный случай.

Теорема 9.2.2. Пусть для дискретного источника U без памяти и меры искажения $d(k; j)$ задана скорость как функция искажения $R(d^*)$ и источник порождает одну букву каждые τ_s секунд. Пусть последовательность L букв источника связана с адресатом с помощью $N = \lfloor L\tau_s/\tau_c \rfloor$ использований дискретного по времени канала и пусть $\bar{d}_L$ является результирующим искажением на букву источника.

а) Если канал является каналом без памяти (с ограничениями или без ограничений на входе) и C — его пропускная способность (в натах) на одно использование, то

$$R(\bar{d}_L) \leqslant \frac{\tau_s}{\tau_c} C. \tag{9.2.18}$$

б) Если канал является каналом с конечным числом состояний и $\bar{C}$ — его верхняя пропускная способность (в натах) на одно использование, то в пределе при $L \to \infty$, $N = L\lfloor \tau_s/\tau_c \rfloor$ имеем

$$R(\bar{d}_\infty) \leqslant \frac{\tau_s}{\tau_c} \bar{C}. \tag{9.2.19}$$

в) Если канал является каналом с конечным числом A состояний и $\underline{C}$ — его нижняя пропускная способность (в натах) на одно использование и если входной ансамбль канала $\mathbf{X}^N$ не зависит от первоначального состояния s_0, то по крайней мере для одного значения s_0

$$R(\bar{d}_L) \leqslant \frac{\tau_s}{\tau_c}\left[\underline{C} + \frac{\log A}{N}\right]. \tag{9.2.20}$$

Обсуждение. Для дискретных по времени каналов без памяти и для неразложимых каналов с конечным числом состояний теорема устанавливает, что в пределе при больших L функция $R(\bar{d}_L)$ меньше или равна пропускной способности канала (в натах) на символ источника. Следовательно, если ордината кривой $R(d^*)$ равна пропускной способности (в натах на символ источника), то соответствующая абсцисса равна нижней границе среднего искажения на букву, независимо от того, какие преобразования проводятся над источником и на входе и выходе канала. Обратно, если абсцисса обозначает требуемое значение критерия верности, то соответствующая ордината является нижней границей пропускной способности канала, требуемой для достижения этого значения критерия верности. Эта теорема обычно называется обращением теоремы кодирования для источников относительно некоторой меры искажения.

Доказательство. Утверждение а. Из равенства (7.2.11), рассматривая $\mathbf{X}^N$, $\mathbf{Y}^N$ соответственно как входные и выходные ансамбли канала, получаем

$$I(\mathbf{U}^L; \mathbf{V}^L) \leqslant I(\mathbf{X}^N; \mathbf{Y}^N). \tag{9.2.21}$$

Согласно (7.2.12) для канала без ограничений на входе и согласно (7.3.5) для канала с ограничением на входе имеем

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) \leqslant NC. \tag{9.2.22}$$

Сочетая (9.2.9), (9.2.21) и (9.2.22), имеем

$$R(\overline{d}_L) \leqslant \frac{N}{L} C. \tag{9.2.23}$$

Так как $N \leqslant L\tau_s/\tau_c$, то отсюда получаем (9.2.18).

Утверждение б. Для заданного начального состояния s_0 и заданного L (9.2.9) утверждает, что

$$R(\overline{d}_L) \leqslant \frac{1}{L} I(\mathbf{U}^L;\ \mathbf{V}^L \mid s_0). \tag{9.2.24}$$

Из (4.6.15) и (4.6.21) имеем

$$I(\mathbf{U}^L;\ \mathbf{V}^L \mid s_0) \leqslant I(\mathbf{X}^N;\ \mathbf{Y}^N \mid s_0) \leqslant N\overline{C}_N. \tag{9.2.25}$$

Следовательно, независимо от начального состояния

$$R(\overline{d}_L) \leqslant \frac{\tau_s}{\tau_c} \overline{C}_N. \tag{9.2.26}$$

Переходя к пределу при $L \to \infty$, получаем $\overline{C}_N \to \overline{C}$ и отсюда следует (9.2.19).

Утверждение в. Из (4.6.15) и (4.6.23) следует, что существует некоторое начальное состояние s_0, для которого

$$I(\mathbf{U}^L;\ \mathbf{V}^L \mid s_0) \leqslant I(\mathbf{X}^N;\ \mathbf{Y}^N \mid s_0) \leqslant N\underline{C}_N. \tag{9.2.27}$$

Согласно теореме 4.6.1

$$\underline{C}_N \leqslant \underline{C} + \frac{\log A}{N}. \tag{9.2.28}$$

Сочетая (9.2.24), (9.2.27) и (9.2.28), получаем (9.2.20). |

Предыдущая теорема несколько неестественна в том, что она относится только к двум классам каналов. Единственная трудность в установлении эквивалентной теоремы для других каналов лежит в нахождении подходящего определения для пропускной способности. Как показано в § 4.6, максимум средней взаимной информации на букву даже для канала с конечным числом состояний может существенно зависеть от начального состояния и от того, известно или нет это начальное состояние передатчику. Для наиболее интересных частных каналов, не входящих в эти классы, таких, как канал с аддитивным гауссовым шумом, изучавшийся в последней главе, сравнительно легко определить пропускную способность как максимум средней взаимной информации на единицу времени в пределе для больших интервалов времени. Как только пропускная способность будет определена и проблема, связанная с влиянием прошлой истории будет каким-то образом обойдена, то так же, как в последней теореме, можно будет найти, что $R(\overline{d}_\infty) \leqslant \tau_s C$, где C — пропускная способность канала в натах в секунду.

9.3. ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ ИСТОЧНИКОВ ПРИ ЗАДАННОМ КРИТЕРИИ ВЕРНОСТИ

Обратимся теперь к установлению того, что $R(d^*)/\ln 2$ является числом двоичных символов на символ источника, требуемых для достижения среднего искажения, произвольно близкого к d^*. Для заданных источника U и алфавита адресата V *код источника* из M кодовых слов с длиной блока L определяется как отображение множества последовательностей длины L источника в множество M *кодовых* слов, где каждое кодовое слово $\mathbf{v}_m = (v_{m1}, \ldots, v_{mL})$ является последовательностью из L букв алфавита адресата. Пример рис. 9.3.1 иллюстрирует код источника с двумя кодовыми словами с длиной блока 3 для двоичных алфавитов источника и адресата.

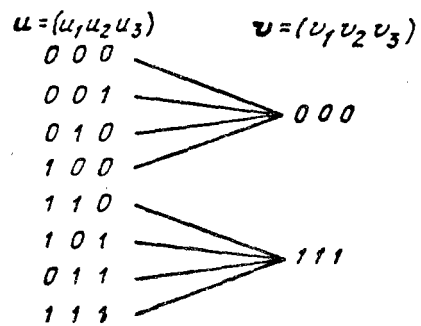


Рис. 9.3.1. Код источника с двумя кодовыми словами и длиной блока 3 (линии указывают отображение **u** в **v**).

Для заданной меры искажения $d(k; j)$ среднее искажение на букву кода источника задается равенством

$$\bar{d}_L = \frac{1}{L} \sum_{\mathbf{u}} Q_L(\mathbf{u})\, D[\mathbf{u}; \mathbf{v}(\mathbf{u})], \tag{9.3.1}$$

где $\mathbf{v}(\mathbf{u})$ — кодовое слово, в которое отображается $\mathbf{u}$ и

$$D(\mathbf{u}; \mathbf{v}) = \sum_{l=1}^{L} d(u_l; v_l). \tag{9.3.2}$$

Для кода источника с M кодовыми словами каждое кодовое слово может быть представлено отличной от других двоичной последовательностью длины $\lceil \log_2 M \rceil$ или с помощью $(1/L)\lceil \log_2 M \rceil$ двоичных символов на символ источника. Из теоремы кодирования для канала с шумами следует, что эти последовательности могут быть переданы с любой надежностью по каналу, пропускная способность которого больше чем $(1/L)\lceil \log_2 M \rceil$ двоичных символов на символ источника. Следовательно, основная задача сводится к нахождению, насколько малым может быть сделано $\bar{d}_L$ при заданных L и M.

С этого момента, и это не должно быть удивительно, мы попытаемся решить эту задачу, анализируя поведение случайно выбранного множества кодовых слов. Пусть $P(j|k)$ — заданное множество переходных вероятностей между буквами источника и адресата. Для удобства будем называть дискретный канал без памяти с этим переходным вероятностным *тест-каналом* и, если на $P(j|k)$ для заданного d^* достигается $R(d^*)$, то будем говорить, что на соответствующем тест-канале достигается $R(d^*)$ для этого d^*. Выходные вероятности $\omega(j)$ для заданного тест-канала равны

$$\omega(j) = \sum_k Q(k)\, P(j \mid k).$$

Для любого заданного тест-канала рассмотрим ансамбль кодов источника, в котором каждая буква каждого кодового слова выбирается независимо с распределением вероятности $\omega(j)$. Для заданного множества кодовых слов $\mathbf{v}_1, \ldots, \mathbf{v}_M$ из ансамбля каждая последовательность источника $\mathbf{u}$ отображается в такое кодовое слово $\mathbf{v}_m$, для которого $D(\mathbf{u}; \mathbf{v}_m)$ минимально по m.

В последующих рассуждениях одновременно будут рассматриваться две различные вероятностные меры на входных и выходных последовательностях; одна — на ансамбле тест-канала, другая — на ансамбле случайного кода. Для ансамбля тест-канала вероятностная мера на входных последовательностях $\mathbf{u} = (u_1, \ldots, u_L)$ и выходных последовательностях $\mathbf{v} = (v_1, \ldots, v_L)$ задается выражением $Q_L(\mathbf{u})\, P_L(\mathbf{v} \mid \mathbf{u})$, где

$$Q_L(\mathbf{u}) = \prod_{l=1}^{L} Q(u_l), \tag{9.3.3}$$

$$R_L(\mathbf{v} \mid \mathbf{u}) = \prod_{l=1}^{L} P(v_l \mid u_l). \tag{9.3.4}$$

В этих выражениях $Q(u_l)$ и $P(v_l \mid u_l)$ — вероятности источника и тест-канала соответственно. Взаимная информация равна

$$I(\mathbf{u}; \mathbf{v}) = \ln \frac{P_L(\mathbf{v} \mid \mathbf{u})}{\omega_L(\mathbf{v})} = \sum_{l=1}^{L} I(u_l; v_l), \tag{9.3.5}$$

где

$$\omega_L(\mathbf{v}) = \sum_{\mathbf{u}} Q_L(\mathbf{u})\, P_L(\mathbf{v} \mid \mathbf{u}) = \prod_{l=1}^{L} \omega(v_l). \tag{9.3.6}$$

Другой ансамбль — это ансамбль кодов, в котором M кодовых слов выбраны независимо, с распределением вероятностей $\omega_L(\mathbf{v})$, последовательность источника выбрана с тем же распределением вероятностей $Q_L(\mathbf{u})$, как и выше, и для каждого кода в ансамбле $\mathbf{u}$ отображается в такое $\mathbf{v}_m$, обозначаемое $\mathbf{v}(\mathbf{u})$, которое минимизирует $D(\mathbf{u}; \mathbf{v}_m)$ на $1 \leqslant m \leqslant M$.

Лемма 9.3.1. Пусть для заданных источника, меры искажения и тест-канала $P_c[D > L\hat{d}]$ — вероятность в указанном выше ансамбле кодов с M кодовыми словами длины L того, что $D[\mathbf{u}; \mathbf{v}(\mathbf{u})]$ (т. е. искажение между $\mathbf{u}$ и кодовым словом, в которое оно отображается), превосходит заданное число $L\hat{d}$. Тогда

$$P_c[D > L\hat{d}] \leqslant P_t(A) + \exp\left[-Me^{-L\hat{R}}\right], \tag{9.3.7}$$

где $\hat{R}$ — произвольное положительное число; A — множество $\mathbf{u}, \mathbf{v}$ последовательностей, задаваемое соотношением

$$A = \{\mathbf{u}; \mathbf{v}: \text{ или } I(\mathbf{u}; \mathbf{v}) > L\hat{R} \text{ или } D(\mathbf{u}; \mathbf{v}) > L\hat{d}\}, \tag{9.3.8}$$

и $P_t(A)$ — вероятность A в ансамбле тест-канала.

Прежде чем доказывать лемму, используем ее для доказательства следующей теоремы кодирования источника.

Теорема 9.3.1. Пусть $R(d^*)$ — скорость как функция искажения для дискретного источника без памяти с конечной мерой искажения. Для любого $d^* \geqslant 0$, любого $\delta > 0$ и любой достаточно большой длины блока L существует код источника с $M \leqslant \exp\{L[R(d^*) + \delta]\}$ кодовыми словами, для которого среднее искажение на букву удовлетворяет неравенству

$$\bar{d}_L \leqslant d^* + \delta. \tag{9.3.9}$$

Доказательство. Применим лемму к тест-каналу, для которого достигается $R(d^*)$ при заданном d^*, выбирая $\hat{d}$ равным $d^* + \delta/2$ и $\hat{R}$ равным $R(d^*) + \delta/2$. Среднее искажение на букву $\bar{\bar{d}}_L$ по ансамблю кодов леммы удовлетворяет соотношению

$$\bar{\bar{d}}_L \leqslant d^* + \delta/2 + P_c[D > L(d^* + \delta/2)] \max_{k,\, j} d(k; j). \tag{9.3.10}$$

Это соотношение вытекает из того, что искажение на букву ограничивается сверху величиной $d^* + \delta/2$, когда $D[\mathbf{u}, \mathbf{v}(\mathbf{u})] \leqslant L(d^* + \delta/2)$, и величиной $\max_{k,\, j} d(k; j)$ в других случаях. Для $M = \lfloor \exp[LR(d^*) + L\delta] \rfloor$ имеем

$$Me^{-L\hat{R}} \geqslant \{\exp[LR(d^*) + L\delta] - 1\} \times$$
$$\times \exp\{-L[R(d^*) + \delta/2]\} \geqslant e^{L\delta/2} - 1.$$

Таким образом, (9.3.7) принимает вид

$$P_c[D > L(d^* + \delta/2)] \leqslant P_t(A) + \exp(-e^{L\delta/2} + 1). \tag{9.3.11}$$

В соответствии с определением A

$$P_t(A) \leqslant \Pr\{I(\mathbf{u}; \mathbf{v}) > L[R(d^*) + \delta/2]\} +$$
$$+ \Pr[D(\mathbf{u}; \mathbf{v}) > L(d^* + \delta/2)], \tag{9.3.12}$$

где вероятности справа берутся по ансамблю тест-канала. Так как каждая из величин $I(\mathbf{u}; \mathbf{v})$ и $D(\mathbf{u}; \mathbf{v})$ равна сумме L независимых одинаково распределенных случайных величин со средними $R(d^*)$ и d^* соответственно, то по неравенству Чебышева

$$P_t(A) \leqslant \frac{4\sigma_I^2}{L\delta^2} + \frac{4\sigma_d^2}{L\delta^2}, \tag{9.3.13}$$

где σ_I^2 — дисперсия $I(u; v)$ и σ_d^2 — дисперсия $d(u; v)$ для источника, порождающего одну букву, и вероятностей тест-канала. Отсюда

$$P_c[D > L(d^* + \delta/2)] \leqslant \frac{4\sigma_I^2}{L\delta^2} + \frac{4\sigma_d^2}{L\delta^2} + \exp(-e^{L\delta/2} + 1). \tag{9.3.14}$$

Из этого неравенства видно, что последнее выражение в (9.3.10) стремится к 0 с возрастанием L для любого $\delta > 0$. Следовательно, для достаточно больших L

$$\bar{\bar{d}}_L \leqslant d^* + \delta. \tag{9.3.15}$$

Так как по крайней мере одно кодовое слово в ансамбле должно иметь искажение столь же малое, как среднее искажение, то теорема доказана. |

Заметим, что кодовые слова данного кода могут быть представлены $\lceil L\,[R\,(d^*) + \delta]/\ln 2\rceil$ двоичными символами, или не более чем $[R\,(d^*) + \delta]/\ln 2 + 1/L$ двоичными символами на символ источника. Следовательно, взяв L достаточно большим, можно подойти сколь угодно близко к среднему искажению d^*, используя сколь угодно близкое к $R\,(d^*)/\ln 2$ число двоичных символов на символ источника.

Доказательство леммы. Вероятность $P_c\,(D > L\hat{d})$ можно представить в виде

$$P_c\,(D > L\hat{d}) = \sum_{\mathbf{u}} Q_L(\mathbf{u})\, P_c\,(D > L\hat{d}\,|\,\mathbf{u}). \tag{9.3.16}$$

Для заданного $\mathbf{u}$ определим $A_{\mathbf{u}}$ как множество $\mathbf{v}$, для которых пара $\mathbf{u}$, $\mathbf{v}$ содержится в A:

$$A_{\mathbf{u}} = \{\mathbf{v}\text{: или } I\,(\mathbf{u};\mathbf{v}) > L\hat{R} \text{ или } D\,(\mathbf{u};\mathbf{v}) > L\hat{d}\}. \tag{9.3.17}$$

Отметим, что для заданного $\mathbf{u}$ имеем $D\,[\mathbf{u};\ \mathbf{v}\,(\mathbf{u})] > L\hat{d}$ только тогда, когда $D\,(\mathbf{u};\mathbf{v}_m) > L\hat{d}$ при всех m, $1 \leqslant m \leqslant M$, и, следовательно, только тогда, когда $\mathbf{v}_m \in A_{\mathbf{u}}$ при всех m. Так как $\mathbf{v}_m$ выбраны независимо то

$$P_c\,(D > L\hat{d}\,|\,\mathbf{u}) \leqslant \Big[1 - \sum_{\mathbf{v}\in A_{\mathbf{u}}^c} \omega_L(\mathbf{v})\Big]^M, \tag{9.3.18}$$

где сумма берется по всем $\mathbf{v}$ из дополнения $A_{\mathbf{u}}$. Для $\mathbf{v} \in A_{\mathbf{u}}^c$ имеем

$$I\,(\mathbf{u};\mathbf{v}) = \ln \frac{P_L(\mathbf{v}\,|\,\mathbf{u})}{\omega_L(\mathbf{v})} \leqslant L\hat{R}, \tag{9.3.19}$$

$$\omega_L(\mathbf{v}) \geqslant P_L(\mathbf{v}\,|\,\mathbf{u})\, e^{-L\hat{R}}, \tag{9.3.20}$$

$$P_c\,(D > L\hat{d}\,|\,\mathbf{u}) \leqslant \Big[1 - e^{-L\hat{R}} \sum_{\mathbf{v}\in A_{\mathbf{u}}^c} P_L(\mathbf{v}\,|\,\mathbf{u})\Big]^M. \tag{9.3.21}$$

Теперь потребуются следующие неравенства:

$$[1 - \alpha x]^M = \exp\,[M \ln (1 - \alpha x)] \leqslant \exp\,(-M\alpha x), \tag{9.3.22}$$

$$\exp\,(-M\alpha x) \leqslant 1 - x + e^{-M\alpha}, \quad 0 \leqslant x \leqslant 1. \tag{9.3.23}$$

Равенство (9.3.23), очевидно, удовлетворяется при $x = 0$ и $x = 1$ и так как левая часть выпукла $\cup$ по x, а правая часть линейна по x, то оно

удовлетворяется при $0 \leqslant x \leqslant 1$. Применяя эти неравенства к (9.3.21) и полагая x равным сумме по $\mathbf{v}$, выводим

$$P_c(D > L\hat{d} \mid \mathbf{u}) \leqslant 1 - \sum_{\mathbf{v} \in A_{\mathbf{u}}^c} P_L(\mathbf{v} \mid \mathbf{u}) + \exp\left(-Me^{-L\hat{R}}\right). \qquad (9.3.24)$$

Подставляя это выражение в (9.3.16), получаем

$$P_c(D > L\hat{d}) \leqslant \sum_{\mathbf{u}} Q_L(\mathbf{u}) \left[\sum_{\mathbf{v} \in A_{\mathbf{u}}} P_L(\mathbf{v} \mid \mathbf{u}) + \exp\left(-Me^{-L\hat{R}}\right) \right] = \qquad (9.3.25)$$

$$= P_t(A) + \exp\left(-Me^{-L\hat{R}}\right). \qquad (9.3.26)$$

Нетрудно увидеть, что только что доказанная теорема кодирования для источника и различные ее обобщения играют ту же роль в теории передачи с заданным критерием верности, какую играет теорема кодирования для каналов с шумами. Этот результат даже при беглом знакомстве с ним кажется довольно удивительным.

Теорема 9.3.1 рассматривает только меры с конечным искажением, и это ограничение используется в (9.3.10) при оценке искажения для тех маловероятных последовательностей источника, для которых искажение каждого кодового слова больше чем $d^* + \delta/2$. Следующая теорема применима к мерам с бесконечным искажением.

Теорема 9.3.2. Пусть $R(d^*)$ — скорость как функция искажения для дискретного источника без памяти с произвольной мерой искажения отдельной буквы. Для любых $d^* \geqslant 0$, $\delta > 0$ и достаточно большого L существует код источника, для которого энтропия множества кодовых слов $H(\mathbf{V}^L)$ удовлетворяет неравенству $H(\mathbf{V}^L) \leqslant L[R(d^*) + \delta]$ и среднее искажение на букву удовлетворяет неравенству $\bar{d}_L \leqslant d^* + \delta$.

Доказательство. Применим лемму к тест-каналу, на котором достигается $R(d^*)$, выбирая $\hat{d}$ равным $d^* + \delta$ и $\hat{R}$ равным $R(d^*) + \delta/3$. Для ансамбля с $M = \lfloor \exp[LR(d^*) + 2L\delta/3] \rfloor$ кодовыми словами, как и в (9.3.14), имеем

$$P_c[D > L(d^* + \delta)] \leqslant \frac{9\sigma_I^2}{L\delta^2} + \frac{\sigma_d^2}{L\delta^2} + \exp\left(-e^{L\delta/3} + 1\right). \qquad (9.3.27)$$

Имеем $\sigma_d^2 < \infty$, поскольку тест-канал всем переходам с бесконечным искажением соотносит вероятность, равную нулю. В любом частном коде этого ансамбля имеется множество B последовательностей источника, которые не могут быть закодированы с искажением $d^* + \delta$ или меньшим на букву, и должен быть по крайней мере один код, для которого $P(B)$ ограничена сверху правой частью (9.3.27). Возьмем этот код и добавим к нему по одному кодовому слову $\mathbf{v}$ для каждого $\mathbf{u}$ из B, выбирая $\mathbf{v}$ так, чтобы $D(\mathbf{u}; \mathbf{v}) = 0$. Отображая каждое $\mathbf{u}$, не принадлежащее B, в одно из первоначальных M слов с искажением на букву $d^* + \delta$ или меньшем и отображая каждое $\mathbf{u}$ из B в одно из новых кодовых слов с искажением 0, имеем $\bar{d}_L \leqslant d^* + \delta$. Первоначальные M слов кода имеют общую вероятность $1 - P(B)$. Энтропия множества кодовых слов оценивается сверху с помощью предположения, что все перво-

начальные кодовые слова равновероятны и все новые слова равновероятны. Так как имеется не более K^L новых слов, то

$$H(\mathbf{V}^L) \leqslant [1-P(B)] \ln M + P(B) \ln K^L + \mathscr{H}[P(B)] \leqslant$$
$$\leqslant L\left\{R(d^*) + 2\delta/3 + P(B)\ln K + \frac{\mathscr{H}[P(B)]}{L}\right\}. \qquad (9.3.28)$$

Так как $P(B)$ стремится к 0 при возрастании L, то можно выбрать L столь большим, что $H(\mathbf{V}^L) \leqslant L[R(d^*) + \delta]$, что завершает доказательство теоремы. |

Из теоремы кодирования источников неравномерными кодами (см. гл. 3) следует, что это множество кодовых слов может быть закодировано двоичным неравномерным кодом и даст среднюю длину не более чем $[H(\mathbf{V}^L) + \ln 2]/(L \ln 2)$ двоичных символов на символ источника.

Подытоживая сказанное, мы видим, что источник может быть закодирован числом двоичных символов на символ источника, сколь угодно мало превышающим $R(d^*)/\ln 2$, и так, что среднее искажение на букву в последовательности, восстановленной для адресата, будет сколь угодно мало превышать d^*.

Основное отличие между бесконечной и конечной мерой искажения возникает, когда мы пытаемся передать закодированный выход источника по каналу с шумами. Для конечных мер искажения существует максимальное искажение, которое может возникнуть, когда будет сделана ошибка при декодировании в канале и, следовательно, вклад ошибок канала в среднее искажение стремится к нулю, когда вероятность ошибки при передаче по каналу стремится к нулю. Для бесконечной меры искажения, если *какое-либо* используемое кодовое слово имеет бесконечное искажение по отношению к какой-либо последовательности источника, то в канале, все переходные вероятности которого отличны от нуля, искажение наступит с ненулевой вероятностью и, следовательно, среднее искажение бесконечно.

Вернемся теперь к теореме 9.3.1 и исследуем, как быстро скорость кода (т. е. $(\ln M)/L$) может сходиться к $R(d^*)$ при возрастании L. Для этой цели удобнее использовать центральную предельную теорему, а не неравенство Чебышева. Имеем

$$\Pr\{I(\mathbf{u};\mathbf{v}) \geqslant L[R(d^*) + \delta/2]\} \approx \Phi\left(-\frac{\sqrt{L}\,\delta}{2\sigma_I}\right), \qquad (9.3.29)$$

$$\Pr[D(\mathbf{u};\mathbf{v}) \geqslant L(d^* + \delta/2)] \approx \Phi\left(-\frac{\sqrt{L}\delta}{2\sigma_d}\right), \qquad (9.3.30)$$

где

$$\Phi(-x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \leqslant \exp\left(-\frac{x^2}{2}\right), \quad x \geqslant 0.$$

Следовательно, $P_t(A)$ в (9.3.12) приближенно оценивается сверху выражением

$$P_t(A) \lesssim \exp\left(-\frac{L\delta^2}{8\sigma_I^2}\right) + \exp\left(-\frac{L\delta^2}{8\sigma_d^2}\right). \qquad (9.3.31)$$

Подставляя это выражение в (9.3.11) и (9.3.10), получаем

$$\bar{\bar{d}}_L \leqslant d^* + \delta/2 + \left[\exp\left(-\frac{L\delta^2}{8\sigma_I^2}\right) + \right.$$
$$\left. + \exp\left(-\frac{L\delta^2}{8\sigma_d^2}\right) + \exp\left(-e^{L\delta/2}+1\right)\right] \max_{k;\, j} d(k;\, j).$$

Выбирая*) $\delta = 2(\sigma_I + \sigma_d) \sqrt{(\ln L)/L}$, находим, что $\bar{\bar{d}}_L$ сходится к d^* с возрастанием L, как $\sqrt{(\ln L)/L}$ и $(\ln M)/L$ сходится к $R(d^*)$ таким же образом.

Это означает, что для любой заданной точки на кривой $R(d^*)$ оценка скорости и искажения для кода источника длины L лежит на линии с наклоном 45°, проходящей через заданную точку, и стремится к ней, как $\sqrt{(\ln L)/L}$. Так как рассматриваемое d^* можно изменять с L так, чтобы или $\bar{d}_L$ ($\bar{d}_L > 0$), или $(\ln M)/L$ было фиксированным, то δ может быть опущено или из границы для M или из границы для $\bar{d}_L$ в теореме 9.3.1 при $\bar{d}_L > 0$. Легко видеть, что та же скорость сходимости имеет место и в теореме 9.3.2.

Используя более тонкие методы, Пилк (1967) установил существование кодов со скоростью $R(d^*)$, для которых искажение сходится к d^* с возрастанием длины блока, как $(\ln L)/L$.

9.4. ВЫЧИСЛЕНИЕ $R(d^*)$

Обратимся теперь к нахождению множества переходных вероятностей (т. е. тест-канала), на которых достигается $R(d^*)$. Обычный подход к минимизации функции $\mathcal{I}(\mathbf{Q}; \mathbf{P})$ при ограничении $\bar{d} \leqslant d^*$ состоит в использовании множителя Лагранжа, скажем ρ, и минимизации

$$R_0(\rho, \mathbf{P}) = \mathcal{I} + \rho\bar{d} \stackrel{\Delta}{=} \sum_{k,\, j} Q(k) P(j|k) \left[\ln \frac{P(j|k)}{\omega(j)} + \rho d(k; j)\right] \qquad (9.4.1)$$

по всем выборам множества переходных вероятностей $\mathbf{P}$, где

$$\omega(j) = \sum_k Q(k) P(j|k).$$

Другие ограничения

$$P(j|k) \geqslant 0, \quad \sum_j P(j|k) = 1$$

рассмотрим позднее. Варьируя $\rho \geqslant 0$, мы увидим, что можно найти $R(d^*)$ для всех значений d^*. В этой задаче множитель ρ имеет геометрический смысл величины наклона кривой $R(d^*)$ в точке, порождаемой этим значением ρ.

*) Для того чтобы удостовериться, что ошибка аппроксимации с помощью центральной предельной теоремы стремится к нулю быстрее, чем δ при этом выборе, см. Феллер (1966), т. 2, гл. XVI, § 6, теорема 1.

Для того чтобы показать это, рассмотрим $\mathcal{I}(\mathbf{Q}; \mathbf{P})$ и $\bar{d}$ для любого частного $\mathbf{P}$ как точку на графике с ординатой $\mathcal{I}$ и абсциссой $\bar{d}$. Прямая с наклоном $-\rho$, проходящая через эту точку, пересечет $\mathcal{I}$-ось в точке $\mathcal{I}(\mathbf{Q}; \mathbf{P}) + \rho\bar{d}$. Распределение $\mathbf{P}$, которое минимизирует $R_0(\rho, \mathbf{P})$, будет минимизировать ординату этой точки пересечения оси $\mathcal{I}$, и все точки кривой $R(d^*)$ будут лежать на прямой или выше прямой с наклоном $-\rho$, проходящей через эту точку пересечения. Следовательно, эта прямая касается кривой $R(d^*)$. В дальнейшем будет показано, что для конечной меры искажения наклон $R(d^*)$ не может изменяться разрывно, кроме, быть может, точки $d^*_{\max}$.

Для того чтобы в действительности провести минимизацию $R_0(\rho, \mathbf{P})$ по $\mathbf{P}$, целесообразно ввести множители Лагранжа для ограничений

$$\sum_j P(j \mid k) = 1$$

при всех k. Удобно ввести эти множители в виде $-Q(k) \ln \frac{f_k}{Q(k)}$, что означает, что нужно минимизировать выражение

$$F(\rho, \mathbf{P}, \mathbf{f}) = \sum_{k,j} Q(k) P(j \mid k) \left[\ln \frac{P(j \mid k)}{\sum_i Q(i) P(j \mid i)} + \right.$$
$$\left. + \rho d(k; j) - \ln \frac{f_k}{Q(k)}\right] \tag{9.4.2}$$

и затем выбирать $\mathbf{f} = (f_0, \ldots, f_{K-1})$ так, чтобы

$$\sum_j P(j \mid k) = 1$$

для всех k. Дифференцируя F, получаем

$$\frac{\partial F(\rho, \mathbf{P}, \mathbf{f})}{\partial P(j \mid k)} = Q(k)\left[\ln \frac{P(j \mid k)}{\omega(j)} + \rho d(k; j) - \ln \frac{f_k}{Q(k)}\right], \tag{9.4.3}$$

где $\omega(j) = \sum_k Q(k) P(j \mid k)$.

Следовательно, условия на $\mathbf{P}$, которые приводят к стационарной точке для F, принимают вид (для всех k, j):

$$\ln \frac{P(j \mid k)}{\omega(j)} + \rho d(k; j) - \ln \frac{f_k}{Q(k)} = 0, \tag{9.4.4}$$

$$P(j \mid k) = \frac{\omega(j) f_k}{Q(k)} e^{-\rho d(k; j)}. \tag{9.4.5}$$

Если умножить обе части (9.4.5) на $Q(k)$, просуммировать по k и сократить на

$$\omega(j) = \sum_k Q(k) P(j \mid k),$$

то получим

$$1 = \sum_k f_k e^{-\rho d(k; j)} \quad \text{для всех } j. \tag{9.4.6}$$

Точно так же, суммируя (9.4.5) по j и используя ограничение $\sum_j P(j|k) = 1$, получаем

$$1 = \frac{f_k}{Q(k)} \sum_j \omega(j)\, e^{-\rho d(k;\, j)} \quad \text{для всех } k. \tag{9.4.7}$$

Равенства (9.4.6) дают J линейных уравнений относительно переменных f_k, а (9.4.7) дают K линейных уравнений относительно $\omega(j)$. Если $\mathcal{J} = K$, то обычно эти уравнения можно решить и затем найти $P(j|k)$ из (9.4.5). Так как $\mathcal{J}(\mathbf{Q}; \mathbf{P})$ выпукло $\smile$ по $\mathbf{P}$, то $F(\rho, \mathbf{P}, \mathbf{f})$ также выпукло $\smile$ и решение дает минимум.

Трудность указанного выше подхода заключается в том, что получающиеся $P(j|k)$ не обязательно будут неотрицательными. В следующей теореме этот подход излагается строго с учетом ограничения $P(j|k) \geqslant 0$. В ней даются необходимые и достаточные условия, налагаемые на множество переходных вероятностей $\mathbf{P}$, которые минимизируют $R_0(\rho, \mathbf{P})$ и определяют удобную нижнюю границу для $R_0(\rho, \mathbf{P})$ и, следовательно, для $R(d^*)$.

Теорема 9.4.1. Для заданного источника с энтропией $H(U)$ и заданной меры искажения пусть

$$R_0(\rho, \mathbf{P}) = \sum_{k;\, j} Q(k) P(j|k) \left[\ln \frac{P(j|k)}{\sum_i Q(i) P(j|i)} + \rho d(k; j)\right].$$

Здесь и всюду в дальнейшем сумма берется только по тем k, j, для которых $\mathbf{P}(j|k) > 0$. Тогда для любого $\rho > 0$

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) = H(U) + \max_{\mathbf{f}} \sum_k Q(k) \ln f_k, \tag{9.4.8}$$

где максимум берется по всем $\mathbf{f} = (f_0, \ldots, f_{K-1})$ с неотрицательными компонентами, удовлетворяющими ограничениям

$$\sum_k f_k e^{-\rho d(k;\, j)} \leqslant 1, \quad 0 \leqslant j \leqslant J-1. \tag{9.4.9}$$

Необходимые и достаточные условия, накладываемые на $\mathbf{f}$, при которых достигается минимум в (9.4.8), сводятся к тому, что существует множество неотрицательных чисел $\omega(0), \ldots, \omega(J-1)$, удовлетворяющих (9.4.7), и что (9.4.9) удовлетворяется с равенством для каждого j с $\omega(j) > 0$. Выраженный через $\mathbf{f}$ и $\boldsymbol{\omega}$ вектор $\mathbf{P}$, задаваемый (9.4.5), минимизирует $R_0(\rho, \mathbf{P})$.

Обсуждение. Одно из следствий из (9.4.8) заключается в том, что для любого множества $f_k \geqslant 0$, которое удовлетворяет (9.4.9),

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) \geqslant H(U) + \sum_k Q(k) \ln f_k. \tag{9.4.10}$$

Так как уже было показано, что

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) - \rho d^* \leqslant R(d^*),$$

то это означает, что для любого $\rho \geqslant 0$ и любого $\mathbf{f}$, удовлетворяющего (9.4.9) для этого ρ,

$$R(d^*) \geqslant H(U) + \sum_k Q(k) \ln f_k - \rho d^*. \tag{9.4.11}$$

Это соотношение задает нижнюю границу для $R\ (d^*)$ и фактически дает $R\ (d^*)$, если ρ и $\mathbf{f}$ выбраны оптимально.

Попытка фактического нахождения $\min\limits_{\mathbf{P}} R_0\ (\rho, \mathbf{P})$ приводит нас к ситуации, весьма похожей на ту, которая была при отыскании пропускной способности. Можно попытаться решить (9.4.9) относительно $\mathbf{f}$, сначала угадывая множество j, для которых должно иметь место равенство; затем, используя полученное $\mathbf{f}$ для решения (9.4.7) относительно $\omega\ (j)$ и после этого, находя $P\ (j \mid k)$ с помощью (9.4.5). Суммируя (9.4.5) по j и используя (9.4.7), видим, что $P\ (j \mid k)$ действительно являются переходными вероятностями. Умножая (9.4.5) на $Q\ (k)$, суммируя по k и используя (9.4.9), находим, что $\omega\ (j)$ действительно являются выходными вероятностями $\sum\limits_k Q\ (k) P\ (j \mid k)$. Хотя (9.4.8) выражает $\min R_0\ (\rho, \mathbf{P})$ только через $\mathbf{f}$, нужно еще решить (9.4.7), чтобы убедиться, что $\omega\ (j) \geqslant 0$ и что (9.4.9) удовлетворяется с равенством для каждого j с $\omega\ (j) > 0$. В следующем параграфе эта процедура будет применена к важному примеру. Другой подход к нахождению $\min R_0\ (\rho, \mathbf{P})$ в случае, если источник и мера искажения достаточно симметричны, состоит в угадывании решения и проверке, что это решение удовлетворяет условиям теоремы. И в еще одном последнем подходе используется то, что так как $\sum\limits_k Q\ (k) \ln f_k$ выпукло $\frown$ по $\mathbf{f}$, то $R_0\ (\rho, \mathbf{P})$ выпукло $\smile$ по $\mathbf{P}$ и множества ограничений выпуклы; при этом относительно легко вычислить минимум или максимум численными методами с помощью вычислительной машины.

Следует отметить, что в силу строгой выпуклости $\sum\limits_k Q\ (k) \ln f_k$ по $\mathbf{f}$ значение $\mathbf{f}$, на котором достигается максимум в (9.4.8), является единственным. Вместе с тем $\mathbf{P}$ и $\boldsymbol{\omega}$ не обязательно являются единственными.

Для того чтобы дать простую интерпретацию максимизирующего значения $\mathbf{f}$ в теореме, заметим, что для $\mathbf{P}$ и $\mathbf{f}$, удовлетворяющих (9.4.5), вероятность входа при заданном выходе (иногда называемая обратной переходной вероятностью) равна

$$P_b\ (k \mid j) = f_k\, e^{-\rho d\ (k; j)}.$$

Тогда соотношение (9.4.9) с равенством представляет собой условие, что эти выражения действительно являются условными вероятностями, а (9.4.7) представляет собой условие, что эти обратные вероятности согласуются с входными вероятностями.

Доказательство теоремы. Сначала покажем, что для всех $\mathbf{P}$ и $\mathbf{f}$, удовлетворяющих заданным ограничениям,

$$R_0\ (\rho, \mathbf{P}) \geqslant H\ (U) + \sum_k Q\ (k) \ln f_k. \tag{9.4.12}$$

Рассмотрим функцию

$$F\ (\rho, \mathbf{P}, \mathbf{f}) = \sum_{k,\ j} Q\ (k)\, P(j \mid k) \left[\ln \frac{P\ (j \mid k)}{\sum\limits_i Q\ (i)\, P\ (j \mid i)} + \right.$$

$$+\rho d(k;\ j)-\ln\frac{f_k}{Q(k)}\Big]. \tag{9.4.13}$$

Отделяя третье выражение в (9.4.13) от первых двух и суммируя по j, получаем

$$F(\rho,\ \mathbf{P},\ \mathbf{f})=R_0(\rho,\ \mathbf{P})-\sum_k Q(k)\ln\frac{f_k}{Q(k)}=R_0(\rho,\ \mathbf{P})-$$

$$-H(U)-\sum_k Q(k)\ln f_k. \tag{9.4.14}$$

Покажем теперь, что $F(\rho,\ \mathbf{P},\ \mathbf{f})$ неотрицательна, используя обычное неравенство $\ln x\leqslant x-1$. Из (9.4.13) имеем

$$-F(\rho,\mathbf{P},\mathbf{f})=\sum_{k,j}Q(k)P(j\mid k)\ln\left[\frac{\omega(j)f_k}{P(j\mid k)Q(k)}e^{-\rho d(k\mid j)}\right]\leqslant \tag{9.4.15}$$

$$\leqslant\sum_{k,j}\omega(j)f_k e^{-\rho d(k;\ j)}-\sum_{k,j}Q(k)P(j\mid k). \tag{9.4.16}$$

Суммируя сначала по k и используя затем ограниченне (9.4.9), получаем

$$-F(\rho,\ \mathbf{P},\ \mathbf{f})\leqslant\sum_j\omega(j)-\sum_j\omega(j)=0. \tag{9.4.17}$$

Из (9.4.17) и (9.4.14) выводнм (9.4.12).

Неравенство (9.4.12) переходит в равенство тогда и только тогда, когда оба неравенства, ведущие от (9.4.15) к (9.4.17), переходят в равенства, или тогда и только тогда, когда

$$\frac{\omega(j)f_k}{P(j\mid k)Q(k)}e^{-\rho d(k;\ j)}=1\ \text{для всех}\ P(j\mid k)>0, \tag{9.4.18}$$

$$\sum_k f_k e^{-\rho d(k;\ j)}=1\ \text{для всех}\ \omega(j)>0. \tag{9.4.19}$$

Если обе части (9.4.18) умножить на $P(j\mid k)$ и просуммировать по j, то получим (9.4.7), так что условия теоремы необходимы для равенства в (9.4.12). Аналогично, если неотрицательные числа $\omega(0),\ \ldots,\ \omega(J-1)$ удовлетворяют (9.4.7) и если (9.4.19) удовлетворяется, то, как было уже показано ранее, $P(j\mid k)$, заданные формулой (9.4.5), являются переходными вероятностями с выходными вероятностями $\omega(j)$. В соответствии с (9.4.5) выбранные значения удовлетворяют (9.4.18), так что условия теоремы достаточны для равенства в (9.4.12).

Для завершения доказательства следует показать существование $\mathbf{P}$ и $\mathbf{f}$, удовлетворяющих заданным ограничениям и удовлетворяющих (9.4.12) с равенством. Пусть $\mathbf{P}$ минимизирует*) $R_0(\rho,\ \mathbf{P})$ в допустимой области, в которой $\mathbf{P}$ — множество переходных вероятностей. Для каждого k выберем некоторое j, скажем $j(k)$, для которого $P(j\mid k)>0$, и определим f_k равенством

$$\ln\frac{P[j(k)\mid k]}{\omega[j(k)]}+\rho d(k;\ j(k))-\ln\frac{f_k}{Q(k)}=0. \tag{9.4.20}$$

*) Такое $\mathbf{P}$ должно существовать, так как область, по которой проводится минимизация, замкнута и ограничена, и R_0 непрерывна по $\mathbf{P}$ в этой области.

В оставшейся части доказательства будем считать это $\mathbf{f} = (f_0, \ldots, f_{K-1})$ фиксированным. Рассмотрим функцию $F(\rho, \mathbf{P}, \mathbf{f})$, определенную в (9.4.13), и заметим, что из (9.4.14) следует, что так как рассматриваемое $\mathbf{P}$ минимизирует $R_0(\rho, \mathbf{P})$, то $\mathbf{P}$ минимизирует также $F(\rho, \mathbf{P}, \mathbf{f})$. Для любых k, j с $\omega(j) > 0$ и $d(k; j) < \infty$ имеем

$$\frac{\partial F(\rho, \mathbf{P}, \mathbf{f})}{\partial P(j \mid k)} = \left[\ln \frac{P(j \mid k)}{\omega(j)} + \rho d(k; j) - \ln \frac{f_k}{Q(k)}\right] Q(k). \quad (9.4.21)$$

Если это выражение отрицательно (положительно) для каких-либо k, j (с $\omega(j) > 0$, $d(k, j) < \infty$), то приращение, увеличивающее (уменьшающее) это $P(j \mid k)$, и приращение, уменьшающее (увеличивающее) $P(j(k) \mid k)$ (для которого $\partial F / \partial P[j(k) \mid k] = 0$), будет приводить к убыванию $F(\rho, \mathbf{P}, \mathbf{f})$ в противоречии с предположением, что $\mathbf{P}$ минимизирует $F(\rho, \mathbf{f}, \mathbf{P})$. Отсюда следует, что для минимизирующих $\mathbf{P}$

$$\ln \frac{P(j \mid k)}{\omega(j)} + \rho d(k; j) - \ln \frac{f_k}{Q(k)} = 0 \quad (9.4.22)$$

для всех k, j с $\omega(j) > 0$, $d(k; j) < \infty$. Также $P(j \mid k) = 0$, если или $\omega(j) = 0$, или $d(k; j) = \infty$. Следовательно, для рассматриваемых $\mathbf{P}$, $\mathbf{f}$ (9.4.5) удовлетворяется для всех k, j. Умножая (9.4.5) на $Q(k)$ и суммируя по k, видим, что (9.4.9) удовлетворяется с равенством для j, таких, что $\omega(j) > 0$. Таким образом, $\mathbf{P}$ и $\mathbf{f}$ удовлетворяют достаточным условиям для равенства в (9.4.12). Доказательство еще не закончено, так как следует показать, что $\mathbf{f}$ лежит в области, соответствующей заданным ограничениям, т. е. что (9.4.9) также удовлетворяется для j, таких, что $\omega(j) = 0$. Предположим, что $\omega(j') = 0$ для заданного j', и определим $\mathbf{P}'$ через минимизирующее $\mathbf{P}$ следующими равенствами:

$$P'(j' \mid k) = \varepsilon \frac{f_k}{Q(k)} e^{-\rho d(k; j)} \text{ для всех } k, \quad (9.4.23)$$

$$P'(j(k) \mid k) = P(j(k) \mid k) - P'(j' \mid k) \text{ для всех } k, \quad (9.4.24)$$

$$P'(j \mid k) = P(j \mid k) \text{ для всех других } j, k. \quad (9.4.25)$$

Для достаточно малых $\varepsilon > 0$ вектор $\mathbf{P}'$ представляет собой множество переходных вероятностей. Пусть $F_j(\rho, \mathbf{P}, \mathbf{f})$ задается равенством

$$F_j(\rho, \mathbf{P}, \mathbf{f}) = \sum_k Q(k) P(j \mid k) \left[\ln \frac{P(j \mid k)}{\omega(j)} + \rho d(k; j) - \ln \frac{f_k}{Q(k)}\right]. \quad (9.4.26)$$

Тогда

$$F(\rho, \mathbf{P}', \mathbf{f}) - F(\rho, \mathbf{P}, \mathbf{f}) = F_{j'}(\rho, \mathbf{P}', \mathbf{f}) + \sum_{j \neq j'} [F_j(\rho, \mathbf{P}', \mathbf{f}) - F_j(\rho, \mathbf{P}, \mathbf{f})]. \quad (9.4.27)$$

Так как $\partial F_j(\rho, \mathbf{P}, \mathbf{f}) / \partial P(j(k) \mid k) = 0$, то, очевидно, в (9.4.27) сумма по $j \neq j'$ не имеет членов первого порядка по ε, так что с точностью до членов первого порядка по ε

$$F(\rho, \mathbf{P}', \mathbf{f}) - F(\rho, \mathbf{P}, \mathbf{f}) = \sum_k Q(k) P'(j' \mid k) \ln \left[\frac{\varepsilon}{\sum_k Q(k) P'(j' \mid k)}\right] =$$

$$= \sum_k Q(k) P'(j' \mid k) \ln \frac{1}{\sum_k f_k e^{-\rho d(k; j')}}. \qquad (9.4.28)$$

Так как **P** минимизирует $F(\rho, \mathbf{P}, \mathbf{f})$, то правая часть (9.4.28) неотрицательна, откуда следует

$$\sum_k f_k e^{-\rho d(k; j')} \leqslant 1,$$

что завершает доказательство теоремы. |

Следствие 9.4.1. Для конечной меры искажения с $d^*_{\max} > 0$ наклон $R(d^*)$ непрерывен при $0 < d^* < d^*_{\max}$ и стремится к $-\infty$ при d^*, стремящемся к 0.

Доказательство. Как было показано, $R(d^*)$ выпукла $\smile$ и $\min_{\mathbf{P}} R_0(\rho, \mathbf{P})$ является точкой пересечения оси R с касательной к $R(d^*)$ с наклоном $-\rho$. Если наклон имеет разрывы или если он сходится к конечному пределу при $d^* \to 0$, то на кривой $R(d^*)$ должна быть точка, которая является точкой касания различных касательных с наклонами из некоторой области. Любое **P**, при котором достигается эта точка на кривой $R(d^*)$, должно тогда минимизировать $R_0(\rho, \mathbf{P})$ для этой области наклонов. Для завершения доказательства остается показать, что если **P** минимизирует $R_0(\rho, \mathbf{P})$ для двух различных значений ρ, скажем ρ и ρ', то $\mathcal{I}(\mathbf{Q}; \mathbf{P}) = 0$. Пусть $\mathbf{f}'$ — максимизирующее **f** для ρ'; имеем

$$P(j \mid k) = \frac{\omega(j) f_k}{Q(k)} e^{-\rho d(k; j)} = \frac{\omega(j)}{Q(k)} f'_k e^{-\rho' d(k; j)}. \qquad (9.4.29)$$

Следовательно, если $\omega(j) > 0$, то

$$\frac{f_k}{f'_k} = e^{(\rho - \rho') d(k; j)}. \qquad (9.4.30)$$

Это означает, что $d(k; j)$ не зависит от j для всех j с $\omega(j) > 0$ и, следовательно, как вытекает из (9.4.29), $Q(k) P(j \mid k) = \omega(j) \alpha(k)$, где $\alpha(k) = f_k \exp[-\rho(d; j)]$ не зависит от j. Следовательно, $P(j \mid k) = \omega(j)$ и $\mathcal{I}(\mathbf{Q}; \mathbf{P}) = 0$. |

Задача 9.1 дает пример кривой $R(d^*)$, у которой наклон теряет непрерывность при d^*_{max}, а задача 9.4 показывает, что ограничение на конечность меры искажения в следствии является необходимым.

Следствие 9.4.2. $R(d^*)$ является непрерывной функцией d^* при $d^* \geqslant 0$.

Доказательство. Так как $R(d^*)$ выпукла $\smile$ и не возрастает с d^* при $d^* \geqslant 0$, то, как известно, функция $R(d^*)$ должна быть всюду непрерывна, за исключением, быть может, точки $d^* = 0$. Итак, поскольку $R(d^*) \leqslant H(U)$ для $d^* \geqslant 0$ и поскольку $R(d^*)$ не может убывать с убыванием d^*, то, как известно, существует

$$\lim_{d^* \to 0^+} R(d^*).$$

Следовательно, остается только показать, что

$$R(0) = \lim_{d^* \to 0+} R(d^*).$$

Для этого рассмотрим новую меру искажения $d'(k; j)$, задаваемую равенством

$$d'(k; j) = \begin{cases} 0, & \text{если } d(k; j) = 0, \\ \infty & \text{во всех других случаях.} \end{cases} \tag{9.4.31}$$

Скорость как функция искажения для $d'(k; j)$ равна $R'(d^*) = R(0)$, $d^* \geqslant 0$, так как среднее искажение для любого $\mathbf{P}$, равно или 0 или ∞, и $R(0)$ является минимумом $\mathcal{I}(\mathbf{Q}; \mathbf{P})$, для которого среднее искажение равно 0. Следовательно, функция

$$\min_{\mathbf{P}} R'_0(\rho, \mathbf{P})$$

для $d'(k; j)$ также равна $R(0)$ при всех $\rho > 0$, и, используя (9.4.8) для

$$\min_{\mathbf{P}} R'_0(\rho, \mathbf{P})$$

при любом $\rho > 0$, получаем

$$R(0) = H(U) + \max_{f'} \sum_k Q(k) \ln f'_k \tag{9.4.32}$$

при ограничении

$$\sum_{k:d(k;j)=0} f'_k \leqslant 1 \text{ для всех } j. \tag{9.4.33}$$

Взяв ρ достаточно большим, можно аппроксимировать $\mathbf{f}'$, которое максимизирует (9.4.32) при условии (9.4.33), сколь угодно точно с помощью $\mathbf{f}(\rho)$, которое удовлетворяет

$$\sum_k f_k(\rho)\, e^{-\rho d(k;j)} \leqslant 1 \text{ для всех } j. \tag{9.4.34}$$

Следовательно, для любого $\varepsilon > 0$ можно выбрать ρ достаточно большим, так что

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) \geqslant H(U) + \sum_k Q(k) \ln f_k(\rho) \geqslant R(0) - \varepsilon. \tag{9.4.35}$$

Так как

$$R(d^*) \geqslant \min_{\mathbf{P}} R_0(\rho, \mathbf{P}) - \rho d^*,$$

то имеем

$$\min_{d^* \to 0} R(d^*) \geqslant R(0) - \varepsilon \text{ при любом } \varepsilon > 0.$$

Так как $R(d^*) \leqslant R(0)$ для $d^* \geqslant 0$, то это завершает доказательство. |

Следствие 9.4.3. При любом заданном d^* для достижения $R(d^*)$ требуется не более $K + 1$ букв алфавита адресата. Для любой части кривой $R(d^*)$, для которой наклон не является постоянным, можно использовать не более K букв адресата.

Доказательство. Для любого заданного ρ пусть $\mathbf{f}$ максимизирует правую часть (9.4.8) при условии (9.4.9). Из теоремы следует, что $\mathbf{P}$, которое минимизирует $R_0(\rho, \mathbf{P})$, может быть найдено из (9.4.5), где выходные вероятности удовлетворяют равенствам

$$\frac{f_k}{Q(k)} \sum_j \omega(j)\, e^{-\rho d(k;j)} = 1 \text{ для всех } k. \tag{9.4.36}$$

Это является множеством K линейных уравнений с J неизвестными и имеется по крайней мере одно решение с $\omega(j) \geqslant 0$. А тогда, точно так же как в следствии 3 § 4.5, доказывается, что имеется решение, для которого только K чисел $\omega(j) > 0$. Если прямая $\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) - \rho d^*$ касается кривой $R(d^*)$ только в одной точке, то это решение должно дать эту точку. Вместе с тем, если $\min R_0(\rho, \mathbf{P}) - \rho d^*$ является касательной к $R(d^*)$ на целом интервале, то для того чтобы найти $\mathbf{P}$, которое приводит к заданному d^*, требуется дополнительное ограничение

$$\sum_{k,j} Q(k) P(j \mid k)\, d(k; j) = d^*.$$

Используя (9.4.5) и (9.4.9) с равенством при $\omega(j) > 0$, получаем

$$\sum_j \omega(j) \sum_k f_k e^{-\rho d(k;j)}\, d(k; j) = d^*. \tag{9.4.37}$$

Из (9.4.37) и (9.4.36) получаем $K + 1$ линейных уравнений с J неизвестными и, как раньше, должно существовать решение, для которого только $K + 1$ из $\omega(j)$ положительны. |

9.5. МОДИФИКАЦИЯ ОБРАЩЕНИЯ ТЕОРЕМЫ КОДИРОВАНИЯ ДЛЯ КАНАЛА С ШУМАМИ

Как было уже отмечено, функция искажения

$$d(k; j) = \begin{cases} 0, & k = j, \\ 1 & \text{во всех других случаях,} \end{cases} \tag{9.5.1}$$

при $K = J$ является приемлемой мерой искажения для изучения ошибок при воспроизведении выхода источника. Действительно, средняя вероятность ошибки на букву источника $\langle P_e \rangle$, которая изучалась в гл. 4, является просто средним искажением на букву для указанной выше меры искажения. В этом параграфе вычисляется $R(d^*)$ для произвольного дискретного источника без памяти и этой меры искажения и, таким образом, отыскивается минимальная вероятность ошибки на букву источника, которая может быть достигнута для скоростей источника, бóльших пропускной способности. Будет показано, что нижняя граница $\langle P_e \rangle$, приведенная в гл. 4, равна фактически минимуму вероятности ошибки для некоторого диапазона пропускных способностей каналов, меньших энтропии источника. Это вычисление будет иметь некоторый дополнительный смысл, так как будет показывать, как

результаты последней главы могут быть применены. Неравенства, задающие ограничения

$$\sum_k f_k e^{-\rho d(k;j)} \leqslant 1,$$

упрощаются для этой меры искажения и имеют вид

$$f_j + \left[\left(\sum_{k=1}^{K-1} f_k\right) - f_j\right] e^{-\rho} \leqslant 1;\ 0 \leqslant j \leqslant J-1. \tag{9.5.2}$$

Из симметрии следует, что все эти неравенства могут быть удовлетворены с равенством, если положить все f_k равными одному и тому же значению, скажем f_0. Тогда

$$f_k = f_0 = [1 + (K-1)e^{-\rho}]^{-1}. \tag{9.5.3}$$

Используя это **f** в нижней границе (9.4.10), получаем

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) \geqslant H(U) - \ln[1 + (K-1)e^{-\rho}]. \tag{9.5.4}$$

Тогда для всех $\rho > 0$ имеем

$$R(d^*) \geqslant -\rho d^* + H(U) - \ln[1 + (K-1)e^{-\rho}]. \tag{9.5.5}$$

Правая часть максимизируется по ρ, в результате давая точную границу при ρ, удовлетворяющем равенству

$$d^* = \frac{(K-1)e^{-\rho}}{1 + (K-1)e^{-\rho}}, \tag{9.5.6}$$

$$\rho = \ln(K-1) + \ln\frac{1-d^*}{d^*}. \tag{9.5.7}$$

Подставляя (9.5.7) в (9.5.5) и производя преобразования, выводим

$$R(d^*) \geqslant H(U) - \mathcal{H}(d^*) - d^* \ln(K-1), \tag{9.5.8}$$

где

$$\mathcal{H}(d^*) = -d^* \ln d^* - (1-d^*) \ln(1-d^*). \tag{9.5.9}$$

В соединении с теоремой 9.2.2 результат (9.5.8) эквивалентен для дискретных источников без памяти обращению теоремы кодирования для канала с шумом, т. е. теореме 4.3.4. Для того чтобы достигнуть вероятность ошибки на символ источника, равную d^*, канал должен иметь пропускную способность, по крайней мере, равную правой части (9.5.8), или, что эквивалентно, для канала такой пропускной способности, d^* — нижняя граница вероятности ошибки на символ.

Теперь найдем условия, при выполнении которых (9.5.8) удовлетворяется и переходит в равенство. Из теоремы 9.4.1 следует, что (9.5.4) удовлетворяется с равенством, если для (9.5.7) существует решение с $\omega(j) \geqslant 0$, которое для этой меры искажения принимает вид

$$\omega(k) + (1 - \omega(k))\, e^{-\rho} = Q(k)/f_k,\ 0 \leqslant k \leqslant K-1. \tag{9.5.10}$$

Для $f_k = f_0 = [1 + (K-1)e^{-\rho}]^{-1}$

$$\omega(k) = \frac{Q(k)[1 + (K-1)e^{-\rho}] - e^{-\rho}}{1 - e^{-\rho}}. \tag{9.5.11}$$

Все $\omega(k)$ неотрицательны, если

$$Q(k) \geqslant \frac{1}{e^{\rho}+(K-1)} \text{ для всех } k. \tag{9.5.12}$$

Следовательно, (9.5.4) удовлетворяется с равенством для всех достаточно больших ρ. Так как для каждого $\rho>0$

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) - \rho d^*$$

имеет с кривой $R(d^*)$, по крайней мере, одну общую точку, и так как эта точка задается (9.5.6) для всех ρ, удовлетворяющих (9.5.12), то имеем

$$R(d^*) = H(U) - \mathcal{H}(d^*) - d^* \ln(K-1) \tag{9.5.13}$$

для

$$d^* \leqslant (K-1)Q_{min}, \tag{9.5.14}$$

где (9.5.14) вытекает из (9.5.12) и (9.5.7) и где Q_{min} — наименьшее значение $Q(k)$. Из этого результата и теорем 9.2.2 и 9.3.1 получаем следующую теорему.

Теорема 9.5.1. Пусть задан дискретный источник без памяти с энтропией $H(U)$ нат, с алфавитом объема K и минимальной вероятностью буквы Q_{min} и пусть задано, что источник связан с адресатом дискретным каналом без памяти с пропускной способностью C (в натах на символ источника), тогда для любого $d^* \leqslant (K-1)Q_{min}$ вероятность ошибки на символ источника d^* может быть всегда достигнута с помощью соответствующего кодирования, если

$$C > H(U) - \mathcal{H}(d^*) - d^* \ln(K-1), \tag{9.5.15}$$

и никакими способами не может быть достигнута, если

$$C < H(U) - \mathcal{H}(d^*) - d^* \ln(K-1). \tag{9.5.16}$$

Для простоты в теореме рассматриваются лишь дискретные каналы без памяти. Она, очевидно, остается справедливой всегда, когда C может быть выражена как максимум средней взаимной информации и вместе с тем интерпретирована с помощью теоремы кодирования. Весьма примечательно, что этот минимум вероятности ошибки может быть однозначно определен с помощью столь малого числа параметров.

Теперь вычислим $R(d^*)$ для больших значений d^*. Для простоты обозначений предположим, что вероятности букв упорядочены

$$Q(0) \geqslant Q(1) \geqslant \ldots \geqslant Q(K-1). \tag{9.5.17}$$

Как из физических соображений, так и из рассмотрения выражения (9.5.11) можно предположить, что если какая-либо из выходных вероятностей равна 0, то она должна относиться к тем буквам, которые соответствуют наименее вероятным входам. Таким образом, мы принимаем, что для некоторого m, которое будет выбрано позднее, $\omega(k) = 0$ при $k \geqslant m$ и $\omega(k) > 0$ при $k \leqslant m-1$. Для $k \geqslant m$ равенство (9.5.10) принимает вид

$$Q(k) = f_k e^{-\rho}, \quad k \geqslant m. \tag{9.5.18}$$

Неравенство $f_j + (\Sigma f_k - f_j)e^{-\rho} \leqslant 1$, задающее ограничение, должно удовлетворяться с равенством при $j \leqslant m-1$. Следовательно, все f_j должны быть одни и те же, скажем f_0, при $j \leqslant m-1$ и $f_j \leqslant f_0$ при $j \geqslant m$. Находя выражение, задающее ограничение для $j = 0$, и используя (9.5.18), получаем

$$f_0 [1 + (m-1) e^{-\rho}] + \sum_{k=m}^{K-1} Q(k) = 1, \tag{9.5.19}$$

$$f_k = f_0 = \frac{S_m}{1 + (m-1) e^{-\rho}}, \quad k \leqslant m-1, \tag{9.5.20}$$

где $S_m = \sum_{k=0}^{m-1} Q(k)$.

Из (9.5.18) видим, что $f_m \geqslant f_{m+1} \geqslant \ldots \geqslant f_{K-1}$ и ограничение $f_j \leqslant f_0$ имеет место для $j \geqslant m$, если

$$Q(m) \leqslant \frac{S_m e^{-\rho}}{1 + (m-1) e^{-\rho}}. \tag{9.5.21}$$

Вектор $\mathbf{f}$, задаваемый (9.5.18) и (9.5.20), максимизирует выражение $\sum_k Q(k) \ln f_k$, если все $\omega(k)$, определяемые (9.5.10), неотрицательны. Так как $\omega(m-1) \leqslant \omega(m-2) \leqslant \ldots \leqslant \omega(0)$, то это требует только, чтобы

$$Q(m-1) \geqslant \frac{S_m e^{-\rho}}{1 + (m-1) e^{-\rho}}. \tag{9.5.22}$$

Следовательно, для области ρ, где (9.5.21) и (9.5.22) удовлетворяются, заданное $\mathbf{f}$ приводит к соотношению

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) = H(U) + \sum_{k=0}^{m-1} Q(k) \ln \frac{S_m}{1 + (m-1) e^{-\rho}} + \sum_{k=m}^{K-1} Q(k) \ln [Q(k) e^{\rho}]. \tag{9.5.23}$$

Теперь, зная $\min_{\mathbf{P}} R_0(\rho, \mathbf{P})$ в некоторой области ρ, определяем $R(d^*)$ в соответствующей области наклонов. Параметр ρ связан с d^* соотношением

$$d^* = \frac{\partial \min R_0(\rho, \mathbf{P})}{\partial \rho} = S_m \frac{(m-1) e^{-\rho}}{1 + (m-1) e^{-\rho}} + 1 - S_m. \tag{9.5.24}$$

Для ρ и d^*, связанных (9.5.24), имеем

$$R(d^*) = \min_{\mathbf{P}} R_0(\rho, \mathbf{P}) - \rho d^*.$$

После некоторых преобразований это выражение приводится к виду

$$R(d^*) = S_m [H(U_m) - \mathcal{H}(\hat{d}) - \hat{d} \ln (m-1)], \tag{9.5.25}$$

где $H(U_m)$ — энтропия редуцированного ансамбля с вероятностями $Q(0)/S_m, \ldots, Q(m-1)/S_m$ и $\hat{d}$ определяется равенством

$$\hat{d} = \frac{d^* - (1 - S_m)}{S_m}. \tag{9.5.26}$$

Подставляя (9.5.24) в (9.5.21) и (9.5.22), находим, что это решение справедливо для заданного m, если

$$mQ(m) + \sum_{k=m+1}^{K-1} Q(k) \leqslant d^* \leqslant (m-1)Q(m-1) + \sum_{k=m}^{K-1} Q(k). \tag{9.5.27}$$

Заметим, что для $m = K - 1$ нижняя граница для d^* совпадает с верхней границей для d^*, если d^* удовлетворяет (9.5.13). Также верхняя граница для d^* при каком-либо значении m (отличном от $K - 1$) совпадает с нижней границей при соседнем значении, меньшем m.

Наконец, верхняя граница для d^* при $m = 2$ равна d_{max}, так что для любого d^*, большего чем то значение, для которого (9.5.13) задает $R(d^*)$, (9.5.27) определяет m, для которого (9.5.25) задает $R(d^*)$.

Имеется простое физическое истолкование (9.5.25). Для заданного m мы пытаемся представить источник, используя только буквы адресата от 0 до $m - 1$. С вероятностью $1 - S_m$ источник порождает одну из букв от m до $K - 1$, при этом искажение равно единице и нет никакого смысла в затрате какой-либо информации на эти буквы. Следовательно, минимум средней взаимной информации равняется умноженному на S_m минимуму при условии появления какой-либо из букв от 0 до $m - 1$. Из (9.5.26) следует, что $\hat{d}$ можно интерпретировать как минимум среднего искажения при условии появления одной из букв от 0 до $m - 1$. Для этого условного искажения член в скобках равен как раз минимуму условной средней взаимной информации, что можно было ожидать из (9.5.13).

9.6. ДИСКРЕТНЫЕ ПО ВРЕМЕНИ ИСТОЧНИКИ С НЕПРЕРЫВНЫМИ АМПЛИТУДАМИ

Рассмотрим теперь источники, для которых выход представляет собой последовательность статистически независимых, одинаково распределенных непрерывных, действительных, случайных величин..., u_{-1}, u_0, u_1, ... Предположим, что распределение букв источника в каждый момент описывается плотностью вероятности $q(u)$. Выход источника должен быть представлен у адресата последовательностью действительных чисел ..., v_{-1}, v_0, v_1, ..., и имеется мера искажения $d(u; v)$, принимающая числовые значения для всех действительных чисел u и v и определяющая искажение, если выход источника u представляется у адресата значением v. Примем, как и раньше, что $d(u; v) \geqslant 0$ и что для каждого u имеется v (обычно $v = u$), для которого $d(u; v) = 0$. Наиболее употребительными мерами искажения являются разностные меры искажения, у которых $d(u; v)$ является функцией только разности $v - u$ и наиболее распространенной разностной мерой является $d(u; v) = (v - u)^2$.

Функция $R(d^*)$ *для такого источника определяется равенством*

$$R(d^*) = \inf I(U;\ V), \tag{9.6.1}$$

где нижняя грань берется по всем совместным вероятностным мерам на пространстве UV *при ограничениях, что* $q(u)$ *— плотность вероятности на* U *и что среднее искажение не более* d^*. Такое же определение будет использоваться для $R(d^*)$ и в случае, когда U — произвольное пространство, а не только действительная прямая, и когда вероятностная мера на U (и заданная σ-алгебра подмножеств) произвольна, а не только плотность вероятности.

Для этих источников и мер искажения кривая $R(d^*)$ невозрастающая и выпуклая $\smile$ по d^*. Это можно понять, пользуясь теми же соображениями, что и в дискретном случае, и используя здесь интегралы вместо сумм. Основное различие между $R(d^*)$, рассматриваемым здесь, и $R(d^*)$ для дискретного случая состоит в том, что, как правило, в непрерывном случае $\lim\limits_{d^* \to 0} R(d^*) = \infty$ (ср. рис. 9.2.1 и 9.7.1).

Теорема 9.2.1 (которая утверждает, что если $\bar{d}_L$ — среднее искажение на букву в последовательности из L букв источника, то $(1/L)I(\mathbf{U}^L;\ \mathbf{V}^L) \geqslant R(\bar{d}_L)$) также имеет место, когда U — произвольное пространство с произвольной вероятностной мерой. Однако, когда энтропия $\mathbf{U}^L$ не определена, доказательство должно быть видоизменено следующим образом:

$$I(\mathbf{U}^L;\ \mathbf{V}^L) = \sum_{l=1}^{L} I(U_l;\ \mathbf{V}^L \,|\, U_1 \cdots U_{l-1}),$$

$$I(U_l;\ \mathbf{V}^L \,|\, U_1 \cdots U_{l-1}) = I(U_l;\ \mathbf{V}^L U_1 \cdots U_{l-1}) -$$

$$- I(U_l;\ U_1 \cdots U_{l-1}) = I(U_l;\ \mathbf{V}^L U_1 \cdots U_{l-1}) \geqslant I(U_l;\ V_l),$$

$$I(\mathbf{U}^L;\ \mathbf{V}^L) \geqslant \sum_{l=1}^{L} I(U_l;\ V_l). \tag{9.6.2}$$

В остальном доказательство проводится, как и ранее.

Теорема 9.6.1. Теорема 9.2.2 (обращение теоремы кодирования для источников, связанных с мерой искажения) применима в общем случае ко всем дискретным по времени источникам без памяти с мерой искажения, заданной для отдельных букв.

Доказательство такое, как в теореме 9.2.2, и поэтому опускается.

Теорема кодирования для непрерывных по амплитуде источников также вполне аналогична соответствующей теореме для дискретных источников. Для заданного источника и тест-канала (т. е. заданной вероятностной меры на пространстве UV) опять рассматривается ансамбль независимо выбранных кодовых слов, каждая буква в котором выбирается в соответствии с вероятностной мерой на V. Лемма 9.3.1 применима здесь без изменений. Если источник и тест-канал описываются совместной плотностью вероятности, то доказательство леммы может быть модифицировано просто заменой всех сумм интегралами и всех вероятностей плотностями. Для произвольной совместной

меры вероятности (полагая, что $I(u; v)$ и $D(u; v)$ измеримы и $I(U; V) < \infty$) применимо такое же доказательство, опирающееся на общие результаты теории меры.

Теорема 9.6.2. Пусть $R(d^*)$ — скорость как функция искажения для дискретного по времени источника с мерой искажения $d(u; v)$ и пусть существует конечное множество букв адресата $a_1, \ldots, a_J$, такое, что $\overline{\min_j d(u; a_j)}$ конечно, где математическое ожидание берется по ансамблю источника сообщений. Для любых $d^* > 0$, $\delta > 0$ и достаточно большого L существует код источника, для которого энтропия множества кодовых слов удовлетворяет неравенству $H(\mathbf{V}^L) \leqslant L[R(d^*) + \delta]$ и среднее искажение на букву удовлетворяет неравенству $\bar{d}_L \leqslant d^* + \delta$. Кроме того, если $\ln J < R(d^*)$, то существует код с $M' \leqslant \exp[LR(d^*) + L\delta]$ кодовыми словами и средним искажением $\bar{d}_L \leqslant d^* + \delta$.

Обсуждение. Интерпретация $H(\mathbf{V}^L)$ и M с помощью числа двоичных символов на символ источника, требуемых для представления источника, аналогична той, которая была в дискретном случае. Условие $\overline{\min_j d(u; a_j)} < \infty$ означает, что имеется способ разбиения (или квантования) выхода источника на конечное число областей с конечным средним искажением. Если это условие не имеет места, то любой блоковый код с конечным числом кодовых слов должен иметь бесконечное среднее искажение, так как он использует лишь конечное множество букв адресата.

Доказательство. Выберем тест-канал (т. е. совместную вероятностную меру), для которого $\bar{d} \leqslant d^*$ и $I(U; V) \leqslant R(d^*) + \delta/4$. Применим лемму 9.3.1 к этому тест-каналу, выбирая $\hat{d} = d^* + \delta/2$, $\hat{R} = R(d^*) + \delta/2$ и

$$M = \left\lfloor \exp\left(LR(d^*) + \frac{3\delta}{4}L\right)\right\rfloor.$$

Тогда, как и в (9.3.11), имеем

$$P_c[D > L(d^* + \delta/2)] \leqslant P_t(A) + \exp[-e^{L\delta/4} + 1], \tag{9.6.3}$$

где

$$P_t(A) \leqslant \Pr\{I(\mathbf{u}; \mathbf{v}) > L[R(d^*) + \delta/2]\} + \Pr\{D(\mathbf{u}; \mathbf{v}) > L(d^* + \delta/2)\} \leqslant \tag{9.6.4}$$

$$\leqslant \Pr\{I(\mathbf{u}; \mathbf{v}) > L(I(U; V) + \delta/4)\} + \Pr\{D(\mathbf{u}; \mathbf{v}) > L(\bar{d} + \delta/2)\}. \tag{9.6.5}$$

Так как $I(U, V)$ и $\bar{d}$ конечны, то закон больших чисел утверждает, что для фиксированного δ вероятность $P_t(A)$ стремится к 0 при возрастании L и, следовательно,

$$\lim_{L\to\infty} P_c[D > L(d^* + \delta/2)] = 0. \tag{9.6.6}$$

Пусть для любого заданного кода B — множество последовательностей источника, для которых $D\,[\mathbf{u};\ \mathbf{v}\,(\mathbf{u})] > L\,(d^* + \delta/2)$. Для любого L некоторый код в ансамбле с заданным M удовлетворяет неравенству $P\,(B) \leqslant P_c\,[D > L\,(d^* + \delta/2)]$. Добавим к этому коду множество J^L кодовых слов, одно кодовое слово для каждой последовательности из L букв алфавита $a_1, \ldots, a_J$. Отобразим последовательности источника, не содержащиеся в B, в первоначальные M кодовых слов с искажением на букву, не большим $d^* + \delta/2$. Отобразим последовательности источника из B в ближайшее из добавочных J^L кодовых слов. Энтропия всего множества кодовых слов ограничена так же, как в (9.3.28),

$$H(\mathbf{V}^L) \leqslant L\left\{R\,(d^*) + \delta/2 + P\,(B)\ln J + \frac{\mathcal{H}\,[P\,(B)]}{L}\right\}. \tag{9.6.7}$$

Если $\ln J \leqslant R\ (d^*)$, то общее число кодовых слов равно

$$M' = M + J^L \leqslant 2\exp L\,[R\,(d^*) + \delta/2]. \tag{9.6.8}$$

Следовательно, для достаточно больших L удовлетворяются ограничения на $H\,(\mathbf{V}^L)$ и M'.

Пусть x_B — случайная величина,

$$x_B = \begin{cases} 1, & \mathbf{u} \in B, \\ 0, & \mathbf{u} \notin B, \end{cases} \tag{9.6.9}$$

и пусть z_l, $1 \leqslant l \leqslant L$, являются одинаково распределенными случайными величинами

$$z_l = \min_j d\,(u_l;\ a_j).$$

Искажение на букву для последовательности $\mathbf{u} \in B$ имеет вид $\frac{1}{L}\sum_{l=1}^{L} z_l$.

Среднее искажение на букву для заданного кода, таким образом, ограничено выражением

$$\bar{d}_L \leqslant d^* + \delta/2 + \overline{\frac{x_B}{L}\sum_{l=1}^{L} z_l}, \tag{9.6.10}$$

где $d^* + \delta/2$ — верхняя граница искажения, связанная с $\mathbf{u} \notin B$, а последнее выражение является искажением, связаным с $\mathbf{u} \in B$. Пусть z' — некоторое число, которое будет выбрано позднее, и пусть для каждого l, $1 \leqslant l \leqslant L$,

$$x_l = \begin{cases} 1, & z_l \leqslant z', \\ 0, & z_l > z'. \end{cases} \tag{9.6.11}$$

Тогда

$$\overline{x_B z_l} = \overline{x_B z_l x_l} + \overline{x_B z_l (1 - x_l)} \leqslant \bar{x}_B z' + \overline{z_l (1 - x_l)}. \tag{9.6.12}$$

Первое из указанных выше слагаемых было ограничено сверху на основе того, что так как $x_l = 0$ для $z_l > z'$, то $x_l z_l \leqslant z'$. Второе слагае-

мое было ограничено сверху с помощью неравенства $x_B \leqslant 1$. Используя определения x_B и x_l, имеем

$$P(B) = \bar{x}_B; \quad \overline{z_l(1-x_l)} = \int_{z'}^{\infty} z_l \, dF(z_l), \tag{9.6.13}$$

где $F(z_l)$ — функция распределения z_l. Подставляя (9.6.12) и (9.6.13) в (9.6.10), получаем

$$\bar{d}_L \leqslant d^* + \delta/2 + z' P(B) + \int_{z'}^{\infty} z dF(z). \tag{9.6.14}$$

Так как по предположению $\bar{z}_l < \infty$, то последний интеграл в (9.6.14) стремится к 0, когда z' стремится к ∞; следовательно, последний интеграл меньше чем $\delta/4$ при достаточно больших z'. Для такого z' имеем $z' P(B) \leqslant \delta/4$ при достаточно больших L. Следовательно, при достаточно больших L имеем $d_L \leqslant d^* + \delta$. |

Хотя эта теорема показывает, что коды источника могут иметь скорость, сколь угодно близкую к $R(d^*)$ при средних искажениях, сколь угодно близких к d^*, однако условия теоремы не достаточно сильны, чтобы можно было утверждать, что искажения, близкие к d^*, могут быть достигнуты после передачи по каналу с шумами. Требуемые дополнительные условия приведены в следующей теореме.

Теорема 9.6.3. Если дискретный по времени источник без памяти с мерой искажения $d(u; v)$ удовлетворяет условию $\overline{d(u; v)} < \infty$ для каждого v, где математическое ожидание взято по ансамблю источника, и если этот источник связан с адресатом с помощью канала, имеющего пропускную способность C нат на букву источника, и для которого произвольно малая вероятность ошибки может быть достигнута при любой скорости передачи данных, меньшей C, то может быть достигнуто среднее искажение на букву, произвольно близкое к d^*, где $C = R(d^*)$. В более общем случае условие $\overline{d(u; v)} < \infty$ заменяется на условие, что $R(d^*)$ может быть аппроксимирована сколь угодно точно на совместных мерах вероятностей, для которых равна нулю вероятность множества таких v, что $\overline{d(u; v)} = \infty$.

Доказательство. Покажем, что для любого $\delta > 0$ может быть достигнуто среднее искажение на букву $\bar{d} < d^* + \delta$. Если $C = 0$, то теорема тривиальна, так что будем предполагать, что $C = R(d^*) > 0$. Из условий теоремы следует, что d_{max} конечно, где d_{max} — наименьшее значение, для которого $R(d_{max}) = 0$. Следовательно, $d^* < d_{max}$ и $R(d^*)$ строго убывает по d^*, где $C = R(d^*)$. Пусть $d_1 = d^* + \delta/2$ и пусть δ_1 равно наименьшему из чисел $1/2\,[C - R(d_1)]$ и $\delta/4$. Из теоремы 9.6.2 следует, что можно выбрать код столь большой блоковой длины L, что число кодовых слов удовлетворяет неравенству

$$M \leqslant \exp[LR(d_1) + L\delta_1] \leqslant \exp[L(C - \delta_1)] \tag{9.6.15}$$

и искажение на букву удовлетворяет неравенству

$$d_L \leqslant d_1 + \delta_1 \leqslant d^* + 3/4\delta, \qquad (9.6.16)$$

и для всех v из кода $\overline{d(u; v)} < \infty$.

Так как имеется конечное число букв в этом заданном коде, то $\hat{d} < \infty$ можно выбрать равным максимуму $\overline{d(u; v)}$ по буквам v из кода. Теперь среднее искажение на последовательность при условии, что кодовое слово источника $\mathbf{v}_m$ ошибочно воспроизводится каналом как $\mathbf{v}_{m'}$, равно среднему искажению между $\mathbf{v}_{m'}$ и теми $\mathbf{u}$, которые отображаются в $\mathbf{v}_m$. Оно ограничено сверху $L\hat{d}/\Pr(\mathbf{v}_m)$. Так как вероятность ошибочного декодирования в канале может быть сделана произвольно малой (быть может, одновременным кодированием многих блоков источника), то вклад в среднее искажение из-за ошибок в канале может быть сделан сколь угодно малым для заданного кода источника, и среднее искажение на букву, включающее ошибки в канале, удовлетворяет неравенству $\overline{d}_L \leqslant d^* + \delta$. |

Для того чтобы глубже понять, почему в теореме требуется, чтобы $\overline{d(u; v)} < \infty$ для всех v, полезно рассмотреть код, для которого $\overline{d(u; v_{l,m})} = \infty$, скажем для l-й буквы m-го кодового слова кода. Предположим, что этот код источника используется для передачи по дискретному по времени каналу без памяти с минимальной переходной вероятностью $P_{min} > 0$. Если для кода канала с длиной блока N имеется какой-либо выход канала, который декодируется в $\mathbf{v}_m$, то при любом выборе u_l последовательность $\mathbf{v}_m$ (и, следовательно, $v_{l,m}$) будет возникать с вероятностью, не меньшей P_{min}^N. Тогда среднее искажение для l-й буквы кода удовлетворяет соотношению

$$\overline{d(l)} \geqslant \int q(u_l) P_l(v_{l,m} \mid u_l)\, d(u_l; v_{l,m})\, du_l \geqslant$$
$$\geqslant P_{min}^N \int q(u_l)\, d(u_l; v_{l,m})\, du_l = \infty.$$

При попытке вычислить функцию $R(d^*)$ для любых заданных непрерывного по амплитуде источника и меры искажения удобно (как и в § 9.4) иметь дело с функцией

$$R_0(\rho, \mathbf{P}) = I(U; V) + \rho d, \qquad (9.6.17)$$

где через $\mathbf{P}$ обозначен заданный тест-канал, а через $I(U; V)$ и $\overline{d}$ обозначены средняя взаимная информация и среднее искажение для совокупности источника и тест-канала. Как и ранее, нижняя грань $R_0(\rho, \mathbf{P})$ по всем тест-каналам (т. е. по всем совместным вероятностным мерам с соответствующей входной вероятностной мерой) равна координате точки пересечения оси R с касательной наклона $-\rho$ к кривой $R(d^*)$. Если $q(u)$ — плотность вероятности букв источника, то нижняя граница $R_0(\rho, \mathbf{P})$ теоремы 9.4.1 принимает вид

$$R_0(\rho, \mathbf{P}) \geqslant \int q(u) \ln \frac{f(u)}{q(u)}\, du, \qquad (9.6.18)$$

где $f(u)$ удовлетворяет ограничению

$$\int f(u)\, e^{-\rho d(u;} \quad du \leqslant 1 \text{ для всех } v. \qquad (9.6.19)$$

Необходимые и достаточные условия, накладываемые на $f(u)$ и на плотность переходной вероятности $p(v \mid u)$ для того, чтобы в (9.6.18) имело место равенство, заключаются в том, что функция $\omega(v) \geqslant 0$ удовлетворяет соотношению

$$\int \omega(v)\, e^{-\rho d(u;\, v)}\, dv = \frac{q(u)}{f(u)} \text{ для всех } u \tag{9.6.20}$$

и что (9.6.19) удовлетворяется с равенством для всех v, для которых $\omega(v) > 0$. Тогда

$$p(v \mid u) = \frac{\omega(v) f(u)}{q(u)}\, e^{-\rho d(u;\, v)}. \tag{9.6.21}$$

Доказательство этих утверждений совпадает с доказательством первой половины теоремы 9.4.1 при замене сумм на интегралы. Однако мы не можем показать, что всегда существуют функции $p(v \mid u)$ и $f(u)$, для которых (9.6.18) удовлетворяется с равенством. Мы не можем даже доказать, что к равенству можно приблизиться путем все более и более тщательного выбора $p(v \mid u)$ и $f(u)$, хотя это последнее утверждение кажется верным. К счастью, для важного примера, изложенного в следующем параграфе, равенство может быть достигнуто и нижняя граница дает возможность легко вычислить $R(d^*)$.

9.7. ГАУССОВСКИЕ ИСТОЧНИКИ С КВАДРАТИЧНО-РАЗНОСТНЫМ ИСКАЖЕНИЕМ

Рассмотрим источник, выход которого представляет собой последовательность статистически независимых одинаково распределенных гауссовских случайных величин $\ldots, u_{-1}, u_0, u_1, \ldots$, каждая из которых имеет плотность вероятности

$$q(u) = \frac{1}{\sqrt{2\pi A}} \exp\left(-\frac{u^2}{2A}\right). \tag{9.7.1}$$

Найдем для этого источника скорость как функцию искажения $R(d^*)$ с мерой искажения $d(u; v) = (u - v)^2$. Найдем сначала нижнюю границу для $\min R_0(\rho, \mathbf{P})$ [точки, в которых ось R пересекается с касательной наклона $-\rho$ к кривой $R(d^*)$] и используем эту границу для построения нижней границы для $R(d^*)$. Затем покажем, что нижняя граница равна $R(d^*)$ и рассмотрим получающийся тест-канал. Из (9.6.18) имеем

$$\min R_0(\rho, \mathbf{P}) \geqslant \int_{-\infty}^{\infty} q(u) \ln \frac{f(u)}{q(u)}\, du, \tag{9.7.2}$$

где $f(u)$ — любая функция, удовлетворяющая ограничению

$$\int_{-\infty}^{\infty} f(u) \exp\left[-\rho (u - v)^2\right] du \leqslant 1 \text{ для всех } v. \tag{9.7.3}$$

Заменой переменных $y = u - v$ приводим этот интеграл к виду

$$\int f(y+v)\exp[-\rho y^2]dy.$$

Поэтому (9.7.3) может быть удовлетворено с равенством для всех v, если $f(u)$ постоянная*)

$$f(u) = \sqrt{\rho/\pi}. \tag{9.7.4}$$

Подставляя эту формулу в (9.7.2) и интегрируя, получаем

$$\min R_0(\rho, \mathbf{P}) \geqslant {}^1/_2 \ln(2\rho e A). \tag{9.7.5}$$

Для любого ρ имеем $R(d^*) \geqslant \min R_0(\rho, \mathbf{P}) - \rho d^*$ и, следовательно

$$R(d^*) \geqslant {}^1/_2 \ln(2\rho e A) - \rho d^*. \tag{9.7.6}$$

Максимизация правой части (9.7.6) по ρ дает

$$\rho = \frac{1}{2d^*}, \tag{9.7.7}$$

$$R(d^*) \geqslant \frac{1}{2}\ln\frac{eA}{d^*} - \frac{1}{2} = \frac{1}{2}\ln\frac{A}{d^*}. \tag{9.7.8}$$

Для $d^* > A$ граница в (9.7.8) отрицательна и может быть заменена на $R(d^*) \geqslant 0$. Вместе с тем, отображением всех u в $v = 0$ достигается среднее искажение $\bar{d} = A$ с нулевой средней взаимной информацией, так что

$$R(d^*) = 0, \quad d^* \geqslant A. \tag{9.7.9}$$

Далее покажем, что (9.7.8) удовлетворяется с равенством при $d^* < A$. Для этого сначала покажем, что (9.7.5) удовлетворяется с равенством при $\rho \geqslant 1/(2A)$, и найдем соответствующий тест-канал. Необходимое условие для равенства в (9.7.5) состоит в том, что имеется решение с $\omega(v) \geqslant 0$ уравнения

$$\int_{-\infty}^{\infty} \omega(v)\exp[-\rho(u-v)^2]\,dv = \frac{q(u)}{f(u)} \text{ для всех } u, \tag{9.7.10}$$

$$\int_{-\infty}^{\infty} \omega(v)\sqrt{\rho/\pi}\exp[-\rho(u-v)^2]\,dv = q(u). \tag{9.7.11}$$

Левая часть (9.7.11) является сверткой гауссовской плотности вероятности с дисперсией $1/(2\rho)$ с функцией $\omega(v)$, а правая часть является гауссовской плотностью вероятности с дисперсией A. Следовательно, (9.7.11) удовлетворяется для всех u при

$$\omega(v) = \frac{1}{\sqrt{2\pi A - \pi/\rho}}\exp\left(-\frac{v^2}{2A - 1/\rho}\right). \tag{9.7.12}$$

Функция $\omega(v)$ является плотностью вероятности при $\rho > 1/(2A)$, сходится к δ-функции при $\rho \to 1/(2A)$ и не существует как действительное

*) Заметим, что постоянное значение для $f(u)$ будет удовлетворять (9.7.3) с равенством для любой разностной меры искажения и что эта постоянная не зависит от $q(u)$.

решение (9.7.11) при $\rho < 1/(2A)$. Отсюда следует, что (9.7.5) удовлетворяется с равенством при $\rho > 1/(2A)$ и, следовательно, что

$$R(d^*) = \begin{cases} \frac{1}{2} \ln \frac{A}{d^*}; & d^* < A, \\ 0; & d^* \geqslant A. \end{cases} \tag{9.7.13}$$

График этой функции изображен на рис. 9.7.1. Координата пересечения оси R и касательной наклона $-\rho$ к этой кривой задается для всех $\rho > 0$ равенством

$$\min_{\mathbf{P}} R_0(\rho, \mathbf{P}) = \begin{cases} {}^1/_2 \ln(2\rho e A); & \rho > 1/(2A), \\ \rho A; & \rho \leqslant 1/2A. \end{cases} \tag{9.7.14}$$

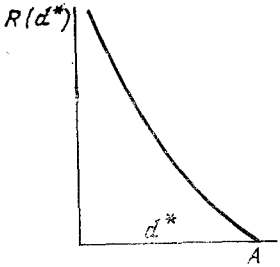


Рис. 9.7.1. Скорость как функция искажения для гауссовского дискретного по времени источника с дисперсией A и с квадратично-разностным искажением.

Так как эти рассуждения, доказывающие равенство, строились на довольно абстрактных идеях § 9.4 и 9.6, то найдем тест-канал, задаваемый (9.7.12), и покажем, что он действительно дает искажение $d^* = 1/(2\rho)$ и среднюю взаимную информацию $1/2 \ln (A/d^*)$. Из (9.6.21) следует, что переходная плотность вероятности тест-канала равна

$$p(v \mid u) = \frac{\omega(v)}{q(u)} p_b(u \mid v),$$

где обратная переходная вероятность $p_b(u \mid v)$ задается выражением

$$p_b(u \mid v) = f(u) \exp[-\rho(u-v)^2] = \tag{9.7.15}$$

$$= \sqrt{\rho/\pi} \exp[-\rho(u-v)^2]. \tag{9.7.16}$$

Если временно представить себе, что v является входом канала, то $p_b(u \mid v)$ можно считать распределением, порожденным аддитивной гауссовской случайной величиной z с нулевым средним значением и дисперсией $1/(2\rho)$. При таком истолковании $u = v + z$, где v и z — независимые гауссовские случайные величины (рис. 9.7.2) и тест-канал является «обращенным» каналом с аддитивным гауссовым шумом. Тогда среднее искажение равно как раз среднеквадратическому значению z

$$\overline{d} = \overline{(u-v)^2} = \overline{z^2} = 1/(2\rho). \tag{9.7.17}$$

Средняя взаимная информация равна

$$I(U; V) = \frac{1}{2} \ln \left[1 + \frac{\overline{v^2}}{\overline{z^2}}\right] = \frac{1}{2} \ln \frac{\overline{u^2}}{\overline{d}} = \frac{1}{2} \ln \frac{A}{\overline{d}}. \tag{9.7.18}$$

В каком-то смысле интуитивно более удовлетворительны вид тест-канала изображен на рис. 9.7.3, где выход v получается сначала сложением u и независимой гауссовской случайной величины w с нулевым средним и дисперсией $\bar{d}A/(A-\bar{d})$ и затем умножением результата на $(A-\bar{d})/A$. Так как u и v совместно гауссовские, эквивалентность этих рисунков показывает, что $\overline{v^2}$ и $\overline{uv}$ одни и те же для каждого рисунка. Можно также показать, что умножение является в точности той операцией, которая требуется для образования оценки u с минимальной дисперсией ошибки по сумме $u+w$. Теоремы кодирования 9.6.2 и 9.6.3 для источника применимы к этому источнику, так как для любого заданного v среднее относительно $q(u)$ значение $d(u; v)$ равно $A+v^2$.

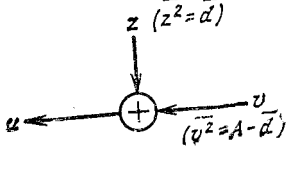


Рис. 9.7.2. Тест-канал (обращенный вид).

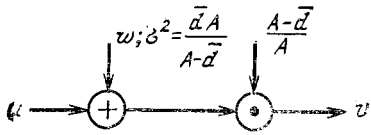


Рис. 9.7.3. Тест-канал (прямой вид).

Имеется один особенно важный канал, а именно дискретный по времени канал с аддитивным гауссовым шумом с ограничением на энергию, при передаче по которому можно достигнуть без кодирования минимума среднеквадратической ошибки d^*, задаваемой $C=R(d^*)$. Нужно просто усилить выход источника до энергии, соответствующей ограничению на входе канала, и затем соответственно ослабить выход канала. Более интересный вариант этой задачи возникает, когда используется канал, скажем, N раз для каждого выхода источника. Это, конечно, математическое моделирование ситуации, когда гауссовский случайный процесс с некоторой полосой частот должен быть передан с помощью непрерывного по времени канала с аддитивным шумом и полосой частот в N раз большей полосы частот источника. В этой ситуации для достижения предельного искажения, определяемого теоремой кодирования для источника, должны быть использованы кодирование источника и кодирование для канала. В таких случаях для достижения сравнительно малого искажения может быть использована хорошо знакомая и просто реализуемая техника частотной, фазовой и импульсно-кодовой модуляций. Однако, как мы увидим вскоре, если в нашем распоряжении имеется бесшумная обратная связь от выхода канала к выходу источника, то возможно достигнуть минимально возможного искажения, по существу, без кодирования.

Пусть σ^2 — дисперсия шума w в канале и пусть A — значение энергии, задающее ограничение на входе канала. Определим величину $\bar{d}$ равенством $\bar{d}=\sigma^2A/(A+\sigma^2)$, и заметим, что $\sigma^2=\bar{d}A/(A-\bar{d})$. Удобно умножить выход канала на $(A-\bar{d})/A$ так, что канал принимает вид тест-канала рис. 9.7.3. Пусть $x_1, \ldots, x_N$ — N входов канала, соответст-

вующих одной букве источника, и $y_1, \ldots, y_N$ — соответствующие выходы (справа от умножителя). Пусть $z_n = x_n - y_n$, $1 \leqslant n \leqslant N$ заметим, что если каждый вход представляет собой гауссовскую случайную величину с нулевым средним и дисперсией A, то обращенный канал на рис. 9.7.2 эквивалентен каналу на рис. 9.7.3 и каждая z_n является гауссовской случайной величиной с нулевым средним, не зависит от y_n и имеет дисперсию $\bar{d}$.

Примем вначале, что выход источника u имеет дисперсию A, равную ограничению на энергию в канале. Рассмотрим следующий выбор входов канала:

$$x_1 = u,$$
$$x_n = z_{n-1}\sqrt{A/\bar{d}}, \quad 2 \leqslant n \leqslant N. \tag{9.7.19}$$

Так как x_1 — гауссовская случайная величина с нулевым средним и дисперсией A, то отсюда следует, что z_1 — независимая от y_1 гауссовская случайная величина с нулевым средним и с дисперсией $\bar{d}$. Так как $x_2 = z_1\sqrt{A/\bar{d}}$, то отсюда следует, что x_2 — гауссовская случайная величина с нулевым средним и дисперсией A. Продолжая эти рассуждения, видим, что все x_n являются гауссовскими случайными величинами с нулевыми средними и дисперсиями A, и каждая z_n — независимая от y_n гауссовская случайная величина с нулевым средним с дисперсией $\bar{d}$. Заметим, что эта схема передачи требует, чтобы передатчик знал $(n-1)$-й принятый символ перед передачей n-го символа.

Предположим, что приемник для каждого n находит оценку v_n символа источника u с помощью соотношений

$$v_1 = y_1,$$
$$v_n = v_{n-1} + y_n\left(\frac{\bar{d}}{A}\right)^{(n-1)/2}, \quad 2 \leqslant n \leqslant N. \tag{9.7.20}$$

Покажем теперь, что v_n можно представить в виде

$$v_n = u - z_n\left(\frac{\bar{d}}{A}\right)^{(n-1)/2}. \tag{9.7.21}$$

Это устанавливается по индукции. Для $n = 1$ имеем $v_1 = y_1 = x_1 - z_1 = u - z_1$, что согласуется с (9.7.21). Теперь примем, что (9.7.21) справедливо для $n-1$ и подставим это выражение вместо v_{n-1} в (9.7.20). Используя соотношения $y_n = x_n - z_n$ и (9.7.19) для x_n, найдем, что (9.7.21) справедливо для n.

Из (9.7.21) видно, что z_n пропорционально ошибке в оценке при n-й передаче. Следовательно, при каждой последующей передаче передается нормированное по амплитуде значение предыдущей ошибки [см. (9.7.19)]. Из (9.7.20) видно, что приемник использует каждый принятый сигнал для исправления ошибки предыдущей передачи.

Как следует из (9.7.21), конечное среднее искажение после N передач равно

$$\bar{d}_N = \overline{(v_N - u)^2} = \bar{d}\,(\bar{d}/A)^{N-1}. \tag{9.7.22}$$

Отсюда следует, что $1/2 \ln (A/d_N) = N/2 \ln (A/\bar{d})$. Это равно пропускной способности канала, умноженной на N, и, таким образом, $\bar{d}_N$ — минимально возможное искажение для этих источника и канала, как следует из теоремы кодирования для источника. Если дисперсия источника отлична от величины, ограничивающей энергию на входе канала, то можно просто изменить масштаб перед первой передачей и восстановить масштаб на последней оценке, так что отношение дисперсии источника к конечному искажению останется равным тому же самому оптимальному значению. Приведенный выше результат принадлежит Элайсу (1961).

Указанный выше метод использования обратной связи для передачи гауссовской случайной величины с минимальным среднеквадра-

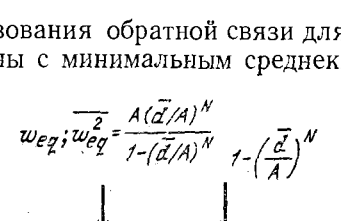


Рис. 9.7.4. Модель для передачи с обратной связью гауссовской случайной величины.

тичным искажением тесно связан с методом передачи цифровых данных по гауссовскому каналу с обратной связью, открытым Шелквийком и Кайлатом (1966). Для того чтобы вывести этот результат, установим сначала, что для каждого n, $1 \leqslant n \leqslant N$, величины v_n и z_n в (9.7.21) статистически независимы. Чтобы увидеть это, напомним, что z_n и y_n независимы. Теперь, если применить индукцию и принять, что z_{n-1} и v_{n-1} независимы, то из (9.7.19) следует, что x_n и v_{n-1} независимы. Так как z_n является линейной комбинацией x_n и w_n, то z_n и v_{n-1} независимы. Следовательно, из (9.7.20) находим, что z_n и v_n независимы, если z_{n-1} и v_{n-1} независимы. Для завершения доказательства остается заметить, что, очевидно, z_1 и v_1 независимы. Исходя из этого результата, всю передачу от u до v_N можно представить с помощью рис. 9.7.2, рассматривая z_N $(\bar{d}/A)^{(N-1)/2}$ как шум. Эквивалентно этому, передачу можно представить, как изображено на рис. 9.7.4.

Так как эквивалентный шум w_{eq} на рис. 9.7.4 является гауссовым и он не зависит от u, то рис. 9.7.4 можно использовать для изображения схемы передачи при использовании (9.7.19) независимо от того, является u гауссовской случайной величиной или нет*). Теперь предположим, что требуется передать одно сообщение из множества M сообщений. Эти сообщения можно закодировать в множество чисел от $-\sqrt{A}$ к $\sqrt{A}$ следующим образом:

*) Студент, которого беспокоит абстрактный характер этого рассуждения, может выразить z_n через случайные величины w_n — действительные шумы в канале, $z_n = x_n \bar{d}/A - w_n(A - \bar{d})/A$. С использованием этого выражения и (9.7.19) для каждого значения n величина z_n может быть выражена как линейная комбинация u и w_n, $1 \leqslant n \leqslant N$. Сочетая это с (9.7.21), можно непосредственно проверить модель, изображенную на рис. 9.7.4.

$$m \to u_m = -\sqrt{A} + \frac{2m\sqrt{A}}{M-1}, \quad 0 \leqslant m \leqslant M-1. \tag{9.7.23}$$

Тогда декодирование можно осуществить, отображая принятое число $r_N = v_N/[1 - (\bar{d}/A)^N]$ в ближайшую точку u_m сообщения. Если шум $|w_{eq}|$ меньше чем $\sqrt{A}/(M-1)$, то ошибки не произойдет. Следовательно, для каждого сообщения m

$$P_{e,m} \leqslant 2\Phi\left[-\frac{\sqrt{A}}{(M-1)\sqrt{\overline{w_{eq}^2}}}\right]; \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Взяв $M = e^{NR}$, где R — скорость в натах на символ канала, и вспоминая, что пропускная способность канала в натах равна $C = {}^1/_2 \ln(A/\bar{d})$, получаем отсюда

$$P_{e,m} \leqslant 2\Phi\left[-e^{NC-NR}\left(\frac{\sqrt{1-e^{-2NC}}}{1-e^{-NR}}\right)\right]. \tag{9.7.24}$$

При $R < C$ выражение, стоящее в круглых скобках, ограничено снизу 1, и, используя границу (8.2.38) для $\Phi(-x)$, находим

$$P_{e,m} \leqslant \sqrt{\frac{2}{\pi}}\, e^{-N(C-R)} \exp\left[-\frac{e^{2N(C-R)}}{2}\right]. \tag{9.7.25}$$

Если сообщения равновероятны, то расстояние между сообщениями может быть увеличено до $2\sqrt{3A}/\sqrt{M^2-1}$ без нарушения ограничения на мощность при первой передаче. При этом получающаяся вероятность ошибки ограничена выражением

$$P_{e,m} \leqslant \sqrt{\frac{2}{3\pi}}\, e^{-N(C-R)} \exp\left[-\frac{3e^{2N(C-R)}}{2}\right].$$

Значение и необычность этого результата заключаются в том, что $P_{e,m}$ убывает по двойной экспоненте с ростом N, а не по обычной экспоненте, подобно всем остальным нашим результатам о вероятности ошибки.

Источники, порождающие гауссовские случайные процессы

Распространим теперь результаты для гауссовских дискретных по времени источников на источник, выход которого является стационарным гауссовским случайным процессом с нулевым средним и корреляционной функцией $\mathcal{R}_1(\tau)$. Рассмотрим выход источника на интервале $(-T/2, T/2)$ и представим его с помощью разложения Карунена—Лоева

$$u(t) = \sum_i u_i \theta_i(t), \quad -T/2 \leqslant t \leqslant T/2, \tag{9.7.26}$$

где ортонормальные на интервале $-T/2 \leqslant t \leqslant T/2$ функции $\theta_i(t)$ являются решениями уравнения

$$\int_{-T/2}^{T/2} \mathcal{R}_1(t_1 - t_2)\,\theta_i(t_2)\,dt_2 = \lambda_i\,\theta_i(t_1). \tag{9.7.27}$$

Как было показано в гл. 8, u_i — статистически независимые гауссовские случайные величины с нулевыми средними и дисперсиями

$$\overline{u_i^2} = \lambda_i. \tag{9.7.28}$$

Предположим, что нужно представить $u(t)$ у адресата с помощью функции $v(t)$, и определим искажение на единицу времени между $u(t)$ и $v(t)$ на интервале $(-T/2,\ T/2)$ равенством

$$d_T[u(t);\ v(t)] = \frac{1}{T}\int_{-T/2}^{T/2} [u(t) - v(t)]^2\,dt. \tag{9.7.29}$$

Если $v(t)$ разложено по тем же ортонормальным функциям, как $u(t)$, т. е. $v(t) = \Sigma\ v_i\theta_i(t)$, то

$$d_T[u(t);\ v(t)] = \frac{1}{T}\sum_i (u_i - v_i)^2. \tag{9.7.30}$$

Для любой заданной вероятностной меры на последовательности u_i и последовательности v_i, для которой u_i — независимые гауссовские случайные величины с нулевыми средними и с $\overline{u_i^2} = \lambda_i$ для каждого i, можно рассмотреть взаимную информацию в натах на единицу времени $(1/T)\ I(\mathbf{U};\ \mathbf{V})$ между $\mathbf{u}$ и $\mathbf{v}$ последовательностями и среднее искажение $\overline{d_T}$ на единицу времени, задаваемое как среднее значение выражения (9.7.30) по $\mathbf{u}$ и $\mathbf{v}$. Скорость как функция искажения для заданного источника и заданного интервала T тогда определяется как

$$R_T(d^*) = \inf \frac{1}{T}\,I(\mathbf{U};\ \mathbf{V}), \tag{9.7.31}$$

где нижняя грань берется по вероятностным мерам, для которых u_i — независимые гауссовские случайные величины с нулевыми средними и с $\overline{u_i^2} = \lambda_i$ и для которых $\overline{d_T} \leqslant d^*$.

Найдем сначала $R_T(d^*)$, затем возьмем предел

$$R(d^*) = \lim_{T\to\infty} R_T(d^*) \tag{9.7.32}$$

и, наконец, покажем, что $R(d^*)$ допускает то же самое истолкование, как $R(d^*)$ для дискретных по времени источников без памяти.

Для того чтобы вычислить $R_T(d^*)$, определим сначала

$$R_{0,T}(\rho,\ \mathbf{P}) = \frac{1}{T}\,I(\mathbf{U};\ \mathbf{V}) + \rho\overline{d_T}, \tag{9.7.33}$$

где $\mathbf{P}$ в (9.7.33) обозначает совместную вероятностную меру с данной статистикой источника, а $I(\mathbf{U};\ \mathbf{V})$ и $\overline{d_T}$ вычисляются по этой мере. Так

как u_i статистически независимы, то можно, используя те же рассуждения, как и в (9.6.2), получить

$$\frac{1}{T} I(\mathbf{U}; \mathbf{V}) \geqslant \frac{1}{T} \sum_{i=1}^{\infty} I(U_i; V_i) \tag{9.7.34}$$

с равенством, если каждая пара $U_i V_i$ статистически независима от других пар. Аналогично из (9.7.30)

$$\bar{d}_T = \frac{1}{T} \sum_{i=1}^{\infty} \overline{(u_i - v_i)^2}. \tag{9.7.35}$$

Следовательно,

$$R_{0,\,T}(\rho, \mathbf{P}) \geqslant \frac{1}{T} \sum_{i=1}^{\infty} [I(U_i; V_i) + \rho \overline{(u_i - v_i)^2}] \tag{9.7.36}$$

с равенством, если каждая пара статистически независима от других пар. Как было показано в (9.7.14), минимум каждого отдельного слагаемого этого типа имеет вид

$$\min [I(U_i; V_i) + \rho \overline{(u_i - v_i)^2}] =$$

$$= g(\lambda_i; \rho) \overset{\Delta}{=} \begin{cases} {}^1/_2 \ln (2\rho e \lambda_i); & \rho > 1/(2\lambda_i). \\ \rho\lambda_i, \quad ; & \rho \leqslant 1/(2\lambda_i). \end{cases} \tag{9.7.37}$$

Левая часть (9.7.36) минимизируется, когда каждый совместный ансамбль $U_i V_i$ выбирается удовлетворяющим (9.7.37) и когда каждая пара не зависит от всех других пар. Имеем

$$\inf_{\mathbf{P}} R_{0,\,T}(\rho, \mathbf{P}) = \frac{1}{T} \sum_i g(\lambda_i, \rho), \tag{9.7.38}$$

$$R_T(d_T^*) = \min_{\rho \geqslant 0} \left[\frac{1}{T} \sum_i g(\lambda_i, \rho) - \rho d_T^* \right]. \tag{9.7.39}$$

Замечая, что $g(\lambda_i, \rho)$ дифференцируема по ρ (даже на границе $\rho = 1/(2\lambda_i)$), можно положить производную по ρ правой части (9.7.39) равной нулю, получая, что для минимизирующего ρ,

$$d_T^* = \frac{1}{T} \left[\sum_{i\,:\,\lambda_i > 1/(2\rho)} \frac{1}{2\rho} + \sum_{i\,:\,\lambda_i \leqslant 1/(2\rho)} \lambda_i \right]. \tag{9.7.40}$$

Подставновка (9.7.40) в (9.7.39) дает

$$R_T(d_T^*) = \frac{1}{T} \sum_{i\,:\,\lambda_i > 1/(2\rho)} \frac{1}{2} \ln (2\rho\lambda_i). \tag{9.7.41}$$

Можно увидеть, что (9.7.40) и (9.7.41) являются параметрическими уравнениями, определяющими d_T^* и $R_T(d_T^*)$ через ρ, где $-\rho$ — наклон $R_T(d_T^*)$. Для достаточно малых ρ, чтобы удовлетворить условию $\lambda_i \leqslant 1/(2\rho)$ при всех i, имеем $R_T(d_T^*) = 0$ и $d_T^* = (1/T)\Sigma\lambda_i$ является

средней мощностью выхода источника, которую в дальнейшем будем обозначать через d^*_{max}.

Заметим теперь, что (9.7.40) и (9.7.41) представляются каждое в виде $1/T$, умноженной на сумму по i функций собственных значений. В (9.7.40) функция равна $1/(2\rho)$ при $\lambda_i > 1/(2\rho)$ и равна λ_i при $\lambda_i \leqslant 1/(2\rho)$. В (9.7.41) функция равна $^1/_2 \ln(2\rho\lambda_i)$ при $\lambda_i > 1/(2\rho)$ и равна 0 при $\lambda_i \leqslant 1/(2\rho)$. Лемма 8.5.3 применима к обеим этим функциям и для любого фиксированного $\rho > 0$ можно перейти к пределу при $T \to \infty$, получая параметрические уравнения

$$d^* = \lim_{T\to\infty} d^*_T = \frac{1}{2\rho} \int_{f\,:\,F(f)>1/(2\rho)} df + \int_{f\,:\,F(f)\leqslant 1/(2\rho)} F(f)\,df, \tag{9.7.42}$$

$$R(d^*) = \lim_{T\to\infty} R_T(d^*) = \int_{f\,:\,F(f)>1/(2\rho)} \frac{1}{2} \ln[2\rho F(f)]\,df, \tag{9.7.43}$$

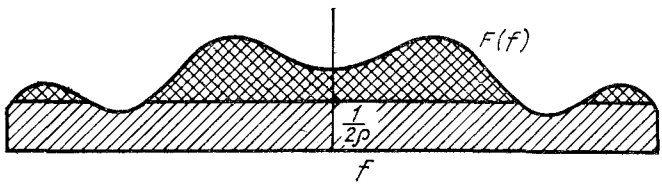


Рис. 9.7.5. Интерпретация интегралов для d^* и $R(d^*)$ для источника, порождающего гауссовский стационарный случайный процесс.

где $F(f) = \int \mathscr{R}_1(\tau) e^{j2\pi f\tau}\,d\tau$ — спектральная плотность источника. Рис. 9.7.5 иллюстрирует смысл этих интегралов. Заметим, что d^* — площадь заштрихованной области и что область интегрирования для $R(d^*)$ совпадает с областью частот под штриховкой накрест.

Тест-канал, для которого достигается эта функция $R(d^*)$, может быть представлен или с помощью рис. 9.7.6, который является непрерывным аналогом рис. 9.7.2, или с помощью рис. 9.7.7, который является непрерывным аналогом с небольшими изменениями рис. 9.7.3. Шумовой процесс $z(t)$ на каждом рисунке имеет спектральную плотность, задаваемую заштрихованной частью на рис. 9.7.5. Фильтры в прямом тест-канале на рис. 9.7.7 являются физически нереализуемыми, но, добавляя достаточную задержку в представлении $v(t)$, их

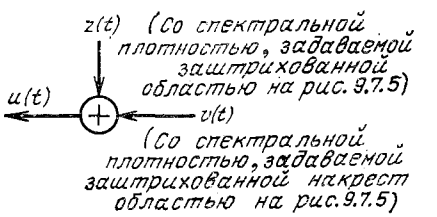


Рис. 9.7.6. Обращенный тест-канал для источника, порождающего стационарный гауссовский случайный процесс.

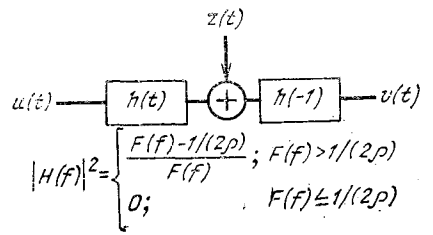


Рис. 9.7.7. Прямой тест-канал.

можно аппроксимировать сколь угодно близко физически реализуемыми фильтрами. Следует напомнить, что эти тест-каналы являются исскуственными ограничениями в теории передачи с заданным искажением. В отличие от дискретного по времени гауссовского источника тест-канал для источника, порождающего гауссовский случайный процесс, в общем случае не используется источником со скоростью, равной пропускной способности (при этом принимаются соответствующие мощностные ограничения на входе канала). Следовательно, даже если имеется соответствующий тест-канал, пригодный для достижения искажения d^* со скоростью $R(d^*)$, то можно достигнуть среднего искажения, меньшего чем d^*, если использовать дополнительную обработку данных и увеличить среднюю взаимную информацию в канале выше $R(d^*)$. Основная цель этих тест-каналов — дать легко запоминаемую картину того, что должно выполнять эффективное представление гауссовского случайного процесса. В частности, представление должно полностью игнорировать те частотные компоненты источника, на которых спектральная плотность мала.

Функция $R(d^*)$ определяется как предел по весьма большому интервалу времени минимума средней взаимной информации на единицу времени между $u(t)$ и $v(t)$ для заданного среднего искажения d^* на единицу времени. Из рассуждений, таких же как в теореме 9.2.2, ясно, что если источник связан с адресатом с помощью канала с пропускной способностью C нат в единицу времени, то среднее искажение, которое может быть достигнуто при любых преобразованиях источника и сигналов, ограничено снизу d^*, для которого $C = R(d^*)$. Теорема кодирования для источника также применима здесь, однако, к сожалению, ее доказательство отличается в ряде мест от данного в теореме 9.6.2 и потому оно будет приведено ниже.

Теорема 9.7.1. Пусть выходом источника является стационарный гауссовский случайный процесс $u(t)$ с корреляционной функцией $\mathcal{R}_1(\tau)$ и интегрируемой и ограниченной спектральной плотностью $F(f)$. Пусть мера искажения между $u(t)$ и $v(t)$ на интервале $(-T/2, T/2)$ задается (9.7.29). Тогда для любого $d^* > 0$, любого $\delta > 0$ и любого достаточно большого T существует код с $M \leqslant \exp[TR(d^*) + T\delta]$ кодовыми словами [где $R(d^*)$ задается (9.7.42) и (9.7.43)] такой, что среднее искажение на единицу времени между $u(t)$, $-T/2 \leqslant t \leqslant T/2$, и кодовым словом, в которое отображается $u(t)$, меньше или равно $d^* + \delta$.

Доказательство. Если $d^* \geqslant d^*_{max}$, то теорема тривиальна, так как код может состоять из одного-единственного кодового слова $v(t) = 0$. Если $d^* < d^*_{max}$, то выбираем ρ так, чтобы удовлетворить (9.7.42) для заданного d^*. Пусть для любого заданного $\delta > 0$ интервал времени T произволен, но достаточно велик, чтобы удовлетворить условию

$$R_T(d^*_T) \leqslant R(d^*) + \delta/4, \tag{9.7.44}$$

$$d^*_T \leqslant d^* + \delta/4, \tag{9.7.45}$$

где d_T^* и $R_T(d_T^*)$ задаются (9.7.40) и (9.7.41). Рассмотрим теперь ансамбль кодов с

$$M' = \lfloor \exp[TR(d^*) + T\delta] \rfloor - 1 \tag{9.7.46}$$

кодовыми словами. Кодовые слова выбираются независимо в соответствии с выходными вероятностями тест-канала, что приводит к $R_T(d_T^*)$, т. е. кодовые слова задаются соотношением

$$v_m(t) = \sum_l v_{m,l}\theta_l(t), \quad 1 \leqslant m \leqslant M',$$

где коэффициенты $v_{l,m}$ — независимые гауссовские случайные величины с нулевыми средними и с

$$\overline{v_{l,m}^2} = \begin{cases} \lambda_l - 1/(2\rho), & \lambda_l > 1/(2\rho), \\ 0, & \lambda_l \leqslant 1/(2\rho). \end{cases} \tag{9.7.47}$$

Пусть L — число собственных значений λ_l, которые больше чем $1/(2\rho)$. Тогда для каждого кодового слова $v_{l,m} = 0$ при всех $l > L$ и искажение между любым кодовым словом $v_m(t)$ и $u(t)$ задается выражением

$$D[u(t);\ v_m(t)] = \sum_{l=1}^{L}(u_l - v_{m,l})^2 + \sum_{l=L+1}^{\infty} u_l^2. \tag{9.7.48}$$

Отсюда видно, что для любого заданного множества кодовых слов и заданного $u(t)$ кодовое слово с минимальным искажением определяется только по первым L компонентам суммы $\Sigma u_l\theta_l(t)$. Заметим также, что вклад в среднее искажение на единицу времени от всех компонент, за исключением первых L, равен $\frac{1}{T}\sum_{l=L+1}^{\infty}\lambda_l$. Следовательно, если положить $\mathbf{u} = (u_1, \ldots, u_L)$, $\mathbf{v}_m = (v_{m,1}, \ldots, v_{m,L})$ и

$$D_1(\mathbf{u};\ \mathbf{v}_m) = \sum_{l=1}^{L}(u_l - v_{m,l})^2,$$

то среднее искажение в единицу времени для какого-либо кода из ансамбля можно представить следующим образом:

$$\bar{d}_T = \frac{1}{T}\left[\overline{D_1(\mathbf{u};\ \mathbf{v}(\mathbf{u}))} + \sum_{l=L+1}^{\infty}\lambda_l\right], \tag{9.7.49}$$

где $\mathbf{v}(\mathbf{u})$ — кодовое слово, в которое отображается $\mathbf{u}$. Далее определим $\hat{R}$ и $\hat{d}$ равенствами

$$\hat{R} = R_T(d_T^*) + \delta/2, \tag{9.7.50}$$

$$\hat{d} = \frac{1}{T}\sum_{l=1}^{L}\frac{1}{2\rho} + \frac{\delta}{2}. \tag{9.7.51}$$

Пусть $P_c(D_1 > T\hat{d})$ — вероятность в ансамбле кодов и функций, порождаемых источником того, что $D_1[\mathbf{u};\ \mathbf{v}(\mathbf{u})] > T\hat{d}$.

Как и в лемме 9.3.1,

$$P_c(D_1 > T\hat{d}) \geqslant P_t(A) + \exp(-M' e^{-T\hat{R}}), \tag{9.7.52}$$

где

$$A = \{\mathbf{u}, \mathbf{v} : I(\mathbf{u}; \mathbf{v}) > T\hat{R} \quad \text{или} \quad D_1(\mathbf{u}; \mathbf{v}) > T\hat{d}\} \tag{9.7.53}$$

и $P_t(A)$ — вероятность события A в ансамбле тест-канала. Далее оценим сначала сверху правую часть (9.7.52) и используем эту границу для получения верхней границы для $\overline{D_1[\mathbf{u}; \mathbf{v}(\mathbf{u})]}$.

Так как для тест-канала $I(\mathbf{U}; \mathbf{V}) = TR_T(d_T^*)$ и

$$\overline{D_1(\mathbf{u}; \mathbf{v})} = \sum_{l=1}^{L} \frac{1}{2\rho},$$

то $P_t(A)$ можно оценить сверху выражениями

$$P_t(A) \leqslant \Pr[I(\mathbf{u}; \mathbf{v}) > I(\mathbf{U}; \mathbf{V}) + (\delta/2)T] + \\ + \Pr[D_1(\mathbf{u}; \mathbf{v}) > \overline{D_1(\mathbf{u}; \mathbf{v})} + (\delta/2)T]. \tag{9.7.54}$$

Из неравенства Чебышева следует, что

$$P_t(A) \leqslant \frac{4\mathbf{D}[I(\mathbf{u}; \mathbf{v})]}{\delta^2 T^2} + \frac{4\mathbf{D}[D_1(\mathbf{u}; \mathbf{v})]}{\delta^2 T^2}. \tag{9.7.55}$$

Так как пары u_l, v_l независимы в ансамбле тест-канала, то

$$\mathbf{D}[I(\mathbf{u}; \mathbf{v})] = \sum_{j=1}^{L} \mathbf{D}[I(u_l; v_l)]. \tag{9.7.56}$$

Вспоминая, что v_l и $u_l - v_l$ — независимые гауссовские случайные величины с дисперсиями $\lambda_l - 1/(2\rho)$ и $1/(2\rho)$ соответственно, получаем

$$I(u_l; v_l) = \ln \frac{P_l(u_l \mid v_l)}{q_l(u_l)} = \frac{1}{2} \ln(2\rho\lambda_l) - \rho(u_l - v_l)^2 + \frac{u_l^2}{2\lambda_l}, \tag{9.7.57}$$

$$\mathbf{D}[I(u_l; v_l)] = 1 - \frac{1}{2\lambda_l \rho} \leqslant 1, \tag{9.7.58}$$

$$\mathbf{D}[I(\mathbf{u}; \mathbf{v})] \leqslant L. \tag{9.7.59}$$

Аналогично

$$\mathbf{D}[D_1(\mathbf{u}; \mathbf{v})] = \sum_{l=1}^{L} \mathbf{D}[(u_l - v_l)^2] = \\ = \sum_{l=1}^{L} \frac{2}{(2\rho)^2} = \frac{L}{2\rho^2}, \tag{9.7.60}$$

где использовано то обстоятельство, что $u_l - v_l$ — гауссовская случайная величина с дисперсией $1/(2\rho)$. Подставляя эти выражения в (9.7.55), находим

$$P_t(A) \leqslant \frac{4L[1 + 1/(2\rho^2)]}{\delta^2 T^2}. \tag{9.7.61}$$

На основании (9.7.46), (9.7.44) и (9.7.50) получаем

$$M' \geqslant \exp T\,[\hat{R} + \delta/4] - 2. \tag{9.7.62}$$

Подставляя (9.7.61) и (9.7.62) в (9.7.52), выводим

$$P_c(D_1 > T\hat{d}) \leqslant \frac{4L\,[1+1/(2\rho^2)]}{\delta^2 T^2} + \exp\left(-e^{T\delta/4} + 2\right). \tag{9.7.63}$$

Теперь рассмотрим некоторый код из ансамбля, для которого (9.7.63) удовлетворяется, и положим

$$B = \{\mathbf{u}\colon D_1\,[\mathbf{u};\ \mathbf{v}\,(\mathbf{u})] > T\hat{d}\}. \tag{9.7.64}$$

Добавим теперь к уже выбранным M' кодовым словам добавочное кодовое слово $v_0\,(t) = 0$ и заметим, что $M = M' + 1$ кодовых слов удовлетворяют условиям теоремы. Если $\mathbf{u} \in B$, то отобразим $\mathbf{u}$ в $v_0\,(t)$, в противном случае отобразим $\mathbf{u}$ в ближайшее кодовое слово. Для этого нового кода имеем

$$\overline{D_1\,[\mathbf{u};\ \mathbf{v}\,(\mathbf{u})]} \leqslant T\hat{d} + P(B) \sum_{l=1}^{L} \overline{d}_{l,\,B}, \tag{9.7.65}$$

где $\overline{d}_{l,\,B}$ — среднее искажение, или среднее значение u_l^2, при условии, что $\mathbf{u} \in B$. Как и в теореме 9.6.2, $\overline{d}_{l,\,B}$ можно оценить сверху, предполагая, что все большие значения $u^2{}_l$ относятся к последовательностям $\mathbf{u}$ из B. Так как $q_l\,(u_l) = q_l\,(-u_l)$, то это приводит к

$$P(B)\,\overline{d}_{l,\,B} \leqslant 2 \int_{u_l'}^{\infty} u_l^2\, q_l\,(u_l)\, du_l, \tag{9.7.66}$$

где u'_l определяется равенством

$$P(B) = 2 \int_{u_l'}^{\infty} q_l\,(u_l)\, du_l. \tag{9.7.67}$$

Используя неравенство Шварца для (9.7.66), получаем

$$P(B)\,\overline{d}_{l,\,B} \leqslant \sqrt{\left[2\int_{u_l'}^{\infty} u_l^4\, q_l\,(u_l)\, du_l\right]\left[2\int_{u_l'}^{\infty} q_l\,(u_l)\, du_l\right]} \leqslant$$

$$\leqslant \sqrt{\left[\int_{-\infty}^{\infty} u_l^4\, q_l\,(u_l)\, du_l\right] P(B)} = \lambda_l \sqrt{3P(B)}. \tag{9.7.68}$$

Подставляя (9.7.68) в (9.7.65) и строя границу сверху с помощью суммирования по всем l, находим

$$\overline{D_1\,[\mathbf{u};\ \mathbf{v}\,(\mathbf{u})]} \leqslant T\,[\hat{d} + \sqrt{3P(B)}\, d^*_{max}], \tag{9.7.69}$$

где $d^*_{max} = (1/T)\,\Sigma\lambda_l$ — средняя мощность источника. Подставляя (9.7.69), (9.7.51), (9.7.40) и (9.7.45) в (9.7.49), получаем

$$\overline{d}_T \leqslant d^* + \frac{3\delta}{4} + \sqrt{3P(B)\,d^*_{max}}. \tag{9.7.70}$$

Наконец, $P(B)$ оценивается сверху правой частью (9.7.63). Доказательство будет завершено, если будет показано, что это выражение стремится к 0 при $T \to \infty$. Заметим, что L является функцией T и из теоремы Каца, Мардока и Сеге следует (см. лемму 8.5.2), что

$$\lim_{T\to\infty} \frac{L}{T} = \int\limits_{f:\,F(f)\geqslant 1/(2\rho)} df. \tag{9.7.71}$$

Если $F(f) = 1/(2\rho)$ в ненулевой области частот, то (9.7.71) не справедливо, но L/T может быть ограничена сверху в пределе при больших T, если использовать меньшее значение ρ в (9.7.71). Следовательно, $P(B)$ стремится к нулю при $T \to \infty$, и для достаточно больших T имеем $\overline{d}_T \leqslant d^* + \delta$, что завершает доказательство. |

9.8. ДИСКРЕТНЫЕ ЭРГОДИЧЕСКИЕ ИСТОЧНИКИ

В этом параграфе рассматривается дискретный стационарный эргодический источник с алфавитом (0, 1, .., $K-1$). Выход источника ..., u_{-1}, u_0, u_1, ... представляется у адресата последовательностью букв ..., v_{-1}, v_0, v_1, ..., каждая из которых выбирается из алфавита (0, 1, .., $J-1$). Предположим, что имеется мера искажения $d(\mathbf{u}, v_0)$ между последовательностью букв источника и отдельными буквами адресата. Например, если адресата интересует только появление пар последовательных единиц на выходе двоичного источника, то приемлемая мера искажения должна бы иметь вид

$$d(u_{-1}, u_0; v_0) = \begin{cases} 0, & u_{-1}\cdot u_0 = v_0, \\ 1 & \text{в других случаях.} \end{cases}$$

В этом примере искажение для l-й буквы выходной последовательности должно иметь вид $d(u_{l-1}, u_l; v_l)$, где d — указанная выше функция. Для общей меры искажения $d(\mathbf{u}; v_0)$ можно задать искажение для l-й буквы адресата, сначала определяя l-й сдвиг входной последовательности $\mathbf{u}$ равенством $T^l\mathbf{u} = \mathbf{u}'$, где $u'_n = u_{n+l}$. В обозначениях этого оператора сдвига искажение между $\mathbf{u}$ и v_l определяется как $d(T^l\mathbf{u}; v_l)$. Наконец, с помощью меры искажения $d(\mathbf{u}; v_0)$ определим общее искажение между $\mathbf{u}$ и последовательностью $v_0, v_1, \ldots, v_L$ как

$$D(\mathbf{u}; v_0, \ldots, v_L) = \sum_{l=1}^{L} d(T^l\mathbf{u}; v_l). \tag{9.8.1}$$

Для простоты примем в дальнейшем, что $d(\mathbf{u}; v_0)$ неотрицательна и ограничена. Однако не будем предполагать, что для каждого $\mathbf{u}$ можно выбрать v_0, для которого $d(\mathbf{u}; v_0) = 0$. Меры искажения только что

определенного класса назовем *аддитивными инвариантными во времени* мерами искажения.

В последующем изложении сначала определим скорость как функцию искажения для таких источников и мер искажения. Затем докажем теорему кодирования для источника, показав, что эта скорость как функция искажения допускает такое же истолкование, что и скорость как функция искажения для дискретных источников без памяти с мерой искажения для одиночной буквы.

Пусть $\mathbf{u} = (\ldots, u_{-1}, u_0, u_1, \ldots)$ обозначает бесконечную последовательность источника и пусть $\mathbf{u}_L = (u_1, \ldots, u_L)$ обозначает последовательность L букв из $\mathbf{u}$. Аналогично пусть $\mathbf{v}_L = (v_1, \ldots, v_L)$ обозначает соответствующую последовательность букв адресата. Рассмотрим передачу последовательности $\mathbf{u}_L$ по каналу с шумами; пусть $\mathbf{v}_L$ — принимаемая последовательность. Канал и любые преобразования, сопровождающие передачу, можно описать с помощью переходной вероятности L-го порядка $P_L(\mathbf{v}_L \mid \mathbf{u}_L)$. Примем, что при заданном $\mathbf{u}_L$ принятая последовательность статистически не зависит от других букв последовательности бесконечной длины $\mathbf{u}$, т. е. что $P(\mathbf{v}_L \mid \mathbf{u}) = P_L(\mathbf{v}_L \mid \mathbf{u}_L)$. В соединении с вероятностной мерой источника $P_L(\mathbf{v}_L \mid \mathbf{u}_L)$ определяет среднюю взаимную информацию $I(U_1, \ldots, U_L; V_1, \ldots, V_L)$ между последовательностями $\mathbf{u}_L$ и $\mathbf{v}_L$. Аналогично $P_L(\mathbf{v}_L \mid \mathbf{u}_L)$ определяет среднее значение общего искажения*) между $\mathbf{u}$ и $\mathbf{v}_L$, т. е. $\overline{D(\mathbf{u}; \mathbf{v}_L)}$.

Скорость как функция искажения L-го порядка для заданных источника и меры искажения определяется как

$$R_L(d^*) = \min_{P_L : \frac{1}{L}\overline{D} \leqslant d^*} \frac{1}{L} I(U_1, \ldots, U_L; \; V_1, \ldots, V_L). \qquad (9.8.2)$$

Минимизация проводится по всем заданиям переходных вероятностей $P_L(\mathbf{v}_L \mid \mathbf{u}_L)$, таким, что среднее искажение на букву $(1/L)\overline{D(\mathbf{u}; \mathbf{v}_L)}$ не превосходит d^*. Для тех ситуаций, когда $\min_{v_0} d(\mathbf{u}; v_0)$ не равен 0 при всех $\mathbf{u}$, множество ограничений в приведенной выше минимизации может быть пусто для малых d^* и в этих случаях полагаем $R_L(d^*)$ равной ∞. Так же как и в § 9.2, показывается, что функция $R_L(d^*)$ неотрицательная невозрастающая и выпуклая $\cup$ по d^*. Скорость как функция искажения для источника определяется как

$$R(d^*) = \lim_{L \to \infty} R_L(d^*). \qquad (9.8.3)$$

Следующая простая теорема утверждает, что этот предел существует, а также что для любого L значение $R_L(d^*)$ ограничивает сверху $R(d^*)$.

*) Это среднее будет существовать, если при любом выборе v_0 функция $d(\mathbf{u}; v_0)$ будет измеримой на борелевском поле множеств входных последовательностей, на котором вероятностная мера источника определена. Этот класс содержит любую функцию, представляющую какой-либо интерес для практических целей.

Теорема 9.8.1.

$$\inf_L R_L(d^*) = \lim_{L\to\infty} R_L(d^*). \tag{9.8.4}$$

Доказательство. Пусть L и n — произвольные положительные целые числа и для заданного d^* пусть $P_L(\mathbf{v}_L \mid \mathbf{u}_L)$ и $P_n(\mathbf{v}_n \mid \mathbf{u}_n)$ — переходные вероятности, на которых достигаются $R_L(d^*)$ и $R_n(d^*)$ соответственно. Рассмотрим тест-канал, по которому передаются $L + n$ символов и используется P_L для первых L символов и независимо P_n — для последних n, т. е.

$$P_{L+n}(\mathbf{v}_{L+n} \mid \mathbf{u}_{L+n}) =$$

$$= P_L(\mathbf{v}_L \mid \mathbf{u}_L) P_n(v_{L+1}, \ldots, v_{L+n} \mid u_{L+1}, \ldots, u_{L+n}). \tag{9.8.5}$$

Обозначим через $\mathbf{U}_1$ ансамбль первых L букв источника $u_1, \ldots, u_L$ и через $\mathbf{U}_2$ ансамбль последующих n букв $u_{L+1}, \ldots, u_{L+n}$. Аналогично, пусть $\mathbf{V}_1$ и $\mathbf{V}_2$ обозначают ансамбли первых L и последующих n букв адресата соответственно. С помощью тех же рассуждений, как в теореме 4.2.1, имеем

$$I(\mathbf{U}_1\mathbf{U}_2;\ \mathbf{V}_1\mathbf{V}_2) = H(\mathbf{V}_1\mathbf{V}_2) -$$
$$-H(\mathbf{V}_1\mathbf{V}_2 \mid \mathbf{U}_1\mathbf{U}_2),$$
$$H(\mathbf{V}_1\mathbf{V}_2) \leqslant H(\mathbf{V}_1) + H(\mathbf{V}_2).$$

А также из независимости каналов вытекает [см. (9.8.5)]

$$H(\mathbf{V}_1\mathbf{V}_2 \mid \mathbf{U}_1\mathbf{U}_2) = H(\mathbf{V}_1 \mid \mathbf{U}_1) + H(\mathbf{V}_2 \mid \mathbf{U}_2).$$

Отсюда следует, что

$$I(\mathbf{U}_1\mathbf{U}_2;\ \mathbf{V}_1\mathbf{V}_2) \leqslant I(\mathbf{U}_1;\ \mathbf{V}_1) + I(\mathbf{U}_2;\ \mathbf{V}_2). \tag{9.8.6}$$

По определению P_L и P_n и в силу стационарности источника правая часть (9.8.6) равна $LR_L(d^*) + nR_n(d^*)$. Также из (9.8.1) вытекает, что общее среднее искажение $L + n$ символов равно общему среднему искажению первых L символов, сложенному с общим средним искажением последних n символов. Следовательно, среднее искажение на символ для $L + n$ символов не более d^*, и $(L + n) R_{L+n}(d^*)$ является нижней границей левой части (9.8.6). Следовательно,

$$(L+n) R_{L+n}(d^*) \leqslant LR_L(d^*) + nR_n(d^*).$$

Теорема вытекает из леммы 4А.2. |

Теперь легко видеть, что функция $R(d^*)$, так же как и $R_L(d^*)$, не отрицательная не возрастающая с d^* и выпуклая $\smile$.

Обращение теоремы кодирования для источника, данное в теореме 9.2.2, применимо без изменения для источников и мер искажения, рассмотренных здесь. Однако сама теорема кодирования требует некоторых дополнительных рассмотрений. Сначала установим вспомогательный результат, который полезен сам по себе.

Теорема 9.8.2. Пусть $R_1(d^*)$ — скорость как функция искажения первого порядка для дискретного эргодического источника с аддитивной инвариантной во времени мерой искажения. Для любого d^*, тако-

го, что $R_1(d^*) < \infty$, любого $\delta > 0$ и любого достаточно большого L существует блоковый код источника длины L с $M \leqslant \exp[LR_1(d^*) + L\delta]$ кодовыми словами, для которого среднее искажение на букву удовлетворяет неравенству

$$\bar{d}_L \leqslant d^* + \delta. \tag{9.8.7}$$

Для доказательства теоремы нам потребуется следующая лемма.

Лемма 9.8.1. Пусть выход ..., u_{-1}, u_0, u_1, ... дискретного эргодического источника пропускается через дискретный канал без памяти с выходом ..., v_{-1}, v_0, v_1, ... Тогда совместный процесс ..., $u_{-1}v_{-1}$, u_0v_0, u_1v_1, ... является эргодическим.

Мы опустим формальное доказательство этой леммы, обобщение которой доказано Вольфовицем (1961) (§ 10.3) и Файнстейном (1958). Идея доказательства состоит в следующем. Пусть $\mathbf{u}'_l$ — любая частная последовательность l букв источника и $\mathbf{v}'_l$ — какая-либо последовательность l выходов канала. Так как источник эргодический, то относительная частота появления $\mathbf{u}'_l$ на всех начальных позициях в последовательности источника длины $L > l$ стремится к $Q_l(\mathbf{u}'_l)$, т. е. к вероятности $\mathbf{u}'_l$, с вероятностью 1 при $L \to \infty$. Так как канал без памяти, то относительная частота $\mathbf{v}'_l$ по тем начальным позициям, где $\mathbf{u}'_l$ появляется, стремится к $P_l(\mathbf{v}'_l \mid \mathbf{u}'_l)$ с вероятностью 1. Следовательно, относительная частота $\mathbf{u}'_l$, $\mathbf{v}'_l$ сходится к $Q_l(\mathbf{u}'_l)P_l(\mathbf{v}'_l/\mathbf{u}'_l)$, что достаточно для эргодичности.

Доказательство теоремы. Пусть $P(j \mid k)$ — переходная вероятностная мера, на которой достигается $R_1(d^*)$ для заданного d^*, и пусть $Q(k)$ обозначает вероятность отдельной буквы источника. Пусть

$$\omega(j) = \sum_k Q(k) P(j \mid k).$$

Для любого L рассмотрим ансамбль кодов с $M = \lfloor \exp[LR_1(d^*) + L\delta] \rfloor$ кодовыми словами, в которых каждая буква каждого кодового слова выбрана независимо с распределением вероятностей $\omega(j)$. В любом коде из ансамбля каждая последовательность источника $\mathbf{u}$ отображается в кодовое слово, которое минимизирует искажение с $\mathbf{u}$. Пусть

$$P_L(\mathbf{v}_L \mid \mathbf{u}_L) = \prod_{l=1}^{L} P(v_l \mid u_l),$$

пусть

$$\omega_L(\mathbf{v}_L) = \prod_{l=1}^{L} \omega(v_l)$$

и определим

$$I_1(\mathbf{u}_L;\ \mathbf{v}_L) = \ln \frac{P_L(\mathbf{v}_L \mid \mathbf{u}_L)}{\omega_L(\mathbf{v}_L)}. \tag{9.8.8}$$

Заметим, что $I_1(\mathbf{u}_L;\ \mathbf{v}_L)$ не является взаимной информацией между $\mathbf{u}_L$ и $\mathbf{v}_L$ при использовании источника на входе тест-канала без памяти

с переходными вероятностями, задаваемыми $P(j|k)$, так как $\omega_L(\mathbf{v}_L)$ не является вероятностью $\mathbf{v}_L$ выхода тест-канала.

Пусть $\hat{d}$ и $\hat{R}$ произвольны ; определим

$$A = \{\mathbf{u},\ \mathbf{v}_L\colon \text{ или } I_1(\mathbf{u}_L;\ \mathbf{v}_L) > L\hat{R} \text{ или } D(\mathbf{u};\ \mathbf{v}_L) > L\hat{d}\}. \qquad (9.8.9)$$

Пусть $P_t(A)$ — вероятность A в ансамбле тест-канала. Заметим, что этот ансамбль содержит бесконечные последовательности на выходе источника, но содержит последовательности адресата $\mathbf{v}_L$ только конечной длины. Пусть $P_c(D > L\hat{d})$ — вероятность в ансамбле кодов и ансамбле выходов источника того, что искажение (между $\mathbf{u}$ и кодовым словом, в которое оно отображается) превысит $L\hat{d}$. Тогда из доказательства леммы 9.3.1 следует, что

$$P_c(D > L\hat{d}) \leqslant P_t(A) + \exp\left(-M\mathrm{e}^{-L\hat{R}}\right). \qquad (9.8.10)$$

Читателю сейчас следует просмотреть это доказательство. Существенная разница состоит в том, что $\omega_L(\mathbf{v}_L)$ было в лемме как выходной вероятностной мерой для тест-канала, так и вероятностной мерой, с которой выбирались кодовые слова. Если заметить, что только последнее свойство было использовано в доказательстве, то становится очевидным, что здесь применимо неравенство (9.8.10).

Теперь пусть $\hat{R} = R_1(d^*) + \delta/2$ и $\hat{d} = d^* + \delta/2$. Как и в теореме 9.3.1, среднее искажение на букву $\bar{d}_L$ по ансамблю кодов удовлетворяет неравенству

$$\bar{d}_L \leqslant d^* + \delta/2 + P_c[D > L(d^* + \delta/2)] \sup_{\mathbf{u},\, v_0} d(\mathbf{u};\ v_0). \qquad (9.8.11)$$

Так как $d(\mathbf{u};\ v_0)$ по предположению ограниченно, то теорема будет доказана, если будет показано, что $P_c[D > L(d^* + \delta/2)]$ сходится к 0 при $L \to \infty$. Как в теореме 9.3.1 последнее слагаемое в (9.8.10) стремится к 0 с возрастанием L. Первое слагаемое оценивается сверху выражениями

$$P_t(A) \leqslant \Pr\{I_1(\mathbf{u}_L;\ \mathbf{v}_L) > L[R_1(d^*) + \delta/2]\} + \\ + \Pr\{D(\mathbf{u};\ \mathbf{v}_L) > L[d^* + \delta/2]\}, \qquad (9.8.12)$$

$$\Pr\{I_1(\mathbf{u}_L;\ \mathbf{v}_L) > L[R_1(d^*) + \delta/2]\} = \\ = \Pr\left[\frac{1}{L}\sum_{l=1}^{L} \ln\frac{P(v_l|u_l)}{\omega(v_l)} > R_1(d^*) + \delta/2\right]. \qquad (9.8.13)$$

Вместе с тем, так как совместный процесс u, v является эргодическим, то из (3.5.13) вытекает, что с вероятностью 1

$$\lim_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L} \ln\frac{P(v_l|u_l)}{\omega(v_l)} = \overline{\ln\frac{P(v_l|u_l)}{\omega(v_l)}} = R_1(d^*). \qquad (9.8.14)$$

Следовательно, вероятность в (9.8.13) стремится к 0 с возрастанием L. Такие же рассуждения применимы к последнему слагаемому в (9.8.12), что завершает доказательство. |

Один из наиболее интересных аспектов предыдущей теоремы состоит в том, что как $R_1(d^*)$, так и ансамбль кодов, использованный в доказательстве, зависит только от меры искажения и вероятностей отдельной буквы источника. Это указывает весьма ясно, что если хороший код источника конструируется только на основе знания вероятностей отдельной буквы источника, то можно с уверенностью сказать, что любая память в источнике только уменьшит среднее искажение по отношению к тому, которое ожидается в случае источника без памяти. К сожалению, это утверждение нельзя сделать точным и не очень трудно указать примеры кодов, для которых среднее искажение источника с памятью превышает среднее искажение для источника без памяти с теми же самыми вероятностями отдельных букв.

Было показано, что для дискретного эргодического источника можно найти коды источника со скоростями, произвольно близкими к $R_1(d^*)$ и со средними искажениями на букву, сколь угодно близкими к d^*. Перейдем теперь к изложению более сильного результата, когда $R_1(d^*)$ может быть заменена на $R(d^*)$. Путь доказательства заключается в том, что сначала выбирается n достаточно большим, так чтобы $R_n(d^*)$ было близко к $R(d^*)$. Затем рассматриваются последовательности из n букв источника как единые буквы «супериcточника», который порождает одну букву каждые n единиц времени. Затем применим предыдущую теорему к этому суперисточнику и после соответствующей нормировки по n получим требуемый результат. Имеется небольшая трудность в этой процедуре: суперисточник не обязательно будет эргодическим. Чтобы обойти ее, сначала покажем, что суперисточник может быть разбит не более чем на n «эргодических компонент» и затем применим предыдущую теорему в каждой компоненте отдельно.

Предположим теперь, что дискретный эргодический источник имеет алфавит объема K и рассмотрим суперисточник n-го порядка с алфавитом объема K^n, где каждая буква суперисточника является последовательностью n букв первоначального источника. Каждой последовательности $\mathbf{u} = (\ldots, u_{-1}, u_0, u_1, \ldots)$ первоначального источника соответствует последовательность $\mathbf{u}' = (\ldots, u'_{-1}, u'_0, u'_1, \ldots)$ суперисточника, где $u'_1 = (u_1, u_2, \ldots, u_n)$, $u'_2 = (u_{n+1}, \ldots, u_{2n})$ и т. д. Обозначим это соответствие между последовательностями первоначального источника и суперисточника в виде $\mathbf{u} \leftrightarrow \mathbf{u}'$. Так как сдвиг на n единиц времени в первоначальном источнике соответствует сдвигу на одну единицу времени для суперисточника, то имеем $T^{ln}\mathbf{u} \leftrightarrow T^l\mathbf{u}'$ для $\mathbf{u} \leftrightarrow \mathbf{u}'$. Аналогично, любое множество S последовательностей первоначального источника соответствует множеству S' суперисточника и $T^{ln}S \leftrightarrow T^lS'$. Напомним, что, как указано в § 3.5, инвариантное множество S является множеством, для которого $TS = S$, и что источник эргодический тогда и только тогда, когда любое измеримое инвариантное множество имеет вероятность 1 или 0.

Предположим теперь, что S'_0 — инвариантное множество последовательностей суперисточника и что его вероятность $Q'(S'_0)$ больше 0. Соответствующее множество $S_0 \leftrightarrow S'_0$ первоначального источника имеет вероятность $Q(S_0) = Q'(S'_0)$ и удовлетворяет равенству $T^nS_0 = S_0$. Определим множества S_i, $1 \leqslant i \leqslant n-1$, равенствами

$$S_i = T^i S_0. \tag{9.8.15}$$

Так как источник стационарный, то $Q(S_i) = Q(S_0)$ для $1 \leqslant i \leqslant n-1$. Пусть теперь множество A является объединением множеств S_0, S_1, ..., S_{n-1}. Тогда имеем

$$TA = T\left(\bigcup_{i=0}^{n-1} S_i\right) = \bigcup_{i=0}^{n-1} TS_i = S_1 \cup S_2 \cup \ldots \cup TS_{n-1}. \tag{9.8.16}$$

Поскольку $TS_{n-1} = TT^{n-1}S_0 = S_0$, то это сводится к равенству $TA = A$. Так как источник эргодический и $Q(A) > 0$, то должно быть $Q(A) = 1$ и, следовательно, $Q(S_0) = Q'(S'_0) \geqslant 1/n$. Таким образом, показано, что каждое измеримое инвариантное множество для супериcточника имеет вероятность 0 или вероятность, не меньшую $1/n$.

Далее предположим, что S_0' — инвариантное множество последовательностей супериcточника и что B' — инвариантное подмножество S'_0 с $0 < Q'(B') < Q'(S'_0)$. Пусть $S'_0 - B'$ — последовательности, содержащиеся в S'_0, но не содержащиеся в B', заметим, что $T(S'_0 - B') = TS'_0 - TB' = S'_0 - B'$, так что $S'_0 - B'$ — также инвариантное подмножество S'_0. Из предыдущего результата следует, что $1/n \leqslant Q'(B') \leqslant Q'(S'_0) - 1/n$. Если теперь представить, что B' играет роль S' и повторить приведенные выше рассуждения, то увидим, что постепенно мы должны попасть в S'_0, которое не имеет инвариантного подмножества B' с $0 < Q'(B') < Q'(S'_0)$. Будем называть такое S'_0 *эргодической компонентой* супериcточника. Источник, который порождает супербуквы в соответствии с условной вероятностной мерой Q' при условии, что задано S'_0, является, очевидно, эргодическим источником, так как при условии, что задано S'_0, каждое инвариантное подмножество S'_0 имеет вероятность 0 или 1.

Пусть теперь $S_0 \leftrightarrow S'_0$, пусть $S_i = T^i S_0$ и пусть $S_i \leftrightarrow S'_i$. Каждое S'_i должно также образовывать эргодическую компоненту супериcточника, т. е. если B'_i — подмножество S'_i и $TB'_i = B'_i$, то для соответствующего множества B_i первоначального источника имеем $T^n B_i = B_i$. Положив, что $B_0 = T^{-i}B_i$, получаем, что B_0 — подмножество S_0 и $T^n B_0 = B_0$. Следовательно, B'_0 — подмножество S'_0 и $TB'_0 = B'_0$. Так как $Q'(B'_i) = Q'(B'_0)$, то будет или $Q'(B'_i) = 0$ или $Q'(B'_i) = Q'(S'_i)$. Наконец, рассмотрим пересечение $S'_i \cap S'_j$. Это пересечение является инвариантным множеством и является подмножеством как множества S'_i, так и множества S'_j. Таким образом, или $Q'(S'_i \cap S'_j) = 0$ или $Q'(S'_i \cap S'_j) = Q'(S'_i)$. В последнем случае S'_i и S'_j — одни и те же множества (исключая, быть может, разность вероятности нуль). Легко видеть, что если S'_i и S'_j совпадают, то совпадают S'_{i+k} и S'_{j+k}, где $i+k$ и $j+k$ взяты по модулю n. Отсюда легко следует, что число различных эргодических компонент, скажем n', является делителем n и что в качестве эргодических компонент могут быть взяты S'_0, S'_1, ..., $S'_{n'-1}$. В этом случае каждая эргодическая компонента имеет вероятность $1/n'$. В случае, представляющем наибольший физический интерес, $n' = 1$ и супериcточник с самого начала эргодический.

Следующая лемма подытоживает результаты.

Л е м м а 9. 8. 2. Рассмотрим суперисточник n-го порядка, буквы которого — последовательности из n букв дискретного эргодического источника. Множество последовательностей суперисточника может быть разбито на n' эргодических компонент, каждая из которых имеет вероятность $1/n'$, где n' является делителем n. Компоненты не пересекаются, за исключением, быть может, множеств вероятности нуль. Множества последовательностей S_0, S_1, ..., $S_{n'-1}$ первоначального источника, соответствующие этим эргодическим компонентам, можно связать соотношениями $T(S_i) = S_{i+1}$, $0 \leqslant i \leqslant n'-2$ и $T S_{n'-1} = S_0$.

В качестве примера рассмотрим двоичный источник, для которого выход образован парами одних и тех же символов. С вероятностью $^1/_2$ каждый символ на четной позиции получается в результате независимого равновероятного выбора символов алфавита, а каждый символ с нечетным номером совпадает с предыдущим символом с четным номером. Аналогично с вероятностью $^1/_2$ каждый символ с нечетным номером получается в результате независимого равновероятного выбора символов, а каждый символ с четным номером совпадает с предыдущим символом с нечетным номером. Это эргодический источник, однако суперисточник второго порядка имеет две компоненты. На одной компоненте суперисточник не имеет памяти и порождает 00 с вероятностью $^1/_2$ и 11 с вероятностью $^1/_2$. На второй компоненте все четыре пары символов равновероятны, однако последний символ любой пары совпадает с первым символом следующей пары. Очевидно, что в этом примере выход какой-либо компоненты (в символах первоначального источника) статистически эквивалентен выходу другой компоненты, сдвинутой во времени на один символ.

Как было показано, вообще суперисточник n-го порядка, соответствующий эргодическому источнику, может быть разбит на n' эргодических компонент, где n' является делителем n. В последующем изложении удобно рассматривать их как n эргодических компонент, где только n' среди них различны. Определим суперисточник i-й фазы $0 \leqslant i \leqslant n-1$ как источник, который порождает последовательности из S'_i в соответствии с первоначальной вероятностной мерой на последовательностях супербукв при условии наступления S'_i. Тогда для $n' < n$ суперисточники i-й, $(i + n')$-й, $(i + 2n')$-й фаз и т. д. тождественны. Первоначальный суперисточник моделируется как совокупность n эргодических суперисточников, каждый из которых используется во всей бесконечной последовательности с вероятностью $1/n$.

Используем теперь эту модель в конструкции кодов эргодического источника. Для заданного n пусть $R_n(d^*)$—скорость как функция искажения n-го порядка и для заданного d^* с $R_n(d^*) < \infty$ пусть $P_n(\mathbf{v}_n \mid \mathbf{u}_n)$ обозначает множество переходных вероятностей, на которых достигается $R_n(d^*)$. В терминах суперисточника n-го порядка $\mathbf{u}_n$ является единичной буквой u'_1, и аналогично можно рассматривать $\mathbf{v}_n$ как отдельную букву v'_1 в супералфавите последовательностей n букв адресата. Полагая, что U' — ансамбль отдельных супербукв и V' — ансамбль отдельных супербукв адресата, определяемый P_n,

имеем

$$R_n(d^*) = \frac{1}{n} I(U'; V'). \tag{9.8.17}$$

Аналогично

$$d^* \geqslant \frac{1}{n} \overline{D(\mathbf{u}; \mathbf{v}_n)} = \frac{1}{n} \overline{d'(\mathbf{u}'; v'_1)}, \tag{9.8.18}$$

где d' определяется равенством $D(\mathbf{u}; \mathbf{v}_n) = d'(\mathbf{u}'; v'_1)$ и неравенство возникает из-за возможности $R_n(d^*) = 0$.

Пусть теперь $I(U'; V' \mid i)$ — средняя взаимная информация между супербуквой суперисточника i-й фазы, $0 \leqslant i \leqslant n-1$, и буквой супералфавита адресата при использовании переходных вероятностей P_n. Так как средняя взаимная информация в канале — выпуклая $\frown$ функция входных вероятностей, то

$$I(U'; V') \geqslant \frac{1}{n} \sum_{i=0}^{n-1} I(U'; V' \mid i). \tag{9.8.19}$$

Аналогично среднее значение d' является линейной функцией вероятностной меры на u', так что

$$\overline{d'(\mathbf{u}'; v'_1)} = \frac{1}{n} \sum_{i=0}^{n-1} \overline{d'_i}, \tag{9.8.20}$$

где $\overline{d'_i}$ — среднее значение $d'(\mathbf{u}'; v'_1)$ для суперисточника i-й фазы.

Заметим теперь, что $I(U'; V' \mid i)$ является верхней границей для скорости как функции искажения первого порядка для суперисточника i-й фазы, вычисленной при среднем искажении $\overline{d'_i}$. Так как суперисточник i-й фазы является эргодическим, то теорема 9.8.2 применима и для любого $\delta > 0$ при любом достаточно большом L имеется множество $M_i \leqslant \exp[LI(U'; V' \mid i) + L\delta]$ кодовых слов длины L (супербукв), для которых среднее искажение на супербукву для этого источника не больше $\overline{d'_i} + \delta$. Для заданного $\delta > 0$ пусть L достаточно велико, так что для каждой из n фаз такой код может быть выбран, и рассмотрим это множество из n кодов.

Удобно представлять себе кодовые слова в этих кодах не как последовательности L супербукв, а как последовательности Ln букв первоначального алфавита адресата. Будем рассматривать такой i-й код, выбранный выше для суперисточника i-й фазы, как i-й малый код. Используем теперь эти n малых кодов для построения нового множества n больших кодов с длиной блока $Ln^2 + n$, как показано на рис. 9.8.1. Построенный i-й большой код, $0 \leqslant i \leqslant n-1$, кодирует выход суперисточника i-й фазы. Как показано на рис. 9.8.1, для каждого l, $0 \leqslant l \leqslant n-1$, множество букв на позициях от $l[nL+1]+1$ до $l[nL+1]+Ln$ является произвольным кодовым словом из $(i+l)$-го малого кода, где $i+l$ берется по модулю n. Буквы на позициях $Ln+1$, $2[Ln+1]$, ..., $n[Ln+1]$ являются фиксированными буквами алфавита адресата. Таким образом, общее число кодовых слов

в i-м большом коде равно $\prod_{i=0}^{n-1} M_i$, где M_i — число кодовых слов в i-м малом коде. Напомним теперь, что из леммы 9.8.2 следует, что если S_i — множество последовательностей букв первоначального алфавита источника, порожденное супериспочником i-й фазы, то $TS_i = S_{i+1}$. Отсюда следует, что для супериспочника i-й фазы вероятностная мера на буквах позиций от $l(Ln+1)+1$ до $l(Ln+1)+Ln$ такая же, как для супериспочника $(i+l)$-й фазы на позициях от 1 до Ln. Следовательно, среднее искажение на букву между супериспочником i-й фазы и i-м большим кодом на позициях $l(Ln+1)+1$ до $l(Ln+1)+Ln$ составляет $1/n$ часть среднего искажения на супербукву между супериспочником $(i+l)$-й фазы и $(i+l)$-м малым кодом (где $i+l$ взята по модулю n). Выберем малые коды так, чтобы среднее искажение на букву было ограничено сверху выражением $(1/n)[\overline{d'_{i+l}}+\delta]$. Замечая, что искажение между каждой из n фиксированных позиций в коде и источником ограничено сверху $\sup d(\mathbf{u}; v_0)$, находим, что среднее искажение на букву между i-м супериспочником и i-м большим кодом ограничено сверху выражением

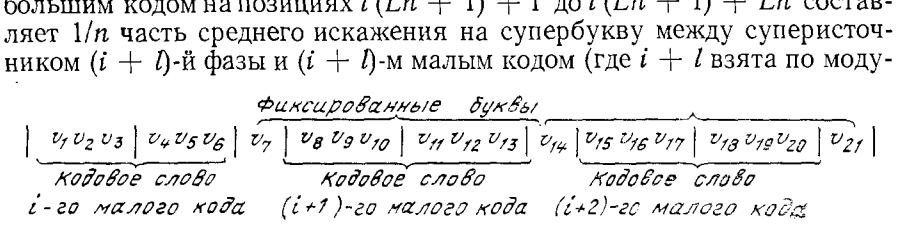


Рис. 9.8.1. Конструирование большого кода из малого кода; $n=3$, $L=2$.

$$\frac{n \sup d(\mathbf{u}; v_0) + \sum_{l=0}^{n-1} L\left(\overline{d'_{i+l}}+\delta\right)}{Ln^2+n}. \tag{9.8.21}$$

Можно дальше построить границу сверху, уменьшая знаменатель до Ln^2 и применяя (9.8.18) и (9.8.20), в результате получим

$$\frac{\sup d(\mathbf{u}; v_0)}{Ln} + d^* + \frac{\delta}{n}. \tag{9.8.22}$$

Следовательно, для достаточно больших L среднее искажение на букву между супериспочником i-й фазы и i-м большим кодом ограничено сверху $d^*+\delta$, где $\delta > 0$ произвольно.

Если все кодовые слова этих n больших кодов объединить в один код, то независимо от того, какой эргодический супериспочник действует, среднее искажение на букву будет не больше $d^*+\delta$. Общее число кодовых слов равно

$$M = n\prod_{i=0}^{n-1} M_i \leqslant n\exp\left\{\sum_{i=0}^{n-1}[LI(U';\ V'|i)+\delta]\right\} \leqslant$$
$$\leqslant n\exp[LnI(U';\ V')+Ln\delta] = n\exp[Ln^2 R_n(d^*)+Ln\delta] \leqslant$$
$$\leqslant \exp\left\{(Ln^2+n)\left[R_n(d^*)+\frac{\delta}{n}+\frac{\ln n}{Ln^2+n}\right]\right\}. \tag{9.8.23}$$

Таким образом показано, что для любого $n > 1$, любого $\delta > 0$ и достаточно большого L можно найти коды длины $Ln^2 + n$ с

$$M \leqslant \exp\{[(Ln^2 + n)[R_n(d^*) + \delta]\}$$

и средним искажением на букву, меньшим или равным $d^* + \delta$. Кроме того, очевидно, ограничение, состоящее в том, что длина блока должна иметь вид Ln^2+n, не является необходимым, так как для блока любой достаточно большой длины L' можно найти наибольшее L, такое, что $Ln^2 + n \leqslant L'$, далее для этого L найти код, удовлетворяющий (9.8.22) и (9.8.23), и затем добавить последовательность, содержащую не более $n^2 - 1$ фиксированных символов, к концу каждого кодового слова. Среднее искажение на букву возрастает, таким образом, не более чем на $\sup d(\mathbf{u}; v_0)/L$, что для больших L пренебрежимо мало. Наконец, так как n может быть выбрано достаточно большим, так чтобы $R_n(d^*)$ было сколь угодно близко к $R(d^*)$, то тем самым доказана следующая фундаментальная теорема.

Теорема 9.8.3. Пусть $R(d^*)$ — скорость как функция искажения для дискретного эргодического источника с аддитивной инвариантной во времени мерой искажения. Для любого d^* с $R(d^*) < \infty$, любого $\delta > 0$ и достаточно больших L существует блоковый код источника длины L с $M \leqslant \exp[LR(d^*) + L\delta]$ кодовыми словами и средним искажением на букву, не большим $d^* + \delta$.

При желании терпеливый и настойчивый читатель может соединить метод доказательства этой теоремы с результатами § 9.3 и 9.6 и снять как условие, что $d(\mathbf{u}; v_0)$ ограничено, так и условие, что источник дискретный.

ИТОГИ И ВЫВОДЫ

В этой главе рассмотрена проблема воспроизведения выхода источника у адресата при выполнении заданного критерия верности. Источник был описан с помощью множества возможных выходов источника и вероятностной меры на этих выходах. За критерий верности было выбрано среднее искажение на букву или единицу времени между источником и адресатом. С точки зрения теории мера искажения является неотрицательной функцией выхода источника и его воспроизведенного варианта. С точки зрения приложений, конечно, мера искажения должна быть выбрана так, чтобы она отражала в некотором смысле стоимость для потребителя любого заданного воспроизведения любого заданного выхода источника.

Мы начали с дискретных источников без памяти с мерами искажения отдельной буквы и затем обобщали результаты на все более и более общие случаи. В каждом случае сначала определялась скорость как функция искажения $R(d^*)$ и затем этой функции придавался смысл с помощью доказательства теоремы кодирования для источника и ее обращения. То, что говорится в теореме кодирования для источника, по существу, сводится к тому, что для заданного d^* выход источника может быть закодирован с помощью $R(d^*)/\ln 2$ двоичных символов на символ источника (или на единицу времени, в зависимости от

единиц измерения $R(d^*)$) и что эти двоичные символы могут быть преобразованы в буквы адресата таким образом, что среднее искажение на букву (или на единицу времени) будет сколь угодно близко к d^*. Этот же результат остается справедливым, если двоичные символы кодируются и передаются по каналу с шумами и пропускной способностью, большей чем $R(d^*)$, где пропускная способность выражается в натах на букву источника (или в натах на единицу времени). Обращение теоремы устанавливает, что если выход источника передается по каналу с пропускной способностью, меньшей чем $R(d^*)$, то независимо от преобразований выхода источника и сообщений, поступающих адресату, среднее искажение на букву (или на единицу времени) должно быть больше d^*.

Читатель должен заметить много аналогий между полученными здесь результатами для кодирования источника и теорией кодирования для канала с шумами, однако может быть полезно обратить здесь внимание на некоторые отличия. Теорема кодирования для канала с шумами связывает достижимую вероятность ошибочного декодирования P_e с длиной кодового блока и скоростью кода R. Было найдено, что для фиксированного R, меньшего чем пропускная способность канала, P_e убывает экспоненциально с увеличением длины блока. При кодировании источника эквивалентными параметрами, представляющими интерес, являются среднее искажение на букву $\bar{d}$, длина кодового блока L и скорость кода R. Здесь для фиксированного R при возрастании L число $\bar{d}$ убывает к d^*, задаваемому $R = R(d^*)$. Сходимость $\bar{d}$ к d^* происходит не медленнее чем $\sqrt{(\ln L)/L}$ и, по-видимому, не быстрее чем $1/L$ [см. Пилк (1967)]. Таким образом, при кодировании источника следует затратить весьма много усилий (в смысле увеличения длины блока) для того, чтобы достичь весьма незначительного уменьшения $\bar{d}$, в то же время для каналов с шумами весьма умеренное возрастание длины блока может привести к сильному убыванию вероятности ошибки.

В качестве примера (быть может, нетипичного) того, сколь немногое может быть достигнуто утонченным кодированием источника, можно рассмотреть гауссовский дискретный по времени источник с квадратично-разностной мерой искажения. Гоблик (1967) и Макс (1960) рассматривали среднее искажение, которое может быть достигнуто с помощью квантования (т. е. кода источника с единичной длиной блока). Они нашли, что если квантованные буквы закодированы без шумов (с помощью метода Хаффмана), то для малых значений искажения среднее искажение примерно лишь на $^1/_4$ дБ больше минимального, предсказываемого кривой $R(d^*)$. Если также не использовать кодирование Хаффмана, то потеря все же будет меньше 1 дБ.

Другое основное отличие между кодированием для канала и кодированием источника проявляется в трудности получения приемлемых вероятностных моделей и разумных мер искажения для источников, представляющих практический интерес. В силу этой трудности вообще неясно, будет ли теоретический подход полезным в таких задачах, как

преобразование речи в дискретные данные или сужение полосы частот в телевидении.

ИСТОРИЧЕСКИЕ ЗАМЕЧАНИЯ И ССЫЛКИ

Большинство результатов этой главы принадлежит Шеннону (1959). Доказательства леммы 9.3.1 и теоремы 9.3.1 используют как методы первоначального доказательства Шеннона, так и последующие методы, развитые Гобликом и Стиглитцем. Они проще, чем первоначальные методы, и приводят к лучшим результатам о сходимости к $R(d^*)$ для кодов с увеличивающейся длиной блока. Распространение теории на случай, когда $d(u; v)$ принимает бесконечные значения, и теорема 9.3.2 взяты у Пинкстона (1967). Теорема 9.4.1 публикуется здесь впервые, хотя нижняя граница $R_0(\rho, \mathbf{P})$ в (9.4.10) была выведена Шенноном в частном случае квадратично-разностной меры искажения. Обращение теоремы кодирования для канала с шумами в § 9.5 взято у Пинкстона (1967). Вычисление скорости как функции искажения для гауссовского дискретного по времени источника проведено Шенноном (1948), и скорость как функция искажения для гауссовского случайного процесса найдена Колмогоровым (1956). По-видимому, теорема кодирования для источника, порождающего гауссовский случайный процесс, ранее не появлялась в литературе. Результаты § 9.8 новые, хотя аналогичная теорема, не требующая, чтобы источник был дискретным, но требующая выполнения несколько более сильного условия чем эргодичность, была получена Гобликом (1967).

Глава 2

2.1. Три события E_1, E_2, E_3, определенные на одном и том же пространстве, имеют вероятности $P(E_1) = P(E_2) = P(E_3) = 1/4$.

Пусть E_0 является событием, состоящим в том, что имеют место одно или более событий E_1, E_2, E_3.

(а) Найти вероятность $P(E_0)$, если:

1) События E_1, E_2, E_3 несовместны.

2) События E_1, E_2, E_3 статистически независимы.

3) События E_1, E_2, E_3 совпадают.

(б) Найти максимальные значения, которые может принимать $P(E_0)$, когда:

1) Ничего не известно относительно независимости или несовместности событий E_1, E_2, E_3.

2) Известно, что события E_1, E_2, E_3 являются попарно независимыми, т. е. что вероятность наступления событий E_i и E_j равна $P(E_i)P(E_j)$, $1 \leqslant i \neq j \leqslant 3$; но ничего не известно относительно вероятности наступления всех трех событий вместе.

Указание: используйте диаграммы Венна.

2.2. Нечестный игрок использует шулерскую игральную кость, которая с вероятностью $2/3$ показывает цифру 1, а цифры от 2 до 6 показывает каждую с вероятностью $1/15$. Он неудачно опустил свою игральную кость в ящик с двумя правильными игральными костями и не может разделить их. Он выбирает случайно одну кость из ящика, бросает ее и появляется цифра 1. При условии этого результата найти вероятность того, что он выбрал шулерскую кость. После этого он бросает кость еще раз и она снова показывает 1. Чему равна вероятность (после этого второго бросания) того, что выбранная кость была шулерской?

2.3. Пусть x и y — дискретные случайные величины.

(а) Доказать, что математическое ожидание суммы x и y, $\overline{x + y}$, равно сумме математических ожиданий $\bar{x} + \bar{y}$.

(б) Доказать, что если x и y статистически независимы, то x и y также не коррелированы (по определению x и y некоррелированы, если $\overline{xy} = \bar{x}\,\bar{y}$). Придумать пример, в котором x и y статистически зависимы, но некоррелированы, и другой пример, в котором x и y статистически зависимы и коррелированы (т. е. $\overline{xy} \neq \bar{x}\,\bar{y}$).

(в) Показать, что если x и y статистически независимы, то дисперсия суммы равна сумме дисперсий. Остается ли справедливым это утверждение, если x и y некоррелированы, но не статистически независимы?

2.4. (а) Одним из видов неравенства Чебышева является следующее утверждение: для любой случайной величины x, которая принимает только неотрицательные значения, и для любого $\delta > 0$

$$\Pr[x \geqslant \delta] \leqslant \frac{\bar{x}}{\delta}.$$

Показать, что это неравенство справедливо и что для любого *заданного* $\delta > 0$ можно указать случайную величину, для которой это неравенство удовлетворяется с равенством. *Указание*: представить $\bar{x}$ в виде суммы и затем ограничить область суммирования.

(б) Пусть y — случайная величина со средним значением $\bar{y}$ и дисперсией σ_y^2. Считая, что рассмотренной выше случайной величиной x является $(y - \bar{y})^2$,

показать, что для любого $\varepsilon > 0$

$$\Pr[|y-\bar{y}| \geqslant \varepsilon] \leqslant \frac{\sigma_y^2}{\varepsilon^2}.$$

(в) Пусть $z_1, z_2, \ldots, z_N$ — последовательность статистически независимых одинаково распределенных случайных величин со средним значением $\bar{z}$ и дисперсией σ_z^2. Пусть y_N — выборочное среднее значение

$$y_N = \frac{1}{N}\sum_{n=1}^{N} z_n.$$

Показать, что для любого $\varepsilon > 0$

$$\Pr[|y_N-\bar{y}_N| \geqslant \varepsilon] \leqslant \frac{\sigma_z^2}{N\varepsilon^2}. \qquad (*)$$

Указание: сначала найдите дисперсию Σz_n, используя задачу 2.3 (в).

Найдите предел при $N \to \infty$ вероятности $\Pr[|y_N - \bar{y}_N| \geqslant \varepsilon]$ для любого фиксированного $\varepsilon > 0$ (этот результат называется законом больших чисел).

(г) Пусть случайные величины $z_1, \ldots, z_N$, определенные в пункте (в), соответствуют N независимым испытаниям в эксперименте. Пусть E — некоторое заданное событие и пусть $z_n = 1$, если E происходит при n-м испытании и $z_n = 0$ в противном случае. Выразить словами утверждение неравенства (*) в этом случае и вычислить $\bar{y}_N$ и σ_z^2. Найти также точное выражение для $\Pr[|y_N - \bar{y}_N| \geqslant \varepsilon]$.

2.5. Источник производит статистически независимые равновероятные буквы из алфавита (a_1, a_2) со скоростью 1 буква каждые 3 с. Эти буквы передаются по двоичному симметричному каналу так, что каждую секунду передается один символ; буква источника a_1 кодируется в кодовое слово 000 и буква a_2 кодируется в кодовое слово 111*). Если на выходе канала в течение интервала, равного 3 с, принимается одна из последовательностей 000, 001, 010, 100, то декодируется a_1, в противном случае декодируется a_2. Пусть $\varepsilon < 1/2$ является переходной вероятностью канала (см. рис. 1.3.1).

(а) Для каждой возможной последовательности из трех принятых символов на интервале, соответствующем передаче заданной буквы источника, найти условную вероятность того, что a_1 была на выходе источника.

(б) Используя пункт (а), показать, что приведенное выше правило декодирования минимизирует вероятность неправильного решения.

(в) Найти вероятность неправильного решения (использование для этого пункта (а) не является самым простым способом).

(г) Предположим, что источник работает с более низкой скоростью, производя одну букву за каждые $2n + 1$ секунд; a_1 кодируются в $2n + 1$ символов 0 и a_2 в $2n + 1$ символов 1. Найти правило решения, минимизирующее вероятность неправильного решения при декодировании. Найти предел этой вероятности неправильного решения при $n \to \infty$.

Указание: используйте закон больших чисел.

2.6. Среди всех женщин X имеются $1/4$ блондинок, $1/2$ брюнеток и $1/4$ шатенок; блондинки никогда не опаздывают на свидания; шатенки всегда опаздывают; а каждая брюнетка всегда подбрасывает симметричную монету, чтобы решить, следует ли торопиться или не спешить на каждое свидание.

(а) Какое количество информации содержится в утверждении «x, элемент X, прибыла вовремя» относительно каждого из следующих предложений:

---

*) В условиях задач и их решениях не делается различия в написании элементов поля, элементов последовательностей и натуральных чисел (в отличие от основного текста). (*Прим. ред.*)

1) x — блондинка;
2) x — брюнетка;
3) x — шатенка?

(б) Сколько информации содержится в утверждении «x, элемент X, пришла вовремя на три свидания подряд» относительно предложения «x — брюнетка»?

2.7. Для передачи по двоичному симметричному каналу с переходной вероятностью ε (см. рис. 2.2.1) множество из восьми равновероятных сообщений кодируется в следующее множество из восьми кодовых слов:

$$\begin{aligned} x_1 &= 0000, & x_5 &= 1001, \\ x_2 &= 0011, & x_6 &= 1010, \\ x_3 &= 0101, & x_7 &= 1100, \\ x_4 &= 0110, & x_8 &= 1111. \end{aligned}$$

Пусть на выходе канала принимается последовательность $\mathbf{y} = 0000$, определить:

(а) Количество информации о $\mathbf{x}_1$, содержащейся в первом принятом символе.

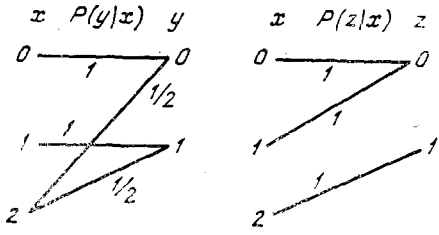
(б) Дополнительное (условное) количество информации о $\mathbf{x}_1$, содержащееся во втором принятом символе, третьем символе, четвертом символе.

2.8. Рассмотрим ансамбль последовательностей $x_1, x_2, \ldots, x_N$, состоящих из N двоичных символов. Каждая последовательность, содержащая четное число единиц, имеет вероятность 2^{-N+1}, а каждая последовательность с нечетным числом единиц имеет нулевую вероятность. Найти средние взаимные информации

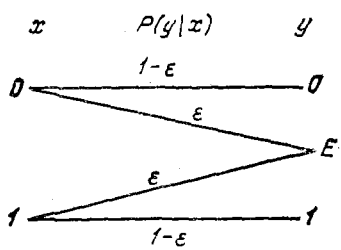
$$I(X_2; X_1),\ I(X_3; X_2 | X_1), \ldots,\ I(X_N; X_{N-1} | X_1 \ldots X_{N-2}).$$

Проверить ваш результат для $N = 3$.

2.9. Источник X производит буквы из алфавита, содержащего три символа, с вероятностями $P_X(0) = 1/4$, $P_X(1) = 1/4$, $P_X(2) = 1/2$. Каждая буква источника x непосредственно передается сразу по двум каналам; y и z выходы каналов, а переходные вероятности указаны ниже.



К задаче 2.9



К задаче 2.10

(Заметьте, что эту передачу можно было бы рассматривать как передачу по одному каналу с выходом yz.)

Найти $H(X)$, $H(Y)$, $H(Z)$, $H(YZ)$, $I(X; Y)$, $I(X; Z)$, $I(X; Y|Z)$, $I(X; YZ)$. Дать толкование этим выражениям взаимной информации.

2.10. Двоичный стирающий канал, который изображен выше, является каналом с «шумом» особенно простого типа, при котором переданные символы могут быть «стерты», но никогда не могут быть приняты ошибочно.

(а) Пусть $P_X(0) = p = 1 - P_X(1)$. Выразить $I(X; Y)$ через p и найти значение p, которое максимизирует $I(X; Y)$. Для этого максимизирующего p найти $I(x; y)$ во всех точках ансамбля XY, а также найти $I(X; Y)$.

(б) Предположим, что последовательность статистически независимых и равновероятных двоичных символов требуется передать по двоичному стирающему каналу. Предположим далее, что имеется бесшумный канал обратной связи от приемника, по которому отправитель узнает, как был принят каждый символ. Рассмотреть следующую стратегию передачи последовательности с абсолютной

надежностью. Посылать символы по порядку по каналу, повторяя каждый символ, если он был стерт, до тех пор, пока он не будет правильно принят. Найти среднее число переданных символов при одном использовании канала.

Замечание. Интуитивно ясно, что рассмотренная здесь схема передачи приводит к оптимальному использованию канала. Заметьте фундаментальную роль, которую играет $I(X; Y)$.

2.11. Пусть $x = a_1$ означает событие, состоящее в том, что шар при игре в рулетку останавливается в «красной» лунке, а $x = a_2$ аналогично означает, что шар останавливается в «черной». Предположим, что игорный дом не берет налог, т. е. что пари, заключенное на один доллар, относительно того, будет ли лунка красной или черной, приносит один дополнительный доллар в случае успеха. Будем считать, что $P_X(a_1) = P_X(a_2) = {}^1/_2$.

Крупье рулетки разработал план для обмана игорного дома. После многих лет терпеливого изучения он научился частично предсказывать цвет, наблюдая путь шара вплоть до самого последнего момента, когда пари может быть еще заключено. Сообщая эти сведения сообщнику, крупье рассчитывает использовать свое провидение, чтобы получить кругленькую сумму ко времени, когда он уйдет в отставку.

Пусть y означает сигнал крупье; покашливание, $y = b_1$ означает предсказание красного, а подмигивание, $y = b_2$, означает предсказание черного. Предположим, что $P_{X|Y}(a_1 | b_1) = P_{X|Y}(a_2 | b_2) = {}^3/_4$ и что последовательные вращения независимы.

(а) Найти $I(X; Y)$.

(б) Сообщник имеет некоторый начальный капитал C_0. Он решил ставить фиксированную долю $1 - q$ его текущего капитала на предсказываемый цвет при каждом последовательном пари и долю q на другой цвет (заметьте, что это эквивалентно ставке доли $1 - 2q$ на предсказываемый цвет и отказу ставить $2q$).

Показать, что после N игр капитал сообщника является случайной величиной

$$C_N = C_0 \prod_{n=1}^{N} [2(1-q)]^{z_n} [2q]^{1-z_n},$$

где $z_n = 1$, если предсказание правильно, и 0 в противном случае. Определим скорость роста:

$$E_N = \frac{1}{N} \log_2 \frac{C_N}{C_0}; \quad C_N = C_0 2^{NE_N}.$$

Найти значения q, которые максимизируют математические ожидания $\overline{C}_N$ и $\overline{E}_N$. Сравните максимальное значение математического ожидания скорости роста $\overline{E}_N$ с $I(X; Y)$.

(в) Если бы вы были сообщником, какое бы значение q вы использовали и почему?

Указание: применим ли закон больших чисел к E_N или C_N при $N \to \infty$?

См. Дж. Л. Келли (1956) для знакомства с детальным анализом такого типа задач.

2.12. Выборочное пространство ансамбля X состоит из неотрицательных целых чисел. Найти вероятности $P_X(n)$, $n = 0, 1, \ldots$, которые максимизируют $H(X)$ при условии, что среднее значение X

$$\sum_{n=0}^{\infty} nP_X(n)$$

равно заданному значению A. Вычислить максимальное значение $H(X)$.

2.13. Запись погоды в некотором городе дается в приводимой ниже таблице; числа указывают относительную частоту соответствующих событий.

Предсказание	На самом деле: Дождь	На самом деле: Нет дождя
Дождь	$^1/_8$	$^3/_{16}$
Нет дождя	$^1/_{16}$	$^{10}/_{16}$

Умный студент заметил, что составитель прогноза прав только в $^{12}/_{16}$ случаев, но мог бы быть правым в $^{13}/_{16}$ случаев, предсказывая всегда отсутствие дождя. Студент объяснил ситуацию и написал заявление о приеме на работу в бюро прогнозов, но начальник бюро прогнозов, который был специалистом по теории информации, отклонил его заявление. Почему?

2.14. Пусть X — ансамбль, состоящий из M точек $a_1, \ldots, a_M$, и пусть $P_X(a_M) = \alpha$. Показать, что

$$H(X) = \alpha \log \frac{1}{\alpha} + (1-\alpha) \log \frac{1}{1-\alpha} + (1-\alpha) H(Y),$$

где Y является ансамблем из $M-1$ точек с $a_1, \ldots, a_{M-1}$ с вероятностями

$$P_Y(a_j) = P_X(a_j)/(1-\alpha); \quad 1 \leqslant j \leqslant M-1.$$

Показать далее, что

$$H(X) \leqslant \alpha \log \frac{1}{\alpha} + (1-\alpha) \log \frac{1}{1-\alpha} + (1-\alpha) \log (M-1)$$

и найти условие, при котором выполняется равенство.

2.15. Энтропия дискретного ансамбля рассматривается как мера неопределенности. Обосновать это истолкование, доказав, что любое преобразование вероятностей двух элементов ансамбля, которое делает эти вероятности более близкими друг к другу, увеличивает энтропию ансамбля.

2.16. Дать пример совместного ансамбля XY, где X имеет выборочным пространством (a_1, a_2), Y имеет выборочным пространством (b_1, b_2), и

$$H(X) + \sum_{k=1}^{2} P_{X|Y}(a_k \mid b_j) \log P_{X|Y}(a_k \mid b_j)$$

положительно при $j = 1$ и отрицательно при $j = 2$.

2.17. Пусть $a_1, \ldots, a_K$ является множеством несовместных событий и пусть $P(a_1), \ldots, P(a_K)$ и $Q(a_1), \ldots, Q(a_K)$ два различных набора вероятностей, приписанных этим событиям

$$\left[\sum_{k=1}^{K} P(a_k) = \sum_{k=1}^{K} Q(a_k) = 1 \right].$$

Доказать следующие утверждения:

(а) $$\sum_{k=1}^{K} P(a_k) \log \frac{P(a_k)}{Q(a_k)} \geqslant 0.$$

(б) $$\sum_{k=1}^{K} \frac{[P(a_k)]^2}{Q(a_k)} \geqslant 1.$$

Выражение в пункте (а) часто называется энтропией P относительно Q.

2.18. Рассмотрите последовательные каналы, изображенные на рис. 2.3.2, и предположите, что $I(X; Z) = I(X; Y)$, т. е. что никакая информация о входе не теряется во втором канале. Определите 2 буквы в ансамбле Z, например c_i и c_l, как эквивалентные, если $P_{X|Z}(a_k|c_i) = P_{X|Z}(a_k|c_l)$ при всех k.

(а) Показать, что $I(X; Z) = I(X; Y)$ тогда и только тогда, когда из неравенств $P_{Y|Z}(b_j|c_i) > 0$ и $P_{Y|Z}(b_j|c_l) > 0$ для какой-либо буквы b_j алфавита Y следует, что c_i и c_l эквивалентны. Другими словами, второй канал не разрушает никакой информации о X, если никакие неэквивалентные буквы алфавита Z не спутываются во втором канале.

(б) Показать, что если $I(X; Z) = I(X; Y)$, то тот же результат остается справедливым при всех выборах входного распределения $P_X(a_k)$; предполагается, что $P_X(a_k) > 0$ при всех значениях k (если второй канал понимается как приемник, то такой приемник часто называют достаточным приемником).

2.19. Пусть XYZ является дискретным совместным ансамблем. Установить, справедливы или нет следующие неравенства; в случае справедливости найти условия выполнения равенств:

(а) $I(XY; Z) \geqslant I(X; Z)$.

(б) $H(XY|Z) \geqslant H(X|Z)$.

(в) $I(X; Z|Y) \geqslant I(Z; Y|X) - I(Z; Y) + I(X; Z)$.

(г) $H(XYZ) - H(XY) \leqslant H(XZ) - H(X)$.

2.20. Пусть X, Y и Z являются ансамблями, каждый из которых содержит по два элемента так, что восемь элементов совместного ансамбля XYZ могут быть рассмотрены как вершины единичного куба.

(а) Найти совместное распределение вероятностей $P(x, y, z)$, для которого $I(X; Y) = 0$ и $I(X; Y|Z) = 1$ бит.

(б) Найти совместное распределение вероятностей $P(x, y, z)$, для которого $I(X; Y) = 1$ бит и $I(X; Y|Z) = 0$.

Характерным в этой задаче является то, что нет общего неравенства между $I(X; Y)$ и $I(X; Y|Z)$.

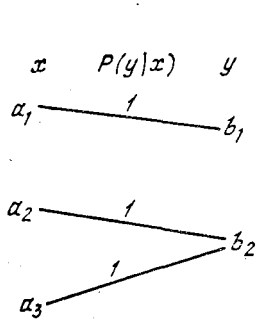
2.21. В совместном ансамбле XY взаимная информация $I(x; y)$ является случайной величиной. В этой задаче будет рассматриваться дисперсия этой случайной величины $\mathbf{D}[I(x; y)]$.

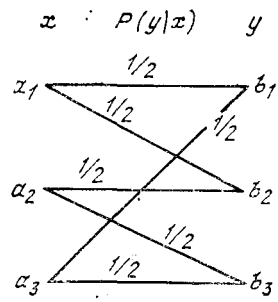
(а) Доказать, что $\mathbf{D}[I(x; y)] = 0$ тогда и только тогда, когда существует некоторая постоянная α такая, что при всех x, y, для которых $P(x, y) > 0$,

$$P(x, y) = \alpha P_X(x) P_Y(y).$$

(б) Выразить $I(X; Y)$ через α и истолковать частный случай $\alpha = 1$.

(в) Для каждого из изображенных ниже каналов найти распределение вероятностей $P_X(x)$, для которого $I(X; Y) > 0$, а $\mathbf{D}[I(x; y)] = 0$. Найти $I(X; Y)$.





2.22. По определению, гауссовская случайная величина (г. с. в.) с нулевым средним значением и единичной дисперсией имеет плотность вероятности $p_X(x) = (1/\sqrt{2\pi}) \exp[-x^2/2]$. То что эта плотность вероятности со вторым моментом,

равным единице, следует из табличных интегралов (для вывода см. Феллер (1950), т. 1, гл. VII, § 1)

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = 1, \qquad (*)$$

$$\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx = 1. \qquad (**)$$

(а) Показать, что если x г. с. в. с нулевым средним значением и единичной дисперсией, то $y = a + bx$ имеет среднее значение a, дисперсию b^2 и плотность

$$P_Y(y) = \frac{1}{\sqrt{2\pi b^2}} \exp\left\{-\frac{(y-a)^2}{2b^2}\right\}.$$

Случайная величина, имеющая эту плотность, называется г. с. в. со средним значением a и дисперсией b^2.

(б) Пусть y и z — статистически независимые г. с. в. с нулевыми средними значениями и дисперсиями b^2 и c^2 соответственно. Взяв свертку функций $p_Y(y)$ и $p_Z(z)$, показать, что $w = x + y$ является г. с. в. с нулевым средним значением и дисперсией $b^2 + c^2$.

Указание: дополните до полного квадрата показатель экспоненты и используйте (*).

(в) Показать, что характеристическая функция

$$\varphi_Y(\omega) = \overline{e^{j\omega y}} = \int_{-\infty}^{\infty} p_Y(y)\, e^{j\omega y}\, dy; \quad j = \sqrt{-1};$$

г. с. в. с нулевым средним значением и дисперсией b^2 задается равенством $\varphi_Y(\omega) = \exp[-\omega^2 b^2/2]$.

(г) При условии, что y и z статистически независимы, показать, что $w = y + z$ имеет характеристическую функцию $\varphi_W(\omega) = \varphi_Y(\omega)\,\varphi_Z(\omega)$. Используйте это для получения независимого вывода утверждения пункта (б).

2.23. Физический канал с тепловым шумом часто может быть представлен следующей моделью. Вход канала представляет собой последовательность импульсов; каждый импульс имеет фиксированную длительность T и фиксированную величину амплитуды x, $|x| = \sqrt{S}$, но произвольного знака. Приемник усредняет выход канала на интервале каждого импульса и выдает выход y, задаваемый равенством $y = x + z$, где z — усредненный на интервале шум. Предполагается, что z представляет собой г.с. в., не зависящую от x и имеющую нулевое среднее значение и дисперсию σ^2 (для белого шума со спектральной плотностью $N_0/2$ имеем $\sigma^2 = N_0/2T$),

$$p_Z(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{z^2}{2\sigma^2}\right).$$

(а) Найти вероятность того, что знак y противоположен знаку x и вычертить график зависимости этой вероятности от S/σ^2.

(б) Показать, что определенная выше вероятность стремится к $^1/_2 - \sqrt{S/(2\pi\sigma^2)}$ при $S/\sigma^2 \to 0$ и к $\sqrt{\sigma^2/(2\pi S)}\exp[-S/(2\sigma^2)]$ при $S/\sigma^2 \to \infty$.

Указание: стандартные границы для «хвостов» гауссовского распределения можно найти у Феллера (1950), т. 1, гл. VIII, § 1.

2.24. Пусть непрерывный совместный ансамбль XY имеет плотность вероятности $p_{XY}(x, y)$ и отдельные плотности $p_X(x)$ и $p_Y(y)$. Показать, что $I(X; Y) \geqslant 0$ и привести пример, в котором $H(X)$, задаваемая равенством (2.4.24), отрицательна.

2.25. Входом канала является фаза x, $0 \leqslant x \leqslant 2\pi$, а выходом канала — фаза y. Выход задается равенством $y = x + z$, где z представляет собой случайную величину, моделирующую шум, которая независима от входа x и имеет плотность вероятности $p_Z(z)$. Сумму $x + z$ можно представлять себе как остаток по модулю 2π (т. е. в обычном смысле сложения фаз). Пусть $p_X(x) = 1/2\pi$, $0 \leqslant x \leqslant 2\pi$.

(а) Выразить $I(X; Y)$ через $p_Z(z)$.

(б) Положить $p_Z(z) = 1/(b - a)$ при $a < z \leqslant b$ и найти $I(X; Y)$; предполагается, что $b - a \leqslant 2\pi$. Объясните, почему полученный вами ответ зависит только от $b - a$, а не от b и a раздельно.

2.26. Входной ансамбль X канала состоит из чисел $+1$ и -1, используемых с вероятностями $P_X(+1) = P_X(-1) = {}^1/_2$. Выход y является суммой входа x и не зависящего от него случайного шума z с *плотностью* вероятности $p_Z(z) = {}^1/_4$ при $-2 < z \leqslant 2$ и $p_Z(z) = 0$ при всех остальных z. Другими словами, условная плотность вероятности y при условии, что задана x, определяется равенствами $p_{Y|X}(y|x) = {}^1/_4$ при $-2 < y - x \leqslant 2$ и $p_{Y|X}(y|x) = 0$ в других случаях.

(а) Найти и вычертить график плотности вероятности на выходе канала.

(б) Найти $I(X; Y)$.

(в) Предположим, что выход преобразуется в процесс с дискретными значениями u, определяемыми следующим образом: $u = 1$ при $y > 1$; $u = 0$ при $-1 < y \leqslant 1$; $u = -1$ при $y \leqslant -1$. Найти $I(X; Y)$ и объяснить результат аналогично тому, как это сделано в конце § 2.3.

2.27. Пусть X, Y и Z — двоичные ансамбли; рассмотрим следующее распределение вероятностей на совместном ансамбле:

$$P_{XYZ}(0, 0, 0) = P_{XYZ}(1, 1, 1) = {}^1/_2.$$

(а) Показать, что $I(X; Y|Z) = 0$.

(б) Показать, что $\sup I(X_p; Y_p|Z_p) = 1$ бит, где верхняя грань берется по всем разбиениям каждого ансамбля.

Указание: разбиение может содержать только одно событие, т. е. совпадает со всем выборочным пространством для этого ансамбля.

(в) Показать, что для произвольного совместного ансамбля XYZ величина $I(X; Y|Z)$ не равна $\sup I(X_p; Y_p|Z_p)$, когда $I(X; Y) > I(X; Y|Z)$.

Глава 3

3.1. Источник производит последовательность статистически независимых двоичных символов с вероятностями $P(1) = 0{,}005$, $P(0) = 0{,}995$. Берутся по 100 символов одновременно и двоичное кодовое слово сопоставляется каждой последовательности из 100 символов, содержащей 3 или менее единиц.

(а) Для случая, когда все кодовые слова имеют одну и ту же длину, найти минимальную длину, требуемую для того, чтобы сопоставить множество кодовых слов указанным последовательностям.

(б) Найти вероятность появления последовательности источника, которой не соответствует кодовое слово.

(в) Используйте неравенство Чебышева для того, чтобы оценить вероятность появления последовательности, которой не соответствует кодовое слово, и сравните результат с пунктом (б).

3.2. Источник производит статистически независимые двоичные символы с вероятностями $P(0) = {}^3/_4$, $P(1) = {}^1/_4$. Рассмотрите последовательности $\mathbf{u}$ из L последовательных символов и связанное с ними неравенство

$$\Pr\left[\left|\frac{I(\mathbf{u})}{L} - H(U)\right| \geqslant \delta\right] \leqslant \varepsilon, \qquad (*)$$

где $H(U)$ — энтропия источника.

(а) Найти L_0, такое, что (*) справедливо при $L > L_0$, когда $\delta = 0{,}05$, $\varepsilon = 1/10$.

Указание: используйте неравенство Чебышева.

(б) Повторить вычисления для $\delta = 10^{-3}$, $\varepsilon = 10^{-6}$.

(в) Пусть A является множеством последовательностей $\mathbf{u}$, для которых

$$\left| \frac{I(\mathbf{u})}{L} - H(U) \right| < \delta.$$

Найти верхнюю и нижнюю границы числа последовательностей в A, когда $L = L_0$ для случаев, описанных в пунктах (а) и (б).

3.3. Источник имеет алфавит из 4 букв. Вероятности букв и два возможных множества двоичных кодовых слов для источника приведены ниже.

Буква	Вероятность	Код I	Код II
a_1	0,4	1	1
a_2	0,3	01	10
a_3	0,2	001	100
a_4	0,1	000	1000

Для *каждого* кода ответьте на следующие вопросы (не требуются доказательства или численные подтверждения).

(а) Удовлетворяет ли код свойству префикса?

(б) Является ли код однозначно декодируемым?

(в) Чему равна взаимная информация о событии, что буква источника равна a_1, содержащаяся в событии, что первый символ кодового слова равен 1?

(г) Чему равна средняя взаимная информация о букве источника, содержащаяся в первом символе кодового слова? Дайте эвристическое объяснение цели, с которой используется первая буква в кодовых словах кода II.

3.4. Код не является однозначно декодируемым тогда и только тогда, когда существует конечная последовательность, которая может быть разбита двумя различными способами на последовательности кодовых слов. То есть должна произойти ситуация, такая, как изображена ниже

A_1	A_2	$A_3 \ldots A_m$			
B_1	B_2	B_3	B_4	...	B_n

где A_i (и B_i) — кодовое слово. Заметим, что B_1 должно быть префиксом A_1, и при этом образуется некоторый «повисший суффикс». Каждый повисший суффикс в этой последовательности в свою очередь должен быть либо префиксом кодового слова, либо иметь кодовое слово в качестве префикса и давать другой повисший суффикс. Наконец, последний повисший суффикс последовательности должен сам быть кодовым словом. Таким образом, можно предложить следующий способ проверки однозначности декодирования [который на самом деле является критерием Сардинаса — Паттерсона (1953)]. Постройте множество S всех возможных повисших суффиксов. Код является однозначно декодируемым тогда и только тогда, когда S не содержит кодовых слов.

(а) Составить точные правила для построения множества S.

(б) Предположите, что длины кодовых слов равны m_i, $i = 1, 2, \ldots, M$. Найти хорошую границу сверху для числа элементов в множестве S.

(в) Определить, какой их следующих кодов однозначно декодируем:

$\{0,\ 10,\ 11\}$ $\{00,\ 01,\ 10,\ 11\}$
$\{0,\ 01,\ 11\}$ $\{110,\ 11,\ 10\}$
$\{0,\ 01,\ 10\}$ $\{110,\ 11,\ 100,\ 00,\ 10\}$
$\{0,\ 01\}$

(г) Для каждого однозначно декодируемого кода из (в) построить, если это возможно, бесконечно длинную кодовую последовательность с известным начальным символом, такую, что она может быть разложена на кодовые слова двумя различными способами. (Это иллюстрирует тот факт, что однозначная декодируемость не всегда означает конечную декодируемость.) Доказать, что такая последовательность не возникает для префиксного кода.

3.5. Рассмотрите следующий метод построения двоичных кодовых слов для ансамбля сообщений U с распределением вероятности $P(u)$. Пусть $P(a_k) \leqslant P(a_j)$ для $k > j \geqslant 1$; определим

$$Q_i = \sum_{k=1}^{i-1} P(a_k) \quad \text{при } i > 1; \quad Q_1 = 0.$$

Кодовое слово, соответствующее сообщению a_i, составляется с помощью «десятичного» разложения $Q_i < 1$ в двоичной системе (т. е. $^1/_2 \to 100\ldots$, $^1/_4 \to 0100\ldots$, $^5/_8 \to 10100\ldots$,) и усечения этого разложения после первых n_i символов, где n_i является наименьшим целым числом, большим или равным $I(a_i)$ бит.

(а) Построить двоичные кодовые слова для множества из восьми сообщений, появляющихся с вероятностями $^1/_4$, $^1/_4$, $^1/_8$, $^1/_8$, $^1/_{16}$, $^1/_{16}$, $^1/_{16}$, $^1/_{16}$.

(б) Доказать, что описанный выше метод во всех случаях приводит к множеству кодовых слов, удовлетворяющему свойству префикса, и что для средней длины $\bar{n}$ справедливо неравенство $H(U) \leqslant \bar{n} < H(U) + 1$.

3.6. Иногда требуется закодировать множество сообщений источника «алфавитным» двоичным кодом. Двоичный код называется алфавитным, если для любых i, k при $i < k$ кодовое слово для a_i (рассматриваемое как двоичное «десятичное» разложение) меньше, чем кодовое слово для a_k. Если $P(a_k) \geqslant P(a_i)$ для всех i и k при $i < k$, то можно использовать процедуру, описанную в задаче 3.5. Если это условие не выполнено, то можно поступить следующим образом. Пусть

$$Q_i = \sum_{k=1}^{i-1} P(a_k) + \frac{1}{2} P(a_i) \quad \text{при } i > 1; \quad Q_1 = 0.$$

Кодовое слово для сообщения a_i составляется усечением двоичного «десятичного» разложения Q_i после первых n_i символов, где n_i является наименьшим целым числом, большим или равным $I(a_i) + 1$. Доказать, что код, построенный согласно указанному здесь правилу, является алфавитным, удовлетворяет свойству префикса и для его средней длины справедливо неравенство $\bar{n} \leqslant H(U) + 2$.

3.7. (а) Показать, что теоремы 3.2.1 и 3.2.2 справедливы при $K = \infty$.

Указание: покажите, что из теоремы 3.2.2 следует, что однозначная декодируемость влечет за собой выполнение условия

$$\sum_{i=1}^{k} D^{-n_i} \leqslant 1$$

для всех конечных k, затем перейти к пределу.

(б) Показать, что теорема 3.3.1 справедлива для ансамбля источника с бесконечным счетным алфавитом и конечной энтропией.

3.8. Теорема кодирования для источника утверждает, что для источника с энтропией $H(U)$ можно каждой букве алфавита сопоставить двоичное кодовое слово таким образом, чтобы образовался префиксный код со средней длиной $\bar{n} < H(U) + 1$.

Показать на примере, что эта теорема не может быть улучшена. Т. е. для любого $\varepsilon > 0$ найти источник, для которого минимальная длина $\bar{n}$ удовлетворяет $\bar{n} \geqslant H(U) + 1 - \varepsilon$.

3.9. Рассмотрите два дискретных источника без памяти. Источник 1 имеет алфавит из 6 букв с вероятностями 0,3,0,2, 0,15, 0,15, 0,1, 0,1. Источник 2 имеет алфавит из 7 букв с вероятностями 0, 3, 0,25, 0,15, 0,1, 0,1, 0,05, 0,05. Построить двоичный код Хаффмана и троичный код Хаффмана для каждого множества сообщений. Найти среднее число кодовых букв на букву источника для каждого кода.

3.10. Рассмотрите двоичный код Хаффмана для источника 1 из задачи 3.9. Пусть N_L является случайной величиной, представляющей общее число кодовых букв, порождаемых последовательностью L букв источника, поступающих на кодер.

(а) Найти $\lim\limits_{L\to\infty}(N_L/L)$ и четко указать, в каком смысле этот предел существует.

(б) Предположить, что двоичный код строится с использованием множества всех последовательностей длины K из того же самого источника, что и множество сообщений, и пусть $N_{LK}(K)$ является случайной величиной, обозначающей общее число кодовых букв, порождаемых последовательностью из LK букв источника, поступающих на кодер. Найти

$$\lim_{K\to\infty}\lim_{L\to\infty}\frac{N_{LK}(K)}{LK}$$

и указать, в каком смысле этот предел существует.

3.11. Для каждого множества сообщений задачи 3.9 найти префиксный код с минимальной средней длиной при условии, что первая буква каждого кодового слова должна быть 0 или 1 и каждая последовательная кодовая буква может быть 0, 1 или 2. Найти общее правило для построения префиксных кодов с минимальной средней длиной при указанном ограничении и объяснить, почему оно работает.

3.12. Для источника 1 задачи 3.9 найти двоичный код с минимальной средней длиной, который удовлетворяет *свойству суффикса*. Свойство суффикса состоит в том, что никакое кодовое слово не совпадает с никаким суффиксом никакого другого кодового слова. Показать, что код, обладающий свойством суффикса, всегда однозначно декодируем, и показать, что минимальная средняя длина, взятая по всем кодам, обладающим свойством суффикса, для данного источника равна средней длине кода Хаффмана для того же источника.

3.13. Множество из 8 сообщений с вероятностями 0,2, 0,15, 0,15, 0,1, 0,1, 0,1, 0,1, 0,1 должно быть закодировано троичным префиксным кодом. Построить два множества кодовых слов, длины которых имеют одно и то же минимальное среднее значение, но различные дисперсии. Найти общую среднюю длину и каждую из дисперсий. Указать причины, по которым тот или другой код предпочтительнее для применения.

3.14. Дискретный источник без памяти с алфавитом объема K имеет энтропию $H(U)$ бит. Множество троичных кодовых слов, удовлетворяющее свойству префикса, найдено для букв источника, и средняя длина кодовых слов равна

$$\bar{n} = \frac{H(U)}{\log_2 3}.$$

(а) Доказать, что вероятность каждой буквы источника имеет вид 3^{-i}, где i — целое число, зависящее от буквы.

(б) Доказать, что число сообщений в алфавите источника нечетно.

3.15. Источник имеет алфавит из K букв и все буквы равновероятны. Эти буквы кодируются двоичными кодовыми словами по методу Хаффмана (т. е. так, чтобы минимизировать среднюю длину кодового слова).

Пусть j и x выбраны так, что $K = x2^j$, где j является целым числом и $1 \leqslant x < 2$.

(а) Имеют ли какие-либо кодовые слова длины, неравные j или $j + 1$? Почему?

(б) Выразить с помощью j и x число кодовых слов, имеющих длину j.

(в) Чему равна средняя длина кодового слова?

3.16. Определим рассогласование между двоичным кодом и ансамблем сообщений U как $\bar{n} - H(U)$. Показать, что рассогласование между U и оптимальным двоичным кодом для U всегда больше или равно рассогласованию между U' и оптимальным двоичным кодом для U', где U' — редуцированный ансамбль для U, рассмотренный в § 3.4.

3.17. В нижеследующей задаче изложен метод, который называется кодированием длин серий. Этот метод является не только полезным сам по себе, но он также позволяет понять более точно причину, по которой метод кодирования Хаффмана является оптимальным. Источник U производит последовательность независимых двоичных символов с вероятностями $P(0) = 0{,}9$, $P(1) = 0{,}1$. Будем кодировать эти последовательности в два этапа, сначала подсчитывая число нулей между последовательными единицами на выходе источника, и затем кодируя эти длины серий двоичными кодовыми словами. Первый этап кодирования отображает последовательности источника в промежуточные числа по следующему правилу:

Последовательность источника	Промежуточные числа (число нулей)	Последовательность источника	Промежуточные числа (число нулей)
1	0	·	·
01	1	·	·
001	2	·	·
0001	3	00000001	7
		00000000	8

Таким образом, приводимая ниже последовательность кодируется следующим образом:

```
1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1
0,    2,              8,    2,0,      4
```

Последний этап кодирования сопоставляет кодовое слово из одного двоичного символа промежуточному числу 8 и кодовые слова из четырех двоичных символов — другим промежуточным числам.

(а) Показать справедливость (настолько подробно, насколько вам будет удобно) того, что получившийся код однозначно декодируем.

(б) Найти среднее число $\bar{n}_1$ букв источника, приходящихся на промежуточное число.

(в) Найти среднее число $\bar{n}_2$ кодовых двоичных символов, приходящихся на промежуточное число.

(г) Используя закон больших чисел, показать, что для очень длинной последовательности букв источника отношение числа кодовых двоичных символов к числу букв источника с большой вероятностью будет близко к $\bar{n}_2/\bar{n}_1$. Сравнить это отношение со средним числом кодовых символов на букву источника для кода Хаффмана, кодирующего сразу четыре буквы источника.

3.18. (а) Источник имеет пять букв со следующими вероятностями: $P(a_1) = 0{,}3$; $P(a_2) = 0{,}2$; $P(a_3) = 0{,}2$; $P(a_4) = 0{,}15$; $P(a_5) = 0{,}15$. Эти буквы должны быть закодированы двоичным кодом для передачи по каналу без шумов. Передача 0 занимает 1 с, а передача 1 занимает 3 с. Используя метод проб и ошибок, найти префиксный код, который минимизирует среднее время, требуемое для передачи буквы источника, и вычислить это минимальное среднее время.

(б) Любой такой код может быть представлен деревом, в котором длина ребра пропорциональна времени, требующемуся для передачи соответствующей буквы. Показать, что у кода, минимизирующего среднее время передачи, вероят-

ности, соответствующие промежуточным и концевым узлам, не должны увеличиваться с увеличением длины.

3.19. Имеется некоторое число M пенни и известно, что $M - 1$ из них имеют один и тот же вес. M-я пенни может иметь тот же вес, что и остальные, может быть тяжелее, может быть легче. Имеются балансные весы, на которых можно сравнить вес любых двух групп пенни. Нужно найти отличающуюся пенни, если таковая имеется, и определить, тяжелее ли она или легче остальных. Найти максимальное значение M, для которого задача может быть решена с помощью n взвешиваний, и подробно описать процедуру взвешивания, если:

(а) В дополнение к M пенни имеется, для сравнения, стандартная пенни.

(б) Стандартная пенни отсутствует.

Предложение: рассмотрите $2M + 1$ возможных ответов как множество равновероятных исходов и сопоставьте каждому исходу кодовое слово, представляющее результаты последовательных взвешиваний.

3.20. Определим условную энтропию дискретного стационарного источника с алфавитом объема K:

$$H_{L|L}(U) = \frac{1}{L} H(U_{2L} \dots U_{L+1} | U_L \dots U_1).$$

(а) Доказать, что $H_{L|L}(U)$ не возрастает вместе с L.

(б) Доказать, что

$$\lim_{L \to \infty} H_{L|L}(U) = H_\infty(U).$$

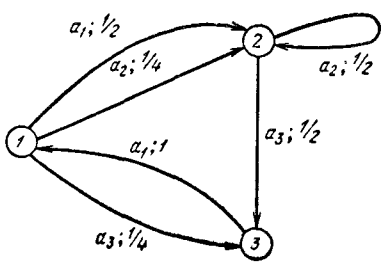
(в) Рассмотрим метод кодирования для источника, при котором для каждой из K^L возможных последовательностей из предыдущих L букв, поступивших в кодер, строится код Хаффмана для множества возможных последовательностей из L последующих букв, поступающих в кодер. Сформулировать аналог теоремы 3.5.2 для этого метода.

3.21. Рассмотрим стационарный эргодический двоичный источник, последовательность на выходе которого обозначим через $\mathbf{u} = (\dots, u_{-1}, u_0, u_1, \dots)$. Предположим, что символы, имеющие четные номера, $\dots, u_2, u_0, u_2, \dots$, суть статистически независимые равновероятные двоичные случайные величины. С вероятностью $1/2$ $u_{2n+1} = u_{2n}$ для всех n и с вероятностью $1/2$ $u_{2n-1} = u_{2n}$ для всех n. Будем считать, что пары букв источника (u_{2n}, u_{2n+1}) кодируются в соответствии с правилом

$$\begin{aligned} 00 &\to 0 \\ 11 &\to 10 \\ 01 &\to 110 \\ 10 &\to 111 \end{aligned}$$

Показать, что с вероятностью $1/2$ число кодовых букв на букву источника в длинной последовательности стремится к $3/4$ и с вероятностью $1/2$ стремится к $9/8$.

3.22. (а) Найти стационарные вероятности состояний $q(j)$ для марковского источника, указанного ниже, и найти стационарные вероятности отдельных букв.



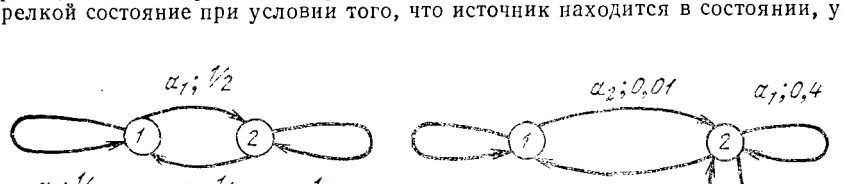
(б) Найти условную энтропию буквы на выходе источника при условии, что задано текущее состояние источника $H(U|s = j)$ при $1 \leqslant j \leqslant 3$.

(в) Найти энтропию на букву последовательности источника $H_\infty(U)$.

(г) Для любого состояния источника $s = j$ найти оптимальный неравномерный двоичный код, чтобы кодировать буквы алфавита источника, для которых $P_j(a_k) > 0$. Показать, что весь код (при выборе кодовых слов в соответствии с текущим состоянием и выходом) является однозначно декодируемым даже тогда, когда выбрано одно кодовое слово нулевой длины для состояния 3.

(д) Подсчитать среднее число кодовых букв на букву источника $\bar{n}$. Сформулировать общие условия, при которых $\bar{n} = H_\infty(U)$ для такого метода кодирования.

3.23. Для указанных ниже источников введем в рассмотрение следующий процесс Маркова, состояния которого отмечены кружками. В течение каждой единицы времени источник находится в определенном состоянии, производит букву (a_1 или a_2) и переходит в новое состояние. Число, написанное на каждой стрелке, означает вероятность произвести указанную букву и перейти в указанное стрелкой состояние при условии того, что источник находится в состоянии, указанном на хвосте стрелки. Заметим, что такие источники не являются марковскими, так как новое состояние не определяется текущим состоянием и выходной буквой. Источники такого типа часто называются источниками, порождаемыми марковскими источниками.



(а) Показать, что источник *1*, изображенный на рисунке, является тонко замаскированным двоичным источником с независимыми равновероятными буквами и имеет энтропию 1 бит на букву (неосторожное применение (3.6.21) могло бы дать энтропию, равную нулю).

(б) Показать, что энтропия второго источника (и любого такого источника) ограничена неравенствами

$$H(U_l|U_{l-1}\ldots U_1 S_1) \leqslant H_\infty(U) \leqslant H(U_l|U_{l-1}\ldots U_1).$$

Показать, что нижняя граница не убывает с ростом l, а верхняя граница не возрастает с ростом l. (Для того чтобы найти $H_\infty(U)$ для частного источника, приведенного здесь, см. Гилберт (1960), где развит метод разложения в ряды.)

Глава 4

4.1. Пропускные способности каналов из множества N дискретных каналов без памяти равны $C_1, C_2, \ldots, C_N$. Пусть эти каналы соединены параллельно в том смысле, что в каждую единицу времени по каждому каналу передается и принимается произвольный символ. Следовательно, вход $\mathbf{x} = (x_1, \ldots, x_N)$ параллельного соединения каналов представляет собой последовательность N компонент, каждая из которых является входом одного из каналов, а выход $\mathbf{y} = (y_1, \ldots, y_N)$ представляет собой последовательность N компонент, каждая из которых является выходом одного из каналов. Доказать, что пропускная способность параллельного соединения каналов равна $\sum_{n=1}^{N} C_n$.

Предполагается, что выход каждого канала статистически зависит *лишь* от входа этого канала, т. е.

$$P(\mathbf{y} \mid \mathbf{x}) = \prod_n P(y_n \mid x_n).$$

Не делайте предположения, что входы независимы.

Указание: см. доказательство теоремы 4.2.1.

4.2. Выход канала проходит через устройство обработки информации без потери информации, т. е. $I(X; Y) = I(X; Z)$, где X — ансамбль на входе канала; Y — ансамбль на выходе канала, а Z — выход устройства обработки информации. Привести пример, в котором $H(Y) > H(Z)$, и привести пример, в котором $H(Y) < H(Z)$. Устройство обработки информации не обязательно должно быть неслучайным.

4.3. Другое доказательство теоремы 4.3.1 проводится следующим образом. Можно интерпретировать совместный ансамбль UV как ансамбль UVE, где выборочными точками E являются события e и c, причем e соответствует ошибке ($u \neq v$), а c соответствует правильному приему ($u = v$). Тогда

$$H(U \mid V) = H(UE \mid V) = H(E \mid V) + H(U \mid VE) = H(E|V) + P_e H(U \mid Ve) +$$
$$+ (1 - P_e) H(U \mid Vc) \leqslant H(E) + P_e H(U \mid Ve) \leqslant H(E) + P_e \log(M - 1).$$

Обосновать те из указанных выше соотношений, которые не являются очевидными.

4.4. Источник производит статистически независимые равновероятные двоичные символы со скоростью один символ в секунду. Выход источника кодируется и передается по двоичному симметричному каналу с вероятностью ошибки ε, $0 < \varepsilon < 1/2$. Каждую секунду передается один символ канала. Используя обращение теоремы кодирования, найти границу сверху и снизу для $\langle P_e \rangle$. Объяснить, почему невозможно сделать вероятность ошибки больше, чем полученная вами верхняя граница. Сравнить ваши границы с $\langle P_e \rangle$ для случая, когда кодирование и декодирование не производится.

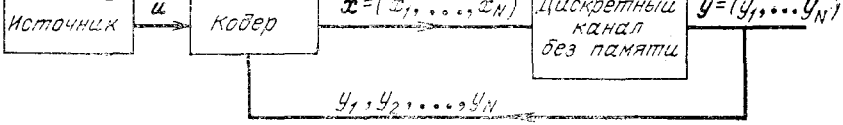
4.5. Дискретный источник без памяти производит один символ каждую секунду и имеет энтропию 2 бита. Выход источника кодируется и передается по каналу с пропускной способностью C_t бит в секунду.

(а) Предполагая, что алфавит источника имеет объем $M = 4$, тщательно вычертить зависимость от C_t нижней границы вероятности ошибки на символ источника, которая дана в теореме 4.3.4.

(б) Для любого заданного C_t найти канал, для которого указанная нижняя граница точно достигается без какого-либо кодирования.

(в) Предположите теперь, что объем алфавита источника неизвестен, но что $C_t = 1$ и $\langle P_e \rangle = 10^{-6}$. Найти нижнюю границу для M, используя теорему 4.3.4.

4.6. В этой задаче показывается, что пропускная способность дискретного канала без памяти не увеличивается при введении обратной связи от приемника к передатчику. Рассмотрите схему, изображенную ниже.



Каждый символ на выходе канала после его получения приемником посылается назад к кодеру и может оказывать влияние на выбор последующих входов канала. Докажем, что

$$I(\mathbf{U}; \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n; Y_n) \leqslant NC.$$

Это можно показать с помощью следующих соотношений (ваша задача — показать, что каждое из этих соотношений справедливо):

$$I(\mathbf{U}; \mathbf{Y}^N) = \sum_{n=1}^{N} I(\mathbf{U}; Y_n | Y_{n-1} \dots Y_1),$$

$$I(\mathbf{U}; Y_n | Y_1 \dots Y_{n-1}) \leqslant I(\mathbf{U}X_n; Y_n | Y_1 \dots Y_{n-1}) =$$
$$= I(X_n; Y_n | Y_1 \dots Y_{n-1}) \leqslant I(X_n; Y_n).$$

С помощью этого результата показать, что обращение теоремы кодирования остается в силе для дискретного канала без памяти с обратной связью.

Замечание: этот результат несправедлив для каналов с памятью.

4.7. В теореме 4.3.1 была найдена нижняя граница для P_e, выраженная через $H(U|V)$. Предположим здесь, что декодер проводит декодирование с минимальной вероятностью ошибки, и найдем верхнюю границу для P_e. Будем считать, что U и V имеют одно и то же выборочное пространство, скажем $a_1, \dots, a_K$; при декодировании с минимальной вероятностью ошибки имеет место следующее соотношение: $P_{U|V}(a_i|a_i) \geqslant P_{U|V}(a_k|a_i)$ при всех $k \neq i$. Используя это неравенство, показать, что

$$P_e \leqslant H(U|V) \text{ натуральных единиц.} \qquad (*)$$

Указание: пусть W является произвольным ансамблем с вероятностями q_i. Покажите, что (в натуральных единицах)

$$H(W) \geqslant \sum_i q_i (1 - q_i) \geqslant 1 - q_{max}. \qquad (**)$$

Пусть

$$H(U | v) = -\sum_u P(u | v) \ln P(u | v)$$

и пусть $P_e(v)$ является вероятностью ошибки при условии, что декодировано сообщение v. Используйте (**) для того, чтобы показать, что $H(U|v) \geqslant P_e(v)$, и используйте этот результат, чтобы установить справедливость (*).

Усложнение задачи. Показать, что (**) можно заменить на более сильное неравенство для $H(W)$ в битах: $H(W) \geqslant 2(1 - q_{max})$ и, таким образом, $P_e \leqslant {}^1/_2 H(U|V)$ бит.

4.8. Используя неравенство $\log z \leqslant (z - 1) \log e$, показать справедливость (4.3.16).

4.9. Доказать, что неравенства (4.4.4) и (4.4.5) справедливы.

Указание [для (4.4.4)]: пусть α, β — две точки на интервале и пусть $\delta = \lambda\alpha + (1 - \lambda)\beta$ является точкой, лежащей между α и β. Используйте затем разложение в ряд Тейлора при любом x

$$f(x) = f(\delta) + (x - \delta) f'(\delta) + \frac{(x - \delta)^2}{2} f''(y)$$

при некотором y между x и δ. Изобразите также это на рисунке.

Указание [для (4.4.5)]: используйте индукцию по L.

4.10. Пусть $Q_1(x)$ и $Q_2(x)$ — некоторые распределения вероятностей на одном и том же дискретном выборочном пространстве и пусть $H_1(X)$ и $H_2(X)$ — соответствующие им энтропии.

(а) Показать что $Q(x) = \lambda Q_1(x) + (1 - \lambda) Q_2(x)$ при $0 \leqslant \lambda \leqslant 1$ является распределением вероятностей на выборочном пространстве.

(б) Пусть $H(X)$ является энтропией, соответствующей распределению вероятностей $Q(x)$ из пункта (а). Доказать, что

$$H(X) \geqslant \lambda H_1(X) + (1 - \lambda) H_2(X)$$

и найти условия равенства.

(в) Дать геометрическую интерпретацию результата пункта (б).

4.11. Функция $f(\alpha)$ определена в выпуклой области R векторного пространства. Доказать, что f является выпуклой $\cap$ тогда и только тогда, когда функ-

ция $f(\lambda\alpha_1 + (1-\lambda)\alpha_2)$ является выпуклой функцией λ, $0 \leqslant \lambda \leqslant 1$, при всех α_1, α_2 из R.

4.12. Пусть $f_i(\alpha)$, $i \in I$ (I — некоторое произвольное множество индексов) является множеством функций, каждая из которых выпукла $\cup$ и убывает с ростом действительного аргумента α. Предположить, что $f(\alpha) = \sup_{i \in I} f_i(\alpha)$ везде конечна и доказать, что $f(\alpha)$ выпукла $\cup$ и убывает по α

4.13. Рассмотрите функцию двух переменных $f(\alpha) = \alpha_1(2-\alpha_1) - (\alpha_2+1)^2$ в области $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$.

(а) Показать, что $f(\alpha)$ выпукла $\cap$ в этой области.

(б) Найти максимум $f(\alpha)$ в этой области и объяснить, почему он действительно является максимумом.

4.14. Канал с аддитивным гауссовым шумом часто можно моделировать с помощью множества параллельных каналов с дискретным временем; известно, что это множество параллельных каналов имеет пропускную способность

$$C(S_1, \ldots, S_L) = \sum_{l=1}^{L} {}^1/_2 \log\left(1 + \frac{S_l}{N_l}\right),$$

где N_l — мощность шума в l-м канале и S_l — мощность сигнала в l-м канале. Предположим, что общая мощность сигнала равна S и она может быть разделена между каналами любым образом. Пусть α_l — доля мощности, которая выделена для l-го канала, так что $S_l = \alpha_l S$.

(а) Найти необходимые и достаточные условия, накладываемые на эти доли $\alpha_1, \ldots, \alpha_L$, при которых максимизируется $C(S_1, \ldots, S_L)$ при соблюдении ограничений

$$\alpha_l \geqslant 0, \quad \sum_l \alpha_l = 1.$$

(б) Найти максимизирующие значения S_1, S_2 и S_3 в случае $L = 3$, $S = 2$, $N_1 = 1$, $N_2 = 2$, $N_3 = 3$.

Указание: используйте теорему 4.4.1.

4.15. *Полезные для теории информации неравенства.* Пусть в последующих неравенствах a_i, b_i, P_i, Q_i являются неотрицательными числами, определенными для конечного множества значений i, $1 \leqslant i \leqslant A$. Пусть

$$\sum_i P_i = \sum_i Q_i = 1.$$

Пусть s и r — положительные числа, а λ принадлежит интервалу $0 < \lambda < 1$. Доказать справедливость следующих неравенств (различные доказательства каждого из них можно найти у Харди, Литтльвуда и Полиа (1934)).

(а)
$$\sum_i Q_i^{\lambda} P_i^{1-\lambda} \leqslant 1$$

с равенством тогда и только тогда, когда $P_i = Q_i$ при всех i.

Указание: покажите, что левая часть неравенства является выпуклой функцией $\cup$ по λ, и рассмотреть случаи, когда $\lambda \to 0$, $\lambda \to 1$.

(б) Неравенство Гёльдера:

$$\sum_i a_i b_i \leqslant \left(\sum_i a_i^{1/\lambda}\right)^{\lambda} \left[\sum_i b_i^{1/(1-\lambda)}\right]^{1-\lambda}$$

с равенством тогда и только тогда, когда при некотором c, $a_i^{1-\lambda} = b_i^{\lambda} c$ при всех i.

Указание: полагая

$$Q_i = a_i^{1/\lambda} \Big/ \left(\sum_i a_i^{1/\lambda}\right)$$

и

$$P_i = b_i^{1/(1-\lambda)} \Big/ \left[\sum_i b_i^{1/(1-\lambda)}\right],$$

используйте пункт (а). В частном случае $\lambda = 1/2$, неравенство называется неравенством Коши; интегральный аналог называется неравенством Шварца.

(в) Вариант неравенства Гёльдера:

$$\sum_i P_i a_i b_i \leqslant \left(\sum_i P_i a_i^{1/\lambda}\right)^{\lambda}\left[\sum_i P_i b_i^{1/(1-\lambda)}\right]^{1-\lambda}$$

с равенством тогда и только тогда, когда при некотором c, $P_i a_i^{1/\lambda} = P_i b_i^{1/(1-\lambda)}$ c при всех i.

Указание: используйте $P_i^{\lambda} a_i$ вместо a_i и $P_i^{1-\lambda} b_i$ вместо b_i в пункте (б).

(г) $(\Sigma P_i a_i)^r \leqslant \sum_i P_i a_i^r$, $r > 1$,

$$(\Sigma P_i a_i)^r \geqslant \sum_i P_i a_i^r, \quad r < 1$$

с равенством тогда и только тогда, когда значения a_i, соответствующие $P_i > 0$, равны.

Указание: при $r > 1$ используйте пункт (в) с $b_i = 1$; при $r < 1$ используйте a_i^r вместо a_i в пункте (в).

(д) $\left(\sum_i P_i a_i^r\right)^{1/r} \leqslant \left(\sum_i P_i a_i^s\right)^{1/s}$, $0 < r < s$

с равенством тогда и только тогда, когда значения a_i, соответствующие $P_i > 0$, равны.

Указание: используйте пункт (в) при $b_i = 1$, $\lambda = r/s$.

(е) $\left(\sum_i a_i\right)^r \leqslant \sum_i a_i^r$, $r \leqslant 1$,

$$\left(\sum_i a_i\right)^r \geqslant \sum_i a_i^r, \quad r \geqslant 1,$$

с равенством тогда и только тогда, когда $r = 1$ или если только одно a_i не равно нулю. Отметить отличие пунктов (г) и (е).

Указание: положите $P_i = a_i/\Sigma a_i$ и рассмотрите отношение правых слагаемых к левой части.

(ж) Неравенство Минковского. Пусть a_{jk} — множество неотрицательных чисел, $1 \leqslant j \leqslant J$, $1 \leqslant k \leqslant K$. Тогда

$$\left[\sum_j \left(\sum_k a_{jk}\right)^{1/r}\right]^r \leqslant \sum_k \left(\sum_j a_{jk}^{1/r}\right)^r, \quad r < 1,$$

$$\left[\sum_j \left(\sum_k a_{jk}\right)^{1/r}\right]^r \geqslant \sum_k \left(\sum_j a_{jk}^{1/r}\right)^r, \quad r > 1$$

с равенством тогда и только тогда, когда при некоторых $\{\alpha_j\}$, $\{\beta_k\}$ имеем $\alpha_{jk} = \alpha_j \beta_k$ для всех j, k.

Указание: при $r < 1$ используйте

$$\left(\sum_k a_{jk}\right)^{1/r} = \left(\sum_k a_{jk}\right)\left(\sum_i a_{ji}\right)^{(1-r)/r},$$

чтобы получить $\sum_j \left(\sum_k a_{jk}\right)^{1/r} = \sum_k \left[\sum_j a_{jk}\left(\sum_i a_{ji}\right)^{(1-r)/r}\right]$.

Использование неравенства Гёльдера для выражения в квадратных скобках при каждом k дает

$$\sum_j \left(\sum_k a_{jk}\right)^{1/r} \leqslant \sum_k \left(\sum_j a_{jk}^{1/r}\right)^r \left[\sum_j \left(\sum_i a_{ji}\right)^{1/r}\right]^{1-r}.$$

Разделив обе части этого неравенства на выражение, стоящее в квадратных скобках, получаем искомый результат. При $r > 1$ следует положить $r' = 1/r$ и использовать пункт (ж) при $r' < 1$, подставляя $a_{jk}^{1/r}$ вместо a_{jk}.

(з) Вариант неравенства Минковского

$$\left[\sum_j Q_j \left(\sum_k a_{jk}\right)^{1/r}\right]^r \leqslant \sum_k \left(\sum_j Q_j a_{jk}^{1/r}\right)^r, \quad r < 1$$

с обратным неравенством при $r > 1$.

Указание: используйте $Q_j^r a_{jk}$ вместо a_{jk} в пункте (ж). Заметьте, что если $a_1, \ldots, a_K$ является множеством случайных величин, принимающих значения $a_{j1}, a_{j2}, \ldots, a_{jK}$ на j-м элементе выборочного пространства, с вероятностью Q_j, то результат пункта (з) можно выразить следующим образом:

$$\left[\overline{\left(\sum_k a_k\right)^{1/r}}\right]^r \leqslant \sum_k \left[\overline{a_k^{1/r}}\right]^r, \quad r < 1,$$

с обратным неравенством при $r > 1$.

4.16. (а) Пусть X и Y являются входным и выходным ансамблями соответственно дискретного канала без памяти (ДКБП). Показать, что $H(Y)$ является выпуклой $\cap$ функцией входного вектора вероятностей.

Указание: рассмотрите выходные векторы вероятностей, получающихся из входных векторов вероятностей.

(б) Привести пример ДКБП, для которого выпуклость, указанная в пункте (а), не является строгой.

(в) Вновь рассмотреть ДКБП и показать, что — $H(Y|X)$ является линейной функцией входного вектора вероятностей.

(г) Объединяя пункты (а) и (в), показать, что $I(X; Y)$ является выпуклой $\cap$ функцией входного вектора вероятностей.

4.17. Пусть для произвольного распределения вероятностей $\mathbf{Q}_0$ в ДКБП

$$I_0(x=k;\ Y) = \sum_j P(j|k) \log \frac{P(j|k)}{\sum_i Q_0(i) P(j|i)}.$$

Показать, что

$$\sum_k Q_0(k) I_0(x=k;\ Y) \leqslant C \leqslant \max_k I_0(x=k;\ Y). \qquad (*)$$

Используйте это для того, чтобы показать, что

$$C = \min_{\mathbf{Q}_0} \max_k I_0(x=k;\ Y), \qquad (**)$$

Указание: пусть $I_1(X; Y)$ соответствует некоторому $\mathbf{Q}_1 \neq \mathbf{Q}_0$; взяв частную производную $I_0(X; Y)$ по $Q_0(k)$, покажите, что

$$I_1(X;\ Y) \leqslant \sum_k Q_1(k) I_0(x=k;\ Y).$$

Заметьте, что при отыскании C с помощью численных методов на вычислительной машине, (*) дает границы того, насколько близко аппроксимируется C при каком-либо заданном $\mathbf{Q}_0$ и дает критерий для прекращения вычислений, когда C аппроксимируется достаточно хорошо.

4.18. (а). Рассмотрим n (вообще говоря, различных) ДКБП с пропускными способностями $C_1, C_2, \ldots, C_n$. Соответствующая им «сумма» каналов определяется как канал, входным и выходным алфавитами которого являются объединения алфавитов исходных каналов, т. е. в сумме каналов все n каналов можно использовать, но только так, что в любой заданный момент времени можно передачу

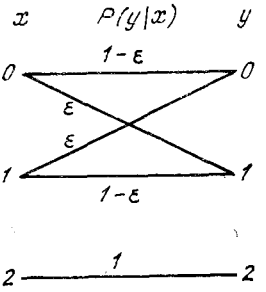
вести только по одному каналу. Показать, что пропускная способность суммы каналов имеет вид

$$C = \log_2 \sum_{i=1}^{n} 2^{C_i}$$

и найти q_i — вероятность использования i-го канала. Истолковать C как среднее значение пропускных способностей отдельных каналов, сложенное с информацией, содержащейся в выборе канала.

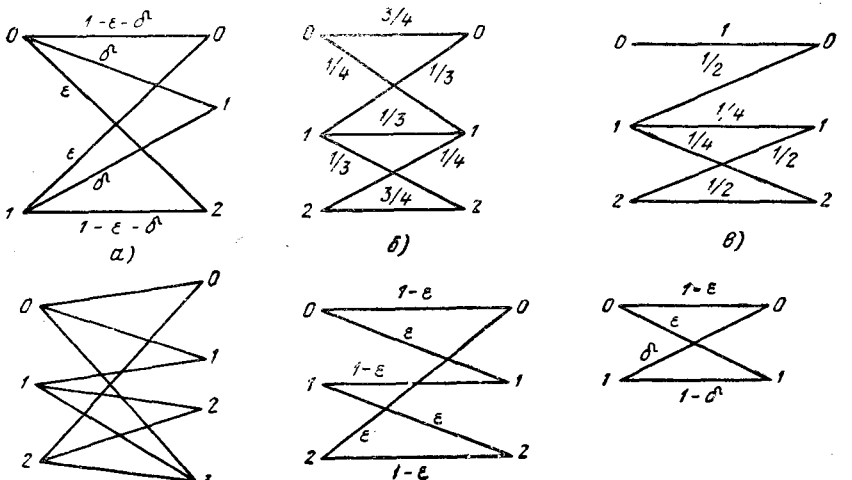
Указание: запишите входное распределение вероятностей в виде $q_i\, Q_i(k)$, где $Q_i(k)$ является распределением вероятностей на входе i-го канала при условии, что используется i-й канал. После этого применить теорему 4.5.1.

(б) Используйте полученный выше результат для того, чтобы найти пропускную способность канала, изображенного ниже.



4.19. ДКБП называется аддитивным по модулю K, если его входным и выходным алфавитами является множество 0,1, ..., $K-1$ и выход y связан с входом x и шумом z равенством $y = x \oplus z$. Шум принимает значения 0, 1, ..., $K-1$ и статистически не зависит от входа, а сложение $x \oplus z$ производится по модулю K (т. е. принимается $x + z$ или $x + z - K$ в зависимости от того, какое из значений лежит между 0 и $K-1$).

(а) Показать, что $I(X; Y) = H(Y) - H(Z)$.



К задаче 4.20

(б) Выразить пропускную способность через $H(Z)$ и найти входное распределение вероятностей, дающее максимум.

4.20. Найти пропускную способность и оптимальное входное распределение вероятностей для каждого из изображенных выше ДКБП.

Указание: только канал е) требует выполнения соответствующих алгебраических преобразований.

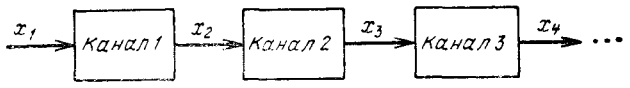
4.21. Предположим, что ДКБП обладает тем свойством, что для некоторого входного распределения вероятностей $Q(k) > 0$, $0 \leqslant k \leqslant K - 1$, дисперсия $I(x; y)$ равна 0. Доказать, что на этом входном распределении достигается пропускная способность канала (см. задачу 2.21).

4.22. Рассмотрим опять канал, описанный в задаче 2.23. Найти выражение для пропускной способности этого канала, рассматривая вначале в качестве выхода канала принятую величину y, а затем рассматривая в качестве выхода канала знак y. Показать, что первая пропускная способность больше второй. Показать, что в пределе при $S \to 0$ первая пропускная способность стремится к $S/(2\sigma^2)$, а вторая стремится к $S/(\pi\sigma^2)$.

4.23. Проверить справедливость значений для $\underline{C}$ и $\overline{C}$, приведенных для каналов, представленных на рис. 4.6.3 — 4.6.5.

4.24. Привести пример двоичного канала, для которого $\underline{C} = \overline{C} = (\overline{C}_N + 1)/N$ для всех N, а также канала с двумя состояниями, в котором имеется только память, связанная с межсимвольной интерференцией и для которого $(\underline{C} - 1)/N = \underline{C} = \overline{C}$ для всех N.

4.25. Рассмотрим полубесконечное последовательное соединение дискретных каналов (см. рисунок, помещенный ниже), в котором при любом положительном n величина x_n является входом n-го канала и выходом $(n-1)$-го канала. Предположим, что шумы в каналах статистически независимы.



(а) Показать, что последовательность $x_1, x_2, \ldots$ представляет собой неоднородную цепь Маркова.

(б) Предположим, что каждый канал имеет входной и выходной алфавиты объема K и что имеется $\delta > 0$, не зависящее от n, такое, что n-й канал в соединении имеет по крайней мере одну выходную букву, x_{n+1}, для которой $P(x_{n+1}|x_n) \geqslant \delta$ при любых входах x_n. Показать, что $I(X_1; X_n)$ стремится к нулю с ростом n и эта информация ограничена сверху выражением $2K(1-\delta)^n \log e$.

Указание: примените лемму 4.6.2.

4.26. Предположим, что канал с конечным числом состояний, в котором имеется только память, связанная с межсимвольной интерференцией (т. е s_n является функцией s_{n-1} и x_n), переходит в некоторое известное состояние. Показать, что он может перейти в это состояние не более чем после $2^A - A - 1$ входов, где A — число состояний.

Указание: пусть для последовательности входов, которая приводит канал в некоторое известное состояние, B_n обозначает множество возможных состояний канала после n-го входа.

Покажите, что если последовательность B_n имеет какие-либо повторения до достижения известного состояния, то последовательность входов может быть укорочена.

Глава 5

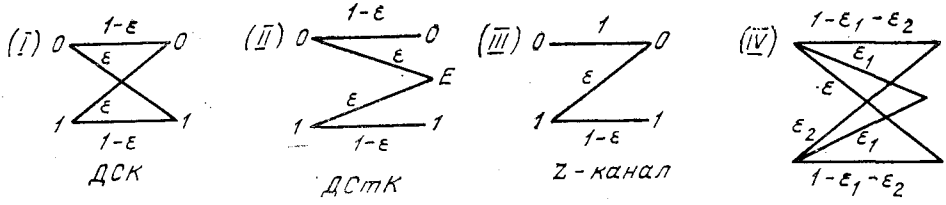
5.1. Будем считать, что целые числа от 1 до M кодируются в последовательности входов канала $\mathbf{x}_1, \ldots, \mathbf{x}_M$. Пусть $\mathbf{y}$ будет последовательностью на выходе канала и пусть заданы $P_N(\mathbf{y}|\mathbf{x}_m)$, $1 \leqslant m \leqslant M$. В случае, когда стоимость деко-

дирования переданного сообщения m в m' задается функцией $C(m, m')$, а распределение вероятностей на входных целых числах от 1 до M задается функцией $Q(m)$, найти правило декодирования, которое приводит к минимальной средней стоимости.

5.2. (а) Вычислить

$$g_n(s) = \sum_{y_n} P(y_n \mid 0)^{1-s} P(y_n \mid 1)^s$$

для каждого из следующих каналов и для каждого из них минимизировать $g_n(s)$ на $0 \leqslant s \leqslant 1$. Используйте этот результат для получения верхней границы вероятности ошибочного декодирования по максимуму правдоподобия $P_{e,m}(m = 1,2)$, достижимой при использовании кода с двумя кодовыми словами $\mathbf{x}_1$ (последовательность из N нулей) и $\mathbf{x}_2$ (последовательность из N единиц).



(б) Для первых трех указанных выше каналов найти точные выражения для вероятности ошибки. Найдите численные значения этих выражений с точностью до двух значащих цифр (используя в случае необходимости формулу Стирлинга $N! \approx \sqrt{2\pi N}(N/e)^N$ при $N = 25$ и $\varepsilon = 0{,}1$ и сравните их с границей из пункта (а).

(в) Показать, что в ДСК для больших четных N вероятности $P_{e,1}$ и $P_{e,2}$ [см. (5.3.14) и (5.3.15)] приближенно задаются равенствами

$$P_{e,1} \approx \sqrt{\frac{2}{\pi N}}\left(\frac{1-\varepsilon}{1-2\varepsilon}\right)[2\sqrt{\varepsilon(1-\varepsilon)}]^N; \quad P_{e,2} \approx P_{e,1}\left(\frac{\varepsilon}{1-\varepsilon}\right).$$

Показать, что для больших нечетных N имеем

$$P_{e,1} = P_{e,2} \approx \sqrt{\frac{2\varepsilon}{\pi N(1-\varepsilon)}}\left(\frac{1-\varepsilon}{1-2\varepsilon}\right)[2\sqrt{\varepsilon(1-\varepsilon)}]^N.$$

Указание: используйте то, что в окрестности $i = N/2$ значение $\binom{N}{i}$ изменяется медленно по сравнению с $\varepsilon^i(1-\varepsilon)^{N-i}$.

(г) Для Z-канала найдите верхнюю границу и оцените $P_{e,m}(m = 1,2)$ для кода с двумя кодовыми словами, в котором $\mathbf{x}_1$ образовано из $N/2$ нулей, за которыми следуют $N/2$ единиц, а x_2 образовано из $N/2$ единиц, за которыми следуют $N/2$ нулей. Заметьте, что для остальных каналов это изменение кода не дает различия.

5.3. (а) Найти границу (5.5.10) для средней вероятности ошибки по ансамблю кодов с двумя кодовыми словами для каналов из задачи 5.2. Найти минимумы границ по $Q(k)$.

(б) Для того же самого ансамбля и тех же самых каналов найти верхнюю границу для $1/N \ln P_{e,m}$, т. е. для среднего показателя экспоненты вероятности ошибки, используя $Q(0) = Q(1) = 1/2$, и истолковать полученный результат как границу вероятности ошибки для типичного кода из ансамбля.

5.4. Заданные M сообщений кодируются в последовательности двоичных символов длины N; M последовательностей выбираются независимо из 2^N возможных последовательностей с равномерным распределением вероятностей. Так как любая ошибка, которая происходит при декодировании принятой последо-

вательности по максимуму правдоподобия, происходит потому, что одно или больше из $M - 1$ неправильных сообщений являются более правдоподобными, чем правильное сообщение, то используйте границу средней вероятности ошибки, полученную вами для случая кода с двумя случайно выбранными словами, и тот факт, что вероятность наступления одного или большего числа событий меньше или равна сумме их вероятностей, для того чтобы получить границу вероятности ошибки при передаче M сообщений в каждом из четырех каналов задачи 5.2. Найти максимальную скорость передачи $R = (\ln M)/N$, при которой ваша граница будет экспоненциально убывающей с ростом N.

5.5. Пусть

$$z = \sum_{n=1}^{N} x_n,$$

где x_n — статистически независимые одинаково распределенные случайные величины со средним значением, равным нулю, и дисперсией, равной единице.

(а) Считая, что x_n являются гауссовскими случайными величинами, оценить $\Pr(z \geqslant N)$. (Используйте оценку

$$\int_{y=W}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}\, dy \approx \frac{\sigma}{\sqrt{2\pi}W} e^{-W^2/2\sigma^2},$$

справедливую при больших W.) Сравните полученный результат с границей Чернова и обычным неравенством Чебышева.

(б) Пусть $x_n = 1$ с вероятностью $1/2$ и $x_n = -1$ с вероятностью $1/2$. Показать, что при $N = 100$ имеем $\Pr(z \geqslant N) = 2^{-100}$. Сравните это с границей Чернова для $\Pr(z \geqslant N)$, неравенством Чебышева и оценкой, получаемой на основе центральной предельной теоремы. (Это дает хороший пример для того, чтобы понять, почему слово «*центральной*» стоит в названии центральной предельной теоремы.)

5.6. Доказать, что в (5.6.10) минимум по $s > 0$ достигается при $s = 1/(1 + \rho)$.

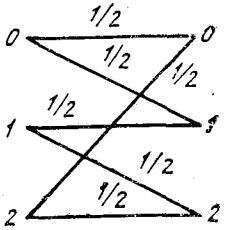
Указание: используйте неравенство Гёльдера (см. задачу 4.15) для того, чтобы показать, что

$$\left[\sum_{\mathbf{x}} Q_N(\mathbf{x})\, P_N(\mathbf{y} \mid \mathbf{x})^{1/1+\rho}\right]^{1+\rho} \leqslant$$

$$\leqslant \left[\sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m)\, P_N(\mathbf{y} \mid \mathbf{x}_m)^{1-s\rho}\right] \left[\sum_{\mathbf{x}} Q_N(\mathbf{x})\, P_N(\mathbf{y} \mid \mathbf{x})^{s}\right]^{\rho}.$$

5.7. Предположим, что два кодовых слова длины N для канала, изображенного ниже, выбираются так, что каждая буква каждого кодового слова выбирается независимо с распределением вероятности $Q(0) = Q(1) = Q(2) = 1/3$.

(а) Найти среднюю по ансамблю вероятность ошибки в предположении, что используется декодирование по максимуму правдоподобия. В случае неопределенности предположите, что декодер делает ошибку. Не применяйте здесь теорему кодирования; начните с начала и используйте специфику этого канала.



Указание: пусть посылается $\mathbf{x}_1$ и принимается $\mathbf{y}$. Подсчитать, сколько последовательностей на входе канала $\mathbf{x}$ могут привести к $\mathbf{y}$ и найти $P_N(\mathbf{y}|\mathbf{x}_1)$ для каждой из них. Затем найти вероятность того, что кодовое слово $\mathbf{x}_2$ было выбрано в этом множестве.

(б) Используйте ваш результат, полученный в пункте (а), для того чтобы ограничить сверху вероятность ошибки для кода с M словами, выбираемыми независимо из того же самого ансамбля. Используйте аддитивную границу для множества неправильных сообщений. Сравните ваш ответ с тем, что утверждает теорема кодирования.

(в) Предположим, что код состоит из двух кодовых слов $\mathbf{x}_1 = (0, 0, 0, \ldots, 0)$ и $\mathbf{x}_2 = (1, 1, 1, \ldots, 1)$. Найти вероятность ошибки для этого кода при той же самой процедуре декодирования, что и в пункте (а). Объясните отличие в ваших ответах в пункте (а) и в пункте (в).

5.8. При рассмотрении двоичных каналов часто полезно иметь точные границы для биномиальных коэффициентов $\binom{N}{j}$.

(а) Показать, что при $j \geqslant 1$, $N - j \geqslant 1$

$$\sqrt{\frac{N}{8j(N-j)}} \leqslant \binom{N}{j} e^{-N\mathcal{H}(j/N)} < \sqrt{\frac{N}{2\pi j(N-j)}},$$

где

$$\mathcal{H}(j/N) = -\frac{j}{N}\ln\frac{j}{N} - \left(1-\frac{j}{N}\right)\ln\left(1-\frac{j}{N}\right).$$

Указание: используйте формулу Стирлинга (см. Феллер (1950), т. 1, гл. II, § 9)

$$N! = \sqrt{2\pi N}\left(\frac{N}{e}\right)^N \exp(\varepsilon_N),$$

где ε_N убывает с ростом N и находится в пределах $0 < \varepsilon_N < \frac{1}{12N}$. Нижняя граница должна быть проверена с помощью прямых вычислений, когда наибольшее из чисел j и $N - j$ равно или меньше двух. Заметьте, что в пределе при больших j и $N - j$ верхняя граница достигается.

(б) Используя пункт (а), доказать более слабые границы

$$\sqrt{\frac{1}{2N}} \leqslant \binom{N}{j} e^{-N\mathcal{H}(j/N)} < 1,$$

$$\binom{2N-1}{N} \geqslant \frac{1}{\sqrt{4N}} 2^{2N-1}.$$

Указание: для последнего неравенства покажите, что $\binom{2N}{N} = 2\binom{2N-1}{N}$.

(в) Пусть ε произвольно, $0 < \varepsilon < 1$, $j < N$ и $j/N > \varepsilon$. Показать, что

$$\binom{N}{j}\varepsilon^j(1-\varepsilon)^{N-j} \leqslant \sum_{n=j}^{N}\binom{N}{n}\varepsilon^n(1-\varepsilon)^{N-n} \leqslant$$

$$\leqslant \frac{j(1-\varepsilon)}{j(1-\varepsilon)-(N-j)\varepsilon}\binom{N}{j}\varepsilon^j(1-\varepsilon)^{N-j}.$$

Указание: для верхней границы покажите, что

$$\binom{N}{n+1} < \binom{N}{n}\left(\frac{N-n}{n}\right),$$

и использовать это для того, чтобы установить неравенство

$$\binom{N}{j+m} < \binom{N}{j}\left(\frac{N-j}{j}\right)^m.$$

Затем просуммировать по j как геометрическую прогрессию. Используйте этот результат совместно с результатом пункта (а) для того, чтобы получить верхнюю и нижнюю границы для

$$\sum_{n=j}^{N} \binom{N}{n} \varepsilon^n (1-\varepsilon)^{N-n}.$$

(г) Пусть

$$w = \sum_{n=1}^{N} x_n,$$

где x_n — статистически независимые случайные величины, принимающие значение 1 с вероятностью ε и значение 0 с вероятностью $1-\varepsilon$. Показать, что

$$\Pr[w \geqslant j] = \sum_{n=j}^{N} \binom{N}{j} \varepsilon^j (1-\varepsilon)^{N-j}$$

и использовать границу Чернова для того, чтобы показать, что

$$\Pr[w \geqslant j] \leqslant \exp\left\{N\left[\mathscr{H}\left(\frac{j}{N}\right) + \frac{j}{N}\ln\varepsilon + \frac{N-j}{N}\ln(1-\varepsilon)\right]\right\}.$$

Сравнить это с вашим результатом в пункте (в).

5.9. (Теорема кодирования для ДСК, использующая биномиальные коэффициенты.) В ДСК с вероятностью ошибки ε рассмотрим ансамбль блоковых кодов с длиной N и с $M = e^{NR}$ кодовыми словами, в котором буквы в кодовых словах выбираются независимо с $Q(0) = Q(1) = 1/2$. Для любого кода из ансамбля рассмотрим декодер, который при заданном $\mathbf{y}$ выбирает сообщение m, для которого $d(\mathbf{x}_m; \mathbf{y})$ минимально, где $d(\mathbf{x}_m; \mathbf{y})$ является расстоянием Хэмминга между $\mathbf{x}_m$ и $\mathbf{y}$ (т. е. равно числу символов, в которых отличаются $\mathbf{x}_m$ и $\mathbf{y}$).

(а) Показать, что этот декодер является декодером по максимуму правдоподобия.

(б) При условии, что на входе кодера было сообщение m, показать, что

$$\Pr[d(\mathbf{x}_m, \mathbf{y}) = i] = \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i},$$

$$\Pr[d(\mathbf{x}_{m'}, \mathbf{y}) = i] = \binom{N}{i} 2^{-N}, \text{ при всех } m' \neq m,$$

где вторая вероятность взята по ансамблю кодов.

(в) Показать, что для заданного сообщения m

$$\Pr[\text{ошибка} \mid d(\mathbf{x}_m, \mathbf{y}) = i] \leqslant \begin{cases} (M-1)\sum_{n=0}^{i}\binom{N}{n}2^{-N} \leqslant \\ 1, \end{cases}$$

$$\leqslant \begin{cases} \frac{N-i}{N-2i}\binom{N}{i}\exp[-N(\ln 2 - R)]; \quad i < \frac{N}{2}, \\ 1. \end{cases}$$

Указание: используйте ваши результаты в задаче 5.7. (в) при $\varepsilon = 1/2$ и $i = N - j$.

(г) Используя пункты (б) и (в), показать, что средняя по ансамблю вероятность ошибки ограничена при любом $j \leqslant N/2$ неравенством

$$\bar{P}_e \leqslant \sum_{i=0}^{j-1} \frac{N-i}{N-2i} \exp\left[-N(\ln 2 - R)\right] \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} + \\ + \sum_{i=j}^{N} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}.$$

(д) Выбрать j так, чтобы удовлетворялись неравенства

$$\mathscr{H}\left(\frac{j-1}{N}\right) < \ln 2 - R \leqslant \mathscr{H}\left(\frac{j}{N}\right)$$

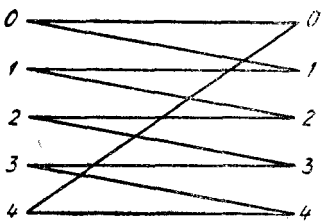
и построить границу сверху для каждой из этих сумм, либо с помощью метода геометрической прогрессии [см. задачу 5.7 (в)], либо оценивая сверху сумму наибольшим слагаемым, умноженным на число слагаемых в сумме. Внимательно следите за тем, выполняется ли неравенство

$$\left(\frac{j}{N-j}\right)^2 < \frac{\varepsilon}{1-\varepsilon}.$$

Показать, что вами получена та же самая экспонента по N, как и в примере 1 § 5.6.

5.10. Найти показатель экспоненты случайного кодирования $E_r(R)$ (в параметрической форме) для двоичного стирающего канала с вероятностью стирания ε. Построить график $E_r(R)$ при $\varepsilon = 1/2$, выбрав некоторые определенные числовые значения $E_r(0)$, $R_{cr} = E_0(\rho)/\partial\rho\,|_{\rho=1}$ и $E_r(R_{cr})$. Дать геометрическую интерпретацию $E_r(R)$ подобную той, которая была дана на рис. 5.6.4 для ДСК.

5.11. (а) Найти показатель экспоненты случайного кодирования $E_r(R)$ для изображенного ниже канала с пятью входами, пятью выходами и всеми переходными вероятностями, равными $1/2$.



(б) Найти код с длиной блока 1 и скоростью $R = \ln 2$ (т. е. со скоростью один двоичный символ на одно использование канала), такой, для которого $P_e = 0$. Найти код с длиной блока 2 и $R = (\ln 5)/2$ (т. е. с пятью кодовыми словами), такой, что $P_e = 0$.

Замечание: пропускная способность канала с нулевой ошибкой C_0 определяется как наибольшая скорость, для которой $P_e = 0$ может быть достигнута при конечной длине блока [Шеннон (1956)]. Таким образом, вы показали, что $C_0 \geqslant (\ln 5)/2$ для этого канала. Неизвестным является то, будет ли C_0 строго больше, чем $(\ln 5)/2$.

5.12. Пусть $\mathbf{Q}$ будет распределением вероятностей на входе, на котором достигается пропускная способность дискретного канала без памяти. Показать, что $E_r(R, \mathbf{Q})$ ведет себя как $\alpha(C - R)^2$, и подсчитать постоянную α.

5.13. Показать, что в произвольном ДКБП с двоичным входом функция $E_0(\rho, \mathbf{Q})$ достигает максимума по $\mathbf{Q}$ при $\rho = 1$ в точке $Q(0) = Q(1) = {}^1/_2$.

5.14. (Декодирование со стираниями и ошибками.) Рассмотрим ДКБП с переходными вероятностями $P(j|k)$. Пусть $Q(k)$ обозначает входные вероятности, на которых достигается пропускная способность, и пусть

$$\omega(j) = \sum_k Q(k) P(j \mid k).$$

Рассмотрим ансамбль кодов с длиной блока N и с $M = 2^{NR}$ кодовыми словами, в которых все символы кодовых слов выбраны независимо с распределением вероятности $Q(k)$.

Рассмотрим декодер, который работает следующим образом. При заданной принятой последовательности $\mathbf{y}$ декодер вычисляет

$$I_m = \ln \frac{P_N(\mathbf{y} \mid \mathbf{x}_m)}{\omega_N(\mathbf{y})} = \sum_{n=1}^{N} \ln \frac{P(y_n \mid x_{m,n})}{\omega(y_n)}$$

для каждого кодового слова $\mathbf{x}_m$, $1 \leqslant m \leqslant M$. Декодер сравнивает все I_m с TN, где T — некоторый фиксированный порог. Если имеется одно и только одно значение m, для которого $I_m \geqslant TN$, то декодер декодирует это сообщение. В противном случае декодер производит стирающий символ и не декодирует никакого сообщения.

Пусть $\overline{P}_1$ будет вероятностью в ансамбле кодов того, что переданное кодовое слово не удовлетворяет порогу, и пусть $\overline{P}_2$ будет вероятностью того, что одно или большее число других кодовых слов удовлетворяют порогу.

(а) Примените границу Чернова для того, чтобы показать, что

$$\overline{P}_1 \leqslant \exp[-N\alpha],$$

$$\overline{P}_2 \leqslant \exp[-N(\alpha + T - R)],$$

$$\alpha = \max_{0 \leqslant s \leqslant 1} -\ln\left[\sum_{j,k} Q(k)\,\omega(j)^s P(j \mid k)^{1-s} e^{sT}\right].$$

(б) Показать, что $\alpha > 0$ при $T < C$, где C пропускная способность канала в натуральных единицах, и что $\alpha \to 0$ при $T \to C$.

(в) Показать, что $\overline{P}_1 + \overline{P}_2$ является верхней границей вероятности стирания при декодировании и что $\overline{P}_2$ является верхней границей для вероятности ошибочного декодирования. Изобразить графически $\alpha + T - R$ как функцию R в пределе при $T \to C$ и проведите сравнение с показателем экспоненты случайного кодирования. Указать, что это означает для канала, в котором имеется бесшумная обратная связь от приемника к передатчику. Подобная, но более сильная граница для вероятности стирания и ошибки имеется у Форни (1968).

5.15. В предыдущей задаче было отмечено, что вероятность того, что не произошло правильное декодирование (т. е. что была либо ошибка, либо стирание) ограничена сверху неравенством $\overline{P}_e \leqslant \overline{P}_1 + \overline{P}_2$; если положить $T = R$, то $\overline{P}_e \leqslant 2\exp(-N\alpha)$, где

$$\alpha = \max_{0 \leqslant s \leqslant 1} \left\{-sR - \ln\left[\sum_{j,k} Q(k)\,\omega(j)^s P(j \mid k)^{1-s}\right]\right\}. \tag{1}$$

Заметим, что так как $\alpha > 0$ при $R < C$, то это дает очень простое доказательство теоремы кодирования (хотя с ненаилучшей экспонентой ошибки).

(а) Заменить s на $\rho/(1+\rho)$ и показать, что в двоичном симметричном канале (1) сводится к

$$\alpha = \max_{\rho \geqslant 0} \frac{1}{1+\rho}[E_0(\rho) - \rho R].$$

Сравнить α с $E_r(R)$ графическим методом подобно тому, как это было сделано на рис. 5.6.3.

(б) Используя неравенство Гёльдера [см. задачу 4.15. (б)], для суммы по j в (1), показать, что в общем ДКБП

$$\alpha \geqslant \max_{\rho \geqslant 0} \frac{1}{1+\rho} [E_0(\rho) - \rho R].$$

(в) Положив $s = \rho$ в (1) и использовав неравенство Гёльдера для

$$\sum_k Q(k) P(j \mid k)^{1/(1+\rho)},$$

показать, что $\alpha \leqslant E_r(R)$.

5.16. *Совместная теорема кодирования для источника и канала.*

(а) Пусть $P_N(\mathbf{y}|\mathbf{x})$ будут переходными вероятностями для последовательностей длины N в дискретном канале; рассмотрим ансамбль кодов, в котором M кодовых слов выбраны независимо, каждое с распределением вероятностей $Q_N(\mathbf{x})$. Пусть сообщения кодируются в эти кодовые слова и имеют распределение вероятностей q_m, $1 \leqslant m \leqslant M$; рассмотрим декодер по максимуму апостериорной вероятности, который при заданном $\mathbf{y}$ выбирает m, для которого $q_m P_N(\mathbf{y}|\mathbf{x}_m)$ максимально. Пусть

$$\overline{P}_e = \sum_m q_m \overline{P}_{e,\,m}$$

будет средней по этому ансамблю сообщений и кодов вероятностью ошибки. Видоизменяя доказательство теоремы 5. 6. 1, там где это необходимо, показать, что

$$\overline{P}_e \leqslant \left[\sum_{m=1}^{M} q_m^{1/(1+\rho)}\right]^{1+\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^{1/(1+\rho)}\right]^{1+\rho}. \qquad (1)$$

(б) Пусть канал является каналом без памяти с переходными вероятностями $P(j|k)$, пусть буквы кодовых слов выбраны независимо с распределением вероятностей $Q(k)$ и пусть сообщениями являются последовательности длины L дискретного источника без памяти U с распределением вероятности $\pi(i)$, $0 \leqslant i \leqslant A-1$. Показать, что (1) равносильно неравенству

$$\overline{P}_e \leqslant \exp\{-NE_0(\rho, \mathbf{Q}) + LE_s(\rho)\}, \qquad (2)$$

$$E_s(\rho) = (1+\rho) \ln\left[\sum_{i=0}^{A-1} \pi(i)^{1/(1+\rho)}\right].$$

(в) Показать, что $E_s(0) = 0$, $\left.\dfrac{\partial E_s(\rho)}{\partial \rho}\right|_{\rho=0} = H(U)$ в натуральных единицах и что $E_s(\rho)$ — строго возрастающая функция ρ (если $\pi(i) < 1$ для всех i).

(г) Пусть $\lambda = L/N$ и пусть $N \to \infty$ при фиксированном λ. Показать, что $\overline{P}_e \to 0$, если $\lambda H(U) < C$ при соответствующем выборе $Q(k)$.

(д) Показать, что граница (2) равносильна (5.6.13) в случае, когда $\pi(i) = 1/A$ при $0 \leqslant i \leqslant A-1$, и равносильна положительной части утверждения теоремы кодирования для источника 3.1.1 (за исключением экспоненциальной сходимости здесь) в случае, когда канал является каналом без шума. Указанные выше результаты были получены автором и впервые использованы в 1964 г.

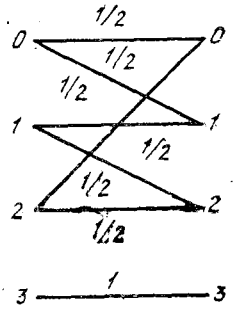
5.17. Рассмотрим сумму каналов (как в задаче 4.18), соответствующую множеству из n ДКБП. Пусть $E_{0,\,i}(\rho) = \max_{\mathbf{Q}} E_{0,\,i}(\rho, \mathbf{Q})$ для i-го канала из множества ДКБП и пусть $E_0(\rho) = \max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})$ для суммы каналов. Пусть q_i — вероятность использования i-го канала, так что на этом распределении вероятности

достигается максимум $E_0(\rho)$, и пусть $Q_i(k)$ является вероятностью использования входа k в i-м канале при условии, что используется i-й канал. Показать, что

$$\exp\left[\frac{E_0(\rho)}{\rho}\right]=\sum_{i=1}^{n}\exp\left[\frac{E_{0,\,i}(\rho)}{\rho}\right],$$

$$q_i=\frac{\exp\left[\dfrac{E_{0,\,i}(\rho)}{\rho}\right]}{\sum\limits_{i=1}^{n}\exp\left[\dfrac{E_{0,\,i}(\rho)}{\rho}\right]}.$$

Применить ваш результат для нахождения $E_0(\rho)$ канала, изображенного на рисунке.



5.18. Следующее доказательство теоремы кодирования не позволяет получить точную границу вероятности ошибки, найденной в § 5.6, но выявляет с большей ясностью значение пропускной способности канала. Пусть $P(j|k)$ обозначают переходные вероятности в ДКБП, пусть $Q(k)$ — распределение на входе, на котором достигается пропускная способность, и пусть

$$\omega(j)=\sum_k Q(k)\,P(j\mid k).$$

Пусть для блока любой длины N

$$P_N(\mathbf{y}\mid\mathbf{x})=\prod_{n=1}^{N}P(y_n\mid x_n),\quad Q_N(\mathbf{x})=\prod_n Q(x_n),$$

$$\omega_N(\mathbf{y})=\prod_n\omega(y_n).$$

Предположим, что R является произвольной скоростью, $R<C$, и для каждого N выберем код из $M=\lceil e^{NR}\rceil$ кодовых слов с помощью независимого выбора слов с распределением вероятности $Q_N(\mathbf{x})$. Пусть $\varepsilon=(C-R)/2$ и для каждого N определим «типичное» множество T_N как множество пар $\mathbf{x}$, $\mathbf{y}$, для которых

$$\left|\frac{1}{N}\,I(\mathbf{x};\,\mathbf{y})-C\right|\leqslant\varepsilon,$$

где

$$I(\bar{\mathbf{x}};\,\bar{\mathbf{y}})=\ln\frac{P_N(\mathbf{y}\mid\mathbf{x})}{\omega_N(\mathbf{y})}.$$

Для любого N и любого кода из ансамбля рассмотрим декодер, который при заданном $\mathbf{y}$ выбирает m, для которого $(\mathbf{x}_m, \mathbf{y})$ принадлежит T_N. Если такие кодовые слова отсутствуют или их больше чем одно, то будем считать, что произошла ошибка.

(а) Показать, что при заданном N вероятность ошибки при условии, что на кодер поступило сообщение m, удовлетворяет неравенству

$$\overline{P}_{e,\,m} \leqslant \Pr\left[(\mathbf{x}_m,\ \mathbf{y}) \notin T_N \mid m\right] + \sum_{m' \neq m} \Pr\left[(\mathbf{x}_{m'},\ \mathbf{y}) \in T_N \mid m\right].$$

(б) Показать, что в ансамбле кодов $\Pr[(\mathbf{x}_m, \mathbf{y}) \notin T_N | m]$ стремится к нулю при N, стремящемся к ∞.

Указание: рассмотрите $I(\mathbf{x}_m; \mathbf{y})$ как сумму независимых одинаково распределенных случайных величин и используйте закон больших чисел.

(в) Показать, что в ансамбле кодов при любом $m' \neq m$ имеем

$$\Pr\left[(\mathbf{x}_{m'},\ \mathbf{y}) \in T_N \mid m\right] \leqslant \exp\left[-N(C-\varepsilon)\right].$$

Указание: покажите, что

$$\Pr\left[(\mathbf{x}_{m'},\ \mathbf{y}) \in T_N \mid m\right] = \sum_{(\mathbf{x}_{m'},\ \mathbf{y}) \in T_N} \left[Q_N(\mathbf{x}_{m'})\,\omega_N(\mathbf{y})\right].$$

Затем покажите, что при $(\mathbf{x}_{m'},\ \mathbf{y}) \in T_N$

$$\omega_N(\mathbf{y}) \leqslant P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \exp\left[-N(C-\varepsilon)\right].$$

(г) На основе результатов пунктов (а), (б) и (в) показать, что $\overline{P}_{e,\,m}$ стремится к нулю для каждого m при N, стремящемся к ∞.

5.19. Пусть $\mathbf{x}_1, \ldots, \mathbf{x}_M$ является множеством кодовых слов с длиной блока N, которые используются в ДСК с вероятностью ошибки ε. Пусть $d(\mathbf{x}_m, \mathbf{x}_{m'})$ — расстояние Хэмминга между $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ (т. е. число позиций, в которых $\mathbf{x}_m$ отличается от $\mathbf{x}_{m'}$).

(а) Используя границу Чернова и аддитивную границу, показать, что для декодера по максимуму правдоподобия

$$P_{e,\,m} \leqslant \sum_{m' \neq m} \exp\left[d(\mathbf{x}_m,\ \mathbf{x}_{m'}) \ln \sqrt{(4\varepsilon(1-\varepsilon))}\right].$$

(б) Минимальное расстояние в коде определяется как

$$d_{min} = \min_{m \neq m'} d(\mathbf{x}_m,\ \mathbf{x}_{m'}).$$

Показать, что при всех m

$$P_{e,\,m} \leqslant (M-1) \exp\left[d_{min} \ln \sqrt{(4\varepsilon(1-\varepsilon))}\right].$$

(в) Рассмотрим код с заданным минимальным расстоянием $d_{min} = d$, выбираемый в соответствии со следующей процедурой.

I. Запишем все 2^N различных двоичных последовательностей длины N.

II. Выберем произвольное кодовое слово из этого списка.

III. Удалим из списка только что выбранное кодовое слово и все последовательности длины N, находящиеся на расстоянии $d-1$ или меньшем от этого кодового слова.

Если список окажется пустым, то произвести остановку, в противном случае перейти к шагу (II).

Показать, что число кодовых слов, выбранных таким образом, удовлетворяет неравенству

$$M \geqslant 2^N \left[\sum_{i=0}^{d-1} \binom{N}{i}\right]^{-1}.$$

Оно называется границей Гилберта (1952).

(г) Объединяя полученные выше результаты с результатами задачи 5.8, показать, что при $d \leqslant N/2$ имеем

$$M \geqslant \exp\left\{N\left[\ln 2 - \mathscr{H}\left(\frac{d_{min}-1}{N}\right)\right]\right\}.$$

Показать, что отсюда вытекает, что при любой скорости $R = (\ln M)/N < \ln 2$ существует код с минимальным расстоянием $d_{min} \geqslant \delta N$, где $\delta < 1/2$ определяется из уравнения $\mathscr{H}(\delta) = \ln 2 - R$.

(д) Объединяя результаты пунктов (б) и (г), получить границу вероятности ошибки в виде $P_e \leqslant \exp[-NE(R)]$ и выписать параметрические выражения, определяющие $E(R)$. Показать, что при нулевой скорости эта граница согласуется с (5.5.1). Вычертить график $E(R)$ при $\varepsilon = 0{,}01$ и произвести сравнение с $E_r(R)$. [Заметьте, что, как показано в § 5.7, $E(R) < E_{ex}(R)$].

5.20. Предположим, что условия теоремы 5.6.1 видоизменены так, что декодер при заданной принятой последовательности y выдает список L из сообщений $m_1, \ldots, m_L$, для которых $P_N(\mathrm{y}|\mathrm{x}_m)$ являются наибольшими (здесь L — заданное целое число). Если переданное сообщение не входит в список, то считается, что произошла ошибка при декодировании списком. Пусть $\overline{P}_{L,e}$ является вероятностью ошибки при декодировании списком.

(а) Показать, что для любого ρ_0, $0 \leqslant \rho_0 \leqslant 1$, и любого $s > 0$ имеем

$$\Pr\left[\begin{matrix}\text{ошибка при}\\ \text{декодирова-}\\ \text{нии списком}\end{matrix}\middle| m, \mathrm{x}_m, \mathrm{y}\right] \leqslant \binom{M-1}{L}^{\rho_0}\left\{\sum_{\mathrm{x}} Q_N(\mathrm{x})\left[\frac{P_N(\mathrm{y}|\mathrm{x})}{P_N(\mathrm{y}|\mathrm{x}_m)}\right]^s\right\}^{L\rho_0}.$$

(б) Оценить сверху $\binom{M-1}{L}$ с помощью $(M-1)^L$, положить $\rho = L\rho_0$ и показать, что при $0 \leqslant \rho \leqslant L$ справедливо неравенство

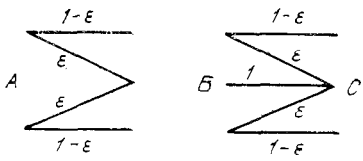
$$\overline{P}_{L,e} \leqslant (M-1)^{\rho}\sum_{\mathrm{y}}\left(\sum_{\mathrm{x}} Q_N(\mathrm{x})\,P_N(\mathrm{y}|\mathrm{x})^{1/1+\rho}\right)^{1+\rho}.$$

(в) Показать, что для ДКБП последнее неравенство принимает вид:

$$\overline{P}_{L,e} \exp\left[-N\left\{\max_{0\leqslant\rho\leqslant L}\left[-\rho R + E_0(\rho, \mathbf{Q})\right]\right\}\right].$$

Изобразите графически выражение, стоящее в фигурных скобках, как функцию R при заданном L и проведите сравнение с $E_r(R, \mathbf{Q})$.

5.21. Рассмотрим следующие возможности для передачи данных из точки A в точку C (см. рисунок, изображенный ниже).



I. В точке B каждый выход первого канала непосредственно соединяется с соответствующим входом второго канала и данные передаются с использованием блокового кода длины N со скоростью R.

II. Блоковый код длины $N/2$ со скоростью R используется в первом канале, и данные декодируются в точке B и вновь кодируются блоковым кодом длины $N/2$ со скоростью R для передачи по второму каналу с использованием только двух внешних входов.

III. Двоичные символы источника непосредственно передаются из A в C при соответствующем соединении входов и выходов в точке B. С помощью бесшумного обратного канала передача организуется так, что символы источника повторяются вновь всегда тогда, когда принимается средний выход в точке C.

(а) Найти показатели экспонент случайного кодирования при передачах I и II и сравнить эти методы передачи на основе найденных экспонент.

(б) Применить границу Чернова для оценки сверху вероятности того, что меньше чем RN символов источника будут переданы за N использований канала при передаче III, и сравнить с двумя первыми методами передачи.

5.22. Дискретный канал без памяти имеет переходные вероятности $P(j \mid k)$. К сожалению, декодер для канала является декодером максимального правдоподобия, построенным при ошибочном представлении, что переходные вероятности равны $P'(j|k)$. Т. е. сообщение m декодируется, если $P'_N(\mathbf{y}|\mathbf{x}_m) > P'_N(\mathbf{y}|\mathbf{x}_m)$ для всех $m' > m$, где

$$P'_N(\mathbf{y} \mid \mathbf{x}_m) = \prod_n P'(y_n \mid x_n).$$

Найти верхнюю границу для средней вероятности ошибочного декодирования в ансамбле кодов, в котором буквы кодовых слов выбраны независимо с распределением вероятности $Q(k)$. Ваша граница должна иметь вид

$$\overline{P}_e \leqslant \exp\{-N[-\rho R + f(\rho, \mathbf{Q}, \mathbf{P}, \mathbf{P}')]\} \text{ при } 0 \leqslant \rho \leqslant 1.$$

Найти пример переходных вероятностей $\mathbf{P}$ и $\mathbf{P}'$, для которых f является неположительной, и объяснить, почему это не удивительно. См. Стиглитц (1966).

5.23. В заданном ДКБП с переходными вероятностями $P(j|k)$ разложить $E_0(\rho, \mathbf{Q})$ при $\mathbf{Q}$, на котором достигается пропускная способность, в степенной ряд для того, чтобы показать, что

$$\overline{P}_e \leqslant \exp\left\{-N\left[\frac{(C-R)^2}{2\alpha}\right]\right\} \tag{I}$$

при $C - R \ll \alpha$, где α является верхней границей для $-E_0''(\rho, \mathbf{Q})$ при $0 \leqslant \rho \leqslant 1$. Показать, что

$$-E_0'(\rho, \mathbf{Q}) = \sum_j \omega_j \sum_k q_{kj} \ln\left[\frac{Q(k)}{P(j \mid k)^{1/(1+\rho)}}\right], \tag{II}$$

где

$$\omega_j = \frac{\alpha_j^{1+\rho}}{\sum_j \alpha_j^{1+\rho}}, \quad q_{kj} = Q(k)\frac{P(j \mid k)^{1/(1+\rho)}}{\alpha_j},$$

$$\alpha_j = \sum_k Q(k) P(j \mid k)^{1/(1+\rho)}.$$

Взяв производную в (II), показать, что

$$-E_0''(\rho, \mathbf{Q}) \leqslant \sum_j \omega_j \sum_k q_{kj}\left(\ln\frac{Q(k)}{q_{kj}}\right)^2.$$

На основе этого показать, что

$$\overline{P}_e \leqslant \exp\left\{-N\left[\frac{(C-R)^2}{8/e^2 + 4[\ln J]^2}\right]\right\},$$

где J — объем выходного алфавита.

5.24. Рассмотрим ДКБП, в котором пропускная способность с нулевой ошибкой равна нулю (т. е. $R_{x,\infty}$, определенная равенством (5.7.16), равна нулю). Показать что

$$\lim_{R \to 0} E_{ex}(R, \mathbf{Q}) = -\sum_{k,i} Q(k) Q(i) \ln\left[\sum_j \sqrt{P(j \mid k) P(j \mid i)}\right].$$

Указание: покажите сначала, что $\lim_{R\to 0} E_{ex}(R, \mathbf{Q}) = \lim_{\rho\to\infty} E_x(\rho, \mathbf{Q})$. Далее используйте либо правило Лопиталя, либо разложите $E_x(\rho, \mathbf{Q})$ в ряд по степеням $1/\rho$.

5.25. Показать, что $R_{x,\infty}$ [см. (5.7.16)] задается равенством $R_{x,\infty} = \ln L$, где L — объем наибольшего множества I целых чисел, такого, что для всех $i \in I$, $k \in I$, $i \neq k$ имеем $\varphi_{k,i} = 0$ [см. (5.7.15)]. (Другими словами, I является наибольшим множеством входов, таким, что никакой выход не может быть достигнут из более чем одного входа этого множества; при использовании только этого множества входов каждая выходная буква однозначно определяет вход).

Указание: предположите, что $\varphi_{0,1} = 1$ и покажите, что

$$\sum_{k,i} Q(k) Q(i) \varphi_{k,i} = [Q(0) + Q(1)]^2 +$$

$$+ 2 \sum_{i=2}^{K-1} Q(i) [Q(0) \varphi_{0,i} + Q(1) \varphi_{1,i}] + \sum_{k=2}^{K-1} \sum_{i=2}^{K-1} Q(k) Q(i) \varphi_{k,i}.$$

Показать, что при любых заданных значениях $Q(2), \ldots, Q(K-1)$ это выражение достигает минимума по $Q(0)$ и $Q(1)$ либо при $Q(0)=0$, либо при $Q(1) = 0$. Используйте это для того, чтобы показать, что

$$\sum_{k,i} Q(k) Q(i) \varphi_{k,i}$$

достигает минимума при $Q(k)$, неравных нулю только на некотором множестве I, для которого $\varphi_{k,i} = 0$ при всех $k \in I$, $i \in I$, $k \neq i$.

5.26. Для канала из задачи 5.11 выбрать длину блока $N=2$ и $R = (\ln 5)/2$. Показать, что выражение для $P_{e,m}$ из (5.7.7) не достигает минимума по ρ и $Q_N(\mathbf{x})$ на произведении распределений.

Указание: внимательно рассмотрите ваше решение пункта (б) задачи 5.11.

5.27. Показать, что $E_x(\rho, \mathbf{Q})$ является положительной при всех $\rho > 0$ и что

$$\partial E_x(\rho, \mathbf{Q})/\partial\rho \leqslant -\ln \sum_k [Q(k)]^2.$$

5.28. Вычислить и изобразить на графике $E_x(\rho, \mathbf{Q})$ и $E_0(\rho, \mathbf{Q})$ для двоичного канала без шума при $Q(0) = 0,1$ и $Q(1) = 0,9$. Построить графики $E_{ex}(R, \mathbf{Q})$ и $E_r(R, \mathbf{Q})$.

5.29. Показать, что для любых двух входов ДКБП $E_x(\rho, \mathbf{Q})$ достигает максимума по $\mathbf{Q}$ на $Q(0) = Q(1) = 1/2$.

Указание: заметьте, что слагаемые в двойной сумме по k и i в определении E_x принимают только два различных значения: одно при $k = i$ и другое при $k \neq i$. Этот результат и результат следующей задачи были получены Джелинеком (1968).

5.30. (а) Показать, что если матрица A с элементами

$$a_{ik} = \left[\sum_j \sqrt{P(j|k) P(j|i)}\right]^{1/\rho}$$

является неотрицательно определенной, то

$$\sum_{k,i} Q(k) Q(i) \left[\sum_j \sqrt{P(j|k) P(j|i)}\right]^{1/\rho}$$

является выпуклой $\cup$ по $\mathbf{Q}$.

(б) Показать, что из этого также следует, что матрица с элементами

$$\left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}|\mathbf{x}) P_N(\mathbf{y}|\mathbf{x}')}\right]^{1/\rho},$$

имеющими индексы $\mathbf{x}$, $\mathbf{x}'$, также неотрицательно определена, и поэтому

$$\sum_{\mathbf{x},\,\mathbf{x}'} Q_N(\mathbf{x})\,Q_N(\mathbf{x}')\left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}\mid\mathbf{x})\,P_N(\mathbf{y}\mid\mathbf{x}')}\,\right]^{1/\rho}$$

является выпуклой $\cup$ функцией $\mathbf{Q}_N$.

Указание: покажите, что собственные значения последней матрицы являются произведениями N собственных значений первой матрицы. Покажите на основе этого, что указанная выше сумма по $\mathbf{x}$ и $\mathbf{x}'$ достигает минимума на произведении распределений.

Указание: см. пример 4 § 5.6.

5.31. Показать, что в пределе для канала с очень большим шумом

$$\lim_{R\to 0} E_{ex}(R, \mathbf{Q})$$

[см. (5.7.20)] сходится к $E_r(R, \mathbf{Q})$. Используйте этот результат, чтобы показать, что в таком канале $E_{ex}(R, \mathbf{Q})$ стремится к $E_r(R, \mathbf{Q})$ при $R \leqslant R_{cr}$.

5.32. Пусть для любого заданного входа и метода декодирования $P_{e,m}$ обозначает вероятность ошибки для m-го кодового слова и

$$P_{max} = \max_{1\leqslant m\leqslant M} P_{e,\,m}.$$

Пусть $P_{max}(N, M)$ является минимальным значением P_{max} для заданного канала при вариации по всем кодам с длиной блока N и с M кодовыми словами. Показать, что

$$P_{max}(N, M) \geqslant P_e(N, M)$$

и что

$$P_{max}(N, M) \leqslant 2P_e(N, 2M),$$

где $P_e(N, M)$ определена в § 5.8.

Указание: см. следствие 2 теоремы 5.6.2.

5.33. Доказать, что граница сферической упаковки для ДСК (5.8.19) остается справедливой в случае, если между приемником и передатчиком имеется обратная связь (т. е. если передатчик знает, что было получено, и каждый переданный символ является функцией как сообщения, так и предыдущих принятых символов).

Указание: пусть опять Y_m будет множеством принятых последовательностей, декодированных в сообщение m; покажите, что эти множества взаимно исключают друг друга. Пусть $z(y)$ является последовательностью ошибок для y, т. е., если $\mathbf{y}\in Y_m$ и $\mathbf{x}_m(\mathbf{y})$ является переданной последовательностью для сообщения m и принятой последовательностью $\mathbf{y}$, то полагаем $\mathbf{z}(\mathbf{y}) = \mathbf{x}_m(\mathbf{y}) + \mathbf{y}$. Покажите, что если для заданного m $\mathbf{y}\in Y_m$ и $\mathbf{y}'\in Y_m$ и $\mathbf{y}\neq\mathbf{y}'$, то $\mathbf{z}(\mathbf{y})\neq\mathbf{z}(\mathbf{y}')$.

5.34. Показать, что параметр A в теореме 5.8.5 удовлетворяет неравенству:

$$A \leqslant \frac{4}{e^2} + \frac{2(\ln J)(\ln J + 2/e)}{\min\limits_{j}\max\limits_{k} P(j\mid k)}.$$

Указание: заметьте, что

$$A \leqslant \max_k \sum_j P(j\mid k)\left[\ln\frac{P(j\mid k)}{\omega(j)}\right]^2,$$

$$\sum_j P(j\mid k)\left[\ln\frac{P(j\mid k)}{\omega(j)}\right]^2 \leqslant \sum_{j\,:\,P(j\mid k)<\omega(j)} P(j\mid k)\times$$

$$\times\left[\ln\frac{\omega(j)}{P(j\mid k)}\right]^2 + \sum_{j\,:\,P(j\mid k)>\omega(j)} P(j\mid k)\ln\frac{P(j\mid k)}{\omega(j)}\ln\frac{1}{\omega(j)}.$$

Для первой суммы используйте неравенство

$$[\ln x]^2 \leqslant \frac{4}{e^2} x \quad \text{при} \quad x \geqslant 1.$$

Для второй суммы используйте (5.8.60) для того, чтобы показать, что

$$\ln \frac{1}{\omega(j)} \leqslant \frac{C + H(Y \mid x = k)}{P(j \mid k)} \quad \text{при всех } k.$$

5.35. Используйте границу Чернова вместо неравенства Чебышева в (5.8.67) для того, чтобы показать, что при фиксированной $R > C$ имеем

$$P_e(N, \lceil e^{NR} \rceil) \geqslant 1 - 2\exp[-N\alpha(R)],$$

где $\alpha(R) > 0$ при $R > C$.

5.36. Примените центральную предельную теорему в форме Берри — Эссена (см. Феллер (1968), гл. XVI, § 5) вместо неравенства Чебышева в (5.8.67). Предположите, что канал такой, что выражение $\ln[P(j|k)/\omega(j)]$ не зависит от j при любом k. Пусть δ — произвольное действительное число (положительное или отрицательное) и пусть $R(\delta, N) = C + \delta/\sqrt{N}$. Показать, что для любого $\varepsilon > 0$ существует такое $N(\delta, \varepsilon)$, что для всех $N \geqslant N(\delta, \varepsilon)$

$$P_e\{N, \lceil \exp[NR(\delta, N)] \rceil\} \geqslant f(\delta) - \varepsilon,$$

где функция $f(\delta)$ является положительной при всех δ, возрастающей по δ, и $f(0) = 1/2$.

5.37. (а) Пусть задан канал с конечным числом состояний. Предположите, что передатчик знает начальное состояние и использует отдельный код для каждого начального состояния. Пусть декодер использует то же самое правило декодирования, как и в (5.9.1) — (5.9.5), показать, что в этом случае можно изменить порядок взятия минимума и максимума в (5.9.5) [см. Юдкин (1967)].

(б) Предположим, что приемник знает начальное состояние и декодирует сообщение m, которое максимизирует $P_N(\mathbf{y}/\mathbf{x}, s_0)$. Показать, что (5.9.5) также справедливо и в этом случае, и что множитель $A^{1+\rho}$ можно опустить в выражении для границы.

(в) Для канала, изображенного на рис. 4.6.3, показать, что минимакс в (5.9.5) достигается на входном распределении, соответствующем равновероятным независимым входам. Показать, что максмин в пункте (а) достигается на том же самом распределении, на котором достигается $\bar{C}$ (см. § 4.6).

5.38. Рассмотрим канал с двоичными входом и выходом, в котором выходные символы y_n связаны со входными символами x_n равенством $y_n = x_n \oplus z_n$. Предположим, что шумовая последовательность $z_1, z_2, \ldots$ не зависит от входа и представляет собой выход марковского источника с эргодической последовательностью состояний (см. § 3.6).

(а) Показать, что пропускная способность канала равна $\log 2 - H_\infty(Z)$, где $H_\infty(Z)$ является энтропией шумовой последовательности [см. (3.6.21)]. Показать, что эта пропускная способность больше, чем пропускная способность ДСК с вероятностью ошибки $P_{z_n}(1)$.

(б) Рассмотрим ансамбль кодов с $Q_N(\mathbf{x}) = 2^{-N}$ для всех $\mathbf{x}$. Показать, что

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) = \rho \ln 2 - \frac{1+\rho}{N} \ln \sum_{\mathbf{z}} P_{\mathbf{z}}(\mathbf{z} \mid s_0)^{1/(1+\rho)}.$$

(в) Пусть $[\alpha(\rho)]$ — матрица порядка $A \times A$ с элементами

$$\alpha_{l,i} = \sum_{z_n} P(z_n, s_n = i \mid s_{n-1} = l)^{1/(1+\rho)}.$$

Пусть $\lambda(\rho)$ является наибольшим собственным значением матрицы $[\alpha(\rho)]$. Показать, что

$$\lim_{N\to\infty} E_{0,N}(\rho, Q_N, s_0) = \rho \ln 2 - (1+\rho)\ln\lambda(\rho).$$

5.39. Рассмотрим класс неразложимых каналов с конечным числом состояний, в которых отсутствует память, связанная с межсимвольной интерференцией (т. е. последовательность состояний не зависит от входа) и в которых s_n является функцией y_n при каждом n. Например, канал, изображенный на *рис. 5.9.1, принадлежит этому классу. Показать*, что для любого канала из этого класса, пропускная способность достигается на независимых одинаково распределенных входах и что пропускная способность равна пропускной способности ДКБП с

$$P(y_n \mid x_n) = \sum_{s_{n-1}} P(y_n \mid x_n, s_{n-1})\, q(s_{n-1}).$$

Глава 6

6.1. Код отображает пары информационных символов в кодовые слова длины 5 по следующему правилу:

Информационные последовательности	Кодовые слова
00	00000
01	01101
10	10111
11	11010

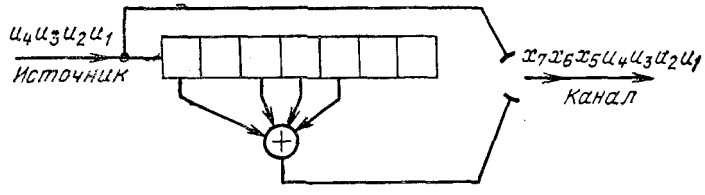
(а) Показать, что полученный код является систематическим кодом с проверкой на четность и выразить каждый символ кодового слова в виде линейной *комбинации информационных символов*.

(б) Найти для этого кода порождающую и проверочную матрицы.

(в) Привести таблицу декодирования для случая декодирования по максимуму правдоподобия в ДСК с переходной вероятностью $\varepsilon < 1/2$.

(г) Сколько конфигураций 1, 2, и 3 ошибок исправляется при использовании этой таблицы декодирования? Сколько вообще существует конфигураций 1, 2 и 3 ошибок? Найти вероятность неправильного декодирования.

6.2. Представленное на рисунке устройство используется для кодирования двоичных символов при передаче по двоичному симметричному каналу с пере-



ходной вероятностью $\varepsilon < 1/2$. Первоначально регистр сдвига заполнен нулями; затем в регистр поступают четыре информационных символа и одновременно передаются по каналу. После этого передаются три проверочных символа; перед

вычислением каждого проверочного символа все четыре информационных символа сдвигаются в регистре сдвига на одну позицию вправо по сравнению с предыдущим положением.

Найти проверочную матрицу, порождающую матрицу, таблицу декодирования и вероятность ошибочного декодирования для этого кода.

6.3. (а) Показать, что в коде с проверкой на четность либо все кодовые слова содержат четное число единиц, либо половина кодовых слов содержит нечетное число единиц и половина — четное.

(б) Пусть $x_{m,n}$ — n-й символ в m-м кодовом слове кода с проверкой на четность. Показать, что при любом заданном n или ровно половина, или все $x_{m,n}$ равны нулю. Объяснить, как можно улучшить код, если все $x_{m,n}$ при данном n равны нулю.

(в) Показать, что среднее число единиц в кодовом слове, усредненное по всем кодовым словам блокового кода длины N с проверкой на четность, не должно превышать $N/2$.

6.4. Вес двоичной последовательности длины N определяется как число единиц в этой последовательности. Расстояние (Хэмминга) между двумя двоичными последовательностями длины N определяется как вес их суммы по модулю 2.

(а) Пусть $\mathbf{x}_1$ — произвольное кодовое слово в блоковом коде длины N с проверкой на четность и пусть $\mathbf{x}_0$ — кодовое слово, целиком состоящее из нулей, которое соответствует информационной последовательности, целиком состоящей из нулей. Показать, что при любом $n \leqslant N$ число кодовых слов, лежащих на расстоянии n от $\mathbf{x}_1$, совпадает с числом кодовых слов, лежащих на расстоянии n от $\mathbf{x}_0$.

(б) Наименьшее расстояние d_{min} в двоичном коде определяется как наименьшее расстояние между всеми возможными парами кодовых слов. Показать, что для кода с проверкой на четность d_{min} равно наименьшему весу кодовых слов, отличных от $\mathbf{x}_0 = (0, 0, \ldots, 0)$.

(в) Показать, что если двоичный код с наименьшим расстоянием d_{min} используется в каком-либо канале с двоичной выходной последовательностью, то можно исправлять все комбинации, содержащие менее чем $d_{min}/2$ ошибок.

Указание: покажите, что если произошло менее чем $d_{min}/2$ ошибок, то принятая последовательность будет ближе в смысле расстояния Хэмминга к переданному кодовому слову, чем к любому другому кодовому слову.

(г) Показать, что если двоичный код с наименьшим расстоянием d_{min} используется в двоичном стирающем канале, то можно исправлять все комбинации, содержащие менее чем d_{min} стираний.

6.5. Показать, что если код с проверкой на четность имеет нечетный наименьший вес, то прибавление проверочного символа, участвующего в проверке каждого символа в коде, увеличивает наименьший вес на 1.

6.6. (Граница Хэмминга.) Исправляющая способность двоичного блокового кода определяется как наибольшее целое e, такое, что все комбинации не более чем e ошибок в блоке могут быть исправлены.

(а) Сколько различных синдромных последовательностей имеется в (N, L)-коде с проверкой на четность? Сколько различных конфигураций из j ошибок могут появиться в последовательности N символов? Показать на основе этого, что исправляющая способность e для (N, L)-кода с проверкой на четность должна удовлетворять неравенству

$$\sum_{j=0}^{e} \binom{N}{j} \leqslant 2^{N-L}.$$

(б) Показать, что в произвольном двоичном коде с длиной блока N и M кодовыми словами e должно удовлетворять соответствующему неравенству

$$\sum_{j=0}^{e} \binom{N}{j} \leqslant \frac{2^N}{M}.$$

Указание: рассмотреть множество последовательностей, расположенных на расстоянии, не большем e от каждого кодового слова и показать, что эти множества не должны пересекаться.

(в) Используя пункт (б) и задачу 6.4. (в), показать, что

$$\sum_{j=0}^{\left\lfloor \frac{d_{min}-1}{2} \right\rfloor} \binom{N}{j} \leqslant \frac{2^N}{M}.$$

6.7. (Граница Плоткина) (а). Используя решение задачи 6.3 (в), показать, что средний вес ненулевого кодового слова в $(N,\ L)$-коде с проверкой на четность не превышает $\frac{N}{2}\frac{2^L}{2^L-1}$. Показать отсюда, что справедливо следующее неравенство для d_{min}:

$$d_{min} \leqslant \frac{N}{2}\left(\frac{2^L}{2^L-1}\right). \tag{I}$$

Заметим, что эта оценка согласуется с оценкой (5.8.1), справедливой для произвольных двоичных кодов.

(б) Приведенная выше оценка эффективна, когда L мало по сравнению с N. При больших значениях L более точной является приводимая ниже оценка. Показать, что при всех j, $1 \leqslant j \leqslant L$,

$$d_{min} \leqslant \frac{N-L+j}{2}\left(\frac{2^j}{2^j-1}\right). \tag{II}$$

Указание: рассмотрите 2^j кодовых слов в коде, у которых первые $L-j$ информационных символов равны 0. Считая это множество кодовых слов с опущенными первыми $L-j$ информационными символами $(N-L+j,\ j)$-кодом с проверкой на четность, примените оценку, полученную в пункте (а).

Дополнение. Показать, что (II) выполняется для любого двоичного кода с длиной блока N и числом кодовых слов, равным 2^L.

(в) Пусть N и d_{min} фиксированы и $N \geqslant 2d_{min} - 1$. Показать, что число проверочных символов должно удовлетворять неравенству

$$N-L \geqslant 2d_{min}-2-\lfloor \log_2 d_{min} \rfloor.$$

Указание: положите $j = 1 + \lfloor \log_2 d_{min} \rfloor$ и воспользуйтесь тем, что числа $N-L$, d_{min} и j целые.

(г) Показать, что при фиксированном $d_{min}/N < 1/2$ при переходе к пределу $N \to \infty$ получим, что скорость в двоичных единицах $R = L/N$ должна удовлетворять неравенству

$$R \leqslant 1 - \frac{2d_{min}}{N}.$$

6.8. (Граница Варшамова (1957) — Гилберта.) Рассмотрим следующий метод построения проверочной матрицы для кода с проверкой на четность. При заданном числе r проверочных символов возьмем диагональную матрицу $r \times r$ и начнем добавлять строки по r двоичных символов выше этой совокупности из r строк. До некоторого заданного числа d имеется гарантия того, что каждая вновь выбираемая строка не является линейной комбинацией каких-либо $d-2$ ранее выбранных строк. Процедура заканчивается, когда не найдется более строк, удовлетворяющих этому условию; длина блока N в коде равна общему числу строк, из которых состоит матрица.

(а) Показать, что наименьшее расстояние построенного таким образом кода не меньше d.

(б) Показать, что общее число линейных комбинаций из $d-2$ строк равно

$$\sum_{i=0}^{d-2} \binom{N}{i};$$

используя это, показать, что N должно удовлетворять соотношению

$$\sum_{i=0}^{d-2} \binom{N}{i} \geqslant 2^r = 2^{N-L},$$

где L — число информационных символов в коде.

Замечание. Из этого неравенства следует нижняя граница для наименьшего расстояния, которого можно достичь в кодах с проверкой на четность (и, следовательно, в произвольном двоичном коде). Заметим, что эта граница несколько сильнее, чем граница Гилберта в задаче 5.19, поскольку она увеличивает достижимое значение d_{min} на 1.

6.9. Рассмотрим два кода с проверкой на четность. Код I строится следующим образом:

$$x_1 = u_1, \qquad x_4 = u_1 \oplus u_2,$$
$$x_2 = u_2, \qquad x_5 = u_1 \oplus u_3,$$
$$x_3 = u_3, \qquad x_6 = u_2 \oplus u_3,$$
$$x_7 = u_1 \oplus u_2 \oplus u_3.$$

Код II строится совершенно аналогично, за исключением того, что $x_6 = u_2$.

(а) Найти порождающую и проверочную матрицы для кода I.

(б) Найти таблицу декодирования для кода I для ДСК с переходной вероятностью $\varepsilon < 1/2$.

(в) Найти точное выражение вероятности ошибочного декодирования для кодов I и II. Какая из этих вероятностей больше?

(г) Найти d_{min} для кодов I и II.

(д) Построить опровергающий пример для утверждения, что если наименьшее расстояние одного (N, L)-кода с проверкой на четность больше, чем наименьшее расстояние другого (N, L)-кода с проверкой на четность, то первый код имеет в ДСК меньшую вероятность ошибки.

6.10. Рассмотрим код с проверкой на четность, у которого строки порождающей матрицы не являются линейно независимыми. Показать, что некоторая ненулевая информационная последовательность отображается в нулевое кодовое слово. Используя это, показать, что для каждой информационной последовательности существует по крайней мере одна другая информационная последовательность, которая отображается в то же самое кодовое слово. Ясно, что такие коды не интересны для практики.

6.11. Рассмотрим две проверочные матрицы H_1 и H_2, столбцы которых образуют одно и то же пространство (т. е. у которых одно и то же множество линейных комбинаций столбцов).

(а) Показать, что последовательность $\mathbf{x}$ удовлетворяет соотношению $\mathbf{x}H_2 = 0$ тогда и только тогда, когда $\mathbf{x}H_1 = 0$ и, следовательно, H_1 и H_2 соответствуют одному и тому же множеству кодовых слов.

(б) Показать, что синдромы двух шумовых последовательностей, вычисленные по H_2, не совпадают (т. е. $\mathbf{e}_1 H_2 \neq \mathbf{e}_2 H_2$) тогда и только тогда, когда не совпадают синдромы, вычисленные по H_1. Используя это, показать, что таблицы декодирования, построенные по H_2 и по H_1, исправляют одно и то же множество шумовых последовательностей.

(в) Допустим, что число столбцов H_1 равно r и мощность наибольшего множества линейно независимых столбцов, равна $r' < r$. Показать, что число кодовых слов в коде равно $2^{N-r'}$ и что число строк в таблице декодирования равно $2^{r'}$.

6.12. Рассмотрим два кода с проверкой на четность с одной и той же длиной блока. Будем интерпретировать множество кодовых слов в каждом коде как группу сложения по модулю 2. Показать, что множество последовательностей, являющихся кодовыми словами в обоих кодах, образует подгруппу в каждой из рассмотренных выше групп. Интерпретируя последовательности в этой подгруппе как код с проверкой на четность, опишите процедуру нахождения проверочной матрицы нового кода через проверочные матрицы первоначальных кодов.

6.13. Показать, что число элементов любой конечной группы, в которой каждый элемент является обратным для самого себя, равно 2^n, где n — некоторое целое число, и что эта группа изоморфна группе, образуемой последовательностями из n двоичных символов с операцией сложения по модулю 2. (Две группы называются изоморфными, если существует взаимнооднозначное соответствие между их элементами, которое сохраняет групповую операцию.)

Указание: сначала покажите методом доказательства от противного, что эта группа абелева. Затем рассмотрите подгруппу из двух элементов. Далее покажите, что элементы любой подгруппы и единственный смежный класс этой подгруппы образуют новую подгруппу с удвоенным числом элементов.

6.14. Пусть 1, a, a^2, ..., a^5 обозначают элементы циклической группы порядка 6. Найти порядок каждого элемента группы и выписать все подгруппы этой группы.

6.15. (а) Написать таблицы сложения и умножения для поля целых чисел 0, 1, 2, 3, 4, в котором сложение и умножение производится по модулю 5.

(б) Доказать теорему Ферма: для любого простого числа p и любого целого числа a, не делящегося на p, остаток от деления a^{p-1} на p по модулю p равен единице, т. е. $R_p(a^{p-1}) = 1$.

Указание: рассмотрите поле элементов по модулю p и определите элемент поля **a** как $R_p(a)$; исследуйте мультипликативный порядок **a**.

6.16. Пусть $D^4 + D^2 + D + 1$ — многочлен над $GF(2)$. Выразить его как произведение двух неприводимых нормированных сомножителей. Какими соображениями вы руководствовались при решении этой задачи?

6.17. Рассмотрим поле, элементы которого являются многочленами степени 1 или меньше с коэффициентами, принадлежащими $GF(3)$; умножение элементов поля определяется как умножение многочленов по модулю $D^2 + 1$.

(а) Доказать, что $D^2 + 1$ — неприводимый многочлен над $GF(3)$.

(б) Написать таблицы сложения и умножения для этого поля.

6.18. Рассмотрим код с троичными символами, в котором каждое кодовое слово $\mathbf{x} = (x_1, x_2, x_3, x_4)$ порождается троичной информационной последовательностью $u = (u_1, u_2)$ согласно правилам

$$x_1 = u_1, \qquad x_3 = u_1 \oplus u_2,$$

$$x_2 = u_2, \qquad x_4 = u_1 \oplus 2u_2,$$

где $\oplus$ означает сложение по модулю 3.

(а) Найти порождающую и проверочную матрицы для этого кода.

(б) Составить таблицу декодирования (т. е. таблицу соответствия синдромов шумовым последовательностям) таким образом, чтобы минимизировать вероятность неправильного декодирования. При этом предполагается, что передача производится в троичном симметричном канале с $P(j|k) = \varepsilon < 1/3$ при $j \neq k$ и $P(j|k) = 1 - 2\varepsilon$ при $j = k$.

(в) Показать, как для любого числа $m \geqslant 2$ проверочных символов построить проверочную матрицу линейного кода с троичными символами, чтобы таблица декодирования содержала нулевую последовательность, все шумовые последовательности с одной ошибкой и не содержала бы ни одной последовательности более чем с одной ошибкой. Выразить длину блока для этих кодов как функцию m.

(г) Повторить пункт (в) для линейных кодов с символами из произвольного поля $GF(q)$.

6.19. Пусть приведенная ниже матрица 5×8 является *порождающей* матрицей для двоичного кода с проверкой на четность, у которого число информационных символов равно 5, а число проверочных символов равно 3,

$$\begin{bmatrix} 1&0&0&0&0&1&1&1 \\ 0&1&0&0&0&1&0&0 \\ 0&0&1&0&0&0&1&0 \\ 0&0&0&1&0&0&0&1 \\ 0&0&0&0&1&1&1&1 \end{bmatrix}.$$

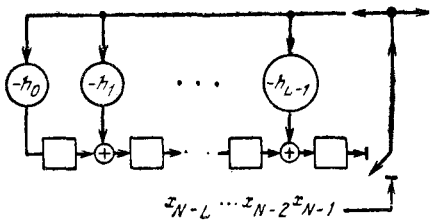
Показать, что этот код циклический и найти порождающий и проверочный многочлены.

6.20. Рассмотреть циклический код длины 7 с порождающим многочленом $D^3 + D + 1$. Представить схемы двух кодирующих устройств, одно из которых включает трехразрядный регистр сдвига, а другое — четырехзарядный регистр сдвига. Показать, что этот код способен исправлять все единичные ошибки и потому эквивалентен коду Хэмминга длины 7, исправляющему единичные ошибки.

6.21. Пусть $g(D) = g_m D^m + \ldots + g_0$ — порождающий многочлен циклического кода. Показать, что $g_0 \neq 0$.

6.22. Показать, что представленное на рисунке устройство является циклическим (N, L)-кодером с проверочным многочленом $h(D)$.

Ключ замыкается вертикально для L информационных символов $x_{N-1}, \ldots, x_{N-L}$, затем замыкается горизонтально. Регистр первоначально пуст.



Указание: найдите, что будет храниться в крайнем правом разряде регистра после того, как L информационных символов сойдут с регистра и поступят в линию.

Замечание: это устройство имеет два практических преимущества перед устройством на рис. 6.5.4: во-первых, сложение по модулю 2 может производиться непосредственно в разрядах регистра и, во-вторых, отпадает необходимость в последовательном сложении по модулю 2, что важно при передаче с высокой скоростью.

6.23. Предположим, что двоичный циклический (N, L)-код используется в канале с пакетами стираний. т. е. при передаче кодового слова $x_{N-1}, \ldots, x_1, x_0$ принимается последовательность вида $\mathbf{y} = x_{N-1}, \ldots, x_l, e, e, \ldots, e, x_j, \ldots, x_2, x_0$. Иными словами, правильно принимаются все символы, кроме серии, состоящей из нескольких стертых символов.

(а) Показать, что если число стертых символов не превышает $N - L$, то всегда можно выполнить правильное декодирование.

(б) Показать, что если стерто более чем $N - L$ символов, то декодер максимального правдоподобия всегда будет производить решение в условиях неопределенности.

(в) Начертить блок-схему простейшего декодера, который может исправлять все конфигурации $N - L$ или меньшего числа последовательных стираний (предполагается, что $L > N/2$).

6.24. Расстоянием (Хэмминга) между двумя последовательностями N символов, принадлежащих $GF(q)$, называется число позиций, в которых эти последовательности отличаются, а весом последовательности называется число ненулевых символов последовательности. Наименьшим расстоянием d_{min} линейного кода с символами из $GF(q)$ называется минимум расстояний между всеми парами кодовых слов.

(а) Показать, что d_{min} равно наименьшему весу ненулевых кодовых слов. Показать, что можно исправить все конфигурации менее чем $d_{min}/2$ ошибок (см. задачу 6.4).

(б) Показать, что для линейных кодов с символами из $GF(q)$ справедливы приведенные ниже обобщения оценок d_{min}, которые были получены в задачах 6.6 — 6.8.

Указание: используйте те же рассуждения, что и в задачах 6.6—6.8.

I. Граница Хэмминга $\sum_{j=0}^{\left\lfloor \frac{d_{min}-1}{2} \right\rfloor} (q-1)^j \binom{N}{j} \leqslant q^{N-L}$.

II. Граница Плоткина $d_{min} \leqslant [N(q-1)/q](q^L/(q^L-1))$.

Для любого j, $1 \leqslant j \leqslant L$ $d_{min} \leqslant \frac{(N-L+j)(q-1)}{q} \quad \frac{qj}{qj-1}$.

Для $N \geqslant (qd_{min}-1)(q-1)$. Пусть $j = \lfloor \log_q d_{min} \rfloor + 1$.

Тогда $N-L \geqslant \frac{qd_{min}-1}{q-1} - 1 - \log_q d_{min}$.

III. Граница Варшамова — Гилберта. Существует линейный (N, L)-код такой, что

$$\sum_{i=0}^{d_{min}-2} \binom{N}{i} (q-1)^i \geqslant q^{N-L}.$$

6.25. Проверить справедливость (6.6.1) для поля многочленов над $GF(2)$ по модулю многочлена $D^2 + D + 1$.

Указание: чтобы избежать путаницы, представьте элементы поля как многочлены по t.

6.26. Пусть α — примитивный элемент $GF(q)$. Показать, что мультипликативный порядок α^i равен $(q-1)/[\text{НОД}(i, q-1)]$ или, это эквивалентно, НОК $(i, q-1)/i$*). Показать, что если $q-1$ делится на n, то существует n элементов поля, мультипликативный порядок которых равен n или делителю n, и что эти элементы образуют циклическую группу по умножению.

6.27. Предположим, что данный элемент α из $GF(p^n)$ имеет минимальный многочлен [над $GF(p)$] степени $m < n$.

(а) Показать, что в минимальном подполе $GF(p^n)$, содержащем α, имеется p^m элементов. Выразить все элементы этого подполя через α и целые элементы поля.

Указание: просмотрите доказательство теоремы 6.6.4.

(б) Показать, что n должно делиться на m.

(в) Показать, что подполе, определенное в пункте (а), является единственным подполем $GF(p^n)$, содержащим p^m элементов.

Указание: воспользуйтесь тем, что $D^{p^m-1} - 1$ имеет в $GF(p^n)$ лишь $p^m - 1$ корней.

(г) Предположим, что элемент β из $GF(p^n)$ имеет мультипликативный порядок j. Показать, что степень i минимального многочлена [над $GF(p)$] элемента β такова, что $p^i - 1$ делится на j.

6.28. Рассмотреть поле $GF(2^4)$ многочленов по модулю $D^4 + D + 1$. Найти многочлены этого поля, принадлежащие подполю $GF(2^2)$.

6.29. Рассмотреть поле $GF(2^4)$ многочленов по модулю $D^4 + D + 1$.

*) НОД — наибольший общий делитель, НОК — наименьшее общее кратное.

(а) Найти минимальный многочлен $f_\alpha(D)$ элемента $\alpha = t^3 + 1$, используя то обстоятельство, что α, α^2, α^4 и α^8 — множество всех корней $f_\alpha(D)$ и, следовательно, $f_\alpha(D) = (D - \alpha)(D - \alpha^2)(D - \alpha^4)(D - \alpha^8)$.

(б) Повторить пункт (а) путем решения уравнения $f_\alpha(\alpha) = 0$. Точнее, положив $f_\alpha(D) = f_0 + f_1 D + \ldots + f_4 D^4$, разрешить уравнение $f_0 + f_1(t^3 + 1) + f_2(t^3 + 1)^2 + \ldots + f_4(t^3 + 1)^4 = 0$. Оно может быть интерпретировано как система 4 уравнений относительно двоичных неизвестных, одно из которых включает t^3, другое t^2, третье t^1 и последнее t^0. Так как $f_\alpha(D)$ — неприводимый многочлен, то можно положить $f_0 = 1$ и провести решение относительно двоичных неизвестных f_1, f_2, f_3, f_4.

6.30. (а) Найти порождающий многочлен для исправляющего две ошибки двоичного БЧХ-кода длины 15 и для исправляющего три ошибки БЧХ-кода длины 15.

(б) Допустим, что два двоичных БЧХ-кода определяются тем, что $\alpha, \ldots, \alpha^{d-1}$ являются корнями для всех кодовых слов. Но для одного кода α является примитивным элементом $GF(2^n)$ с минимальным многочленом $f_1(D)$, а для другого кода α является примитивным элементом $GF(2^n)$ с другим минимальным многочленом $f_2(D)$. Показать, что множество кодовых слов для одного кода может быть получено путем некоторой фиксированной перестановки символов кодовых слов другого кода.

6.31. (а) Показать, что наименьшее расстояние d_{min} линейного кода с L информационными символами и длиной блока N не должно превышать $N - L + 1$.

(б) Кодом Рида-Соломона называется БЧХ-код с $m = 1$. Показать, что все коды Рида-Соломона удовлетворяют с равенством указанной выше верхней границе для d_{min}.

(в) Предположим, что код Рида-Соломона используется в стирающем канале (т. е. каждый символ на выходе канала или совпадает с соответствующим входным символом, или является стертым символом). Показать, что если в блоке произошло менее чем d_{min} стираний, то декодирование всегда правильно, и если произошло $d_{min} + i$, $i \geqslant 0$ стираний, то существует ровно q^{i+1} кодовых слов, которые могут быть приняты в качестве декодированного слова. Здесь q — объем входного алфавита.

(г) Для кода Рида-Соломона с длиной блока N и объемом входного алфавита q найти общее число кодовых слов, у которых число ненулевых элементов точно равно d_{min}.

6.32. Рассмотрим двоичный БЧХ-код с произвольным m, примитивным элементом α, $r = 1$ и $d = 3$. Показать, что он совпадает с кодом Хэмминга в циклическом виде, длина блока которого равна $2^m - 1$. Решить (6. 7. 27) для этого случая и показать, что решение эквивалентно методу декодирования при помощи таблицы декодирования, которая ранее была получена для кодов Хэмминга.

6.33. Воспользоваться итеративным алгоритмом и найти двоичный регистр сдвига минимальной длины, который генерирует многочлен [над $GF(2)$] $1 + D^3 + D^4 + D^7$; найти двоичный регистр сдвига минимальной длины, генерирующий $D^2 + D^5 + D^6$ [над $GF(2)$].

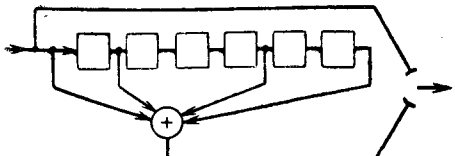
6.34. Многочлен $f(D) = D^4 + D + 1$ является примитивным над $GF(2)$. Пусть α является многочленом t в поле многочленов по модулю $f(D)$ (см. рис. 6.6.3). Рассмотреть БЧХ-код с $q = 2$, $m = 4$, с α, определенным выше, $r = 1$ и $d = 5$. Нарисовать блок-схему устройства для вычисления синдромного многочлена $S(D)$, представляя все S_i как элементы поля многочленов по модулю $f(D)$.

6.35. Показать, что $A_n(D) = [C_n(D)\, S(D)]_0^{n-1}$ при всех $n \geqslant 0$ [см. (6.7.69)].

Указание: просмотрите доказательство утверждений (в) и (г) теоремы 6.7.3.

6.36. Показать, что из (6.7.21) следует (6.7.22). Заметим, что для этого достаточно доказать, что если производная многочлена над конечным полем определяется (6.7.21) и если $f(D) = g(D)h(D)$, то $f'(D) = g'(D)h(D) + g(D)h'(D)$.

6.37. Начертить блок-схему исправляющего две ошибки порогового декодера для сверточного кодера, представленного ниже.



6.38. Начертить блок-схему порогового декодера, исправляющего две ошибки и обнаруживающего три ошибки, для систематического сверточного кодера со скоростью (в битах), равной $^1/_2$, и с проверочными символами, порождаемыми по следующему правилу:

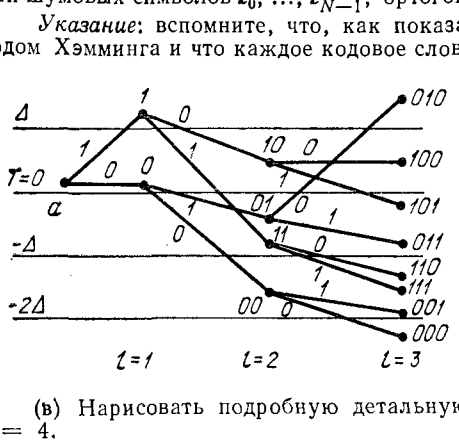
$$x_n^{(2)} = u_n + u_{n-6} + u_{n-7} + u_{n-9} + u_{n-10}.$$

Указание: постройте множество из пяти линейных комбинаций шумовых символов, ортогональных к $z_1^{(1)}$, и использовать значения этих линейных комбинаций в качестве входных символов в пороговом устройстве.

6.39. Метод порогового декодирования можно использовать для блоковых кодов, так же как и для сверточных кодов. Рассмотреть код максимальной длины с L информационными символами и длиной блока $N = 2^L - 1$.

(а) Показать, что декодер может вычислять $(N - 1)/2$ линейных комбинаций шумовых символов $z_0, \ldots, z_{N-1}$, ортогональных к z_{N-1}.

Указание: вспомните, что, как показано в § 6.6., дуальный код является кодом Хэмминга и что каждое кодовое слово в дуальном коде соответствует проверочному соотношению кода максимальной длины. Вспомните также, что в коде Хэмминга каждая последовательность (в частности, имеющая вес 2) находится на расстоянии не больше 1 от некоторого кодового слова.



(б) Показать, что если пороговый декодер рассчитан на исправление z_{N-1}, тот же пороговый декодер может использоваться для исправления всех шумовых символов. Показать, что таким образом можно исправлять комбинации $(N-3)/4$ или менее ошибок.

(в) Нарисовать подробную детальную блок-схему такого декодера при $L = 4$.

6.40. Последовательный декодер выходит из узла a с нулевым порогом и движется к узлу 0, лежащему в дереве принятых цен ниже. Записать последовательность проверяемых узлов, просмотренных декодером до первой проверки узла дерева на глубине $l = 3$. Предполагается, что при движении вперед декодер всегда выбирает ребро, отмеченное символом 0.

6.41. (а) Пусть Γ_l—цена l-го узла вдоль правильного пути в дереве принятых цен для последовательного декодера. Найти среднее значение по ансамблю кодов и шумов в канале цены Γ_l. Воспользоваться границей Чернова для оценки сверху $\Pr[\Gamma_l \leqslant iA]$ и показать, что $\Pr[\Gamma_l \leqslant iA]$ стремится к 0 экспоненциально с ростом l при любом смещении $B < C$.

(б) Пусть Γ_l' цена l-го узла в дереве принятых цен, которая соответствует некоторой информационной последовательности, отличающейся от переданной в первом символе. Применить к $\Pr[\Gamma_l' \geqslant iA]$ границу Чернова. Показать, что полученная граница отличается от границы для $\Pr[\Gamma_l < iA]$ множителем $\exp[-\nu LB - iA]$. Показать, что если $R \leqslant B < C$, то вероятность того, что какая-либо из цен Γ_l' превысит iA, убывает с ростом iA экспоненциально.

6.42. Границы вероятности ошибки и среднего числа вычислений в § 6.9 производят впечатление несколько искусственных, поскольку их вывод основы-

вался на использовании либо меняющегося во времени кода, либо кода с бесконечным кодовым ограничением. В этой задаче изменим эти нереальные предположения на новые. Точнее, будет использоваться ансамбль кодов, введенный перед леммой 6.9.1, с конечной длиной кодового ограничения, равной L подблокам. Изменим также алгоритм декодирования следующим образом. При любом $l (l > L)$, как только декодер произведет первую F-проверку узла на глубине l в дереве принятых цен, он окончательно принимает гипотезу о символах источника в $(l - L + 1)$-м подблоке и полагает $\Gamma_{l-L} = -\infty$. Другими словами, декодер не может менять гипотезы о символах, от которых узел максимального проникновения декодера в дерево удален более, чем на длину кодового ограничения. Будем говорить, что в n-м узле произошла ошибка, если декодер когда-либо совершит F-проверку на глубине $n + L$ в узле какого-либо неправильного пути, впервые ответвляющегося от правильного в n-м узле.

(а) Показать, что среднее значение $\overline{W}_n$ ограничено сверху правой частью неравенства (6.9.23)*).

Указание: сначала покажите, что передаваемые символы, соответствующие двум последовательностям источника, отличающимися в $(n + 1)$-м подблоке, статистически независимы в промежутке от $(n + 1)$-го до $(n + L)$-го подблоков. Затем следуйте процедуре, рассмотренной при выводе границы (6.9.23).

(б) Показать, что вероятность ошибки в n-м узле при $B \leqslant E_0(1, \mathbf{Q})$ удовлетворяет неравенству

$$P_{e,\,n} \leqslant (L+1)\, e^{\Delta/2} \left\{ -\nu L \left[\frac{E_0(1, \mathbf{Q}) + B}{2} - R \right] \right\}.$$

Сравнить этот результат с результатом (6.9.40).

6.43. (а) Показать, что γ_i и γ_i, определяемые в (6Б.2) и (6Б.4), статистически независимы, если матрица переходных вероятностей такова, что каждая строка является перестановкой любой другой строки и каждый столбец является перестановкой любого другого столбца и если входные символы равновероятны (этот класс каналов является подклассом симметричных каналов, рассмотренных в § 4.5, и включает в себя ДСК).

(б) Показать, что для этого класса каналов результат леммы 6.9.3 может быть усилен следующим образом:

$$\Pr\left[\Gamma'_l \geqslant \Gamma_{min} + (i-2)\,\Delta\right] \leqslant \exp\left[-\frac{(i-2)}{2}\Delta - \nu l \frac{E_0(1, \mathbf{Q}) + B}{2} \right].$$

Указание: воспользоваться вашим результатом из пункта (а) для доказательства того, что Γ'_l и Γ_{min} статистически независимы. Затем соответствующим образом измененные рассуждения, использованные при выводе соотношений (6Б.14) — (6Б.22), примените не к $\min \Gamma_{n,\,l}$, а к Γ_{min}.

(в) Воспользуйтесь полученным выше результатом для получения более точной границы для $\overline{W}_n$, аналогичной (6.9.30), и границы для $P_{e,\,n}$, аналогичной (6.9.40) — (6.9.42).

6.44. Снова рассмотрите двоичный симметричный канал с $Q(0) = Q(1) = 1/2$ так, что Γ'_l и Γ_{min} статистически независимы. Пусть $\overline{W_0(u)}$ — среднее число проверок «вперед», выполненных декодером, при условии, что $\Gamma_{min} = u$. Предположим, что смещение B равно скорости R и что $R > E_0(1, \mathbf{Q})$. Показать, что

$$\overline{W_0(u)} \leqslant A \exp\left(\frac{-u}{1+\rho_r} \right),$$

где ρ_r — решение уравнения $\rho_r R = E_0(\rho_r)$ и где A — не более чем линейно убывающая функция u.

Указание: покажите, что

$$\Pr\left[\Gamma'_l \geqslant u + (i-2)\,\Delta\right] \leqslant \exp\left\{ -\frac{1}{1+\rho} \left[u + (i-2)\,\Delta + Bl\nu + E_0(\rho)\, l\nu \right] \right\}$$

*) При оценке числа операций автор предполагает, что ошибка декодирования отсутствует. *(Прим. ред.)*

при $\rho \geqslant 0$. При суммировании по l выберите $\rho = \rho_r$ при малых значениях l и малое значение ρ при больших l.

6.45. Рассмотрим случайное блуждание $S_n = \sum_{i=1}^{n} z_i$, где случайные величины z_i одинаково распределены и $\overline{z_i} < 0$.

Пусть $u < 0$ фиксировано и пусть N (случайная величина) равна минимальному значению n, для которых $S_n \leqslant u$ и пусть S_N равно значению S_n при этом n. Взяв производную от тождества Вальда (6Б. 29) по r при $r = 0$, показать, что $\overline{N} = \overline{S_N}/\overline{z}$. Пусть z_{min} равно минимальному значению, принимаемому случайной величиной z_i; показать, что $u + z_{min} < \overline{N}\overline{z} \leqslant u$.

6.46. Рассмотрим двоичный код с произвольно большой длиной блока и двумя кодовыми словами. Показать, что исправление всех пакетов длины g при защитном интервале g является невозможным.

Указание: рассмотрите два типа шумовых последовательностей, приведенных на рис. 6.10.2, и покажите, что $\mathbf{z}_1$ и $\mathbf{z}_2$ можно выбрать таким образом, что $\mathbf{x}_1 \oplus \mathbf{z}_1 = \mathbf{x}_2 \oplus \mathbf{z}_2$.

Глава 7.

7.1. Вход канала x образован числами $+1$ и -1, используемыми с вероятностями $P_X(+1) = P_X(-1) = 1/2$. Выход y представляет собой сумму x и независимой шумовой случайной величины z с плотностью вероятности $p_Z(z) = 1/4$ при $-2 < z \leqslant 2$ и $p_Z(z) = 0$ при других значениях z. Другими словами, условная плотность вероятности y при заданном x определяется равенствами $p_{Y|X}(y|x) = 1/4$ при $-2 < y - x \leqslant 2$ и $p_{Y|X}(y|x) = 0$ при других значениях x, y.

(а) Найти плотность вероятности на выходе канала и изобразить ее на графике.

(б) Найти $I(X; Y)$.

(в) Предположим, что выход преобразуется в новую дискретную случайную величину z, определяемую равенствами $z = 1$ при $y > 1$; $z = 0$ при $-1 < y \leqslant 1$; $z = -1$ при $y \leqslant -1$. Найти $I(X; Z)$ и дать наглядное толкование результата.

(г) Пусть X, Y — дискретные ансамбли и Z — новый ансамбль с элементами, определенными по элементам Y-ансамбля, $z = z(y)$. Показать, что если $P[x|z(y)] = P(x|y)$, то $I(X; Z) = I(X; Y)$.

7.2. Вход дискретного по времени канала без памяти образован числами $+1$ и -1, а выход, принимающий действительные значения, определяется переходной плотностью вероятности

$$p(y \mid 1) = \begin{cases} \dfrac{1}{a+b} \exp(-y/a); & y > 0, \\ \dfrac{1}{a+b} \exp(y/b); & y \leqslant 0, \end{cases}$$

и $p(-y|-1) = p(y|1)$. Постоянные a и b произвольны, $a > b > 0$. (Оказывается, что эта плотность вероятности возникает при рассмотрении релеевского канала с замираниями, если входы представляют собой ортогональные функции с равной энергией и выход равен логарифму отношения правдоподобия для выходного сигнала. См. задачу 8.21.)

(а) Найти выражение для пропускной способности канала и вычислить ее предельные значения при $b/a \to 1$ и $b/a \to 0$.

(б) Вычислить $E_0(1)$*).

*) Здесь $E_0(\rho) = \max_{\mathbf{Q}} E_0(\rho, \mathbf{Q})$.

(в) Показать, что декодирование по максимуму правдоподобия сводится к выбору m, которое минимизирует выражение

$$\sum_{n\,:\,x_{m,\,n}\,y_n<0} |y_n|.$$

(г) Вычислить вероятность ошибки при отсутствии кодирования и при использовании приемника максимального правдоподобия.

7.3. Интервал (0,1) является входным алфавитом дискретного по времени канала без памяти, а интервал (0,1) и символ стирания E — выходным алфавитом. Для каждого входа x, $0 \leqslant x \leqslant 1$, с вероятностью $1/2$ выход y принимает значение x и с вероятностью $1/2$ выход равен символу E.

(а) Найти пропускную способность этого канала, функцию $E_0(\rho)$ и экспоненту случайного кодирования $E_r(R)$. Отметьте, в частности, что $E_0(\rho)$ разрывна при $\rho = 0$.

Указание: рассмотрите случай, когда используется только конечное множество входных букв и перейдите к пределу, когда объем этого множества стремится к бесконечности.

(б) Показать, что для M равновероятных сообщений полученная оценка точно в $M/(M-1)$ раз больше действительно минимальной достижимой вероятности ошибочного декодирования.

7.4. Алфавитами на входе и выходе дискретного по времени канала без памяти является множество фазовых углов, $0 \leqslant x < 2\pi$. В канале действует аддитивный фазовый шум z, где z не зависит от входа x и имеет плотность вероятности $p_Z(z)$, которая отлична от нуля только для $0 \leqslant z < 2\pi$. Выход канала y равен сумме $x + z$ по модулю 2π. Например, если $x = 7\pi/4$ и $z = 3\pi/2$, то $y = x + z - 2\pi = 5\pi/4$.

(а) Показать, что C и $E_0(\rho)$ достигаются на входной плотности вероятности $p_X(x) = 1/(2\pi)$, $0 \leqslant x < 2\pi$.

(б) Найти C, $E_0(\rho)$ и $E_r(R)$ в следующих двух случаях:

1) $p_Z(z) = 1/\alpha$; $0 \leqslant z < \alpha$, и $p_Z(z) = 0$ при других значениях z.

2) $$p_Z(z) = \frac{\alpha\, e^{-\alpha z}}{1 - e^{-2\pi\alpha}}; \quad 0 \leqslant z < 2\pi.$$

7.5. Дискретный по времени канал без памяти имеет вход x, ограниченный интервалом $(-A, A)$, и аддитивный шум z с плотностью вероятности $p_Z(z) = 1/2$ при $-1 < z \leqslant 1$ и $p_Z(z) = 0$ при других значениях z.

(а) Найти при $A = 1/2$ пропускную способность канала и входное распределение, приводящее к ней. Показать, что по множеству входов, которые приводят к пропускной способности, канал эквивалентен двоичному стирающему каналу. Найти экспоненту случайного кодирования $E_r(R)$ для этого канала и проверить, что она такая же, как и для ДСтК.

Указание: угадайте входное распределение и проверьте, что оно приводит к пропускной способности; это наиболее легкий путь получить здесь нужный результат.

(б) Показать, что для произвольного нецелого A средняя взаимная информация и $E_0(\rho, \mathbf{Q})$ максимизируются на дискретном распределении, которое для каждого целого i, $0 \leqslant i < n$ задается соотношением

$$Q(A-2i) = Q(-A+2i) = \frac{n-i}{n(n+1)},$$

где $n = \lceil A \rceil$.

Изобразить на графике плотность вероятности выхода при таком распределении на входе и найти C и $E_r(R)$ (в параметрическом виде).

(в) Найти максимизирующее распределение для целых A и интерпретировать его как предел для случая нецелых A.

7.6. N независимых дискретных по времени каналов с аддитивным гауссовым шумом соединены параллельно. Дисперсия шума n-го канала задается равенством $\sigma_n^2 = n^2$ для каждого n. Энергия на входе ограничена условием

$$\sum_{n=1}^{N} \mathscr{E}_n/n \leqslant 5.$$

(а) Найти пропускную способность параллельного соединения и найти значения $\mathscr{E}_n$, при которых достигается пропускная способность для случаев $N = 2$, $N = 4$, $N = \infty$.

Указание: сначала изменением масштаба сигнала и шума в каждом канале сведите задачу к той, которая решена в § 7.5.

(б) Для случая $N = \infty$ найти критическую скорость R_{cr} и найти $E_r(R_{cr})$ и $E_r(0)$.

(в) Изменить ограничение, приняв его в виде

$$\sum_{n=1}^{N} \mathscr{E}_n/n \leqslant 50,$$

и найти новые значения $\mathscr{E}_n$, при которых достигается пропускная способность для $N = \infty$.

7.7. Рассматривается множество N параллельных дискретных по времени каналов с гауссовым шумом и дисперсиями шума $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$. Предположим, что выбираются два слова для разового использования параллельного множества каналов

$$x_1=(\sqrt{\mathscr{E}_1}, \sqrt{\mathscr{E}_2}, \ldots, \sqrt{\mathscr{E}_N}) \text{ и } x_2=(-\sqrt{\mathscr{E}_1}, -\sqrt{\mathscr{E}_2}, \ldots, -\sqrt{\mathscr{E}_N}).$$

(а) Найти точно выражение для вероятности ошибки при использовании декодирования по максимуму правдоподобия. Ответ будет «хвостом» соответствующего гауссовского распределения.

(б) При ограничении

$$\sum_{n=1}^{N} \mathscr{E}_n \leqslant \mathscr{E},$$

найти значения $\mathscr{E}_n$, $1 \leqslant n \leqslant N$, которые минимизируют вероятность ошибки.

(в) Сравните ответ с показателем экспоненты при нулевой скорости в границе случайного кодирования для процедуры с выбрасыванием.

Глава 8.

8.1. (а) Пусть $z(t)$ — гауссовский случайный процесс с нулевым средним и пусть дисперсия случайной величины $\int x(t)\, z(t)dt$ конечна для всех $x(t)$ из L_2. Показать, что это означает существование некоторого числа M такого, что дисперсия $\int x(t)z(t)dt$ меньше или равна M для всех нормированных функций $x(t)$.

Предполагается, что для любой последовательности $\{x_i(t)\}$, для которой существуют l. i. m. $x_i(t)$ и l.i.m. $\int x_i(t)z(t)dt$, удовлетворяется условие l. i. m. $\int x_i(t)\, z\ (t)\ dt = \int[\text{l. i. m. } x_i(t)]\, z(t)dt$.

Указание: см. § 117 работы Рисса и Надь (1955).

(б) Используйте полученный выше результат в случае, когда $z(t)$ определена как и выше; $\{\varphi_i(t)\}$ — полное множество ортонормальных функций;

$$x(t) = \sum x_i \varphi_i(t)$$

—функция из L_2 и $z_i = \int \varphi_i(t)\, z(t)\, dt$, чтобы показать, что

$$\lim_{k\to\infty} \overline{\left[\int x(t)\, z(t)\, dt - \sum_{i=1}^{k} x_i z_i\right]^2} = 0.$$

8.2. По определению, гауссовский случайный процесс $z(t)$ стационарный, если для всех функций $x(t)$ из L_2 и для всех τ дисперсия $\int x(t)z(t)\, dt$ равна дисперсии $\int x(t+\tau)z(t)dt$. По определению, спектральная плотность стационарного гауссовского случайного процесса, если она существует, задается выражением

$$S(f) = \lim_{k\to\infty} \overline{\left[\int_0^{k/f} \sqrt{2f/k}\, (\cos 2\pi f t)\, z(t)\, dt\right]^2}.$$

Показать, что $S(f)$ меньше или равна M, определенному в задаче 8.1.

8.3. Пусть $z(t)$ — случайный процесс с нулевым средним и корреляционной функцией $\mathcal{R}(t, \tau)$. Показать, что $\mathcal{R}(t, \tau)$ непрерывна тогда и только тогда, когда

$$\lim_{\varepsilon\to 0} \overline{[z(t) - z(t+\varepsilon)]^2} = 0 \text{ для всех } t.$$

Указание: для того чтобы установить это, примените неравенство Шварца к $\mathcal{R}(t, \tau) - \mathcal{R}(t+\varepsilon, \tau)$.

8.4. Пусть $x(t)$ и $X(f)$ — пара преобразований Фурье,

$$X(f) = \int x(t)\, e^{-j2\pi f t}\, dt, \quad x(t) = \int X(f)\, e^{j2\pi f t}\, df.$$

Показать, что если $X(f)$ абсолютно интегрируема (т. е. $\int_{-\infty}^{\infty} |X(f)|\, df < \infty$), то $x(t)$ непрерывна.

Указание: замечая что

$$x(t) - x(t-\varepsilon) = \int X(f)\, [e^{j2\pi f t}]\, [1 - e^{-j2\pi f \varepsilon}]\, df,$$

оцените интеграл отдельно для больших f и для малых f и покажите, что его значение стремится к 0 при ε, стремящемся к 0.

8.5. Пусть $z(t)$ — определенный в § 8.1 гауссовский случайный процесс с нулевым средним; пусть $h(t)$ — функция из L_2 и пусть $y(t)$ случайный процесс, определяемый равенством $y(t) = \int h(t-\tau)z(\tau)\, d\tau$. Показать, что корреляционная функция $\mathcal{R}_y(t, \tau)$ процесса $y(t)$ непрерывна.

Указание: пусть M — верхняя граница для дисперсии $\int x(t)z(t)dt$ при любых нормированных $x(t)$ (см. задачу 8.1.), покажите, что

$$\overline{[y(t) - y(t+\varepsilon)]^2} \leqslant M \int [h(t) - h(t+\varepsilon)]^2\, dt.$$

Далее, определяя $w(\varepsilon) = \int h(t)h(t+\varepsilon)dt$ и используя задачу 8.4, покажите, что $w(\varepsilon)$ непрерывна. На основе этого покажите, что

$$\lim_{\varepsilon\to\infty} \int [h(t) - h(t-\varepsilon)]^2\, dt = 0.$$

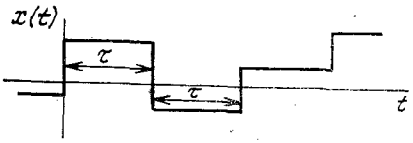
Наконец, используйте задачу 8.3.

8.6. Пусть $z(t)$ — стационарный гауссовский случайный процесс с нулевым средним и корреляционной функцией $\mathcal{R}(\tau)$. Пусть $y(t)$ — другой случайный процесс, статистически независимый от $z(t)$, для которого при каждом t случайная величина $y(t)$ равномерно распределена между -1 и $+1$. Пусть $y(t)$ для каждого t статистически не зависит от $y(t)$ при всех других значениях t. Показать, что в соответствии с определением, данным в § 8.1, $z(t) + y(t)$ является гауссовским случайным процессом, однако для каждого t, $z(t) + y(t)$ — негауссовская случай-

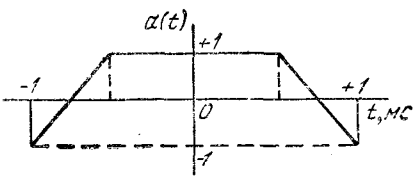
ная величина. (Заметим, что с физической точки зрения случайный процесс $y(t)$ никогда не может наблюдаться, так как любое измерительное устройство должно производить некоторое усреднение. Однако с математической точки зрения эта задача указывает на характер патологических эффектов, которые возникают при рассмотрении случайных процессов, не имеющих непрерывную корреляционную функцию.)

8.7. Канал с белым аддитивным гауссовым шумом имеет ограничение на мощность входа S и спектральную плотность шума $N_0/2$. Предположим, что при некотором заданном τ вход канала является постоянным внутри каждого интервала времени длины τ и изменяется только в моменты времени, отделенные интервалом τ.

(а) Найти соответствующее ортонормальное разложение для входа и представить канал в виде дискретного по времени канала. Найти пропускную способность канала в натах в секунду (при указанном выше ограничении) и найти предел пропускной способности при $\tau \to 0$.



К задаче 8.7



К задаче 8.9

(б) Ввести теперь дополнительное ограничение, состоящее в том, что вход может принимать только значения $\pm\sqrt{S}$, и найти пропускную способность при $\tau \to 0$.

(в) Теперь предположите, что при ограничении на вход, указанном в пункте (б), приемник может только вычислять интеграл от выхода в течение каждого интервала τ секунд и запоминать только знак интеграла. Другими словами, канал сводится к ДСК при передаче одного символа канала каждые τ секунд. Найти пропускную способность (в нат/с) при $\tau \to 0$. (Заметим, что все эти каналы являются «каналами с большим шумом» в пределе при $\tau \to 0$, и что вся кривая показателя экспоненты в зависимости от скорости определяется пропускной способностью.)

8.8. В канале белый гауссовый шум имеет спектральную плотность $N_0/2$. Передаваемый сигнал ограничен тем, что должен иметь среднюю мощность S и для каждого целого i при t между i и $i+1$ секундами должен иметь следующий вид:

$$\sum_{m=M_1}^{M_2} x_m(i) \cos[2\pi \cdot 100mt + \varphi_m(i)]; \quad i \leqslant t < i+1.$$

Функции $x_m(i)$ и $\varphi_m(i)$ произвольны, за исключением ограничения на мощность, и они не должны изменяться в пределах интервала продолжительностью в одну секунду.

(а) Найти пропускную способность канала (в нат/с) для произвольных целых $M_2 > M_1 > 0$.

(б) Установить взаимосвязь между ответом в пункте (а) и пропускной способностью канала с белым гауссовым шумом и с ограничением на мощность и полосу частот на входе.

(в) Найти предел в ответе к пункту (а) при $M_2 \to \infty$ и M_1 фиксированном.

(г) Какой смысл пропускной способности, найденной в пункте (а)? Объяснить, что понимается под блоковым кодом для этого канала при этом множестве ограничений.

8.9. Рассмотрим систему связи, в которой передаваемая функция $x(t)$ образована смещениями во времени изображенной на рисунке базисной функции

$$x(t) = \sum_n x_n a(t - nT), \quad T = 2 \text{ мкс}, \tag{1}$$

где x_n — произвольные случайные величины, средняя передаваемая мощность которых ограничена сверху значением S. Принятая функция равна $x(t)$, сложенной с аддитивным белым гауссовым шумом, имеющим спектральную плотность $N_0/2$.

(а) Найти пропускную способность канала в натуральных единицах в секунду при ограничении на мощность и условии (1).

(б) Найти то же, что и в пункте (а), если время повторения T в (1) равно 1 мкс; следовательно, допуская перекрытие базисных функций.

8.10. Непрерывный по времени канал имеет аддитивный белый гауссов шум со спектральной плотностью $N_0/2$. Вход представляет собой последовательность синусоидальных импульсов длительности T, имеющих вид

$$\sqrt{2S} \sin\left(\frac{2\pi k}{T} t + \varphi\right); \quad 0 < t \leqslant T,$$

где φ может принимать значения 45°, 135°, —135°, —45°. Принятый на интервале T сигнал декодируется по максимуму правдоподобия в один из четырех сигналов.

(а) Найти соответствующее ортонормальное разложение для этих четырех функций и представить указанные четыре функции коэффициентами в этих разложениях.

(б) Представить принятые функции с помощью того же самого разложения и на графике, используя эти коэффициенты как оси, показать, какие принятые сигналы должны быть декодированы в каждый из переданных сигналов.

(в) Показать, что операция декодирования может быть разбита на пару решений и что это разбиение соответствует преобразованию дискретного канала с 4 входами и 4 выходами в пару параллельных независимых двоичных симметричных каналов; найти переходные вероятности для этих двоичных симметричных каналов.

(г) Найти выражение для пропускной способности в битах в секунду этого соединения каналов и найти предел при $T \to 0$.

8.11. Пусть L статистически независимых каналов с белым гауссовым шумом соединяют передатчик, расположенный в точке A, с приемником, расположенным в точке B. Спектральная плотность мощности шума l-го канала равна $N_0(l)/2$. Передатчик ограничен по мощности значением SL. Найти пропускную способность этой совокупности каналов при каждом из следующих условий.

(а) Кодер может послать произвольные сигналы по каждому из L каналов и может произвольно распределить мощность между каналами. Приемник принимает каждый из L сигналов раздельно и может обрабатывать их любым образом.

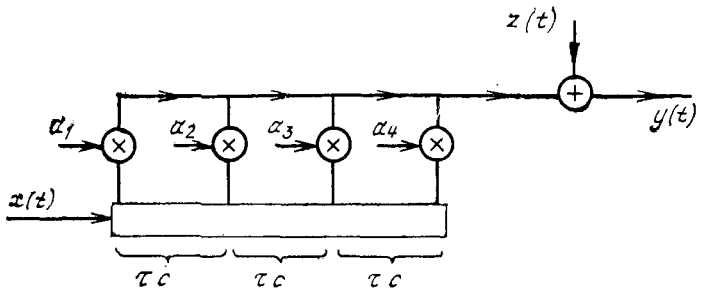
(б) Передатчик должен посылать один и тот же сигнал по каждому каналу, т. е. $s_1(t) = s_2(t) = \ldots = s_L(t)$ и $s_l(t)$ имеет мощность S при $1 \leqslant l \leqslant L$. Приемник принимает все сигналы раздельно и может обрабатывать их любым образом.

(в) Передатчик посылает один и тот же сигнал по каждому каналу и зашумленные сигналы складываются вместе перед приемом.

Указание: в пункте (б) представьте вход и каждый выход в виде ортонормальных разложений и покажите, что при рассмотрении соответствующей линейной комбинации выходов не происходит потери информации.

8.12. Предположим, что одно из двух равновероятных сообщений длительности T должно быть послано по каждому из трех каналов предыдущей задачи. Произвести оптимальный выбор этих двух сигналов в каждом из случаев (при ограничении на мощность) и найти выражение для получающейся вероятности ошибки.

8.13. На рисунке изображена модель простого диспергирующего во времени канала связи. Шум $z(t)$ аддитивный, белый и гауссов со спектральной плотностью $N_0/2$. Усиления известны и являются положительными постоянными.



$$y(t) = z(t) + \sum_{i=1}^{4} a_i x(t + \tau - i\tau)$$

(а) Пусть задано множество кодовых слов $x_1(t), \ldots, x_M(t)$. Найти и изобразить блок-схему декодера по максимуму правдоподобия. Предполагается, что значения $a_1, \ldots, a_4$ известны.

(б) Пусть *средняя* мощность переданного сигнала ограничена значением S и пусть полоса частот неограничена. Найти пропускную способность канала в натах в секунду.

Указание: найдите отношение мощности выходного сигнала к входной мощности для синусоиды с большой длительностью и частотой, кратной $1/\tau$.

(в) Найти и изобразить на графике наибольшее достижимое $E(R)$, такое, что $P_e \leqslant \exp[-TE(R)]$ может быть достигнута в пределе при больших T (здесь R — скорость передачи в натах в секунду).

8.14. Биортогональный код, состоящий из M кодовых слов (M четно), определяется как код, для которого первые $M/2$ кодовых слов являются ортогональными словами равной энергии, а последние $M/2$ кодовых слов противоположны по знаку первым $M/2$, т. е.

$$x_m(t) = \sqrt{\mathscr{E}}\,\varphi_m(t); \quad 1 \leqslant m \leqslant M/2,$$
$$x_m(t) = -\sqrt{\mathscr{E}}\,\varphi_{m-M/2}(t); \quad M/2+1 \leqslant m \leqslant M.$$

(а) Показать, что верхняя граница (8.2.43) и (8.2.44) вероятности ошибки для ортогонального кода при передаче по каналу с белым гауссовым шумом со спектральной плотностью $N_0/2$ применима также к биортогональному коду.

(б) Рассмотрим декодер, который для каждого m, $1 \leqslant m \leqslant M/2$, сравнивает $y_m = \int y(t)\,\varphi_m(t)\,dt$ с фиксированным порогом A. Когда $|y_m| \geqslant A$ только для одного m, то результатом декодирования является m, если $y_m \geqslant A$, и $m + M/2$, если $y_m \leqslant -A$. В других случаях декодер отказывается от декодирования. Пусть P_a — вероятность того, что декодер отказывается от декодирования и пусть P_e — вероятность ошибочного декодирования. Показать, что $P_a \leqslant P_1 + P_2$ и $P_e \leqslant P_1 P_2$, где

$$P_1 = \int_{-\infty}^{A} \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(y-\sqrt{\mathscr{E}})^2}{N_0}\right] dy,$$

$$P_2 = M \int_{A}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{y^2}{N_0}\right] dy.$$

(в) Найти значение A, которое минимизирует приведенную выше границу P_a, и выразить получающиеся границы P_a и P_e через показатель экспоненты, зависящий от скорости, так же как в (8.2.43) и (8.2.44).

(г) Пусть $A = \sqrt{\mathscr{E}} - \varepsilon$, где ε мало. Вновь выразить P_a и P_e через показатель экспоненты, зависящий от скорости. Изобразите на графике показатели экспонент в зависимости от скорости для пунктов (а), (в), (г).

8.15. (а) Определим расстояние между функциями $x(t)$ и $y(t)$ выражением

$$\sqrt{\int [x(t)-y(t)]^2\,dt}.$$

Показать, что расстояние между любой парой кодовых слов симплексного кода совпадает с расстоянием между любой другой парой кодовых слов.

Показать, что это расстояние равно квадратному корню верхней границы среднего квадрата расстояния, заданной (8.2.27). Показать, что отсюда следует, что симплексный код оптимален в смысле максимизации расстояния между двумя ближайшими кодовыми словами при заданном числе кодовых слов и заданном ограничении на энергию.

(б) Пусть $\mathbf{x}_m = (x_{m,1}, \ldots, x_{m,N})$ — m-е кодовое слово в двоичном коде максимальной длины (см. § 6.6.) Пусть $\varphi_1(t), \ldots, \varphi_N(t)$ — ортонормальные функции. Показать, что множество функций

$$x_m(t) = \sum_{n=1}^{N} (2x_{m,n}-1)\,\varphi_n(t), \quad 1 \leqslant m \leqslant N+1,$$

образует симплексный код.

8.16. Рассмотрим дискретный по времени канал, состоящий из канала с аддитивным гауссовым белым шумом (спектральной плотности $N_0/2 = 1$) и модулятора цифровых данных. На каждом интервале продолжительности T_0 модулятор передает одну функцию из множества K ортогональных функций, имеющих энергии ST_0 и ограниченных во времени заданным интервалом. Рассмотрим выход канала на заданном интервале как выходы K согласованных фильтров, каждый из которых согласован с одной из функций модулятора. Показать, что функция $E_0(1, \mathbf{Q})$ для этого канала при $Q(k) = 1/K$, $0 \leqslant k \leqslant K-1$, задается выражением

$$E_0(1, \mathbf{Q}) = \frac{ST_0}{4} - \ln\left[1+\frac{1}{K}\left(e^{ST_0/4}-1\right)\right].$$

Показать, что для этого дискретного по времени канала при любой скорости R (в натах в секунду) и длины блока N существуют коды с

$$P_e \leqslant \exp\left[TR - T\frac{S}{4}\left\{1-\frac{4}{ST_0}\ln\left[1+\frac{e^{ST_0/4}-1}{K}\right]\right\}\right],$$

где $T = NT_0$.

Показать, что при $R < S/8$, и при любом фиксированном T_0 эта экспонента аппроксимирует экспоненту для ортогональных кодовых слов при больших K. Обсудить качественно, каким должно быть K, чтобы эта аппроксимация была хорошей.

Указание: выходным алфавитом для этого канала является множество K-мерных векторов, отсюда

$$E_0(\rho, \mathbf{Q}) = -\ln \int\cdots\int \left\{\sum_{k=0}^{K-1} Q(k)\, p(y_0, \ldots, y_{K-1} \mid x=k)^{1/1+\rho}\right\}^{1+\rho} dy_0 \ldots dy_{K-1}.$$

При $\rho = 1$ квадрат можно раскрыть и проинтегрировать в замкнутой форме. См. Возенкрафт и Кеннеди (1966).

8.17. Можно освободиться от коэффициентов в границах $P_{e,m}$ (8.2.43) и (8.2.44) для ортогональных кодовых слов, заменяя (8.2.35) выражением

$$Q(y_m) \leqslant [(M-1)\,\Phi(-y_m)]^\rho \leqslant (M-1)^\rho \exp\left[-\frac{y_m^2\rho}{2}\right], \quad y_m \geqslant 0,$$

справедливым для любого ρ, $0 \leqslant \rho \leqslant 1$. Подставить это выражение в (8.2.32), проинтегрировать и, оптимизируя по ρ, вывести экспоненциальную границу (8.2.43) и (8.2.44).

Замечание: граница не будет иметь коэффициентов, которые присутствуют в (8.2.43) и (8.2.44).

8.18. Пусть задано множество ортогональных кодовых слов

$$x_m(t) = \sqrt{2S}\cos\left(\frac{2\pi mt}{T}\right), \quad 0 \leqslant t \leqslant T; \quad 1 \leqslant m \leqslant M,$$

и предположим, что при посылке сообщения m принятая функция имеет вид

$$y(t) = \sqrt{2S}\cos\left[\frac{2\pi mt}{T} + \theta\right] + z(t);\ 0 \leqslant t \leqslant T,$$

где θ — случайная фаза, равномерно распределенная между 0 и 2π, и $z(t)$ — белый гауссов шум со спектральной плотностью $N_0/2$. Пусть

$$y_{m,1} = \int_0^T y(t)\sqrt{\frac{2}{T}}\cos\left(\frac{2\pi mt}{T}\right)dt,$$

$$y_{m,2} = \int_0^T y(t)\sqrt{\frac{2}{T}}\sin\left(\frac{2\pi mt}{T}\right)dt$$

и будем считать, что декодирование заключается в выборе m, которое максимизирует

$$r_m = y_{m,1}^2 + y_{m,2}^2.$$

(а) Показать, что вероятность ошибки, при условии, что передано сообщение m, задается выражением

$$P_{e,m} = \iint p(y_{m,1}, y_{m,2} \mid m)\Pr[r_{m'} \geqslant r_m, \text{ какое-либо } m' \neq m \mid r_m, m]\,dy_{m,1}\,dy_{m,2} \leqslant$$
$$\leqslant (M-1)^{\rho}\iint p(y_{m,1}, y_{m,2} \mid m)\Pr[r_{m'} \geqslant r_m \mid r_m, m]^{\rho}\,dy_{m,1}\,dy_{m,2} \qquad (1)$$

для любого ρ, $0 \leqslant \rho \leqslant 1$.

(б) Показать, что для переданного m величина $r_{m'}$ имеет плотность вероятности $N_0^{-1}\exp[-r_{m'}/N_0]$ и, следовательно, $\Pr[r_{m'} \geqslant r_m \mid r_m, m] = \exp[-r_m/N_0]$. Подставляя это выражение в (1), показать, что правая часть (1) равна

$$\frac{(M-1)^{\rho}}{1+\rho}\exp\left[-\frac{\rho ST}{N_0(1+\rho)}\right]. \qquad (2)$$

Указание: сначала выполните интегрирование при условии $\theta = 0$ и затем покажите, что этот результат сохраняется для любых значений θ.

(в) Показать, что для $M = 2$, $\rho = 1$ выражение (2) дает точное значение $P_{e,m}$. Для произвольного M величину $(M-1)^{\rho}/(1+\rho)$ следует оценить сверху M^{ρ} и показать, что

$$P_{e,m} \leqslant \begin{cases} \exp[-T(\sqrt{C} - \sqrt{R})^2]; & {}^1/_4 C \leqslant R \leqslant C, \\ \exp\left[-T\left(\frac{C}{2} - R\right)\right]; & R \leqslant {}^1/_4 C, \end{cases}$$

где $C = S/N_0$ [см. Зеттерберг (1968)].

8.19. Проверить справедливость (8.5.89). Следует начать с непрерывности $\widetilde{E}_{\infty}(B, \rho)$ и показать, что для любого $\varepsilon > 0$ существуют δ_1 и δ_2, такие, что

$$\widetilde{E}_{\infty}(B + \delta_2, \rho - \delta_1) \geqslant \widetilde{E}_{\infty}(B, \rho) - \varepsilon,$$
$$\widetilde{S}_{\infty}(B + \delta_2, \rho - \delta_1) \leqslant \widetilde{S}_{\infty}(B, \rho),$$
$$\widetilde{R}_{\infty}(B + \delta_2) \geqslant \widetilde{R}_{\infty}(B).$$

Затем показать, что для достаточно больших T соотношение (8.5.85), используемое при аргументах $B + \delta_2$, $\rho - \delta_1$, приводит к (8.5.89).

8.20. Показать, что при постоянном R значение E, рассматриваемое как функция S (см. рис. 8.5.9), имеет всюду непрерывные производные.

8.21. Вход канала на интервале $(-T/2, T/2)$ равен одной из двух функций $x_1(t) = A\cos(2\pi f_1 t)$ или $x_2(t) = A\cos(2\pi f_2 t)$. Выход канала на интервале $(-T/2, T/2)$, когда посылается i-я функция ($i = 1, 2$), равен

$$y(t) = v_{1,i}\cos 2\pi f_i t + v_{2,i}\sin 2\pi f_i t + z(t),$$

где $z(t)$ — белый гауссов шум со спектральной плотностью $N_0/2$ и $v_{1,i}$ и $v_{2,i}$ — независимые гауссовские случайные величины с нулевыми средними и с дисперсиями $\mathscr{E}/T$; заметим, что $\mathscr{E}$ — общая средняя энергия сигнала, принятого на интервале.

(а) Пусть

$$y_{1,i} = \int_{-T/2}^{T/2} y(t)\,\frac{2}{\sqrt{T}}\cos(2\pi f_i t)\,dt,$$

$$y_{2,i} = \int_{-T/2}^{T/2} y(t)\,\frac{2}{\sqrt{T}}\sin(2\pi f_i t)\,dt.$$

Показать, что если сообщение i — послано, то $y_{1,i}$ и $y_{2,i}$ — независимые гауссовские случайные величины, каждая из которых имеет дисперсию $N_0 + \mathscr{E}$, и что $y_{1,j}$ и $y_{2,j}$ ($j \neq i$) — независимые гауссовские случайные величины с дисперсиями N_0. (Примите, что $f_1 T$ и $f_2 T$ — целые числа.)

(б) Пусть $\mathbf{y} = (y_{1,1}, y_{2,1}, y_{1,2}, y_{2,2})$. Найти

$$r(\mathbf{y}) = \ln\frac{p(\mathbf{y} \mid x_1(t))}{p(\mathbf{y} \mid x_2(t))}.$$

Показать, что плотность вероятности r, при условии, что передано $x_1(t)$, равна

$$p(r \mid x_1(t)) = \begin{cases} \dfrac{N_0(N_0+\mathscr{E})}{\mathscr{E}(2N_0+\mathscr{E})}\exp\left[-\dfrac{N_0}{\mathscr{E}}r\right]; & r \geqslant 0, \\ \dfrac{N_0(N_0+\mathscr{E})}{\mathscr{E}(2N_0+\mathscr{E})}\exp\left[\dfrac{N_0+\mathscr{E}}{\mathscr{E}}r\right]; & r \leqslant 0. \end{cases}$$

Показать, что $p(r|x_2(t)) = p(-r|x_1(t))$.

(в) Использовать результаты, полученные в пункте (б), для того чтобы показать, что при применении декодирования по максимуму правдоподобия вероятность ошибки в решении того, какая из этих функций была передана, задается равенством

$$P_e = \frac{1}{2 + \mathscr{E}/N_0}.$$

(г) Сравнить этот результат с границей вероятности ошибки (8.6.22), положив $\lambda_1 = \mathscr{E}$ и $\lambda_j = 0$ для $j > 1$.

(д) В интегральном уравнении (8.6.7) положим $T_1 = T$ и примем, что $\mathscr{R}(\tau) = \mathscr{E}/T$ при всех τ из области $(-T, T)$. Показать, что отсюда вытекает, что $\lambda_1 = \mathscr{E}$ и $\lambda_j = 0$ для $j > 1$. Указать качественно, как приведенное выше предположение отражается на величине допплеровского разброса в канале в сравнении с T^{-1}.

8.22. (а) Используя модель канала связи и обозначения § 8.6, показать, что декодирование по максимуму правдоподобия может быть выполнено, если выбирать сообщение m, для которого максимально выражение

$$\sum_{i=1}^{2}\sum_{j=1}^{\infty} y_{i,m,j}^2\,\frac{\lambda_j}{N_0+\lambda_j}.$$

(б) Показать, что указанная выше сумма равна

$$\sum_{i=1}^{2} \int_{-T_1/2}^{T_1/2} \left\{ \int_{-T_1/2}^{T_1/2} y_{i,\,m}(\tau) \left[\sum_j \sqrt{\frac{\lambda_j}{N_0+\lambda_j}}\, \varphi_j(\tau)\, \varphi_j(t) \right] d\tau \right\}^2 dt.$$

Показать, что выражение в квадратных скобках может быть истолковано как импульсный отклик соответственно выбранного изменяющегося во времени линейного фильтра.

8.23. Предположим, что собственные значения λ_j в (8.6.7) обладают тем свойством, что

$$\lambda_j = \begin{cases} \lambda; & 1 \leqslant j \leqslant n, \\ 0; & j > n, \end{cases}$$

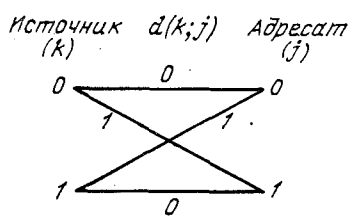
где $n = ST_1/\lambda$. Для $\rho = 1$ изобразить на графике $E_0(\rho, T)$ как функцию λ/N_0, считая, что ST фиксировано. Изобразить также

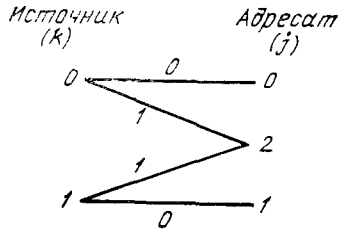
$$\lim_{\rho \to 0} \frac{E_0(\rho, T)}{\rho},$$

как функцию λ/N_0 при фиксированном ST. Какое заключение можно сделать из этого о наиболее благоприятных значениях отношения сигнал/шум на степень свободы в принятом сигнале (т. е. λ/N_0) для низких и высоких скоростей передачи?

Глава 9

9.1. Источник порождает независимые равновероятные двоичные символы. Найти и изобразить на графике скорость как функцию искажения для этого источника с каждой мерой искажения, указанной ниже. Опущенные переходы в диаграмме соответствуют бесконечным искажениям.





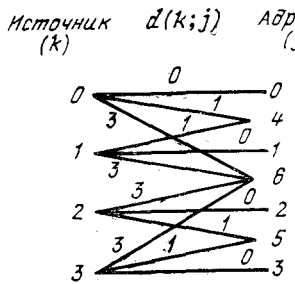
Указание: используя симметрию, попытайтесь угадать $P(j \mid k)$ и проверьте ваш результат, используя выпуклость $\mathcal{I}(\mathbf{Q}, \mathbf{P})$ по $\mathbf{P}$. Заметим, что вторая мера искажения дает пример не строго выпуклой $\cup$ функции $R(d^*)$.

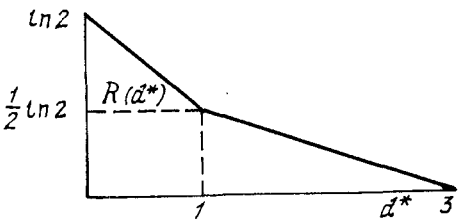
9.2. (а) Рассмотреть следующую схему кодирования источника и первую меру искажения из предыдущей задачи. Разобьем выход источника на последовательности из 7 символов в каждой. Для заданного (7.4) — кода Хэмминга с проверкой на четность закодируем каждую последовательность из 7 символов в 4 информационных символа кодового слова, ближайшего к заданной последовательности. У адресата, представим последовательность источника выбранным кодовым словом. Скорость для такой схемы равна $^4/_7 \ln 2$ нат на букву источника. Найти среднее искажение для такой схемы и сравнить его с $R(d^*)$.

(б) Для произвольного l использовать ту же схему с $(2^l - 1,\ 2^l - l - 1)$-кодом Хэмминга и найти скорость и среднее искажение.

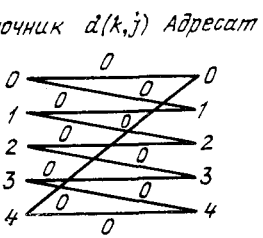
9.3. Для источника и второй меры искажения из задачи 9.1 найти простую схему кодирования, для которой скорость для любого заданного среднего искажения равна $R(d^*)$, вычисленной для этого среднего искажения.

9.4. Источник порождает независимые равновероятные буквы из алфавита объема 4. Показать, что скорость как функция искажения для источника и меры искажения, заданной на рисунке, имеет указанный вид.





9.5. Источник порождает независимые равновероятные буквы алфавита из 5 букв. Мера искажения показана на рисунке (где опущены переходы, соответствующие бесконечному искажению, и включены переходы с нулевым искажением).



(а) Найти скорость как функцию искажения для этого источника и меры искажения.

(б) Показать, что для любой скорости $R > \ln(5/2)$ существуют коды с достаточно большой длиной блока N, содержащие не более чем e^{NR} кодовых слов и имеющие нулевое искажение.

Указание: используйте лемму 9.3.1 и заметьте, что если по ансамблю кодов $P_c(D > 0) < 5^{-N}$, то по крайней мере один код должен иметь нулевое искажение [см. Пинкстон (1967)].

В некотором смысле эта задача двойственна задаче 5.11 (б). Удивительно, однако, что в той задаче C_0 неизвестно, в то время как здесь соответствующий результат так прост.

9.6. Источник порождает независимые буквы из троичного алфавита с вероятностями $Q(0) = 0{,}4$; $Q(1) = 0{,}4$; $Q(2) = 0{,}2$. Алфавит адресата троичный, и мера искажения имеет вид $d(k; j) = 0$ для $k = j$ и $d(k; j) = 1$ для $k \neq j$. Найти $R(d^*)$ и изобразить ее графически. Найти и изобразить графически как функцию C минимальную вероятность ошибки на символ источника, которая может быть достигнута при передаче этого источника по каналу с пропускной способностью C.

9.7. Рассмотреть дискретный по времени источник без памяти, выход которого представляет собой последовательность действительных случайных величин, имеющих плотность вероятности $q(u)$, дисперсию A и энтропию $H(U) = -\int q(u) \ln q(u)\, du$. Показать, что для $d^* \leqslant A$ функция $R(d^*)$ располо-

жена между следующими границами:

$$H(U) - {}^1/_2 \ln (2\pi e d^*) \leqslant R(d^*) \leqslant {}^1/_2 \ln \frac{A}{d^*}.$$

Указание: для вывода нижней границы проследите рассуждения, проводившиеся при переходе от (9.7.2) до (9.7.8). Для вывода верхней границы рассмотрите тест-канал, изображенный на рис. 9.7.3.

9.8. (а) Найти скорость как функцию искажения $R(d^*)$ для источника, выход которого представляет собой стационарный гауссовский случайный процесс со спектральной плотностью

$$F(f) = \begin{cases} A; & |f| \leqslant W_1, \\ 0; & |f| > W_1. \end{cases}$$

Мера искажения является квадратично-разностной.

(б) Найти пропускную способность канала с аддитивным белым гауссовым шумом с ограниченной мощностью S, спектральной плотностью шума $N_0/2$ и ограниченной полосой частот W_2.

(в) Найти выражение для минимальной среднеквадратической ошибки, которая может быть достигнута при передаче последовательности, порожденной источником пункта (а) по каналу пункта (б). Ввести безразмерные параметры и построить график этой минимальной среднеквадратической ошибки как функции W_2.

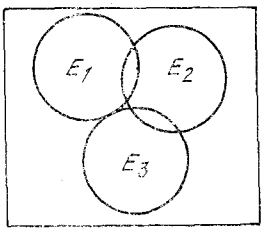
2.1. ($а_1$) Для несовместных событий $P(E_0) = P(E_1) + P(E_2) + P(E_3)$ так, что $P(E_0) = {}^3/_4$.

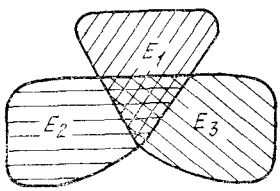
($а_2$) Для статистически независимых событий $1 - P(E_0)$ является вероятностью того, что не произойдет ни одного из событий; эта вероятность равна произведению вероятностей того, что не произойдет каждое из событий. Поэтому $1 - P(E_0) = ({}^3/_4)^3$ и $P(E_0) = {}^{27}/_{64}$.

($а_3$) Если $E_0 = E_1 = E_2 = E_3$, то $P(E_0) = {}^1/_4$.

($б_1$) Как следует из диаграммы Венна, представленной ниже, $P(E_0)$, очевидно, достигает максимума, когда события несовместны, так что $P(E_0) = {}^3/_4$.

($б_2$) Пересечение каждой пары множеств имеет вероятность ${}^1/_{16}$. Как следует из рисунка, $P(E_0)$ достигает максимума тогда, когда все эти попарные пересечения совпадают; в этом случае $P(E_0) = {}^3/_4 - {}^1/_{16}$.





2.2. Пусть L будет событием, состоящим в том, что выбрана шулерская игральная кость, и пусть H является событием, состоящим в том, что выбрана правильная игральная кость. Пусть A_i является событием, состоящим в том, что цифра i появляется при первом бросании и B_i является событием, состоящим в том, что цифра i появляется при втором бросании. Нам задано, что $P(L) = {}^1/_3$; $P(H) = {}^2/_3$; $P(A_1|L) = {}^2/_3$; $P(A_i|L) = {}^1/_{15}$ $(2 \leqslant i \leqslant 6)$; $P(A_i/H) = {}^1/_6$ $(1 \leqslant i \leqslant 6)$. Тогда

$$P(L \mid A_1) = \frac{P(L, A_1)}{P(A_1)} = \frac{P(A_1 \mid L)\,P(L)}{P(A_1 \mid L)\,P(L) + P(A_1 \mid H)\,P(H)} = {}^2/_3.$$

Это является условной вероятностью того, что шулерская кость выбрана при условии, что была цифра 1 при первом бросании. Рассматривая два последовательных бросания, мы сделаем предположение, исходящее из физического механизма бросания кости, что исходы последовательных бросаний данной кости являются независимыми. Поэтому $P(A_1 B_1 \mid L) = ({}^2/_3)^2$ и $P(A_1 B_1 \mid H) = ({}^1/_6)^2$. Отсюда следует, как и ранее, что $P(L \mid A_1 B_1) = {}^{16}/_{17}$.

2.3. (а) $\overline{x+y} = \sum_{x,y} (x+y) P_{XY}(x, y) = \sum_{x,y} x P_{XY}(x, y) + \sum_{x,y} y P_{XY}(x, y) =$

$$= \sum_x x P_X(x) + \sum_y y P_Y(y) = \bar{x} + \bar{y}.$$

Заметим, что статистическая независимость не является необходимой здесь и что вывод распространяется на недискретные случайные величины, если существуют их математические ожидания.

(б) $\overline{xy} = \sum_{x,y} xy P_{XY}(x, y) = \sum_{x,y} xy P_X(x)\, P_Y(y) = \sum_x x P_X(x) \sum_y y P_Y(y) = \bar{x}\,\bar{y}$.

Заметим, что в первой строке равенства была использована статистическая независимость. Пусть x и y принимают лишь значения ± 1 и 0. Пример некоррелированных, но зависимых случайных величин имеет место, когда

$$P_{XY}(1,0) = P_{XY}(0,1) = P_{XY}(-1,0) = P_{XY}(0,-1) = {}^1/_4.$$

Пример коррелированных и зависимых случайных величин имеет место, когда

$$P_{XY}(1,1) = P_{XY}(-1,-1) = {}^1/_2.$$

(в) Используя (а), будем иметь

$$\sigma_{x+y}^2 = \overline{(x-\bar{x}+y-\bar{y})^2} = \overline{(x-\bar{x})^2} + \overline{2(x-\bar{x})(y-\bar{y})} + \overline{(y-\bar{y})^2}.$$

Согласно (а) стоящее в середине выражение равно $2[\overline{xy} - \bar{x}\,\bar{y}]$. Для некоррелированных случайных величин оно равно нулю и поэтому $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$.

2.4.(а) $\bar{x} = \sum_x xP_X(x) \geqslant \sum_{x \geqslant \delta} xP_X(x)$,

где знак неравенства поставлен потому, что отброшенное выражение в последней сумме является неотрицательным. Оценивая снизу x с помощью δ в последней сумме, получаем

$$\bar{x} \geqslant \delta \sum_{x \geqslant \delta} P_X(x) = \delta \Pr(x \geqslant \delta), \quad \Pr(x \geqslant \delta) \leqslant \bar{x}/\delta.$$

При заданном $\delta > 0$ неравенство удовлетворяется с равенством, когда x — двоичная случайная величина, принимающая только значения 0 и δ.

(б) Если подставить ε^2 вместо δ в пункте (а), то получим

$$\Pr[(y-\bar{y})^2 \geqslant \varepsilon^2] \leqslant \frac{\sigma_y^2}{\varepsilon^2}.$$

Но $(y-\bar{y})^2 \geqslant \varepsilon^2$ равносильно $|y-\bar{y}| \geqslant \varepsilon$.

(в) Применяя индукцию в 2.3 (в), получаем, что дисперсия суммы независимых случайных величин равна сумме дисперсий отдельных случайных величин. Отсюда $\sum z_n$ имеет дисперсию $N\sigma_z^2$,

$$\overline{(y_N-\bar{y}_N)^2} = \frac{1}{N^2}\overline{\left(\sum z_n - \sum \bar{z}_n\right)^2} = \frac{N\sigma_z^2}{N^2} = \frac{\sigma_z^2}{N}.$$

Подставляя это выражение вместо дисперсии y_N в выражение пункта (б), получаем

$$\Pr[|y_N - \bar{y}_N| \geqslant \varepsilon] \leqslant \frac{\sigma_z^2}{N\varepsilon^2}, \tag{1}$$

$$\lim_{N\to\infty} \Pr[|y_N - \bar{y}_N| \geqslant \varepsilon] = 0 \quad \text{при любом } \varepsilon > 0.$$

(г) Здесь y_N равна умноженному на $1/N$ числу наступлений события E в N экспериментах, или, другими словами, y_N является относительной частотой наступления E в выборке. Поэтому (1) означает, что событие, состоящее в том, что относительная частота E отличается от вероятности E не более, чем на некоторое малое число ε, имеет вероятность, которая стремится к нулю с ростом N. Пусть p является вероятностью события E. Тогда $\bar{y}_N = p$ и $\sigma_z^2 = p(1-p)$. Наконец, так как событие E может произойти i раз в N испытаниях $\binom{N}{i}$ различными способами, каждый из которых имеет вероятность $p^i(1-p)^{N-i}$, то

$$\Pr\left[\sum_{n=1}^{N} z_n = i\right] = \binom{N}{i} p^i (1-p)^{N-i}.$$

Так как неравенство $|y_N - \bar{y}_N| \geqslant \varepsilon$ равносильно $|z_n - pN| \geqslant \varepsilon N$, то

$$\Pr\left[|y_N - \bar{y}_N| \geqslant \varepsilon\right] = \sum_{i=0}^{\lfloor pN-\varepsilon N \rfloor} \binom{N}{i} p^i (1-p)^{N-i} + \sum_{t=\lceil pN+\varepsilon N \rceil}^{N} \binom{N}{i} p^i (1-p)^{N-i}.$$

2.5 (а) Для любой заданной принятой последовательности y имеем

$$P(a_1 \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid a_1)\ P(a_1)}{P(\mathbf{y} \mid a_2)\ P(a_2) \dotplus P(\mathbf{y} \mid a_1)\ P(a_1)}.$$

y	$P(a_1 \mid \mathbf{y})$	y	$P(a_1 \mid \mathbf{y})$
000	$\dfrac{(1-\varepsilon)^3}{(1-\varepsilon)^3 \dotplus \varepsilon^3}$	111	$\dfrac{\varepsilon^3}{(1-\varepsilon)^3 \dotplus \varepsilon^3}$
001	$1-\varepsilon$	110	ε
010	$1-\varepsilon$	101	ε
100	$1-\varepsilon$	011	ε

(б) Обозначая через e событие, состоящее в том, что принято неправильное решение, для любого частного решения получаем

$$P(e) = \sum_{\mathbf{y}} P(e \mid \mathbf{y})\, P(\mathbf{y}).$$

Так как $P(\mathbf{y})$ не зависит от правила решения, то $P(e)$ достигает минимума, когда для каждого y выбирается такая буква источника, которая минимизирует $P(e \mid \mathbf{y})$, т. е. буква источника a_i, которая максимизирует $P(a_i \mid \mathbf{y})$. Так как $P(a_2 \mid \mathbf{y}) = 1 - P(a_1 \mid \mathbf{y})$ и так как величины, стоящие слева в приведенной выше таблице, больше $1/2$, а величины, стоящие справа, меньше чем $1/2$, то данное правило минимизирует $P(e)$.

(в) Неправильное решение происходит тогда, когда 2 или 3 из трех принятых символов являются неправильными. Поэтому $P(e) = 3\varepsilon^2(1-\varepsilon) + \varepsilon^3$.

(г) Вероятность неправильного решения достигает минимума при выборе a_1, если появляются n или меньше нулей, и при выборе a_2 во всех остальных случаях. Так как вероятность того, что любой заданный символ принят неправильно, меньше чем $1/2$, и так как неправильное решение происходит тогда, когда более половины символов приняты неправильно, то закон больших чисел (см. задачу 2.4) утверждает, что $P(e) \to 0$ при $n \to \infty$.

2.6. Пусть Y является ансамблем событий p (быстрая) и t (медлительная). Тогда

(а) $$I_{X;Y}(\text{блондинка};\ p) = \log \frac{P_{Y \mid X}(p \mid \text{блондинка})}{P_Y(p)} = \log \frac{1}{(1/4)\,1 + (1/2)\,(1/2) + (1/4)\,0} = 1 \text{ бит}.$$

Подобно этому $I_{X;Y}$ (брюнетка; p) $= 0$; $I_{X;Y}$ (шатенка; p) $= -\infty$ (не может быть шатенка).

(б) $$I_{X;Y_1Y_2Y_3}(\text{брюнетка};\ ppp) = \frac{\Pr(\text{брюнетка не опоздала все 3 раза})}{\Pr(\text{элемент } X \text{ не опоздал все 3 раза})} = \log \frac{1/8}{1/4 + 1/16} = -\log 5/2.$$

2.7. (а) $I_{\mathbf{X};\,Y_1}(\mathbf{x}_1;\,0) = \log \dfrac{\Pr[\text{1-й принятый символ} = 0 \mid \text{передавался } \mathbf{x}_1]}{\Pr[\text{1-й принятый символ} = 0]} =$

$$= \log \frac{1-\varepsilon}{1/2} = \log[2(1-\varepsilon)].$$

Здесь был использован результат, который очевиден из-за симметрии и который также легко вывести: если входами двоичного симметричного канала являются символы 0 или 1 с равными вероятностями, то выходами также являются символы 0 или 1 с равными вероятностями.

(в) Заметим, что первые три входных символа для канала являются статистически независимыми равновероятными двоичными символами. Так как канал является каналом без памяти, то легко получить, что первые три выхода также статистически независимы и равновероятны. Следовательно,

$$I_{\mathbf{X};\,Y_2 \mid Y_1}(x_1;\,0 \mid 0) = I_{\mathbf{X};\,Y_3 \mid Y_1 Y_2}(x_1;\,0 \mid 0,0) = \log[2(1-\varepsilon)].$$

Усредняя, наконец, по всем кодовым словам, найдем, что вероятность приема $\mathbf{y} = (0, 0, 0, 0)$ равна $1/8\,[(1-\varepsilon)^4 + 6(1-\varepsilon)^2\varepsilon^2 + \varepsilon^4]$. Поэтому

$$P_{Y_4 \mid Y_1 Y_2 Y_3}(0 \mid 0,0,0) = (1-\varepsilon)^4 + 6(1-\varepsilon)^2\varepsilon^2 + \varepsilon^4$$

и

$$I_{\mathbf{X};\,Y_4 \mid Y_1 Y_2 Y_3}(\mathbf{x}_1;\,0 \mid 0,0,0) = \log \frac{1-\varepsilon}{(1-\varepsilon)^4 + 6(1-\varepsilon)^2\varepsilon^2 + \varepsilon^4}.$$

2.8. Первые $N-1$ двоичных символов статистически не зависят друг от друга, а N-й определяется первыми $N-1$, так что

$$I(X_n;\,X_{n-1} \mid X_1, \ldots, X_{n-2}) = \begin{cases} 0 & \text{при } n < N, \\ 1 \text{ бит при } & n = N. \end{cases}$$

Заметим, что это рассуждение не зависит от порядка символов в последовательности. Другими словами, никакое множество из $N-2$ символов не дает никакой информации относительно x_N, хотя при любом заданном множестве из $N-2$ символов оставшийся символ разрешает всю неопределенность относительно x_N.

2.9. $H(X) = 3/2$ бит, $I(X;Y) = 1/2$ бит,
$H(Y) = 1$ бит, $I(X;Z) = 1$ бит,
$H(Z) = 1$ бит, $I(X;Y \mid Z) = 1/2$ бит,
$H(YZ) = 2$ бит, $I(X;YZ) = 3/2$ бит.

Выражение $I(X;\,Y \mid Z)$ истолковывается как средняя взаимная информация, содержащаяся в Y относительно X после того, как становится известным Z. В этом примере условие, что задано Z, является несущественным.

2.10. (а) С помощью прямых вычислений и упрощения полученного выражения будем иметь

$$I(X;Y) = (1-\varepsilon)\left[p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}\right].$$

Либо выполняя непосредственную оптимизацию по p, либо используя теорему 2.3 1, можно найти, что $p = 1/2$ максимизирует $I(X;\,Y)$. При $p = 1/2$ имеем $I(X;\,Y) = (1-\varepsilon)$ бит и

$$I_{X;Y}(0;\,0) = I_{X;\,Y}(1;\,1) = 1 \text{ бит}; \quad I_{X;Y}(x;\,E) = 0.$$

(б) Для данной стратегии передача символа источника заканчивается каждый раз, когда в канале не возникают стирания. Поэтому среднее число символов источника, переданных при одном использовании канала, равно $(1-\varepsilon)$. Согласно закону больших чисел при большом числе использований канала N с высокой вероятностью (близкой к 1) число переданных символов источника близко к $(1-\varepsilon)\,N$.

2. 11. (а) $I(X;Y) = 1 + {}^3/_4 \log {}^3/_4 + {}^1/_4 \log {}^1/_4 = 0{,}189$ бит.

(б) Капитал в конце n-й игры C_n равен удвоенной ставке на выигравший цвет, т. е. $C_n = C_{n-1}[2(1-q)]^{z_n}[2q]^{1-z_n}$.

Решая это рекуррентное соотношение, получаем

$$C_N = C_0 \prod_{n=1}^{N} [2(1-q)]^{z_n}[2q]^{1-z_n},$$

$$E_N = \frac{1}{N}\sum_{n=1}^{N} [z_n \log[2(1-q)] + (1-z_n)\log 2q].$$

Так как $z_n = 1$ с вероятностью ${}^3/_4$ и 0 с вероятностью ${}^1/_4$ и так как математическое ожидание произведения независимых случайных величин равно произведению математических ожиданий, то

$$\bar{C}_N = C_0\left\{\frac{3}{4}\left[2(1-q)\right] + \frac{1}{4}2q\right\}^N = C_0\left[\frac{3}{2} - q\right]^N,$$

$$\bar{E}_N = \frac{3}{4}\log[2(1-q)] + \frac{1}{4}\log(2q).$$

Из рассмотрения выражения для $\bar{C}_N$ находим, что $\bar{C}_N$ максимизируется при $q = 0$. Дифференцируя $\bar{E}_n$ по q, получим, что единственный максимум имеет место при $q = 1/4$. Поэтому

$$\max_q \bar{C}_N = C_0\left(\frac{3}{2}\right)^N, \quad \max_q \bar{E}_N = 1 + \frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4} = I(X;Y).$$

(в) Заметим, что при любом заданном q значение E_N является выборочным средним множества N одинаково распределенных независимых случайных величин $z_n \log[2(1-q)] + (1-z_n)\log 2q$. Поэтому, в силу закона больших чисел $\lim\limits_{N\to\infty} \Pr[|E_N - \bar{E}_N| > \varepsilon] = 0$ при любом $\varepsilon > 0$. В обозначениях C_N это результат утверждает, что

$$\lim_{N\to\infty} \Pr\left\{C_0 2^{N[\bar{E}_N - \varepsilon]} \leqslant C_N \leqslant C_0 2^{N[\bar{E}_N + \varepsilon]}\right\} = 1.$$

Другими словами, значение $\bar{E}_N$ определяет те близкие пределы, в которых как C_N, так и E_N при больших N находятся с высокой вероятностью. Игрок, который использует $q = 1/4$ (которое максимизирует $\bar{E}_N$), будет с подавляющей вероятностью иметь бóльший капитал после достаточно большого числа игр, чем игрок, который использует какое-либо другое значение q. Дополнительное проникновение в эту своеобразную ситуацию может быть получено с помощью рассмотрения того, что случится при $q = 0$ (которое максимизирует $\bar{C}_N$). В этом случае весь капитал ставится каждый раз на предсказанный цвет и любое появление другого цвета уменьшает капитал до 0. Таким образом, после N игр капитал равен $C_0 2^N$ с вероятностью $({}^3/_4)^N$. При увеличении N вероятность выигрыша экспоненциально убывает, однако выигрыши столь велики, когда они возникают, что математическое ожидание C_N велико. Это является примером экстремальной ситуации, в которой математический термин «ожидание» не имеет смысла, обычно вкладываемого в него в русском языке.

2.12. Требуется найти максимум $H(X)$ при двух ограничениях $\sum nP_X(n) = A$ и $\sum P_X(n) = 1$. Не будем учитывать дополнительное условие $P_X(n) \geqslant 0$ для всех $n \geqslant 0$ и будем надеяться, что решение, удовлетворяющее другим ука-

занным условиям, удовлетворяет также этим последним неравенствам. Используя множители Лагранжа, будем иметь

$$\frac{\partial}{\partial P(n)}\left[\sum_{q=0}^{\infty} -P(n)\log P(n) - \lambda \sum_{n=0}^{\infty} nP(n) - \gamma \sum_{n=0}^{\infty} P(n)\right] = 0,$$

$$-\log P(n) - \log e - \lambda n - \gamma = 0, \quad P(n) = e^{2-\lambda n - \gamma} = Bx^n,$$

где B и x должны быть выбраны так, чтобы удовлетворялись эти условия. Имеем

$$\sum P(n) = \frac{B}{1-x} = 1, \quad \sum nP(n) = \frac{Bx}{(1-x)^2} = A.$$

Отсюда

$$P_X(n) = \frac{1}{1+A}\left(\frac{A}{1+A}\right)^n, \qquad n = 0, 1, \ldots$$

То что это распределение действительно максимизирует $H(X)$ (а не является просто стационарной точкой), легче всего проверить с помощью свойств выпуклых функций, описанных в § 4.4.

2.13. Постоянное предсказание отсутствия дождя не дает никакой информации о погоде. Не следует думать, однако, что синоптик, в идеале, должен заботиться о максимизации средней взаимной информации в своих прогнозах.

2.14. Так как $P_X(a_M) = \alpha$, то

$$H(X) = -\alpha\log\alpha - \sum_{i=1}^{M-1} P_X(a_i)\log P_X(a_i) = -\alpha\log\alpha -$$

$$-(1-\alpha)\sum_{i=1}^{M-1}\frac{P_X(a_i)}{1-\alpha}\left[\log\frac{P_X(a_i)}{1-\alpha} + \log(1-\alpha)\right] = -\alpha\log\alpha -$$

$$-(1-\alpha)\log(1-\alpha) + (1-\alpha)H(Y).$$

Из теоремы 2.3.1 следует, что $H(Y) \leqslant \log(M-1)$, поэтому

$$H(X) \leqslant -\alpha\log\alpha - (1-\alpha)\log(1-\alpha) + (1-\alpha)\log(M-1).$$

2.15. Пусть X является ансамблем с вероятностями $P(a_k)$, $1 \leqslant k \leqslant K$. Без потери общности можно предположить, что $P(a_1) > P(a_2)$ и пусть Y является ансамблем с вероятностями

$$P(a_1) - \varepsilon, \quad P(a_2) + \varepsilon, \quad P(a_3), \ldots, P(a_K),$$

где $0 < \varepsilon < \dfrac{P(a_1) - P(a_2)}{2}$. Покажем теперь, что $H(Y) > H(X)$. Имеем

$$H(X) - H(Y) = -P(a_1)\log P(a_1) - P(a_2)\log P(a_2) +$$

$$+ [P(a_1) - \varepsilon]\log[P(a_1) - \varepsilon] + [P(a_2) + \varepsilon]\log[P(a_2) + \varepsilon] =$$

$$= P(a_1)\log\frac{P(a_1) - \varepsilon}{P(a_1)} + P(a_2)\log\frac{P(a_2) + \varepsilon}{P(a_2)} - \varepsilon\log\frac{P(a_1) - \varepsilon}{P(a_2) + \varepsilon}.$$

Используя неравенство $\log x \leqslant (x-1)\log e$, получим

$$H(X) - H(Y) \leqslant (\log e)[P(a_1) - \varepsilon - P(a_1) + P(a_2) + \varepsilon - P(a_2)] -$$

$$-\varepsilon\log\frac{P(a_1) - \varepsilon}{P(a_2) + \varepsilon} = -\varepsilon\log\frac{P(a_1) - \varepsilon}{P(a_2) + \varepsilon} < 0.$$

2.16. $P_{XY}(a_1, b_1) = 1/2,$

$P_{XY}(a_1, b_2) = 1/4,$

$P_{XY}(a_2, b_2) = 1/4.$

В этой задаче нужно показать, что возможны случаи, когда наблюдения некоторых букв из Y-ансамбля увеличивают неопределенность относительно X-ансамбля. Легко, однако, показать, с помощью того же самого доказательства, что и в теореме 2.3.5, что сумма

$$\sum_k P(a_k \mid b_j) \log \frac{P(a_k \mid b_j)}{P(a_k)}$$

всегда неотрицательна. Дальнейшие рассмотрения, связанные с этими частичными средними значениями, см. Blachman N. M., IEEE Trans., **IT-14**. January, 1968, 27—31.

2.17. (а) Это неравенство равносильно неравенству

$$\sum_{k=1}^{K} P(a_k) \log \frac{Q(a_k)}{P(a_k)} \leqslant 0,$$

которое сразу же следует из неравенства $\log x \leqslant (\log e)(x-1)$.

(б) Вновь используя то же самое неравенство, будем иметь

$$0 \leqslant \sum_k P(a_k) \log \frac{P(a_k)}{Q(a_k)} \leqslant [\log e] \left[\sum_k \frac{P^2(a_k)}{Q(a_k)} - \sum_k P(a_k) \right],$$

$$0 \leqslant \sum_k \frac{P^2(a_k)}{Q(a_k)} - 1, \quad \sum_k \frac{P^2(a_k)}{Q(a_k)} \geqslant 1.$$

2.18. (а) Согласно (2.3.15) для последовательных каналов, изображенных на рис. 2.3.2, $I(X; Y \mid Z) = 0$. Это означает, что X и Y становятся независимыми при условии, что задано значение Z, т. е. для каждой пары yz, имеющей ненулевую вероятность,

$$P(x \mid yz) = P(x \mid z). \tag{I}$$

Из (2.3.17) следует, что $I(X; Z) = I(X; Y)$ для последовательных каналов тогда и только тогда, когда $I(X; Z \mid Y) = 0$, что имеет место тогда и только тогда, когда для каждой пары yz, имеющей ненулевую вероятность,

$$P(x \mid yz) = P(x \mid y). \tag{II}$$

Из (I) и (II) получаем, что $I(X; Z) = I(X; Y)$ тогда и только тогда, когда для каждой пары yz, имеющей ненулевую вероятность,

$$P(x \mid z) = P(x \mid y) \text{ для всех } x \in X. \tag{III}$$

Предположим теперь, что $P_{Y \mid Z}(b_j \mid c_i) > 0$ и $P_{Y \mid Z}(b_j \mid c_l) > 0$. Так как условные вероятности определены только тогда, когда события, входящие в условия, имеют ненулевые вероятности, то yz-пары $b_j c_i$ и $b_j c_l$ имеют ненулевые вероятности, и, если $I(X; Y) = I(X; Z)$, то из (III) следует, что

$$P_{X \mid Z}(a_k \mid c_i) = P_{X \mid Y}(a_k \mid b_j) = P_{X \mid Z}(a_k \mid c_l)$$

при всех k, и c_i и c_l эквивалентны. Обратно, пусть предположение, что $P_{YZ}(b_j, c_i) > 0$ и $P_{YZ}(b_j, c_l) > 0$, означает, что c_i и c_l эквивалентны. Тогда из (I) следует, что для этих пар yz, имеющих ненулевую вероятность, $P(x \mid yz)$ не зависит от z и поэтому справедливо (II), откуда следует $I(X; Y) = I(X; Z)$.

(б) Если c_i и c_l являются эквивалентными для данного распределения на входе с $P_X(a_k) > 0$, то для всех k имеем

$$P_{X \mid Z}(a_k \mid c_i) = \frac{P_{Z \mid X}(c_i \mid a_k) P_X(a_k)}{P_Z(c_i)} = \frac{P_{Z \mid X}(c_l \mid a_k) P_X(a_k)}{P_Z(c_l)}.$$

Пусть $\alpha = P_Z(c_i) / P_Z(c_l)$. Тогда для всех k имеем

$$P_{Z \mid X}(c_i \mid a_k) = \alpha P_{Z \mid X}(c_l \mid a_k). \tag{IV}$$

Если $Q_X(a_k) > 0$ является другим распределением на входе и Q_Z — соответствующее ему распределение на выходе, то из (IV) получаем $Q_Z(c_i) = \alpha Q_Z(c_l)$.

При этом $\dfrac{P_{Z|X}(c_i|a_k)\,Q_X(a_k)}{Q_Z(c_i)} = \dfrac{P_{Z|X}(c_l|a_k)\,Q_X(a_k)}{Q_Z(c_l)}$

и c_i, c_l остаются эквивалентными при этом новом распределении. Теперь из результатов пункта (а) следует, что если $I(X; Y) = I(X; Z)$ для первоначального распределения, то тот же самый результат справедлив и для нового распределения.

2.19. (а) $I(XY; Z) = I(X; Z) + I(Y; Z|X)$.

Так как $I(Y; Z|X) \geqslant 0$, то $I(XY; Z) \geqslant I(X; Z)$. Из теоремы 2.3.3 следует, что неравенство переходит в равенство тогда и только тогда, когда при условии, что задано любое значение x, имеет место статистическая независимость y и z.

(б) $H(XY|Z) = H(X|Z) + H(Y|XZ)$. Поэтому $H(XY|Z) \geqslant H(X|Z)$ с равенством тогда и только тогда, когда y однозначно определяется x и z.

(в) $I(XY; Z) = I(X; Z) + I(Z; Y|X) = I(Z; Y) + I(X; Z|Y)$. После соответствующей перестановки членов можно увидеть, что заданное неравенство всегда удовлетворяется с равенством.

(г) $H(XYZ) - H(XY) = H(Z|XY)$, $H(XZ) - H(X) = H(Z|X)$.

Из (2.3.13) следует, что $H(Z|X) \geqslant H(Z|XY)$ с равенством тогда и только тогда, когда при условии, что задано любое значение x, имеет место статистическая независимость y и z. Таким образом, данное неравенство справедливо и переходит в равенство при указанных выше условиях.

2.20. (а) $P_{XYZ}(0, 0, 0) = P_{XYZ}(0, 1, 1) = P_{XYZ}(1, 0, 1) = P_{XYZ}(1, 1, 0) = 1/4$.

При этом имеем $I(X; Y) = 0$ и $I(X; Y|Z) = 1$ бит.

(б) $P_{XYZ}(0, 0, 0) = P_{XYZ}(1, 1, 1) = 1/2$.

При этом $I(X; Y) = 1$ бит и $I(X; Y|Z) = 0$.

2.21. (а) $\mathbf{D}[I(x; y)] = 0$ тогда и только тогда, когда $I(x; y)$ принимает одно и то же значение для всех пар xy, имеющих положительную вероятность, т. е. если $\log P(x, y)/P_X(x)\,P_Y(y) = \log \alpha$ для всех xy, для которых $P(x, y) > 0$; $P(x, y) = \alpha P_X(x)\,P_Y(y)$.

(б) После усреднения $I(x; y)$ получим $I(X; Y) = \log \alpha$.

При $\alpha = 1$ имеет место статистическая независимость X и Y, и $I(X; Y) = 0$.

(в) Для первого канала $P_X(a_1) = 1/2$, и $P_X(a_2)$ и $P_X(a_3)$ произвольны. Имеем $I(X; Y) = 1$ бит. Для второго канала $P_X(a_i) = 1/3$, $1 \leqslant i \leqslant 3$, и $I(X; Y) = \log 3/2$.

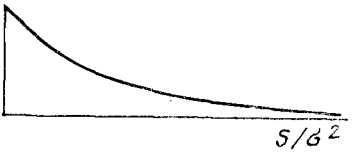
2.22. См. какой-либо элементарный учебник, в котором рассматриваются случайные величины с непрерывным множеством значений.

2.23. (а) $p_{Y|X}(y|-\sqrt{S}) = \dfrac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\dfrac{(y+\sqrt{S})^2}{2\sigma^2}\right]$,

$$\Pr[y > 0 \mid x = -\sqrt{S}] = \int_0^\infty \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(y+\sqrt{S})^2}{2\sigma^2}\right] dy =$$

$$= \int_{\sqrt{S}/\sigma}^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] du,$$

где была сделана подстановка $u=\dfrac{y+\sqrt{S}}{\sigma}$. Точно так же можно показать, что $\Pr[y<0 \mid x=\sqrt{S}]$ имеет то же самое значение.



(б) $$\Pr[\operatorname{sign} y \neq \operatorname{sign} x]=\int\limits_{\sqrt{S/\sigma^2}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du=$$

$$=\frac{1}{2}-\int\limits_{0}^{\sqrt{S/\sigma^2}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du\approx\frac{1}{2}-\sqrt{\frac{S}{2\pi\sigma^2}};\quad \sqrt{S/\sigma^2}\text{ — мало},$$

где было использовано приближенное равенство $\exp -\alpha \approx 1$ для малых α. Известно (Феллер, т. 1, гл. VII, § 1), что для больших значений y справедливо

$$\int\limits_{y}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du\approx\frac{1}{\sqrt{2\pi}y}\exp\left(-\frac{y^2}{2}\right).$$

Положив $y=\sqrt{S/\sigma^2}$, получаем искомый результат.

2.24. (а) $$-I(X;Y)=\int\int p_{XY}(x,y)\log\frac{p_X(x)\,p_Y(y)}{p_{XY}(x,y)}\,dx\,dy\leqslant$$

$$\leqslant(\log e)\iint p_{XY}(x,y)\left[\frac{p_X(x)\,p_Y(y)}{p_{XY}(x,y)}-1\right]dx\,dy=0.$$

(б) Пусть

$$p_X(x)=\begin{cases}1/A \text{ при } 0\leqslant x\leqslant A,\\ 0 \text{ во всех остальных случаях.}\end{cases}$$

Тогда $H(X)=\log A$ и это выражение отрицательно при $A<1$.

2.25. (а) Заметим, что в силу симметрии $p_Y(y)=\dfrac{1}{2\pi}$, $0\leqslant y\leqslant 2\pi$. Имеем

$$I(X;Y)=\int\limits_X p_X(x)\int\limits_Y p_{Y\mid X}(y\mid x)\log\frac{p_{Y\mid X}(y\mid x)}{p_Y(y)}\,dy\,dx.$$

Теперь имеем $p_{Y\mid X}(y\mid x)=p_Z(y-x)$, где вычитание производится по модулю 2π. Таким образом,

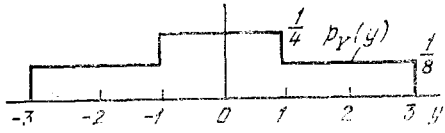
$$\int\limits_Y dy\,p_{Y\mid X}(y\mid x)\log p_{Y\mid X}(y\mid x)=-H(Z),\quad I(X;Y)=\log 2\pi-H(Z).$$

(б) Если $p_Z(z)=\dfrac{1}{b-a}$, $a\leqslant z<b$, то $H(Z)=\log(b-a)$ и

$$I(X;Y)=\log 2\pi-\log(b-a)=\log\frac{2\pi}{b-a}.$$

Различие между аддитивным шумом, находящимся в интервале от a до b, и аддитивным шумом, находящимся в интервале от 0 до $b-a$, соответствует известному фазовому сдвигу a в принятой фазе. Это не меняет $I(X;Y)$, и фактически это различие можно устранить с помощью изменения нуля опорной фазы на приемнике.

2.26. (а)



(б) $I(X;Y) = 1/2$ бит.

(в) $I(X;U) = 1/2$ бит.

Это означает, что рассматриваемое преобразование не разрушает информацию, содержащуюся на выходе относительно X. Это не удивительно, так как для всех y из интервала (1,3) значение x должно быть $+1$, и тот же самый результат получается при $u = 1$. Точно так же при любом заданном y из интервала $(-1, 1)$ значение x с равными вероятностями будет $+1$ или -1, и тот же самый результат получается при $u = 0$.

2.27. (а) Нам задан ансамбль $P_{XYZ}(0,0,0) = P_{XYZ}(1,1,1) = 1/2$.

Либо производя вычисления, либо замечая, что z непосредственно определяет x, получаем $I(X;Y \mid Z) = 0$.

(б) Если X и Y разбиваются на отдельные события 0 и 1, то имеем $X_p = X$ и $Y_p = Y$. Если Z состоит только из одного события, которое представляет собой все выборочное пространство, то $I(X_p;Y_p \mid Z_p) = I(X;Y) = 1$ бит.

(в) Рассматривая Z, состоящие только из одного события, которое представляет собой все выборочное пространство, будем иметь $I(X_p;Y_p \mid Z_p) = I(X_p;Y_p)$. Поэтому получим

$$\sup I(X_p;Y_p \mid Z_p) \geqslant \sup I(X_p;Y_p) = I(X;Y).$$

Таким образом, если $I(X;Y) > I(X;Y \mid Z)$, то

$$\sup I(X_p;Y_p \mid Z_p) > I(X;Y \mid Z).$$

3.1. (а) Имеются $1 + 100 + \binom{100}{2} + \binom{100}{3} = 166\,701$ последовательность с 3 или меньшим числом единиц, и кодовые слова должны быть сопоставлены каждой из них. Так как $2^{17} < 166\,701 < 2^{18}$, то двоичные кодовые слова должны иметь длину 18.

(б) Pr [последовательность содержит 4 или больше единиц] $=$

$$= \sum_{i=4}^{100} \binom{100}{i} (0{,}995)^{100-i} (0{,}005)^i \approx 0{,}0018.$$

(в) Используя неравенство Чебышева в виде, данном в задаче 2.4 (а), и обозначая через i число единиц в последовательности, получаем

$$\Pr[i \geqslant 4] \leqslant \frac{\bar{i}}{4} = \frac{1}{8}.$$

Используя обычную форму неравенства из задачи 2.4 (б), получим

$$\Pr[i \geqslant 4] \leqslant \Pr[|i - \bar{i}| \geqslant 4 - \bar{i}] \leqslant \frac{100\,(0{,}005)\,(0{,}995)}{(3{,}5)^2} = 0{,}04.$$

Заметим, что указанные оба вида границы являются довольно слабыми.

3.2. Согласно неравенству Чебышева [см. задачу 2.4 (в)], имеем

$$\Pr\left[\left|\frac{I(u)}{L} - H(U)\right| \geqslant \delta\right] \leqslant \frac{\mathrm{D}[I(u)]}{L\delta^2},$$

где u — одиночная буква источника и

$$\mathrm{D}\,[I(u)] = \frac{3}{4}\left[\log\frac{4}{3}\right]^2 + \frac{1}{4}(\log 4)^2 - H^2(U) \approx 0{,}471.$$

(а) $L_0 \approx 1884$.

Заметим, что результаты будут отличными, если использовать логарифмы с другими основаниями

(б) $L_0 \approx 0{,}471 \cdot 10^{12}$.

Как можно было бы ожидать, неравенство Чебышева является здесь очень слабым.

(в) Из (3.1.13) и (3.1.14) имеем

$$(1-\varepsilon)\,2^{L[H(U)-\delta]} \leqslant A \leqslant 2^{L[H(U)+\delta]},$$

$$10^{403} \leqslant A \leqslant 10^{516{,}5} \quad \text{для пункта (а)},$$

$$10^{1{,}146 \cdot 10^{10}} \leqslant A \leqslant 10^{1{,}149 \cdot 10^{10}} \quad \text{для пункта (б)}.$$

3.3. (а) Код I удовлетворяет свойству префикса, а код II не удовлетворяет.

(б) Оба кода являются однозначно декодируемыми (появление 1 в коде II всегда означает начало нового кодового слова).

(в) Для кода I имеем $I(a_1; 1) = -\log 0{,}4$. Для кода II имеем $I(a_1; 1) = 0$.

(г) Для кода I имеем $I(U; X_1) = -0{,}4 \log 0{,}4 - 0{,}6 \log 0{,}6 = 0{,}971$ бит. Для кода II имеем $I(U; X_1) = 0$.

Первая буква каждого кодового слова кода II дает информацию о предыдущем сообщении.

3.4. (а) Возьмите сначала каждую пару кодовых слов, для которой одно является префиксом другого, и для каждой такой пары запишите повисший суффикс, который остается, когда префиксное слово устраняется из начальной части более длинного слова. Например, повисший суффикс для пары 01, 01110 есть 110. Рассмотрим далее кодовые слова вместе с повисшими суффиксами и для каждой пары, состоящей из одного повисшего суффикса и одного кодового слова, в которой одно является префиксом другого, запишем новый повисший суффикс. Нет нужды, конечно, записывать суффиксы, которые уже были записаны. Будем продолжать делать это с каждым новым повисшим суффиксом, добавленным к списку, до тех пор пока либо нельзя будет сформировать новый повисший суффикс (в этом случае код будет однозначно декодируемым), либо один из повисших суффиксов будет кодовым словом (в этом случае код не является однозначно декодируемым).

(б) Каждый повисший суффикс, сформированный на первом этапе процедуры, должен быть суффиксом кодового слова. Каждый повисший суффикс, который сформирован далее, является либо суффиксом кодового слова, либо суффиксом некоторого ранее сформированного повисшего суффикса. С помощью индукции получаем, что каждый повисший суффикс является суффиксом кодового слова (т. е. если все повисшие суффиксы, найденные вплоть до данного момента, являются суффиксами кодового слова, то следующий повисший суффикс должен быть подобным же). Исключая тривиальный случай кодовых слов из повторений или нулевой длины, получим, что для кодового слова длины m_i существует не более $m_i - 1$ суффиксов этого кодового слова, которые могут появляться как повисшие суффиксы. Следовательно, общее число повисших суффиксов, которые могут появиться, ограничено сверху величиной $\sum_{i=1}^{M} (m_i - 1)$.

(в) Все перечисленные коды, кроме {0, 01, 10}, являются однозначно декодируемыми.

(г) Для кода {0, 01, 11} последовательность 01111111... может быть разложена на $a_1, a_3, a_3, \ldots$ или $a_2, a_3, a_3, \ldots$ Для кода {110, 11, 100, 00,10} последовательность 11000000... может быть разложена на $a_1, a_4, a_4, \ldots$ или $a_2, a_4, a_4, \ldots$ Такие последовательности не могут быть построены для других однозначно декодируемых кодов.

3.5. (а) {00, 01, 100, 101, 1100, 1101, 1110, 1111}.
(б) Заметим, что для любых $k > j$

$$Q_k - Q_j = \sum_{i=j}^{k-1} P(a_i) \geqslant P(a_j) \geqslant 2^{-n_j}.$$

Поэтому кодовые слова для a_k и a_j должны отличаться где-либо в первых n_j позициях, и так как оба они имеют длину, по крайней мере, n_j, то имеет место свойство префикса. Далее для каждого i имеем

$$-\log P(a_i) \leqslant n_i < -\log P(a_i) + 1.$$

Умножая на $P(a_i)$ и суммируя по i, получаем $H(U) \leqslant \bar{n} < H(U) + 1$.
3.6. Для любых $k > j$ теперь имеем

$$Q_k - Q_j = \sum_{i=j+1}^{k-1} P(a_i) + \frac{1}{2} P(a_j) + \frac{1}{2} P(a_k) \geqslant$$
$$\geqslant \frac{1}{2} P(a_j) + \frac{1}{2} P(a_k) \geqslant 2^{-n_j} + 2^{-n_k}.$$

Поэтому кодовые слова должны отличаться в первых n_j или n_k символах в зависимости от того, какое из этих чисел меньше. Отсюда следует, что свойство префикса удовлетворяется и код является алфавитным. Так как $n_i < -\log P(a_i) + 2$, то, усредняя по i, получаем $\bar{n} < H(U) + 2$.
3.7. (а) Если код является однозначно декодируемым, то при любом k первые k кодовых слов должны образовывать однозначно декодируемый код и на основании доказанных теорем

$$\sum_{i=1}^{k} D^{-n_i} \leqslant 1 \text{ при любых } k.$$

Так как левая сторона неравенства увеличивается с ростом k и ограничена 1, то предел существует и ограничен 1. Обратно, при заданном бесконечном множестве длин, которые удовлетворяют неравенству Крафта, можно использовать построение, выполненное в теореме 3.2.1, чтобы построить код, обладающий свойством префикса. Здесь необходимо сделать два замечания. Первое состоит в том, что так как неравенство Крафта удовлетворяется, то существует конечное множество слов каждой длины n_i и поэтому длины могут быть упорядочены; второе состоит в том, что следует начинать (методически) с полного бесконечного дерева.
(б) Теорема 3.3.1 следует из теорем 3.2.1 и 3.2.2 при $K = \infty$ точно так же, как это было при конечном K. Заметим, однако, что если значение $H(U)$ бесконечно, то средняя длина кодового слова также будет бесконечной.
3.8. Пусть $P(a_1) = 1 - \delta$, $P(a_2) = \delta$. Код с этими сообщениями декодируется однозначно только тогда, когда оба слова имеют длину, по крайней мере, равную 1, так, что $\bar{n} \geqslant 1$. Вместе с тем при $\delta \to 0$ имеем $H(U) \to 0$, так что для любого $\varepsilon > 0$ и для достаточно малого δ имеем $\bar{n} \geqslant H(U) + 1 - \varepsilon$.
3.9. Двоичным и троичным кодами для первого источника будут {00, 10, 010, 011, 110, 111} и {0, 10, 11, 12, 20, 21}. Средние длины равны 2,5 и 1,7. Эти коды не являются единственными, а средние длины однозначно определенными. Кодами для вторичного источника будут {00, 10, 010, 110, 111, 0110, 0111} и {0, 1, 20, 21, 220, 221, 222}. Средние длины равны 2,55 и 1,65.
3.10. (а) N_L является суммой длин L кодовых слов, соответствующих L буквам источника. Поэтому N_L представляет собой сумму L независимых одинаково распределенных случайных величин, каждая из которых имеет среднее значение $\bar{n} = 2{,}5$. Таким образом, на основании закона больших чисел искомый предел равен 2,5, или в более точной формулировке, для любого $\varepsilon > 0$ имеем

$$\lim_{L\to\infty} \Pr\left[\left|\frac{N_L}{L} - 2{,}5\right| > \varepsilon\right] = 0.$$

(б) На основании теоремы 3.3.2 получаем

$$\lim_{K\to\infty} \frac{\overline{N_{LK}(K)}}{LK} = H(U),$$

а также, используя те же рассуждения, что и в пункте (а), получаем

$$\lim_{L\to\infty} \Pr\left[\left|\frac{N_{LK}(K)}{LK} - \frac{\overline{N_{LK}(K)}}{LK}\right| > \varepsilon\right] = 0.$$

Поэтому

$$\lim_{K\to\infty}\lim_{L\to\infty} \Pr\left[\left|\frac{N_{LK}(K)}{LK} - H(U)\right| > \varepsilon\right] = 0.$$

3.11. Двумя такими кодами будут {00, 01, 02, 10, 11, 12} и {00, 01, 02, 10, 11, 120, 121}. Общее правило состоит в том, чтобы вначале сгруппировать два наиболее вероятных сообщения, если алфавит источника имеет нечетное число букв, и сгруппировать три наиболее вероятных сообщения, если число букв четное. Затем используется процедура Хаффмана с $D = 3$ до тех пор, пока приведенный ансамбль не будет иметь только два сообщения, и в этом месте используется процедура Хаффмана с $D = 2$. Этот алгоритм можно обосновать точно так же, как обосновывается обычная процедура Хаффмана; единственное отличие состоит в том, что полное дерево с самой низкой двоичной точкой ветвления концов ребер и последующими троичными точками ветвления имеет четное число концевых узлов.

3.12. Заметим, что если кодовое слово в коде, обладающем свойством префикса, инвертируется (т. е. слово $\mathbf{x} = x_1x_2x_3x_4$ заменяется на $x_4x_3x_2x_1$), то результирующий код удовлетворяет свойству суффикса. Код, удовлетворяющий свойству суффикса, должен быть однозначно декодируемым, так как если бы это было не так, то две последовательности кодовых слов с одними и теми же кодовыми буквами можно было бы инвертировать и получить неоднозначно декодируемую последовательность для соответствующего кода, обладающего свойством префикса. Код, обладающий свойством суффикса и имеющий минимальную среднюю длину, может быть построен с помощью отыскания кода Хаффмана на первом этапе и последующей инверсией кодовых слов (эта последняя операция не меняет среднюю длину).

3.13. Двумя такими кодами будут {00, 01, 02, 10, 11, 12, 20, 21} и {0,20, 21, 10, 11, 12, 220, 221}. Оба они имеют среднюю длину 2, но дисперсия длины для первого кода равна 0, а для второго — равна 0,4. Очевидным преимуществом первого кода является то, что при его использовании не возникает проблема ждущей очереди.

3.14. (а) Из (3.3.5) и (3.3.6) следует, что $H(U) = \bar{n}\log_2 3$ тогда и только тогда, когда неравенство

$$\log\frac{3^{-n_k}}{P(a_k)} \leqslant (\log e)\left[\frac{3^{-n_k}}{P(a_k)} - 1\right]$$

удовлетворяется при всех k. Это, в свою очередь, происходит лишь тогда, когда $3^{-n_k} = P(a_k)$.

(б) Если $P(a_k) = 3^{-n_k}$ для всех k, то неравенство Крафта удовлетворяется с равенством. Вместе с тем, если число сообщений является четным, то кодовое дерево не является полным и при этом еще одно кодовое слово могло бы быть добавлено без нарушения неравенства Крафта. Поэтому число сообщений должно быть нечетным.

3.15. (а) Существуют, по крайней мере, два кодовых слова самой большой длины. Если бы длина самого короткого слова отличалась от этой наибольшей длины больше чем на 1, то можно было бы уменьшить длину двух длинных слов на 1 и увеличить длину самого короткого слова на 1 без нарушения неравенства Крафта. Это бы привело к коду, средняя длина которого была бы меньше, чем для первоначального кода, и это противоречит предположению, что первоначальный код был кодом Хаффмана. Таким образом, наибольшая и наименьшая длины отличаются не более чем на 1 и согласно неравенству Крафта длины должны быть равны j и $j+1$.

(б) Пусть L является числом слов с длиной j. Согласно неравенству Крафта, которое должно удовлетворяться с равенством для двоичного кода Хаффмана, имеем

$$L2^{-j}+(x2^j-L)\,2^{-j-1}=1, \quad L=(2-x)\,2^j.$$

(в) $$\bar{n}=\frac{L}{x2^j}j+\frac{x2^j-L}{x2^j}(j+1)=j+\frac{2(x-1)}{x}.$$

3.16. Пусть a_{K-1} и a_K будут объединены в ансамбле U, что даст редуцированный ансамбль U'. Тогда

$$H(U)-H(U')=-P(a_{K-1})\log P(a_{K-1})-P(a_K)\log P(a_K)+$$
$$+[P(a_{K-1})+P(a_K)]\log[P(a_{K-1})+P(a_K)]=$$
$$=[P(a_{K-1})+P(a_K)]\left[q\log\frac{1}{q}+(1-q)\log\frac{1}{1-q}\right],$$

где $$q=\frac{P(a_K)}{P(a_{K-1})+P(a_K)}.$$

В предположении, что логарифмы берутся по основанию 2, энтропия, которая была выписана выше, не больше чем 1, поэтому

$$H(U)-H(U')\leqslant P(a_{K-1})+P(a_K).$$

Так как оптимальный код для U может быть построен из оптимального кода для U' с помощью добавления концевых 0 и 1 к последнему кодовому слову, то

$$\bar{n}-\bar{n}'=P(a_{K-1})+P(a_K).$$

Объединяя эти результаты, получаем $\bar{n}-H(U)\geqslant\bar{n}'-H(U')$.

3.17. (а) Префиксный код может быть использован на последнем этапе кодирования, и если это так, то все этапы, очевидно, являются однозначно декодируемыми и, следовательно, весь код целиком является однозначно декодируемым.

(б) Указанные последовательности источника имеют вероятности 0,1; $(0,9)\cdot(0,1)$; $(0,9)^2\cdot(0,1)$; ...; $(0,9)^7\cdot(0,1)$; $(0,9)^8$. Поэтому

$$\bar{n}_1=\sum_{i=1}^{8} i\,(0,1)\,(0,9)^{i-1}+8\,(0,9)^8=5,6953.$$

(в) $\bar{n}_2=(0,9)^8+4\,(1-0,9^8)=2,7086$.

(г) Пусть $N(i)$ является числом символов источника, порождающих первые i промежуточных символов. Для любого $\varepsilon>0$ имеем

$$\lim_{i\to\infty}\Pr\left[\left|\frac{N(i)}{i}-\bar{n}_1\right|>\varepsilon\right]=0.$$

Аналогично пусть $L(i)$ является числом окончательных кодовых символов, соответствующих первым i промежуточным символам. Имеем

$$\lim_{i\to\infty}\Pr\left[\left|\frac{L(i)}{i}-\bar{n}_2\right|>\varepsilon\right]=0.$$

Отсюда видно, что для любого $\varepsilon > 0$

$$\lim_{i\to\infty} \Pr\left[\left|\frac{L(i)}{N(i)} - \frac{\bar{n}_2}{\bar{n}_1}\right| > \varepsilon\right] = 0,$$

$$\frac{\bar{n}_2}{\bar{n}_1} = 0{,}4756\ldots$$

Средняя длина кода Хаффмана, который кодирует сразу четыре символа, равна 1,9702. Отсюда видно, что код Хаффмана дает оптимальное решение математической задачи при заданном множестве сообщений, однако выбор множества сообщений может быть более важным, чем выбор кодовых слов для данного множества сообщений.

3.18. (а) Оптимальным кодом будет {1, 01, 0000, 001, 0001} среднее время на одну букву источника равно 4,15.

(б) Если вероятность одного из узлов больше, чем вероятность узла, соответствующего более короткому пути, тогда два узла могут быть переставлены (переставляются также части деревьев, исходящих из них) и такая перестановка должна уменьшить среднюю длину.

3.19. Существуют $2M + 1$ возможных вариантов для M монет, соответствующих тому, что одна из M монет будет тяжелой или легкой или все монеты будут стандартными. Имеются 3 возможных исхода каждого взвешивания и 3^n последовательностей исходов n взвешиваний. Отсюда следует, что $2M+1 \leqslant 3^n$; $M \leqslant (3^n - 1)/2$.

(а) Для любого заданного $n \geqslant 1$ возьмем $M = (3^n - 1)/2$. Поместим $(3^{n-1} + 1)/2$ пенни на одной стороне весов и $(3^{n-1} - 1)/2$ пенни и одну стандартную пенни — на другой стороне. Если весы находятся в равновесии, то все, кроме $(3^{n-1} - 1)/2$ оставшихся пенни, имеют стандартный вес и возникает первоначальная задача при n, уменьшенном на 1. Если весы находятся не в равновесии, то у нас будет $(3^{n-1} + 1)/2$ потенциально тяжелых монет и $(3^{n-1} - 1)/2$ потенциально легких монет (или, возможно, наоборот). Поместим $(3^{n-2} + 1)/2$ потенциально тяжелых монет и столько же потенциально легких монет на одну сторону и по $(3^{n-2} - 1)/2$ каждых из них и еще одну стандартную монету на другую сторону. Легко проверить, что после взвешивания (для каждого из трех возможных исходов) будут $(3^{n-2} + 1)/2$ потенциально тяжелых монет и $(3^{n-2} - 1)/2$ потенциально легких монет (или наоборот), и, таким образом, точно такая же стратегия с n, уменьшенным на 1, будет работать при следующем взвешивании.

(б) Секрет успеха (а) состоит в том, что при каждом взвешивании альтернативами были три равных множества. При $M = (3^n - 1)/2$ и в отсутствии стандартной монеты четное число альтернатив будет невозможным, если первое взвешивание будет с нарушением равновесия, и поэтому альтернативы не могут представлять собой равные множества. Покажем теперь, как можно выполнить взвешивания с $M = (3^n - 3)/2$ монетами. Положим вначале $(3^{n-1} - 1)/2$ монет на каждой стороне весов. Если будет равновесие, то останутся $(3^{n-1} - 1)/2$ монет, которым нужно уделить внимание, и теперь уже имеются в распоряжении стандартные монеты. Если весы будут находиться не в равновесии, то следует продолжить, используя стратегию пункта (а)*).

3.20. (а) $H_{L|L}(U) = \frac{1}{L}[H(U_{2L} \mid U_{2L-1} \ldots U_1) +$

$$+ H(U_{2L-1} \mid U_{2L-2} \ldots U_1) + \ldots + H(U_{L+1} \mid U_L \ldots U_1)] \geqslant \quad (1)$$

$$\geqslant H(U_{2L} \mid U_{2L-1} \ldots U_1), \quad (2)$$

*) Не участвующие в первом взвешивании $(3^{n-1} - 1)/2$ монет будут стандартными. (*Прим. ред.*)

где было использовано то, что первое слагаемое в правой части (1) является границей снизу для каждого из остальных слагаемых. Аналогично

$$H_{L|L}(U) = \frac{1}{L}\,[H(U_{2L}\,|\,U_{2L-1}\ldots U_1) + H(U_{2L-1}\ldots U_{L+1}\,|\,U_L\ldots U_1)] \leqslant$$
$$\leqslant \frac{1}{L}\,H_{L|L}(U) + \frac{L-1}{L}\,H_{L-1|L-1}(U),$$

где было использовано (2) для первого слагаемого и стационарность и (2.3.13) для второго слагаемого. Преобразуя, получаем $H_{L|L}(U) \leqslant H_{L-1|L-1}(U)$.

(б) Используя соотношение (2), будем иметь

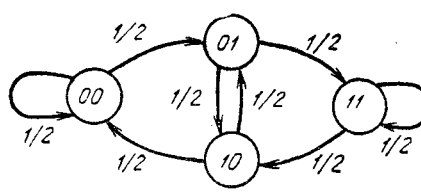
$$H_L(U) \geqslant H_{L|L}(U) \geqslant H(U_{2L}\,|\,U_{2L-1}\ldots U_1).$$

Так как эти границы стремятся к $H_\infty(U)$ при $L \to \infty$, то $H_{L|L}(U) \to H_\infty(U)$.

(в) Используя почти такое же доказательство, как в теореме 3.3.2, получаем

$$\frac{H_{L|L}(U)}{\log D} \leqslant \bar{n} \leqslant \frac{H_{L|L}(U)}{\log D} + 1.$$

3.21. При условии, что $u_{2n+1} = u_{2n}$ для всех n, пары 00 и 11 имеют каждая вероятность $1/2$. Поэтому при условии $u_{2n+1} = n_{2n}$ последовательные пары являются независимыми и число кодовых символов на букву источника стремится с ростом n к $3/4$. Подобно этому при условии, что $u_{2n-1} = u_{2n}$ для всех n, последовательные пары на позициях $2n$ и $2n+1$ являются статистически зависимыми, но они образуют марковский граф, изображенный ниже.



Стационарные вероятности состояний равны $1/4$ для каждого состояния; средняя по времени длина кодового слова равна $9/4$, а число кодовых символов на букву источника стремится к $9/8$.

3.22. (а) $q(1) = q(3) = 2/7$; $q(2) = 3/7$.

$$P(a_1) = 3/7;\ P(a_2) = P(a_3) = 2/7.$$

(б) $H(U\,|\,s=1) = 11/2$ бит, $H(U\,|\,s=2) = 1$ бит, $H(U\,|\,s=3) = 0$ бит.

(в) $H_\infty(U) = 6/7$ бит.

(г) Для состояния 1: $a_1 \to 0$, $a_2 \to 10$, $a_3 \to 11$. Для состояния 2: $a_1 \to 0$, $a_2 \to 1$; состоянию 3 не сопоставляется никакой кодовый символ. Для заданного начального состояния конец первого кодового слова можно установить, и это определит букву источника и, следовательно, следующее состояние и т. д.

(д) $\bar{n}$ равно $6/7$ бит. Имеем $\bar{n} = H_\infty(U)$, если каждый код, относящийся к текущему состоянию, имеет среднюю длину, равную энтропии данного текущего состояния. Поэтому $\bar{n} = H_\infty(U)$ тогда и только тогда, когда вероятности всех букв имеют вид 2^{-n} при некотором неотрицательном целом n.

3.23. (а). Последовательность состояний является стационарной последовательностью статистически независимых случайных величин, принимающих значения 1 и 2, и индексы букв источника совпадают с номерами состояний.

(б) $H(U_l \mid U_{l-1} \dots U_1 S_1) = H(U_{l+1} \mid U_l \dots U_2 S_2) = H(U_{l+1} \mid U_l \dots U_2 S_2 U_1 S_1) \leqslant$
$\leqslant H(U_{l+1} \mid U_l \dots U_1 S_1)$.

Второе из указанных выше равенств является результатом того, что U_1 и S_1 статистически независимы от U_{l+1} при условии, что заданы $S_2 U_2 \dots U_l$ [см. (2.3.13)]. Поэтому $H(U_l \mid U_{l-1} \dots U_1 S_1)$ — неубывающая функция l и согласно теореме 3.5.1 $H(U_l \mid U_{l-1} \dots U_1)$ — невозрастающая функция l. Из (2.3.13) также следует, что

$$H(U_l \mid U_{l-1} \dots U_1 S_1) \leqslant H(U_l \mid U_{l-1} \dots U_1) \quad \text{при всех } l.$$

Так как левая часть не убывает с ростом l, то

$$H(U_l \mid U_{l-1} \dots U_1 S_1) \leqslant \lim_{l \to \infty} H(U_l \mid U_{l-1} \dots U_1) = H_\infty(U).$$

Скорость сходимости можно оценить, если заметить, что

$$\frac{1}{l} H(U_l \dots U_1 \mid S_1) = \frac{1}{l} \sum_{i=1}^{l} H(U_i \mid U_{i-1} \dots U_1 S_1) \leqslant H(U_l \mid U_{l-1} \dots U_1 S_1).$$

Вместе с тем

$$H(U_l \dots U_1) - H(U_l \dots U_1 \mid S_1) = I(S_1;\ U_1 \dots U_l) \leqslant H(S_1) \leqslant \log J,$$

где J является числом состояний. Объединяя эти соотношения и учитывая (3.5.3), получаем

$$H(U_l \mid U_{l-1} \dots U_1) - H(U_l \mid U_{l-1} \dots U_1 S_1) \leqslant \frac{\log J}{l}.$$

4.1. Пусть $Q_N(\mathbf{x})$ является совместным распределением вероятностей N входных символов канала. Для любого такого распределения имеем

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) = H(\mathbf{Y}^N) - H(\mathbf{Y}^N \mid \mathbf{X}^N).$$

Заметим, что $H(\mathbf{Y}^N \mid \mathbf{X}^N)$ является математическим ожиданием величины

$$-\log P(\mathbf{y} \mid \mathbf{x}) = -\sum_{n=1}^{N} \log P(y_n \mid x_n)$$

и поэтому

$$H(\mathbf{Y}^N \mid \mathbf{X}^N) = \sum_{n=1}^{N} H(Y_n \mid X_n).$$

Из (2.3.10) также следует, что

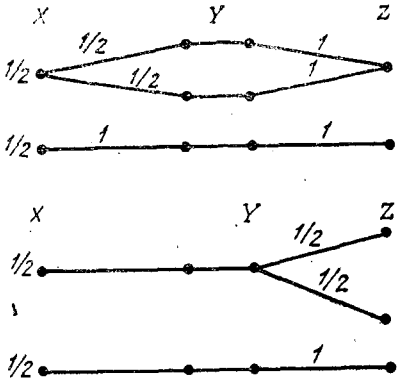
$$H(\mathbf{Y}^N) \leqslant \sum_{n=1}^{N} H(Y_n).$$

Из этих соотношений получаем

$$I(\mathbf{X}^N;\ \mathbf{Y}^N) \leqslant \sum_{n=1}^{N} I(X_n;\ Y_n) \leqslant \sum_{n=1}^{N} C_n. \tag{2}$$

Равенство в (1) имеет место, если входы являются статистически независимыми, и в (2), если, кроме того, входные вероятности выбираются так, чтобы достигалась пропускная способность отдельных каналов. Таким образом, пропускная способность параллельного соединения равна ΣC_n.

4.2



$H(Y) > H(Z)$

$H(Y) < H(Z)$

4.3. Единственным неочевидным соотношением является последнее. Имеем

$$H(U \mid Ve) = \sum_{v} \sum_{u \neq v} \Pr(u, v \mid e) \log \frac{1}{\Pr(u \mid ve)} =$$

$$= \sum_{v} \Pr(v \mid e) \sum_{u \neq v} \Pr(u \mid ve) \log \frac{1}{\Pr(u \mid ve)}.$$

При любом выборе v последняя сумма является энтропией для алфавита $M - 1$ значений u, неравных v, и, таким образом, последняя сумма не более чем $\log(M - 1)$ для любого v. Суммируя по v, получаем

$$H(U \mid V_e) \leqslant \log(M - 1).$$

4.4. Применяя (4.3.22) к источнику и каналу этой задачи, получаем

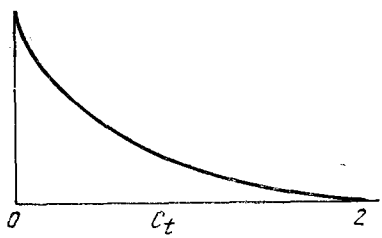
$$\mathscr{H}(\langle P_e \rangle) \geqslant 1 - C = \mathscr{H}(\varepsilon).$$

Это означает, что

$$\varepsilon \leqslant \langle P_e \rangle \leqslant 1 - \varepsilon.$$

Если бы некоторая стратегия давала вероятность ошибки, большую чем $1 - \varepsilon$, то все решения при этой стратегии могли бы быть обращены, что привело бы к вероятности ошибки, меньшей чем ε. На обычном языке студентов это означает, что также трудно дать все неправильные ответы при контрольном испытании по системе «да—нет», как дать все правильные ответы. Без кодирования $\langle P_e \rangle = \varepsilon$; это значит, что нельзя сделать ничего лучшего, чем просто передавать символы без изменения.

4.5. (а) $\langle P_e \rangle \log 3 + \mathscr{H}(\langle P_e \rangle) \geqslant 2 - C_t.$ (1)



(б) $P(j \mid k) = \begin{cases} \varepsilon/3 & \text{при } j \neq k, \\ 1-\varepsilon & \text{при } j = k, \end{cases}$

где $\varepsilon \leqslant 3/4$ удовлетворяет равенству $\varepsilon \log 3 + \mathcal{H}(\varepsilon) = 2 - C_t$.

(в) $\langle P_e \rangle \log (M-1) + \mathcal{H}(\langle P_e \rangle) \geqslant 1$,

$$\log (M-1) \geqslant \frac{1 - 2 \cdot 10^{-5}}{10^{-6}}, \quad M \geqslant 1 + 2^{(10^6 - 20)}.$$

4.6. Первые два соотношения, содержащиеся в доказательстве, получаются непосредственно. Далее

$$I(\mathrm{U}X_n;\ Y_n \mid Y_1 \ldots Y_{n-1}) = I(X_n;\ Y_n \mid Y_1 \ldots Y_{n-1}) + \\ + I(\mathrm{U};\ Y_n \mid X_n Y_1 \ldots Y_{n-1}).$$

В ДКБП, однако, выход канала в любой момент времени зависит только от входа в этот момент, что дает

$$P(y_n \mid x_n, y_1, \ldots, y_{n-1}, \mathrm{u}) = P(y_n \mid x_n). \tag{1}$$

С чисто математической точки зрения это включает в себя дополнительное предположение, что последовательность источника *связана* с последовательностью, которая выдается адресату с помощью N-кратного использования канала

$$P(\mathbf{y} \mid \mathbf{x}, \mathbf{u}) = P(\mathbf{y} \mid \mathbf{x}) = \prod_{n=1}^{N} P(y_n \mid x_n).$$

Просуммируем обе части этого равенства по $y_{n+1} \ldots y_N$. Разделив результат на $P(y_1, \ldots, y_{n-1} \mid \mathbf{x}, \mathbf{u})$, получим (1). Из теоремы 2.3.3 и (1) следует, что $I(\mathrm{U}; Y_n \mid X_n Y_1 \ldots Y_{n-1}) = 0$. Для завершения доказательства воспользуемся тем, что

$$I(X_n;\ Y_n \mid Y_1 \ldots Y_{n-1}) = H(Y_n \mid Y_1 \ldots Y_{n-1}) - H(Y_n \mid X_n Y_1 \ldots Y_{n-1}).$$

Поскольку канал является каналом без памяти, то имеем

$$H(Y_n \mid X_n Y_1 \ldots Y_{n-1}) = H(Y_n \mid X_n)$$

и

$$I(X_n;\ Y_n \mid Y_1 \ldots Y_{n-1}) \leqslant H(Y_n) - H(Y_n \mid X_n) = I(X_n;\ Y_n).$$

Если выход канала $\mathbf{V}^N$ преобразуется в последовательность $\mathbf{V}$, которая поступает к адресату, то

$$I(\mathrm{U};\ \mathbf{V}) \leqslant I(\mathrm{U};\ Y^N) \leqslant \sum_{n=1}^{N} I(X_n;\ Y_n) \leqslant NC.$$

Обращение теоремы кодирования получается теперь точно таким же образом, как это было для каналов без обратной связи.

4.7. Для ансамбля W с вероятностями q_i имеем

$$H(W) = \sum_i - q_i \ln q_i \geqslant \sum_i - q_i (q_i - 1) \geqslant \sum_i q_i (1 - q_{max}) = 1 - q_{max}.$$

Применяя этот результат к $H(U \mid v)$ и замечая, что $P_e(v)$ равно $1 - q_{max}$, будем иметь $H(U \mid v) \geqslant P_e(v)$. После умножения на $P_V(v)$ и суммирования по v, получим $H(U \mid V) \geqslant P_e$. Для получения более сильного результата предположим сначала, что $q_{max} \geqslant 1/2$. При заданном $q_{max} \geqslant 1/2$ величина $H(W)$ достигает минимума по q_i при $q_1 = q_{max}$, $q_2 = 1 - q_{max}$, так что

$$H(W) \geqslant -q_{max} \log_2 q_{max} - (1 - q_{max}) \log_2 (1 - q_{max}).$$

Выражение, стоящее справа, равно $2(1 - q_{max})$ при $q_{max} = 1/2$ и $q_{max} = 1$. В силу выпуклости оно всюду ограничено снизу $2(1 - q_{max})$. В случае, когда

$q_{max} < 1/2$, имеем

$$H_i(W) = \sum_i q_i \log \frac{1}{q_i} \geqslant \sum_i q_i \log \frac{1}{q_{max}} = -\log q_{max}.$$

Это выражение равно $2(1-q_{max})$ в точке $q_{max} = 1/2$, но имеет производную $-(\log e)/q_{max} < -2$ и, таким образом, превосходит $2(1-q_{max})$ при $q_{max} < 1/2$.

4.8. $$\frac{1}{L}\sum_{l=1}^{L} \mathcal{H}(P_{e,l}) = -\frac{1}{L}\sum_l [P_{e,l}\log P_{e,l} + (1-P_{e,l})\log(1-P_{e,l})],$$

$$\mathcal{H}(P_e) = -P_e \log P_e - (1-P_e)\log(1-P_e) =$$

$$= -\frac{1}{L}\sum_l [P_{e,l}\log P_e + (1-P_{e,l})\log(1-P_e)],$$

$$\frac{1}{L}\sum \mathcal{H}(P_{e,l}) - \mathcal{H}(P_e) = \frac{1}{L}\sum_l P_{e,l}\log\frac{P_e}{P_{e,l}} +$$

$$+(1-P_{e,l})\log\frac{1-P_e}{1-P_{e,l}} < \frac{\log e}{L}\sum_l \left\{P_{e,l}\left(\frac{P_e}{P_{e,l}} - 1\right) + \right.$$

$$\left. +(1-P_{e,l})\left[\frac{1-P_e}{1-P_{e,l}}\right]\right\} = 0.$$

4.9. Эта задача сформулирована не очень аккуратно, но ее цель показать, что если $d^2f(\alpha)/d\alpha^2 \leqslant 0$ на некотором интервале, то $f(\alpha)$ является выпуклой $\cap$ на этом интервале. Для любых α и β из этого интервала и любого Θ, $0 \leqslant \Theta \leqslant 1$, пусть $\delta = \Theta\alpha + (1-\Theta)\beta$. Используя разложение в ряд Тейлора в окрестности δ, будем иметь

$$f(\alpha) = f(\delta) + (\alpha-\delta)f'(\delta) + \frac{(\alpha-\delta)^2}{2} f''(y).$$

при некотором y между δ и α. Так как $f''(y) \leqslant 0$, то

$$f(\alpha) \leqslant f(\delta) + (\alpha-\delta)f'(\delta). \tag{1}$$

Аналогично

$$f(\beta) \leqslant f(\delta) + (\beta-\delta)f'(\delta). \tag{2}$$

Имеем

$$\Theta f(\alpha) + (1-\Theta) f(\beta) \leqslant f(\delta) + [\Theta(\alpha-\delta) + (1-\Theta)(\beta-\delta)] f'(\delta) = f(\delta)$$

в силу того, что $\Theta\alpha - \Theta\delta + (1-\Theta)\beta - (1-\Theta)\delta = \delta - \delta = 0$. Если вторая производная строго отрицательна, то неравенство (1) является строгим и выпуклость является строгой. В обозначениях рис. 4.4.2 (1) и (2) связывают ординаты концевых точек кривой и величины отрезков прямых $\Theta = 0$ и 1, отсекаемых касательной к кривой в точке $\Theta = \delta$.

Равенство (4.4.5) тривиально при $L = 1$ и равносильно определению выпуклости при $L = 2$. Предположим теперь, что (4.4.5) справедливо для $L \leqslant n-1$; покажем, что оно справедливо для $L = n$, и по индукции это завершит доказательство. Имеем

$$\sum_{i=1}^{n} \Theta_i f(\alpha_i) = (1-\Theta_n)\sum_{i=1}^{n-1} \frac{\Theta_i}{1-\Theta_n} f(\alpha_i) + \Theta_n f(\alpha_n) \leqslant$$

$$\leqslant (1-\Theta_n) f\left(\sum_{i=1}^{n} \frac{\Theta_i}{1-\Theta_n}\alpha_i\right) + \Theta_n f(\alpha_n) \leqslant \tag{3}$$

$$\leqslant f\left(\sum_{i=1}^{n} \Theta_i \alpha_i\right). \tag{4}$$

Неравенство (3) следует из справедливости (4.4.5) для $L = n - 1$ и того, что сумма $\Theta_i/(1 - \Theta_n)$ по $1 \leqslant i \leqslant n - 1$ равна 1. Неравенство (4) следует из определения выпуклости.

4.10. **(а)** Так как $0 \leqslant \lambda \leqslant 1$, то $\lambda Q_1(x) + (1-\lambda) Q_2(x) \geqslant 0$ для любого x из выборочного пространства. Также имеем

$$\sum_x [\lambda Q_1(x) + (1-\lambda) Q_2(x)] = \lambda \sum_x Q_1(x) + (1-\lambda) \sum_x Q_2(x) = \lambda + (1-\lambda) = 1.$$

(б)
$$H(X) = -\sum_x Q(x) \log Q(x) =$$
$$= -\sum_x [\lambda Q_1(x) \log Q(x) + (1-\lambda) Q_2(x) \log Q(x)],$$
$$\lambda H_1(X) + (1-\lambda) H_2(X) - H(X) = \lambda \sum_x Q_1(x) \log \frac{Q_1(x)}{Q(x)} +$$
$$+ (1-\lambda) \sum_x Q_2(x) \log \frac{Q_2(x)}{Q(x)}.$$

Используя неравенство $\log z \leqslant (\log e)(z - 1)$, можно заметить, что правая часть написанного выше равенства отрицательна.

(в) Этот результат утверждает, что энтропия является выпуклой $\cap$ функцией распределения вероятности.

4.11. Утверждение, что $f(\lambda\alpha_1 + (1-\lambda)\alpha_2)$ является выпуклой $\cap$ по λ, $0 \leqslant \lambda \leqslant 1$, означает, что для любых λ_1 и λ_2 из единичного интервала и любого Θ из единичного интервала справедливо соотношение

$$\Theta f[\lambda_1\alpha_1 + (1-\lambda_1)\alpha_2] + (1-\Theta) f[\lambda_2\alpha_1 + (1-\lambda_2)\alpha_2] \leqslant$$
$$\leqslant f\{\Theta[\lambda_1\alpha_1 + (1-\lambda_1)\alpha_2] + (1-\Theta)[\lambda_2\alpha_1 + (1-\lambda_2)\alpha_2]\}.$$

Если f является выпуклой в области R и α_1 и α_2 принадлежит R, то $\lambda_1\alpha_1 + (1-\lambda_1)\alpha_2$ и $\lambda_2\alpha_1 + (1-\lambda_2)\alpha_2$ принадлежит R и написанное выше неравенство удовлетворяется. Вместе с тем, если $f[\lambda\alpha_1 + (1-\lambda)\alpha_2]$ является выпуклой $\cap$ по λ, то написанное выше неравенство удовлетворяется при $\lambda_1 = 1$, $\lambda_2 = 0$ и это является определением того, что $f(\alpha)$ является выпуклой $\cap$ по α.

4.12. Пусть $\alpha_1 < \alpha_2$ и пусть $\varepsilon > 0$ являются произвольными. Так как $f(\alpha) = \sup f_i(\alpha)$, то существует i, для которого $f_i(\alpha_2) > f(\alpha_2) - \varepsilon$. Поскольку $f_i(\alpha_1) \cap f_i(\alpha_2)$, то будем иметь $f(\alpha_1) \geqslant f_i(\alpha_1) > f_i(\alpha_2) > f(\alpha_2) - \varepsilon$. Так как ε может быть произвольно малым, то $f(\alpha_1) \geqslant f(\alpha_2)$. Заметим, что $f(\alpha)$ не должна быть обязательно строго убывающей по α, как показывает пример $f_i(\alpha) = e^{-\alpha/i}$, где множеством индексов является множество положительных целых чисел. Здесь $f(\alpha) = 1$ при $\alpha \geqslant 0$.

Для произвольных α_1, α_2 и Θ, $0 < \Theta < 1$, и произвольного $\varepsilon > 0$ существует такое i, что

$$f_i[\Theta\alpha_1 + (1-\Theta)\alpha_2] > f[\Theta\alpha_1 + (1-\Theta)\alpha_2] - \varepsilon.$$

Поэтому

$$\Theta f(\alpha_1) + (1-\Theta) f(\alpha_2) \geqslant \Theta f_i(\alpha_1) + (1-\Theta) f_i(\alpha_2) > f[\Theta\alpha_1 + (1-\Theta)\alpha_2] - \varepsilon.$$

Так как $\varepsilon > 0$ произвольно, то функция $f(\alpha)$ выпуклая $\cup$. Выпуклость не обязательно является строгой даже тогда, когда каждая из $f_i(\alpha)$ является строго выпуклой; это можно опять показать с помощью $f_i(\alpha) = e^{-\alpha/i}$.

4.13. **(а)**
$$f[\Theta\alpha + (1-\Theta)\beta] = [\Theta\alpha_1 + (1-\Theta)\beta_1][2 - \Theta\alpha_1 - (1-\Theta)\beta_1] -$$
$$- [\Theta\alpha_2 + (1-\Theta)\beta_2]^2,$$
$$\frac{d^2 f[\Theta\alpha + (1-\Theta)\beta]}{d\Theta^2} = -2[\alpha_1 - \beta_1]^2 - 2[\alpha_2 - \beta_2]^2 < 0.$$

Используя задачу 4.11 и соотношение (4.4.4), можно показать, что из полученного соотношения следует выпуклость $\cap$ функции $f(\alpha)$.

(б) Для этой частной функции и области можно провести максимизацию по α_1 и α_2 раздельно. Максимум по α_1 достигается при $\alpha_1 = 1$. Поскольку $-(\alpha_2 + 1)^2$ является убывающей функцией α_2 при $\alpha_2 \geqslant 0$, то максимум по α_2 достигается при $\alpha_2 = 0$. Таким образом, $f(\alpha)$ достигает максимума (по $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$) при $\alpha = (1, 0)$ и $\max f(\alpha) = 0$. Можно проверить, что это решение удовлетворяет (4.4.8) и (4.4.9).

4.14. (а) Рассмотрим C как функцию долей мощностей $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_L)$:

$$C(\alpha) = \sum_{l=1}^{L} \frac{1}{2} \log\left(1 + \frac{\alpha_l S}{N_l}\right).$$

Беря вторую производную от $\log\left(1 + \frac{\alpha_l S}{N_l}\right)$ по α_l, можно увидеть, что эта функция является выпуклой $\cap$ по α_l и, следовательно, является выпуклой в области определения α. Таким образом, $C(\alpha)$ является выпуклой $\cap$ по α согласно свойству 1, изложенному на стр. 101. Область, на которой нужно найти максимум $C(\alpha)$, задается условиями $\alpha_l \geqslant 0$, $\Sigma\, \alpha_l = 1$, которые формально определяют область, в которой α является вектором вероятностей. Поэтому согласно теореме 4.4.1 необходимое и достаточное условие того, что максимум достигается на α, состоит в том, что для некоторого λ

$$\frac{\partial C(\alpha)}{\partial \alpha_l} \begin{cases} = \lambda & \text{при всех } l, \text{ для которых } \alpha_l > 0, \\ \leqslant \lambda & \text{при всех } l, \text{ для которых } \alpha_l = 0. \end{cases}$$

Беря производную, преобразуя выражение и обозначая $K = \frac{S \log e}{2\lambda}$ (эта величина является просто постоянной, которую нужно определить), получаем

$$N_l + S_l \begin{cases} = K & \text{при } S_l > 0, \\ \geqslant K & \text{при } S_l = 0. \end{cases}$$

Это значит, что $S_l = K - N_l$ при $N_l < K$ и $S_l = 0$ при $N_l \geqslant K$. Постоянная K должна быть выбрана так, чтобы $\sum S_l = S$.

(б) Простейшим способом решить систему неравенств является метод проб и ошибок. Предположим вначале, что $S_l > 0$ при всех l, что приводит к $S = \sum S_l = 3K - \sum N_l$. Отсюда получим $K = 8/3$ и $S_3 = -1/3$. Это противоречит условию, поэтому предположим теперь, что $S_3 = 0$; это приводит к $K = 5/2$ и $S_1 = 3/2$, $S_2 = 1/2$, $S_3 = 0$.

Эти значения удовлетворяют системе и, таким образом, являются максимизирующим решением.

4.15. См. Харди, Литтльвуд и Полиа (1934) или какое-либо другое пособие по неравенствам.

4.16. (а) Пусть $Q_1(k)$, $0 \leqslant k \leqslant K - 1$, и $Q_2(k)$ являются двумя распределениями вероятностей на входе канала. Пусть

$$\omega_i(j) = \sum_k Q_i(k)\, P(j \mid k); \quad i = 1, 2, \tag{1}$$

будут связанными с ними распределениями вероятностей на выходе.

Для заданного Θ, $0 < \Theta < 1$, пусть $Q(k) = \Theta Q_1(k) + (1 - \Theta)\, Q_2(k)$ и пусть

$$\omega(j) = \sum_k Q(k)\, P(j \mid k) \tag{2}$$

будет связанным с $Q(k)$ выходным распределением. Из (1) и (2) следует, что

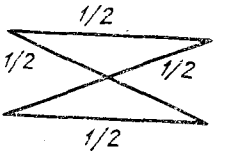
$$\omega(j) = \Theta\omega_1(j) + (1 - \Theta)\, \omega_2(j). \tag{3}$$

Нам нужно показать, что

$$\Theta H_1(Y)+(1-\Theta)\,H_2(Y) \leqslant H(Y), \tag{4}$$

где $H_1(Y)$, $H_2(Y)$ и $H(Y)$ являются энтропиями $\omega_1(j)$, $\omega_2(j)$ и $\omega(j)$ соответственно. В задаче 4.10 (б) было показано, что (4) справедливо и поэтому $H(Y)$ является выпуклой $\cap$ в области векторов вероятностей на входе.

(б) Заметим здесь, что $H(Y)$ является строго выпуклой $\cap$ по распределению вероятностей на выходе, но не является строго выпуклой по распределению вероятностей на входе. Какой-либо пример, в котором входные вероятности могут быть изменены без изменения выходных вероятностей, был бы ответом на этот пункт задачи.



(в) $$-H(Y\mid X)=\sum_k Q(k)\sum_j P(j\mid k)\log P(j\mid k).$$

Это выражение является линейным по $Q(k)$ и, следовательно, выпуклым $\cap$ по $\mathbf{Q}$.

(г) Согласно свойству 1 на стр. 101 $I(X;Y)=H(Y)-H(Y\mid X)$ является выпуклой $\cap$ по Q. Недостатком этого доказательства по сравнению с доказательством в теореме 4.4.2 является то, что оно не распространяется на недискретные каналы, в которых $H(Y)$ не определено.

4.17. Средняя взаимная информация для входного распределения вероятности $\mathbf{Q}_0$ равна

$$I_0(X;Y)=\sum Q_0(k)\,I(x=k;Y)\leqslant C.$$

Пусть $J(\mathbf{Q})$ обозначает среднюю взаимную информацию между X и Y при входном распределении $\mathbf{Q}$. Переписывая соотношение, определяющее выпуклость $\cap$ для $J(\mathbf{Q})$, будем иметь

$$J(\mathbf{Q}_1)\leqslant J(\mathbf{Q}_0)+\frac{J[\Theta\mathbf{Q}_1+(1-\Theta)\,\mathbf{Q}_0]-J(\mathbf{Q}_0)}{\Theta}.$$

Переходя к пределу при $\Theta\to 0$, получаем

$$J(\mathbf{Q}_1)\leqslant J(\mathbf{Q}_0)+\sum_k \frac{\partial J(\mathbf{Q}_0)}{\partial Q_0(k)}\,[Q_1(k)-Q_0(k)].$$

Это соотношение утверждает, что функция $J(\mathbf{Q})$ лежит ниже гиперплоскости, которая касается $J(\mathbf{Q})$ в точке $\mathbf{Q}_0$. Выполняя дифференцирование [см. (4.5.5)], получаем

$$J(\mathbf{Q}_1)\leqslant \sum_k Q_1(k)\,I_0(x=k;Y).$$

Отсюда, обозначая через $\mathbf{Q}_1$ такое $\mathbf{Q}$, на котором достигается пропускная способность, будем иметь

$$C\leqslant \max_k I_0(x=k;Y). \tag{1}$$

Наконец, для $\mathbf{Q}_0$, на котором достигается пропускная способность, согласно теореме 4.5.1 имеем

$$C=\max_k I_0(x=k;Y). \tag{2}$$

Из (1), (2) следует, что

$$C=\min_{\mathbf{Q}_0}\max_k I_0(x=k;Y).$$

4.18. (а) Пусть $Q_i(k)$ является входным распределением, на котором достигается пропускная способность i-го канала, и пусть $P_i(j\mid k)$ представляет со-

бой переходные вероятности i-го канала. Тогда взаимная информация, соответствующая k-му входу i-го канала, равна

$$I(x=(k, i); Y) = \sum_j P_i(j \mid k) \log \frac{P_i(j \mid k)}{q_i \sum_l Q_i(l) P_i(j \mid l)} =$$

$$= C_i + \log \frac{1}{q_i} \quad \text{при } Q_i(k) > 0.$$

Применяя теорему 4.5.1 к сумме каналов, получаем

$$C = C_i + \log \frac{1}{q_i} \quad \text{при всех } i, \tag{1}$$

$$q_i = 2^{C_i - C}.$$

Отсюда, используя условие $\sum q_i = 1$, выводим

$$C = \log_2 \sum_i 2^{C_i}.$$

Умножая далее обе части (1) на q_i и суммируя по i, получим

$$C = \sum_i q_i C_i + \sum_i q_i \log \frac{1}{q_i}.$$

(б) Рассматривая канал как сумму каналов со входами 0,1 от первого канала и 2 от второго, получаем

$$C_1 = 1 - \mathcal{H}(\varepsilon) \text{ [бит]}, \quad C_2 = 0,$$

$$C = \log_2 \left[1 + 2^{1 - \mathcal{H}(\varepsilon)}\right] = \log_2 \left[1 + 2\varepsilon^{\varepsilon}(1-\varepsilon)^{1-\varepsilon}\right].$$

4.19. (а) $I(X; Y) = H(Y) - H(Y \mid X)$,

$$H(Y \mid X) = \sum_k Q(k) \sum_j P(j \mid k) \log \frac{1}{P(j \mid k)}.$$

Замечая, что $P(j \mid k) = P_Z\big((j-k) \bmod K\big)$, получаем, что

$$\sum_j P(j \mid k) \log \frac{1}{P(j \mid k)} = \sum_{i=0}^{K-1} P_Z(i) \log \frac{1}{P_Z(i)} = H(Z)$$

и, таким образом, $I(X; Y) = H(Y) - H(Z)$.

(б) Пропускная способность отыскивается с помощью максимизации $H(Y)$. В силу симметрии шума по отношению к различным входам равновероятные входы дают равновероятные выходы и, следовательно, пропускную способность

$$C = \log K - H(Z).$$

4.20. Каналы а), г) и д) являются симметричными, и пропускную способность в каждом из них достигается на равновероятных входах. Небольшое усилие позволяет догадаться, что пропускная способность для каналов (б) и в) достигается на $Q(0) = Q(2) = 1/2$, $Q(1) = 0$. Пропускные способности (в битах) равны

(а) $[1-\delta]\left[1 - \mathcal{H}\left(\frac{\varepsilon}{1-\delta}\right)\right]$,

(б) 3/4,

(в) 1,

(г) $2/3 \log_2 3/2$,
(д) $\log_2 3 - \mathcal{H}(\varepsilon)$.

Канал е) дан для того, чтобы продемонстрировать трудность непосредственного вычисления пропускной способности даже для простейшего канала, у которого нет упрощающей симметрии. Можно использовать метод, представленный (4.5.9)—(4.5.12), или прямые вычисления на основе теоремы 4.5.1; оба они требуют приблизительно одинаковое количество усилий. Получим

$$C = \frac{\delta \mathcal{H}(\varepsilon) + \varepsilon \mathcal{H}(\delta)}{\rho} + \log_2 \left[2^{-\mathcal{H}(\delta)/\rho} + 2^{-\mathcal{H}(\varepsilon)/\rho}\right],$$

$$Q_0 = \frac{1}{\rho}\left[-\delta + \frac{2^{\mathcal{H}(\delta)/\rho}}{2^{\mathcal{H}(\delta)/\rho} + 2^{\mathcal{H}(\varepsilon)/\rho}}\right], \quad \rho = 1 - \varepsilon - \delta.$$

4.21. Если $\mathbf{D}[I(x; y)] = 0$, то $I(x; y)$ является постоянной, скажем, C, на тех парах x, y, для которых $Q(x) P(y \mid x) \neq 0$. Так как по предположению $Q(x) > 0$ для всех входов x, то $I(x; y) = C$ для всех x, y, для которых $P(y \mid x) \neq 0$. Поэтому $P(y \mid x) I(x; y) = P(y \mid x) C$ для всех x, y. Получаем

$$I(x = k; Y) = \sum_j P(j \mid k) I(k; j) = \sum_j P(j \mid k) C = C.$$

Теорема 4.5.1 показывает теперь, что пропускная способность достигается на $Q(k)$.

4.22. Нетрудно проверить, что теоремы 4.4.2 и 4.5.1 применимы к каналу с дискретным входом и недискретным выходом. Поэтому, используя симметрию, получаем, что пропускная способность достигается на равновероятных входах. Пусть C_1 и C_2 являются пропускными способностями каналов, в которых используются y и знак y в качестве выходов соответственно. Имеем $C_1 > C_2$ в соответствии с рассуждениями о последовательном соединении каналов на стр. 41. При рассмотрении лишь знака y теряется информация относительно того, насколько вероятным является каждый вход. Позднее при изучении кодирования будет показано, что эта информация является важной, несмотря на то, что при отсутствии кодирования она не может быть использована при приеме одиночных двоичных символов. Имеем

$$C_1 = \int_{-\infty}^{\infty} p_Z(z - \sqrt{S}) \ln \frac{p_Z(z - \sqrt{S})}{\frac{1}{2}\left[p_Z(z - \sqrt{S}) + p_Z(z + \sqrt{S})\right]} dz =$$

$$= \int_{-\infty}^{\infty} p_Z(z - \sqrt{S}) \ln \left[\frac{2}{1 + e^{-2z\sqrt{S}/\sigma^2}}\right] dz \quad \text{(натуральных единиц)}. \qquad (1)$$

Разлагая логарифм в ряд по степеням $\sqrt{S}$, получаем

$$\ln \frac{2}{1 + e^{-2z\sqrt{S}/\sigma^2}} = \frac{\sqrt{S} z}{\sigma^2} - \frac{z^2 S}{2\sigma^4} + R_z(S),$$

$$|R_z(S)| \leqslant \frac{|z^3|}{3\sigma^6} S^{3/2}.$$

Подставляя это выражение в (1) и выполняя интегрирование, будем иметь

$$C_1 = \frac{S}{2\sigma^2} + R(S),$$

где $R(S)$ ограничено сверху постоянной, умноженной на $S^{3/2}$, и стремится к нулю при $S \to 0$.

Используя знак y в качестве выхода, получаем двоичный симметричный канал с переходной вероятностью ε, которая удовлетворяет неравенствам

$$\frac{1}{2} - \sqrt{\frac{S}{2\pi\sigma^2}} \leqslant \varepsilon \leqslant \frac{1}{2} - \sqrt{\frac{S}{2\pi\sigma^2}} + \frac{S^{3/2}}{6\sqrt{2\pi}\,\sigma^3}.$$

Верхняя граница может быть получена таким же методом, как и в задаче 2.23, т. е. с помощью неравенств

$$1 \geqslant e^{-y^2/(2\sigma^2)} \geqslant 1 - \frac{y^2}{2\sigma^2}.$$

Положив $\delta = \frac{1}{2} - \varepsilon$, получаем

$$C_2 = \left(\frac{1}{2} - \delta\right) \ln \frac{1/2 - \delta}{1/2} + \left(\frac{1}{2} + \delta\right) \ln \frac{1/2 + \delta}{1/2}.$$

Разлагая с точностью до второго порядка по δ (при этом обязательно сохраняется первый порядок по S), получаем

$$C_2 = 2\delta^2 = \frac{S}{\pi\sigma^2} \text{ натуральных единиц.}$$

4.23. Для канала, изображенного на рис. 4.6.3, и для четного N пусть $Q_n(k)$ *обозначает вероятность входа* k *при* n*-м использовании канала и пусть*

$$q = \sum_{n=1}^{N/2} [Q_{2n-1}(0) + Q_{2n-1}(1) + Q_{2n}(2) + Q_{2n}(3)] \frac{1}{N}.$$

При условии, что канал начинает работать в состоянии 0, значение q является средней по последовательности N символов вероятностью использования одного из неискаженных символов (т. е. 0 или 1 при нечетных посылках и 2 или 3 при четных посылках). Аналогично $1 - q$ является такой же вероятностью при условии, что начальным состоянием было 1. При заданном q и при условии, что начальным состоянием является 0, легко понять, используя выпуклость, что максимальная средняя взаимная информация на символ равна $q \log (2/q) + (1 - q) \log [1/(1 - q)]$ и достигается при статистически независимых входах, для которых $Q_{2n-1}(0) = Q_{2n-1}(1) = Q_{2n}(2) = Q_{2n}(3) = q/2$ при $1 \leqslant n \leqslant N/2$. Аналогично при условии, что начальным состоянием является 1, максимальная взаимная средняя информация на символ равна $(1 - q) \log [2/(1 - q)] + q \log 1/q$. Максимальное значение минимума из этих двух выражений равно $^3/_2$ бит и имеет место при $q = {}^1/_2$. Чтобы убедиться в этом, заметим, что любое изменение q уменьшает либо то, либо другое выражение.

Для канала, изображенного на рис. 4.6.4, используем тот же самый метод, положив $q = \frac{1}{N} \sum_{n=1}^{N} Q_n(0)$. Тогда получим (в битах)

$$I_Q(\mathbf{X}^N; \mathbf{Y}^N \mid s_0 = 0) \leqslant N[-q \log_2 q - (1-q) \log_2 (1-q)],$$

$$I_Q(\mathbf{X}^N; \mathbf{Y}^N \mid s_0 = 1) \leqslant N[(1-q) \log_2 3].$$

Оба неравенства удовлетворяются с равенством на независимых одинаково распределенных входах с $Q(0) = q$, $Q(1) = Q(2) = Q(3) = (1 - q)/3$. Правые части неравенств равны при $q = 0{,}391$, и любое изменение q уменьшает то или другое выражение.

Для канала, изображенного на рис. 4.6.5, можно заметить, что если канал начинает работать с состояния 1, то выход не содержит информации о входе

и $\underline{C} = 0$. Если канал начинает работать с состояния 0, то можно передать один бит при одном использовании канала, если использовать входы 0 и 1 независимо и с равными вероятностями. Следовательно, $\overline{C} \geqslant 1$ бит. Так как выход является двоичным, то также $\overline{C} \leqslant 1$ бит.

4.24. Пусть $s_n = x_n$ и $y_n = s_{n-1}$. Другими словами, канал является двоичным каналом без шума с задержкой, т. е. $y_n = x_{n-1}$. Входная последовательность $\overline{\mathbf{x}} = (x_1, \ldots, x_N)$ в этом случае точно определит $y_2, \ldots, y_N$ (и y_{N+1}), но никак не связана с y_1. Поэтому $I(\mathbf{X}^N; \mathbf{Y}^N \mid s_0)$ меньше или равна $N-1$ бит с равенством, когда входы являются независимыми и равновероятными. Поэтому $\overline{C}_N = 1 - \frac{1}{N}$ и $\underline{C} = \overline{C} = 1$. Найдем теперь двоичный канал с двумя состояниями, для которого $\underline{C}_N = 1/N = \underline{C} = \overline{C}$ при всех N. Пусть последовательность состояний $\{s_n\}$ будет последовательностью независимых и равновероятных двоичных символов и пусть $y_n = x_n \oplus s_{n-1}$. Тогда $\underline{C}_N = 1/N$, так как условие, состоящее в том, что состояние в момент 0 является известным, позволяет передать один бит информации в первый момент. Очевидно, $\underline{C} = \overline{C} = 0$.

4.25. (а) Под предположением, что имеет место статистическая независимость шумов, в каналах понимается, что при заданном входе в n-й канал выход n-го канала не зависит от входов всех предыдущих каналов, т. е. $P(x_{n+1} \mid x_n) = P(x_{n+1} \mid x_n, \ldots, x_1)$. Таким образом, последовательность $x_1, x_2, \ldots$ является марковской цепью, которая в общем случае является неоднородной (за исключением случая, когда все каналы являются одинаковыми).

(б) Пусть x_1' и x_1'' — какие-либо два символа на входе канала 1. Используя лемму 4.6.2 при $n = 1$ и x_N вместо s_{N-1} получим

$$\sum_{x_N} |P_{X_N \mid X_1}(x_N \mid x_1') - P_{X_N \mid X_1}(x_N \mid x_1'')| \leqslant 2(1-\delta)^{N-2}.$$

Так как каждое слагаемое суммы, в свою очередь, ограничено правой частью неравенства и так как $P_{X_n}(x_N)$ является взвешенным средним $P(x_N \mid x_1)$ по x_1, то получим для всех x_N и x_1':

$$|P_{X_N \mid X_1}(x_N \mid x_1') - P_{X_N}(x_N)| \leqslant 2(1-\delta)^{N-2}, \tag{1}$$

$$I(X_1; X_N) = \sum_{x_1,\, x} P(x_1, x_N) \log \frac{P(x_N \mid x_1)}{P(x_N)} \leqslant$$

$$\leqslant \sum_{x_1,\, x_N} P(x_1, x_N) \left[\frac{P(x_N \mid x_1)}{P(x_N)} - 1\right] \log e =$$

$$= \sum_{x_1,\, x_N} P(x_1 \mid x_N) [P(x_N \mid x_1) - P(x_N)] \log e \leqslant$$

$$\leqslant \sum_{x_1,\, x_N} P(x_1 \mid x_N) 2(1-\delta)^{N-2} \log e = 2K(1-\delta)^{N-2} \log e.$$

Степень $N-2$ может быть заменена на $N-1$, что можно понять, просматривая доказательство леммы 4.6.2, для случая $n = 1$. Однако ее нельзя заменить на N, что можно понять, если рассмотреть случай, когда $N = 2$ и канал является двоичным стирающим каналом с вероятностью стирания, близкой к 1.

4.26. Как предлагается в указании, пусть B_n обозначает множество возможных состояний канала после n-го входа. Если $B_n = B_m$ для $m > n$ при входной последовательности $x_1, x_2, \ldots, x_N$, то входная последовательность $x_1, x_2, \ldots, x_n, x_{m+1}, x_{m+2}, \ldots, x_N$ приведет канал в известное состояние на $m-n$ моментов времени раньше, так как по предположению x_{m+1}, x_{m+2} способны

привести канал в известное состояние, начиная с неизвестного множества состояний $B_n = B_m$. Имеются 2^A различных множеств состояний. Полное множество имеется в момент 0; пустое множество невозможно и не должно быть множества, содержащего одно состояние до того, как канал окажется в известном состоянии. Таким образом, последовательность $B_1, B_2, \ldots$ может проходить не более чем $2^A - A - 2$ различных множеств до того, как она достигнет известного состояния, и поэтому известное состояние может быть достигнуто не более чем за $2^A - A - 1$ переходов.

5.1. Стоимость декодирования последовательности $\mathbf{y}$ как сообщения m' равна

$$C(m' \mid \mathbf{y}) = \sum_m C(m, m') \Pr(m \mid \mathbf{y}) =$$
$$= \sum_m C(m, m') Q(m) P_N(\mathbf{y} \mid \mathbf{x}_m)/\Pr(\mathbf{y}).$$

Так как $\Pr(\mathbf{y})$ не зависит от того, какое решение было сделано, то правилом решения с минимальной стоимостью является следующее: выбрать m', которое минимизирует

$$\sum_m C(m, m') Q(m) P_N(\mathbf{y} \mid \mathbf{x}_m).$$

5.2. (а) (I) $g_n(s) = (1-\varepsilon)^{1-s} \varepsilon^s + \varepsilon^{1-s}(1-\varepsilon)^s$,

$$\min_{0 \leqslant s \leqslant 1} g_n(s) = 2\sqrt{(1-\varepsilon)\varepsilon}, \quad \text{при } s = 1/2,$$

$$P_{e,\,m} \leqslant [4(1-\varepsilon)\varepsilon]^{N/2}.$$

(II) $g_n(s) = \varepsilon$; $\min\limits_{0 \leqslant s \leqslant 1} g_n(s) = \varepsilon$,

$$P_{e,\,m} \leqslant \varepsilon^N.$$

(III) $g_n = \varepsilon^{1-s}$; $\min\limits_{0 \leqslant s \leqslant 1} g_n(s) = \varepsilon$,

$$P_{e,\,m} \leqslant \varepsilon^N.$$

(IV) $g_n(s) = (1-\varepsilon_1-\varepsilon_2)^{1-s} \varepsilon_2^s + \varepsilon_1 + (1-\varepsilon_1-\varepsilon_2)^s \varepsilon_2^{1-s}$;

$$\min_{0 \leqslant s \leqslant 1} g_n(s) = 2\sqrt{(1-\varepsilon_1-\varepsilon_2)\varepsilon_2} + \varepsilon_1,$$

$$P_{e,\,m} \leqslant [2\sqrt{(1-\varepsilon_1-\varepsilon_2)\varepsilon_2} + \varepsilon_1]^N.$$

(б) (I) $P_{e,\,m} = \sum\limits_{i=\frac{N+1}{2}}^{N} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}$; N — нечетное.

См. пункт (в), в котором рассмотрено четное N. По формуле Стирлинга

$$\binom{N}{i} \approx \sqrt{\frac{N}{2\pi i (N-i)}} \binom{N}{i}^i \left(\frac{N}{N-i}\right)^{N-i}, \tag{1}$$

$$\binom{25}{13} \varepsilon^{13}(1-\varepsilon)^{12} = 2{,}60 \cdot 10^{-7},$$

$$\binom{25}{14} \varepsilon^{14}(1-\varepsilon)^{11} = 0{,}25 \cdot 10^{-7},$$

$$\binom{25}{15} \varepsilon^{15}(1-\varepsilon)^{10} = 0{,}02 \cdot 10^{-7},$$

$$P_{e,\,m} = 2{,}87 \cdot 10^{-7}.$$

Граница в этом случае равна

$$P_{e,\,m} \leqslant (0{,}36)^{12{,}5} = 2{,}843 \cdot 10^{-6}.$$

(II) В ДСтК, очевидно, происходит ошибка только тогда, когда все символы стираются. По условию в этом случае декодируется сообщение 2. Таким образом,

$$P_{e,\,1} = (0,\,1)^{25} = 10^{-25}, \quad P_{e,\,2} = 0.$$

Граница будет $P_{e,\,m} \leqslant 10^{-25}$, $m = 1,\ 2$.

(III) Для Z-канала сообщение 2 декодируется, если среди принятых символов имеется единица и сообщение 1 декодируется во всех остальных случаях. Следовательно,

$$P_{e,\,1} = 0, \quad P_{e,\,2} = 10^{-25}.$$

Граница имеет вид $P_{e,\,m} \leqslant 10^{-25}$, $m = 1,\ 2$.

(в) Из (5.3.14) для четного N имеем

$$P_{e,\,1} = \sum_{i=\frac{N}{2}}^{N} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}.$$

Используя $\binom{N}{i+1} = \binom{N}{i}\left(\frac{N-i}{i+1}\right)$, получаем

$$P_{e,\,1} = \binom{N}{N/2} [\varepsilon(1-\varepsilon)]^{N/2} \left\{1 + \left(\frac{N/2}{N/2+1}\right) \times \right.$$
$$\left. \times \left(\frac{\varepsilon}{1-\varepsilon}\right) + \left(\frac{(N/2)\,(N/2-1)}{(N/2+1)\,(N/2+2)}\right)\left(\frac{\varepsilon}{1-\varepsilon}\right)^2 + \ldots\right\}. \qquad (2)$$

Заметим, что при очень большом N выражения в фигурных скобках, содержащие N в отношениях в слагаемых, становятся приближенно равными 1, в то время как $\left(\frac{\varepsilon}{1-\varepsilon}\right)^i \to 0$.

Таким образом,

$$P_{e,\,1} \approx \binom{N}{N/2} [\varepsilon(1-\varepsilon)]^{N/2} \left\{1 + \frac{\varepsilon}{1-\varepsilon} + \left(\frac{\varepsilon}{1-\varepsilon}\right)^2 + \ldots\right\}. \qquad (3)$$

Выражение в фигурных скобках равно $\dfrac{1}{1-\varepsilon/(1-\varepsilon)} = \dfrac{1-\varepsilon}{1-2\varepsilon}$.

С помощью (1) теперь получим

$$P_{e,\,1} \approx \sqrt{\frac{2}{\pi N}}\left(\frac{1-\varepsilon}{1-2\varepsilon}\right)[2\sqrt{\varepsilon(1-\varepsilon)}]^N.$$

Для $P_{e,\,2}$ имеем ту же самую сумму, за исключением того, что суммирование начинается с $i = \frac{N}{2} + 1$ вместо $\frac{N}{2}$. Поэтому для $P_{e,\,2}$ равенства (2) и (3) нужно модифицировать, выбросив первое слагаемое из суммы в фигурных скобках. Это равносильно умножению каждого слагаемого на $\left(\frac{\varepsilon}{1-\varepsilon}\right)$, так что $P_{e,\,2} = \frac{\varepsilon}{1-\varepsilon} P_{e,\,1}$. Если N нечетно, то получаем

$$P_{e,\,m} = \binom{N}{(N+1)/2} \varepsilon^{(N+1)/2} (1-\varepsilon)^{(N-1)/2} \left\{1 + \right.$$

$$+\frac{(N-1)/2}{(N+3)/2}\,\frac{\varepsilon}{1-\varepsilon}+\frac{(N-1)(N-3)}{(N+3)(N+5)}\left(\frac{\varepsilon}{1-\varepsilon}\right)^2+\ldots\Bigg\}.$$

С помощью того же приближения, что и раньше, получим

$$P_{e,\,m}\approx\left[\sqrt{\frac{2}{\pi N}}\,2^N\right]\sqrt{\frac{\varepsilon}{1-\varepsilon}}\,[\varepsilon(1-\varepsilon)]^{N/2}\left\{\frac{1}{1-\dfrac{\varepsilon}{1-\varepsilon}}\right\}.$$

(г) В Z-канале при $\mathbf{x}_1=(0,\,0,\,\ldots,\,0,\,1,\,\ldots,\,1)$ и $\mathbf{x}_2=(1,\,\ldots,\,1,\,0,\,\ldots,\,0)$ граница (5.3.1) приобретает вид

$$P_{e,\,m}\leqslant\min_{0\leqslant s\leqslant 1}\{\varepsilon^{(1-s)N/2},\,\varepsilon^{sN/2}\}=\varepsilon^{N/2}.$$

Отметим, что декодер всегда производит правильное декодирование, если принята какая-нибудь единица, так как принятая в какой-либо момент единица означает, что единица была передана в тот же самый момент. Если были приняты все нули, то декодер декодирует сообщение 2 (по предположению) так, что $P_{e,\,1}=\varepsilon^{N/2}$, $P_{e,\,2}=0$. Это *заметное изменение в вероятности ошибки*, возникающее из-за такого невинного изменения в кодовых словах, является довольно удивительным и оказывается очень важным явлением при отыскании нижних границ вероятности ошибки.

5.3. (а) (I) $\overline{P}_{e,\,m}\leqslant\{[Q(0)\sqrt{1-\varepsilon}+Q(1)\sqrt{\varepsilon}\,]^2+$
$+[Q(1)\sqrt{1-\varepsilon}+Q(0)\sqrt{\varepsilon}\,]^2\}^N.$

Подставляя $1-Q(0)$ вместо $Q(1)$ и минимизируя выражение, стоящее в фигурных скобках по $Q(0)$, находим, что минимум достигается при $Q(0)=\frac{1}{2}$, что следовало ожидать в силу симметрии. При $Q(0)=1/2$ имеем

$$\overline{P}_{e,\,m}\leqslant\left\{\frac{1}{2}\,[\sqrt{1-\varepsilon}+\sqrt{\varepsilon}\,]^2\right\}^N=\left[\frac{1}{2}+\sqrt{\varepsilon(1-\varepsilon)}\right]^N.$$

(II) $\overline{P}_{e,\,m}\leqslant\{(1-\varepsilon)[Q^2(0)+Q^2(1)]+\varepsilon\}^N.$

Минимум по $\mathbf{Q}$ достигается при $Q(0)=1/2$, что дает

$$\overline{P}_{e,\,m}\leqslant\left[\frac{1+\varepsilon}{2}\right]^N.$$

(III) $\overline{P}_{e,\,m}\leqslant\{[Q(0)+\sqrt{\varepsilon}\,Q(1)]^2+(1-\varepsilon)\,Q^2(1)\}^N.$

Минимум по Q достигается при $Q(0)=1/2$. Это довольно удивительно, но в задаче 5.13 будет показано, что этот минимум всегда достигается при $Q(0)=1/2$ для любых каналов, двоичных по входу. При $Q(0)=1/2$ имеем

$$\overline{P}_{e,\,m}\leqslant\left\{\frac{1}{4}(1+\sqrt{\varepsilon}\,)^2+\frac{1}{4}(1-\varepsilon)\right\}^N=\left\{\frac{1}{2}(1+\sqrt{\varepsilon}\,)\right\}^N.$$

(IV) $\overline{P}_{e,\,m}\leqslant\{[Q(0)\sqrt{1-\varepsilon_1-\varepsilon_2}+Q(1)\sqrt{\varepsilon_2}\,]^2+$
$+\varepsilon_1+[Q(1)\sqrt{1-\varepsilon_1-\varepsilon_2}+Q(0)\sqrt{\varepsilon_2}\,]^2\}^N.$

Минимум достигается при $Q(0)=1/2$, что дает

$$\overline{P}_{e,\,m}\leqslant\left\{\frac{1}{2}(\sqrt{1-\varepsilon_1-\varepsilon_2}+\sqrt{\varepsilon_2}\,)^2+\varepsilon_1\right\}^N=$$
$$=\{(1+\varepsilon_1)/2+\sqrt{\varepsilon_2(1-\varepsilon_1-\varepsilon_2)}\}^N.$$

(б) Для выборочного кода из ансамбля можно использовать (5.3.8) и (5.3.7) и получить, что

$$\frac{1}{N}\ln P_{e,m} \leqslant \frac{1}{N}\ln\prod_{n=1}^{N} g_n(s) = \frac{1}{N}\sum_{n=1}^{N}\ln g_n(s) =$$

$$= \frac{1}{N}\sum_{n=1}^{N}\ln\left\{\sum_{y_n} P(y_n \mid x_{1,n})^{1-s} P(y_n \mid x_{2,n})^{s}\right\}. \qquad (1)$$

Так как $x_{1,n}$ и $x_{2,n}$ при каждом n являются независимыми случайными величинами, каждая из которых принимает значения 0 и 1 с вероятностями 1/2, то можно усреднить (1) по ансамблю и получить

$$\frac{1}{N}\overline{\ln P_{e,m}} \leqslant \sum_{k=0}^{1}\sum_{i=0}^{1}\frac{1}{4}\ln\left\{\sum_{j} P(j \mid k)^{1-s} P(j \mid i)^{s}\right\}.$$

Замечая, что при $k = i$ выражение в фигурных скобках равно 1, получим

$$\frac{1}{N}\overline{\ln P_{e,m}} \leqslant \frac{1}{4}\ln\sum_{j} P(j \mid 0)^{1-s} P(j \mid 1)^{s} +$$

$$+ \frac{1}{4}\ln\left\{\sum_{j} P(j \mid 1)^{1-s} P(j \mid 0)^{s}\right\}.$$

Правая часть достигает минимума при $s = 1/2$, давая наиболее точную границу такого вида. Имеем

$$\frac{1}{N}\overline{\ln P_{e,m}} \leqslant \frac{1}{2}\ln\sum_{j}\sqrt{P(j \mid 0)\,P(j \mid 1)}. \qquad (2)$$

Таким образом, для четырех указанных каналов правая часть (2) равна

(I) $\frac{1}{4}\ln[4\varepsilon(1-\varepsilon)]$, (III) $\frac{1}{4}\ln\varepsilon$,

(II) $\frac{1}{2}\ln\varepsilon$, (IV) $\frac{1}{2}\ln\left[2\sqrt{(1-\varepsilon_1-\varepsilon_2)\,\varepsilon_2}+\varepsilon_1\right]$.

Согласно закону больших чисел, правая часть (1) (при $s = 1/2$) при больших N близка к (2) почти для всех кодов из ансамбля, так что для большинства кодов из ансамбля

$$P_{e,m} \lesssim \exp\left\{N\cdot\frac{1}{2}\ln\sum_{j}\sqrt{P(j \mid 0)\,P(j \mid 1)}\right\}.$$

Заметим, что для двоичного симметричного канала это согласуется с (5.5.1).

5.4. Пусть задан ансамбль кодов и на кодер поступает сообщение m и пусть $E_{m\to m'}$ является событием, состоящим в том, что $P_N(y \mid x_{m'}) \geqslant P_N(y \mid x_m)$. В силу того, что ошибочное декодирование может произойти только тогда, когда для некоторого $m' \neq m$ происходит $E_{m\to m'}$, то имеем

$$\overline{P}_{e,m} \leqslant \Pr\left[\bigcup_{m'\neq m} E_{m\to m'}\right] \leqslant \sum_{m'\neq m}\Pr\left[E_{m\to m'}\right].$$

Из (5.5.10) находим, что

$$\Pr\left[E_{m\to m'}\right] \leqslant \left\{\sum_{j}\left[\sum_{k} Q(k)\sqrt{P(j \mid k)}\right]^{2}\right\}^{N} \text{ при каждом } m' \neq m.$$

Таким образом, суммируя по $M-1$ возможным значениям $m' \neq m$, получаем

$$\overline{P}_{e,m} \leqslant (M-1)\left\{\sum_j\left[\sum_k Q(k)\sqrt{P(j|k)}\right]^2\right\}^N.$$

Используя неравенство $M-1<M=e^{NR}$, можно переписать это в виде

$$\overline{P}_{e,m} \leqslant \exp\left\{N\left[R+\ln\sum_j\left(\sum_k Q(k)\sqrt{P(j|k)}\right)^2\right]\right\}.$$

Следовательно, граница экспоненциально убывает по N при

$$R<-\ln\sum_j\left[\sum_k Q(k)\sqrt{P(j|k)}\right]^2. \tag{1}$$

Положив $Q(0)=1/2$, получим, что правые части (1) для каждого из каналов, рассмотренных в задаче 5.2, равны

(I) $-\ln\left[\frac{1}{2}+\sqrt{\varepsilon(1-\varepsilon)}\right]$, (III) $-\ln\frac{1+\sqrt{\varepsilon}}{2}$;

(II) $-\ln\frac{1+\varepsilon}{2}$, (IV) $-\ln\left[\frac{1+\varepsilon}{2}+\sqrt{\varepsilon_2(1-\varepsilon_1-\varepsilon_2)}\right]$.

5.5. (а) Здесь z является гауссовской случайной величиной с нулевым средним значением и дисперсией N. Поэтому

$$\int_N^\infty \frac{1}{\sqrt{2\pi N}}e^{-\frac{z^2}{2N}}dz \approx \frac{1}{\sqrt{2\pi N}}e^{-N/2}. \tag{1}$$

Границей Чернова для этого случая будет

$$\Pr[z\geqslant N] \leqslant e^{-sN}\int p_z(z)\,e^{sz}\,dz = e^{-sN+s^2N/2}.$$

После минимизации по $s\geqslant 0$ получаем наиболее точную границу такого вида при $s=1$

$$\Pr[z\geqslant N] \leqslant e^{-N/2}.$$

Неравенство Чебышева дает

$$\Pr[z\geqslant N] \leqslant \Pr[|z|\geqslant N] \leqslant \frac{N}{N^2}=\frac{1}{N}. \tag{2}$$

Можно заметить, что граница Чернова имеет правильную экспоненциальную зависимость от N и что неравенство Чебышева является здесь довольно слабым.

(б) Имеем $z\geqslant N$ только тогда, когда $x_n=1$ при всех n, $1\leqslant n\leqslant N$, и, таким образом, $\Pr[z\geqslant N]=2^{-N}$.

Границей Чернова будет

$$\Pr[z\geqslant N] \leqslant e^{-sN}\left[\frac{1}{2}e^s+\frac{1}{2}e^{-s}\right]^N = 2^{-N}(1+e^{-2s})^N.$$

Минимум по $s\geqslant 0$ здесь достигается при $s=\infty$, что дает $\Pr[z\geqslant N]\leqslant 2^{-N}$. Приближение с помощью центральной предельной теоремы дается правой частью (1). Заметим, что центральная предельная теорема утверждает, что функция распределения z сходится к гауссовской функции распределения, но она ничего

не говорит о том, какова доля ошибок на хвостах. Отличие между выражением (1) и 2^{-N} исчезает, но экспоненциальная зависимость от N является неверной. Неравенство Чебышева задается (2) и является еще худшим приближением.

5.6. В соответствии с неравенством Гёльдера [см. задачу 4.15 (в)], полагая $\lambda = 1/(1+\rho)$, получаем

$$\left[\sum_i Q_i a_i b_i\right]^{1+\rho} \leqslant \left(\sum Q_i a_i^{1+\rho}\right)\left(\sum Q_i b_i^{(1+\rho)/\rho}\right).$$

Чтобы согласовать это с соотношением, содержащимся в указании, положим

$$a_i^{1+\rho} = P_N(\mathbf{y}\mid\mathbf{x})^{1-s\rho}, \qquad b_i^{(1+\rho)/\rho} = P_N(\mathbf{y}\mid\mathbf{x})^{s}.$$

Это значит, что $a_i b_i = P_N(\mathbf{y}\mid\mathbf{x}_m)^{1/(1+\rho)}$ и поэтому

$$\left[\sum_{\mathbf{x}} Q_N(\mathbf{x})\, P_N(\mathbf{y}\mid\mathbf{x})^{1/(1+\rho)}\right]^{1+\rho} \leqslant \left[\sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m)\, P_N(\mathbf{y}\mid\mathbf{x}_m)^{1-s\rho}\right]\times$$
$$\times\left[\sum_{\mathbf{x}} Q_N(\mathbf{x})\, P_N(\mathbf{y}\mid\mathbf{x})^{s}\right]^{\rho}.$$

Так как левая часть равна правой части при $s = 1/(1+\rho)$, то получаем, что правая часть достигает минимума при $s = 1/(1+\rho)$. Это показывает, что каждое слагаемое в сумме по $\mathbf{y}$ в (5.6.10) достигает минимума по s при $s = 1/(1+\rho)$.

Применение неравенства Гёльдера здесь довольно типично. Сначала возникает желание построить доказательство и затем используется неравенство Гёльдера, чтобы выполнить это доказательство.

5.7. (а) Имеются две входные буквы, которые приводят к каждой из выходных букв в заданном канале, и, следовательно, имеются 2^N последовательностей на входе, приводящих к любой заданной последовательности на выходе. Заметим, что $P_N(\mathbf{y}\mid\mathbf{x}) = 2^{-N}$ является одной и той же величиной для всех таких $\mathbf{x}$. Вероятность того, что $\mathbf{x}_2$ принадлежит этому множеству 2^N последовательностей, равна $2^N/3^N$ или $(2/3)^N$.

(б) Для M кодовых слов можно применить аддитивную границу к результату, полученному в пункте (а); будем иметь

$$\overline{P_{e,m}} \leqslant (M-1)\,(2/3)^N. \tag{1}$$

Согласно (5.6.11) $\overline{P_{e,m}} \leqslant (M-1)^{\rho}\,(2/3)^{\rho N}$. Минимум правой части по ρ, $0 \leqslant \rho \leqslant 1$, достигается при $\rho = 1$, что приводит к тому же самому результату, что и в (1).

(в) При $\mathbf{x}_1 = (0,0,\ldots,0)$, $\mathbf{x}_2 = (1,1,\ldots,1)$ ошибка происходит только тогда, когда принимается (1, 1, ..., 1). Поэтому $P_{e,1} = 2^{-N}$, $P_{e,2} = 0$. Отличие возникает потому, что этот код является наилучшим кодом из двух кодовых слов, а (1) получено для ансамбля кодов.

5.8. (а) Применение формулы Стирлинга дает

$$\binom{N}{j} = \sqrt{\frac{N}{2\pi j\,(N-j)}}\left(\frac{N}{j}\right)^{j}\left(\frac{N}{N-j}\right)^{N-j}\exp\left[\varepsilon_N - \varepsilon_j - \varepsilon_{N-j}\right],$$

что равносильно

$$\binom{N}{j} e^{-N\mathscr{H}(j/N)} = \sqrt{\frac{N}{2\pi j\,(N-j)}}\exp\left[\varepsilon_N - \varepsilon_j - \varepsilon_{N-j}\right].$$

Так как $\varepsilon_N < \varepsilon_j$ и $\varepsilon_{N-j} > 0$, то, замечая, что $\exp[\varepsilon_N - \varepsilon_j - \varepsilon_{N-j}] < 1$, получаем искомую верхнюю границу. Для нижней границы $\varepsilon_N > 0$ и

$$\exp\left[\varepsilon_N - \varepsilon_j - \varepsilon_{N-j}\right] > \exp\left[-\frac{1}{12j} - \frac{1}{12\,(N-j)}\right].$$

Заметим теперь, что $-\frac{1}{12j}-\frac{1}{12(N-j)}\geqslant-\frac{1}{9}$ с исключением случаев $j=1$, $N-j=1$; $j=1$, $N-j=2$ и $j=2$, $N-j=1$. Таким образом, за этими исключениями имеем

$$\exp\left[\varepsilon_N-\varepsilon_j-\varepsilon_{N-j}\right]>\exp\left(-\frac{1}{9}\right)>\sqrt{\frac{2\pi}{8}} \quad \text{и}$$

$$\binom{N}{j}e^{-N\mathcal{H}(j/N)}>\sqrt{\frac{N}{2\pi j(N-j)}}\sqrt{\frac{2\pi}{8}}=\sqrt{\frac{N}{8j(N-j)}}.$$

Нижняя граница должна быть оценена численно для указанных выше исключительных случаев и можно увидеть, что неравенство имеет место при $j=1$ и $N-j=1$.

(б) Минимизируя $\sqrt{\frac{N}{8j(N-j)}}$ по j, $0<j<N$, находим, что минимум достигается при $j=N/2$, что дает $\sqrt{\frac{N}{8j(N-j)}}\geqslant\sqrt{\frac{1}{2N}}$. Аналогично $\sqrt{\frac{N}{2\pi j(N-j)}}$ максимизируется по целым j, $1\leqslant j\leqslant N-1$, и достигает максимального значения при $j=1$, и $\sqrt{\frac{N}{2\pi j(N-j)}}<\sqrt{\frac{1}{2\pi}}<1$.

Заметим, что $\sqrt{\frac{1}{2\pi}}$ является границей, в которую не входят ни N, ни j. Далее отметим, что

$$\binom{2N-1}{N}=\frac{1}{2}\binom{2N}{N}\geqslant\frac{1}{2}e^{2N\mathcal{H}(1/2)}\sqrt{\frac{1}{4N}}=\sqrt{\frac{1}{4N}}2^{2N-1}.$$

(в). Нижняя граница является очевидной, а верхняя граница следует непосредственно из указания; получаем

$$\binom{N}{n+1}=\binom{N}{n}\left(\frac{N-n}{n+1}\right)<\binom{N}{n}\left(\frac{N-n}{n}\right),$$

$$\binom{N}{j+m}<\binom{N}{j+m-1}\left(\frac{N-(j+m-1)}{j+m-1}\right)<\binom{N}{j+m-1}\left(\frac{N-j}{j}\right).$$

По индукции будем иметь

$$\binom{N}{j+m}\leqslant\binom{N}{j}\left(\frac{N-j}{j}\right)^m.$$

Из пунктов (а) и (в) следует, что

$$\sqrt{\frac{N}{8j(N-j)}}\exp\left\{N\mathcal{H}(j/N)+j\ln\varepsilon+(N-j)\ln(1-\varepsilon)\right\}\leqslant$$

$$\leqslant\sum_{n=j}^{N}\binom{N}{n}\varepsilon^n(1-\varepsilon)^{N-n}\leqslant\sqrt{\frac{N}{2\pi j(N-j)}}\times$$

$$\times\frac{j(1-\varepsilon)}{j(1-\varepsilon)-(N-j)\varepsilon}\exp\left\{N\mathcal{H}\left(\frac{j}{N}\right)+j\ln\varepsilon+(N-j)\ln(1-\varepsilon)\right\}.$$

(г) Первый результат получается легко. Для границы Чернова имеем

$$\Pr[w \geqslant j] \leqslant e^{-sj}[\varepsilon e^{s} + (1-\varepsilon)]^{N} \text{ при всех } s \geqslant 0.$$

Правая часть достигает минимума по $s \geqslant 0$ при

$$s = \ln\left[\frac{(1-\varepsilon)\, j/N}{\varepsilon\,(1-j/N)}\right].$$

После преобразования получаем

$$\Pr[w \geqslant j] \leqslant \exp\left\{N\left[\mathscr{H}\left(\frac{j}{N}\right) + \frac{j}{N}\ln\varepsilon + \frac{N-j}{N}\ln(1-\varepsilon)\right]\right\}.$$

Сравнивая с (в), получаем, что граница Чернова дает правильную экспоненциальную зависимость от N.

5.9. (а) $\Pr(\mathbf{y} \mid \mathbf{x}_m) = \varepsilon^{d(\mathbf{x}_m, \mathbf{y})}(1-\varepsilon)^{N-d(\mathbf{x}_m, \mathbf{y})}$.

Это выражение убывает с ростом $d(\mathbf{x}_m, \mathbf{y})$ так, что выбор m для минимизации $d(\mathbf{x}_m, \mathbf{y})$ равносилен выбору m для максимизации $\Pr(\mathbf{y} \mid \mathbf{x}_m)$.

(б) Когда сообщение m поступает на кодер и какое-нибудь частное значение $\mathbf{x}_m$ передается по каналу, то $d(\mathbf{x}_m, \mathbf{y})$ является числом ошибок, которые происходят в канале. Следовательно, $\Pr[d(\mathbf{x}_m, \mathbf{y}) = i]$ просто равна вероятности i ошибок для N символов. Поэтому

$$\Pr[d(\mathbf{x}_m, \mathbf{y}) = i] = \binom{N}{i}\varepsilon^{i}(1-\varepsilon)^{N-i}.$$

Точно так же при каждом $m' \neq m$ кодовое слово $\mathbf{x}_{m'}$ статистически не зависит как от $\mathbf{x}_m$, так и от переходов в канале и поэтому $\mathbf{x}_{m'}$ не зависит от $\mathbf{y}$ при условии, что m поступает на кодер. Так как каждая компонента $\mathbf{x}_{m'}$ является 0 или 1 с равными вероятностями, независимо от $\mathbf{y}$, то отсюда следует, что каждая компонента $\mathbf{x}_{m'}$ отличается от соответствующей компоненты $\mathbf{y}$ с вероятностью $1/2$, так что $\Pr[d(\mathbf{x}_{m'}, \mathbf{y}) = i]$ —вероятность i отличий в N символах—равна $\binom{N}{i}2^{-N}$.

(в) Ошибка происходит только тогда, когда $d(\mathbf{x}_{m'}, \mathbf{y}) \leqslant d(\mathbf{x}_m, \mathbf{y})$ при $m' \neq m$. Ограничивая сверху вероятность объединения этих событий суммой вероятностей, получим

$$\Pr[\text{ошибка} \mid d(\mathbf{x}_m, \mathbf{y}) = i] \leqslant \sum_{m' \neq m} \Pr[d(\mathbf{x}_{m'}, \mathbf{y}) \leqslant i \mid d(\mathbf{x}_m, \mathbf{y}) = i].$$

Однако, как уже было показано, на ансамбле кодов $d(\mathbf{x}_{m'}, \mathbf{y})$ не зависит от $d(\mathbf{x}_m, \mathbf{y})$ и поэтому

$$\Pr[d(\mathbf{x}_{m'}, \mathbf{y}) \leqslant i \mid d(\mathbf{x}_m, \mathbf{y}) = i] = \sum_{n=0}^{i}\binom{N}{n}2^{-N}.$$

Так как имеются $M-1$ возможных значений для $m' \neq m$ и так как вероятность не превосходит 1, то

$$\Pr[\text{ошибка} \mid d(\mathbf{x}_{m'}, \mathbf{y}) = i] \leqslant \begin{cases} (M-1)\displaystyle\sum_{n=0}^{i}\binom{N}{n}2^{-N}, \\ 1. \end{cases}$$

Заметим, что при малых i первая из указанных выше границ точнее, чем вторая. Наконец, применяя результат задачи 5.8 (в) с $\varepsilon = 1/2$, $i < N/2$, получим

$$\sum_{n=0}^{i}\binom{N}{n}2^{-N} = \sum_{n=N-i}^{N}\binom{N}{n}2^{-N} \leqslant \frac{N-i}{N-2i}\binom{N}{i}2^{-N}.$$

Ограничивая $(M-1)$ сверху с помощью $M = e^{NR}$, получаем искомый результат.

(г) $$\overline{P}_{e,m} = \sum_i \Pr[d(\mathbf{x}_m, \mathbf{y}) = i] \Pr[\text{ошибка} \mid d(\mathbf{x}_m, \mathbf{y}) = i]. \quad (1)$$

Возьмем $\delta < 1/2$ так, чтобы удовлетворялось равенство $\mathscr{H}(\delta) = \ln 2 - R$. Предположим, что $R < C$, что эквивалентно $\varepsilon < \delta$. Пусть j является целым числом, удовлетворяющим неравенствам

$$\frac{j-1}{N} < \delta \leqslant \frac{j}{N}. \quad (2)$$

Используя первую границу из (в) при $i \leqslant j-1$ и вторую при $i \geqslant j$, получаем

$$\overline{P}_{e,m} \leqslant \sum_{i=0}^{j-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i} \left(\frac{N-i}{N-2i}\right) \binom{N}{i} \times$$
$$\times e^{-N[\ln 2 - R]} + \sum_{i=j}^{N} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}. \quad (3)$$

(д) Согласно (2) можно ограничить сверху выражение $\left(\frac{N-i}{N-2i}\right)$ в первой сумме с помощью $\frac{1-\delta}{1-2\delta}$. Так как $j/N \geqslant \delta > \varepsilon$, то можно оценить сверху вторую сумму с помощью результатов задачи 5.8 (в); получим

$$\overline{P}_{e,m} \leqslant \frac{1-\delta}{1-2\delta} \sum_{i=0}^{j-1} \binom{N}{i}^2 \varepsilon^i (1-\varepsilon)^{N-i} e^{-N[\ln 2 - R]} + \frac{j(1-\varepsilon)}{j(1-\varepsilon) - (N-j)\varepsilon} \times$$
$$\times \sqrt{\frac{N}{2\pi j (N-j)}} \exp\left\{N\left[\mathscr{H}\left(\frac{j}{N}\right) + \frac{j}{N}\ln\varepsilon + \frac{N-j}{N}\ln(1-\varepsilon)\right]\right\}. \quad (4)$$

Рассмотрим теперь случай, когда

$$\left(\frac{\delta}{1-\delta}\right)^2 < \frac{\varepsilon}{1-\varepsilon}. \quad (5)$$

С помощью тех же самых рассуждений, что и в задаче 5.8 (в), при $i \leqslant j-1$ получим

$$\binom{N}{i}^2 < \binom{N}{j-1}^2 \left(\frac{j-1}{N-j+1}\right)^{2(j-1-i)} < \binom{N}{j-1}^2 \left(\frac{\delta}{1-\delta}\right)^{2(j-1-i)}.$$

Применяя эту границу для $\binom{N}{i}^2$, устремим затем нижний предел суммирования i к $-\infty$ и, суммируя геометрическую прогрессию, получаем

$$\sum_{i=0}^{j-1} \binom{N}{i}^2 \varepsilon^i (1-\varepsilon)^{N-i} < \frac{\binom{N}{j-1}^2 \varepsilon^{j-1} (1-\varepsilon)^{N-j+1}}{1 - \left(\frac{\delta}{1-\delta}\right)^2 \left(\frac{\varepsilon}{1-\varepsilon}\right)} \leqslant$$

$$\leqslant \frac{N\exp\left\{N\left[2\mathscr{H}\left(\frac{j-1}{N}\right)+\frac{j-1}{N}\ln\varepsilon+\frac{N-j+1}{N}\ln(1-\varepsilon)\right]\right\}}{2\pi(j-1)(N-j+1)\left[1-\left(\frac{\delta}{1-\delta}\right)^2\left(\frac{\varepsilon}{1-\varepsilon}\right)\right]} \leqslant \tag{6}$$

$$\leqslant \frac{\exp\{N[2\mathscr{H}(\delta)+\delta\ln\varepsilon+(1-\delta)\ln(1-\varepsilon)]\}}{2\pi N\delta(1-\delta)\left[1-\left(\frac{\delta}{1-\delta}\right)^2\left(\frac{\varepsilon}{1-\varepsilon}\right)\right]}. \tag{7}$$

Справедливость последнего неравенства может быть показана с помощью дифференцирования экспоненциальной части (6) по j и последующего использования (5) для того, чтобы показать, что производная является положительной. Если продифференцировать экспоненциальную и неэкспоненциальную части второй суммы в (4) по j, то можно заметить, что обе производные будут отрицательными и j можно заменить на δN. Подставляя это и (7) в (4), используя $\mathscr{H}(\delta)$ вместо $[\ln 2 - R]$ и объединяя слагаемые, получаем

$$\overline{P}_{e,\,m} \leqslant A\exp\{-N[-\mathscr{H}(\delta)-\delta\ln\varepsilon-(1-\delta)\ln(1-\varepsilon)]\},$$

где

$$A=\frac{1}{\delta 2\pi N(1-2\delta)\left[1-\left(\frac{\delta}{1-\delta}\right)^2\left(\frac{\varepsilon}{1-\varepsilon}\right)\right]}+\frac{1-\varepsilon}{\delta-\varepsilon}\sqrt{\frac{\delta}{2\pi N(1-\delta)}}.$$

Заметим, что в случае, когда имеет место (5), показатель экспоненты, приведенный выше, согласуется с показателем в (5.6.41). Далее рассмотрим случай, когда

$$\left(\frac{\delta}{1-\delta}\right)^2 \geqslant \frac{\varepsilon}{1-\varepsilon}.$$

В этом случае наибольшее слагаемое первой суммы в (4) может быть в любой точке интервала $(0,\ j-1)$. Результат может быть выражен очень просто, если вновь возвратиться к равенству (1). Будем иметь

$$\overline{P}_{e,\,m} \leqslant \sum_{i=0}^{N}\binom{N}{i}\varepsilon^i(1-\varepsilon)^{N-i}(M-1)\sum_{n=0}^{i}\binom{N}{n}2^{-N} =$$

$$=(M-1)2^{-N}\sum_{i=0}^{N}\binom{N}{i}\left(\sqrt{\varepsilon}\right)^i\left(\sqrt{1-\varepsilon}\right)^{N-i}\sum_{n=0}^{i}\binom{N}{n}\left(\sqrt{\varepsilon}\right)^i\left(\sqrt{1-\varepsilon}\right)^{N-i} \leqslant$$

$$\leqslant (M-1)2^{-N}\sum_{i=0}^{N}\binom{N}{i}\left(\sqrt{\varepsilon}\right)^i\left(\sqrt{1-\varepsilon}\right)^{N-i}\sum_{n=0}^{i}\binom{N}{n}\left(\sqrt{\varepsilon}\right)^n\left(\sqrt{1-\varepsilon}\right)^{N-n} \leqslant$$

$$\leqslant (M-1)2^{-N}[\sqrt{\varepsilon}+\sqrt{1-\varepsilon}]^N[\sqrt{\varepsilon}+\sqrt{1-\varepsilon}]^N \leqslant$$

$$\leqslant \exp\{-N[-R+\ln 2-2\ln(\sqrt{\varepsilon}+\sqrt{1-\varepsilon})]\}. \tag{8}$$

Заметим, что это согласуется с (5.6.45).

Главное, что следует понять в этой задаче, состоит в том, что даже в очень простом случае ДСК вывод этих экспоненциальных границ для ошибки, использующий прямые комбинаторные методы, является довольно скучным и требует большого терпения. В то же время тщательное изучение этого примера может дать значительное понимание того, почему получаются результаты на этом пути.

5.10. В ДСтК с вероятностью стирания ε имеем

$$E_0(\rho,\ \mathbf{Q}) = -\ln\{\varepsilon+(1-\varepsilon)[Q(0)^{(1+\rho)}+Q(1)^{(1+\rho)}]\}.$$

Максимальное значение по $\mathbf{Q}$ достигается при $Q(0) = 1/2$, давая

$$\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) = -\ln[\varepsilon + 2^{-\rho}(1-\varepsilon)].$$

При $\dfrac{(1-\varepsilon)\ln 2}{1+\varepsilon} \leqslant R \leqslant (1-\varepsilon)\ln 2$ имеем параметрические уравнения (см. (5.6.31))

$$R = \frac{2^{-\rho}(1-\varepsilon)\ln 2}{\varepsilon + 2^{-\rho}(1-\varepsilon)},$$

$$E_r(R) = -\ln[\varepsilon + 2^{-\rho}(1-\varepsilon)] - \frac{\rho 2^{-\rho}(1-\varepsilon)\ln 2}{\varepsilon + 2^{-\rho}(1-\varepsilon)}.$$

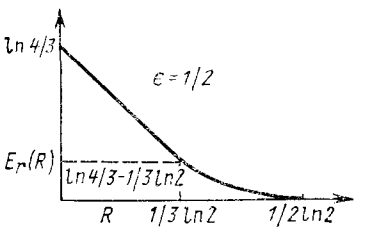
При $R < \dfrac{(1-\varepsilon)\ln 2}{1+\varepsilon}$ имеем $E_r(R) = \ln 2 - \ln(1+\varepsilon) - R$.

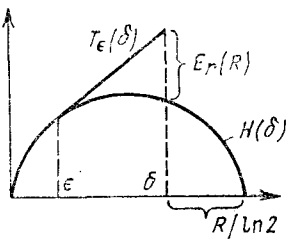
Пусть $\delta = \dfrac{\varepsilon}{\varepsilon + 2^{-\rho}(1-\varepsilon)}$; $\rho \ln 2 = \ln\left(\dfrac{1-\varepsilon}{\varepsilon}\dfrac{\delta}{1-\delta}\right)$.

В обозначениях δ параметрические уравнения примут вид

$$R = (1-\delta)\ln 2,$$

$$E_r(R) = T_\varepsilon(\delta) - H(\delta) \quad \text{[см. (5.6.44)]}.$$





5.11. (а) В силу симметрии канала ясно, что $E_0(\rho, \mathbf{Q})$ достигает максимума при $Q(k) = 1/5$, $0 \leqslant k \leqslant 4$. Поэтому

$$\max_{\mathbf{Q}} E_0(\rho, \mathbf{Q}) = \rho \ln(5/2),$$

$$E_r(R) = \ln(5/2) - R; \quad R \leqslant \ln(5/2).$$

(б) При $N = 1$ получаем нулевую вероятность ошибки, когда $\mathbf{x}_1 = 0$, $\mathbf{x}_2 = 2$. При $N = 2$ нулевую вероятность ошибки дают следующие пять кодовых слов:

$$(0, 0), (1,2), (2, 4), (3, 1), (4, 3).$$

5.12. Пусть $E_0(\rho)$ равно значению $E_0(\rho, \mathbf{Q})$ для распределения $\mathbf{Q}$, на котором достигается пропускная способность. Следует рассмотреть два случая. В вырожденном случае, в котором $E_0''(0) = 0$, имеем $E_r(R, \mathbf{Q}) \approx C - R$. В общем случае, когда $E_0''(0) < 0$, можно использовать два первых члена ряда Тейлора для $E_0(\rho)$ в окрестности $\rho = 0$. Получим

$$E_0(\rho) = \rho E_0'(0) + \frac{\rho^2}{2} E_0''(\rho_1) \text{ при некотором } \rho_1 \text{ из } 0 \leqslant \rho_1 \leqslant \rho.$$

Аналогично $E_0'(\rho) = E_0'(0) + \rho E_0''(\rho_2)$ при некотором ρ_2, $0 \leqslant \rho_2 \leqslant \rho$. Параметрические уравнения для $E_r(R, \mathbf{Q})$ имеют вид

$$R = E_0'(\rho) = C + \rho E_0''(\rho_2),$$

$$E_r(R,\ \mathbf{Q}) = E_0(\rho) - \rho E_0'(\rho) = \rho^2\left[\frac{E_0''(\rho_1)}{2} - E_0''(\rho_2)\right], \tag{1}$$

Решая (1) относительно ρ, получаем

$$E_r(R,\ \mathbf{Q}) = (C-R)^2\left[\frac{E_0''(\rho_1)}{2E_0''(\rho_2)^2} - \frac{1}{E_0''(\rho_2)}\right].$$

При $R \to C$ имеем $\rho \to 0$ и, следовательно, $\rho_1 \to 0$, $\rho_2 \to 0$. Получим

$$\lim_{R\to C}\frac{E_r(R,\ \mathbf{Q})}{(C-R)^2} = -\frac{1}{2E_0''(0)}.$$

Таким образом, для R, близких к C, имеем $E_r(R,\ \mathbf{Q}) \approx \alpha\,(C-R)^2$, где

$$\alpha = -\frac{1}{2E_0''(0)}. \tag{2}$$

Для несимметричных каналов поведение $E_r(R)$ не обязательно описывается таким же образом. Для $E_r(R)$ следовало бы заменить $E_0''(0)$ в (2) на

$$\left.\frac{\partial^2 \max\limits_{\mathbf{Q}}[E_0(\rho,\ \mathbf{Q})]}{\partial\rho^2}\right|_{\rho=0}.$$

5.13. $E_0(1,\ \mathbf{Q}) = -\ln\sum_j [Q(0)\sqrt{P(j\,|\,0)} + Q(1)\sqrt{P(j\,|\,1)}]^2 =$

$$= -\ln\sum_j [Q(0)^2 P(j\,|\,0) + Q(1)^2 P(j\,|\,1) + 2Q(0)Q(1)\sqrt{P(j\,|\,0)P(j\,|\,1)}] =$$

$$= -\ln\left[Q(0)^2 + Q(1)^2 + 2Q(0)Q(1)\sum_j \sqrt{P(j\,|\,0)P(j\,|\,1)}\right]. \tag{1}$$

Подставляя $1 - Q(0)$ вместо $Q(1)$ в написанное выше выражение и выполняя дифференцирование по $Q(0)$, получаем, что точка, соответствующая $Q(0) = {}^1/_2$, стационарна. Беря вторую производную, находим, что $Q(0) = {}^1/_2$ максимизирует (1), если $\sum_j \sqrt{P(j\,|\,0)P(j\,|\,1)} \leqslant 1$. Однако, как показано в задаче 4.15 (а), $\lambda = {}^1/_2$, это последнее неравенство всегда справедливо.

5.14. (а) $\overline{P}_1 = \Pr[I_m < TN \mid \text{сообщение } m]$. При условии, что задано m, величины $\ln\dfrac{P(y_n\,|\,x_{m,n})}{\omega(y_n)}$ образуют последовательность независимых случайных величин с распределением вероятностей $Q(x_{m,n})P(y_n\,|\,x_{m,n})$. Заметим, что $\ln\dfrac{P(y_n\,|\,x_{m,n})}{\omega(y_n)}$ принимает только конечное число значений с ненулевыми вероятностями. Следовательно, из (5.4.16) при $r < 0$ получаем

$$\overline{P}_1 \leqslant e^{-rTN}\left\{\sum_{k,j} Q(k)P(j\,|\,k)\exp\left[r\ln\frac{P(j\,|\,k)}{\omega(j)}\right]\right\}^N,$$

где суммирование ведется по k, j, для которых $Q(k)P(j\,|\,k) > 0$. Полагая $s = -r$, при любом $s > 0$ получаем

$$\overline{P}_1 \leqslant \exp\left\{-N\left\{-sT - \ln\left[\sum_{k,j} Q(k)\omega(j)^s P(j\,|\,k)^{1-s}\right]\right\}\right\}.$$

Замечая, что граница, очевидно, выполняется для $s = 0$, получаем $\overline{P}_1 \leqslant e^{-N\alpha}$ для α, определенном в условии задачи.

Используя аддитивную границу, имеем

$$\overline{P}_2 \leqslant \sum_{m'\neq m} \Pr[I_{m'} \geqslant TN \mid \text{сообщение } m].$$

При заданном сообщении m в ансамбле кодов кодовые слова $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ и принятая последовательность $\mathbf{y}$ имеют распределение вероятностей $Q_N(\mathbf{x}_m)\,P_N(\mathbf{y}\mid\mathbf{x}_m)\times$ $\times Q_N(\mathbf{x}_{m'})$. Таким образом $\mathbf{y}$ и $\mathbf{x}_{m'}$ статистически независимы и имеют распределение вероятности $\omega_N(\mathbf{y})\,Q_N(\mathbf{x}_{m'})$. Отсюда и из (5.4.15) при любом $r>0$ получаем

$$\overline{P}_2 \leqslant (M-1)\,e^{-rTN}\left\{\sum_{k,j} Q(k)\,\omega(j)\exp\left[r\ln\frac{P(j\mid k)}{\omega(j)}\right]\right\}^N.$$

Полагая $s=1-r$ и ограничивая сверху $M-1$ с помощью e^{RN}, получаем, что при $s<1$,

$$\overline{P}_2 \leqslant \exp\left\{-N\left\{-R+(1-s)\,T-\ln\left[\sum_{k,j} Q(k)\,\omega(j)^s\,P(j\mid k)^{1-s}\right]\right\}\right\}.$$

При $s=1$ этот результат является тривиально справедливым. Таким образом,

$$\overline{P}_2 \leqslant \exp\left[-N(\alpha+T-R)\right].$$

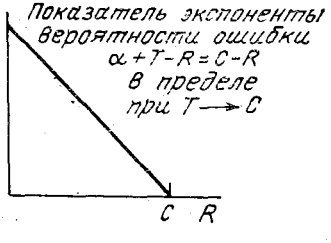
(б) Пусть $\mu(s)=\ln\left\{\sum_{k,j} Q(k)\,P(j\mid k)\exp\left[s\ln\frac{\omega(j)}{P(j\mid k)}\right]\right\}$.

Заметим, что $\alpha=\max\limits_{0\leqslant s\leqslant 1}[-sT-\mu(s)]$ и что $\mu(s)$ является семиинвариантной производящей функцией моментов взятой со знаком минус взаимной информации, поэтому $\mu'(0)=-C$ и отсюда следует, что при любом $T<C$ значение $-sT-\mu(s)$ является положительным для достаточно малого $s>0$. Следовательно, $\alpha>0$. Из выпуклости $\mu(s)$ следует также, что $-sT-\mu(s)$ достигает максимума по s при $T=-\mu'(s)$. При T, стремящемся к C, максимизирующее значение s убывает к 0 и $-sT-\mu(s)$ стремится к 0.

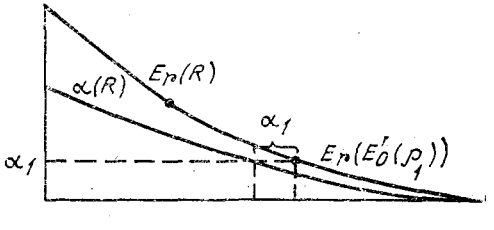
(в) Для того чтобы произошло стирание, когда передавалось сообщение m, должно быть либо $I_m<TM$, либо $I_{m'}\geqslant TN$ при некотором $m'\neq m$. Таким образом, Pr (стирание) $\leqslant \overline{P}_1+\overline{P}_2$.

Для того чтобы произошла ошибка, $I_{m'}$ должно быть большим или равным TN в точности для одного значения $m'\neq m$ и I_m должно быть меньше чем TN. Вероятность этого ограничена вероятностью того, что $I_{m'}\geqslant TN$ для одного или большего числа $m'\neq m$, т. е. Pr (ошибка) $\leqslant \overline{P}_2$.

Отметим, что если декодер модифицируется так, чтобы производить стирание, когда $I_m<TN$ при всех m, и декодировать такое m, для которого I_m является наибольшим среди остальных, то можно было бы показать, что Pr (стирание) $\leqslant \overline{P}_1$, но отличие является несущественным в том интересном случае, когда $\overline{P}_2 \ll \overline{P}_1$.



К задаче 5.14 (в).



К задаче 5.15 (а).

Следствием этого является то, что можно достичь много меньшей вероятности ошибки при заданной скорости и длине блока, когда стертые символы могут быть переданы вновь. Истинная скорость, с которой передаются данные, понижается до $R/(1-e^{-N\alpha})$ из-за повторений передачи, но эта потеря в скорости

пренебрежимо мала при больших N, даже когда α довольно мало. В качестве еще одного практического следствия укажем на то, что все легко реализуемые методы декодирования при больших N имеют механизм обнаружения ошибок, т. е. они не могут найти никакого наиболее вероятного кодового слова, когда шум слишком велик.

5.15. (а) В ДСК $\omega(j)=1/2$, $Q(0)=Q(1)=1/2$ и

$$\alpha=\max_{0\leqslant s\leqslant 1}\{-sR+s\ln 2-\ln[(1-\varepsilon)^{1-s}+\varepsilon^{1-s}]\}=$$

$$=\max_{\rho\geqslant 0}\left\{-\frac{\rho R}{1+\rho}+\frac{\rho}{1+\rho}\ln 2-\frac{1}{1+\rho}\ln[(1-\varepsilon)^{1/(1+\rho)}+\varepsilon^{1/(1+\rho)}]^{1+\rho}\right\}=$$

$$=\max_{\rho\geqslant 0}\frac{1}{1+\rho}[-\rho R+E_0(\rho)].$$

Производя максимизацию по ρ, получаем параметрические уравнения

$$\alpha(R)=E_0(\rho)-\rho E_0'(\rho),$$
$$R=E_0'(\rho)-[E_0(\rho)-\rho E_0'(\rho)]. \qquad (1)$$

Используя графический метод, проиллюстрированный на рис. 5.6.3, можно увидеть, что ординаты точек всех прямых линий, огибающая которых изображает $E_r(R)$, умножаются на $1/(1+\rho)$ и огибающая полученных прямых линий изображает $\alpha(R)$. Более изящная интерпретация, основанная на (1), показывает, что $\alpha(R)=E_r[E_0'(\rho)]$.

(б) $$\alpha=\max_{\rho\geqslant 0}\left\{-\frac{\rho}{1+\rho}R-\ln\left[\sum_{j,k}Q(k)\,\omega(j)^{\rho/(1+\rho)}\,P(j\mid k)^{1/(1+\rho)}\right]\right\}. \qquad (1)$$

Применим неравенство Гёльдера [см. задачу 4.15 (б)] к выражению в квадратных скобках. Получим

$$\sum_j\omega(j)^{\rho/(1+\rho)}\left(\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho)}\right)\leqslant\left[\sum_j\omega(j)^{\rho/[(1+\rho)\lambda]}\right]^{\lambda}\times$$
$$\times\left[\sum_j\left(\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho)}\right)^{1/(1-\lambda)}\right]^{1-\lambda}=$$
$$=\left\{\sum_j\left(\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho)}\right)^{1+\rho}\right\}^{1/(1+\rho)}, \qquad (2)$$

где $\lambda=\rho/(1+\rho)$, и затем замечено, что сумма в первом сомножителе равна 1. Взяв логарифм от каждой части (2) и изменив знак, получаем

$$-\ln\sum_{j,k}\omega(j)^{\rho/(1+\rho)}\,Q(k)\,P(j\mid k)^{1/(1+\rho)}\geqslant$$
$$\geqslant\frac{1}{1+\rho}E_0(\rho).$$

Подставляя это выражение в (1), получаем $\alpha\geqslant\dfrac{1}{1+\rho}[-\rho R+E_0(\rho)]$.

(в) $$\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho)}=\sum_k Q(k)\,P(j\mid k)^{\rho/(1+\rho)}\times$$
$$\times P(j\mid k)^{(1-\rho)/(1+\rho)}\leqslant\left[\sum_k Q(k)\,P(j\mid k)\right]^{\rho/(1+\rho)}\times$$
$$\times\left[\sum_k Q(k)\,P(j\mid k)^{1-\rho}\right]^{1/(1+\rho)},$$

где было использовано неравенство Гёльдера с $\lambda=\rho/(1+\rho)$. Используя полученное неравенство, находим

$$E_0(\rho)=-\ln\sum_j\left(\sum_k Q(k)\,P(j\mid k)^{1/(1+\rho)}\right)^{1+\rho}\geqslant$$

$$\geqslant -\ln \sum_j \omega(j)^{\rho} \sum_k Q(k) P(j \mid k)^{1-\rho}.$$

Поэтому

$$E_r(R) = \max_{0 \leqslant \rho \leqslant 1} [-\rho R + E_0(\rho)] \geqslant \max_{0 \leqslant \rho \leqslant 1} \left\{ -\rho R - \ln \left[\sum_{j,k} \omega(j)^{\rho} Q(k) P(j \mid k)^{1-\rho} \right] \right\} = \alpha$$

5.16. **(а)** Как и в (5.6.5) имеем

$$\overline{P}_{e,\,m} = \sum_{\mathbf{x}_m} \sum_{\mathbf{y}} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m) \Pr[\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}]. \tag{1}$$

Для заданных m, $\mathbf{x}_{m'}$, $\mathbf{y}$ пусть $A_{m'}$ обозначает событие, состоящее в том, что $\mathbf{x}_{m'}$ выбрано таким образом, что

$$q_{m'} P_N(\mathbf{y} \mid \mathbf{x}_{m'}) \Pr \geqslant q_m P_N(\mathbf{y} \mid \mathbf{x}_m).$$

Имеем

$$P(A_{m'}) \leqslant \sum_{\mathbf{x}_{m'}} Q_N(\mathbf{x}_{m'}) \left[\frac{q_{m'} P_N(\mathbf{y} \mid \mathbf{x}_{m'})}{q_m P_N(\mathbf{y} \mid \mathbf{x}_m)} \right]^s. \tag{2}$$

Как и в (5.6.7), имеем

$$\Pr[\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}] \leqslant \left[\sum_{m' \neq m} P(A_{m'}) \right]^{\rho} \leqslant$$

$$\leqslant \left\{ \frac{\left(\sum_{m' \neq m} q_{m'}^s \right) \sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^s}{q_m^s P_N(\mathbf{y} \mid \mathbf{x}_m)^s} \right\}^{\rho}.$$

Подставляя это в (1) и производя преобразования, получаем

$$\overline{P}_{e,\,m} \leqslant q_m^{-s\rho} \left[\sum_{m' \neq m} q_{m'}^s \right]^{\rho} \times$$

$$\times \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m)^{1-s\rho} \right) \left(\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^s \right)^{\rho}.$$

Положив $s = 1/(1+\rho)$ и распространяя суммирование по m' на значение m, получим

$$\overline{P}_e = \sum_m q_m \overline{P}_{e,\,m} \leqslant \left(\sum_m q_m^{1/(1+\rho)} \right)^{1+\rho} \times$$

$$\times \sum_{\mathbf{y}} \left(\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^{1/(1+\rho)} \right)^{1+\rho}. \tag{3}$$

(б) Если сообщение m соответствует последовательности источника $\mathbf{u} = (u_1, \ldots, u_L)$, то $q_m = \prod_{l=1}^{L} \pi(u_l)$ и

$$\left(\sum_m q_m^{1/(1+\rho)} \right)^{1+\rho} = \left\{ \sum_{u_1, \ldots u_L} \prod_{l=1}^{L} \pi(u_l)^{1/(1+\rho)} \right\}^{1+\rho} =$$

$$= \left[\sum_i \pi(i)^{1/(1+\rho)} \right]^{(1+\rho)L}.$$

Преобразуя последнее выражение в (3), так же как в (5.6.11), будем иметь

$$\overline{P}_e \leqslant \exp[-NE_0(\rho, \mathbf{Q}) + LE_s(\rho)],$$

$$E_s(\rho) = (1+\rho) \ln \sum_i \pi(i)^{1/(1+\rho)}.$$

(в) Так как $\sum_i \pi(i) = 1$, то $E_s(0) = 0$. После некоторых преобразований получим

$$\frac{\partial E_s(\rho)}{\partial \rho} = -\frac{\sum_i \pi(i)^{1/(1+\rho)} \ln \dfrac{\pi(i)^{1/(1+\rho)}}{\sum_l \pi(l)^{1/(1+\rho)}}}{\sum_l \pi(l)^{1/(1+\rho)}}.$$

При $\rho = 0$ это выражение равно $H(U)$, а при $\rho > 0$ является энтропией множества с вероятностями $\pi(i)^{1/(1+\rho)} / \sum_l \pi(l)^{1/(1+\rho)}$. Если эта энтропия не равна нулю (т. е. если равенство $\pi(i) = 1$ не выполняется), то $E_s(\rho)$ возрастает вместе с ρ. Тривиальное видоизменение леммы 5Б.1 также показывает, что $E_s(\rho)$ является выпуклой $\cup$.

(г) $$\overline{P}_e \leqslant \exp\{-N[E_0(\rho, \mathbf{Q}) - \lambda E_s(\rho)]\}. \tag{4}$$

При $\mathbf{Q}$, на котором достигается пропускная способность, имеем

$$\left.\frac{\partial [E_0(\rho, \mathbf{Q}) - \lambda E_0(\rho)]}{\partial \rho}\right|_{\rho=0} = C - \lambda H(U).$$

Если это выражение положительно, то выражение, стоящее в квадратных скобках (4), является положительным для достаточно малых $\rho > 0$ и поэтому $\overline{P}_e \to 0$ при $N \to \infty$. Отметим, что в наших предыдущих рассмотрениях кодирование для источника и кодирование для канала изучалось отдельно. Представленный здесь результат имеет некоторые методические преимущества при отыскании соотношения между вероятностью ошибки и кодовым ограничением, относящимся как к кодированию для источника, так и к кодированию для канала.

(д) При $\pi(i) = 1/A$, $0 \leqslant i \leqslant A-1$, имеем

$$E_s(\rho) = (1+\rho) \ln \left[A \left(\frac{1}{A} \right)^{1/(1+\rho)} \right] = \rho \ln A.$$

Поэтому $\lambda E_s(\rho) = \rho \dfrac{\ln A^L}{N} = \rho R.$

Для канала без шума (т. е. для канала, в котором для каждой выходной буквы j имеется только одна входная буква k, для которой $P(j|k) > 0$) имеем

$$E_0(\rho, \mathbf{Q}) = \rho \ln K,$$

где K является объемом алфавита на входе, а $\mathbf{Q} = \left(\frac{1}{K}, \ldots, \frac{1}{K} \right)$. Из результата пункта (б) следует, что P_e экспоненциально убывает по L при фиксированном L/N, если $H(U) < \dfrac{N \ln K}{L}$. Поэтому доказанное равносильно положительному утверждению теоремы 3.1.1 с дополнительным утверждением об экспоненциальной сходимости по L.

5.17. Пусть $P_i(j|k)$ будут переходными вероятностями в i-м канале и пусть $Q_i(k)$ является распределением, на котором достигается максимум $E_{0,i}(\rho, \mathbf{Q})$ при заданном ρ. Обозначим

$$\alpha_{i,j}(\mathbf{Q}_i) = \sum_k Q_i(k) P_i(j|k)^{1/(1+\rho)}. \tag{1}$$

Согласно (5.6.37) имеем

$$\sum_l P_l(j|k)^{1/(1+\rho)} \alpha_{l,j}(\mathbf{Q})^\rho > \sum_l \alpha_{l,j}(\mathbf{Q})^{1+\rho} = e^{-E_{0,l}(\rho)} \tag{2}$$

с равенством при i, k, для которых $Q_i(k) > 0$. Покажем, что на некотором множестве вероятностей $\{q_i\}$ вероятности q_iQ_i (k) максимизируют E_0 (ρ, **Q**). Используя (1), находим, что необходимым и достаточным условием максимума E_0 (ρ, **Q**) для суммы каналов будет

$$q_i^{\rho} \sum_j P_i(j \mid k)^{1/(1+\rho)} \alpha_{ij}(\mathbf{Q})^{\rho} \geqslant \sum_{i,j} q_i^{1+\rho} \alpha_{ij}(\mathbf{Q})^{1+\rho}. \tag{3}$$

Это показывает, что $q_i > 0$ при всех i. Подставляя (2) и (3) и используя k, для которого $Q_i(k) > 0$, получаем

$$q_i^{\rho} e^{-E_{0,i}(\rho)} = \sum_i q_i^{1+\rho} e^{-E_{0,i}(\rho, \mathbf{Q}_i)} = e^{-E_0(\rho)}. \tag{4}$$

Следовательно,

$$q_i = \frac{e^{E_{0,i}(\rho)/\rho}}{e^{E_0(\rho)/\rho}}. \tag{5}$$

Наконец, используя то, что $\Sigma q_i = 1$, получаем из (5), что

$$e^{E_0(\rho)/\rho} = \sum_i e^{E_{0,i}(\rho)/\rho}, \tag{6}$$

из (5) и (6), очевидно, вытекают соотношения, указанные в задаче.

Данный канал является суммой каналов с $E_{0,1}(\rho) = \rho \ln(3/2)$ и $E_{0,2}(\rho) = 0$. Таким образом, $E_0(\rho) = \rho \ln\left[\frac{3}{2} + 1\right] = \rho \ln(5/2)$.

Заметим, что этот канал является одним из довольно необычных каналов, которые рассматривались на стр. 159; в этом канале функция $E_r(R)$ является линейной. Фактически, как ясно из (6), для любой суммы каналов, в которой все составляющие $E_r(R)$ являются линейными функциями, функция $E_r(R)$ также является линейной.

5.18. (а) Для данного декодера правильное декодирование в случае передачи сообщения m происходит, если $(\mathbf{x}_m, \mathbf{y}) \in T_N$ и $(\mathbf{x}_{m'}, \mathbf{y}) \notin T_N$ при всех $m' \neq m$. Поэтому вероятность ошибки равна вероятности объединения следующих событий: $(\mathbf{x}_m, \mathbf{y}) \notin T_N$ или $(\mathbf{x}_{m'}, \mathbf{y}) \in T_N$, $m' = 1$, или $2, \ldots,$ или M $(m' \neq m)$. Таким образом,

$$\overline{P}_{e,m} \leqslant \Pr[(\mathbf{x}_m, \mathbf{y}) \notin T_N \mid m] + \sum_{m' \neq m} \Pr[(\mathbf{x}_{m'}, \mathbf{y}) \in T_N \mid m]. \tag{1}$$

(б) $$\Pr[(\mathbf{x}_m, \mathbf{y}) \notin T_N \mid m] = \Pr\left[\left|\frac{1}{N} \sum_{n=1}^{N} \ln \frac{P(y_n \mid x_{m,n})}{\omega(y_n)} - C\right| > \varepsilon \mid m\right].$$

При заданном m величины $\ln \dfrac{P(y_n \mid x_{m,n})}{\omega(y_n)}$ при $1 \leqslant n \leqslant N$ образуют последовательность независимых одинаково распределенных случайных величин с распределением вероятностей $Q(x_{m,n}) P(y_n \mid x_{m,n})$ и средним значением C. Теперь согласно закону больших чисел имеем

$$\lim_{N \to \infty} \Pr[(\mathbf{x}_m, \mathbf{y}) \notin T_N \mid m] = 0. \tag{2}$$

(в) $$\Pr[(\mathbf{x}_{m'}, \mathbf{y}) \in T_N \mid m] = \sum_{(\mathbf{x}_{m'}, \mathbf{y}) \in T_N} [Q_N(\mathbf{x}_{m'}) \omega_N(\mathbf{y}_N)], \tag{3}$$

так как при заданном m величина $\mathbf{y}_N$ выбирается независимо от $\mathbf{x}_m$ с распределением вероятности $\omega_N(\mathbf{y}) = \sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m)$. Согласно определению

типичного множества T_N имеем

$$\left|\frac{1}{N}\ln\frac{P_N(\mathbf{y}\mid\mathbf{x}_{m'})}{\omega_N(\mathbf{y}_N)}-C\right|\leqslant\varepsilon \text{ при } (\mathbf{x}_{m'},\ \mathbf{y})\in T_N,$$

$$\frac{1}{N}\ln\frac{P_N(\mathbf{y}\mid\mathbf{x}_{m'})}{\omega_N(\mathbf{y}_N)}\geqslant C-\varepsilon \text{ при } (\mathbf{x}_{m'},\ \mathbf{y})\in T_N,$$

$$\omega_N(\mathbf{y}_N)\leqslant P_N(\mathbf{y}\mid\mathbf{x}_m)\,e^{-N(C-\varepsilon)} \text{ при } (\mathbf{x}_{m'},\ \mathbf{y})\in T_N.$$

Подставляя эти выражения в (3) и строя границу сверху с помощью суммирования по всем $\mathbf{x}_{m'}$, $\mathbf{y}$, получаем

$$\Pr[(\mathbf{x}_{m'},\ \mathbf{y})\in T_N\mid m]\leqslant e^{-N(C-\varepsilon)}. \tag{4}$$

(г) $$\sum_{m'\neq m}\Pr[(\mathbf{x}_m,\ \mathbf{y})\in T_N\mid m]\leqslant$$

$$\leqslant(M-1)\,e^{-N(C-\varepsilon)}\leqslant e^{-N(C-R-\varepsilon)}=e^{-N\varepsilon}. \tag{5}$$

Используя (5) и (2) совместно с (1), будем иметь

$$\lim_{N\to\infty}\overline{P}_{e,m}=0.$$

Заметим, что сходимость не зависит от m и поэтому

$$\lim_{N\to\infty}\overline{P}_e=\lim_{N\to\infty}\frac{1}{N}\sum_m\overline{P}_{e,m}=0.$$

С помощью того же самого доказательства, что и в следствии 2 на стр. 157, можно показать, что для любой $R<C$, любом $\delta>0$ и достаточно большом N существуют коды, для которых

$$P_{e,m}\leqslant\delta \text{ при всех } m,\ 1\leqslant m\leqslant\lfloor\exp NR\rfloor.$$

5.19. (а) Заметим, что в ДСК $P_N(\mathbf{y}|\mathbf{x})=\varepsilon^{d(\mathbf{x},\mathbf{y})}(1-\varepsilon)^{N-d(\mathbf{x},\mathbf{y})}$, где $d(\mathbf{x},\mathbf{y})$ равно числу позиций, в которых $\mathbf{x}$ и $\mathbf{y}$ отличаются друг от друга. Отсюда можно заметить, что при $\varepsilon<1/2$ декодирование по максимуму правдоподобия равносильно выбору такого m, которое минимизирует $d(\mathbf{x}_m,\mathbf{y})$ (т. е. выбору ближайшего кодового слова). Вероятность того, что $P_N(\mathbf{y}|\mathbf{x}_{m'})\geqslant P_N(\mathbf{y}|\mathbf{x}_m)$, равна вероятности того, что $\mathbf{y}$ совпадает с $\mathbf{x}_{m'}$, по крайней мере, во стольких же позициях, как с и $\mathbf{x}_m$, или, другими словами, вероятности того, что $\mathbf{y}$ отличается от $\mathbf{x}_m$, по крайней мере, в половине из $d(\mathbf{x}_m,\mathbf{x}_{m'})$ позиций, в которых $\mathbf{x}_m$ и $\mathbf{x}_{m'}$ отличаются друг от друга. Следовательно, из (5.3.13) будем иметь

$$\Pr[P_N(\mathbf{y}\mid\mathbf{x}_{m'})\geqslant P_N[(\mathbf{y}\mid\mathbf{x}_m)\mid m]\leqslant[2\sqrt{\varepsilon(1-\varepsilon)}]^{d(\mathbf{x}_m,\,\mathbf{x}_{m'})}.$$

Для того чтобы произошла ошибка при передаче $\mathbf{x}_m$, значение $P_N(\mathbf{y}\mid\mathbf{x}_{m'})$ должно быть больше или равно значению $P_N(\mathbf{y}\mid\mathbf{x}_m)$, по крайней мере, для одного $m'\neq m$. Имеем

$$P_{e,m}\leqslant\sum_{m'\neq m}\exp\{d(\mathbf{x}_m,\mathbf{x}_{m'})\ln[2\sqrt{\varepsilon(1-\varepsilon)}]\}.$$

(б) Так как $\ln[2\sqrt{\varepsilon(1-\varepsilon)}]<0$, то $P_{e,m}$ можно оценить сверху с помощью оценки снизу $d(\mathbf{x}_m,\mathbf{x}_{m'})$. Поэтому

$$P_{e,m}\leqslant(M-1)\exp\{d_{min}\ln[2\sqrt{\varepsilon(1-\varepsilon)}]\},\quad 1\leqslant m\leqslant M. \tag{1}$$

(в) Каждое новое кодовое слово вызывает устранение из списка не более $\sum_{i=0}^{d-1}\binom{N}{i}$ слов. Поэтому, если после выбора M кодовых слов

$$M\sum_{i=0}^{d-1}\binom{N}{i}<2^N,$$

то можно выбрать следующее кодовое слово. Таким образом, после окончания процедуры $M\sum_{i=0}^{d-1}\binom{N}{i}\geqslant 2^N$.

(г) Из задачи 5.8 (г) при $\varepsilon=1/2$ имеем

$$\sum_{i=0}^{d-1}\binom{N}{i}2^{-N}=\sum_{i=N-d+1}^{N}\binom{N}{i}2^{-N}\leqslant\exp\left\{N\left[\mathscr{H}\left(\frac{d-1}{N}\right)-\ln 2\right]\right\}.$$

Следовательно, число слов, которое выбирается с минимальным расстоянием d_{min}, удовлетворяет границе

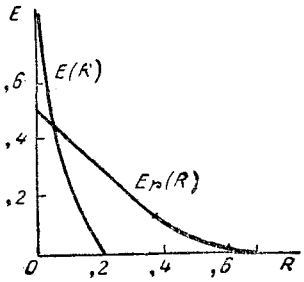
$$M\geqslant\left[2^{-N}\sum_{i=0}^{d_{min}-1}\binom{N}{i}\right]^{-1}\geqslant\exp\left\{N\left[\ln 2-\mathscr{H}\left(\frac{d_{min}-1}{N}\right)\right]\right\}. \qquad (2)$$

Пусть $R<\ln 2$ и пусть $\delta<1/2$ удовлетворяет равенству $\mathscr{H}(\delta)=\ln 2-R$. Выберем нужное значение d_{min} так, чтобы $\frac{d_{min}-1}{N}<\delta\leqslant\frac{d_{min}}{N}$.

(д) Из (2) следует, что можно выбрать код с минимальным расстоянием d_{min} и с $M\geqslant\exp\{N[\ln 2-\mathscr{H}(\delta)]\}=e^{NR}\geqslant M-1$ кодовыми словами. Для такого кода неравенство (1) приводится к виду

$$P_{e,m}\leqslant\exp\{-NE(R)\},\quad R=\ln 2-\mathscr{H}(\delta),$$
$$E(R)=-R-\delta\ln[2\sqrt{\varepsilon(1-\varepsilon)}]=-\ln 2+\mathscr{H}(\delta)-\delta\ln[2\sqrt{\varepsilon(1-\varepsilon)}].$$

При $\varepsilon=0{,}01$ получим следующее графическое изображение.



5.20. Пусть $A(m_1, m_2, \ldots, m_L)$ будет событием, состоящим в том, что $P_N(\mathbf{y}\mid\mathbf{x}_{m_l})\geqslant P_N(\mathbf{y}\mid\mathbf{x}_m)$ при всех l, $1\leqslant l\leqslant L$. Для возникновения ошибки при декодировании списком должно произойти событие $A(m_1, \ldots, m_L)$ при некоторых значениях $m_1, \ldots, m_L$. Следовательно,

$$\Pr(\text{ошибка при декодировании списком}\mid m, \mathbf{x}_m, \mathbf{y})\leqslant\Pr[\bigcup A(m_1, \ldots, m_L)]\leqslant[\sum\Pr(A(m_1,\ldots,m_L))]^{\rho_0},\quad 0<\rho_0\leqslant 1, \qquad (1)$$

где объединение и сумма берутся по всем наборам L различных целых чисел,

не равных m. Так как $\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_L}$ являются независимыми, то

$$\Pr[A(m_1, \ldots, m_L)] = \prod_{l=1}^{L} \Pr[P_N(\mathbf{y} \mid \mathbf{x}_{m_l}) \geqslant P_N(\mathbf{y} \mid \mathbf{x}_m) \mid m, \mathbf{x}_m, \mathbf{y}],$$

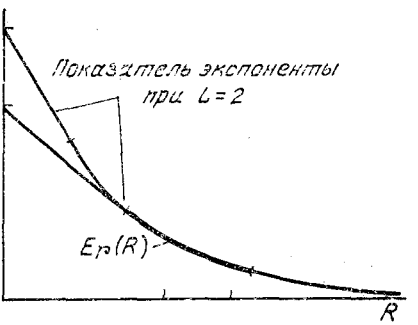
$$\Pr[A(m_1, \ldots, m_L)] \leqslant \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) \frac{P_N(\mathbf{y} \mid \mathbf{x})^s}{P_N(\mathbf{y} \mid \mathbf{x}_m)^s} \right]^L. \qquad (2)$$

В силу того, что имеется $\binom{M-1}{L}$ различных наборов по L целых чисел, можно подставить (2) в (1) и получить (при $0 < \rho \leqslant 1$), $\Pr[\text{ошибка при декодировании списком} \mid m, \mathbf{x}_m, \mathbf{y}] \leqslant \binom{M-1}{L}^{\rho_0} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) \frac{P_N(\mathbf{y} \mid \mathbf{x})^s}{P_N(\mathbf{y} \mid \mathbf{x}_m)^s} \right]^{\rho_0 L}$.

Положив $\rho = \rho_0 L$, $s = 1/(1+\rho)$, и используя неравенство $\binom{M-1}{L}^{\rho_0} \leqslant (M-1)^{\rho}$, получим $\overline{P}_{L,e,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x})^{1/(1+\rho)} \right]^{1+\rho}$ так же, как это было сделано в теореме 5.6.1. Заметим, однако, что это справедливо для всех ρ из интервала $0 \leqslant \rho \leqslant L$, так как $\rho = \rho_0 L$. В ДКБП это сводится к

$$\overline{P}_{L,e,m} \leqslant \exp\left[-N \left[\max_{0 \leqslant \rho \leqslant L} [-\rho R + E_0(\rho, \mathbf{Q})]\right]\right]$$

точно так же, как это было сделано в теореме 5.6.2.



5.21. (а) Для случая I весь канал экивалентен ДСтК с вероятностью стирания $\varepsilon_i = 2\varepsilon - \varepsilon^2$. Показатель экспоненты случайного кодирования был найден в задаче 5.10. Для случая II имеем

$$\overline{P}_e \leqslant 2 \exp\left[-N \frac{E_r(R)}{2}\right],$$

где $E_r(R)$ является показателем экспоненты случайного кодирования для ДСтК с вероятностью стирания ε. Заметим, что пропускная способность больше для случая II, и поэтому он предпочтительнее при больших скоростях. В пределе при нулевой скорости с помощью небольших вычислений можно найти, что случай I предпочтительнее при малых ε, а случай II при больших ε.

(б) В случае III символ стирается с вероятностью $\varepsilon_i = 2\varepsilon - \varepsilon^2$. Выражая скорость в натуральных единицах [в соответствии с (а)], находим вероятность того, что на выходе появятся менее чем $\frac{R}{\ln 2} N$ правильных символов при

N-кратном использовании канала или, другими словами, вероятность более чем δN стираний при N-кратном использовании канала, где

$$\delta = 1 - \frac{R}{\ln 2}.$$

Используя задачу 5.8 (г), находим, что вероятность того, что не будут получены $RN/\ln 2$ правильных символов, равна

$$P_e \leqslant \exp\{N[\mathscr{H}(\delta) + \delta \ln \varepsilon_i + (1-\delta)\ln(1-\varepsilon_i)]\}.$$

Это та же самая граница, как и в случае I при $R_{cr} \leqslant R < C$ и эта экспонента больше при $R < R_{cr}$.

5.22. Так же как и в доказательстве теоремы 5.6.1, при заданных m, $\mathbf{x}_m$, $\mathbf{y}$ обозначим через $A_{m'}$ событие, состоящее в том, что $P'_N(\mathbf{y} \mid \mathbf{x}_{m'}) \geqslant P'_N(\mathbf{y} \mid \mathbf{x}_m)$. Тогда

$$\Pr(\text{ошибка} \mid m, \mathbf{x}_m, \mathbf{y}) \leqslant \Pr\Big(\bigcup_{m' \neq m} A_{m'}\Big) \leqslant \left[(M-1)\sum_{\mathbf{x}} Q_N(\mathbf{x}) \frac{P'_N(\mathbf{y} \mid \mathbf{x})^s}{P'_N(\mathbf{y} \mid \mathbf{x}_m)^s}\right]^\rho.$$

В силу того, что $P_N(\mathbf{y} \mid \mathbf{x})$ задает переходные вероятности в канале,

$$\overline{P}_{e,m} \leqslant (M-1)^\rho \sum_{\mathbf{x}_m, \mathbf{y}} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m) P'_N(\mathbf{y} \mid \mathbf{x}_m)^{-s\rho} \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P'_N(\mathbf{y} \mid \mathbf{x})^s\right]^\rho.$$

Положив $s = 1/(1+\rho)$ (что не является наилучшим выбором при $P'_N \neq P_N$), получим

$$\overline{P}_{e,m} \leqslant \exp\{-N[-\rho R + f[\rho, Q_1, \mathbf{P}, \mathbf{P}']]\},$$

$$f = -\ln \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}_m} Q_N(\mathbf{x}_m) P_N(\mathbf{y} \mid \mathbf{x}_m) P'_N(\mathbf{y} \mid \mathbf{x}_m)^{-\rho/(1+\rho)}\right] \times$$
$$\times \left[\sum_{\mathbf{x}} Q_N(\mathbf{x}) P'_N(\mathbf{y} \mid \mathbf{x})^{1/(1+\rho)}\right]^\rho.$$

В ДСК, полагая $P(1 \mid 0) = P(0 \mid 1) = 1 - \varepsilon$ и $P'(1 \mid 0) = P'(0 \mid 1) = \varepsilon$, $\varepsilon < 1/2$, получаем, что значение f будет отрицательным и, в действительности, алгоритм декодирования строится так, что всегда выбирается наиболее вероятное кодовое слово.

5.23. Эта задача довольно трудоемкая, но ее результат часто оказывается полезным. Разлагая $E_0(\rho, \mathbf{Q})$ в окрестности $\rho = 0$ с точностью до членов второго порядка, получаем

$$E_0(\rho, \mathbf{Q}) = \rho C + \frac{\rho^2}{2} E''_0(\rho_1, \mathbf{Q})$$

при некотором ρ_1, $0 \leqslant \rho_1 \leqslant \rho$,

$$E_0(\rho, \mathbf{Q}) \leqslant \rho C - \frac{\rho^2}{2}\alpha,$$

$$E_r(R, \mathbf{Q}) \geqslant E_0(r, \mathbf{Q}) - \rho R \geqslant \rho(C-R) - \frac{\rho^2}{2}\alpha;\ 0 \leqslant \rho \leqslant 1,$$

$$E_r(R, \mathbf{Q}) \geqslant \frac{(C-R)^2}{2\alpha} \text{ при } (C-R) \leqslant \alpha, \tag{1}$$

где было положено $\rho = (C-R)/\alpha$. Теперь $E_0(\rho, \mathbf{Q})$ можно представить в виде

$$\left.\begin{aligned} E_0(\rho, \mathbf{Q}) &= -\ln \sum_j \alpha_j^{1+\rho}, \\ \alpha_j &= \sum_k Q(k) P(j \mid k)^{1/(1+\rho)}, \end{aligned}\right\} \tag{2}$$

$$-E_0'(\rho, \mathbf{Q}) = \frac{\sum_j a_j^{1+\rho}\left[\ln \alpha_j + \frac{(1+\rho)}{\alpha_j}\frac{\partial \alpha_j}{\partial \rho}\right]}{\sum_j \alpha_j^{1+\rho}} =$$

$$= \sum_j \omega_j \left[\ln \alpha_j - \sum_k q_{kj} \ln P(j \mid k)^{1/(1+\rho)}\right] = \sum_{j,k} \omega_j q_{kj} \ln \frac{Q(k)}{q_{kj}}, \qquad (3)$$

где

$$\omega_j = \frac{\alpha_j^{1+\rho}}{\sum_j \alpha_j^{1+\rho}}; \quad q_{kj} = \frac{Q(k) P(j \mid k)^{1/(1+\rho)}}{\alpha_j}. \qquad (4)$$

Найдем теперь $-E_0''(\rho, Q)$, используя (3),

$$-E_0''(\rho, \mathbf{Q}) = -\sum_{j,k} \frac{\partial}{\partial \rho}(\omega_j q_{kj}) \ln \frac{Q(k)}{q_{kj}} - \sum_{j,k} \omega_j \frac{\partial}{\partial \rho}(q_{kj}). \qquad (5)$$

Так как $\sum_k q_{kj} = 1$ при всех ρ, j, то можно заметить, что $\sum_k \frac{\partial}{\partial \rho} q_{kj} = 0$ и вторая из написанных выше сумм равна нулю. Используя (4) (а также (2), чтобы ввести член E_0'), будем иметь

$$\frac{\partial}{\partial \rho}(\omega_j q_{kj}) = \omega_j q_{kj}\left[\ln \alpha_j + \frac{\rho \alpha_j'}{\alpha_j} - \frac{1}{1+\rho} \ln P(j \mid k)^{1/(1+\rho)} - E_0'(\rho, \mathbf{Q})\right] =$$

$$= \omega_j q_{kj}\left[\ln \alpha_j - \frac{\rho}{1+\rho} \sum_i q_{ij} \ln P(j \mid i)^{1/(1+\rho)} - \frac{1}{1+\rho} \ln P(j \mid k)^{1/(1+\rho)} - \right.$$

$$\left. - E_0'(\rho, \mathbf{Q})\right] = \omega_j q_{kj}\left[\frac{\rho}{1+\rho} \sum_i q_{ij} \ln \frac{Q(i)}{q_{ij}} + \frac{1}{1+\rho} q_{kj} \ln \frac{Q(k)}{q_{kj}} - E_0'(\rho, \mathbf{Q})\right].$$

Подставляя это в (5), получаем

$$-E_0''(\rho, \mathbf{Q}) = \sum_j \omega_j \frac{\rho}{1+\rho}\left[\sum_k q_{kj} \ln \frac{Q(k)}{q_{kj}}\right]^2 + \frac{1}{1+\rho} \sum_{k,j} \omega_j q_{kj}\left[\ln \frac{Q(k)}{q_{kj}}\right]^2 -$$

$$- [E_0'(\rho, \mathbf{Q})]^2.$$

Используя неравенство задачи 4.15 (г) для выражения в первых квадратных скобках и затем преобразуя первые два выражения, будем иметь

$$-E_0''(\rho, \mathbf{Q}) \leqslant \sum_{k,j} \omega_j q_{kj}\left[\ln \frac{Q(k)}{q_{kj}}\right]^2 - [E_0'(\rho, \mathbf{Q})]^2. \qquad (6)$$

Так как последнее выражение отрицательно, то оно может быть отброшено и в результате получим

$$-E_0''(\rho, \mathbf{Q}) \leqslant \sum_{k,j} \omega_j q_{kj}\left(\ln \frac{Q(k)}{q_{kj}}\right)^2. \qquad (7)$$

Заметим, что $-[E_0'(1, \mathbf{Q})]^2$ всегда можно вновь поместить в окончательную границу для $-E_0''$, если возникает необходимость иметь более сложную, но более точную границу.

Далее используем то, что $(\ln x)^2 < \frac{4}{e^2} x$ при $x \geqslant 1$. Это можно показать, замечая, что выражение $(\ln x)^2/x$ в области $x \geqslant 1$ достигает максимума, равного $4/e^2$. Таким образом,

$$-E_0''(\rho, \mathbf{Q}) \leqslant \frac{4}{e^2} \sum_{k,j} \omega_j q_{kj} \frac{Q(k)}{q_{kj}} + \sum_{\substack{k,j: \\ Q(k)/q_{kj} < 1}} \omega_j q_{kj}\left(\ln \frac{Q(k)}{q_{kj}}\right)^2. \qquad (8)$$

Используя также (4), получаем

$$\left(\ln\frac{Q(k)}{q_{kj}}\right)^2=\left[\ln\frac{\alpha_j}{P(j\mid k)^{1/(1+\rho)}}\right]^2\leqslant(\ln\alpha_j)^2, \tag{9}$$

где последнее неравенство справедливо при $Q(k)/q_{kj}<1$ в силу того, что в этом случае отрицательная величина внутри квадратных скобок уменьшается.

Из (4) и (2) имеем

$$\ln\alpha_j=\frac{1}{1+\rho}\ln\alpha_j^{1+\rho}=\frac{1}{1+\rho}[\ln\omega_j-E_0(\rho,\mathbf{Q})]. \tag{10}$$

Сочетая (8)—(10), получаем

$$-E_0''(\rho,\mathbf{Q})\leqslant\frac{4}{e^2}+\sum_{\substack{k,j:\\ Q(k)/q_{kj}<1}}\omega_j q_{kj}\left[\frac{\ln\omega_j-E_0(\rho,\mathbf{Q})}{1+\rho}\right]^2\leqslant$$

$$\leqslant\frac{4}{e^2}+\sum_j\omega_j\left[\frac{\ln\omega_j-E_0(\rho,\mathbf{Q})}{1+\rho}\right]^2.$$

Можно произвести дальнейшую оценку этого выражения, максимизируя его по ω_j при условии, что $\sum\omega_j=1$. Отметим, что максимум достигается при $\omega_j=1/J$, давая

$$-E_0''(\rho,\mathbf{Q})\leqslant\frac{4}{e^2}+\left[\frac{\ln J+E_0(\rho,\mathbf{Q})}{1+\rho}\right]^2. \tag{11}$$

Заметим, наконец, что $E_0(\rho,\mathbf{Q})\leqslant\rho C\leqslant\rho\ln J$. Подставляя в (11), получаем $-E_0''(\rho,\mathbf{Q})\leqslant\frac{4}{e^2}+(\ln J)^2$. Используя это неравенство для оценки α в (1) и замечая, что $C-R\leqslant\alpha$ для всех R, $0\leqslant R\leqslant C$, получаем

$$E_r(R\ \mathbf{Q})\geqslant\frac{(C-R)^2}{\frac{8}{e^2}+2(\ln J)^2}. \tag{12}$$

Отметим, что эта граница немного точнее того результата, который следовало доказать.

5.24. В силу того, что $E_x(\rho,\mathbf{Q})$ — неубывающая функция ρ, и так как $E_{ex}(R,\mathbf{Q})=\sup\limits_{\rho\geqslant1}[E_x(\rho,\mathbf{Q})-\rho R]$, получим $E_{ex}(R,\mathbf{Q})\leqslant\lim\limits_{\rho\to\infty}E_x(\rho,\mathbf{Q})$ при всех R. Кроме того, если $\lim\limits_{\rho\to\infty}E_x(\rho,\mathbf{Q})$ конечен, то при любом $\varepsilon>0$ можно выбрать ρ так, чтобы $E_x(\rho,\mathbf{Q})\geqslant\lim\limits_{\rho\to\infty}E_x(\rho,\mathbf{Q})-\varepsilon/2$ и поэтому при R, достаточно малом, $E_x(\rho,\mathbf{Q})-\rho R\geqslant\lim\limits_{\rho\to\infty}E_x(\rho,\mathbf{Q})-\varepsilon$. Следовательно, при всех достаточно малых $R>0$ имеем

$$\lim_{\rho\to\infty}E_x(\rho,\mathbf{Q})\geqslant E_{ex}(R,\mathbf{Q})\geqslant\lim_{\rho\to\infty}E_x(\rho,\mathbf{Q})-\varepsilon$$

и

$$\lim_{R\to0}E_{ex}(R,\mathbf{Q})=\lim_{\rho\to\infty}E_x(\rho,\mathbf{Q}).$$

Положив $\delta=1/\rho$, получим

$$E_x(1/\delta,\mathbf{Q})=\frac{-\ln\sum\limits_{k,i}Q(k)Q(i)\left[\sum\limits_j\sqrt{P(j\mid k)P(j\mid i)}\right]^\delta}{\delta}.$$

Замечая, что сумма по j является строго положительной при всех i, k (см. стр. 171), можно найти предел при $\delta \to 0$ по правилу Лопиталя; в результате получим

$$\lim E_x(1/\delta, \mathbf{Q}) = -\sum_{k,i} Q(k) Q(i) \ln \sum_j \sqrt{P(j|k) P(j|i)}.$$

5.25. Так как $\varphi_{k,i} = 1$ при $\sum_j \sqrt{P(j|k) P(j|i)} > 0$ и $\varphi_{k,i} = 0$ во всех остальных случаях, то $\varphi_{k,i} = \varphi_{i,k}$. Также $\varphi_{1,1} = \varphi_{0,0} = 1$. Следовательно,

$$\sum_{k,i} Q(k) Q(i) \varphi_{k,i} = Q^2(0) + 2Q(0) Q(1) + Q^2(1) +$$

$$+ \sum_{k=2}^{K-1} \sum_{i=0}^{1} Q(k) Q(i) \varphi_{k,i} + \sum_{k=0}^{1} \sum_{i=2}^{K-1} Q(k) \tilde{Q}(i) \varphi_{k,i} +$$

$$+ \sum_{k>2} \sum_{i>2} Q(k) Q(i) \varphi_{k,i} = [Q(0) + Q(1)]^2 +$$

$$+ 2 \sum_{i=2}^{K-1} Q(i) [Q(0) \varphi_{0,i} + Q(1) \varphi_{1,i}] + \sum_{k>2} \sum_{i>2} Q(k) Q(i) \varphi_{k,i}.$$

Если $Q(2), \ldots, Q(K-1)$ зафиксированы, а $Q(0)$ и $Q(1)$ варьируются при постоянной их сумме (так, что $\sum Q(k) = 1$), то заметим, что написанная выше функция является линейной при этой вариации. Поэтому минимум при $Q(0) \geqslant 0$, $Q(1) \geqslant 0$ достигается либо при $Q(0) = 0$, либо при $Q(1) = 0$. Отсюда видно, что для любого вектора вероятностей $\mathbf{Q}$ можно уменьшить (или не изменить) $\sum Q(k) Q(i) \varphi_{k,i}$ с помощью последовательного пересмотра каждой пары i, k ($i \neq k$), для которой $\varphi_{k,i} = 1$, и замены либо $Q(k)$, либо $Q(i)$ на 0. В конце этой процедуры множество i (обозначим его через I), для которых $Q(i) > 0$, удовлетворяет условию $\varphi_{k,i} = 0$ при всех $k \in I$, $i \in I$, $i \neq k$. Для такого множества I

$$\sum_{k,i} Q(k) Q(i) \varphi_{k,i} = \sum_{i \in I} Q^2(i).$$

В этом множестве $\sum_{i \in I} Q^2(i)$ достигает минимума, когда все $Q(i)$, $i \in I$, равны друг другу. Поэтому минимум достигается на некотором множестве I (допустим, состоящим из L элементов) и равен $1/L$. Очевидно, что минимум достигается на наибольшем таком множестве и

$$R_{X,\infty} = \max_{\mathbf{Q}} -\ln \sum_{k,i} Q(k) Q(i) \varphi_{k,i} = \ln L.$$

5.26. Заметим, что $R_{X,\infty}$ из задачи 5.25 равен $\ln 2$. Поэтому при любой $R > \ln 2$ и любом распределении, которое является произведением распределений, выражение $E_{ex}(R, \mathbf{Q})$ является конечным. Вместе с тем, если рассмотреть пары букв в канале как буквы в произведении каналов, то, используя задачи 5.11 (б) и 5.25, можно найти, что $R_{X,\infty}$ для произведения каналов равна $\ln 5$. Производя нормировку на буквы первоначального канала, отсюда получим скорость $(1/2) \ln 5$. Поэтому, используя входное распределение, при котором последовательные пары букв будут статистически зависимыми, получим, что граница (5.7.7) равна 0 при $R < (1/2) \ln 5$.

5.27. Первая часть задачи представляет собой часть теоремы 5.7.2 и она доказана в приложении 5Б; однако $E_x(\rho, \mathbf{Q})$ равно нулю, если $\mathcal{I}(\mathbf{Q}; \mathbf{P})$ равно нулю, и это непосредственно следует из определения. Имеем

$$\frac{\partial E_x(\rho, \mathbf{Q})}{\partial \rho} = -\ln \sum_{k,i} Q(k) Q(i) \left[\sum_j \sqrt{P(j|k) P(j|i)}\right]^{1/\rho} +$$

$$+\frac{\sum\limits_{k,\,j} Q(k)\,Q(i)\left[\sum\limits_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}\ln\left[\sum\limits_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}}{\sum\limits_{k,\,i} Q(k)\,Q(i)\left[\sum\limits_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}}.$$

В пределе при $\rho \to 0$ достигается максимум, так как $E_x(\rho, \mathbf{Q})$ является выпуклой $\cap$. В этом пределе второе из написанных выше выражений стремится к нулю, а первое выражение может быть ограничено сверху, если заметить, что

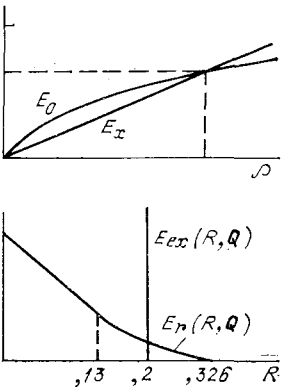
$$\lim_{\rho\to 0}\left[\sum_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho}\begin{cases}=1 & \text{при } k=i,\\ >0 & \text{при } k\neq i.\end{cases}$$

Используя эти границы, получаем

$$\frac{\partial E_x(\rho, \mathbf{Q})}{\partial\rho} < \frac{\partial E_x(\rho, \mathbf{Q})}{\partial\rho}\Big|_{\rho=0} < -\ln\sum_{k} Q(k)^2.$$

5.28. В ДСК без шума при $\mathbf{Q} = (0{,}1;\ 0{,}9)$ имеем

$$E_x(\rho, \mathbf{Q}) = -\rho\ln[Q(0)^2 + Q(1)^2] = -\rho\ln 0{,}82,$$
$$E_0(\rho, \mathbf{Q}) = -\ln[Q(0)^{1+\rho} + Q(1)^{1+\rho}].$$



Главное в этой задаче состоит в том, чтобы показать, что если на $\mathbf{Q}$ не достигается максимум $E_r(R, \mathbf{Q})$, то можно получить $E_{ex}(R, \mathbf{Q}) > E_r(R, \mathbf{Q})$ при $R > \dfrac{\partial E_0(\rho, \mathbf{Q})}{\partial\rho}\Big|_{\rho=1}$. Это невозможно для оптимального $\mathbf{Q}$, так как функция надежности равна $E_r(R, \mathbf{Q})$ при $R > \partial E_0(\rho, \mathbf{Q})/\partial\rho|_{\rho=1}$.

5.29. В ДКБП с двумя входами

$$E_x(\rho, \mathbf{Q}) = -\rho\ln\sum_{k=0}^{1}\sum_{i=0}^{1} Q(k)\,Q(i)\left[\sum_{j}\sqrt{P(j\mid k)\,P(j\mid i)}\right]^{1/\rho} =$$
$$= -\rho\ln\left\{Q^2(0) + Q^2(1) + 2Q(0)\,Q(1)\left[\sum_{j}\sqrt{P(j\mid 1)\,P(j\mid 0)}\right]^{1/\rho}\right\}. \qquad (1)$$

Подставляя $Q(1) = 1 - Q(0)$ и дифференцируя выражение, стоящее в фигурных скобках, по $Q(0)$, находим стационарную точку $Q(0) = 1/2$. Вторая производная выражения в фигурных скобках равна

$$4\left\{1 - \left[\sum_{j}\sqrt{P(j\mid 1)\,P(j\mid 0)}\right]^{1/\rho}\right\}. \qquad (2)$$

В соответствии с задачей 4.15 (а) выражение в квадратных скобках в (2) меньше, чем 1, если только не имеет место равенство $P(j|1) = P(j|0)$ при всех j. Следовательно, $Q(0) = {}^1/_2$ минимизирует выражение, стоящее в фигурных скобках в (1), и поэтому максимизирует $E_x(\rho, \mathbf{Q})$ при всех $\rho > 0$.

5.30. (а) Матрица A с элементами

$$a_{ik} = \left[\sum_j \sqrt{P(j|k)\,P(j|i)}\right]^{1/\rho}$$

является в соответствии с гипотезой неотрицательно определенной. Это означает, что для любого вектора $\mathbf{x} = (x_0, \ldots, x_{K-1})$

$$\sum_{k,\,i} x_k\, x_i\, a_{ik} \geqslant 0. \tag{1}$$

Нужно показать, что $\sum\limits_{k,\,i} Q(k)\,Q(i)\,a_{ik}$ является выпуклой $\cup$ функцией вектора вероятностей $\mathbf{Q}$. Из задачи 4.11 следует, что для этого достаточно показать, что для любых векторов вероятностей $\mathbf{Q}$ и $\mathbf{q}$

$$\sum_{k,\,i} [\lambda Q(k) + (1-\lambda)\,q(k)]\;[\lambda Q(i) + (1-\lambda)\,q(i)]\,a_{ik} \tag{2}$$

является выпуклой $\cup$ функцией λ. Вторая производная от (2) по λ равна

$$2\sum_{k,\,i} [Q(k) - q(k)]\,[Q(i) - q(i)]\,a_{ik}.$$

Из (1) следует, что это выражение неотрицательно, так что $\sum\limits_{k,\,i} Q(k)\,Q(i)\,a_{ik}$ является выпуклой $\cup$ функцией.

(б) Поскольку A является симметричной $K \times K$ матрицей, то она имеет K ортогональных собственных векторов, допустим $\xi_m = (\xi_m(0), \ldots, \xi_m(K-1))$, $1 \leqslant m \leqslant K$, с собственными значениями λ_m, $1 \leqslant m \leqslant K$:

$$\sum_{k=0}^{K-1} a_{ik}\,\xi_m(k) = \lambda_m\,\xi_m(i); \qquad 1 \leqslant m \leqslant K,$$

$$\sum_{k=0}^{K-1} \xi_m(k)\,\xi_{m'}(k) = 0; \qquad m \neq m'.$$

Рассмотрим теперь $K^N \times K^N$-матрицу B со строкой $\mathbf{x}$ и столбцом $\mathbf{x}'$ для каждой входной последовательности и

$$\left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}|\mathbf{x})\,P_N(\mathbf{y}|\mathbf{x}')}\right]^{1/\rho} =$$

$$= \prod_{n=1}^{N} \left[\sum_{y_n} \sqrt{P(y_n|x_n)\,P(y_n|x_n')}\right]^{1/\rho}$$

в качестве элемента, стоящего на пересечении строки $\mathbf{x}$ и столбца $\mathbf{x}'$. Для заданной последовательности $\xi_{m_1}, \xi_{m_2}, \ldots, \xi_{m_N}$ введенных выше собственных векторов пусть η будет вектором с K^N компонентами $\eta(\mathbf{x}) = \xi_{m_1}(x_1)\,\xi_{m_2}(x_2) \ldots \xi_{m_N}(x_N)$ (т. е. одна компонента для каждой входной последовательности $\mathbf{x}$). Теперь

$$\sum_{\mathbf{x}} \eta(\mathbf{x}) \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y}|\mathbf{x})\,P_N(\mathbf{y}|\mathbf{x}')}\right]^{1/\rho} =$$

$$= \prod_{n=1}^{N} \sum_{x_n} \xi_{m_n}(x_n) \left[\sum_{y_n} \sqrt{P(y_n \mid x_n) P(y_n \mid x_n')} \right]^{1/\rho} =$$

$$= \prod_{n=1}^{N} \lambda_{m_n} \xi_{m_n}(x_n') = \left[\prod_{n=1}^{N} \lambda_{m_n} \right] \eta(\mathbf{x}').$$

Поэтому $\eta(\mathbf{x})$ является собственным вектором B с собственным значением $\prod_{n=1}^{N} \lambda_{m_n} \geqslant 0$. Можно получить K^N таких собственных векторов с помощью рассмотрения всех возможных наборов $m_1, m_2, \ldots, m_N$ и, если

$$\eta'(\mathbf{x}) = \xi_{m_1'}(x_1) \ldots \xi_{m_N'}(x_N),$$

то

$$\sum_{\mathbf{x}} \eta(\mathbf{x}) \eta'(\mathbf{x}) = \prod_{n=1}^{N} \sum_{x_n} \xi_{m_n}(x_n) \xi_{m_n'}(x_n) = 0,$$

если $m_n \neq m_n'$ для некоторого n, $1 \leqslant n \leqslant N$.

Таким образом, найдены K^N ортогональных собственных векторов матрицы B, каждый из которых соответствует неотрицательному собственному значению. Поэтому B является неотрицательно определенной матрицей и

$$\sum_{\mathbf{x}, \mathbf{x}'} Q_N(\mathbf{x}) Q_N(\mathbf{x}') \left[\sum_{\mathbf{y}} \sqrt{P_N(\mathbf{y} \mid \mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x}')} \right]^{1/\rho}$$

является выпуклой $\cup$ функцией по $\mathbf{Q}$. Теперь теорема 4.4.1 дает необходимые и достаточные условия для максимума этой функции, взятой со знаком минус; непосредственно проверяется то, что произведение распределений $Q_N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n)$ удовлетворяет этим условиям, если $Q(x_n)$ выбрано так, чтобы максимизировать

$$-\sum_{k, i} Q(k) Q(i) \left[\sum_{j} \sqrt{P(j \mid k) P(j \mid i)} \right]^{1/\rho}.$$

5.31. Напомним, что согласно задаче 5.24

$$\lim_{R \to 0} E_{ex}(R, \mathbf{Q}) = -\sum_{k, i} Q(k) Q(i) \ln \sum_{j} \sqrt{P(j \mid k) P(j \mid i)}. \tag{1}$$

Так же как и в примере 3 на стр. 163, определим ε_{jk} с помощью равенства

$$P(j \mid k) = \omega_j (1 + \varepsilon_{jk}), \tag{2}$$

где ω_j является распределением вероятностей и $\varepsilon_{jk} \ll 1$. Производя разложение в ряд до второго порядка по ε_{jk}, получаем

$$\ln \sum_{j} \sqrt{P(j \mid k) P(j \mid i)} = \ln \sum_{j} \omega_j \sqrt{(1 + \varepsilon_{jk})(1 + \varepsilon_{ji})} \approx$$

$$\approx \ln \sum_{j} \omega_j \left[1 + \frac{\varepsilon_{jk} + \varepsilon_{ji}}{2} + \frac{\varepsilon_{jk} \varepsilon_{ji}}{4} - \frac{\varepsilon_{jk}^2 + \varepsilon_{ji}^2}{8} \right] = \tag{3}$$

$$= \ln \left\{ 1 + \sum_{j} \omega_j \left[\frac{\varepsilon_{jk} \varepsilon_{ji}}{4} - \frac{\varepsilon_{jk}^2 + \varepsilon_{ji}^2}{8} \right] \right\}, \tag{4}$$

$$\ln \sum_{j} \sqrt{P(j \mid k) P(j \mid i)} \approx \sum_{j} \omega_j \left[\frac{\varepsilon_{jk} \varepsilon_{ji}}{4} - \frac{\varepsilon_{jk}^2 + \varepsilon_{ji}^2}{8} \right]. \tag{5}$$

При переходе от (3) к (4) было использовано (2) для того, чтобы убедиться, что $\sum_j \omega_j \varepsilon_{jk} = 0$. Подставляя выражение (5) в (1) и производя соответствующие преобразования, получаем

$$\lim_{R \to 0} E_{ex}(R, \mathbf{Q}) = -\sum_j \frac{\omega_j}{4} \left[\left(\sum_k Q(k)\, \varepsilon_{jk} \right)^2 - \sum_k Q(k)\, \varepsilon_{jk}^2 \right]. \qquad (6)$$

Сравнивая это с (5.6.48) и (5.6.49), будем иметь

$$\lim_{R \to 0} E_{ex}(R, \mathbf{Q}) = E_0(1, \mathbf{Q}) \text{ с точностью до члена второго порядка по } \varepsilon_{jk}. \qquad (7)$$

Так как $E_{ex}(R, \mathbf{Q})$ уменьшается при увеличении R быстрее, чем $E_r(R, \mathbf{Q})$, то из этого следует, что $E_{ex}(R, \mathbf{Q}) = E_r(R, \mathbf{Q})$ при $R \leqslant R_{cr}$, по крайней мере, с той же точностью, что и приближение, сделанное в (7).

5.32. Поскольку $P_e(N, M)$ является минимумом $\frac{1}{M} \sum_{m=1}^{M} P_{e,m}$ по всем кодам с длиной блока N и с M кодовыми словами и так как для каждого такого кода $P_{max} \geqslant \frac{1}{M} \sum_{m=1}^{M} P_{e,m}$, то получим

$$P_{max}(N, M) \geqslant P_e(N, M).$$

Теперь при заданном M рассмотрим код с $2M$ кодовыми словами и длиной блока N и вероятностью ошибки $P_e(N, 2M)$. Устраним M кодовых слов, для которых $P_{e,m}$ является наибольшей. Заметим, что невозможно, чтобы $P_{e,m}$ была больше, чем $2P_e(N, 2M)$ для каждого из этих устраненных слов. Поэтому для каждого из оставшихся слов при первоначально заданных областях декодирования имеем $P_{e,m} \leqslant 2P_e(N, 2M)$. Отсюда следует, что $P_{max}(N, M) \leqslant 2P_e(N, 2M)$.

5.33. Декодер отображает выходные последовательности в сообщения и $\mathbf{y} \in Y_m$ тогда и только тогда, когда $\mathbf{y}$ отображается в сообщение m. Поэтому области декодирования $Y_1, \ldots, Y_M$ являются непересекающимися. Далее заметим, что для данного сообщения m и принятой последовательности $\mathbf{y}$ каждый символ переданного кодового слова однозначно определяется по m и предыдущим принятым символам. Следовательно, m и $\mathbf{y}$ однозначно задают кодовое слово $\mathbf{x}_m(\mathbf{y})$. Это значит, что при заданном m только шумовая последовательность $\mathbf{y} \oplus \mathbf{x}_m(\mathbf{y})$ может привести к тому, чтобы было принято $\mathbf{y}$ и, таким образом, различные $\mathbf{y} \in Y_m$ соответствуют различным шумовым последовательностям. Пусть теперь $M_{n,m}$ будет числом шумовых последовательностей с n единицами, которые правильно декодируются в случае, когда на кодер поступает сообщение m. Так как Y_m являются непересекающимися и каждое $\mathbf{y} \in Y_m$ соответствует различным шумовым последовательностям, то ограничения (5.8.15) и (5.8.16) для $A_{n,m}$ выполняются. Граница сферической упаковки при наличии обратной связи совпадает с этой границей без обратной связи, которая представлена (5.8.19).

5.34. Из (5.8.69) имеем

$$A = \max_k \sum_j P(j \mid k) \left[\ln \frac{\omega(j)}{P(j \mid k)} \right]^2, \qquad (1)$$

где $\omega(j) = \sum_k Q(k) P(j \mid k)$ с $\mathbf{Q}$, на котором достигается пропускная способность. При $\omega(j) \geqslant P(j \mid k)$ используем неравенство $(\ln x)^2 \leqslant \frac{4}{e^2} x$ для $x \geqslant 1$.

Это неравенство следует непосредственно из того, что $(\ln x)^2 - \frac{4}{e^2} x$ при $x \geqslant 1$ имеет минимум, равный 0, который достигается в точке $x = e^2$. При $\omega_j < P(j \mid k)$ получаем

$$\left[\ln \frac{\omega(j)}{P(j \mid k)} \right]^2 = \left[\ln \frac{P(j \mid k)}{\omega(j)} \right]^2 \leqslant \left[\ln \frac{P(j \mid k)}{\omega(j)} \right] \left[\ln \frac{1}{\omega(j)} \right].$$

Используя эти неравенства и полагая $B_k = \{j : \omega(\,) < P(j \mid k)\}$, выводим

$$\sum_j P(j \mid k) \left[\ln \frac{\omega(j)}{P(j \mid k)} \right]^2 \leqslant \frac{4}{e^2} \sum_{B_k^c} \omega(j) + \sum_{B_k} P(j \mid k) \ln \frac{P(j \mid k)}{\omega(j)} \ln \frac{1}{\omega(j)} \leqslant$$

$$\leqslant \frac{4}{e^2} + \left[\max_j \ln \frac{1}{\omega(j)} \right] \sum_{B_k} P(j \mid k) \ln \frac{P(j \mid k)}{\omega(j)}. \tag{2}$$

Из (5.8.60) будем иметь

$$\sum_j P(j \mid k) \ln \frac{P(j \mid k)}{\omega(j)} \leqslant C \quad \text{при всех } k, \tag{3}$$

$$\sum_j P(j \mid k) \ln \frac{1}{\omega(j)} \leqslant C + \sum_j P(j \mid k) \ln \frac{1}{P(j \mid k)} \leqslant 2 \ln J \quad \text{при всех } k.$$

Таким образом,

$$\ln \frac{1}{\omega(j)} \leqslant \frac{2 \ln J}{P(j \mid k)} \quad \text{при всех } k. \tag{4}$$

Отсюда следует, что (4) справедливо для k, которое максимизирует $P(j \mid k)$. Имеем

$$\max_j \ln \frac{1}{\omega(j)} \leqslant \max_j \left[\frac{\ln J}{\max_k P(j \mid k)} \right] = \frac{2 \ln J}{\min_j [\max_k P(j \mid k)]}. \tag{5}$$

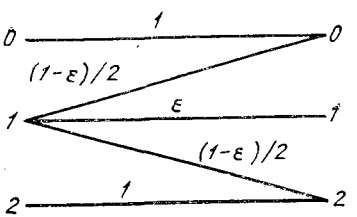
Для последнего выражения в (2), используя (3), будем иметь

$$\sum_{B_k} P(j \mid k) \ln \frac{P(j \mid k)}{\omega(j)} \leqslant C + \sum_{B_k^c} P(j \mid k) \ln \frac{\omega(j)}{P(j \mid k)}. \tag{6}$$

Используя неравенство $\ln x \leqslant \frac{2}{e} x$ при $x \geqslant 1$, получаем следующую границу сверху для (6):

$$\sum_{B_k} P(j \mid k) \ln \frac{P(j \mid k)}{\omega(j)} \leqslant C + \frac{2}{e} \leqslant \ln J + \frac{2}{e}. \tag{7}$$

Подставляя (5) и (7) в (2), получаем искомый результат. Для того чтобы понять, почему нельзя выразить границу для A лишь через объем алфавита, рассмотрим изображенный ниже канал, где ε произвольно мало.



После некоторых вычислений можно найти, что для пропускной способности нужно, чтобы Q (1) $\approx e^{-1/\varepsilon}$. Более того, максимизирующим входом при отыскании A будет $k = 1$ и $A \approx 1/\varepsilon$. Таким образом, A нельзя ограничить лишь в терминах объема алфавита.

5.35. Применяя границу Чернова (5.4.15) к (5.8.66), получаем

$$\sum_{\mathbf{y}\in B_m} P_N(\mathbf{y}\mid \mathbf{x}_m) \leqslant e^{-sN(C+\varepsilon)}\,[g(s)]^N \quad \text{при любом } s\geqslant 0, \tag{1}$$

$$g(s)=\sum_{k,\,j} Q(k)\,P(j\mid k)\exp\left\{s\ln\frac{P(j\mid k)}{\sum\limits_i Q(i)\,P(j\mid i)}\right\}.$$

Границу (1) можно переписать в виде

$$\sum_{\mathbf{y}\in B_m} P_N(\mathbf{y}\mid \mathbf{x}_m) \leqslant \exp\{-N[s(C+\varepsilon)-\ln g(s)]\},\quad s\geqslant 0. \tag{2}$$

Так как $g(s)$ является производящей функцией моментов взаимной информации, то

$$\left.\frac{d\ln g(s)}{ds}\right|_{s=0}=C.$$

Поэтому при любом $\varepsilon>0$ значение $s(C+\varepsilon)-\ln g(s)$ является положительным для достаточно малого $s>0$. Подставляя (2) и (5.8.65) в (5.8.64), получаем

$$P_e\leqslant\frac{\exp[N(C+\varepsilon)]}{M}+\exp\{-N[s(C+\varepsilon)-\ln g(s)]\}.$$

Для любой скорости $R>C$ пусть $M=\lceil e^{NR}\rceil$ и пусть $\varepsilon=(R-C)/2$. Тогда

$$P_e(N,\ \lceil e^{NR}\rceil)\geqslant 1-\exp\left[-N\left(\frac{R-C}{2}\right)\right]+$$

$$+\exp\left\{-N\left[s\left(C+\frac{R-C}{2}\right)-\ln g(s)\right]\right\}\geqslant 1-2\exp[-N\alpha(R)],$$

$$\alpha(R)=\min\left[\frac{R-C}{2},\ \max_{s\geqslant 0}\left[s\left(C+\frac{R-C}{2}\right)-\ln g(s)\right]\right]>0.$$

5.36. При заданном кодовом слове $\mathbf{x}_m$ взаимная информация $I(\mathbf{x}_m;\mathbf{y})$ является суммой N независимых случайных величин:

$$I(\mathbf{x}_m;\mathbf{y})=\sum_{n=1}^{N} I(x_{m,n};y_n), \tag{1}$$

где $x_{m,n}$ фиксированно, а y_n выбирается в соответствии с распределением вероятности $P(y_n\mid x_{m,n})$. Из (5.8.60) следует, что каждая из этих случайных величин имеет среднее значение, не большее, чем C так, что $I(\mathbf{x}_m;Y^N)$ (среднее значение $I(\mathbf{x}_m;\mathbf{y})$ при фиксированном $\mathbf{x}_m$) удовлетворяет неравенству

$$NC\geqslant I(\mathbf{x}_m;Y^N). \tag{2}$$

Введем теперь обозначение

$$A(x_{m,n})=\mathbf{D}[I(x_{m,n};y_n)]. \tag{3}$$

Отметим, что $A(x_{m,n})$ определяется лишь тем, какой буквой алфавита является $x_{m,n}$. На основании теоремы Берри-Эссена имеем

$$\Pr[I(\mathbf{x}_m;\mathbf{y})\geqslant NC+\delta_1\sqrt{N}\mid\mathbf{x}_m]\leqslant$$

$$\leqslant\Pr[I(\mathbf{x}_m;\mathbf{y})\geqslant I(\mathbf{x}_m;\mathbf{Y}^N)+\delta_1\sqrt{N}\mid\mathbf{x}_m]\leqslant$$

$$\leqslant 1-\Phi\left[\frac{\delta_1\sqrt{N}}{\sqrt{\sum\limits_{n=1}^{N}A(x_{m,n})}}\right]+\frac{33}{4}\,\frac{\lambda}{\sum\limits_{n=1}^{N}A(x_{m,n})}, \tag{4}$$

где Φ является функцией распределения нормированной гауссовской случайной величины, а λ — максимальное значение по $x_{m,n}$ третьего абсолютного центрального момента $I(x_{m,n}; y_n)$, разделенного на $A(x_{m,n})$. Так как буквы $x_{m,n}$ выбираются из конечного алфавита, то можно рассматривать λ как фиксированную положительную постоянную, не зависящую от $\mathbf{x}_m$. По той же самой причине можно определить $\underline{A}$ как $\min\limits_{0\leqslant k\leqslant K-1} A(k)$ и $\overline{A}$ как $\max\limits_{0\leqslant k\leqslant K-1} A(k)$. Так как по предположению $\ln[P(j|k)/\omega(j)]$ не зависит от j при каждом k, то $\underline{A}>0$ и

$$0<\underline{A}\leqslant A(x_{m,n})\leqslant \overline{A}<\infty \quad \text{при всех } x_{m,n}. \tag{5}$$

Используя (5) для того, чтобы ограничить сверху (4), получаем

$$\Pr[I(\mathbf{x}_m; \mathbf{y}) \geqslant NC+\delta_1\sqrt{N}\,|\,\mathbf{x}_m] \leqslant \begin{cases} 1-\Phi\left(\dfrac{\delta_1}{\sqrt{\overline{A}}}\right)+\dfrac{33}{4}\dfrac{\lambda}{\sqrt{N\underline{A}}}; & \delta_1>0; \\ 1-\Phi\left(\dfrac{\delta_1}{\sqrt{\underline{A}}}\right)+\dfrac{33}{4}\dfrac{\lambda}{\sqrt{N\underline{A}}}; & \delta_1<0. \end{cases} \tag{6}$$

Пусть теперь $R=C+\delta/\sqrt{N}$ и выберем $\delta_1=\delta-N^{-1/4}$ и пусть число кодовых слов равно $M=\lceil e^{NR}\rceil$. Рассмотрим R, δ и M как функции N при фиксированном δ. Для любого заданного $\delta>0$ выберем N достаточно большим, так, чтобы $\delta_1>0$. Подставляя (6) в (5.8.66) и (5.8.65) и (5.8.66) в (5.8.62) (с $\varepsilon=\delta_1/\sqrt{N}$), имеем

$$1-P_e \leqslant 1-\Phi\left(\frac{\delta-N^{-1/4}}{\sqrt{\overline{A}}}\right)+\frac{33}{4}\frac{\lambda}{\sqrt{N\underline{A}}}+ \\ +\exp\left\{N\left[C+\frac{\delta_1}{\sqrt{N}}-C-\frac{\delta}{\sqrt{N}}\right]\right\},$$

$$P_e \geqslant \Phi\left(\frac{\delta-N^{-1/4}}{\sqrt{\overline{A}}}\right)-\frac{33}{4}\frac{\lambda}{\sqrt{N\underline{A}}}-\exp[-N^{1/4}].$$

Поэтому для любого $\varepsilon>0$ и достаточно большого N получим

$$P_e \geqslant \Phi\left(\frac{\delta}{\sqrt{\overline{A}}}\right)-\varepsilon, \quad \delta>0.$$

С помощью тех же рассуждений для любого $\delta\leqslant 0$ получим

$$P_e \geqslant \Phi\left(\frac{\delta}{\sqrt{\underline{A}}}\right)-\varepsilon.$$

При $\delta=0$ это сводится к $P_e\geqslant {}^1/_2-\varepsilon$.

5.37. (а) Если передатчик использует различные коды для каждого из A начальных состояний канала, то каждое сообщение m будет сопоставлено с A различными кодовыми словами, допустим, $\mathbf{x}_{m,1}$, $\mathbf{x}_{m,2}$, ..., $\mathbf{x}_{m,A}$. Пусть $\mathbf{x}_m=(\mathbf{x}_{m,1}, \ldots, \mathbf{x}_{m,A})$ образует множество этих A кодовых слов; рассмотрим ансамбль кодов, в котором каждое $\mathbf{x}_m$ выбирается независимо с распределением вероятностей

$$Q(\mathbf{x}_m)=\prod_{i=1}^{A} Q_i(\mathbf{x}_m, i). \tag{1}$$

Другими словами, кодовые слова для различных начальных состояний выбираются независимо и для каждого состояния может быть использовано распределение, отличное от других. Предположим теперь, что A начальных состояний яв-

ляются равновероятными. Тогда вероятность приема последовательности $\mathbf{y}$ при условии, что на кодер поступило сообщение m, равна

$$P(\mathbf{y} \mid \mathbf{x}_m) = \sum_{i=1}^{A} \frac{1}{A} P_N(\mathbf{y} \mid \mathbf{x}_m, i, \quad s_0 = i). \tag{2}$$

С математической точки зрения $P(\mathbf{y}|\mathbf{x})$ определяет канал, а $Q(\mathbf{x}_m)$ — ансамбль независимых одинаково распределенных кодовых слов для этого канала. Поэтому теорема 5.6.1 может быть применена, что дает

$$\overline{P}_{e,\,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q(\mathbf{x})\, P(\mathbf{y} \mid \mathbf{x})^{1/(1+\rho)} \right]^{1+\rho}. \tag{3}$$

Используя (1) и (2), получаем

$$\overline{P}_{e,\,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} \prod_{i=1}^{A} Q_i(\mathbf{x}_i) \left[\sum_{j=1}^{A} \frac{1}{A} P_N(\mathbf{y} \mid \mathbf{x}_j, \; s_0 = j) \right]^{1/(1+\rho)} \right\}^{1+\rho}.$$

Используя неравенство задача 4.15(е), можно перенести знак суммирования по j влево, поставив его перед знаком суммы по $\mathbf{x}$ и затем произвести суммирование по $\mathbf{x}_i$ при всех $i \neq j$; в результате получим

$$\overline{P}_{e,\,m} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left\{ \sum_{j=1}^{A} \sum_{\mathbf{x}_j} Q_j(\mathbf{x}_j) \left[\frac{1}{A} P_N(\mathbf{y} \mid \mathbf{x}_j, \; s_0 = j) \right]^{1/(1+\rho)} \right\}^{1+\rho}$$

Умножая и деля сумму по j на A и рассматривая $1/A$ как распределение вероятности на j, *можно использовать результат задачи* 4.15(г), *чтобы получить*

$$\overline{P}_{e,\,m} \leqslant (M-1)^{\rho} A^{\rho-1} \sum_{j=1}^{A} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}_j} Q_j(\mathbf{x}_j)\, P_N(\mathbf{y} \mid \mathbf{x}_j, \; s_0 = j)^{1/(1+\rho)} \right\}^{1+\rho}.$$

Выбирая $Q_j(\mathbf{x}_j)$ так, чтобы минимизировать правую часть при каждом j, и умножая границу на 4, чтобы получить равномерную границу, для $P_{e,m}$ для некоторого кода будем иметь

$$P_{e,m} \leqslant 4(M-1)^{\rho} A^{\rho-1} \times$$

$$\times \sum_{j=1}^{A} \min_{Q_j(\mathbf{x}_j)} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}_j} Q_j(\mathbf{x}_j)\, P(\mathbf{y} \mid \mathbf{x}_j, \; s_0 = j)^{1/(1+\rho)} \right\}^{1+\rho}.$$

Можно, наконец, сумму по j ограничить сверху максимальным слагаемым, умноженным на A, и заметить, что для каждого начального состояния вероятность ошибки не может быть больше, чем умноженное на A среднее значение. Это приводит к выражению (5.9.5) с переставленными min и max.

(б) Если на приемнике известно состояние, то можно применить теорему 5.6.1 непосредственно для любого заданного начального состояния s_0; в результате получим

$$\overline{P_{e,\,m}(s_0)} \leqslant (M-1)^{\rho} \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} Q_N(\mathbf{x}) P_N(\mathbf{y} \mid \mathbf{x}, \; s_0)^{1/(1+\rho)} \right\}^{1+\rho}. \tag{4}$$

Для того чтобы вывести равномерную границу для $P_{e,\,m}(s_0)$, умножим это выражение на 4 и возьмем максимум по s_0, получим (5.9.5) без коэффициента $A^{1+\rho}$. Если передатчик знает также начальное состояние, то можно произвести минимизацию по $Q_N(\mathbf{x})$ раздельно для каждого s_0, изменяя порядок знаков min и max в (5.9.5).

(в) Доказательство того, что при независимых равновероятных входах достигается минимакс в (5.9.5), аналогично доказательству, приведенному в ре-

шении задачи 4.23; единственное отличие состоит в том, что выпуклость применяется здесь к (5.9.5), а не к средней взаимной информации. Для нахождения максмина заметим, что при заданном s_0 канал эквивалентен паре параллельных каналов без памяти и оптимизирующее входное распределение совпадает с тем, на котором достигается $\overline{C}$.

5.38. (а) $$I(\mathbf{X}^N; \mathbf{Y}^N | s_0) = H(\mathbf{Y}^N | s_0) - H(\mathbf{Y}^N | \mathbf{X}^N, s_0) = H(\mathbf{Y}^N | s_0) - H(\mathbf{Z}^N | s_0).$$

Но $$\lim_{N\to\infty} \frac{1}{N} H(\mathbf{Z}^N | s_0) = H_\infty(Z) \quad \text{и} \quad H(\mathbf{Y}^N | s_0) \leqslant N \ln 2$$

с равенством для равновероятных выходов, которые образуются при равновероятных входах. Поэтому $C = \ln 2 - H_\infty(Z)$.

(б) Заметим, что $P_N(\mathbf{y} | \mathbf{x}, s_0)$ является вероятностью того, что $\mathbf{z}$ принимает значение $\mathbf{y}-\mathbf{x}$ при заданном s_0, т. е. $P_{\mathbf{z}}(\mathbf{y}-\mathbf{x} | s_0)$. Следовательно,

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) = -\frac{1}{N} \ln \sum_{\mathbf{y}} \left\{ \sum_{\mathbf{x}} 2^{-N} P_{\mathbf{z}}(\mathbf{y}-\mathbf{x} | s_0)^{1/(1+\rho)} \right\}^{1+\rho}.$$

Полагая $\mathbf{z} = \mathbf{y} - \mathbf{x}$ и замечая, что для фиксированного $\mathbf{y}$ суммирование по всем $\mathbf{x}$ эквивалентно суммированию по всем $\mathbf{z}$, получаем

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) = -\frac{1}{N} \ln \sum_{\mathbf{y}} \left\{ 2^{-N} \sum_{\mathbf{z}} P_{\mathbf{z}}(\mathbf{z} | s_0)^{1/(1+\rho)} \right\}^{1+\rho} =$$

$$= -\frac{1}{N} \ln \left\{ 2^N \cdot 2^{-N(1+\rho)} \left[\sum_{\mathbf{z}} P_{\mathbf{z}}(\mathbf{z} | s_0)^{1/(1+\rho)} \right] \right\}^{1+\rho} =$$

$$= \rho \ln 2 - \frac{1+\rho}{N} \ln \sum_{\mathbf{z}} P_{\mathbf{z}}(\mathbf{z} | s_0)^{1/(1+\rho)}.$$

(в) Поскольку $\mathbf{z}$ является выходом марковского источника, то $\mathbf{z}$ и s_0 однозначно определяют последовательность состояний $s = (s_1, s_2, \ldots, s_N)$ (см. § 3.6). Следовательно,

$$P(\mathbf{s}, \mathbf{z} | s_0) = \begin{cases} P_{\mathbf{z}}(\mathbf{z} | s_0) & \text{для определенного } \mathbf{s}, \\ 0 & \text{для всех остальных } \mathbf{s}, \end{cases}$$

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) = \rho \ln 2 - \frac{1+\rho}{N} \ln \sum_{\mathbf{s},\mathbf{z}} P(\mathbf{s}, \mathbf{z} | s_0)^{1/(1+\rho)} =$$

$$= \rho \ln 2 - \frac{1+\rho}{N} \ln \sum_{\mathbf{s}} \prod_{n=1}^{N} \sum_{z_n} P(s_n, z_n | s_{n-1})^{1/(1+\rho)}.$$

Положив

$$\alpha(s_{n-1}, s_n) = \sum_{z_n} P(s_n, z_n | s_{n-1})^{1/(1+\rho)},$$

получим

$$E_{0,N}(\rho, \mathbf{Q}_N, s_0) = \rho \ln 2 - \frac{1+\rho}{N} \ln \sum_{\mathbf{s}} \prod_{n=1}^{N} \alpha(s_{n-1}, s_n).$$

Это выражение теперь имеет тот же самый вид, что и выражение в (5.9.37). Также $A \times A$-матрица $[\alpha(\rho)]$ с элементами $\alpha(i, j)$ является неприводимой, так как цепь Маркова по предположению является эргодической. Поэтому с помощью тех же рассуждений, что и при переходе от (5.9.37) к (5.9.44), получим

$$\lim_{N\to\infty} E_{0,N}(\rho, \mathbf{Q}_N, s_0) = \rho \ln 2 - (1+\rho) \ln \lambda(\rho),$$

где $\lambda(\rho)$ является наибольшим собственным значением $[\alpha(\rho)]$.

5.39. Так как s_n является детерминированной функцией y_n, то

$$P(\mathbf{y},\mathbf{s} \mid \mathbf{x},\ s_0) = \begin{cases} P_N(\mathbf{y} \mid \mathbf{x},\ s_0) & \text{для } \mathbf{s}, \text{ соответствующего } \mathbf{y}, \\ 0 & \text{во всех остальных случаях.} \end{cases}$$

Таким образом,

$$H(\mathbf{Y}^N \mid \mathbf{X}^N,\ s_0) = -\overline{\ln P(\mathbf{y},\mathbf{s} \mid \mathbf{x},\ s_0)},$$

где математическое ожидание берется по $\mathbf{x}$ и $\mathbf{y}$ при заданном s_0. Имеем

$$H(\mathbf{Y}^N \mid \mathbf{X}^N,\ s_0) = -\overline{\ln \prod_{n=1}^{N} P(y_n, s_n \mid x_n, s_{n-1})} =$$

$$= -\sum_{n=1}^{N} \overline{\ln P(y_n,\ s_n \mid x_n,\ s_{n-1})}\ ;\ -\overline{\ln P(y_n,\ s_n \mid x_n,\ s_{n-1})} =$$

$$= -\sum_{y_n,\ s_n,\ x_n,\ s_{n-1}} P(y_n,\ s_n,\ x_n,\ s_{n-1} \mid s_0) \ln P(y_n,\ s_n \mid x_n,\ s_{n-1}) =$$

$$= -\sum_{x_n,\ s_{n-1},\ y_n} Q_n(x_n)\, P(s_{n-1} \mid s_0)\, P(y_n \mid x_n,\ s_{n-1}) \times$$

$$\times \ln P(y_n \mid x_n,\ s_{n-1}) = H(Y_n \mid X_n, S_{n-1},\ s_0). \qquad (1)$$

Поскольку последовательность состояний не зависит от входа, то вероятность $P(s_{n-1} \mid s_0)$ не зависит от входного ансамбля и поэтому $H(Y_n \mid X_n,\ S_{n-1},\ s_0)$ зависит от входного ансамбля только через $Q_n(x_n)$. Также имеем

$$H(\mathbf{Y}^N \mid s_0) = \sum_{n=1}^{N} H(Y_n \mid Y_{n-1},\ Y_{n-2},\ \ldots,\ Y_1,\ s_0) =$$

$$= \sum_{n=1}^{N} H(Y_n \mid S_{n-1},\ Y_{n-1},\ \ldots,\ Y_1,\ s_0) \leqslant \sum_{n=1}^{N} H(Y_n \mid S_{n-1},\ s_0). \qquad (2)$$

Замечая, что y_n зависит от предыдущих выходов только через s_{n-1} и что имеет место статистическая зависимость между входными буквами, получаем, что (2) удовлетворяется с равенством при независимых входах. Объединяя (1) и (2), будем иметь

$$I(\mathbf{X}^N,\ \mathbf{Y}^N \mid s_0) \leqslant \sum_{n=1}^{N} I(X_n;\ Y_n \mid S_{n-1},\ s_0)$$

с равенством, когда входы независимы. Наконец, так как канал является неразложимым, то $I(X_n;\ Y_n \mid S_{n-1},\ s_0) \to I(X_n;\ Y_n \mid S_{n-1})$ при $n \to \infty$. Это означает, что

$$C = \max_{Q_n(x_n)} I(X_n;\ Y_n \mid S_{n-1}),$$

где распределение вероятностей для s_{n-1} является стационарным распределением вероятностей состояний $q(s_{n-1})$. Поэтому пропускная способность равна пропускной способности ДКБП с

$$P(y_n \mid x_n) = \sum_{s_{n-1}} P(y_n \mid x_n,\ s_{n-1})\, q(s_{n-1}).$$

6.1. (а)

$$x_1 = u_1, \quad x_3 = u_1 \oplus u_2, \quad x_5 = u_1 \oplus u_2.$$
$$x_2 = u_2, \quad x_4 = u_1,$$

Так как $x_1 = u_1$, $x_2 = u_2$ и все символы являются линейными комбинациями u_1 и u_2, то рассматриваемый код представляет собой систематический код с проверкой на четность.

(б)

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(в)

Синдром	Шумовая последовательность	Синдром	Шумовая последовательность
0 0 0	0 0 0 0 0	0 1 0	0 0 0 1 0
1 1 1	1 0 0 0 0	0 0 1	0 0 0 0 1
1 0 1	0 1 0 0 0	0 1 1	0 0 0 1 1 (или 1 0 1 0 0)
1 0 0	0 0 1 0 0	1 1 0	0 0 1 1 0 (или 1 0 0 0 1)

(г) Правильно декодируются пять конфигураций единичных ошибок и две конфигурации двойных ошибок; ни одна конфигурация тройных ошибок не декодируется правильно. Имеем

$$P_e = 1 - (1-\varepsilon)^5 - 5\varepsilon(1-\varepsilon)^4 - 2\varepsilon^2(1-\varepsilon)^3.$$

6.2. Утверждение этой задачи трудно интерпретировать. Однако при тщательном изучении задачи легко заключить, что при вычислении x_5 в левой части регистра сдвига находятся четыре информационных символа, так что

$$x_5 = u_1 \oplus u_2 \oplus u_4.$$

После того как вычислено x_5, все информационные символы сдвигаются на одну позицию вправо, а в левый разряд регистра подается нуль, так что $x_6 = u_1 \oplus \oplus u_2 \oplus u_3$. Аналогично $x_7 = u_2 \oplus u_3 \oplus u_4$.

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Синдром	Шумовая последовательность	Синдром	Шумовая последовательность
0 0 0	0 0 0 0 0 0 0	1 0 1	0 0 0 1 0 0 0
1 1 0	1 0 0 0 0 0 0	1 0 0	0 0 0 0 1 0 0
1 1 1	0 1 0 0 0 0 0	0 1 0	0 0 0 0 0 1 0
0 1 1	0 0 1 0 0 0 0	0 0 1	0 0 0 0 0 0 1

$$P_e = 1 - (1-\varepsilon)^7 - 7(1-\varepsilon)^6\varepsilon.$$

6.3. (а) Множество кодовых слов образует группу (по сложению по модулю 2), а множество кодовых слов с четным числом единиц образует подгруппу этой группы. Если рассматриваемая подгруппа не исчерпывает всей группы, то множество кодовых слов с нечетным числом единиц является смежным классом по этой подгруппе и, следовательно, имеет то же число элементов, что и сама подгруппа.

(б) При любом заданном n множество кодовых слов, для которых $x_{m,\,n} = 0$, образует подгруппу в группе кодовых слов. Если существуют какие-либо кодовые слова, у которых $x_{m,\,n} = 1$, то множество таких кодовых слов образует смежный класс по рассмотренной выше подгруппе с числом элементов, равным числу элементов этой подгруппы. Отметим, что для любого имеющего смысл кода $x_{m,\,n}$ не должно быть равно нулю при всех m, поскольку в противном случае n-й символ может быть просто исключен из всех кодовых слов, без изменения величины P_e.

(в) Из результатов пункта (б) следует, что общее число единиц в коде не превышает $MN/2$ (оно точно равно $MN/2$ для любого имеющего смысл кода с проверкой на четность). Поэтому среднее число единиц в слове не превышает $N/2$.

6.4. (а) Пусть $\mathbf{x}_1(n), \mathbf{x}_2(n), \ldots, \mathbf{x}_l(n)$ являются кодовыми словами, лежащими на расстоянии n от $\mathbf{x}_0$. Тогда ясно, что $\mathbf{x}_1(n) \oplus \mathbf{x}_1, \ldots, \mathbf{x}_l(n) \oplus \mathbf{x}_1$ различны, лежат на расстоянии n от $\mathbf{x}_1$ и являются кодовыми словами. Если какое-либо другое кодовое слово, например $\mathbf{x}'$, лежит на расстоянии n от $\mathbf{x}_1$, то $\mathbf{x}'' = \mathbf{x}' \oplus \oplus \mathbf{x}_1$ лежит на расстоянии n от $\mathbf{x}_0$. Однако это предположение противоречит исходному, поскольку $\mathbf{x}''$ должно принадлежать множеству $\mathbf{x}_1(n), \ldots, \mathbf{x}_l(n)$ и $\mathbf{x}'' \oplus \oplus \mathbf{x}_1 = \mathbf{x}'$, поэтому число слов на расстоянии n от $\mathbf{x}_0$ равно числу слов на расстоянии n от $\mathbf{x}_1$. Отметим, что этот результат справедлив также для патологического случая, когда существуют два или более кодовых слова, целиком состоящих из нулей. (Последнее может случиться лишь в несистематическом коде, у которого строки порождающей матрицы линейно зависимы. Естественно, что с практической точки зрения такие коды абсолютно бессмысленны.)

(б) Из пункта (а) следует, что если минимальное расстояние кода с проверкой на четность равно d_{min}, то существует кодовое слово, лежащее на расстоянии d_{min} от $\mathbf{x}_0$, и не существует ни одного кодового слова (отличного от $\mathbf{x}_0$), лежащего ближе к $\mathbf{x}_0$. Поэтому вес указанного выше слова равен d_{min} и не существует ни одного другого слова (отличного от $\mathbf{x}_0$), имеющего меньший вес.

(в) Чтобы преобразовать одно кодовое слово в другое, необходимо изменить в нем не менее d_{min} символов. Если при передаче произошло менее чем $d_{min}/2$ ошибок, то для того, чтобы перевести принятую последовательность в кодовое слово, отличное от переданного, необходимо еще изменить в ней более чем $d_{min}/2$ других символов. Поэтому декодер, декодирующий в ближайшее кодовое слово, во всех случаях, когда произошло менее чем $d_{min}/2$ ошибок, декодирует правильно.

(г) Если существует лишь одно кодовое слово, совпадающее с принятой последовательностью во всех нестертых позициях, то стирание может быть исправлено однозначно. Так как никакие два кодовых слова не могут совпадать более чем в $N - d_{min}$ позициях, то декодирование будет правильным во всех случаях, когда произошло менее чем d_{min} стираний.

6.5. Дополнительный проверочный символ изменяет проверочную матрицу H, превращая ее в матрицу H':

$$H' = \left[\begin{array}{c:c} H & \begin{matrix}1\\1\\\vdots\\1\end{matrix} \\ \hdashline 0\ 0\ \ldots & 1 \end{array}\right].$$

Для того чтобы $\mathbf{x}H' = 0$, необходимо, как это следует из рассмотрения последнего столбца H', чтобы $\sum_{n=1}^{N} x_n = 0$; отсюда вытекает, что любое кодовое слово имеет четное число единиц. В свою очередь, из последнего утверждения следует, что

вес ненулевого слова с минимальным весом является четным и что последняя проверка на четность увеличивает на единицу вес каждого обладающего нечетным весом слова первоначального кода.

6.6 (а) Число синдромных последовательностей (N, L)-кода с проверкой на четность равно 2^{N-L} (мы исключаем тривиальный случай, когда строки матрицы G линейно зависимы и когда даже при отсутствии ошибок декодирование может быть неправильным). Число различных шумовых последовательностей, состоящих из e или менее ошибок, равно $\sum_{j=0}^{e} \binom{N}{j}$; для того чтобы исправлять все конфигурации из e или меньшего числа ошибок, необходимо, чтобы все эти последовательности соответствовали различным синдромным последовательностям. Поэтому, для того чтобы исправляющая способность равнялась e, необходимо чтобы

$$\sum_{j=0}^{e} \binom{N}{j} \leqslant 2^{N-L}.$$

(б) Пусть для произвольного двоичного кода существует последовательность, которая лежит на расстоянии e или меньшем от каждого из двух кодовых слов, тогда при приеме этой последовательности, независимо от того, какое кодовое слово было декодировано, может произойти ошибка декодирования, когда число ошибок при передаче было равным или меньшим e, если передавалось другое кодовое слово. Поэтому вокруг каждого кодового слова должно быть множество $\sum_{j=0}^{e} \binom{N}{j}$ последовательностей, лежащих на расстоянии e или меньшем от этого слова, и эти множества, относящиеся к различным кодовым словам, не должны пересекаться. Так как существуют 2^N различных последовательностей и M кодовых слов, то

$$M \sum_{j=0}^{e} \binom{N}{j} \leqslant 2^N .$$

(в) Для кода с четным минимальным расстоянием d_{min} существует последовательность, лежащая на расстоянии $d_{min}/2$ от каждого из двух ближайших кодовых слов, и поэтому $e < d_{min}/2$. Для кода с нечетным минимальным расстоянием существует последовательность, лежащая на расстоянии $(d_{min} + 1)/2$ от одного кодового слова и на расстоянии $(d_{min} - 1)/2$ от другого кодового слова; поэтому снова $e < d_{min}/2$. Сопоставляя этот результат с результатом задачи 6.4. (в), получаем $e = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor$; поэтому из пункта (б) данной задачи следует, что для любого двоичного кода с M кодовыми словами, длиной блока N и минимальным расстоянием d_{min} справедливо

$$\sum_{j=0}^{\left\lfloor \frac{d_{min}-1}{2} \right\rfloor} \binom{N}{j} \leqslant \frac{2^N}{M}$$

6.7. (а) Согласно задаче 6.3. (в), общее число единиц во всех кодовых словах (N, L)-кода не превышает $(2^L N)/2$. Так как число ненулевых слов равно $2^L - 1$ (снова исключаем тривиальный случай, когда строки порождающей матрицы зависимы), то средний вес ненулевого слова не превышает $\frac{N}{2} \frac{2^L}{2^L - 1}$. Так как все ненулевые слова не могут одновременно иметь веса, превышающие средний вес, то $d_{min} \leqslant \frac{N}{2} \frac{2^L}{2^L - 1}$.

(б) По существу, к указанию, приведенному в задаче, нечего добавить. Заметим, что в произвольном двоичном коде с длиной блока N и 2^L кодовыми словами, первые $L - j$ символов кодовых слов могут принимать лишь 2^{L-j} различных значений; поэтому должно существовать по крайней мере 2^j кодовых слов, совпадающих в первых $(L - j)$ символах. Рассматривая оставшиеся $N - L + j$ символов этих 2^j кодовых слов как код, мы убедимся, что два из этих кодовых слов отличаются не более чем в $\frac{N-L+j}{2}\left(\frac{2^j}{2^j-1}\right)$ позициях. Так как эти два кодовых слова совпадают в первых $L - j$ символах, то имеем

$$d_{min} \leqslant \frac{N-L+j}{2}\left(\frac{2^j}{2^j-1}\right) \quad \text{при всех } j,\ 1 \leqslant j \leqslant L.$$

(в) Преобразуя приведенную выше границу, получаем нижнюю границу для числа проверочных символов $N - L$ при заданном d_{min}

$$N-L \geqslant 2d_{min}(1-2^{-j})-j, \quad 1 \leqslant j \leqslant L.$$

При $j = \lfloor \log_2 d_{min} \rfloor + 1$ она принимает вид

$$N-L \geqslant 2d_{min} - 2^{\{\log_2 d_{min} - \lfloor \log_2 d_{min} \rfloor\}} - [\lfloor \log_2 d_{min} \rfloor + 1].$$

Нетрудно увидеть, что $2d_{min} = 2^{[\log_2 d_{min} + 1]}$. Теперь заметим, что все слагаемые правой части, кроме среднего, являются целыми числами. Однако среднее слагаемое строго меньше 2, поэтому $N - L > 2d_{min} - 2 - \lfloor \log_2 d_{min} + 1 \rfloor$. Теперь обе части неравенства целочисленные, так что можно увеличить правую часть на 1 и заменить знак $>$ на $\geqslant$, $N - L \geqslant 2d_{min} - 2 - \lfloor \log_2 d_{min} \rfloor$.

(г) Разделив обе части приведенного выше неравенства на N и подставив $R = L/N$, получаем

$$1 - R \geqslant \frac{2d_{min}}{N} - \frac{\lfloor 2 + \log_2 d_{min} \rfloor}{N}.$$

Зафиксировав R, получим в пределе при $N \to \infty$, что последнее слагаемое исчезает; это приводит к неравенству

$$R \leqslant 1 - \frac{2d_{min}}{N}.$$

6.8. (а) Способ построения кода гарантирует нам, что все множества $d - 1$ строк в матрице являются множествами линейно независимых строк. Если обозначить через $\mathbf{h}_n$ n-ю строку матрицы H, то для того чтобы $\mathbf{x}=(x_1, \ldots, x_N)$ было кодовым словом, необходимо, чтобы $\mathbf{x}H = \sum_{n=1}^{N} x_n \mathbf{h}_n = 0$. Отсюда видно, что для того чтобы $\mathbf{x}$ было кодовым словом, необходимо, чтобы строки H, соответствующие тем позициям, в которых $\mathbf{x}$ имеет единицы, были бы линейно зависимыми. Поэтому не существует ненулевого кодового слова, имеющего вес, меньший или равный $d - 1$, и поэтому $d_{min} \geqslant d$.

(б) Существенным является здесь вопрос об интерпретации, так как, например, можно рассматривать $\mathbf{h}_1 \oplus \mathbf{h}_2$ и как линейную комбинацию первых двух строк, и как линейную комбинацию какого-либо множества строк, в которое входят первые две строки. В интересующей нас задаче нужно при подсчете учесть $\mathbf{h}_1 \oplus \mathbf{h}_2$ (и каждую другую комбинацию) только один раз. Для этого найдем число подмножеств из $d - 2$ или меньшего числа строк, которые могут быть составлены из N строк; оно равно

$$\sum_{i=0}^{d-2} \binom{N}{i}.$$

Если это число линейных комбинаций меньше чем 2^r, т. е. меньше общего числа возможных строк, то можно выбрать другую строку. Поэтому при окончании процесса выбора строк имеем

$$\sum_{i=0}^{d-2} \binom{N}{i} \geqslant 2^r = 2^{N-L}.$$

6.9. (а)

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(б)

Синдром	Шумовая последовательность	Синдром	Шумовая последовательность
0 0 0 0	0 0 0 0 0 0 0	0 1 1 0	1 1 0 0 0 0 0
1 1 0 1	1 0 0 0 0 0 0	1 0 1 0	1 0 1 0 0 0 0
1 0 1 1	0 1 0 0 0 0 0	0 1 0 1	1 0 0 1 0 0 0
0 1 1 1	0 0 1 0 0 0 0	1 0 0 1	1 0 0 0 1 0 0
1 0 0 0	0 0 0 1 0 0 0	1 1 1 1	1 0 0 0 0 1 0
0 1 0 0	0 0 0 0 1 0 0	1 1 0 0	1 0 0 0 0 0 1
0 0 1 0	0 0 0 0 0 1 0	0 0 1 1	0 1 0 1 0 0 0
0 0 0 1	0 0 0 0 0 0 1	1 1 1 0	1 1 0 1 0 0 0

(в) Для кода I $P_e = 1 - (1-\varepsilon)^7 - 7\varepsilon(1-\varepsilon)^6 - 7\varepsilon^2(1-\varepsilon)^5 - \varepsilon^3(1-\varepsilon)^4$.

Для кода II $P_e = 1 - (1-\varepsilon)^7 - 7\varepsilon(1-\varepsilon)^6 - 8\varepsilon^2(1-\varepsilon)^5$.

$$P_e(\text{код I}) - P_e(\text{код II}) = \varepsilon^2(1-\varepsilon)^5 - \varepsilon^3(1-\varepsilon)^4 = \varepsilon^2(1-\varepsilon)^5\left[1 - \frac{\varepsilon}{1-\varepsilon}\right] > 0.$$

Поэтому вероятность ошибки для кода I больше.

(г) Для кода I $d_{min} = 4$, для кода II $d_{min} = 3$.

(д) Рассмотренные два кода дают требуемый контрпример.

6.10. Пусть $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}L$ — строки порождающей матрицы. Так как по предположению они линейно зависимы, то существует двоичная ненулевая последовательность $u_1, u_2, \ldots, uL$, такая, что

$$\sum_{i=1}^{L} u_i \mathbf{g}_i = \mathbf{0}.$$

Таким образом, ненулевая информационная последовательность $\mathbf{u} = (u_1, \ldots, uL)$ отображается в кодовое слово, целиком состоящее из нулей. Если $\mathbf{u}'$ — произвольная информационная последовательность, то информационная последовательность $\mathbf{u}' \oplus \mathbf{u}$ отображается в то же самое кодовое слово, что и $\mathbf{u}'$.

6.11. (а) Пусть $\mathbf{h}_1, \ldots, \mathbf{h}_{N-L}$ — столбцы матрицы H_2. Тогда из соотношения

$\mathbf{x}H_2=0$ вытекает равенство $\mathbf{x}\mathbf{h}_i=0$ при $1 \leqslant i \leqslant N-L$ $\left(\mathbf{x}\mathbf{h}_i=0\right.$ означает, что $\left.\sum_{n=1}^{N} \mathbf{x}_n h_{n,\,i}=0\right)$. Пусть $\mathbf{h}'_1, \ldots, \mathbf{h}'_{N-L}$ — столбцы матрицы H_1. Как известно, при любом i имеем $\mathbf{h}'_i = \sum_{l=1}^{N-L} \alpha_{i,\,l}\,\mathbf{h}_l$ при некотором наборе коэффициентов $\alpha_{i,l}$, поэтому

$$\mathbf{x}\mathbf{h}'_i = \sum_{l=1}^{N-L} \alpha_{i,\,l}\,(\mathbf{x}\cdot\mathbf{h}_l)=0;$$

$$1 \leqslant i \leqslant N-L.$$

Это доказывает, что $\mathbf{x}\,H_2 = 0 \Rightarrow \mathbf{x}H_1 = 0$; аналогичные рассуждения, начинающиеся с рассмотрения H_1, доказывают, что $\mathbf{x}H_1 = 0 \Rightarrow \mathbf{x}H_2 = 0$.

(б) Соотношение $\mathbf{e}_1H_2 \neq \mathbf{e}_2H_2$ выполняется тогда и только тогда, когда $(\mathbf{e}_1 \oplus \mathbf{e}_2)H_2 \neq 0$. Согласно пункту (а), это справедливо тогда и только тогда когда $(\mathbf{e}_1 \oplus \mathbf{e}_2)H_1 \neq 0$, что, в свою очередь, выполняется тогда и только тогда, когда $\mathbf{e}_1H_1 \neq \mathbf{e}_2H_1$. Множество шумовых последовательностей $\{\mathbf{e}_1, \mathbf{e}_2, \ldots.\}$ в таблице декодирования, составленной по H_2, обладает тем свойством, что $\mathbf{e}_i\,H_2 \neq \neq \mathbf{e}_jH_2$ при $i \neq j$. Как показано выше, эти шумовые последовательности также входят в таблицу декодирования, составленную по H_1.

(в) В линейной алгебре известен результат, утверждающий, что если мощность максимального множества линейно независимых столбцов H равен r', то подпространство векторов $\mathbf{x}$, для которых $\mathbf{x}H=0$, имеет размерность $N-r'$. Выберем r' независимых столбцов матрицы H и приведем эту матрицу к систематическому виду путем элементарных операций над столбцами. Получится проверочная матрица, соответствующая коду с $2^{N-r'}$ кодовыми словами и таблицей декодирования с $2^{r'}$ строками. Из пунктов (а) и (б) следует, что код, соответствующий H_1, состоит из тех же кодовых слов, а его таблица декодирования включает те же шумовые последовательности.

6.12. Последовательность, целиком состоящая из нулей, входит в оба кода. Более того, если $\mathbf{x}_1$ и $\mathbf{x}_2$ входят в оба кода, то и $\mathbf{x}_1 \oplus \mathbf{x}_2$ входит в оба кода. Наконец, $\mathbf{x}$ является обратным по сложению элементом для самого себя. Таким образом, пересечение кодовых слов этих двух кодов образует подгруппу каждого кода. Если H_1 и H_2 — проверочные матрицы для первоначальных кодов, то $\mathbf{x}$ является кодовым словом кода—«пересечения» тогда и только тогда, когда одновременно $\mathbf{x}H_1 = 0$ и $\mathbf{x}H_2 = 0$. Это выполняется тогда и только тогда, когда $\mathbf{x}H = 0$, где

$$H=[H_1 \mid H_2].$$

Обычно удобно исключить из H зависимые столбцы и привести H к систематическому виду.

6.13. Пусть a и b — произвольные элементы группы, а e — нейтральный элемент. Так как обратным элементом для ab является элемент ab, то имеем $ab\cdot a\cdot b = e$. Умножение слева на a и справа на b дает $a\cdot b = b\cdot a$; таким образом, группа является абелевой. Теперь заметим, что для любого элемента группы $a \neq e$ пара элементов e и a образует подгруппу, т. е. $a\cdot a=e$, $a\cdot e=a$, $e\cdot a = a$, $e\cdot e=e$. Теперь возьмем любой элемент (отличный от a и e) и заметим, что b и $b\cdot a$ образуют смежный класс по подгруппе $\{e, a\}$. Аналогично $\{e, a, b, b\cdot a\}$ образует новую подгруппу. Продолжим образование новых подгрупп и смежных классов указанным способом. Произведение любых двух элементов смежного класса принадлежит подгруппе, а произведение любого элемента подгруппы на элемент смежного класса является элементом смежного класса. Наконец, так как каждый элемент является обратным для самого себя, то совокупность подгруппы и смежного класса замкнута по групповой операции и по операции образования обратного элемента; поэтому она образует подгруппу с удвоенным числом элементов. Отсюда следует, что число элементов в каждой новой подгруппе является сте-

пенью 2 и в момент, когда вновь образованная под-руппа совпадает с полной группой, число элементов в ней будет равно 2^N, где N—некоторое целое число. Теперь предположим, что элемент **e** соответствует последовательности из N нулей, элемент **a** соответствует $\{1, 0, 0, \ldots, 0\}$, элемент **b** соответствует $\{0,1,0, \ldots, 0\}$ и пусть каждый последующий новый элемент выбирается таким образом, чтобы создавать новый смежный класс, соответствующий другой двоичной последовательности веса 1. Если в качестве групповой операции ввести сложение последовательностей по модулю 2, то станет ясно, что мы придем к изоморфизму. Этот изоморфизм не единствен и так как в качестве **a** может быть выбран любой из $2^N - 1$ элементов, а в качестве **b** — любой из $2^N - 2$ элементов и т. д., то получим

$$\prod_{i=0}^{N-1} (2^N - 2^i)$$

различных способов установления этого изоморфизма.

6.14.

Элемент	1	a	a^2	a^3	a^4	a^5
Порядок	1	6	3	2	3	6

Подгруппами являются

$$\{1, a^2, a^4\}, \quad \{1, a^3\}, \quad \{1\}, \quad \{1, a, a^2, a^3, a^4, a^5\}.$$

6.15. (а)

+	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

·	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) Пусть α — элемент $GF(p)$. Число $p - 1$ делится на мультипликативный порядок α и, следовательно, $\alpha^{p-1} = 1$. Теперь заметим, что при любых целых a и b имеем $R_p(ab) = R_p(a)\,R_p(b)$. Если воспользоваться символом * для обозначения умножения в $GF(p)$, то последнее эквивалентно соотношению $R_p(ab) = R_p(a) * R_p(b)$. По индукции отсюда следует, что

$$R_p(a^{p-1}) = \underbrace{R_p(a) * R_p(a) * \ldots * R_p(a)}_{p-1 \text{ раз}}$$

Поскольку это $(p - 1)$-я степень некоторого элемента $GF(p)$, то произведение равно 1.

6.16. Простейший путь разложения многочлена над $GF(2)$ состоит в переборе всех возможных делителей, имеющих степень вплоть до половины степени многочлена. Производя сначала деление на $D + 1$, получаем

$$D^4 + D^2 + D + 1 = (D^3 + D^2 + 1)(D + 1). \qquad (1)$$

Далее убедимся, что ни $(D + 1)$, ни D не являются делителями $D^3 + D^2 + 1$. Так как многочлен третьей степени не может разлагаться на два или большее число множителей второй или большей степени и так как мы убедились в непригодности всех множителей первой степени, то многочлен $D^3 + D^2 + 1$ неприводим и (1) задает искомое разложение.

6.17. Многочлен $D^2 + 1$ над $GF(3)$ не делится ни на D, ни на $D + 1$, ни на $D + 2$, поэтому $D^2 + 1$ неприводим над $GF(3)$. Заметим, однако, что $D^2 + 1$ — приводимый многочлен над $GF(2)$ (а также над полем комплексных чисел). Поэтому в задачах разложения многочленов необходимо четко определить поле, над которым задаются многочлены.

+	0	1	2	D	$D+1$	$D+2$	$2D$	$2D+1$	$2D+2$
0	0	1	2	D	$D+1$	$D+2$	$2D$	$2D+1$	$2D+2$
1	1	2	0	$D+1$	$D+2$	D	$2D+1$	$2D+2$	$2D$
2	2	0	1	$D+2$	D	$D+1$	$2D+2$	$2D$	$2D+1$
D	D	$D+1$	$D+2$	$2D$	$2D+1$	$2D+2$	0	1	2
$D+1$	$D+1$	$D+2$	D	$2D+1$	$2D+2$	$2D$	1	2	0
$D+2$	$D+2$	D	$D+1$	$2D+2$	$2D$	$2D+1$	2	0	1
$2D$	$2D$	$2D+1$	$2D+2$	0	1	2	D	$D+1$	$D+2$
$2D+1$	$2D+1$	$2D+2$	$2D$	1	2	0	$D+1$	$D+2$	D
$2D+2$	$2D+2$	$2D$	$2D+1$	2	0	1	$D+2$	D	$D+1$

·	0	1	2	D	$D+1$	$D+2$	$2D$	$2D+1$	$2D+2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	D	$D+1$	$D+2$	$2D$	$2D+1$	$2D+2$
2	0	2	1	$2D$	$2D+2$	$2D+1$	D	$D+2$	$D+1$
D	0	D	$2D$	2	$D+2$	$2D+2$	1	$D+1$	$2D+1$
$D+1$	0	$D+1$	$2D+2$	$D+2$	$2D$	1	$2D+1$	2	D
$D+2$	0	$D+2$	$2D+1$	$2D+2$	1	D	$D+1$	$2D$	2
$2D$	0	$2D$	D	1	$2D+1$	$D+1$	2	$2D+2$	$D+2$
$2D+1$	0	$2D+1$	$D+2$	$D+1$	2	$2D$	$2D+2$	D	1
$2D+2$	0	$2D+2$	$D+1$	$2D+1$	D	2	$D+2$	1	$2D$

6.18. (а)

$$G=\begin{bmatrix}1&0&1&1\\0&1&1&2\end{bmatrix},\quad H=\begin{bmatrix}1&1\\1&2\\2&0\\0&2\end{bmatrix}.$$

(Заметим, что $-1=2$ в $GF(3)$, так что указанная H соответствует рис. 6.5, 1).

(б)

Синдром	Шумовая последовательность	Синдром	Шумовая последовательность
0 0	0 0 0 0	2 2	2 0 0 0
1 1	1 0 0 0	2 1	0 2 0 0
1 2	0 1 0 0	1 0	0 0 2 0
2 0	0 0 1 0	0 1	0 0 0 2
0 2	0 0 0 1		

(в + г) Для шумовой последовательности с одной ошибкой, например в n-й позиции, имеем $z_n=k$, где k — некоторый ненулевой элемент из $GF(q)$. Тогда $S=\mathbf{z}H$ равно умноженной на k n-й строке H. Таким образом, чтобы построить код Хэмминга, будем выбирать строки матрицы H по одной и при каждом выборе новой строки будем исключать из дальнейшего рассмотрения эту строку и все произведения этой строки на элемент поля. Когда число проверочных символов равно m, число возможных строк для матрицы H, исключая нулевую строку, равно q^m-1, поэтому число строк, не являющихся произведением некоторой строки на элемент поля, равно

$$\frac{q^m-1}{q-1}=q^{m-1}+q^{m-2}+\ \dots\ +q+1. \tag{1}$$

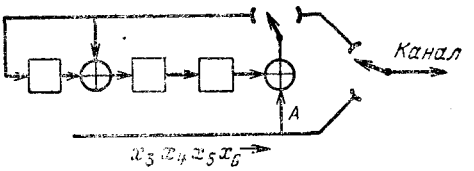
Таким образом, (1) задает зависимость N от m для кодов Хэмминга. Заметим, что при $q = 3$ и $m = 2$ число N, определяемое формулой (1), равно 4, что соответствует пунктам (а) и (б).

6.19. Отметим, что *если* данный код является циклическим, то порождающий многочлен должен соответствовать последней строке порождающей матрицы, т. е. $g(D) = D^3 + D^2 + D + 1$. Разделив $D^8 + 1$ на $g(D)$, получим

$$D^8 + 1 = g(D)[D^5 + D^4 + D + 1].$$

Иными словами, $g(D)$ порождает циклический код с длиной блока 8 и с 5 информационными символами; проверочным многочленом является $h(D) = D^5 + D^4 + D + 1$. Чтобы доказать, что данный код имеет те же самые кодовые слова, что и этот циклический код, нужно показать, что каждой строке данной матрицы соответствует некоторый многочлен, умноженный на $g(D)$. Для этого непосредственной проверкой убедимся, что четвертой строке матрицы соответствует $(D + 1)\,g(D)$, третьей строке $(D^2 + D)\,g(D)$, второй строке $(D^3 + D^2)\,g(D)$ и первой строке $(D^4 + D^3 + 1)\,g(D)$.

6.20. Разделив многочлен $D^7 + 1$ на $g(D) = D^3 + D + 1$ над $GF(2)$, получим, что $h(D) = D^4 + D^2 + D + 1$. Рис. 6.5.5 и 6.5.4 относятся к $GF(2)$ и данные многочлены $g(D)$ и $h(D)$ задают 3- и 4-разрядные реализации регистра сдвига. Например, рис. 6.5. 5 при поступлении данных в точку A принимает вид:



При заполнении регистра информационными символами $x_6, \ldots, x_3$ правый ключ находится в нижней позиции, а верхний ключ — в левой позиции; затем они меняют свое положение. Чтобы убедиться, что получаемый код является кодом Хемминга, можно переписать проверочную матрицу, придав ей вид, указанный на рис. 6.5.3, и заметить, что все строки различны и пробегают все возможные ненулевые значения троек чисел.

$$H = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

6.21. *Доказательство* 1. Если $g_0 = 0$, то многочлен $g_m D^{m-1} + g_{m-1} D^{m-2} + \ldots + g_1$ является циклическим сдвигом $g(D)$ и потому является кодовым словом, степень которого меньше степени $g(D)$. Это противоречит тому, что $g(D)$ является по определению нормированным кодовым словом, имеющим в данном коде минимальную степень.

Доказательство 2. Если $g_0 = 0$, то $g(D) = D[g_m D^{m-1} + \ldots + g_1]$. Однако D не может быть делителем многочлена $D^N - 1$ ни при каком N и, следовательно, $g(D)$ также не может быть делителем. Поэтому $g(D)$ не может быть порождающим многочленом.

6.22. После того, как первый информационный символ выйдет из регистра в линию, в крайнем левом разряде регистра будет храниться $-h_0 x_{N-1}$. В следующий момент времени во втором слева разряде будет храниться $-h_0 x_{N-1}$ —

$-h_1 x_{N-2}$ и аналогично после того, как x_{N-L} поступит в линию, в крайнем правом разряде будет храниться $\sum_{i=0}^{L-1} -h_i x_{N-i-1}$ и

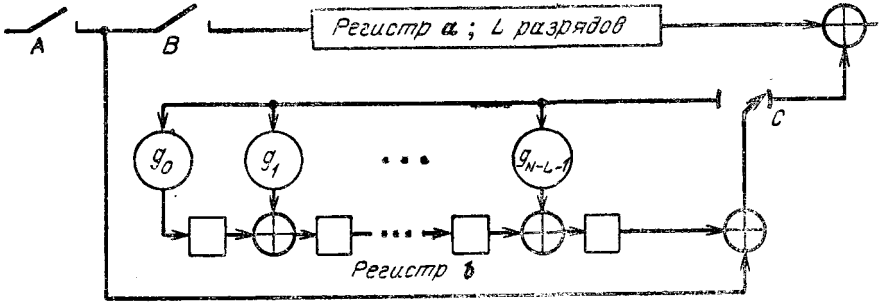
$$x_{N-L-1} = -\sum_{i=0}^{L-1} h_i x_{N-i-1}.$$

Анализируя каждый последующий символ тем же способом, получаем

$$x_{N-L-j} = -\sum_{i=0}^{L-1} h_i x_{N-i-j}.$$

Это означает, что коэффициенты $x(D)$ $h(D)$ при $N-1 \geqslant n \geqslant L$ равны нулю и, следовательно, $x(D)$ является кодовым словом.

6.23. (а + б) Из свойства цикличности следует, что любая циклическая серия последовательных L символов кодового слова может рассматриваться как последовательность информационных символов, а остальные $N-L$ символов могут быть вычислены по ним. Поэтому любая серия не более чем $N-L$ последовательных стираний может быть исправлена, если рассматривать циклически



последующие L символов как информационные. Если произошло $N-L+a$, $a>0$, последовательных стираний, то первые $N-L$ из них могут рассматриваться как проверочные символы и могут быть выбраны так, чтобы образовать кодовое слово при каждом выборе остальных стираний. Поэтому с принятой последовательностью согласуется 2^a кодовых слов и декодирование неоднозначно.

(в) Представим себе, что во время работы изображенное на рисунке устройство производит $N-L$ сдвигов. Принятое слово, у которого стирания заменены нулями, поступает в декодер в течение первых N единиц времени, когда ключ A замкнут. Ключ B замкнут в течение первых L единиц времени и регистр **a** наполняется символами принятой информационной последовательности. В течение последующих $N-L$ единиц времени, когда принимаются проверочные символы, в регистре **a** не происходит сдвигов, затем производится считывание L информационных символов. Регистр **b** сдвигается в каждый момент из общего числа $N+L$ единиц времени. Нетрудно убедиться, что когда первый стертый символ выходит из регистра **a**, в регистре **b** будет содержаться стертая последовательность. [Аналитически это означает, что после $N+i$ сдвигов в регистре **b** будет храниться $R_{g(D)}[D^{N-L+i} z(D)]$, где $z(D) = y(D) - x(D)$ означает стертую часть $x(D)$. Если стерт отрезок от $(N-i-1)$-го символа до $(N-j-1)$-го символа, $i \geqslant j > i-(N-L)$, то $R_{g(D)}[D^{N-L+i} z(D)] = R_{g(D)}[D^{-L+i} z(D)]$, что совпадает с последовательностью стираний.] В этот момент ключ C находится в правом положении и производится исправление. Определение этого момента должно производиться счетчиком, который начинает подсчет в момент, когда первое стирание поступает в декодер и замыкает ключ при счете N.

Единственная нерассмотренная здесь проблема состоит в том, что делать с последовательными блоками принятых символов. Наиболее простой путь ее решения состоит в замене каждого разряда регистра b на два разряда, проведении каждую единицу времени двух сдвигов, и изменении фаз при изменении блоков.

6.24. (а) Задача совпадает с задачей 6.4; объем алфавита не влияет на рассуждения.

(б) I. Линейный (N, L)-код над $GF(q)$ имеет q^{N-L} различных синдромных последовательностей. Число различных расположений i ошибочных символов равно $\binom{N}{i}$ и каждый ошибочный символ может принимать любое из $q-1$ отличных от нуля значений элементов в $GF(q)$. Поэтому число различных конфигураций i ошибок равно $\binom{N}{i}(q-1)^i$. Так как каждой исправимой конфигурации ошибок соответствует один синдром, то для исправления всех конфигураций e ошибок должно выполняться

$$\sum_{i=0}^{e}(q-1)^i\binom{N}{i}\leqslant q^{N-L}.$$

Подставляя вместо e величину $\lfloor (d_{min}-1)/2 \rfloor$, получаем требуемую границу.

II. Рассуждая, как и в задаче 6.3(в), нетрудно убедиться, что самое большее $(q-1)/q$ доля кодовых слов содержит в любой данной позиции ненулевые символы. Так как число кодовых слов равно q^L и в блоке имеется N символов, то общее число ненулевых символов во всех кодовых словах не превышает $Nq^L(q-1)/q$. Поскольку число ненулевых кодовых слов равно q^L-1, то вес по меньшей мере одного из них меньше среднего веса, или

$$d_{min}\leqslant\frac{N(q-1)}{q}\,\frac{q^L}{q^L-1}.$$

Рассматривая совокупность q^j кодовых слов, у которых первые $L-j$ символов нулевые, с помощью тех же рассуждений получаем

$$d_{min}\leqslant\frac{(N-L+j)(q-1)}{q}\,\frac{q^j}{q^j-1}\quad\text{при любом } j,\ 1\leqslant j\leqslant L.$$

Преобразуя это выражение, получаем

$$N-L\geqslant\frac{qd_{min}}{q-1}(1-q^{-j})-j.$$

Положив $j=\lfloor \log_q(qd_{min})\rfloor$, получим

$$N-L\geqslant\frac{qd_{min}}{q-1}-\frac{q^x}{q-1}+x-\log_q(qd_{min}),\qquad(1)$$

где

$$x=\log_q(qd_{min})-\lfloor\log_q(qd_{min})\rfloor.\qquad(2)$$

Отсюда видно, что x принадлежит интервалу $(0,1)$. В этом интервале функция $-\frac{q^x}{q-1}+x$ выпукла $\cap$ по x и принимает при $x=0$ и при $x=1$ значение $-1/(q-1)$; поэтому она ограничена снизу величиной $-1/(q-1)$. Используя эту оценку, из (1) будем иметь

$$N-L\geqslant\frac{qd_{min}-1}{q-1}-\log_q(qd_{min}).$$

III. Воспользуемся процедурой построения, отличающейся от процедуры построения в задаче 6.8. лишь тем, что элементы строк являются теперь элементами $GF(q)$. Общее число линейных комбинаций из $d-2$ или меньшего числа строк, выбираемых из N строк, равно

$$\sum_{i=0}^{d-2} (q-1)^i \binom{N}{i}.$$

В момент окончания процедуры построения эта сумма должна превысить q^{N-L}, и для построенного таким образом кода $d_{min} \geqslant d$, давая требуемую границу.

6.25. Нужно показать, что

$$D^3 - 1 = (D-1)(D-t)[D-(t+1)],$$

где обе части равенства являются многочленами по D над полем $GF(4)$ из элементов $\{0, 1, t, t+1\}$ со сложением и умножением, определяемыми на рис. 6.4.1. Производя умножение в правой части, получаем

$$D^3 + D^2[-1-t-(t+1)] + D[t+(t+1)+t(t+1)] - t(t+1).$$

Выполнив операции в данном поле, получим, например, что $t(t+1)=1$, и приведенное выше выражение сводится к D^3-1, что и требовалось установить.

6.26 (а) Так как элемент α примитивный, то $\alpha^n = 1$ тогда и только тогда, когда n кратно $q-1$, и следовательно, $\alpha^{in} = 1$ тогда и только тогда, когда in кратно $q-1$. Порядок элемента α^i равен наименьшему $n > 0$, при котором $\alpha^{in} = 1$ или, другими словами, наименьшему $n > 0$, такому, что in кратно $q-1$. Определим j равенством $q-1 = [\text{НОД}\,(q-1, i)]\,j$.

Заметим, что j и i взаимно просты, так что n (порядок α^i) должен быть кратен j и, следовательно, $n \geqslant j$. Вместе с тем ij кратно $q-1$, так что $n = j = (q-1)/\text{НОД}\,(q-1, i)$. Из того, что $i(q-1) = [\text{НОД}\,(i, q-1)]\,[\text{НОК}\,(i, q-1)]$, следует, что порядок α^i также равен $[\text{НОК}\,(i, q-1)]/i$.

(б) Если $q-1$ делится на n, то порядки следующих n элементов $\alpha^{(q-1)/n}$, $\alpha^{2(q-1)/n}$, ..., $\alpha^{n(q-1)/n} = 1$ являются делителями n и эти элементы, как нетрудно понять, образуют циклическую группу по умножению. Вместе с тем эти элементы являются единственными элементами, порядки которых являются делителями n, поскольку из пункта (а) следует, что если порядок α^i делит n, то $\text{НОД}\,(q-1, i) = m(q-1)/n$ при некотором m, которое является делителем n. Это означает, однако, что i кратно $m(q-1)/n$; следовательно, α^i является одним из элементов, рассмотренных выше.

6.27. (а) Пусть $f(D)$ — минимальный многочлен элемента α над $GF(p)$. Рассмотрим множество элементов

$$\beta = i_{m-1}\alpha^{m-1} + i_{m-2}\alpha^{m-2} + \ldots + i_0\alpha^0, \qquad (1)$$

где $i_0, i_1, \ldots, i_{m-1}$ — произвольные целые элементы поля. Число таких элементов равно p^m и они все различны [аналогично (6.6.7)]. Вместе с тем, так как $\alpha^m = \sum_{i=0}^{m-1} f_i\alpha^i$, то рассмотренное выше множество замкнуто по умножению и сложению.

Кроме того, $(p-1)\beta = -\beta$ и $\beta^{p^m-2}\beta = 1$, так что $\beta^{-1} = \beta^{p^m-2}$; отсюда следует, что обратные элементы по сложению и умножению принадлежат множеству. Таким образом, рассмотренное выше множество образует подполе из p^m элементов. Оно является минимальным подполем, содержащим α, так как включает лишь α, целые элементы поля и комбинации этих элементов.

(б) Множество p^m-1 ненулевых элементов рассмотренного выше подполя образует подгруппу по умножению p^n-1 ненулевых элементов поля. Поэтому $(p^n-1)/(p^m-1) = p^{n-m} + p^{n-2m} + \ldots$ должен быть целым элементом, откуда следует, что n делится на m.

(в) Так как многочлен $D^{p^m-2}-1$ имеет лишь p^m-1 корней, то в $GF(p^n)$ имеется лишь p^m-1 элементов, мультипликативные порядки которых явля-

ются делителями $p^m - 1$. Так как все эти элементы принадлежат подполю, которое было найдено, то в $GF(p^n)$ не может содержаться никакое другое подполе из p^m элементов.

(г) Из пункта (а) следует, что элемент β принадлежит подполю p^i элементов и потому его мультипликативный порядок должен быть делителем $p^i - 1$. В действительности степень i многочлена $f_\beta(D)$ является минимальным целым числом, таким, что n делится на i и порядок элемента β является делителем $p^i - 1$.

6.28. Заметим, что 0 и 1 принадлежат подполю $GF(2^2)$. Мультипликативный порядок остальных двух элементов равен 3; используя результаты задачи 6.26, убеждаемся, что α^5 и α^{10} являются искомыми элементами. Из рассмотрения рис. 6.6.3 следует, что $\alpha^5 = t^2 + t$ и $\alpha^{10} = t^2 + t + 1$.

6.29. (а) Нужно найти минимальный многочлен $f_\beta(D)$ элемента $\beta = t^3 + 1$ (здесь мы использовали обозначение β, а не α, чтобы можно было воспользоваться рис. 6.6.3, где $\beta = \alpha^{14}$). Тогда получим

$$f_\beta(D) = (D-\beta)(D-\beta^2)(D-\beta^4)(D-\beta^8) = (D-\alpha^{14})(D-\alpha^{13})(D-\alpha^{11})(D-\alpha^7) =$$
$$= [D^2-(\alpha^{14}+\alpha^{13})D+\alpha^{12}][D-\alpha^{11}][D-\alpha^7] = [D^2-\alpha^2 D+$$
$$+\alpha^{12}][D-\alpha^{11}][D-\alpha^7] = [D^3-(\alpha^{11}+\alpha^2)D^2+(\alpha^{12}+\alpha^{13})D-\alpha^8][D-\alpha^7] =$$
$$= (D^3-\alpha^9 D^2+\alpha D-\alpha^8)(D-\alpha^7) = D^4-(\alpha^9+\alpha^7)D^3+$$
$$+(\alpha+\alpha)D^2-(\alpha^8+\alpha^8)D+\alpha^{15} = D^4+D^3+1.$$

(б)

$$f_\beta(\beta) = f_4\beta^4 + f_3\beta^3 + f_2\beta^2 + f_1\beta + 1 = 0. \qquad (1)$$

Используя рис. 6.6.3, получаем $\beta = t^3 + 1$, $\beta^2 = t^3 + t^2 + 1$, $\beta^3 = t^3 + t^2 + t + 1$, $\beta^4 = t^3 + t^2 + t$. Переписывая (1) в виде отдельных равенств для каждой степени t, получаем

$$f_4 + f_3 + f_2 + f_1 = 0 \text{ (члены при } t^3),$$
$$f_4 + f_3 + f_2 = 0 \text{ (члены при } t^2),$$
$$f_4 + f_3 = 0 \text{ (члены при } t),$$
$$f_3 + f_2 + f_1 = 1.$$

Решая эти уравнения [в $GF(2)$], получаем $f_2 = 0$, $f_1 = 0$, $f_3 = 1$, $f_4 = 1$ или $f_\beta(D) = D^4 + D^3 + 1$.

6.30. (а) Пусть α — примитивный элемент на рис. 6.6.3 и пусть $r = 1$. Тогда для кода, исправляющего 2 ошибки, получим
$g(D) = f_1(D) f_3(D) = (D^4 + D + 1)(D^4 + D^3 + D^2 + D + 1) = D^8 + D^7 + D^6 + D^4 + 1$.
Для кода, исправляющего 3 ошибки, приведенный выше многочлен $g(D)$ должен быть умножен на $f_5(D) = D^2 + D + 1$ (см. рис. 6.6.3). Имеем

$$g(D) = D^{10} + D^8 + D^5 + D^4 + D^2 + D + 1.$$

(б) Пусть α — примитивный элемент для первого кода и пусть $\hat{\alpha}$ — примитивный элемент для второго кода. Тогда при некотором i имеем $\hat{\alpha} = \alpha^i$. Для любого кодового слова $\hat{x}(D)$ во втором коде имеем

$$\hat{x}(\hat{\alpha}^j) = 0; \qquad 1 \leqslant j \leqslant d-1,$$
$$\hat{x}(\alpha^{ij}) = 0; \qquad 1 \leqslant j \leqslant d-1. \qquad (1)$$

Теперь рассмотрим перестановку, при которой n-й символ каждого кодового слова второго кода отображается в символ $R_N(in)$ первого кода (т. е. вычет от in по модулю длины блока N). Так как оба элемента α и $\hat{\alpha}$ примитивные, а i и N взаимно простые, то мы имеем дело с перестановкой (другими словами, каждая позиция во втором коде отображается в другую позицию первого кода). Из (1) тогда имеем

$$\sum_{n=0}^{N-1} \hat{x}_n \alpha^{nlj} = \sum_{n=0}^{N-1} x_{R_N(in)} \alpha^{inj} = 0,$$

где $x(D)$ — рассмотренная выше перестановка $\hat{x}(D)$. Полагая $n' = R_N(in)$ и заметив, что n' пробегает все значения от 0 до $N-1$, последнее соотношение представим в следующем виде:

$$\sum_{n'} x_{n'}\, \alpha^{n' j} = 0; \quad 1 \leqslant j \leqslant d-1.$$

6.31. (а) Полагая все информационные символы, кроме одного, равными нулю, получаем кодовое слово веса, не превышающего $N - L + 1$.

(б) Так как для кода Рида-Соломона $m = 1$, минимальный многочлен для α^l равен $D - \alpha^i$. Поэтому при данном d порождающий многочлен имеет степень $d - 1 = N - L$. Так как $d_{min} \geqslant d$ для БЧХ кода, отсюда и из (а) следует, что $d_{min} = N - L + 1$. Заметим, что эти коды полезны лишь при больших объемах алфавита, поскольку $N \leqslant q - 1$.

(в) Множество стираний будет декодировано верно, если не существует двух или более кодовых слов, отличающихся лишь в стертых позициях. Ошибки не произойдет, если число стираний меньше d_{min}. При любых заданных $d_{min} - 1$ позициях не существует двух кодовых слов, которые совпадали бы в оставшихся L позициях. Так как число кодовых слов равно q^L, то это означает, что для каждой совокупности из L позиций существует ровно одно кодовое слово, которое на данной совокупности позиций принимает любые данные возможные значения. Поэтому, если произошло $d_{min} + i$ стираний, то существуют кодовые слова, которые принимают все возможные значения в последних $i + 1$ стертых позициях и имеют поступившие из канала значения во всех нестертых позициях. Это означает, что существуют $q^i + 1$ кодовых слов, согласующихся в нестертых позициях с любой принятой последовательностью с $i + d_{min}$ стираниями.

(г) Из пункта (в) при $i = 0$ следует, то существуют q кодовых слов, символы которых равны нулю для любой заданной совокупности $N - d_{min}$ позиций. Одно из них нулевое кодовое слово, а другие должны быть ненулевыми в остальных d_{min} позициях (так как ни одно ненулевое кодовое слово не может иметь вес, меньший d_{min}). Число таких заданных совокупностей равно $\binom{N}{d_{min}}$ и, следовательно, в коде Рида-Соломона вес d_{min} имеют $(q-1)\binom{N}{d_{min}}$ слов.

6.32. Так как α — примитивный элемент, то длина блока в коде равна $2^m - 1$ и порождающий многочлен имеет степень m. Поскольку все единичные ошибки могут быть исправлены и число проверочных символов равно m, то код должен быть кодом Хэмминга. Соотношение (6.7.27) при $d = 3$ имеет вид

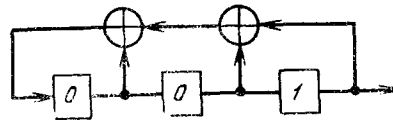
$$\sigma_0 S_1 + \sigma_1 S_0 = 0.$$

Так как σ_0 всегда равно 1, то $\sigma_1 = -S_1/S_0$ (или при отсутствии ошибок $\sigma_1 = 0$). При одной ошибке, например в n_1-й позиции, $S_0 = \alpha^{n_1}$, $S_1 = (\alpha^{n_1})^2$ и, следовательно,

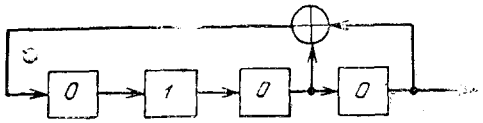
$$\sigma_1 = -\alpha^{n_1}.$$

Это решение для $\sigma(D) = \sigma_0 + \sigma_1 D$ позволяет найти n_1 по σ_1 и, следовательно, исправить ошибку. Таблица декодирования в данном случае, по существу, эквивалентна нахождению n_1 по σ_1.

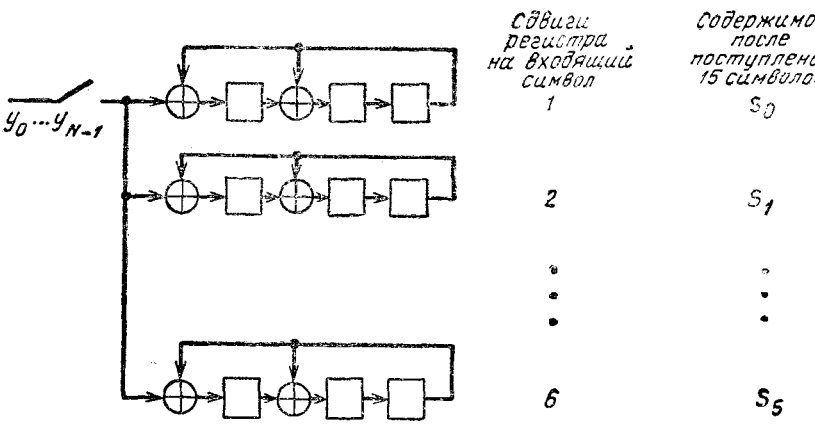
6.33. РСЛОС минимальной длины для $1 + D^3 + D^4 + D^7$ имеет вид, изображенный на рисунке.



Первоначально в регистре хранятся представленные на рисунке символы, а на выходе вначале появляется постоянный член, т. е. 1, затем коэффициент при D, т. е. 0, затем коэффициент при D^2, т. е. 0 и т. д. РСЛОС минимальной длины для $D^2 + D^5 + D^6$ изображена на рисунке, на котором указаны символы, находящиеся в регистре в начальный момент.



6.34. Существует много способов выполнения этого. Приведенная здесь блок-схема, по-видимому, является простейшей по идее, хотя и не самой простой при осуществлении.



Все регистры сдвига первоначально незаполнены; затем поступает y_{N-1}, после чего каждый регистр сдвигается вправо указанное число раз. По окончании этих сдвигов в i-м регистре будет храниться величина $y_{N-1}\alpha^i$, представленная в виде многочлена, у которого члены высшего порядка находятся справа. После этого поступает y_{N-2} и регистры вновь сдвигаются, после чего в i-м регистре остается $y_{N-1}\,\alpha^{2i} + y_{N-2}\alpha^i$. После того как поступит, наконец, y_0 и будут проведены сдвиги, то в i-м регистре будет $y(\alpha^i) = S_{i-1}$, что и требовалось.

6.35. Алгоритм для $A_n(D)$ определяется соотношением

$$A_{n+1}(D) = A_n(D) - \frac{d_n}{d_{k_n}} D^{n-k_n} A_{k_n}(D); \quad n \geqslant 1, \tag{1}$$

с начальными условиями $A_0(D) = 0$, $A_{-1}(D) = -D^{-1}$. Нужно показать, что при каждом $n \geqslant 0$, определенное выше $A_n(D)$ равно

$$A_n(D) = [C_n(D)\, S(D)]_0^{n-1}. \tag{2}$$

Заметим, что (2) справедливо при $n = 0$, так как обе части (2) равны 0. Теперь воспользуемся индукцией. Предположим, что (2) выполняется для $A_1(D)$, $A_2(D)$, ..., $A_n(D)$ и покажем, что это соотношение справедливо для $A_{n+1}(D)$. Используя (6.7.36), получаем

$$[C_{n+1}(D)\,S(D)]_0^n = [C_n(D)\,S(D)]_0^n - \left[\frac{d_n}{d_{k_n}} D^{n-k_n}\, C_{k_n}(D)\,S(D)\right]_0^n.$$

Подставляя значения d_n [см. (6.7.34)] в первое из приведенных выше слагаемых и вынося D^{n-k_n} за скобки справа (заметим, что это требует уменьшения верхнего индекса скобки на $n - k_n$), получаем

$$[C_{n+1}(D)\,S(D)]_0^n = [C_n(D)\,S(D)]_0^{n-1} + \\ + d_n D^n - \frac{d_n}{d_{k_n}} D^{n-k_n} \left[C_{k_n}(D)\,S(D)\right]_0^{k_n}. \qquad (3)$$

Рассмотрим отдельно случай $k_n \geqslant 0$ и $k_n = -1$.

Случай 1. $k_n \geqslant 0$.

По определению d_{k_n},

$$\left[C_{k_n}(D)\,S(D)\right]_0^{k_n} = \left[C_{k_n}(D)\,S(D)\right]_0^{k_n-1} + d_{k_n} D^{k_n}.$$

Подставляя это в (3) и сокращая подобные члены, получаем

$$[C_{n+1}(D)\,S(D)]_0^n = [C_n(D)\,S(D)]_0^{n-1} - \frac{d_n}{d_{k_n}} D^{n-k_n} \left[C_{k_n}\,S(D)\right]_0^{k_n-1}.$$

В силу предположений индукции соотношение (2) справедливо для каждого из стоящих справа слагаемых, так что

$$[C_{n+1}(D)\,S(D)]_0^n = A_n(D) - \frac{d_n}{d_{k_n}} D^{n-k_n} A_{k_n}(D) = A_{n+1}(D).$$

[в силу (1)].

Случай 2. $k_n = -1$.

В этом случае правое слагаемое в (3) равно 0, так что

$$[C_{n+1}(D)\,S(D)]_0^n = [C_n(D)\,S(D)]_0^{n-1} + d_n D^n. \qquad (4)$$

Напомним, что при $k_n = -1$ имеем $d_{k_n} = 1$, $A_{k_n}(D) = -D^{-1}$, поэтому

$$d_n D^n = -\frac{d_n}{d_{k_n}} D^{n-k_n} A_{k_n}(D) \quad \text{для} \quad k_n = -1. \qquad (5)$$

Подставляя (5) в (4) и используя (2) в предположениях индукции, получаем

$$[C_{n+1}(D)\,S(D)]_0^n = A_n(D) - \frac{d_n}{d_{k_n}} D^{n-k_n} A_{k_n}(D) = A_{n+1}(D)$$

[в силу (1)].

Таким образом показано, что (2) справедливо при $n + 1$, что завершает доказательство.

6.36. Пусть $f(D) = g(D)\,h(D)$, где $g(D)$ — многочлен n-й степени над $GF(p^m)$, а $h(D)$ — многочлен k-й степени над $GF(p^m)$. Полагая $f(D) = \sum_{i=0}^{n+k} f_i D^i$, имеем

$$f'(D) = \sum_{i=1}^{n+k} i f_i D^{i-1}.$$

Коэффициент i в приведенном выше выражении может интерпретироваться как i-кратная сумма единиц, так что, быть может, более ясной является запись

$$f'(D) = \sum_{i=0}^{n+k} R_p(i)\, f_i D^{i-1},$$

где $R_p(i)$ интерпретируется как целый элемент поля $GF(p^m)$. Здесь суммирование распространено до $i = 0$, так как соответствующее слагаемое равно 0. Можно выразить каждый из коэффициентов f_i через коэффициенты $g(D)$ и $h(D)$:

$$f_i = \sum_{j=0}^{i} g_j h_{i-j},$$

$$f'(D) = \sum_{i=0}^{n+k} R_p(i) \sum_{j=0}^{i} g_j h_{i-j} D^{i-1}.$$

Заметив, что суммирование приведенной выше двойной суммы производится по всем $j < i$, меняем порядок суммирования, просуммировав по всем $i > j$:

$$f'(D) = \sum_{j=0}^{n+k} \sum_{i=j}^{n+k} R_p(i) g_j h_{i-j} D^{i-1} = \sum_{j=0}^{n} \sum_{i=j}^{n+k} R_p(i) g_j h_{i-j} D^{i-1}.$$

В последнем равенстве использовано то, что $g_j = 0$ при $j > n$. Теперь положим $l = i - j$ и заменим сумму по i суммой по l:

$$f'(D) = \sum_{j=0}^{n} \sum_{l=0}^{n+k-j} R_p(j+l) g_j h_l D^{l+j-1}.$$

Так как $j \leqslant n$ и h_l равно 0 при $l > k$, то во второй сумме суммирование необходимо произвести лишь до $l = k$:

$$f'(D) = \sum_{j=0}^{n} \sum_{l=0}^{k} [R_p(j) + R_p(l)] g_j h_l D^{l+j-1} =$$

$$= \sum_{j=0}^{n} \sum_{l=0}^{k} R_p(j) g_j D^{j-1} h_l D^l + \sum_{j=0}^{n} \sum_{l=0}^{k} R_p(l) g_j D^j h_l D^{l-1} =$$

$$= g'(D) h(D) + g(D) h'(D).$$

Использовав этот результат несколько раз, получим

$$\sigma'(D) = \left[\prod_j (1-U_j D)\right]' = [1-U_1 D]' \prod_{j>1} (1-U_j D) +$$

$$+ (1-U_1 D)\left[\prod_{j>1} (1-U_j D)\right]' = [1-U_1 D]' \prod_{j>1} (1-U_j D) +$$

$$+ (1-U_1 D)[1-U_2 D]' \prod_{j>1} (1-U_j D) + (1-U_1 D)(1-U_2 D) \cdot$$

$$\cdot \left[\prod_{j>2} (1-U_j D)\right]' = \ldots = \sum_{l=1}^{e} [1-U_l D]' \prod_{j \neq l} (1-U_j D) =$$

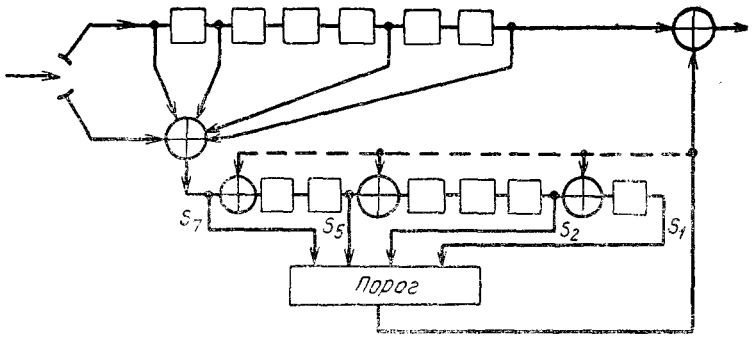
$$= -\sum_{i=1}^{e} U_i \prod_{j \neq i} (1-U_j D).$$

6.37. Представленный здесь кодер можно исследовать так же, как был исследован кодер рис. 6.8.1. Следующие синдромы содержат $z_1^{(1)}$:

$$\begin{aligned} S_1 &= z_1^{(2)} \oplus z_1^{(1)}, \\ S_2 &= z_2^{(2)} \oplus z_2^{(1)} \oplus z_1^{(1)}, \\ S_5 &= z_5^{(2)} \oplus z_5^{(1)} \oplus z_4^{(1)} \oplus z_1^{(1)}, \\ S_7 &= z_7^{(2)} \oplus z_7^{(1)} \oplus z_6^{(1)} \oplus z_3^{(1)} \oplus z_1^{(1)}. \end{aligned} \qquad (1)$$

Заметим, что эти четыре линейные комбинации шумовых символов уже ортогональны к $z_1^{(1)}$ и потому представленный ниже декодер может исправлять все пары ошибок.

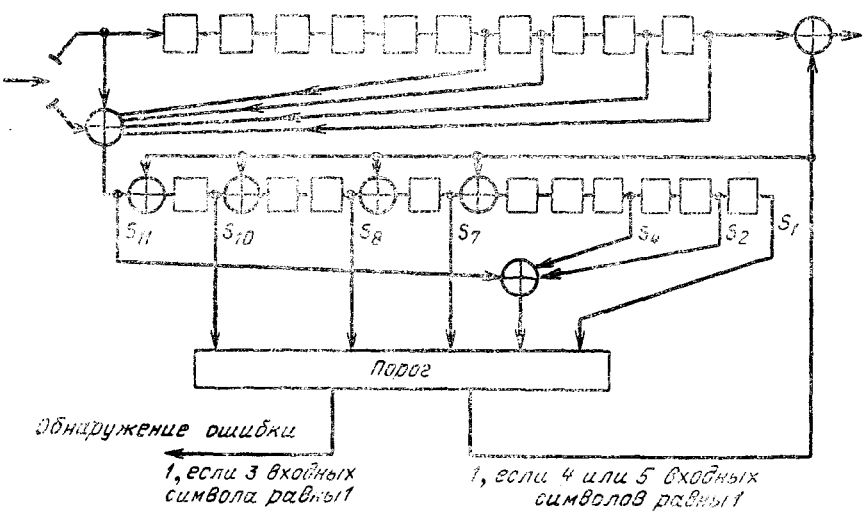
Коды подобного типа, у которых синдромы, проверяющие $z_1^{(1)}$, ортогональны к $z_1^{(1)}$, известны под названием самоортогональных кодов. Интересная особенность таких кодов состоит в том, что линия обратной связи может быть опуще-



на. Это позволяет избежать проблемы распространения ошибок. Чтобы убедиться в этом, просто перепишем четыре равенства (1), включив в них те ошибочные символы, которые уже прошли через декодер. Тогда можно убедиться, что получившиеся линейные комбинации, содержащие также уже исправленные символы, все еще ортогональные к $z_1^{(1)}$. Такое декодирование, не использующее линию обратной связи, называется определенным декодированием (см. Robinson J. P., IEEE Trans., **IT—14**, 121—128, January 1968).

6.38. Проведя то же самое исследование, что и при рассмотрении кодера, представленного на рис. 6.8. 1, получим:

$$S_1 = z_1^{(2)} \oplus z_1^{(1)},$$
$$S_7 = z_7^{(2)} \oplus z_7^{(1)} \oplus z_1^{(1)},$$
$$S_8 = z_8^{(2)} \oplus z_8^{(1)} \oplus z_2^{(1)} \oplus z_1^{(1)},$$



$$S_{10}=z_{10}^{(2)}\oplus z_{10}^{(1)}+z_4^{(1)}\oplus z_3^{(1)}\oplus z_1^{(1)},$$
$$S_{11}=z_{11}^{(2)}\oplus z_{11}^{(1)}\oplus z_5^{(1)}+z_4^{(1)}\oplus z_2^{(1)}\oplus z_1^{(1)}.$$

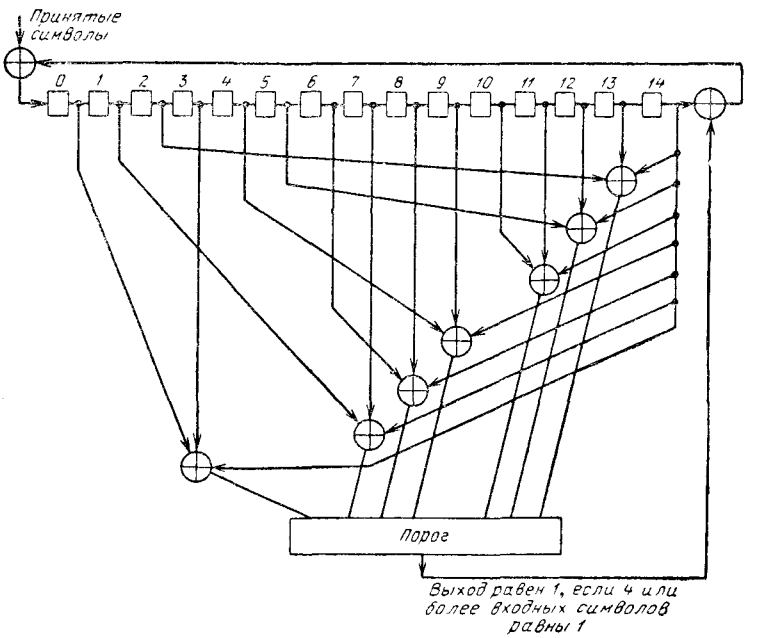
Первые четыре приведенных выше комбинации ортогональны к $z_1^{(1)}$, а последняя может быть включено в это множество, если сложить ее с S_4 и S_2:

$$S_{11}\oplus S_4\oplus S_2=z_{11}^{(2)}\oplus z_{11}^{(1)}\oplus z_5^{(1)}\oplus z_4^{(2)}\oplus z_2^{(2)}\oplus z_1^{(1)}.$$

Тогда декодер имеет вид, представленный на рисунке выше.

6.39. (а) Как показано на стр. 248, дуальным кодом к коду максимальной длины служит код Хэмминга и каждое кодовое слово в дуальном коде является проверочным соотношением в первоначальном коде (т. е. если $\mathbf{h}=(h_0, h_1, \ldots, h_{N-1})$ является кодовым словом дуального кода, то для любого кодового слова $\mathbf{x}=(x_0, \ldots, x_{N-1})$ первоначального кода выполняется соотношение $x_0h_0\oplus x_1h_1\oplus\ldots\oplus x_{N-1}\,h_{N-1}=0$). Напомним теперь, что минимальное расстояние в коде Хэмминга равно 3 и что любая двоичная последовательность из N символов отличается от ближайшего кодового слова не более чем в одной позиции. Поэтому при любом i, $0\leqslant i\leqslant N-1$, можно считать, что в последовательности из N символов на i-й и $(N-1)$-й позициях находится единица, а на остальных позициях — нули. Эта последовательность должна отличаться в одной позиции в точности от одного кодового слова кода Хэмминга и это кодовое слово должно содержать 3 единицы. Поэтому для каждого i, $0\leqslant i<N-1$ существует единственное j, $0\leqslant j<N-1$, такое, что последовательность, содержащая единицы в i-й, j-й и $(N-1)$-й позициях и с нулями в остальных позициях, является кодовым словом кода Хэмминга. Число таких пар (i, j) равно $(N-1)/2$ и каждая из них соответствует одному уравнению из множества $(N-1)/2$ проверочных уравнений кода максимальной длины. Эти проверочные уравнения дают $(N-1)/2$ линейных комбинаций шумовых символов, ортогональных к z_{N-1}.

(б) Поскольку рассматриваемый код циклический, любое устройство, исправляющее z_{N-1}, может быть использовано после циклического сдвига приня-



той последовательности для исправления z_{N-2} и т. д., вплоть до z_0. Так как $(N-1)/2$ нечетно, то при любой конфигурации не более чем $(N-3)/4$ ошибок более половины указанных выше проверочных уравнений будет выполняться, если $z_{N-1}=0$ и более половины из них будет нарушено, если $z_{N-1}=1$. Поэтому z_{N-1} исправляется и, следовательно, после циклического сдвига принятой последовательности можно исправить остальные символы.

(в) Пусть $1+D+D^4$ — порождающий многочлен дуального кода. Кодовое слово веса 3 с единицами на 14-й и 13-й позициях может быть найдено по рис. 6.6.3; заметим, что $\alpha^{14}+\alpha^{13}=\alpha^2$. Это вытекает из того, что α является корнем многочлена $D^{14}+D^{13}+D^2$, поэтому $D^{14}+D^{13}+D^2$ имеет делителем $1+D+D^4$ и, таким образом, является кодовым словом дуального кода. Аналогично, многочлены $D^{14}+D^{12}+D^5$, $D^{14}+D^{11}+D^{10}$, $D^{14}+D^9+D^4$, $D^{14}+D^8+D^6$, $D^{14}+D^7+D$ и $D^{14}+D^3+1$ являются кодовыми словами дуального кода. В регистр сдвига изображенного выше декодера вначале подается полное принятое слово; затем пороговое устройство производит исправление; после этого регистр сдвигается на одну позицию вправо, причем исправленный символ поступает слева. После 14 последовательных исправлений и сдвигов в регистре будет находиться исправленное кодовое слово.

6.40. При данном дереве принятых цен декодер проверит следующую последовательность узлов: a, 0, 00, 01, 0, 1, 10, 11, 1, 10, 100. Заметим, что последний узел имеет меньшую цену в дереве принятых цен, чем узел 010, однако узел 010 никогда не будет просмотрен, хотя все потомки узла 100 в конце концов располагаются ниже нулевого порога.

6.41. (а) В соответствии с соотношением (6.9.2) имеем

$$\Gamma_l=\sum_{i=1}^{l}\sum_{j=1}^{v}\left[\ln\frac{P\left(y_i^{(j)}\mid x_i^{(j)}\right)}{\omega\left(y_i^{(j)}\right)}-B\right].$$

Это выражение является суммой lv независимых, одинаково распределенных случайных величин. Каждая пара {входная буква—выходная буква} выбирается с вероятностью $Q(k)\,P(j|k)$ и поэтому согласно (5.4.16) получим

$$\Pr\left[\Gamma_l\leqslant i\Delta\right]\leqslant e^{-si\Delta}\left[g(s)\right]^{lv};\quad \text{любое } s<0, \tag{1}$$

где

$$g(s)=\sum_{k,j}Q(k)\,P(j\mid k)\exp\left[s\ln\frac{P(j\mid k)}{\omega(j)}-sB\right]=$$
$$=e^{-sB}\sum_{k,j}Q(k)\,P(j\mid k)^{1+s}\,\omega(j)^{-s}. \tag{2}$$

Выбирая распределение на входе, на котором достигается пропускная способность канала, получаем

$$\left.\frac{dg(s)}{ds}\right|_{s=0}=-B+\sum_{k,j}Q(k)\,P(j\mid k)\ln\frac{P(j\mid k)}{\omega(j)}=C-B.$$

Поэтому при любом заданном $B<C$ и отрицательных значениях s, достаточно близких к 0 величина $g(s)<1$. Следовательно, $\Pr[\Gamma_l\leqslant i\Delta]$ стремится к нулю экспоненциально с ростом l. Отметим здесь, что значение $s<0$, минимизирующее правую часть (1), меняется с изменением l, но оно стремится к пределу при $l\to\infty$. Если использовать это предельное значение при всех l, то правая часть (1), равная $e^{-si\Delta}$ при $l=0$, будет убывать экспоненциально по l; величина $e^{-si\Delta}$ больше 1 при $i>0$.

(б)

$$\Gamma_l'=\sum_{i=1}^{l}\sum_{j=1}^{v}\left[\ln\frac{P\left(y_i^{(j)}\mid x_i'^{(j)}\right)}{\omega\left(y_i^{(j)}\right)}-B\right].$$

Это выражение также является суммой $l\nu$ независимых случайных величин, однако здесь в (i, j)-е слагаемое суммы входит передаваемый символ $x_i^{(j)}$ и независимый от $x_i^{(j)}$ символ $x'^{(j)}_i$, который проверяется. Поэтому вероятность пары значений $x_i'^{(j)}$, $y_i^{(j)}$ равна

$$\sum_{x_i^{(j)}} Q\left(x_i^{(j)}\right) P\left(y_i^{(j)} \mid x_i^{(j)}\right) Q\left(x_i'^{(j)}\right) = \omega\left(y^{(j)}\right) Q\left(x_i'^{(j)}\right).$$

Теперь из (5.4.15) выводим, что при любом $r > 0$

$$\Pr\left[\Gamma_l' \geqslant i\Delta\right] \leqslant e^{-ri\Delta}\left\{\sum_{k,j} Q(k)\,\omega(j) \exp\left[r \ln \frac{P(j \mid k)}{\omega(j)} - rB\right]\right\}^{l\nu} =$$

$$= e^{-ri\Delta - l\nu rB}\left[\sum_{k,j} Q(k)\, P(j \mid k)^r\, \omega(j)^{1-r}\right]^{l\nu}.$$

Отсюда, полагая $r = 1 + s$ и используя (2), имеем

$$\Pr\left[\Gamma_l' \geqslant i\Delta\right] \leqslant e^{-i\Delta - l\nu B}\, e^{-si\Delta}\,[g(s)]^{l\nu}, \quad s > -1.$$

Можно заметить, что если не обращать внимания на область изменения s, то эта граница равна умноженной на $\exp[-i\Delta - l\nu B]$ границе вероятности $\Pr[\Gamma_l \leqslant i\Delta]$. Кроме того, легко видеть из рассмотрения (2) (вычисляя $g(s)$ при $s = 0$ и $s = -1$), что $g(s)$ имеет минимум при s, лежащем в пределах от 0 до -1 для $0 < B < C$.

Наконец, число путей на глубине l, отличающихся в первом подблоке от правильного пути, меньше $e^{Rl\nu}$ и $\Pr[\Gamma_l \geqslant i\Delta$ для любого пути на глубине $l] \leqslant$ $\leqslant \exp[-(s+1)\, i\Delta - l\nu\,(B - R)]\, g(s)^{l\nu}$.

При $R \leqslant B$ и таких s, что $g(s) < 1$, эта граница убывает экспоненциально с ростом l. Из этого результата нетрудно сделать вывод, что среднее число вычислений для последовательного декодера при данном конкретном значении Γ_{min} всегда конечно, если $R \leqslant B < C$. Резкое возрастание среднего числа вычислений при $R > E_0(1)$ объясняется тем фактом, что это условное среднее возрастает слишком быстро с возрастанием величины i.

6.42. (а) Предположим, что $\mathbf{u}$ и $\mathbf{u}'$ — информационные последовательности, впервые отличающиеся в $(n+1)$-м подблоке, и пусть $\mathbf{s}$ и $\mathbf{s}'$ — соответствующие им выходные последовательности двоичного сверточного кодера. Пусть $\mathbf{s}'' = \mathbf{s} \oplus \mathbf{s}'$. Рассуждая так же, как и при доказательстве леммы 6.9.1, находим, что $s_i''^{(j)}$, где i изменяется от $n+1$ до $n+L$, а j независимо изменяется от 1 до ν, представляют собой множество νL независимых одинаково распределенных двоичных случайных величин. Так же как и в § 6.9, показывается, что после сложения последовательностей $\mathbf{s}$ и $\mathbf{s}'$ со случайной последовательностью $\mathbf{v}$ отрезки получившихся последовательностей $\mathbf{x}$ и $\mathbf{x}'$ от $(n+1)$-го до $(n+L)$-го подблока статистически независимы и состоят из статистически независимых символов. Если передается $\mathbf{x}$ и цена, принятая для первых $n+l$ подблоков $\mathbf{x}$, равна Γ_{n+l}, а цена, принятая для первых $n+l$ подблоков $\mathbf{x}'$, равна Γ'_{n+l}, то из сказанного выше следует, что $\Gamma_{n+l} - \Gamma_n$ и $\Gamma'_{n+l} - \Gamma_n$ при $l \leqslant L$ имеют те же статистические свойства, что и Γ_l и Γ_l' в лемме 6.9.3. Из леммы 6.9.3 также следует, что $\Gamma_{min} \leqslant$ $\leqslant \min_{1 \leqslant l \leqslant L} \Gamma_l$ и потому соотношение (6.9.16) по-прежнему остается справедливым при замене Γ_{min} на $\min_{1 \leqslant l \leqslant L} \Gamma_l$. Поэтому в рассмотренном здесь случае

$$\Pr\left[\Gamma'_{n+l} - \Gamma_n \geqslant \min_{1 \leqslant l' \leqslant L} \Gamma_{n+l'} - \Gamma_n + (i-2)\,\Delta\right] \leqslant$$

$$\leqslant (l+1) \exp\left[-\frac{(i-2)\,\Delta}{2} - \nu l\, \frac{E_0(1, Q) + B}{2}\right].$$

Из этого результата аналогично тому, как это сделано в основном тексте, получаем (6.9.23). Единственное отличие состоит в том, что в (6.9.13) суммирование проводится только от $l = 0$ до $l = L$.

(б) Ошибка может произойти лишь в том случае, если существует некоторый неправильный путь на глубине $n + L$, ответвляющийся от правильного пути в n-м узле, цена которого превышает $\min\limits_{1 \leqslant l \leqslant L} \Gamma_{n+l} - \Delta$. Вероятность того, что какой-либо отдельный путь удовлетворяет этому условию, определяется формулой 1) при $l = L$ и $i = 1$. Так как число таких путей меньше $e^{LR\nu}$, то

$$P_e \leqslant (L+1) \exp\left[\frac{\Delta}{2} - \nu L\left(\frac{E_0(1, Q)+B}{2} - R\right)\right].$$

6.43. (а)

$$\gamma_i' = \sum_{a=1}^{\nu}\left[\ln \frac{P\left(y_i^{(a)} \mid x_i'^{(a)}\right)}{\omega\left(y_i^{(a)}\right)} - B\right].$$

В силу симметрии заданных условий $\omega(y_i^{(a)}) = \sum\limits_k Q(k)\, P(y_i^{(a)} \mid k) = \dfrac{1}{K}$ (где K — объем алфавита) независимо от символа $y_i^{(a)}$. Аналогично, так как $\mathbf{x}'$ статистически не зависит от $\mathbf{x}$ и $\mathbf{y}$ и так как каждый из символов $x_i'^{(a)}$ принимает любое из K значений входного алфавита независимо от других символов и с равными вероятностями и, следовательно, независимо от $\mathbf{x}$ и $\mathbf{y}$, то каждая величина $P(y_i^{(a)} \mid x_i'^{(a)})$ принимает значения $P(y_i^{(a)} \mid k)$ независимо и с вероятностями $1/K$ при $0 \leqslant k \leqslant K-1$. Поэтому последовательность $\{\gamma_i'\}$ статистически не зависит от $\mathbf{x}$ и $\mathbf{y}$ и, следовательно, не зависит от последовательности $\{\gamma_i\}$. Заметим, что рассмотренный случай представляет собой один из примеров особой ситуации, когда случайные величины $x_i'^{(a)}$, $y_i^{(a)}$ и $P(y_i^{(a)} \mid x_i'^{(a)})$ будут попарно статистически независимыми, но не будут статистически независимыми.

(б) Используя эту независимость, имеем

$$\Pr[\Gamma_l' \geqslant \Gamma_{min} + (i-2)\Delta] = \sum_u \Pr[\Gamma_l' - (i-2)\Delta = u] \Pr[\Gamma_{min} \leqslant u]. \qquad (1)$$

Как и при переходе от (6Б.14) к (6Б.15), получим

$$\Pr[\Gamma_{min} \leqslant u] \leqslant e^{uI/2}. \qquad (2)$$

Подставляя (2) в (1) и используя те же рассуждения, что и при переходе от (6Б.17) к (6Б.22), получаем

$$\Pr[\Gamma_l' \geqslant \Gamma_{min} + (i-2)\Delta] \leqslant \exp\left\{-\frac{(i-2)\Delta}{2} - \frac{\nu l}{2}[E_0(1, Q) + B]\right\}.$$

(в) Границы для $\overline{W}_n$ и $P_{e,\,n}$ можно найти аналогично тому, как это сделано в (6. 9.17) — (6.9.23) и (6.9.38) — (6.9.45). Единственное отличие состоит в том, что в (6.9.17) опускается множитель $(l+1)$, а в (6.9.38) — множитель $(C + L - b + 1)$. В результате получим следующие неравенства:

$$\overline{W}_n \leqslant \frac{4}{1 - \exp\{-\nu[E_0(1, \mathbf{Q}) - R]\}},$$

$$P_{e,\,n} \leqslant \frac{e^{\nu R + \Delta/2}}{\{1 - \exp[-\nu(E_0(1, \mathbf{Q}) - R)]\}^2} \exp\{-\nu L E_0(1, \mathbf{Q})\}.$$

6.44. Используя те же рассуждения, что и при рассмотрении (6.9.13) и 6.9.14), имеем

$$\overline{W_0(u)} < \sum_{l=0}^{\infty} \sum_{m(l)} \sum_{i=1}^{\infty} \Pr[\Gamma'_{m(l)} \geqslant u + (i-2)\Delta]. \tag{1}$$

Из границы Чернова (5.4.15) следует

$$\Pr[\Gamma'_{m(l)} \geqslant u + (i-2)\Delta] \leqslant e^{-s[u+(i-2)\Delta]}[g(s)]^{lv}; \tag{2}$$

$$s > 0 \text{ любое},$$

$$g(s) = \sum_k \sum_j \sum_{k'} Q(k) P(j \mid k) Q(k') \exp\left[s \ln \frac{P(j \mid k)}{\omega(j)} - sB\right] =$$

$$= \left(\frac{1}{2}\right)^{1-s} [(1-\varepsilon)^s + \varepsilon^s] e^{-sB},$$

где ε — вероятность ошибки в ДСК. Полагая $s = 1/(1+\rho)$ или $\rho = (1-s)/s$, приводим это соотношение к виду

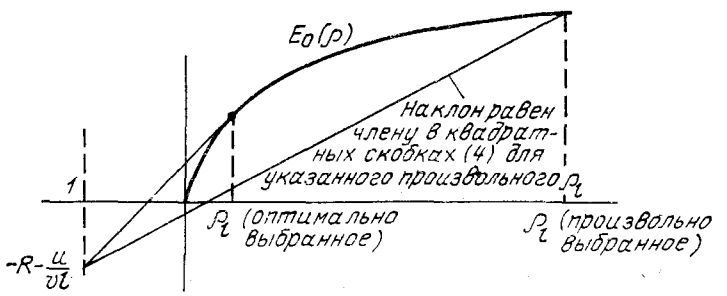
$$g\left(\frac{1}{1+\rho}\right) = \exp\left\{-\frac{1}{1+\rho}[E_0(\rho) + B]\right\}. \tag{3}$$

Подставляя (2) и (3) в (1), вспоминая, что сумма по $m(l)$ содержит не более e^{vlR} слагаемых и используя различные значения $\rho = \rho_l$ при различных l, получаем

$$\overline{W_0(u)} \leqslant \sum_{l=0}^{\infty} \sum_{i=1}^{\infty} \exp\left\{vlR - \frac{1}{1+\rho_l}[u + (i-2)\Delta + Blv + E_0(\rho_l) lv]\right\}.$$

Полагая $B = R$ и суммируя по i, имеем

$$\overline{W_0(u)} \leqslant \sum_{l=0}^{\infty} \frac{e^{\Delta/(1+\rho_l)}}{1 - e^{-\Delta/(1+\rho_l)}} \exp\left\{vlR - vl\left[\frac{u/(vl) + R + E_0(\rho_l)}{1+\rho_l}\right]\right\}. \tag{4}$$



Рисунок, приведенный выше, дает графическое представление о поведении члена, заключенного в квадратные скобки. Из рисунка видно, что при отрицательных u оптимальное значение ρ_l убывает с возрастанием l. В то же время, если выбрать ρ_l равным ρ_r, где $\rho_r R = E_0(\rho_r)$, то можно убедиться, что выражение в фигурных скобках (4) становится равным $-u/(1+\rho_r)$. При больших отрицательных значениях u при некотором l оптимальное значение ρ_l будет близким к ρ_r; это доказывает, что приведенная выше граница $\overline{W_0(u)}$ должна быть по край-

ней мере пропорциональна $\exp[-u/(1+\rho_r)]$ при больших отрицательных значениях u. Это объясняет выбор значений

$$\rho_l=\rho_r; \quad l\leqslant l(u), \qquad \rho_l=\rho'; \quad l>l(u),$$

где ρ' — произвольное число, такое, что $0<\rho'<\rho_r$. Выберем $l(u)$ (вообще говоря, нецелое) как значение l, при котором выражение в фигурных скобках (4) принимает одинаковые значения при $\rho_l=\rho_r$ и при $\rho_l=\rho'$, т. е.

$$l(u)=\frac{-u(\rho_r-\rho')}{\nu[-\rho' R+E_0(\rho')](1+\rho_r)}.$$

Подставляя это выражение в (4) и суммируя отдельно по $l\leqslant l(u)$ и по $l>l(u)$, получаем

$$\overline{W_0(u)}\leqslant A\exp\left[-\frac{u}{1+\rho_r}\right],$$

$$A=[l(u)+1]\frac{e^{\Delta/(1+\rho_r)}}{1-e^{-\Delta/(1+\rho_r)}}+$$

$$+\frac{e^{\Delta/(1+\rho')}}{1-e^{-\Delta/(1+\rho')}}\left[1-\exp\left\{-\frac{\nu}{1+\rho'}[E_0(\rho')-\rho' R]\right\}\right]^{-1}.$$

6.45. Так как $\bar{z}_i<0$, то соотношение (6Б.29) справедливо при $r=0$. Взяв производные от обеих частей (6Б.29) по r, получим

$$0=\sum_{n=1}^{\infty}\sum_{v\leqslant u} f_{0,n}(u,v)\,e^{rv}[g(r)]^{-n}\left[v-n\frac{g'(r)}{g(r)}\right].$$

Полагая $r=0$ и используя тот факт, что $g(0)=1$, $g'(0)=\bar{z}_i$, получаем

$$0=\sum_{n=1}^{\infty}\sum_{v\leqslant u} f_{0,n}(u,v)[v-n\bar{z}_i].$$

Из определения $f_{0,n}(u,v)$ нетрудно видеть, что первый член приведенного выше выражения есть математическое ожидание времени блуждания до первого пересечения барьера, так что $0=\overline{S}_N-N\bar{z}_i$. Если z_{min} является минимальным значением (т. е. максимальное по модулю значение отрицательной величины), принимаемым z, то в момент, когда при блуждании в первый раз будет пересечен барьер, S_n не может принять значения, лежащего ниже барьера на расстоянии, большем чем z_{min}, и $u\geqslant S_n>u+z_{min}$. Подставляя это в (1), получаем $u+z_{min}<\overline{N}\bar{z}_i<u$.

6.46. Пусть $\mathbf{x}_1$ и $\mathbf{x}_2$ — произвольные кодовые последовательности некоторой заданной длины. Пусть $\mathbf{z}_1$ равно $\mathbf{x}_1\oplus\mathbf{x}_2$ во всех позициях последовательности с номерами $2ig+j$, где $0<j\leqslant g-1$, а i — произвольное целое число, и пусть $\mathbf{z}_1$ равно 0 во всех остальных позициях. Пусть $\mathbf{z}_2$ равно $\mathbf{x}_1\oplus\mathbf{x}_2$ во всех позициях с номерами $(2i+1)g+j$, где $0\leqslant j\leqslant g-1$, а i — произвольное целое, и пусть $\mathbf{z}_2$ равно нулю во всех остальных позициях. Тогда $\mathbf{z}_1$ и $\mathbf{z}_2$ представляют собой пакет длины, не большей g с требуемым защитным интервалом, но

$$\mathbf{z}_1\oplus\mathbf{z}_2=\mathbf{x}_1\oplus\mathbf{x}_2; \qquad \mathbf{z}_1\oplus\mathbf{x}_1=\mathbf{z}_2\oplus\mathbf{x}_2.$$

7.1. (а)

$$P_Y(y)=\begin{cases}1/4; & |y|\leqslant 1,\\ 1/8; & 1<|y|\leqslant 3,\\ 0; & |y|>3\end{cases}$$

(б)

$$I(X;\ Y)=P_X(1)\int\limits_{-1}^{3} p_{Y\mid X}(y\mid 1)\log\frac{p_{Y\mid X}(y\mid 1)}{p_Y(y)}\,dy+$$

$$+P_X(-1)\int\limits_{-3}^{1} p_{Y\mid X}(y\mid -1)\log\frac{p_{Y\mid X}(y\mid -1)}{p_Y(y)}\,dy=\frac{1}{2}\ \text{бит}.$$

(в) Канал, определенный $P(y\mid z)$, является ДСтК и

$$I(X;\ Z)=\frac{1}{2}\ \text{бит}.$$

Это показывает, что отображение y в z не разрушает информацию относительно x. Это естественно, так как для каждого $|y|<1$, значения ± 1 принимаются x с равными вероятностями.

7.2. (а) Из симметрии ясно, что пропускная способность достигается на равновероятных входах. Следовательно,

$$C=\int\limits_{-\infty}^{\infty} p(y\mid 1)\ln\left[\frac{p(y\mid 1)}{{}^1/_2 p(y\mid 1)+{}^1/_2 p(y\mid -1)}\right]dy\ \text{нат}=$$

$$=\ln 2-1-\frac{1}{a+b}\int\limits_0^{\infty}\left(e^{-y/a}+e^{-y/b}\right)\ln\left(e^{-y/a}+e^{-y/b}\right)dy.$$

В пределе при $b/a\to 1$ имеем $C=0$, а в пределе при $b/a\to 0$ имеем $C=\ln 2$ нат. Более точное разложение показывает, что

$$C\approx\frac{(a-b)^2}{4a^3}\ \text{для}\ \frac{b}{a}\approx 1.$$

(б) Для частного случая $\rho=1$ функция $E_0(\rho)$ может быть непосредственно вычислена. Результат (оптимизированный равновероятными входами) имеет вид

$$E_0(1)=-\ln\left[1-\frac{1}{2}\left(\frac{a-b}{a+b}\right)^2\right].$$

Для $\frac{b}{a}\approx 1$, $E_0(1)\approx\frac{(a-b)^2}{8a^2}$. Заметим, что это равно половине C. Это следовало ожидать, так как при равновероятных входах рассматриваемый канал удовлетворяет условиям, справедливым для канала с очень большим шумом.

(в) $p(y\mid x)$ можно представить следующим образом:

$$p(y\mid x)=\frac{1}{a+b}\,e^{-|y|/a}\varphi(x,\ y),$$

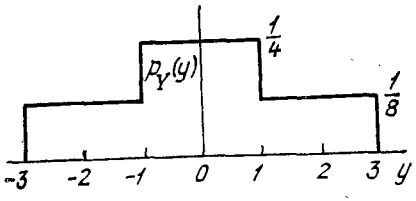
где

$$\varphi(x,\ y)=\begin{cases}1; & xy\geqslant 0,\\ \exp\left\{-|y|\left[\dfrac{1}{b}-\dfrac{1}{a}\right]\right\}; & xy<0.\end{cases}$$

Тогда

$$\ln p(y\mid x_m)=\sum_{n=1}^{N}\ln\left[\frac{1}{a+b}\,e^{-|y_n|/a}\right]+\sum_{n=1}^{N}\ln\varphi(x_{m,\,n},\ y_n).$$

При декодировании по минимуму правдоподобия выбирается m, которое максимизирует приведенное выше выражение. Так как первое слагаемое не является



функцией x_m, то выбирается m, которое максимизирует

$$\sum_{n=1}^{N} \ln\varphi(x_{m,n}, y_n) = \sum_{n;\ x_{m,n} y_n < 0} -|y_n| \left(\frac{1}{b} - \frac{1}{a}\right).$$

Так как $\frac{1}{b} - \frac{1}{a} > 0$, то это эквивалентно выбору m, которое минимизирует

$$\sum_{n;\ x_{m,n} y_n < 0} |y_n|.$$

(г) При отсутствии кодирования принимается $x = 1$ при $y > 0$ и $x = -1$ при $y \leqslant 0$. Следовательно,

$$P_e = \int_{-\infty}^{0} \frac{1}{a+b} e^{y/b}\, dy = \frac{b}{a+b}.$$

7.3. (а) Ограничивая входной алфавит множеством из K букв, получаем стирающий канал с K входами, и как пропускная способность, так и $E_0(\rho)$ достигаются на равновероятных входах. Имеем

$$C(K) = {}^1/_2 \ln K, \qquad E_0(\rho, K) = \ln 2 - \ln[K^{-\rho} + 1].$$

В пределе при $K \to \infty$ получаем

$$C = \infty,$$

$$E_0(\rho) = \begin{cases} \ln 2, & \rho > 0, \\ 0, & \rho = 0, \end{cases}$$

$$E_r(R) = \sup_{0 \leqslant \rho \leqslant 1} [E_0(\rho) - \rho R] = \ln 2 \text{ для всех } R.$$

(б) Для любого кода с M кодовыми словами и длиной блока N полностью стертая последовательность принимается с вероятностью 2^{-N}, и при условии наступления этого события вероятность неправильного декодирования равна $(M-1)/M$. Следовательно,

$$P_e \geqslant \frac{M-1}{M} 2^{-N}.$$

Это выражение отличается от $P_e \leqslant e^{-NE_r(R)} = 2^{-N}$ только множителем $(M-1)/M$.

7.4. (а) Для того чтобы показать, что C достигается при $p_X(x) = 1/2\pi$, $0 \leqslant x < 2\pi$, достаточно показать, что при таком выборе

$$\int_0^{2\pi} p(y|x) \ln \frac{p(y|x)}{\int p_X(x)\, p(y|x)\, dx} dy \tag{1}$$

не зависит от x. Используя равенство

$$\int_{-2\pi+y}^{y} \frac{1}{2\pi} p_Z(y-x)\,dx = \frac{1}{2\pi} \text{ для всех } y,\ 0 \leqslant y < 2\pi,$$

и изменяя порядок интегрирования, интеграл (1) можно привести к выражению

$$C = \int_0^{2\pi} p_Z(z) \ln [2\pi p_Z(z)]\,dz,$$

которое, очевидно, не зависит от x и, следовательно, равно пропускной способности. Аналогичные соображения показывают, что та же плотность удовлетворяет (5.6.37) и

$$E_0(\rho) = -\ln(2\pi) - (1+\rho) \ln \int_0^{2\pi} \frac{1}{2\pi} p_Z(z)^{1/(1+\rho)}\,dz.$$

(б) Для $p_Z(z) = 1/\alpha, \quad 0 \leqslant z < \alpha$, имеем

$$C = \ln(2\pi/\alpha), \qquad E_0(\rho) = \rho \ln(2\pi/\alpha),$$
$$E_r(R) = C - R, \qquad 0 \leqslant R < C.$$

Для $p_Z(z) = \alpha e^{-\alpha z}/(1-e^{-2\pi\alpha}), \quad 0 \leqslant z < 2\pi,$

$$C = \ln \frac{2\pi\alpha}{1-e^{-2\pi\alpha}} - 1 + \frac{2\pi\alpha}{e^{2\pi\alpha}-1},$$

$$E_0(\rho) = \rho \ln 2\pi\alpha + \ln(1-e^{-2\pi\alpha}) - (1+\rho) \ln [(1+\rho)(1-e^{-\frac{2\pi\alpha}{1+\rho}})].$$

Имеем параметрические уравнения для R и $E_r(R)$

$$R = \ln\left[\frac{2\pi\alpha}{e(1+\rho)(1-\exp[-2\pi\alpha/(1+\rho)])}\right] + \frac{2\pi\alpha}{(1+\rho)\{\exp[2\pi\alpha/(1+\rho)]-1\}},$$

$$E_r(R) = \ln\left[\frac{1-\exp(-2\pi\alpha)}{(1+\rho)[1-\exp(-2\pi\alpha/(1+\rho))]}\right] + \rho - \frac{2\pi\alpha\rho}{(1+\rho)[\exp(2\pi\alpha/(1+\rho))-1]}.$$

Они справедливы для $\rho \leqslant 1$, а для меньших R

$$E_r(R) = E_0(1) - R.$$

7.5. (а) Естественно предположить, что входное распределение должно быть частично сосредоточено в концевых точках и, следовательно, при начальном угадывании используем только $x = -1/2$ и $x = 1/2$ с вероятностями $1/2$. Это приводит к выходной плотности, задаваемой выражением

$$p_Y(y) = \begin{cases} 1/2, & |y| \leqslant 1/2, \\ 1/4, & 1/2 < |y| \leqslant 3/2, \\ 0 & \text{во всех других точках,} \end{cases}$$

$$I(x; Y) = \int_{x-1}^{x+1} p(y \mid x) \ln \frac{p(y|x)}{p_Y(y)}\,dy = \frac{1}{2} \ln 2$$

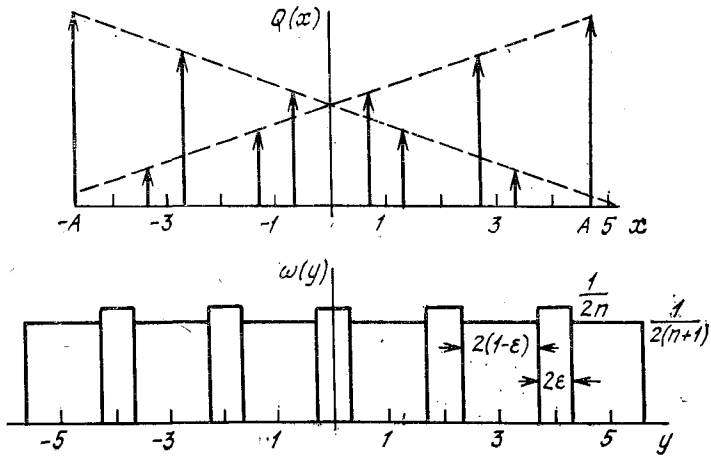
для всех x, $-1/2 \leqslant x \leqslant 1/2$. Выше при выполнении интегрирования необходимо учесть, что $p_Y(y) = 1/4$ точно в половине интервала интегрирования, независимо от x.

Те же соображения, примененные к (5.6.37), подтверждают, что при том же самом входном распределении достигается $E_0(\rho)$, которое равно $E_0(\rho) = \ln 2 - \ln(2^{-\rho} + 1)$. Это выражение совпадает с $E_0(\rho)$ для ДСтК с $\varepsilon = 1/2$ и, следовательно, $E_r(R)$ будет тем же самым (см. решение задачи 5.10). Физически выход $|y| \leqslant 1/2$ не несет информации о входе и, следовательно, может рассматриваться как стирание. Выход $|y| > 1/2$ однозначно определяет вход.

(б) Здесь первый шаг — нахождение выходной плотности вероятности, соответствующей данному входному распределению. Пусть $\varepsilon = n - A$. Входное распределение изображено ниже и выходную плотность можно найти графически, сопоставляя прямоугольник со стороной длины 2 каждой входной точке, взвешивая после этого каждый такой прямоугольник и суммируя.

Для любого x из интервала $[-A, A]$ имеем

$$I(x; Y) = \int_{x-1}^{x+1} p(y|x) \ln \frac{p(y|x)}{\omega(y)}\, dy.$$



Так как $p(y \mid x) = 1/2$ во всей области интегрирования, $\omega(y)$ равна $1/(\alpha n)$ на интервале (или интервалах) общей длины 2ε и $\omega(y)$ равна $1/[2(n+1)]$ в оставшейся области, то имеем

$$I(x; Y) = \varepsilon \ln n + (1-\varepsilon)\ln(n+1) = \ln(n+1) - (n-A)\ln\frac{n+1}{n}.$$

Это выражение одно и то же для всех x из $[-A, A]$, поэтому при этом распределении достигается пропускная способность и

$$C = \ln(n+1) - (n-A)\ln\frac{n+1}{n}.$$

Заметим, что это выражение возрастает с A, кусочно-линейно по A и приближенно равно $\ln(A+1)$ для больших A.

Те же рассуждения применимы к (5.6.37) и показывают, что это распределение максимизирует $E_0(\rho, Q)$. В результате получаем

$$E_0(\rho, Q) = -\ln\left\{(n-A)\left(\frac{1}{n}\right)^{\rho} + (1-n+A)\left(\frac{1}{n+1}\right)^{\rho}\right\}.$$

Выражение для $E_r(R)$ следует из (5.6.31) и (5.6.32). Эти выражения довольно громоздки и ненаглядны, поэтому они не будут даны здесь. Однако в пределе, когда A стремится к целому числу или A становится большим, имеем

$$E_r(R) = \ln(A+1) - R.$$

(в) Для целого A при дискретном распределении

$$Q(A-2i)=1/(A+1), \quad i=0,1,\ldots,K,$$

достигается как пропускная способность, так и $E_r(R)$. В обозначениях рисунка для $Q(x)$ в пункте (б) это соответствует парам импульсов, которые приходят вместе и складываются.

7.6. (а) Из (7.5.7) вытекает, что средняя взаимная информация для заданного набора энергий $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_N$ удовлетворяет неравенству

$$I(X^N; Y^N) \leqslant \sum_{n=1}^{N} {}^1/_2 \ln(1+\mathscr{E}_n/\sigma_n^2)$$

с равенством, если входы независимые гауссовские случайные величины с нулевыми средними. Полагая $\alpha_n=\mathscr{E}_n/n$ и используя $\sigma_n^2 = n^2$, это соотношение приводим к виду

$$I(X^N; Y^N) \leqslant \sum_{n=1}^{N} {}^1/_2 \ln(1+\alpha_n/n).$$

Максимизируя это выражение по α при ограничении $\Sigma\alpha_n \leqslant 5$, получаем [как и в (7.5.3) и (7.5.4)]

$$n+\alpha_n = B, \quad n < B, \tag{1}$$

$$\alpha_n = 0, \quad n \geqslant B. \tag{2}$$

Простейший путь решить эти уравнения состоит в том, чтобы сначала решить их относительно B и $\alpha_1, \ldots, \alpha_N$ одновременно, используя (1) и $\Sigma\alpha_n = 5$ и затем модифицировать решение, положив $\alpha_n = 0$, если данное α_n окажется отрицательным. Для $N = 2$ это приводит к $\alpha_1 = 3$, $\alpha_2 = 2$ и, следовательно, $\mathscr{E}_1 = 3$, $\mathscr{E}_2 = 4$, $N = 2$,

$$C = {}^1/_2 \ln 4 + {}^1/_2 \ln 2 = {}^3/_2 \ln 2.$$

Для $N = 4$ эта же процедура дает $B = 3\,{}^3/_4$, $\alpha_1 = 2^3/_4$, $\alpha_2 = 1^3/_4$, $\alpha_3 = {}^3/_4$, $\alpha_4 = -{}^1/_4$. При этом нарушается ограничение, так что следует положить $\alpha_4 = 0$. Тогда получаем $B = 3^2/_3$, $\alpha_1 = 2^2/_3$, $\alpha_2 = 1^2/_3$, $\alpha_3 = {}^2/_3$. При этом удовлетворяются (1) и (2), так что имеем $\mathscr{E}_1 = 2^2/_3$, $\mathscr{E}_2 = 3^1/_3$, $\mathscr{E}_3 = 2$, $\xi_4 = 0$,

$$C = {}^1/_2 \ln {}^{11}/_3 + {}^1/_2 \ln {}^{11}/_6 + {}^1/_2 \ln {}^{11}/_9 \approx 1{,}05.$$

Это же решение справедливо для $N = \infty$ при $\mathscr{E}_n \approx 0$ для $n > 3$.

(б). Равенство (7.5.20) справедливо для рассматриваемой здесь задачи, однако ограничение $\sum_n \mathscr{E}_n/n \leqslant 5$ приводит к максимизации (7.5.21) при ограничении $\sum_n A_n n = 5$.

Следовательно, эта задача математически идентична решенной в § 7.5, если заменить σ_n^2 на n^2 и $\mathscr{E}_n$ на α_n. Физически, наши действия сводятся к нормированию энергии сигнала и дисперсии шума в каждом подканале.

Используя (7.5.35), можно найти B_{cr} с помощью метода проб и ошибок получаем $B_{cr} \approx 2{,}42$. Подставляя это значение в (7.5.36), находим

$$R_{cr} = {}^1/_2 \ln 2{,}42 + {}^1/_2 \ln 1{,}21 \approx 0{,}537.$$

Также подставляя B_{cr} в (7.5.34) с $\rho = 1$, находим $E_r(R_{cr}) \approx 0{,}206$. Наконец,

$$E_r(0) = E_r(R_{cr}) + R_{cr} \approx 0{,}743.$$

(в) Выражение (1) можно переписать в виде

$$n+\alpha_n = B; \quad n \leqslant \lfloor B \rfloor.$$

Суммируя эти равенства от $n=1$ до $n=\lfloor B' \rfloor$ и учитывая, что $\Sigma\alpha_n=50$, получаем

$$\frac{(\lfloor B \rfloor)(\lfloor B \rfloor+1)}{2}+50=\lfloor B \rfloor B. \tag{3}$$

Учитывая, что B, $\lfloor B \rfloor$ и $\lfloor B \rfloor+1$ сравнительно близки друг к другу, видим, что $B^2/2+50\approx B^2$ или $B\approx 10$. Взяв $\lfloor B \rfloor=10$, уравнение (3) можно решить относительно B, что дает $B=10{,}5$. Следовательно,

$$\mathscr{E}_n=n(10{,}5-n),\quad n\leqslant 10,$$
$$\mathscr{E}_n=0,\qquad\qquad n>10.$$

7.7. (а) Пусть $\mathbf{y}=(y_1, y_2, \ldots, y_N)$ — принятая последовательность. Тогда

$$\ln\frac{p(\mathbf{y}\mid \mathbf{x}_1)}{p(\mathbf{y}\mid \mathbf{x}_2)}=\sum_{n=1}^{N}\left\{-\frac{(y_n-\sqrt{\mathscr{E}_n})^2}{2\sigma_n^2}+\frac{(y_n+\sqrt{\mathscr{E}_n})^2}{2\sigma_n^2}\right\}=\sum_{n=1}^{N}\frac{2y_n\sqrt{\mathscr{E}_n}}{\sigma_n^2}\triangleq\Lambda.$$

Декодер выбирает сообщение 1, если эта величина больше 0 и сообщение 2 в противном случае.

При условии, что сообщение 1 послано, пусть $y_1,\ldots,y_N$ — последовательность независимых гауссовских случайных величин y_n, имеющих средние $\sqrt{\mathscr{E}_n}$ и дисперсии σ_n^2. Следовательно, Λ имеет среднее $\sum_{n=1}^{N}2\mathscr{E}_n/\sigma_n^2$, дисперсию $\sum_{n=1}^{N}4\mathscr{E}_n/\sigma_n^2$ и

$$P_{e,1}=\Pr[\Lambda<0\mid x_1]=\Phi\left(-\sqrt{\sum_{n=1}^{N}\mathscr{E}_n/\sigma_n^2}\right), \tag{1}$$

где $\Phi(\cdot)$ — функция распределения гауссовской случайной величины с нулевым средним и единичной дисперсией. Из симметрии ясно, что $P_{e,2}=P_{e,1}$ и (1) дает вероятность ошибки.

(б) Вероятность P_e минимизируется выбором таких $\mathscr{E}_n$, $1\leqslant n\leqslant N$, которые максимизируют $\sum_{n=1}^{N}\mathscr{E}_n/\sigma_n^2$ при ограничении $\Sigma\mathscr{E}_n\leqslant\mathscr{E}$.

Пусть j — номер канала с наименьшей дисперсией шума, так что $\sigma_j^2\leqslant\sigma_n^2$ для $1\leqslant n\leqslant N$. Тогда

$$\sum_{n=1}^{N}\frac{\mathscr{E}_n}{\sigma_n^2}\leqslant\sum_{n=1}^{N}\frac{\mathscr{E}_n}{\sigma_j^2}\leqslant\frac{\mathscr{E}}{\sigma_j^2}$$

с равенствами всегда, когда $\mathscr{E}_j=\mathscr{E}$ и $\mathscr{E}_n=0$ для всех $n\neq j$. Следовательно, $\mathscr{E}_j=\mathscr{E}$ и $\mathscr{E}_n=0$ для $n\neq j$ минимизируют вероятность ошибки и

$$\min P_e=\Phi\left(-\sqrt{\mathscr{E}/\sigma_j^2}\right)\approx\sqrt{\sigma_j^2/2\pi\mathscr{E}}\exp[-\mathscr{E}/2\sigma_j^2]$$

для $\mathscr{E}\gg\sigma_j^2$ (см. Феллер, т. 1, гл. VII, § 1).

(в) Из (7.5.57) видим, что B — возрастающая функция R' и что при R', стремящемся к 0, B стремится к σ_j^2. Следовательно, из (7.5.58) находим, что

$$E_{ex}(0)=\mathscr{E}/4\sigma_j^2. \tag{2}$$

Заметим, что этот показатель экспоненты равен половине показателя экспоненты, выведенного в пункте (б). Если взглянуть более тщательно на связь между скоростью R и R' [см. (7.5.41) и (7.5.60)], то увидим, что $R\approx R'$ только тогда,

когда $\mathscr{E} \gg \sigma_i^2$ и когда имеется большое число значений n, для которых $\sigma_n^2 \approx \approx \sigma_j^2$. Показатель экспоненты в (2) равен показателю экспоненты, который возникает при малом числе ортогональных кодовых слов и использовании лишь каналов, для которых $\sigma_n^2 = \sigma_j^2$.

8.1 (а) По предположению, если $x_i(t)$, $i = 1, 2, \ldots$ — любая последовательность функций из L^2, аппроксимирующих функцию $x(t)$ из L^2 в том смысле, что

$$\lim_{i \to \infty} \int |x(t) - x_i(t)|^2 \, dt = 0,$$

и если последовательность $x_i = \int x_i(t) z(t) \, dt$, $i = 1, 2, \ldots$, сходится к пределу, скажем x, в том смысле, что $\lim_{i \to \infty} \overline{(x_i - x)^2} = 0$, то $x = \int x(t) z(t) \, dt$.

К этой задаче наиболее естественно подойти, рассматривая функции $x(t)$ из множества L^2 как элементы гильбертового пространства со скалярным произведением $(x(t), y(t)) = \int x(t) y(t) \, dt$. Аналогично случайные величины $x = \int x(t) y(t) \, dt$ можно рассматривать как элементы гильбертового пространства со скалярным произведением $(x, y) = \overline{xy}$. Тогда $x = \int x(t) z(t) \, dt$ можно рассматривать как линейное (быть может, неограниченное) преобразование в гильбертовом пространстве. Доказываемый результат эквивалентен теореме § 117 книги Рисса и Надь (1955).

(б) Так как $\overline{\left[\int x(t) z(t) \, dt\right]^2} \leqslant M$ для всех нормированных $x(t)$, то для произвольных ненормированных $x(t)$ имеем

$$\overline{\left[\int x(t) z(t) \, dt\right]^2} = \left[\int x^2(t) \, dt\right] \overline{\left[\int \frac{x(t)}{\int x^2(\tau) \, d\tau} z(t) \, dt\right]^2} \leqslant M \int x^2(t) \, dt. \qquad (1)$$

Для $x(t) = \sum_{i=1}^{\infty} x_i \varphi_i(t)$ можно использовать условие линейности (8.1.36) и получить

$$\int x(t) z(t) \, dt = \sum_{i=1}^{k} x_i \int \varphi_i(t) z(t) \, dt + \int \left[\sum_{i=k+1}^{\infty} x_i \varphi_i(t)\right] z(t) \, dt.$$

Следовательно,

$$\int x(t) z(t) \, dt - \sum_{i=1}^{k} x_i z_i = \int \left[\sum_{i=k+1}^{\infty} x_i \varphi_i(t)\right] z(t) \, dt. \qquad (2)$$

Применяя (1) к правой части (2), находим

$$\overline{\left[\int x(t) z(t) \, dt - \sum_{i=1}^{k} x_i z_i\right]^2} \leqslant$$

$$\leqslant M \int \left[\sum_{i=k+1}^{\infty} x_i \varphi_i(t)\right]^2 dt = M \sum_{i=k+1}^{\infty} x_i^2. \qquad (3)$$

Так как $\int x^2(t) \, dt = \sum_{i=1}^{\infty} x_i^2$ конечно, то правая часть (3) стремится к нулю с возрастанием k, давая требуемый результат.

8.2. Используя равенство (1), из решения задачи 8.1. (б), имеем

$$\overline{\left[\int_0^{k/f}\sqrt{\frac{2f}{k}}(\cos 2\pi f t)\,z(t)\,dt\right]^2}\leqslant$$

$$\leqslant M\int_0^{k/f}\frac{2f}{k}\cos^2(2\pi f t)\,dt = M\int_0^{k/f}\frac{2f}{k}\left[\frac{1}{2}+\frac{\cos 4\pi f t}{2}\right]dt = M\left[1+\frac{1}{4\pi k}\sin 4\pi k\right]\to$$

$$\to M \text{ при } k\to\infty.$$

8.3. Сначала примем, что $\mathcal{R}(t,\tau)$ непрерывна. Тогда имеем

$$\overline{[z(t)-z(t+\varepsilon)]^2} = \overline{z^2(t)}-\overline{2z(t)z(t+\varepsilon)}+\overline{z^2(t+\varepsilon)} = \mathcal{R}(t,t)-2\mathcal{R}(t,t+\varepsilon)+$$
$$+\mathcal{R}(t+\varepsilon,t+\varepsilon)\to 0 \text{ при } \varepsilon\to 0.$$

Далее примем, что $\lim\limits_{\varepsilon\to 0}\overline{[z(t)-z(t+\varepsilon)]^2}=0$ для всех t. Тогда имеем

$$\mathcal{R}(t,\tau)-\mathcal{R}(t+\varepsilon,\tau)=\overline{[z(t)-z(t+\varepsilon)]\,z(\tau)}\leqslant$$
$$\leqslant\overline{([z(t)-z(t+\varepsilon)]^2)}^{1/2}\,(\overline{z^2(\tau)})^{1/2},$$

где в последней части было использовано неравенство. Шварца. Следовательно, $\lim\limits_{\varepsilon\to 0}\mathcal{R}(t+\varepsilon,\tau)=\mathcal{R}(t,\tau)$ для всех τ и t. Так как $\mathcal{R}(t,\tau)=\mathcal{R}(\tau,t)$, то это также показывает, что

$$\lim_{\varepsilon\to 0}\mathcal{R}(t,\tau+\varepsilon)=\mathcal{R}(t,\tau) \text{ для всех } t,\tau.$$

Используя известные результаты математического анализа, находим, что это эквивалентно тому, что $\mathcal{R}(t,\tau)$ непрерывна.

8.4. Пусть $\delta>0$ произвольно. Так как $\int\limits_{-\infty}^{\infty}|X(f)|\,df<\infty$, то f_1 можно выбрать достаточно большим, так чтобы

$$\int_{-\infty}^{-f_1}|X(f)|\,df+\int_{f_1}^{\infty}|X(f)|\,df<\delta/4.$$

Выберем теперь $\varepsilon>0$ достаточно малым, чтобы

$$|1-e^{-j2\pi f\varepsilon}|\leqslant\delta/\left(2\int|X(f)|\,df\right)$$

для всех $|f|\leqslant f_1$. Тогда

$$|x(t)-x(t-\varepsilon)|=\left|\int X(f)\,e^{j2\pi ft}\,[1-e^{-j2\pi f\varepsilon}]\,df\right|\leqslant\int|X(f)|\,|1-e^{-j2\pi f\varepsilon}|\,df\leqslant$$

$$\leqslant\int_{|f|>f_1}2|X(f)|\,df+\int_{-f_1}^{f_1}\frac{|X(f)|\,\delta}{2\int|X(f)|\,df}\,df\leqslant\frac{\delta}{2}+\frac{\delta}{2}=\delta.$$

Следовательно, для любого $\delta>0$ найдется $\varepsilon_1>0$ такое, что $|x(t)-x(t-\varepsilon)|\leqslant\delta$ для всех $\varepsilon<\varepsilon_1$, т. е. $x(t)$ непрерывно по t.

8.5. $y(t)-y(t+\varepsilon)=\int[h(t-\tau)-h(t+\varepsilon-\tau)]\,z(\tau)\,d\tau$.

Используя соотношение (1) решения задачи 8.1. (б), имеем

$$\overline{[y(t)-y(t+\varepsilon)]^2}\leqslant M\int[h(t-\tau)-h(t+\varepsilon-\tau)]^2\,d\tau=M\int[h(t)-h(t+\varepsilon)]^2\,dt. \quad (1)$$

Преобразование Фурье от $w(\varepsilon)=\int h(t)\,h(t+\varepsilon)\,dt$ равно $|H(f)|^2$, где $H(f)$ преобразование Фурье $h(t)$. Так как $h(t)$ из L_2, то $H(f)$ также из L_2 и $\int |H(f)|^2\,df<\infty$.

Следовательно, используя задачу 8.4, находим, что $w(\varepsilon)$ непрерывно по ε:

$$\int [h(t)-h(t+\varepsilon)]^2\,dt=2\int h^2(t)\,dt-2\int h(t)\,h(t+\varepsilon)\,dt=2[w(0)-w(\varepsilon)]\to$$
$$\to 0 \text{ при } \varepsilon\to 0.$$

Это показывает, что $\lim\limits_{\varepsilon\to 0}\overline{[y(t)-y(t+\varepsilon)]^2}=0$. Отсюда, используя задачу 8.3, находим, что $\mathscr{R}_y$ непрерывна.

8.6. Пусть $x(t)$ — интегрируемая в квадрате функция и $y(t)$ — случайный процесс, такой, что $y(t)$ при каждом t равномерно распределена между -1 и $+1$ и $y(t)$ статистически не зависит от $y(t)$ при всех других значениях t. Желательно рассмотреть случайную величину $\int x(t)\,y(t)\,dt$, однако мы в затруднении, не зная, как понимать этот интеграл. Один из подходов интерпретировать этот интеграл в смысле Римана состоит в рассмотрении предела суммы

$$x_\Delta=\sum_{i=-\infty}^{\infty}\Delta x(i\Delta)\,y(i\Delta)$$

при Δ, стремящемся к нулю. Для всех i, $y(i\Delta)$ — независимые случайные величины с нулевыми средними и дисперсиями $1/3$. Следовательно,

$$\bar{x}_\Delta=0,\quad \overline{x_\Delta^2}=\sum_{i=-\infty}^{\infty}\Delta^2 x^2(i\Delta)\,\frac{1}{3}.$$

Вместе с тем

$$\int x^2(t)\,dt=\lim_{\Delta\to 0}\sum_{i=-\infty}^{\infty}\Delta x^2(i\Delta).$$

Следовательно, для очень малых Δ

$$\overline{x_\Delta^2}\approx\frac{\Delta}{3}\int x^2(t)\,dt,\quad \lim_{\Delta\to 0}\overline{x_\Delta^2}=0.$$

С интуитивной точки зрения процесс $y(t)$ имеет конечную мощность, которая равномерно распределена по всей области частот от $-\infty$ до ∞ и, следовательно, имеет нулевую спектральную плотность на каждой частоте. Если $z(t)$ — гауссовский процесс, то имеем

$$\int x(t)\,[z(t)+y(t)]\,dt=\int x(t)\,z(t)\,dt+\int x(t)\,y(t)\,dt.$$

Так как $\int x(t)\,y(t)\,dt$ имеет нулевое среднее значение и нулевую дисперсию, то эта случайная величина равна нулю с вероятностью 1 и, следовательно,

$$\int x(t)\,[z(t)+y(t)]\,dt=\int x(t)\,z(t)\,dt-\text{гауссовская случайная}$$

величина для каждой интегрируемой в квадрате функции $x(t)$.

Заметим, наконец, что $y(t)$ имеет корреляционную функцию

$$\mathscr{R}_y(t)=\begin{cases}1/3, & t=0,\\ 0, & t\neq 0.\end{cases}$$

Так как $y(t)$ и $z(t)$ независимы, то

$$\mathscr{R}_{y+z}(t)=\begin{cases}1/3+\mathscr{R}_z(0), & t=0,\\ \mathscr{R}_z(t), & t\neq 0.\end{cases}$$

Из этой задачи видно, что не непрерывные корреляционные функции могут привести к весьма патологическим связям между временным описанием случайного процесса (т. е. описания $z(t)$ как семейства случайных величин для произвольных множеств моментов времени) и описанием с помощью линейных функционалов (т. е. описания всех случайных величин вида $\int x(t)\, z(t)\, dt$).

8.7. (а) Соответствующее ортонормальное разложение имеет вид

$$\varphi_i(t) = \begin{cases} \dfrac{1}{\sqrt{\tau}}, & i\tau \leqslant t < (i+1)\tau, \\ 0 & \text{в других точках.} \end{cases}$$

Это разложение может быть дополнено множеством ортонормальных функций, ортогональных ко всем $\varphi_i(t)$. Вход канала ограничен тем, что имеет вид $\sum x_i \varphi_i(t)$, и, конечно, последовательность $\{x_i\}$ можно рассматривать как дискретную по времени последовательность входов. Если канал используется в течение всего временного интервала $T = N\tau$, то вход ограничен линейными комбинациями N ортонормальных функций, и пропускная способность (в натах в секунду) задается формулой (8.2.8):

$$C(\tau) = \frac{N}{2T} \ln\left(1 + \frac{2ST}{N_0 N}\right) = \frac{1}{2\tau} \ln\left(1 + \frac{2S\tau}{N_0}\right).$$

Это выражение, как и следовало ожидать, не зависит от T (если этого не было бы, то следовало бы перейти к пределу при $T \to \infty$);

$$\lim_{\tau \to 0} C(\tau) = S/N_0.$$

Это то же самое выражение, что и для пропускной способности при отсутствии каких-либо других ограничений на входе, кроме мощностных, что естественно, так как при $\tau \to 0$ фактически не остается никаких других ограничений на входную функцию, кроме мощностных.

(б) Если входы ограничены значениями $\pm\sqrt{S}$, то x_i в разложении $\sum x_i \varphi_i(t)$ ограничены значениями $\pm\sqrt{S\tau}$. Рассматривая это как дискретный по времени канал, находим, что это тот же самый канал, что и рассмотренный в задаче 4.22. Здесь ограничение на энергию x_i включает в себя $S\tau$, в то время как в задаче 4.22 было S. Здесь дисперсия шума равна $N_0/2$, в то время как тогда она была равна σ^2. Из решения задачи 4.22 вытекало, что пропускная способность в натах на интервал длительности τ секунд равна

$$C_1 = S\tau/N_0 + R(S\tau),$$

где $R(S\tau)$ стремится к нулю с τ как $\tau^{3/2}$. Следовательно, в пределе при $\tau \to 0$ пропускная способность в натах в секунду равна $C = S/N_0$. Отсюда видно, что условие, состоящее в том, что амплитуда фиксирована, не отражается на пропускной способности в пределе при сколь угодно большом числе степеней свободы.

(в) Здесь приемник интегрирует и квантует принятый сигнал на каждом интервале продолжительности τ секунд, превращая канал в ДСК с переходной вероятностью

$$\varepsilon = \int_{-\infty}^{0} \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(y - \sqrt{S\tau})^2}{N_0}\right] dy \triangleq \Phi\left(-\sqrt{\frac{2S\tau}{N_0}}\right).$$

Пропускная способность вычислена в задаче 4.22 в пределе при $\tau \to 0$, и в натах в секунду она равна $C = 2S/\pi N_0$.

Важно отметить, что из пунктов (б) и (в) следует, что ограничение, сводящееся к тому, что вход двоичный (с достаточным числом степеней свободы) не приводит к какой-либо потере пропускной способности канала, однако двухальтернативное решение на приемнике уменьшает пропускную способность. По су-

ществу, это сводится к отбрасыванию информации о надежности каждого выходного символа. Способность декодера использовать эту информацию для надежного приема каждого бита является обычно основным фактором при выборе кодирующей и декодирующей процедуры для канала с гауссовым шумом.

8.8. (а) Принятый сигнал можно переписать в виде

$$x(t)=\sum_{m=M_1}^{M_2}\left[\frac{1}{\sqrt{2}}x_m(i)\cos\varphi_m(i)\right]\sqrt{2}\cos(200\pi mt)-$$

$$-\sum_{m=M_1}^{M_2}\left[\frac{1}{\sqrt{2}}x_m(i)\sin\varphi_m(i)\right]\sqrt{2}\sin(200\pi mt).$$

Так как $x_m(i)$ и $\varphi_m(i)$ могут быть выбраны произвольно (с соблюдением ограничения на мощность), то $\frac{1}{\sqrt{2}}x_m(i)\cos\varphi_m(i)$ и $\frac{1}{\sqrt{2}}x_m(i)\sin\varphi_m(i)$ могут быть выбраны независимо, но так, чтобы они удовлетворяли ограничению на мощность. Следовательно, функция $x(t)$ на каждом интервале продолжительностью в одну секунду ограничена тем, что она является произвольной линейной комбинацией $2(M_2-M_1+1)$ ортонормальных функций. Из (8.2.8) следует

$$C=(M_2-M_1+1)\ln\left[1+\frac{S}{N_0(M_2-M_1+1)}\right]\text{ нат/с}.$$

(б) Ограничение на $x(t)$ аналогично ограничению на полосу частот, но $x(t)$ согласно пункту (а) не имеет ограниченную полосу частот. Тем не менее пропускная способность совпадает с пропускной способностью канала, имеющего выход, ограниченный полосой частот M_2-M_1+1. Это, конечно, полоса частот, на которой обычно концентрируется энергия $x(t)$.

(в) При $M_2\to\infty$ C стремится к S/N_0.

(г) Блоковый код с M кодовыми словами и длительностью блока T секунд является кодом, в котором каждое из M кодовых слов определяется некоторым выбором значений $x_m(i)$ и $\varphi_m(i)$ для всех m, $M_1\leqslant m\leqslant M_2$, и всех i, $0\leqslant i\leqslant T-1$. Ограничение на мощность заключается в том, что для каждого кодового слова

$$\sum_{i=0}^{T-1}\sum_{m=M_1}^{M_2}x_m^2(i)\leqslant ST.$$

Смысл пропускной способности состоит в том, что для любой скорости $R<C$, взяв длительность блока достаточно большой, можно найти коды со скоростью R и со сколь угодно малой вероятностью ошибки.

8.9. (а) Входная функция $x(t)=\sum_n x_n a(t-nT)$ является линейной комбинацией ортогональных функций; одна функция на 2 мкс. Следовательно, из (8.2.8) имеем

$$C=\frac{1}{4}\ln\left(1+\frac{4S}{N_0}\right)\text{ нат/мкс}.$$

Заметим, что интеграл от квадрата $a(t)$ не имеет отношения к рассматриваемой здесь задаче.

(б) Заметим, что для $T=1$ мкс справедливо $\int a(t)\,a(t-T)\,dt=0$. Следовательно, при $T=1$ мкс базисные функции ортогональны и получается та же самая ситуация, как и в пункте (а), за исключением удвоения числа степеней свободы. Следовательно,

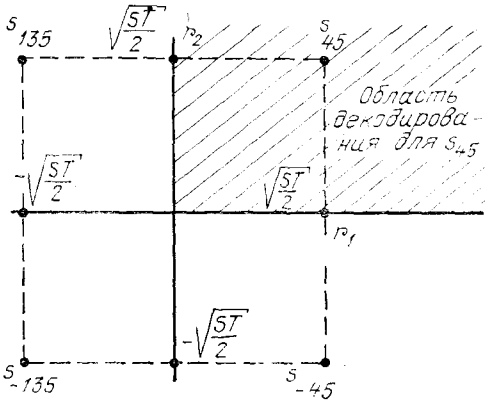
$$C=\frac{1}{2}\ln\left(1+\frac{2S}{N_0}\right)\text{ нат/мкс}.$$

8.10.(а) Пусть

$$\varphi_1(t) = \sqrt{2/T}\sin(2\pi kt/T), \quad 0 \leqslant t \leqslant T,$$
$$\varphi_2(t) = \sqrt{2/T}\cos(2\pi kt/T), \quad 0 \leqslant t \leqslant T.$$

При использовании этих ортонормальных функций возможные входы на интервале $(0, T)$ представляются в виде

$$s_{45}(t) = \sqrt{ST/2}\,\varphi_1(t) + \sqrt{ST/2}\,\varphi_2(t),$$
$$s_{135}(t) = -\sqrt{ST/2}\,\varphi_1(t) + \sqrt{ST/2}\,\varphi_2(t),$$
$$s_{-135}(t) = -\sqrt{ST/2}\,\varphi_1(t) - \sqrt{ST/2}\,\varphi_2(t),$$
$$s_{-45}(t) = \sqrt{ST/2}\,\varphi_1(t) - \sqrt{ST/2}\,\varphi_2(t).$$



(б) $r(t) = r_1\varphi_1(t) + r_2\varphi_2(t) + r_1(t)$, где $r_1(t)$ — часть шума, ортогональная к $\varphi_1(t)$ и $\varphi_2(t)$, и, следовательно, независимая от r_1, r_2 и от сообщения. Декодирование по максимуму правдоподобия заключается в нахождении (r_1, r_2) и декодировании ближайшего сообщения. Как видно из рисунка, это правило сводится к следующему:

при $r_1 \geqslant 0$, $r_2 \geqslant 0$ выбирается $\varphi = 45°$,
при $r_2 < 0$, $r_2 \geqslant 0$ выбирается $\varphi = 135°$,
при $r_1 < 0$, $r_2 < 0$ выбирается $\varphi = -135°$,
при $r_1 \geqslant 0$, $r_2 < 0$ выбирается $\varphi = -45°$.

(в) Приведенное выше правило декодирования зависит только от решения, положительно или отрицательно r_1, и решения, положительно или отрицательно r_2. По условию каждые T секунд вход представляется в виде $x_1\sqrt{ST/2}\varphi_1(t) + x_2\sqrt{ST/2}\,\varphi_2(t)$, где x_1 и x_2 равны $+1$ или -1.

При декодировании возникает ошибка, если x_1 и r_1 противоположны по знаку или, если x_2 и r_2 противоположны по знаку. Эти события независимы и имеют вероятность

$$\Pr[r_i > 0 \mid x_i < 0] = \Pr[r_i < 0 \mid x_i > 0] = \Phi\left(-\sqrt{ST/N_0}\right).$$

Величину x_1 можно рассматривать как вход некоторого ДСК, а x_2 — как вход некоторого параллельного ДСК.

(г) Пропускная способность (в битах в секунду) параллельных ДСК равна

$$C = \frac{2}{T}\left\{1 - \mathcal{H}\left[\Phi\left(-\sqrt{\frac{ST}{N_0}}\right)\right]\right\}.$$

Вычисляя предел этого выражения при $T \to 0$, получаем (см. задачу 4.22)

$$C = \frac{2S}{\pi N_0} \log_2 e \text{ бит/с}.$$

Заметим, что точно так же, как и в задаче 8.7, построение решения на основе знаков r_1 и r_2 соответствует отсутствию кодирования. Однако в системе с кодированием при использовании этого множества базисных функций величины принятых символов являются существенными для декодирования.

8.11. (а) Имеются L параллельных каналов. Если ограничить мощность в l-м канале значением S_l, $1 \leqslant l \leqslant 1$, то пропускная способность параллельного соединения (в предположении, что нет ограничений на полосу частот) равна

$$C(S_1, \ldots, S_L) = \sum_{l=1}^{L} \frac{S_l}{N_0(l)}.$$

Максимизируя это выражение при условии, что $\sum_l S_l \leqslant SL$ и условии, что $S_l \geqslant 0$, $1 \leqslant l \leqslant L$, видим, что S_l должно равняться 0 для всех l, за исключением того l, которое минимизирует $N_0(l)$; в результате получаем

$$C = SL/\min_l N_0(l).$$

(б) Пусть $\{\varphi_i(t)\}$ — множество ортонормальных функций и пусть $S_1(t) = S_2(t) = \ldots = S_L(t)$ задаются разложением $\sum_i x_i \varphi_i(t)$. Пусть функция, принимаемая на выходе l-го канала, равна $\sum_i y_{i,l}\ \varphi_i(t)$. Пусть $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,L})$. Тогда

$$p(\mathbf{y}_i \mid x_i) = \prod_{l=1}^{L} \left[\frac{1}{\sqrt{\pi N_0(l)}} \exp\left\{ -\frac{(y_{i,l} - x_i)^2}{N_0(l)} \right\} \right] =$$

$$= \prod_{l=1}^{L} \left[\frac{1}{\sqrt{\pi N_0(l)}} \right] \exp \sum_{l=1}^{L} \left[-\frac{y_{i,l}^2}{N_0(l)} + \frac{2x_i y_{i,l}}{N_0(l)} - \frac{x_i^2}{N_0(l)} \right].$$

Отсюда видно, что знание $\sum_l \frac{y_{i,l}}{N_0(l)}$ делает возможным вычисление $p(\mathbf{y}_i \mid x_i)/p(\mathbf{y}_i \mid x_i')$ при любых заданных x_i и x_i' без знания отдельных значений $y_{i,l}$. Следовательно, выходы канала могут быть умножены на весовые множители $1/N_0(l)$ и сложены вместе без потери пропускной способности. После такого взвешивания имеем

$$y(t) = \sum_{l=1}^{L} \frac{x(t)}{N_0(l)} + \sum_{l=1}^{L} \frac{z_l(t)}{N_0(l)}.$$

Мощность сигнала теперь равна $S \left[\sum_{l=1}^{L} \frac{1}{N_0(l)} \right]^2$. Спектральная плотность шума равна $\sum_{l=1}^{L} \frac{1}{2N_0(l)}$. Отсюда следует, что

$$C = S \sum_{l=1}^{L} \frac{1}{N_0(l)} \text{ нат/с}.$$

(в) Пусть $s(t)$ — переданная функция и $y(t)$ — сумма принятых функций

$$y(t) = Ls(t) + \sum_{l=1}^{L} z_l(t).$$

Мощность сигнала равна $L^2 S$ и спектральная плотность шума равна $\sum_l \frac{N_0(l)}{2}$. Следовательно,

$$C = L^2 S \Big/ \sum_{l=1}^{L} N_0(l) \text{ нат/с}.$$

8.12. Непосредственный подход к этой задаче заключается в задании двух кодовых слов с помощью ортонормального разложения для L параллельных каналов, нахождении вероятности ошибки и затем оптимизации по кодовым словам. При более простом подходе следует заметить, что каждый из L каналов можно рассматривать как бесконечное множество гауссовских дискретных по времени каналов. Вероятность ошибки не может возрастать, если дисперсии шума в каждом множестве дискретных каналов уменьшаются до дисперсий для множества дискретных каналов с наименьшими дисперсиями. Это, однако, было бы равносильно использованию только лучшего множества каналов. Тогда из (8.2.24) получаем

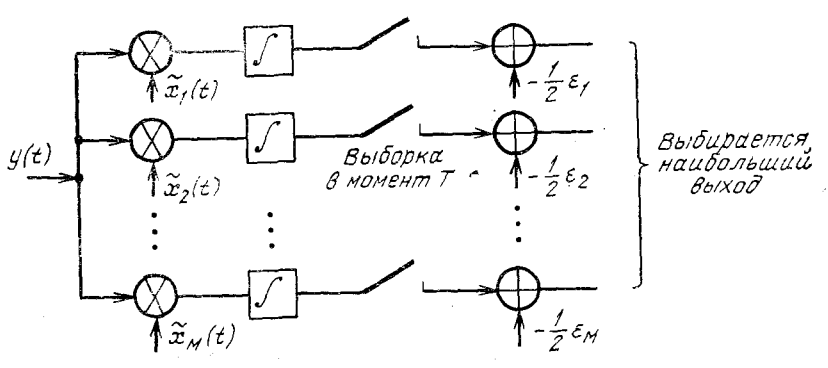
$$P_e = \Phi\left(-\sqrt{\frac{2SLT}{\min_l N_0(l)}}\right) = \Phi\left(-\sqrt{2CT}\right). \tag{1}$$

Далее рассмотрим ситуацию, в которой один и и тот же сигнал должен быть послан по всем каналам. Соображение задачи 8.11, что оптимальный приемник может умножать l-ю функцию на весовой множитель $1/N_0(l)$ и суммировать по l, остается все еще в силе. Это позволяет свести канал к одному каналу с белым гауссовым шумом; при этом все еще справедливо (1), где C теперь — пропускная способность канала задачи 8.11.(б). Аналогично (1) остается справедливым, когда L выходов канала суммируются без взвешивания, однако C будет пропускной способностью канала задачи 8.11.(в).

8.13. (а) Пусть для любого m, $1 \leqslant m \leqslant M$,

$$\tilde{x}_m(t) = \sum_{i=1}^{4} a_i x_m(t + \tau - i\tau).$$

Если $\tilde{x}_m(t)$, $1 \leqslant m \leqslant M$, рассматривать как кодовые слова, то получается канал с аддитивным гауссовым белым шумом. Из (8.2.18) видно, что декодер максимума правдоподобия выбирает m, которое максимизирует $\int y(t)\,\tilde{x}_m(t)\,dt - {}^1/_2 \int \tilde{x}_m^2(t)dt$. Это соответствует следующей схеме, изображенной на рисунке.



(б). Пусть $\tilde{x}(t) = \sum_{i=1}^{4} a_i x(t+\tau-i\tau)$. Тогда

$$\int_{-\infty}^{\infty} \tilde{x}^2(t)\,dt = \sum_{i=1}^{4}\sum_{j=1}^{4} a_i a_j \int_{-\infty}^{\infty} x(t+\tau-i\tau)\,x(t+\tau-j\tau)\,dt.$$

Используя неравенство Шварца и то, что $a_i > 0$, имеем

$$\int \tilde{x}^2(t)\,dt \leqslant \sum_{i=1}^{4}\sum_{i=1}^{4} a_i a_j \int x^2(t)\,dt = \left(\sum_{i=1}^{4} a_i\right)^2 \int x^2(t)\,dt.$$

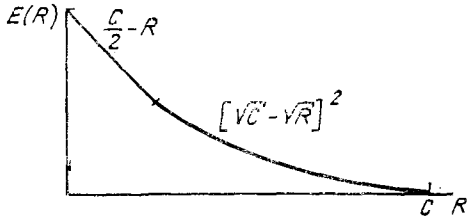
Таким образом, ограничение на мощность S входов линии задержки приводит к ограничению на мощность $(\sum_i a_i)^2$ выходов линии задержки. Следовательно,

$$C \leqslant \left(\sum_{i=1}^{4} a_i\right)^2 S \Big/ N_0.$$

Вместе с тем синусоиды сколь угодно большой длительности с частотами, кратными $1/\tau$, подходят сколь угодно близко к удовлетворению этого мощностного ограничения. Следовательно, при достаточно больших T можно конструировать коды с произвольно большим числом ортогональных кодовых слов, каждое из которых имеет энергию, сколь угодно близкую к $(\sum a_i)^2 ST$, и

$$C = \left(\sum_{i=1}^{4} a_i\right)^2 S \Big/ N_0.$$

(в) Применяя (8.2.43) и (8.2.44) к ортогональным кодам, рассмотренным выше, получаем результат, изображенный на рисунке.



8.14. (а) Сначала выберем масштаб на входе и выходе канала так, чтобы шум имел единичную спектральную плотность. Расписывая $p(\mathbf{y} \mid \mathbf{x}_m)$ для каждого m, как в (8.2.30) и (8.2.31), видим, что декодирование по методу максимума правдоподобия сводится к следующему: вычисляется $y_m = \int y(t)\,\varphi_m(t)\,dt$ для всех m, $1 \leqslant m \leqslant M/2$; выбирается m, для которого $|y_m|$ наибольшее, и декодируется сообщение m, если $y \geqslant 0$, и декодируется сообщение $m + M/2$, если $y_m < 0$.

Предположим теперь, что передано сообщение $m (m \leqslant M/2)$ и принято некоторое значение y_m. Если $y_m < 0$, то определенно будет сделана ошибка, а если $y_m > 0$, то будет сделана ошибка, если $|y_{m'}| \geqslant y_m$ для некоторого $m' \neq m$. Вероятность этого при заданном y_m равна

$$Q(y_m) = 1 + [\Phi(y_m)\Phi(y_m)]^{M/2-1} =$$
$$= 1 - [1 - 2\Phi(-y_m)]^{M/2-1}, \qquad y_m > 0.$$

Аддитивная граница дает

$$Q(y_m) \leqslant (M-2)\,\Phi(-y_m), \quad y_m > 0.$$

Определяя y_0 из $M\exp(-y_0^2/2)=1$, можно верхнюю границу $Q(y_m)$ представить в виде

$$Q(y_m) \leqslant \begin{cases} (M-1)\,\Phi(-y_m), & y_m > y_0, \\ 1, & y_m \leqslant y_0, \end{cases}$$

которая учитывает то, что $Q_m(y) = 1$ для $y_m < 0$. Это та же граница, что и верхняя граница $Q(y_m)$, использованная в (8.2.37); следовательно, как и ранее, получаем (8.2.43). Из симметрии вытекает, что эта же граница применима для $m > M/2$.

Заметим, что при низких скоростях вероятность того, что любое заданное m' (отличное от $m' = m + M/2$) более правдоподобно, чем переданное сообщение m, равна $\Phi(-A/\sqrt{2})$, где $A = \sqrt{2\mathscr{E}/N_0}$. Вероятность того, что $m' = m + M/2$ более правдоподобно, чем m, равна $\Phi(-A) < \Phi(-A/\sqrt{2})$. Затем, используя аддитивную границу, получаем $P_{e,\,m} \leqslant (M-1)\,\Phi(-A/\sqrt{2})$. Это та же граница, как и в (8.2.34), и отсюда (8.2.44) следует, как и ранее.

(б) Откажемся здесь от нормировки, использованной в пункте (а), и примем опять, что спектральная плотность шума равна $N_0/2$. Предположим, что передано сообщение m, $m \leqslant M/2$.

Пусть E_1 — событие, заключающееся в том, что $y_m < A$, и E_2 — событие, заключающееся в том, что для некоторого $m' \neq m$, $|y_{m'}| \geqslant A$. Тогда

$$\Pr[E_1] = \int_{-\infty}^{A} \frac{1}{\sqrt{\pi N_0}} \exp\left[-\frac{(y-\sqrt{\mathscr{E}})^2}{N_0}\right] dy \triangleq P_1, \tag{1}$$

$$\Pr[E_2] = 1 - \left[\int_{-A}^{A} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{y^2}{N_0}\right) dy\right]^{M/2-1} \leqslant$$

$$\leqslant \left(\frac{M}{2}-1\right)\left[1 - \int_{-A}^{A} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{y^2}{N_0}\right) dy\right] =$$

$$= \left(\frac{M}{2}-1\right) 2 \int_{A}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{y^2}{N_0}\right) dy \leqslant \tag{2}$$

$$\leqslant M \int_{A}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{y^2}{N_0}\right) dy \triangleq P_2. \tag{3}$$

Декодер всегда декодирует правильно, если ни E_1, ни E_2 не наступят. Следовательно, P_c — вероятность правильного декодирования удовлетворяет неравенству

$$P_c \geqslant 1 - P_1 - P_2.$$

Так как $1 - P_c = P_a + P_e$, то

$$P_a + P_e \leqslant P_1 + P_2, \quad P_a \leqslant P_1 + P_2.$$

Ошибка произойдет, если $|y_{m'}| \geqslant A$ точно для одного значения $m' \neq m$ и $|y_m| < A$ или если $y_m < -A$ и $|y_{m'}| < A$ для всех $m' \neq m$. Так как y_m не зависит от $y_{m'}$ для всех $m' \neq m$, то

$$P_e \leqslant \Pr[E_1]\Pr[E_2] + \Pr[y_m < -A].$$

Используя (2) для $\Pr(E_2)$, получаем

$$P_e \leqslant P_1 (M-2) \int_{A}^{\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{y^2}{N_0}\right) dy +$$

$$+\int_{-\infty}^{-A}\frac{1}{\sqrt{\pi N_0}}\exp\left[-\frac{(y-\sqrt{\mathscr{E}})^2}{N_0}\right]dy=(M-2)\times$$

$$\times\Phi\left[-\frac{\sqrt{\mathscr{E}}-A}{\sqrt{N_0/2}}\right]\Phi\left[-\frac{A}{\sqrt{N_0/2}}\right]+\Phi\left(-\frac{A+\sqrt{\mathscr{E}}}{\sqrt{N_0/2}}\right).$$

Если A и $\sqrt{\mathscr{E}}-A$ велики по сравнению с $\sqrt{N_0/2}$, то

$$\Phi\left[-\frac{A+\sqrt{\mathscr{E}}}{\sqrt{N_0/2}}\right]\approx\frac{\sqrt{N_0/2}}{(A+\sqrt{\mathscr{E}})\sqrt{2\pi}}\exp\left[-\frac{(A+\sqrt{\mathscr{E}})^2}{N_0}\right]\ll$$

$$\ll\Phi\left(-\frac{\sqrt{\mathscr{E}}-A}{\sqrt{N_0/2}}\right)\Phi\left(-\frac{A}{\sqrt{N_0/2}}\right).$$

Следовательно, в этом случае (который только и представляет интерес)

$$P_e\leqslant M\Phi\left(-\frac{\sqrt{\mathscr{E}}-A}{\sqrt{N_0/2}}\right)\Phi\left(-\frac{A}{\sqrt{N_0/2}}\right)=P_1P_2.$$

(в). Дифференцируя P_1+P_2 по A, можно выбрать A, минимизирующее границу P_a (т. е. P_1+P_2). Получаем

$$\frac{1}{\sqrt{\pi N_0}}\exp\left[-\frac{(A-\sqrt{\mathscr{E}})^2}{N_0}\right]=M\frac{1}{\sqrt{\pi N_0}}\exp\left(-\frac{A^2}{N_0}\right),$$

$$\frac{2\sqrt{\mathscr{E}}\,A}{N_0}-\frac{\mathscr{E}}{N_0}=\ln M,\quad A=\frac{N_0}{2\sqrt{\mathscr{E}}}\ln M+\frac{\sqrt{\mathscr{E}}}{2}.$$

Полагая, что T — длительность сигналов, $R=(\ln M)/T$, $S=\mathscr{E}/T$ и $C=S/N_0$, приведем это выражение к виду

$$A=\frac{\sqrt{N_0CT}}{2}\left(\frac{R}{C}+1\right).$$

Тогда из (1) имеем

$$P_1=\Phi\left(-\frac{\sqrt{\mathscr{E}}-A}{\sqrt{N_0/2}}\right)=\Phi\left[-\sqrt{\frac{T}{2C}}(C-R)\right]\leqslant$$

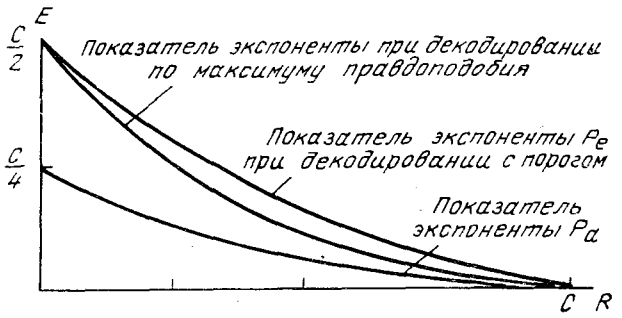
$$\leqslant\sqrt{\frac{C}{\pi T(C-R)^2}}\exp\left[-\frac{T}{4C}(C-R)^2\right],$$

$$P_2=e^{RT}\Phi\left(-\frac{A}{\sqrt{N_0/2}}\right)=e^{RT}\Phi\left[-\sqrt{\frac{T}{2C}}(C+R)\right]\leqslant$$

$$\leqslant\sqrt{\frac{C}{\pi T(C+R)^2}}\exp\left[-\frac{T}{4C}(C-R)^2\right],$$

$$P_a\leqslant\sqrt{\frac{C}{\pi T}}\left(\frac{1}{C-R}+\frac{1}{C+R}\right)\exp\left[-\frac{T}{4C}(C-R)^2\right],$$

$$P_e\leqslant P_1P_2\leqslant\frac{C}{\pi T(C^2-R^2)^2}\exp\left[-\frac{T}{2C}(C-R)^2\right].$$



(г) Для $A = \sqrt{\mathscr{E}} - \varepsilon$

$$P_1 \leqslant \sqrt{\frac{N_0}{\pi\varepsilon^2}} \exp\left[-\frac{\varepsilon^2}{N_0}\right],$$

$$P_2 \leqslant \sqrt{\frac{N_0}{\pi\left(\sqrt{ST}-\varepsilon\right)^2}} \exp\left[-T\left(C-R-2\varepsilon\sqrt{\frac{C}{N_0 T}}+\frac{\varepsilon^2}{N_0 T}\right)\right].$$

При этом выборе P_a зависит от T неэкспоненциально, однако показатель экспоненты P_e равен $C - R$. Здесь наблюдается то же явление, что и указанное в задаче 5.14.

8.15. (а) Пусть $x_1(t)$, ..., $x_M(t)$ — множество ортогональных кодовых слов с энергиями $2E'/N_0$ (здесь опять выбирается масштаб амплитуды, нормирующий шум). Пусть

$$\xi_m(t) = x_m(t) - \frac{1}{M}\sum_{m'=1}^{M} x_{m'}(t)$$

— соответствующие слова симплексного кода. Имеем

$$\int \xi_m^2(t)\,dt = \int x_m^2(t)\,dt - \frac{2}{M}\sum_{m'=1}^{M}\int x_m(t)\,x_{m'}(t)\,dt +$$

$$+\frac{1}{M^2}\sum_{m',\,m''}\int x_{m'}(t)\,x_{m''}(t)\,dt = \frac{2E'}{N_0}\left[1-\frac{2}{M}+\frac{1}{M}\right] = \frac{2E'}{N_0}\left(\frac{M-1}{M}\right).$$

Если определить $2E/N_0$ как энергию $\xi_m(t)$ для каждого m, то

$$E = E'\left(\frac{M-1}{M}\right).$$

(б) Квадрат расстояния между любыми двумя кодовыми словами в симплексном коде задается равенством

$$\int [\xi_m(t) - \xi_k(t)]^2\,dt = \int [x_m(t) - x_k(t)]^2\,dt =$$

$$= \frac{4E'}{N_0} = \frac{4EM}{N_0(M-1)} \quad \text{для всех} \quad m,\ k,\ m \neq k.$$

Это выражение совпадает с верхней границей среднеквадратического расстояния, данного в (8.2.27), и, следовательно, кодовые слова с ограничением на энергию $2E/N_0$ не могут иметь большего минимального расстояния.

(б) Напомним, что в двоичных кодах максимальной длины каждая пара кодовых слов совпадает в $(N-1)/2$ позициях и не совпадает в $(N+1)/2$ позициях. Следовательно,

$$\int [x_m(t) - x_{m'}(t)]^2\,dt = \sum_{n=1}^{N} (2x_{m,n} - 2x_{m',n})^2 = 4\left(\frac{N+1}{2}\right) = 2M$$

для всех $m \neq m'$,

где $M = N+1$ — число кодовых слов. Очевидно, энергия любого кодового слова равна $\int x_m^2(t)\,dt = N = M - 1$. Отсюда видно, что отношение энергии кодового слова к квадрату расстояния равно $(M-1)/2M$, что согласуется с результатом пункта (а) для симплексного кода.

Так как симплексные коды определяются здесь с помощью ассоциированного ортогонального кода, то следует рассмотреть ассоциированный ортогональный код. Пусть $\varphi_M(t)$ — нормированная функция, ортогональная к $\varphi_1(t), \ldots, \varphi_N(t)$. Определим $y_m(t) = x_m(t) + \varphi_M(t)$. Тогда

$$\int y_m(t)\,y_{m'}(t)\,dt = \left[\sum_{n=1}^{N} (2x_{m,n} - 1)(2x_{m',n} - 1)\right] + 1.$$

Указанное выше произведение равно -1 для $(N+1)/2$ слагаемых и $+1$ для оставшихся слагаемых; следовательно, кодовые слова $y_m(t)$ ортогональны и имеют энергию M. Наконец, так как $\sum_m x_m(t) = 0$, то $\varphi_M(t) = 1/M \sum_m y_m(t)$ и, в смысле определения (8.2.28), кодовые слова $x_m(t)$ образуют симплексное множество.

8.16. Когда буква k поступает в модулятор, $\sqrt{ST}\,\varphi_k(t)$ передается, и выход i-го согласованного фильтра приемника равен

$$y_i = \int y(t)\,\varphi_i(t)\,dt = \begin{cases} \sqrt{ST_0} + z_i, & i = k, \\ z_i, & i \neq k, \end{cases}$$

где $z_0, \ldots, z_{K-1}$ — независимые нормированные гауссовские случайные величины. Следовательно,

$$p_{\mathbf{Y}}(y_0, \ldots, y_{K-1} \mid x = k) = \prod_{i=0}^{K-1} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_i - \delta_{ik}\sqrt{ST})^2}{2}\right] =$$

$$= p_{\mathbf{Z}}(\mathrm{y}) \exp[y_k \sqrt{ST_0} - ST_0/2],$$

где $p_{\mathbf{Z}}(\mathrm{y}) = \prod_{i=0}^{K-1} \frac{1}{\sqrt{2\pi}} \exp[-y_i^2/2]$ — совместная плотность K нормированных независимых гауссовских случайных величин.

Имеем

$$E_0(1, \mathrm{Q}) = -\ln \int_{\mathrm{y}} \left\{\sum_{k=0}^{K-1} Q(k) \sqrt{p_{\mathbf{Y}}(\mathrm{y} \mid x = k)}\right\}^2 d\mathrm{y},$$

$$E_0(1, \mathrm{Q}) = -\ln \int \left\{\sum_{k=0}^{K-1} \frac{1}{K} \sqrt{p_{\mathbf{Z}}(\mathrm{y})} \exp\left[\frac{y_k \sqrt{ST_0}}{2} - \frac{ST_0}{4}\right]\right\}^2 d\mathrm{y}.$$

Производя возведение в квадрат суммы в скобках, получаем

$$E_0(1, \mathrm{Q}) = -\ln \int \sum_{i,k} \frac{1}{K^2} p_{\mathbf{Z}}(\mathrm{y}) \exp\left[\frac{y_k \sqrt{ST_0}}{2} + y_i \frac{\sqrt{ST_0}}{2} - \frac{ST_0}{2}\right] d\mathrm{y}.$$

Можно изменить порядок суммирования и интегрирования и проинтегрировать слагаемое за слагаемым. Для $i \neq k$ можно непосредственно проинтегрировать сразу все слагаемые, получаем

$$\int p_{\mathbf{Z}}(\mathbf{y}) \exp\left[\frac{y_k\sqrt{ST_0}}{2} + \frac{y_i\sqrt{ST_0}}{2} - \frac{ST_0}{2}\right] d\mathbf{y} =$$

$$= \int\int \frac{1}{2\pi} \exp\left[-\frac{y_k^2}{2} + \frac{y_k\sqrt{ST_0}}{2} - \frac{y_i^2}{2} + \frac{y_i\sqrt{ST_0}}{2} - \frac{ST_0}{2}\right] dy_k\, dy_i =$$

$$= \exp\left[-\frac{ST_0}{4}\right].$$

Этот же интеграл для $i = k$ равен 1. Следовательно,

$$E_0(1, \mathbf{Q}) = -\ln\left[\frac{K(K-1)}{K^2} e^{-ST_0/4} + \frac{K}{K^2}\right],$$

$$E_0(1, \mathbf{Q}) = \frac{ST_0}{4} - \ln\left[1 + \frac{1}{K}\left(e^{ST_0/4} - 1\right)\right].$$

Из теоремы кодирования вытекает, что существуют коды со скоростью R_1 нат на символ и любой длиной блока N, для которых

$$P_e \leqslant \exp\{-N[E_0(1, \mathbf{Q}) - R_1]\}.$$

Обозначая через $T = NT_0$ — длительность блока, а через $R = R_1/T_0$ — скорость в нат/с, эта граница приводится к виду

$$P_e \leqslant \exp\left[TR - \frac{TS}{4}\left\{1 - \frac{4}{ST_0}\ln\left(1 + \frac{1}{K}\left(e^{ST_0/4} - 1\right)\right)\right\}\right].$$

Когда K возрастает, выражение в фигурных скобках сходится к 1 и для $R < S/8$ это выражение является показателем экспоненты для ортогонального кода (напомним, что мы положили $N_0/2 = 1$, так что $S/8 = C_\infty/4$). Для конечного K член в фигурных скобках может быть истолкован как эффективность η метода дискретной модуляции данных в том смысле, что она определяет ту долю показателя экспоненты ортогонального кода при нулевой скорости, которая достигается при заданном значении K. Эффективность η достигается при

$$K = \frac{e^{ST_0/4} - 1}{e^{(1-\eta)\,ST_0/4} - 1}.$$

При $ST_0/4 \ll 1$ это выражение сводится к $K \approx 1/(1-\eta)$. При $ST_0/4 \gg 1$ имеем $K \approx e^{\eta ST_0/4}$.

8.17. Для $y_m \geqslant 0$ имеем

$$Q(y_m) \leqslant (M-1)^\rho \exp[-y_m^2 \rho/2], \quad 0 \leqslant \rho \leqslant 1.$$

Для $y_m < 0$, $Q(y_m) < 1$. Подставляя эти границы в (8.2.32), получаем

$$P_{e,m} \leqslant \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y_m - A)^2}{2}\right] dy_m +$$

$$+ \int_0^\infty \frac{(M-1)^\rho}{\sqrt{2\pi}} \exp\left[-\frac{(y_m - A)^2}{2} - \frac{y_m^2 \rho}{2}\right] dy_m.$$

Первое слагаемое приведенного выше выражения можно оценить сверху, раскрывая квадрат и пренебрегая перекрестным членом. Второе слагаемое

можно оценить сверху, распространяя нижний предел интегрирования до $-\infty$ и дополняя затем показатель до полного квадрата. Получаем

$$P_{e,\,m} \leqslant \frac{1}{2}\,e^{-A^2/2} + \frac{(M-1)^{\rho}}{\sqrt{1+\rho}} \exp\left[-\frac{\rho}{1+\rho}\,\frac{A^2}{2}\right].$$

Оценивая сверху $(1+\rho)^{-1/2}$ единицей, первое слагаемое — половиной второго, $M-1$ — с помощью e^{TR} и беря $A = \sqrt{2TC_{\infty}}$, это неравенство приводим к виду

$$P_{e,\,m} \leqslant {}^{3}/_{2} \exp\left[-\frac{\rho}{1+\rho}\,TC_{\infty} + \rho RT\right].$$

Минимум по ρ, $0 \leqslant \rho \leqslant 1$, имеет место при $(1+\rho)^2 = C_{\infty}/R$, $C_{\infty}/4 \leqslant R \leqslant C_{\infty}$. Для $R < C_{\infty}/4$ минимум имеет место при $\rho = 1$. Следовательно,

$$P_{e,m} \leqslant {}^{3}/_{2} \exp\left[-T\left(\sqrt{C_{\infty}} - \sqrt{R}\right)^2\right];\quad C_{\infty}/4 \leqslant R \leqslant C_{\infty},$$

$$P_{e,m} \leqslant {}^{3}/_{2} \exp\left[-T\left(C_{\infty}/2 - R\right)\right];\quad R < C_{\infty}/4.$$

8.18.(а) При условии, что послано сообщение m, ошибка произойдет, если $r_{m'} \geqslant r_m$ при каком-либо $m' \neq m$. Следовательно,

$$P_{e,m} = \int dr_m\, p_{r_m}(r_m \mid m)\, \Pr\left[r_{m'} \geqslant r_m,\ \text{при каком-либо } m' \neq m \mid m,\, r_m\right]. \tag{1}$$

При заданном m и любых $m' \neq m$ случайные величины $r_{m'}$ одинаково распределены и, следовательно, используя (5. 6. 2), имеем

$$\Pr\left[r_{m'} \geqslant r_m \text{ при каком-либо } m' \neq m \mid m,\, r_m\right] \leqslant$$

$$\leqslant (M-1)^{\rho} \Pr\left[r_{m'} \geqslant r_m \mid m, r_m\right]^{\rho},$$

где m' в правой части фиксированное, однако произвольное целое число, не равное m, и $0 \leqslant \rho \leqslant 1$. Подставляя это в (1) и используя $r_m = y_{m,1}^2 + y_{m,2}^2$, получаем

$$P_{e,m} \leqslant (M-1)^{\rho} \iint p\,(y_{m,1},\, y_{m,2} \mid m)\, \Pr\left[r_{m'} \geqslant r_m \mid m,\, r_m\right]^{\rho} dy_{m,1}\, dy_{m,2}. \tag{2}$$

(б) При условии, что передано m, $r_{m'} = y_{m',1}^2 + y_{m',2}^2$ — сумма квадратов двух независимых гауссовских случайных величин с дисперсиями $N_0/2$. Следовательно,

$$\Pr\left[r \leqslant r_{m'} \leqslant r + \Delta\right] = \iint dy_{m',1}\, dy_{m',2} \times$$

$$\times\ \frac{1}{\pi N_0} \exp\left[\frac{-y_{m',1}^2 - y_{m',2}^2}{N_0}\right],$$

где интеграл берется по области $r \leqslant y_{m',1}^2 + y_{m',2}^2 \leqslant r + \Delta$. Для малых Δ площадь области интегрирования равна $\pi(r+\Delta) - \pi r$ и подынтегральное выражение постоянно (с точностью до первого порядка по Δ). Следовательно, плотность $r_{m'}$ равна

$$p\left(r_{m'} \mid m \neq m'\right) = \frac{1}{N_0} \exp\left[-r_{m'}/N_0\right], \tag{3}$$

$$\Pr\left[r_{m'} \geqslant r_m \mid m,\, r_m\right] = \exp\left[-\frac{r_m}{N_0}\right].$$

При условии, что фаза принятого сообщения θ равна 0, видим, что $y_{m,1} = \sqrt{ST} + z_{m,1}$ и $y_{m,2} = z_{m,2}$, где $z_{m,1}$ и $z_{m,2}$ — независимые гауссовские случайные величины с нулевыми средними и дисперсиями $N_0/2$. Следовательно,

$$p(y_{m,1}, y_{m,2} \mid \theta = 0) = \frac{1}{\pi N_0} \exp\left[-\frac{(y_{m,1} - \sqrt{ST})^2}{N_0} - \frac{y_{m,2}^2}{N_0} \right]. \tag{4}$$

Подставляя (3) и (4) в (2) (которое теперь рассматривается при условии $\theta = 0$) и вспоминая, что $r_m = y_{m,1}^2 + y_{m,2}^2$, имеем

$$(M-1)^\rho \int\int p(y_{m,1}, y_{m,2} \mid \theta = 0) \times$$

$$\times \Pr[r_{m'} \geqslant r_m \mid m, r_m]^\rho \, dy_{m,1} \, dy_{m,2} = (M-1)^\rho \iint \frac{1}{\pi N_0} \times$$

$$\times \exp\left[-\frac{(y_{m,1} - \sqrt{ST})^2}{N_0} - \frac{y_{m,2}^2}{N_0} - \frac{(y_{m,1}^2 + y_{m,2}^2)\,\rho}{N_0} \right] dy_{m,1} \, dy_{m,2} =$$

$$= \frac{(M-1)^\rho}{1+\rho} \exp\left[-\frac{ST}{N_0} \frac{\rho}{1+\rho} \right], \tag{5}$$

где интеграл вычисляется обычным методом дополнения до полного квадрата.

Наконец, заметим, что приведенное выше подынтегральное выражение как функция, определенная на плоскости $y_{m,1}$, $y_{m,2}$, равно произведению гауссовской функции с максимумом в точке $(\sqrt{ST}, 0)$ на сферически симметричную экспоненциальную функцию с центром в начале координат. Изменение в задании угла θ будет поворачивать центр гауссовской функции, приводя к тому же результату, что и поворот осей. Следовательно, интеграл не зависит от θ.

(в) Для $M = 2$ имеется только одно сообщение m', отличное от переданного m. Поэтому равенство (1) принимает вид

$$P_{e,m} = \int dr_m \, p_{r_m}(r_m \mid m) \Pr[r_{m'} \geqslant r_m \mid m, r_m],$$

где $m' \neq m$. Это выражение такое же, как и правая часть (2) для $\rho = 1$ и $M = 2$, а так как с того места не использовались операции оценивания, то $P_{e,m} = 1/2 \exp[-ST/2N_0]$ для $M = 2$. Для произвольного M

$$P_{e,m} \leqslant \frac{(M-1)^\rho}{(1+\rho)} \exp\left[-\frac{ST}{N_0} \frac{\rho}{1+\rho} \right] \leqslant$$

$$\leqslant M^\rho \exp\left[-\frac{ST}{N_0} \left(\frac{\rho}{1+\rho} \right) \right] = \exp\left[-T \left(C \frac{\rho}{1+\rho} - \rho R \right) \right],$$

где $R = \ln M/T$ и $C = S/N_0$. Это выражение минимизируется по ρ, $0 \leqslant \rho \leqslant 1$, при $(1+\rho)^2 = C/R$ для $C/4 \leqslant R \leqslant C$ и при $\rho = 1$ для $R < C/4$. Имеем

$$P_{e,m} \leqslant \begin{cases} \exp[-T(\sqrt{C} - \sqrt{R})^2], & C/4 \leqslant R \leqslant C, \\ \exp[-T(C/2 - R)], & R < C/4. \end{cases}$$

8.19. Сначала заметим, что из (8.5.86) — (8.5.88) следует, что $\widetilde{S}_\infty(B, \rho)$ $\widetilde{R}_\infty(B)$ и $\widetilde{E}_\infty(B, \rho)$ — непрерывные функции B и ρ. Следовательно, для любых заданных B и ρ и любого заданного $\varepsilon_1 > 0$ существуют $\delta_1 > 0$ и $\delta_3 > 0$, такие что

$$|\widetilde{E}_\infty(B_1, \rho_1) - \widetilde{E}_\infty(B, \rho)| \leqslant \varepsilon_1 \text{ для } |\rho_1 - \rho| \leqslant \delta_1, |B_1 - B| \leqslant \delta_3. \tag{1}$$

Аналогично, так как $\widetilde{S}_\infty\ (B,\ \rho)$ строго возрастает с $\rho(\rho\ >\ 0)$, то имеем

$$\widetilde{S}_\infty\,(B,\ \rho-\delta_1)<\widetilde{S}_\infty(B,\ \rho).$$

Следовательно, существует некоторое $\delta_2\ >\ 0$, удовлетворяющее неравенству $\delta_2\ <\ \delta_3$, такое, что

$$\widetilde{S}_\infty\,(B+\delta_2,\ \rho-\delta_1)<\widetilde{S}_\infty\,(B,\ \rho). \tag{2}$$

Тогда из (1) имеем

$$\widetilde{E}_\infty\,(B+\delta_2,\ \rho-\delta_1)\geqslant\widetilde{E}_\infty\,(B,\ \rho)-\varepsilon_1. \tag{3}$$

Наконец, так как $\widetilde{R}_\infty\,(B)$ возрастает с B, то $\widetilde{R}_\infty\,(B+\delta_2)\geqslant\widetilde{R}_\infty\,(B)$.

Так как $\widetilde{R}_\infty$, $\widetilde{E}_\infty$ и $\widetilde{S}_\infty$ все сходятся к пределу при $T\to\infty$, то можно найти достаточно большое T, так что

$$\widetilde{R}_T\,(B+\delta_2)\geqslant\widetilde{R}_\infty\,(B),$$
$$\widetilde{S}_T\,(B+\delta_2,\ \rho-\delta_1)\leqslant\widetilde{S}_\infty\,(B,\ \rho),$$
$$\widetilde{E}_T\,(B+\delta_2,\ \rho-\delta_1)\geqslant\widetilde{E}_\infty\,(B,\ \rho)-2\varepsilon_1.$$

Следовательно, из (8.5.85) при использовании $B\ +\ \delta_2$ вместо B и $\rho\ -\ \delta_1$ вместо ρ вытекает (8.5.89) для заданных B, ρ, произвольного положительного числа $\varepsilon\ =\ 2\ \varepsilon_1$ и достаточно большого T.

8.20. Для фиксированного R рассмотрим показатель экспоненты вероятности ошибки как функцию мощностного ограничения. При возрастании мощности, начиная с точки, где R равна пропускной способности, сначала следует рассмотреть показатель экспоненты случайного кодирования, затем прямолинейный участок показателя экспоненты и, наконец, показатель экспоненты для процедуры с выбрасыванием. Для показателя экспоненты случайного кодирования имеем равенства (8.5.86) — (8.5.88). При фиксированном R значение B также фиксировано, поскольку оно удовлетворяет равенству $\widetilde{R}_\infty\ (B)\ =\ R$. Из (8.5.88) видно, что $\widetilde{E}_\infty$ можно рассматривать как функцию $\widetilde{S}_\infty$ и ρ. Имеем

$$\frac{\partial\widetilde{E}_\infty}{\partial\widetilde{S}_\infty}=\frac{d\widetilde{E}_\infty}{dS_\infty}\bigg|_\rho+\frac{\partial\widetilde{E}_\infty}{\partial\rho}\Big/\frac{\partial\widetilde{S}_\infty}{\partial\rho},$$

где

$$\widetilde{E}_\infty=\frac{\rho\widetilde{S}_\infty}{2B\,(1+\rho)}-\int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)}\geqslant\frac{1}{B}}{}^{1}\!/_{2}\ln\left[1+\rho-\frac{\rho N(f)}{B\,|H_1(f)|^2}\right]df,$$

$$\frac{\partial\widetilde{E}_\infty}{\partial\rho}=\frac{\widetilde{S}_\infty}{2B\,(1+\rho)^2}-\int{}^{1}\!/_{2}\,\frac{B\,|H_1(f)|^2-N(f)}{(1+\rho)\,B\,|H_1(f)|^2-\rho N(f)}\,df=0,$$

где было использовано определение $\widetilde{S}_\infty$ из (8.5.86). Следовательно,

$$\frac{d\widetilde{E}_\infty}{d\,\widetilde{S}_\infty}=\frac{\partial\widetilde{E}_\infty}{\partial\widetilde{S}_\infty}\bigg|_\rho=\frac{\rho}{2B\,(1+\rho)}. \tag{1}$$

Этот результат справедлив при изменении мощности $\widetilde{S}_\infty$ в области от мощности, при которой $R\ =\ C$ [задаваемой $\widetilde{S}_\infty\ (B,\ 0)$ в (8.5.86)] до мощности $\widetilde{S}_\infty\ (B,\ 1)$, определяемой из (8.5.86). Заметим, что наклон равен нулю при $\widetilde{S}_\infty\ (B,\ 0)$, указывая на непрерывность при меньших значениях S, для которых $E\ =\ 0$. Для $S\ >\ \widetilde{S}_\infty\ (B,\ 1)$ используем (8.5.94) для определения B_{cr} в тер-

минах $S = \widetilde{S}_\infty(B_{cr}, 1)$; показатель экспоненты будет задаваться равенством

$$E = \widetilde{E}_\infty(B_{cr}, 1) + R_{cr} - R,$$

где R_{cr} определяется формулой (8.5.95).

Используя (8.5.88) и (8.5.95), получаем

$$E = \frac{S}{4B_{cr}} + \int\limits_{f:\ \frac{|H_1(f)|^2}{N(f)} \geqslant \frac{1}{B}} \left\{ -1/2 \ln\left[2 - \frac{N(f)}{B_{cr}|H_1(f)|^2}\right] + \right.$$

$$\left. + \frac{1}{2} \ln\left[B_{cr}\frac{|H_1(f)|^2}{N(f)}\right]\right\} df - R,$$

$$\frac{dE}{dS} = \frac{\partial E}{\partial S}\bigg|_{B_{cr}} + \frac{\partial E}{\partial B_{cr}} \Big/ \frac{\partial S}{\partial B_{cr}}.$$

Беря производную $\partial E/\partial B_{cr}$ и используя $S = \widetilde{S}_\infty(B_{cr}, 1)$ из (8.5.94), опять видим, что $\partial E/\partial B_{cr} = 0$, так что

$$\frac{dE}{dS} = \frac{1}{4B_{cr}}. \tag{2}$$

Этот результат справедлив для $\widetilde{S}_\infty(B, 1) \leqslant S \leqslant \widetilde{S}_{x,\infty}(B_x, 1)$, где B_x равно B, которое удовлетворяет (8.5.102) для $R = \widetilde{R}'_{x,\infty}$. Заметим, что B_{cr} в (2) возрастает по S, начиная с B из (1) до B_x. Так как $B_{cr} = B$ при $S = \widetilde{S}_\infty(B, 1)$, то, очевидно, (1) и (2) совпадают при $S = \widetilde{S}_\infty(B, 1)$, что доказывает непрерывность в точке $\widetilde{S}_\infty(B, 1)$. Наконец, из (8.5.103) имеем

$$\frac{d\widetilde{E}_{x,\infty}}{d\widetilde{S}_{x,\infty}} = \frac{1}{4B_x}, \tag{3}$$

где B_x — значение B, определенное из (8.5.102) для $R = \widetilde{R}'_{x,\infty}$. Так как $B_x = B_{cr}$ для $S = \widetilde{S}_{x,\infty}(B_x, 1)$ то (2) и (3) совпадают в этой точке, что доказывает непрерывность в этой точке.

8.21. (а) Пусть послано сообщение i. Тогда

$$y_{1,i} = \int\limits_{-T/2}^{T/2} v_{1,i} \frac{2}{\sqrt{T}} \cos^2(2\pi f_1 t)\, dt + \int\limits_{-T/2}^{T/2} z(t) \frac{2}{\sqrt{T}} \cos(2\pi f_1 t)\, dt = v_{1,i}\sqrt{T} +$$

$$+ \sqrt{2} \int z(t) \frac{2}{\sqrt{T}} \cos(2\pi f_1 t)\, dt.$$

Первое слагаемое имеет дисперсию $\mathscr{E}$, так как $v_{1,i}$ имеет дисперсию $\mathscr{E}/T$; последний интеграл является гауссовской случайной величиной с дисперсией $N_0/2$ и произведение этой случайной величины на $\sqrt{2}$, следовательно, имеет дисперсию N_0. Так как эти слагаемые независимые и гауссовские, то сумма является гауссовской случайной величиной с дисперсией $\mathscr{E} + N_0$. Этот же вывод применим к $y_{2,i}$. Те же доводы также применимы к $y_{1,s}(j \neq i)$, за исключением того, что здесь отсутствует компонента сигнала; это приводит к дисперсии N_0.

$$\text{(б)}\quad p[\mathbf{y} \mid x_1(t)] = \frac{1}{2\pi(N_0 + \mathscr{E})2\pi N_0} \exp\left[-\frac{(y_{1,1}^2 + y_{2,1}^2)}{2(N_0 + \mathscr{E})} - \frac{(y_{1,2}^2 + y_{2,2}^2)}{2N_0}\right],$$

$$p[\mathbf{y} \mid x_2(t)] = \frac{1}{2\pi(N_0 + \mathscr{E})\, 2\pi N_0} \exp\left[-\frac{(y_{1,1}^2 + y_{2,1}^2)}{2N_0} - \frac{(y_{1,2}^2 + y_{2,2}^2)}{2(N_0 + \mathscr{E})}\right].$$

Взяв логарифм от отношения, получаем

$$r(\mathrm{y}) = \frac{(y_{1,1}^2 + y_{2,1}^2)}{2}\left(\frac{1}{N_0} - \frac{1}{N_0+\mathscr{E}}\right) - \frac{(y_{1,2}^2 + y_{2,2}^2)}{2}\left(\frac{1}{N_0} - \frac{1}{N_0+\mathscr{E}}\right).$$

Обозначим первый член приведенного выше выражения через s_1, а второй — через s_2. При заданном $x_1(t)$ член s_1 представляет собой сумму квадратов двух одинаково распределенных гауссовских случайных величин и, следовательно, s_1 имеет экспоненциальное распределение. Среднее значение s_1 равно

$$(N_0+\mathscr{E})\left(\frac{1}{N_0} - \frac{1}{N_0+\mathscr{E}}\right) = \frac{\mathscr{E}}{N_0}.$$

Таким образом,

$$p_1(s_1 \mid x_1(t)) = \begin{cases} \dfrac{N_0}{\mathscr{E}} \exp\left[-\dfrac{N_0}{\mathscr{E}} s_1\right], & s_1 \geqslant 0, \\ 0, & s_1 < 0. \end{cases}$$

Аналогично при заданном $x_1(t)$ распределение s_2 экспоненциально со средним $N_0\left(\dfrac{1}{N_0} - \dfrac{1}{N_0+\mathscr{E}}\right) = \dfrac{\mathscr{E}}{N_0+\mathscr{E}}$ и, следовательно,

$$p_2(s_2 \mid x_1(t)) = \begin{cases} \dfrac{N_0+\mathscr{E}}{\mathscr{E}} \exp\left[-\dfrac{N_0+\mathscr{E}}{\mathscr{E}} s_2\right], & s_2 \geqslant 0, \\ 0, & s_2 < 0. \end{cases}$$

Так как s_2 и s_1 независимы [при заданном $x_1(t)$], то распределение $r = s_1 - s_2$ можно найти с помощью соотношений

$$p(r \mid x_1(t)) = \int p_1(s_1 \mid x_1(t))\, p_2(s_1 - r \mid x_1(t))\, ds_1 =$$

$$= \begin{cases} \dfrac{N_0(N_0+\mathscr{E})}{\mathscr{E}(2N_0+\mathscr{E})} \exp\left[-\dfrac{N_0}{\mathscr{E}} r\right], & r \geqslant 0, \\ \dfrac{N_0(N_0+\mathscr{E})}{\mathscr{E}(2N_0+\mathscr{E})} \exp\left[\dfrac{N_0+\mathscr{E}}{\mathscr{E}} r\right], & r < 0. \end{cases}$$

Из симметрии ясно, что

$$p(-r \mid x_2(t)) = p(r \mid x_1(t)).$$

(в) При декодировании по максимуму правдоподобия декодируется x_1, если $r(\mathbf{y}) \geqslant 0$, и x_2, если $r(\mathbf{y}) < 0$. Следовательно,

$$P_e = P_{e,1} = P_{e,2} = \int_{-\infty}^{0} p(r \mid x_1(t))\, dr =$$

$$= \int_{-\infty}^{0} \frac{N_0(N_0+\mathscr{E})}{\mathscr{E}(2N_0+\mathscr{E})} \exp\left[\frac{N_0+\mathscr{E}}{\mathscr{E}} r\right] dr = \frac{N_0}{2N_0+\mathscr{E}} = \frac{1}{2+\mathscr{E}/N_0}.$$

(г) Из (8.6.22) находим

$$P_{e,m} \leqslant \exp\left[-2\ln\left(1+\frac{\lambda_1}{2N_0}\right) + \ln\left(1+\frac{\lambda_1}{N_0}\right)\right] = \frac{1+\mathscr{E}/N_0}{(1+\mathscr{E}/2N_0)^2}$$

Заметим, что эта граница отличается от точного результата множителем 2 при $\mathscr{E}/N_0 \to 0$ и множителем 4 при $\mathscr{E}/N_0 \to \infty$.

(д) Заметим, что $\varphi_1(t) = 1/\sqrt{T}$ — собственная функция (8.6.7) и что соответствующее собственное значение равно $\lambda_1 = \mathscr{E}$. Это можно увидеть, подставляя эти значения в (8.6.7). Для любой функции $\varphi_j(t)$, ортогональной $\varphi_1(t)$,

$$\int_{-T}^{T} \frac{\mathscr{E}}{T} \varphi_j(t)\, dt = \frac{\mathscr{E}}{T} \sqrt{T} \int \varphi_1(t)\, \varphi_j(t)\, dt = 0.$$

Следовательно, все другие собственные значения равны 0. Если принять, что $\sigma(f)$ — гладкая унимодальная функция, то предположение, что ее преобразование Фурье $\mathscr{R}(\tau)$ постоянно на $(-T, T)$, указывает, что $\sigma(f)$, по существу, равно нулю для $|f| > 1/2T$ или что допплеровский сдвиг мал по сравнению с $1/2T$.

8.22. (а) Из (8.6.15) видно, что декодер максимального правдоподобия выбирает m, которое максимизирует $p_1(y_m)/p_0(y_m)$, где

$$\frac{p_1(\mathbf{y}_m)}{p_0(\mathbf{y}_m)} = \prod_{i=1}^{2} \prod_{j=1}^{\infty} \frac{[2\pi(N_0+\lambda_j)]^{-1/2} \exp[-y_{i,m,j}^2/2(N_0+\lambda_j)]}{(2\pi N_0)^{-1/2} \exp[-y_{i,m,j}/2N_0]}.$$

Беря логарифм от этого выражения и опуская члены, не зависящие от m, выбираем m, которое максимизирует выражение

$$\sum_{i=1}^{2} \sum_{j=1}^{\infty} y_{i,m,j}^2 \left[-\frac{1}{2(N_0+\lambda_j)} + \frac{1}{2N_0} \right] = \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{\infty} y_{i,m,j}^2 \left(\frac{\lambda_j}{N_0+\lambda_j} \right).$$

Опуская множитель $1/2$ (который не отражается при максимизации по m), получаем требуемый результат.

(б) Приведенную выше сумму можно переписать, используя сначала энергетическое соотношение (8.1.16) и затем определение $y_{i,m,j}$ (8.6.10). Получаем

$$\sum_{i=1}^{2} \sum_{j=1}^{\infty} \left(y_{i,m,j} \sqrt{\frac{\lambda_j}{N_0+\lambda_j}} \right)^2 = \sum_{i=1}^{2} \int \left[\sum_j y_{i,m,j} \sqrt{\frac{\lambda_j}{N_0+\lambda_j}}\, \varphi_j(t) \right]^2 dt =$$

$$= \sum_{i=1}^{2} \int \left[\sum_j \int y_{i,m}(\tau)\, \varphi_j(\tau)\, d\tau \sqrt{\frac{\lambda_j}{N_0+\lambda_j}}\, \varphi_j(t) \right]^2 dt =$$

$$= \sum_{i=1}^{2} \int \left[\int y_{i,m}(\tau)\, h(t, \tau)\, d\tau \right]^2 dt,$$

где

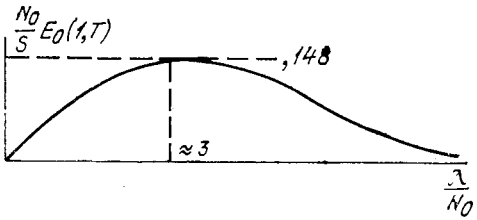
$$h(t, \tau) = \sum_j \sqrt{\frac{\lambda_j}{N_0+\lambda_j}}\, \varphi_j(\tau)\, \varphi_j(t).$$

Это выражение дает требуемую форму ответа, который содержит сначала фильтрацию $y_{i,m}(\tau)$ и затем квадрирование и интегрирование выхода фильтра. Те же операции можно, конечно, выполнить на промежуточной частоте, что устраняет необходимость суммирования по i.

8.23. Пренебрегая тем, что $n = ST_1/\lambda$ может не быть целым числом и для простоты положив $T_1 = T$, имеем

$$E_0(\rho, T) = \frac{S}{\lambda} \left[(1+\rho) \ln\left(1 + \frac{\rho\lambda}{(1+\rho)N_0} \right) - \rho \ln\left(1 + \frac{\lambda}{N_0} \right) \right],$$

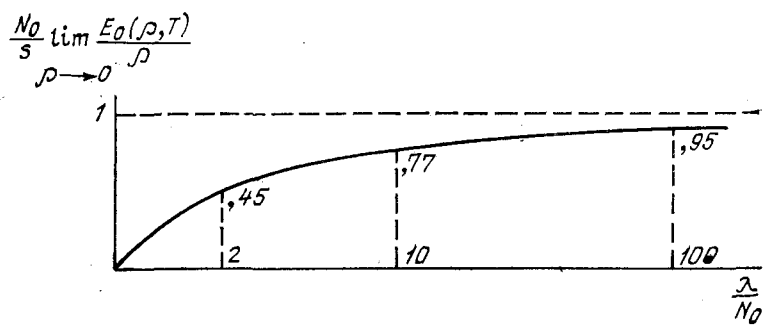
$$E_0(1, T) = \frac{S}{\lambda} \left[\ln \frac{(1+\lambda/2N_0)^2}{(1+\lambda/N_0)} \right] = \frac{S}{\lambda} \ln\left(1 + \frac{\lambda^2}{4N_0(N_0+\lambda)} \right).$$



Для малых λ имеем $E_0(1, T) \approx S\lambda/4N_0^2$; для больших λ заметим, что $E_0(1, T)$ стремится к нулю приближенно как $1/\lambda$. Имеем

$$\frac{E_0(\rho, T)}{\rho} = \frac{S}{\lambda}\left[\left(\frac{1}{\rho}+1\right)\ln\left(1+\frac{\rho\lambda}{(1+\rho)N_0}\right)-\ln\left(1+\frac{\lambda}{N_0}\right)\right],$$

$$\lim_{\rho\to 0}\frac{E_0(\rho, T)}{\rho} = \frac{S}{\lambda}\left[\frac{\lambda}{N_0}-\ln\left(1+\frac{\lambda}{N_0}\right)\right] = \frac{S}{N_0}-\frac{S}{N_0}\left(\frac{N_0}{\lambda}\right)\ln\left(1+\frac{\lambda}{N_0}\right).$$



9.1. Для первой меры искажения угадываем, что выбор $P(j \mid k)$, удовлетворяющий равенству $P(1 \mid 0) = P(0 \mid 1)$, дает $R(d^*)$. Так как $\frac{1}{2} P(1 \mid 0) + \frac{1}{2}P(0 \mid 1)$ равно среднему искажению, то выбираем $d^* = P(1 \mid 0) = P(0 \mid 1)$. Тогда $\mathcal{I}(Q; \mathbf{P})$ равно $\ln 2 - \mathcal{H}(d^*)$, так что $R(d^*) = \ln 2 - \mathcal{H}(d^*)$ нат.

Для того чтобы проверить, что этот выбор $P(j \mid k)$ дает $R(d^*)$, предположим, что $P(j \mid k)$ произвольно и пусть $P'(1|0) = P'(0 \mid 1) = \frac{1}{2} P(1 \mid 0) + \frac{1}{2}P(0 \mid 1)$. Видим, что $P'(j \mid k)$ дает то же самое среднее искажение, что и $P(j \mid k)$, однако в силу выпуклости меньшее значение средней взаимной информации.

Для второго источника положим $P(2|0) = P(2 \mid 1) = d^*$ и $P(0 \mid 0) = P(1 \mid 1) = 1 - d^*$. Тогда $\mathcal{I}(Q; P) = (1 - d^*) \ln 2$ и $R(d^*) = (1 - d^*) \ln 2$. С помощью таких же, как и выше, соображений проверяется, что при этом выборе $P(j \mid k)$ действительно достигается $R(d^*)$.

9.2. (а) Для кода Хэмминга (7,4) 1/8 последовательностей длины 7 являются кодовыми словами и 7/8 лежат на расстоянии один от некоторого кодового слова. Следовательно, среднее искажение на слово равно 7/8, а среднее искажение на букву равно 1/8. Значение $R(d^*)$ для $d^* = \frac{1}{8}$ равно 0,457 ln2; это число, конечно, несколько меньше чем $\frac{4}{7} \ln 2$.

(б) Для произвольного l скорость равна $\dfrac{2^l - l - 1}{2^l - 1} \ln 2$ и среднее искажение равно 2^{-l}.

9.3. Для того чтобы получить требуемое искажение d^*, вычеркните долю d^* символов источника, стоящих на заранее определенных позициях в потоке данных и рассмотрите невычеркнутую часть последовательности, как закодированную последовательность.

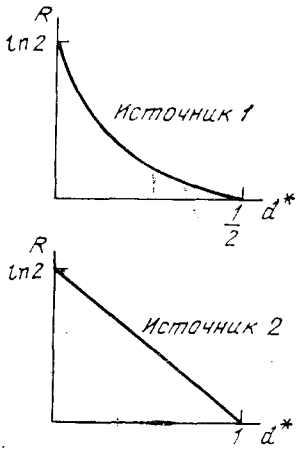
9.4. Используем теорему 9. 4. 1. Из симметрии ясно, что f_k должно не зависеть от k и (9.4.9) сводится к соотношениям:

$$f \leqslant 1 \quad (j=0,\ 1,\ 2,\ 3),$$
$$2f\,e^{-\rho} \leqslant 1 \quad (j=4,\ 5),$$
$$4f\,e^{-3\rho} \leqslant 1 \quad (j=6).$$

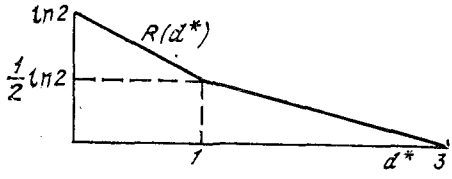
Следовательно, из (9.4.8) имеем

$$\min_{\mathbf{P}} R_0(\rho,\ \mathbf{P}) = \begin{cases} \ln 4, & \rho > \ln 2, \\ \ln 2 + \rho, & \ln 2 \geqslant \rho > {}^1/_2 \ln 2, \\ 3\rho, & {}^1/_2 \ln 2 \geqslant \rho. \end{cases}$$

Взяв огибающую прямых линий $\min_{\mathbf{P}} R_0(\rho,\ \mathbf{P}) - \rho d^*$, получаем функцию, изображенную на рисунке. Заметим, что выход 6 не используется при $d^* \leqslant 1$ и что выходы 0, 1, 2, 3 не используются при $d^* \geqslant 1$.



К задаче 9.1.



К задаче 9.4.

9.5. (а). Заметим, что если ДКБП связывает источник с адресатом, то среднее искажение $\bar{d}$ равно 0, если $P(j \mid k) > 0$ только для k, j, таких, что $d(k;\ j) = 0$, и $\bar{d}$ бесконечно в других случаях. При любом заданном выборе $P(j \mid k)$, для которого $\bar{d} = 0$, имеем $I(U;\ V) = H(U) - H(U \mid V) = \ln 5 - H(U \mid V)$. Для любой буквы адресата U возможны только два входа и, следовательно, $H(U \mid V) \geqslant \ln 2$ и $I(U;\ V) \geqslant \ln (5/2)$.

Так как очевидный интуитивный выбор $P(j \mid k) = {}^1/_2$ для всех j, k, таких, что $d(k;\ j) = 0$, дает $I(U;\ V) = \ln(5/2)$, то этот выбор минимизирует $I(U;\ V)$ при ограничении $\bar{d} = 0$. Следовательно, $R(d^*) = \ln (5/2)$ для всех $d^* < \infty$.

(б) Для любого $R > \ln (5/2)$ пусть $\hat{R} = \ln(5/2)$; $\hat{d} = 0$. Для выбранного тест-канала и длины блока N и для всех $\mathbf{u}$, $\mathbf{v}$, таких, что $P_N(\mathbf{v} \mid \mathbf{u}) > 0$, имеем $I(\mathbf{u};\ \mathbf{v}) = N \ln (5/2)$. Аналогично $D(\mathbf{u};\ \mathbf{v}) = 0$ для всех $\mathbf{u}$, $\mathbf{v}$, для которых $P_N(\mathbf{v} \mid \mathbf{u}) > 0$. Поэтому $P_t(A)$, определенная в лемме 9.3.1, равна 0. Тогда лемма 9.3.1 утверждает, что в ансамбле кодов вероятность $P_c(D > 0)$ того, что искажение больше чем 0, удовлетворяет неравенству

$$P_c(D > 0) \leqslant \exp\left(-M\,e^{-N \ln 5/2}\right).$$

Для $M = \lfloor e^{RN} \rfloor > e^{RN} - 1$ это неравенство принимает вид

$$P_c(D > 0) < \exp\left[(e^{RN} - 1)\,e^{-N \ln 5/2}\right].$$

Так как $R > \ln(5/2)$, то это выражение убывает быстрее, чем экспоненциально с N и, следовательно, для достаточно больших N

$$P_c(D > 0) < 5^{-N}. \tag{1}$$

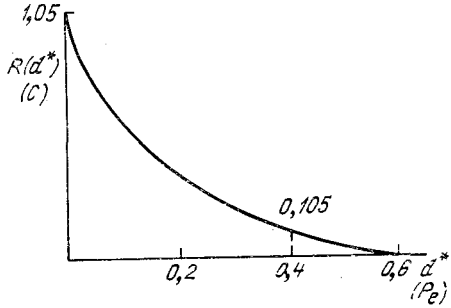
Наконец, для любого заданного кода из ансамбля каждая последовательность источника отображается в кодовое слово либо с нулевым искажением, либо с бесконечным искажением. Так как каждая последовательность источника имеет вероятность 5^{-N}, то для частного кода вероятность того, что искажение больше 0, равна 0 или не меньше 5^{-N}. В силу (1) невозможно, чтобы $\Pr[D > 0]$ было не меньше 5^{-N} для всех кодов из ансамбля, следовательно, по крайней мере для одного кода $\Pr[D > 0] = 0$. Такой код, очевидно, имеет нулевое искажение.

9.6. Для заданного источника имеем $Q_{min} = 0{,}2$. Тогда из (9.5.13) выводим

$$R(d^*) = H(U) - \mathscr{H}(d^*) - d^* \ln 2,$$

для $d^* \leqslant 2Q_{min} = 0{,}4$. Для бо́льших значений d^* имеем [из (9.5.25)]

$$R(d^*) = S_m [H(U_m) - \mathscr{H}(\hat{d}) - \hat{d} \ln (m-1)], \tag{2}$$



где

$$S_m = \sum_{k=0}^{m-1} Q(k), \qquad \hat{d} = \frac{d^* - (1 - S_m)}{S_m},$$

$$H(U_m) = \sum_{k=0}^{m-1} \frac{Q(k)}{S_m} \ln \frac{S_m}{Q(k)}$$

и m выбирается так, чтобы удовлетворить

$$mQ(m) + \sum_{k=m+1}^{K-1} Q(k) < d^* \leqslant (m-1)Q(m-1) + \sum_{k=m}^{K-1} Q(k). \tag{3}$$

Для $m = 2$ выражение (1) принимает вид

$$R(d^*) = 0{,}8 \left[\ln 2 - \mathscr{H}\left(\frac{d^* - 0{,}2}{0{,}8} \right) \right]. \tag{4}$$

Из (3) выводим, что (4) справедливо для $0{,}4 < d^* \leqslant 0{,}6$. Из (4) видно, что для $d^* = 0{,}6$ имеем $R(d^*) = 0$. Для этого источника и меры искажения имеем $d^*_{max} = 0{,}6$ [см. (9.2.7)], так что $R(d^*) = 0$ для $d^* \geqslant 0{,}6$. Заметим, что наклон $R(d^*)$ стремится к нулю при $d^* \to 0{,}6$. Такая ситуация возникает для этой

меры искажения, когда две наиболее вероятные буквы источника имеют равные вероятности. Приведенная здесь кривая является графиком $R(d^*)$ и также дает минимум достижимой P_e на символ как функцию C.

9.7. Как и в основном тексте, имеем соотношения (9.7.2)—(9.7.4) и

$$\min R(\rho, \mathbf{P}) \geqslant \int_{-\infty}^{\infty} q(u) \ln \frac{f(u)}{q(u)}\, du,$$

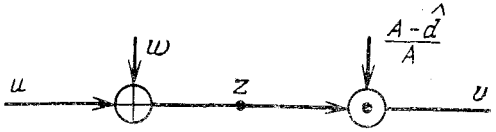
где $f(u) = \sqrt{\rho/\pi}$ удовлетворяет ограничению (9.7.3). Следовательно,

$$\min R_0(\rho, \mathbf{P}) \geqslant \int_{-\infty}^{\infty} q(u) \ln f(u)\, du + H(U) = H(U) + \frac{1}{2} \ln \frac{\rho}{\pi}.$$

Таким образом,

$$R(d^*) \geqslant \min R_0(\rho, \mathbf{P}) - \rho d^* \geqslant H(U) + \frac{1}{2} \ln \frac{\rho}{\pi} - \rho d^*$$

для любого $\rho > 0$. Максимум по ρ имеет место при $\rho = 1/(2d^*)$, давая $R(d^*) \geqslant H(U) - {}^1/_2 \ln (2\pi e\, d^*)$.



Для вывода верхней границы примем, что случайная величина источника u имеет нулевое среднее и что u проходит через канал, изображенный на рисунке, где w — гауссовская случайная величина, не зависящая от u и имеющая нулевое среднее значение и дисперсию $\hat{d}A/(A - \hat{d})$. Примем, что $\hat{d} < A$.

Если среднее значение u отлично от нуля, то это среднее может быть вычтено перед подачей в канал и добавлено на выходе канала, что не повлияет на среднее искажение или взаимную информацию. Имеем

$$\bar{d} = \overline{(u-v)^2} = \overline{\left[u - (u+w)\frac{A-\hat{d}}{A}\right]^2} = \overline{\left[u\frac{\hat{d}}{A} - w\frac{A-\hat{d}}{A}\right]^2} =$$

$$= A\left(\frac{\hat{d}}{A}\right)^2 + \frac{\hat{d}A}{A-\hat{d}}\left(\frac{A-\hat{d}}{A}\right)^2 = \hat{d},$$

$$I(U; V) = I(U; Z) \leqslant \frac{1}{2} \ln \left[1 + \frac{\overline{u^2}}{\overline{w^2}}\right] = \frac{1}{2} \ln \left[1 + \frac{A-\hat{d}}{\hat{d}}\right] = \frac{1}{2} \ln \frac{A}{\hat{d}}.$$

Следовательно, $R(d^*) \leqslant 1/2 \ln (A/d^*)$ при $d^* < A$.

9.8. (а) Уравнения (9.7.42) и (9.7.43) дают $R(d^*)$ для источника, представляющего собой гауссовский случайный процесс, и при среднеквадратической мере искажения. Для спектральной плотности, указанной в задаче и определяемой прямоугольником, эти уравнения приводят к простым выражениям

$$d^* = \frac{1}{2\rho} \int_{-W_1}^{W_1} df = \frac{W_1}{\rho}; \quad \rho > \frac{1}{2A}, \tag{1}$$

$$R(d^*) = \int_{-W_1}^{W_1} \frac{1}{2} \ln [2\rho A]\, df = W_1 \ln (2\rho A). \tag{2}$$

Подставляя ρ из (1), имеем

$$R(d^*) = W_1 \ln (2W_1 A/d^*).$$

(б)

$$C = W_2 \ln (1 + S/(W_2 N_0)).$$

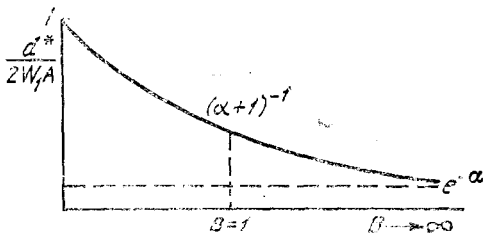
(в) Минимум среднеквадратической ошибки, который может быть достигнут при передаче от источника пункта (а) по каналу пункта (б) при использовании лучшего метода кодирования и декодирования, равен значению d^*, для которого $R(d^*) = C$. Решая это уравнение относительно d^*, получаем минимальную среднеквадратическую ошибку $d^* = 2W_1 A[1 + S/(W_2 N_0)]^{W_2/W_1}$. Для графического изображения этого удобно определить

$$\beta = W_2/W_1, \quad \alpha = S/(W_1 N_0).$$

Тогда

$$\frac{d^*}{2W_1 A} = \left(1 + \frac{\alpha}{\beta}\right)^{-\beta}.$$

Величину β можно интерпретировать как индекс модуляции; α — как энергетическое отношение сигнал/шум в канале на одну степень свободы источника и $d^*/(2W_1 A)$ — как отношение среднеквадратической ошибки к мощности источника.



СПИСОК ОБОЗНАЧЕНИЙ

$\oplus$ Сложение по модулю 2.

$\lceil x \rceil$ Наибольшее целое число, меньшее или равное x (целая часть x).

$\lfloor x \rfloor$ Наименьшее целое число, большее или равное x.

$x \approx y$ x приближенно равно y.

$x \ll y$ x пренебрежимо мало по сравнению с y.

$\bar{z}$ Среднее значение (математическое ожидание) случайной величины z.

$f \triangleq g$ f по определению равно g.

$\|f\|$ Норма функции f, $\|f\| = \sqrt{\int f^2(x)\,dx}$.

$(f \cdot g)$ Скалярное произведение функций f и g, $(f \cdot g) = \int f(x)\, g(x)\, dx$.

$|f(D)|_m^n$ $f_m D^m + f_{m+1} D^{m+1} + \ldots + f_n D^n$ при $n \geqslant m$ и 0 при $m > n$, где $f(D) = \sum f_k D^k$ — многочлен.

A^c Дополнение множества A.

$A \cup B$ Объединение множеств A и B (т. е. множество элементов, содержащихся в A или в B или в обоих множествах).

$\bigcup\limits_m A_m$ $A_1 \cup A_2 \cup A_3 \cup \ldots$

$A \cap B$ Пересечение множеств A и B (т. е. множество элементов, содержащихся в множествах A и B одновременно).

$\|A\|$ Число элементов в множестве A.

$x \in A$ Элемент a содержится в множестве A.

$a : S$ Множества элементов a, таких, что удовлетворяется утверждение S. Например, $\sum\limits_{n:\, x_n > 1} x_n$ является суммой x_n по n, для которых $x_n > 1$.

$A = (a : S)$ A определяется как множество элементов a, для которых удовлетворяется утверждение S.

$S_1 \Rightarrow S_2$ Из утверждения S_1 следует утверждение S_2.

$S_1 \Leftarrow\Rightarrow S_2$ Из утверждения S_1 следует S_2 и из S_2 следует S_1.

$\overline{\qquad}$ Обозначает конец теоремы.

| Обозначает конец доказательства.

НОД Наибольший общий делитель.

$H(\ )$ Энтропия; см. § 2. 2.

$\mathcal{H}(x)$ $-x \log x - (1-x) \log (1-x)$; $0 \leqslant x \leqslant 1$.

$I(\ )$ Информация; см. § 2.2.

НОК Наименьшее общее кратное.

$\lim\limits_{x \to y+}$ Предел при x, стремящемся к y справа.

$\lim\limits_{x \to y-}$ Предел при x, стремящемся к y слева.

$R_n(m)$ Остаток от деления m на n (m и n — положительные целые числа).

$R_{g(D)} f(D)$ Остаток от деления многочлена $f(D)$ на $g(D)$.

$\sup f(x)$ Верхняя грань; наименьшее число, не меньшее $f(x)$ во всей допустимой области значений x. Если существует максимум, то $\sup f(x) = \max f(x)$.

$\inf f(x)$ Нижняя грань; наибольшее число, не превосходящее $f(x)$ во всей допустимой области значений x. Если существует минимум, то $\inf f(x) = \min f(x)$.

ПРИМЕЧАНИЯ РЕДАКТОРОВ

К главе 1.

Первоначальное знакомство с идеями теории информации можно получить по книгам Харкевича (1965), А. Яглома и И. Яглома (1973).

К главе 2.

Колмогоров (1956) предложил наиболее общий подход к построению теории передачи информации и наметил программу его строгого математического обоснования. Этому же посвящена статья Добрушина (1961). Эти две работы рекомендуются математически настроенному читателю, впервые знакомящемуся с теорией информации.

Новые логические основания теории информации, построенные на понятии сложности последовательности, были предложены Колмогоровым (1965, 1969). Колмогоров показал, что такие понятия теории информации как «энтропия» и «информация» могут быть введены без ссылки на теорию вероятностей и тем самым могут быть применены к индивидуальным событиям. Дальнейшее развитие этих идей содержится в статье Звонкина и Левина (1970).

К главе 3.

Постановка задачи кодирования для сокращения длины последовательности букв, вырабатываемых источником сообщений, была обобщена Фитингофом (1966) на случай, когда неизвестны вероятностные свойства источника. Для этого случая была доказана теорема, аналогичная теореме 3.3.1, и построен эффективный универсальный способ кодирования. Аналогичная постановка задачи, когда вероятностные свойства источника полностью или частично неизвестны, рассматривалась Кричевским (1968), получившим асимптотику сходимости к нулю избыточности для различных методов кодирования. Фитингоф (1967) ввел новое минимаксное определение оптимального кодирования (отличное от предложенного Шенноном, состоящего в минимизации $\bar{n}$) и построил соответствующие этому определению оптимальные коды. В работе Бабкина (1971) предложен универсальный метод кодирования, допускающий сравнительно простую реализацию.

Марков (1960, 1961, 1962, 1963) исследовал свойство взаимной однозначности неравномерного кодирования. В его работе построен графический алгоритм определения существования этого свойства. Декодирование неравномерного кода с помощью конечного автомата было рассмотрено Левенштейном (1961) для случая, когда в начальный момент кодирующий и декодирующий автомат синхронизованы и когда такая синхронизация отсутствует. Коды, обладающие свойством синхронизации, были рассмотрены Кирилловым (1959), Левенштейном (1965, 1969, 1971) и другими авторами.

Первое математическое доказательство теоремы Макмиллана было дано Хинчиным (1956).

К главе 4.

Наиболее общее доказательство обращения теоремы кодирования, основанное на алгебраических свойствах информации, принадлежит Колмогорову (1956). Пропускная способность дискретных каналов в предположении, что шум велик (или мал), исследовалась в работе Прелова (1966).

К главе 5.

Первое строгое доказательство теоремы кодирования для стационарного канала с конечной памятью принадлежит Хинчину (1956). В наиболее общей форме теорема кодирования была доказана Добрушиным (1959), который впер-

вые указал, что ее справедливость связана со свойством информационной устойчивости.

Границы для вероятности ошибки в дискретных каналах исследовались в ряде работ советских авторов. Добрушин (1962 б) получил асимптотические выражения для оптимальной вероятности ошибки в каналах, симметричных по входу и выходу. Добрушин (1962 а) и Молчанов (1967) получили выражения для асимптотики оптимальной вероятности ошибки при конечном числе передаваемых сообщений. Границы сферической упаковки для вероятности ошибки в каналах с памятью исследовались Егарминым (1969), который на основе развитого им метода получил выражение для показателя экспоненты в канале с эргодическим марковским аддитивным шумом. Результаты, аналогичные изложенным в § 5.9, были получены Габидулиным (1969).

К главе 6.

Советская литература по теории кодирования насчитывает более тысячи наименований. Опубликован ряд обобщающих монографий. Укажем здесь лишь монографию Колесника и Мирончикова (1968), посвященную детальному рассмотрению циклических кодов и методов их декодирования.

Новый класс линейных кодов, исправляющих ошибки, предложил Гоппа (1970). Коды этого класса задаются некоторым многочленам над полем $GF(q)$. Коды БЧХ являются единственными циклическими кодами, входящими в этот класс. Построение кодов основано на отождествлении исходного пространства двоичных векторов с некоторым множеством рациональных функций. Параметры кода следующие: $n \leqslant 2m$, $k \geqslant n - mt$, $d \geqslant 2t + 1$, где n — длина кода, k — число информационных символов, t — степень многочлена, задающего код. Гоппа показал, что для всех введенных кодов существует схема декодирования, аналогичная алгоритму Питерсона для БЧХ-кодов. Гоппа (1971) показал также, что почти все коды рассматриваемого класса приближаются с ростом n к границе Варшамова—Гилберта.

Один отрицательный результат принадлежит Берману (1967) и состоит в том, что для любого циклического кода, длина которого равна произведению конечного числа простых чисел, минимальное расстояние в коде, деленное на n, стремится к нулю с ростом n, т. е. что среди таких кодов нет кодов, приводящих к экспоненциально убывающей вероятности ошибки.

Среди работ по последовательному декодированию необходимо отметить работу Добрушина (1964), посвященную строгому математическому анализу алгоритма Возенкрафта и работу Кошелева (1966 а), предложившего модификацию этого алгоритма. Пинскер (1965) разработал итеративный алгоритм последовательного декодирования, позволяющий увеличить вычислительную скорость процедуры. Новый алгоритм последовательного декодирования был предложен и рассчитан Зигангировым (1966). Этот алгоритм требует меньшего числа операций, чем рассмотренный в книге алгоритм Фано.

К главе 7.

Теорема кодирования для непрерывных по амплитуде и дискретных по времени каналов без памяти является следствием теоремы кодирования, доказанной Добрушиным (1959).

Пропускная способность общего дискретного по времени гауссовского канала без памяти и с памятью рассматривалась в работе Пинскера (1956). Цыбаковым (1965) были найдены выражения для пропускной способности невырожденных и вырожденных многомерных гауссовских каналов.

К главе 8.

Строгое математическое описание и представление белого шума можно найти в книге Гельфанда и Виленкина (1961).

Теорема отсчетов, представленная в § 8.1, в советской литературе часто называется теоремой Котельникова.

Выражение для количества информации гауссовских процессов исследовалось во многих работах советских авторов. Применимость теоремы Шеннона для гауссовских каналов является следствием теоремы кодирования Добрушина

(1959), информационной устойчивости произвольной пары гауссовских процессов, доказанной Пинскером (1960), и того факта, что максимальное значение количества информации в гауссовском канале достигается на гауссовской паре входного и выходного процессов.

Пятошиным (1968) исследовалась пропускная способность гауссовского канала, когда на входе канала вводилось дополнительное ограничение, состоящее в том, что число входных сигналов не может быть больше некоторого целого числа K.

В работах советских авторов были предложены и исследованы ряд моделей реальных каналов. В работе Сифорова (1958) приведены формулы и оценки пропускной способности канала с замираниями. Было показано, что наличие замирания уменьшает пропускную способность канала не более чем на 17%. Выражения для пропускной способности различных диспергирующих каналов с замираниями исследовались Цыбаковым (1959 а, б).

Изучение пропускной способности каналов при разнесенном приеме было проведено в работах Овсеевича и Пинскера (1961). Было показано, что наличие разнесенного приема ослабляет эффекты диспергирования и замирания в канале.

К главе 9.

Добрушиным (1959) был предложен общий подход к доказательству теоремы кодирования для источников с заданным уровнем верности. По существу, при таком подходе для доказательства теоремы кодирования достаточно установить информационную устойчивость источника. Этот результат позволил Добрушину доказать теоремы кодирования для источников с независимыми значениями и покомпонентными условиями верности. Пинскер (1963 а) доказал, что теорема кодирования имеет место для стационарных вполне эргодических источников. Этот результат был усилен Мартон (1972), обобщившей его на произвольные эргодические источники.

Справедливость теоремы кодирования для произвольных гауссовских источников была установлена Пинскером (1963 б). Было показано, что при критерии верности, определяемом вторыми моментами, минимальное количество информации достигается на гауссовской паре процессов; отсюда и из информационной устойчивости гауссовских процессов следует теорема кодирования.

Ерохиным (1958) были найдены выражения для ε-энтропии дискретных сообщений с вероятностью ошибки в качестве критерия верности.

Выражения для ε-энтропии гауссовских случайных векторов и процессов при среднеквадратическом критерии верности были выписаны Колмогоровым (1956). Пинскер (1963 б) распространил эти результаты на более широкий класс гауссовских процессов при взвешенном среднеквадратическом критерии верности. Цыбаков (1969) получил выражение ε-энтропии, когда критерий верности задается произвольной неотрицательно определенной квадратичной формой.

Метод передачи гауссовской случайной величины, аналогичный предложенному Шелквийком (1966), рассматривался Зигангировым (1967).

Овсеевичем (1970) было показано, что при передаче произвольного гауссовского сообщения по каналу с белым гауссовским шумом при наличии бесшумной обратной связи линейный метод является оптимальным.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ И РЕКОМЕНДУЕМЫЕ КНИГИ

* Амбарцумян Г. А., 1958, К энтропии цепей Маркова. Изв. АН Арм. ССР, Сер. физ.-мат. наук, **11**, 2, 31—40.

Артин (Artin E.), 1946, Galois Theory. Notre Dame Mathematical Lectures, Notre Dame, Indiana.

* Арутюнян Е. А., 1968, Оценка экспоненты вероятности ошибки для полунепрерывного канала без памяти. Пробл. перед. инф., **4**, 4, 37—48.

Ахиезер Н. И., Глазман И. М., 1950, Теория линейных операторов в гильбертовом пространстве. Гостехиздат, М.

* Бабкин В. Ф., 1971, Метод универсального кодирования источника независимых сообщений неэкспоненциальной трудности. Пробл. перед. инф., **7**, 4, 13—21.

* Бакут Л. А., 1959, К теории корректирующих кодов с произвольным основанием. Науч. докл. высш. школы, Радиотехника и электроника, **1**, 26—36.

* Бассалыго Л. А., 1965, Новые верхние границы для кодов, исправляющих ошибки. Пробл. перед. инф., **1**, 4, 41—44.

* Башарин Г. П., 1959, О статистической оценке энтропии последовательности независимых случайных величин. Теор. вероятн. и ее применен., **4**, 3, 361—364.

Бергер, Мандельброт (Berger J. M., Mandelbrot B. M.), 1963, A New Model for Error Clustering in Telephone Circuits. *IBM J. Res. and Dev.* **7**, 224—236.

Берлекэмп (Berlekamp E. R.), 1964, Note of Recurrent Codes. IEEE Trans Inform. Theory, **IT-10**, 257—258.

Берлекэмп (Berlekamp E. R.), 1967, Nonbinary BCH Decoding. IEEE Int. Symp. on Inform. Theory, San Remo, Italy. (См. также гл. 7 и 10 книги Берлекэмпа, 1968).

Берлекэмп (Berlekamp E. R.), 1968, Algebraic Coding Theory. McGraw-Hill, New York. (Русский перевод: Берлекэмп Э., Алгебраическая теория кодирования. «Мир», М., 1971.)

* Берман С. Д., 1967 а, К теории групповых кодов. Кибернетика, 1, 31—39.

* Берман С. Д., 1967 б, Полупростые циклические и абелевы коды. Кибернетика, 3, 21—30.

Биллингслей (Billingsley P.), 1965, Ergodic Theory and Information. Wiley, New York. (Русский перевод: Биллингслей П., *Эргодическая теория и информация*. «Мир», М., 1969.)

Биркгоф и Маклейн (Birkhoff G., Mac Lane S.), 1941, A Survey of Modern Algebra. Macmillan, New York.

Блекуэлл (Blakwell D.), 1957, The Entropy of Functions of Finite-state Markov Chains. Trans. of the First Prague Conf. on Inform. Theory, Statist. Decision Funct. and Random Processes, Publishing House of the Czechoclovak Academy of Sciences, Prague, 13—20.

Блекуэлл (Blakwell D.), 1961, Exponential Error Bounds for Finite State Channels. Proc. Fourth Berkeley Symp. on Math. Stat. and Prob., Univ. of California Press, Berkeley, Calif., **1**, 57—63.

---

* Звездочкой отмечены публикации на русском языке, добавленные редакторами. Как правило, включены работы, непосредственно примыкающие к содержанию книги и относящиеся к публикациям до 1969 г., т. е. к моменту выхода книги.

Блекуэлл, Брейман, Томасян (Blakwell D., Breiman L., Thomasian A. J.), 1958, Proof of Shannon's Transmission Theorem for Finite-state Indecomposable Channels. Ann. Math. Stat, **29**, 1209—1220. (Русский перевод: Блекуэлл Д., Брейман Л., Томасян А., Доказательство теоремы Шеннона о передаче информации для неразложимых каналов с конечным числом состояний. Сб. Математика, 4 : 5, 1960, 123—135.)

Блекуэлл, Брейман, Томасян (Blakwell D., Breiman L., Thomasian A. J.), 1959, The Capacity of a Class of Channels. Ann. Math. Stat, **30**, 1229—1241.

Блекуэлл, Гиршик (Blakwell D., Girshick M. A.), 1954, Theory of Games and Statistical Decisions. Willey, New York. (Русский перевод: Блекуэлл Д., Гиршик М., Теория игр и статистических решений. ИЛ, М., 1958.)

Блюстейн, Жордан (Bluestein G., Jordan K. L), 1963, An Investigation of the Fano Sequential Decoding Algorithm by Computer Simulation. MIT-Lincoln Laboratory, Group Rept. 62G-5.

* Бородин Л. Ф., 1968, Введение в теорию помехоустойчивого кодирования. «Советское радио», М.

Боуз, Рой-Чоудхури (Bose R. C., Ray-Chaudhuri D. K.), 1960 а, On a Glass of Error Correcting Binary Group Codes. Inform. and Control, **3**, 68—79. (Русский перевод: Боуз Р. К., Рой-Чоудхури Д. К., Об одном классе двоичных групповых кодов с исправлением ошибок. Кибернетический сб., вып. 2, ИЛ, М., 1961, 83—94).

Боуз, Рой-Чоудхури (Bose R. C., Ray-Chaudhuri D. K.), 1960 б, Further Results on Error Correcting Binary Group Codes. Inform. and Control, **3**, 279—290. (Русский перевод: Боуз Р. К., Рой-Чоудхури Д. К., Дальнейшие результаты относительно двоичных групповых кодов с исправлением ошибок. Кибернетический сб., вып. 6, ИЛ, М., 1963, 7—19.)

* Бояринов И. М., 1970, Класс циклических кодов, исправляющих пакеты ошибок без пропусков. Пробл. перед. инф., **6**, 4, 91—93.

Брейман (Breiman L.), 1957, The Individual Ergodic Theorem of Information Theory. Ann Math. Stat, **28**, 809—811; исправления к этой статье, 1960, Ann Math. Stat., **31**, 809—810.

Бьюк (Buck R. C.), 1956, Advanced Calculus. McGraw-Hill, New York.

Вагнер (Wagner T. J.), 1968, A Coding Theorem for Abstract Memoryless Channels. Неопубликованные заметки.

Вайнер (Wyner A. D.), 1966, Capacity of the Band-limited Gaussian Channel. Bell System Tech. J., **45**, 359—395.

Вайнер, Эш (Wyner A. D., Ash R. B.), 1963, Analysis of Recurrent Codes. IEEE Trans. Inform. Theory, **IT-9**, 143—156. (Русский перевод: Вайнер Э., Эш Р., Анализ рекуррентных кодов. Кибернетический сб., новая серия, вып. 5, «Мир», М., 1968.)

Вальд (Wald A.), 1947, Sequential Analysis. Wiley, New York. (Русский перевод: Вальд А., Последовательный анализ. Физматгиз, М., 1960).

Ван Трис (Van Trees H. L.), 1965, Comparison of Optimum Angle Modulation Systems and Rate—Distortion Bounds. Proc. IEEE, **53**, 2123—2124.

* Васильев А. М., 1962, О негрупповых плотно упакованных кодах. Пробл. кибернетики, вып. 8, Физматгиз, М., 337—339.

Варшамов Р. Р., 1957, Оценка числа сигнала в кодах с коррекцией ошибок. Докл. Акад. наук СССР, **117**, 5, 739—741.

* Варшамов Р. Р., 1959, О методе линейного кодирования с исправлением ошибок в передаваемых сигналах. Труды научно-технического общества радиотехники и электросвязи им. А. С. Попова, 3, 43—48.

Винер (Wiener N.), 1949, Extrapolation, Interpolation, and Smoothing of Stationary Time Series. MIT Press, Cambridge, Mass., and Wiley, New York.

Витерби (Viterbi A. J.), 1961, On Coded Phase-Coherent Communications. IRE Trans. Space Elect. and Telemetry, **SET-7**, 3—14.

Витерби (Viterbi A. J.), 1967 а, Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. IEEE Trans. Inform. Theory, **IT-13**, 414—422. (Русский перевод: Витерби А., Границы ошибок для сверточных кодов и асимптотически оптимальный алгоритм декодирования. Сб. Некоторые вопросы теории кодирования, «Мир», М., 1970, 142—165).

Витерби (Viterbi A. J.), 1967 б, Performance of an M-ary Orthogonal Communication System Using Stationary Stochastic Signals. IEEE Trans. Inform. Theory, **IT-13**, 414—422.

Возенкрафт (Wozencraft J. M.), 1957, Sequential Decoding for Reliable Communication. 1957 National IRE Convention Record, **5**, part 2, 11—25.

Возенкрафт, Джекобс (Wozencraft J. M., Jacobs I. M.), 1965, Principles of Communication Engineering. Wiley, New York. (Русский перевод: Возенкрафт Дж., Джекобс И., Теоретические основы техники связи. «Мир», М., 1969.)

Возенкрафт, Рейффен (Wozencraft J. M., Reiffen B.), 1961, Sequential Decoding. MIT Press, Cambridge, Mass., and Wiley, New York. (Русский перевод: Возенкрафт Дж., Рейффен Б., Последовательное декодирование. ИЛ, М., 1963).

Возенкрафт, Хорстейн (Wozencraft J. M., Horstein M.), 1961, Coding for Two-way Channels. MIT Research Lab. of Electronics Tech. Rept 383, Cambridge, Mass.

Вольфовиц (Wolfowitz J.), 1957, The Coding of Messages Subject to Chance Errors. Illinois J. of Math., **1**, 591—606.

Вольфовиц (Wolfowitz J.), 1961, Coding Theorems of Information Theory. Springer—Verlag and Prentice-Hall, Englewood Cliffs, N. J. (Русский перевод: Вольфовиц Дж., Теоремы кодирования теории информации. «Мир», М., 1967.)

* Габидулин Э. М., 1967, Границы для вероятности ошибки декодирования при использовании линейных кодов в каналах без памяти. Пробл. перед. инф., **3**, 2, 55—62.

* Габидулин Э. М., 1969, Об оценках вероятности ошибки для некоторых каналов с памятью. Пробл. перед. инф., **5**, 1, 40—46.

Галлагер (Gallager R. G.), 1962, Class Notes Distributed for Courses 6.575, 6.626.MIT, Cambridge, Mass.

Галлагер (Gallager R. G.), 1964, Information Theory. Chapter 4 *in* Mathematics of Physics and Chemistry, H. Margenau and G. M. Murphy (Eds.), Van Nostrand, Princeton N. J., Vol. 2.

Галлагер (Gallager R. G.), 1965 а, A Simple Derivation of the Coding Theorem and some Applications. IEEE Trans. Inform. Theory, **IT-11**, 3—18. (Русский перевод: Галлагер Р. Г. Простой вывод теоремы кодирования и некоторые применения. Кибернетический сб., новая серия, вып. 3, «Мир», М., 1966, 50—90.)

Галлагер (Gallager R. G.), 1965 б, Lower Bounds on the Tails of Probability Distributions. MIT Research Lab. of Electronics, **QPR 77**, 277—291.

Гантмахер Ф. Р., 1967, Теория матриц. «Наука», М.

* Гельфанд И. М., Виленкин Н. Я., 1961, Обобщенные функции, IV. Некоторые применения гармонического анализа. Оснащенные гильбертовы пространства. Физматгиз, М.

* Гельфанд И. М., Колмогоров А. Н., Яглом А. М., 1956, К общему определению количества информации. Докл. Акад. наук СССР, **111**, 4, 745—748.

* Гельфанд И. М., Колмогоров А. Н., Яглом А. М., 1958, Количество информации и энтропия для непрерывных распределений. Труды 3-го Всесоюзного математического съезда, Изд. АН СССР, 3, 521—531.

Гельфанд И. М., Яглом А. М. 1957, О вычислении количества информации о случайной функции, содержащейся в другой такой же функции. Успехи матем. наук, 12, 1, 3—52.

Гилберт (Gilbert E. N.), 1952, A Comparison of Signalling Alphabets. Bell System Tech. J., **31**, 504—522.

Гилберт (Gilbert E. N.), 1960, Capacity of Burst Noise Channel. Bell System Tech. J., **39**, 1253—1256. (Русский перевод: Гилберт Э.Н., Пропускная способность канала с пакетами ошибок. Кибернетический сб., вып. 9, «Мир», М., 1964, 9—12.)

* Глебский Ю. В., 1962, Кодирование с помощью автоматов с конечной внутренней памятью. Пробл. кибернетики, вып. 7, Физматгиз, М., 127—150.

Гоблик (Goblick T. J., Jr.), 1965, Theoretical Limitations on the Transmission of Date from Analogue Sources. IEEE Trans. Inform. Theory, **IT-11**, 558—567.

Гоблик, Холзингер (Goblick T. J., Holsinger J. L.), 1967, Analog Sourse Digitization: A Comparasion of Theory and Practice. IEEE Trans. Inform. Theory, **IT-13**, 323—326.

Голей (Golay M. J. E.), 1954, Binary Coding. IRE Trans. Inform. Theory, **PGIT-4**, 23—28.

Голомб, Гордон, Велч (Golomb S. W., Cordon B., Welch L. R.), 1958, Comma Free Codes. Canad. J. Math., **10**, No 2, 202—209 (Русский перевод: Голомб С. Х., Гордон Б., Велч Л. Р., Коды без запятой. Кибернетический сб., вып. 5, ИЛ, М., 1962, 33—41).

Гоппа В. Д., 1970, Новый класс линейных корректирующих кодов. Пробл. перед. инф., **6**, 3, 24—30.

Гоппа В. Д., 1971, Рациональное представление кодов и (L, g)-коды Пробл. перед. инф., **7**, 3, 41—49.

Гренандер, Сеге (Grenander U., Szego G.), 1958, Toeplitz Forms and their Applications. Univ. of Callifornia Press, Berkeley, Calif. (Русский перевод: Гренандер У., Сеге Г., Тёплицевы формы и их приложения. ИЛ., М., 1961.)

Греттенберг (Grettenberg T. L.), 1968, Exponential Error Bounds for Incoherent Orthogonal Signals. IEEE Trans. Inform. Theory, **IT-14**, 163—164.

Давенпорт, Рут (Davenport W. B., Root W. L.), 1958, Random Signals and Noise. McGraw-Hill, New York. (Русский перевод: Давенпорт В. Б., Рут В. Л., Введение в теорию случайных сигналов и шумов. ИЛ, М., 1960.)

* Демидович Н. Б., 1962, К теории групповых кодов. Пробл. кибернетики, вып. 5, Физматгиз, М., 105—123.

Джекобс (Jacobs I. M.), 1963, The Asymptotic Behavior of Incoherent M-ary Communication Systems. Proc. IRE, **51**, 251—252.

Джекобс, Берлекэмп (Jacobs I. M., Berlekamp E. R.), 1967, A Lower Bound to the Distribution of Computation for Sequential Decoding. IEEE Trans. Inform. Theory, **IT-13**, 167—174. (Русский перевод: Джекобс И., Берлекэмп Е., Нижняя граница распределения количества вычислений для последовательного декодирования. Сб. Некоторые вопросы теории кодирования, «Мир», М., 1970, 230—248.)

Джелинек (Jelinek F.), 1968 а, Evaluation of Expurgated Bound Exponents. IEEE Trans. Inform. Theory, **IT-14**, 501—505.

Джелинек (Jelinek F.), 1968 б, Probabilistic Information Theory. McGraw-Hill, New York.

* Добрушин Р. Л., 1958 а, Передача информации по каналу с обратной связью. Теория вероятн. и ее примен., **3**, 4, 395—419.

* Добрушин Р. Л., 1958 б, Упрощенный метод экспериментальной оценки энтропии стационарной последовательности. Теория вероятн. и ее примен., **3**, 4, 462—464.

Добрушин Р. Л., 1959, Общая формулировка основной теоремы Шеннона в теории информации. Успехи матем. наук, **14**, 6, 3—104.

* Добрушин Р. Л., 1960 а, Асимптотика вероятностей ошибок при передаче информации по каналу без памяти с симметрической матрицей вероятностей перехода. Докл. Акад. наук СССР, **133**, 2, 265—268.

* Добрушин Р. Л., 1960 б, Предельный переход под знаком информации и энтропии. Теория вероятн. и ее примен., **5**, 1, 29—37.

* Добрушин Р. Л., 1961, Математические вопросы шенноновской теории оптимального кодирования информации. Пробл. перед. инф., вып. 10, Изд. Акад. наук СССР, 63—107.

* Добрушин Р. Л., 1962 а, Оптимальные бинарные коды для малых скоростей передачи информации. Теория вероятн. и ее примен., 7, 2, 203—213.

* Добрушин Р. Л., 1962 б, Асимптотические оценки вероятности ошибки при передаче сообщений по дискретному каналу связи без памяти с симметрической матрицей вероятности перехода. Теория вероятн. и ее примен., 7, 3, 283—311.

* Добрушин Р. Л., 1963, Асимптотическая оптимальность групповых и систематических кодов для некоторых каналов. Теория вероятн. и ее примен., **8**, 1, 52—66.

* Добрушин Р. Л., 1964, По поводу последовательного декодирования методом Возенкрафта—Рейффена. Пробл. кибернетики, вып. 12, Физматгиз, М., 113—123.

Дуб (Doob J. L.), 1953, Stochastic Processes. Wiley, New York. (Русский перевод: Дуб Дж. Л., Вероятностные процессы. ИЛ, М., 1956.)

* Дьячков А. Г., 1970, Асимптотика вероятности ошибки при передаче по каналу с белым гауссовским шумом и бесшумной мгновенной обратной связью. Пробл. перед. информ., **6**, 1, 33—44.

* Дьячков А. Г., Пинскер М. С., 1971, Об оптимальном линейном методе передачи по гауссовскому стационарному каналу без памяти с полной обратной связью. Пробл. перед. инф., 7, 2, 38—46.

* Дынькин В. Н., Тененгольц Г. И., 1968, Об одном классе циклических кодов с мажоритарной схемой декодирования. Пробл. перед. инф., **5**, 1, 3—15.

* Егармин В. М., 1969, Нижние и верхние границы вероятности ошибки декодирования для дискретных каналов. Пробл. перед. инф., **5**, 1, 23—39.

* Ерохин В. Д., 1958, ε-энтропия дискретного случайного объекта. Теория вероятн. и ее примен., **3**, 1, 103—107.

* Звонкин А. К., Левин Л. А., 1970, Сложность конечных объектов и обоснование понятия информации и случайности с помощью теории алгоритмов. Успехи матем. наук, **25**, 6, 85—127.

Зеттерберг (Zetterberg L. H.), 1961, Data Transmission over a Noisy Caussian Channel. Trans. Roy. Inst. Technol., Stockholm, No 184.

* Зигангиров К. Ш., 1966, Некоторые последовательные процедуры декодирования. Пробл. перед. инф., **2**, 4, 13—25.

* Зигангиров К. Ш., 1967, Передача по гауссовскому каналу с обратной связью. Проб. перед. инф., **3**, 2, 98—101.

* Зигангиров К. Ш., 1968 а, Процедура последовательного декодирования с экспонентой вероятности ошибки, даваемой случайным кодированием. Пробл. перед. инф., **4**, 2, 83—85.

* Зигангиров К. Ш., 1968 б, Передача сообщений по двоичному симметричному каналу с бесшумной обратной связью (случайное время передачи). Пробл. перед. инф., **4**, 3, 38—47.

* Зигангиров К. Ш., 1971, Последовательная передача от источника с переменной скоростью. Пробл. перед. инф., 7, 2, 114—118.

* Зигангиров К. Ш., Овчинников В. В., 1971, Последовательное декодированне в канале с пакетами ошибок. Пробл. перед. инф., 7, 1, 29—37.

* Зигангиров К. Ш., Пинскер М. С., Цыбаков Б. С., 1967, Последовательное декодирование в непрерывном канале. Пробл. перед. инф., **3**, 4, 5—17.

* Зяблов В. В., 1971, Оценка сложности построения двоичных линейных каскадных кодов. Пробл. перед. инф., 7, 1, 5—13.

Ивадаре (Iwadare Y.), 1967, Simple and Efficient Procedures of Burst-error Corection. Ph. D. Thesis, Univ. of Tokyo.

Истман (Eastman W. L.), 1965, On the Construction of Comma-free Codes. IEEE Trans. Inform. Theory, **IT-11**, 263—267. (Русский перевод: Истман В. Л., О построении кодов без запятой. Кибернетический сб., новая серия, вып. 3, «Мир», М., 1966, 91—100.)

Казами (Kasami T.), 1963, Optimum Shortened Cyclic Codes for Burst—error Correction. IEEE Trans. Inform. Theory, **IT-9**, 105—109.

Кайлат (Kailath T.), 1967, A Projection Method for Signal Detection in Colored Gaussian Noise. IEEE Trans. Inform. Theory, **IT-13**, 441—447.

Кармайкл (Carmichael R. D.), 1956, Introduction to the Theory of Groups of Finite Order. Dover, New York.

Каруш (Karush J.), 1961, A Simple Proof of Inequality of McMillan. IRE Trans. Inform. Theory, **IT-7**, 118.

Кац, Мардок, Сеге (Kac M., Murdock W. L., Szego G.), 1953, On the Eigenvalues of Certain Hermitian Forms. J. Rat. Mech. and Anal., **2**, 767—800.

Келли (Kelly J. L.), 1956, A New Interpretation of Information Rate. Bell System Tech. J., **35**, 917—926.

Келли, Рид, Рут (Kelly E. J., Reed I. S., Root W. L.), 1960, The Detection of Radar Echoes in Noise. J. Soc. Ind. Appl. Math., **8**, 309—341.

Кендалл, Рид (Kendall W. B., Reed I. S.), 1962, Path Invariant Comma-free Codes. IRE Trans. Inform. Theory, **IT-8**, 350—355.

Кеннеди (Kennedy R. S.), 1963, Finite State Binary Symmetric Channels. Sc. D. Thesis, Dept. of E. E., MIT, Cambridge, Mass.

Кеннеди (Kennedy R. S.), 1964, Performance Limitations of Dispersive Fading Channels. International Conference on Microwaves, Circut Theory, and Information Theory, Tokyo; abstract in IEEE Trans. Inform. Theory, **IT-10**, 398.

Кеннеди (Kennedy R. S.), 1969, Fading Dispesive Communication Channels. Wiley, New York. (Русский перевод: Кеннеди Р. Каналы связи с замираниями и рассеянием. «Советское радио», М., 1973).

* Кириллов Н. Е., 1959, Коды с разделительными знаками. Электросвязь, 6, 5—10.

* Кислицын С. С., 1962, Средняя длина двоичного кода с минимальной избыточностью в случае, когда вероятности декодируемых символов близки друг к другу. Теория вероятн. и ее примен., **7**, 3, 342—343.

Кокс, Миллер (Cox D. R., Miller H. D.), 1965, The Theory of Stochastic Processes. Wiley, New York.

Коленберг (Kohlenberg A.), 1965, Randon and Burst Error Control. First IEEE Annual Communications Convention, Boulder, Colo, June 7—9.

* Колесник В. Д., Мирончиков Е. Т., 1968 а, Декодирование циклических кодов. «Связь», М.

* Колесник В. Д., Мирончиков Е. Т., 1968 б, Циклические коды Рида—Малера и их декодирование. Пробл. перед. инф., **4**, 4, 20—25.

Колмогоров А. Н., 1941, Интерполяция и экстраполяция случайных последовательностей. Изв. Акад. наук СССР. Сер. Мат., 5, 3—14.

Колмогоров А. Н., 1956, Теория передачи информации. Сб. Сессия АН СССР по научным проблемам автоматизации производства, Пленарные заседания, М., АН СССР, 66—99.

* Колмогоров А. Н., 1965, Три подхода к определению понятия количества информации. Пробл. перед. инф., 1, 1, 3—11.

* Колмогоров А. Н., 1969. К логическим основам теории информации и теории вероятностей. Пробл. перед. инф., **5**, 3, 3—7.

Котельников В. А., 1956, Теория потенциальной помехоустойчивости. Госэнергоиздат, М.

Коутц (Kotz S.), 1965, Recent Developments in Information Theory, J. Applied Probability.

* Кошелев В. Н., 1966 а, Оценка сложности последовательного декодирования случайных древовидных кодов. Пробл. перед. инф., 2, 2, 12—28.

* Кошелев В. Н., 1966 б, Экспериментальная оценка числа операций при последовательном декодировании с одним порогом. Пробл. перед. инф., **2**, 4, 80—83.

* Кричевский Р. К., 1988, Связь между избыточностью кодирования и достоверностью сведений об источнике. Пробл. перед. инф., **4**, 3, 48—57.

Крамер (Cramer H.), 1937, Random Variables and Probability Distributions. *Cambridge Tracts in Mathematics No 36, Cambridge, England.* (Русский перевод: Крамер Г., Случайные величины и распределения вероятностей. ИЛ, М., 1947).

Крафт (Kraft L. C.), 1949, A Device for Quantizing, Grouping and Coding Amplitude Modulated Pulses, M. S. Thesis, Dept. of E. E., MIT, Cambridge, Mass.

Курант, Гильберт (Courant R., Hilbert D.), 1959, Methods of Mathematical Physics. Vol I. Interscience, New York. (Русский перевод: Курант Р., Гильберт Д., Методы математической физики. Гостехиздат, М.—Л., 1951).

Кун, Тюкер (Kuhn H. W., Tucker A. W.), 1951, Nonlinear Programming, Proc. 2-nd Berkeley Symposium on Math. Stat. and Prob., Univ of Calif. Press, Berkeley, 481—492.

Кэмперман (Kemperman J. H. B.), 1962, Studies in Coding Theory. Неопубликованные заметки, Rochester.

Ландау, Поллак (Landau H. J., Pollak H. O.), 1961, Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-II. Bell System Tech. J., **40**, 65—84.

Ландау, Поллак (Landau H. J., Pollak H. O.), 1962, Prolate Spheroidal Wave Functions, Fourier Analisis, and Uncertainty-III. Bell System Tech. J., **41**, 1295—1336.

* Левенштейн В. И., 1960, Об одном классе систематических кодов. Докл. Акад. наук СССР, 131, 5, 1011—1014.

* Левенштейн В. И., 1961 а, О некоторых свойствах кодовых систем. Докл. Акад. наук СССР, 140, 6, 1274—1277.

* Левенштейн В. И., 1961 б, Применение матриц Адамара к одной задаче кодирования. Пробл. кибернетики, вып. 5, Физматгиз, М., 123—136.

* Левенштейн В. И., 1961 в, Самонастраивающиеся автоматы для декодирования сообщений. Докл. Акад. наук СССР, **141**, 6, 1320—1323.

* Левенштейн В. И., 1965, Двоичные коды с исправлением выпадений и вставок символа 1. Пробл. перед. инф., **1**, 1, 12—25.

* Левенштейн В. И., 1969, Оценка для кодов, обспечивающих исправление ошибок и синхронизацию. Пробл. перед. инф., **5**, 2, 3—13.

* Левенштейн В. И., 1970, О максимальном числе слов в кодах без перекрытий. Пробл. перед. инф., **6**, 4, 88—90.

* Левенштейн В. И., 1971 а, Об одном методе построения квазилинейных кодов, обеспечивающих синхронизацию при наличии ошибок. Пробл. перед. инф., 7, 3, 30—40.

* Левенштейн В. И., 1971 б, О верхних оценках для кодов с фиксированным весом векторов. Пробл. перед. инф., **7**, 4, 3—12.

* Леонтьев В. К., 1968, Об одной гипотезе о кодах Боуза-Чоудхури. Пробл. перед. инф, 4, 1, 83—85.

* Либкинд Л. М., 1965, ε-энтропия дискретных источников сообщений. Пробл. перед. инф., **1**, 3, 48—55.

* Либкинд Л. М., 1967, Двусторонние дискретные каналы связи без памяти. Пробл. перед. инф., **3**, 2, 37—46.

* Линьков Ю. Н., 1965, Вычисление ε-энтропии случайных величин при малых **1**. Пробл. перед. инф., **1**, 2, 18—26.

* Линьков Ю. Н., 1967, Об асимптотических оценках ε-энтропии случайных величин. Кибернетика. Семинар, вып. 1, Киев, 3—9.

* Линьков Ю. Н., 1971, Эпсилон-энтропия случайных процессов с непрерывным временем и дискретным фазовым пространством. Пробл. перед. инф., 7, 2, 16—25.

Лоэв (Loeve M.), 1955, Probability Theory. Van Nostrand, Princeton, N. J. (Русский перевод: Лоэв М., Теория вероятностей. ИЛ, М., 1962).

* Любич Ю. И., 1962, Замечание о пропускной способности дискретного канала связи без шумов. Успехи матем. наук, **17**, 1, 191—198.

Макмиллан (McMillan B.), 1953, The Basic Theorems of Information Theory. Ann Math. Stat., **24**, 196—219.

Макмиллан (McMillan B.), 1956, Two Inequalities Implied by Unique Decipherability, IRE Trans. Inform. Theory, **IT-2**, 115—116. (Русский перевод: Макмиллан Б., Два неравенства, обусловленные однозначностью расшифрования Кибернетический сб., вып. 3, ИЛ, М., 1961, 88—92.)

Макс (Max J.), 1960, Quantizing for Minimum Distortion. IRE Trans. Inform. Theory, **IT-6**, 7—12.

* Марков А. А., 1960, Об алфавитном кодировании I, Докл. Акад. наук СССР, **132**, 3, 521—523.

* Марков А. А., 1961, Об алфавитном кодировании II. Докл. Акад. наук СССР, **139**, 3, 560—561.

* Марков А. А., 1962, Нерекуррентное кодирование. Пробл. кибернетики, вып. 8, Физматгиз, М., 169—186.

* Марков А. А., 1963, Условие полноты для неравномерных кодов. Пробл. кибернетики, вып. 9, Физматгиз, М., 327—332.

* Мартон К., 1971, Асимптотика эпсилон-энтропии дискретных стационарных процессов. Пробл. перед. инф., **7**, 2, 3—15.

* Мартон К., 1972, Информация и информационная устойчивость эргодических источников. Пробл. перед. инф., **8**, 3, 3—8.

* Марчуков А. С., 1968, О суммировании произведений кодов. Пробл. перед. инф., **4**, 2, 11—20.

Метцнер, Морган (Metzner J. J., Morgan K. C.), 1960, Coded Feedback Communication Systems, National Electronics Conference, Chicago, 111., October.

Месси (Massey J. L.), 1963, Threshold Decoding. MIT Press, Cambridge, Mass. (Русский перевод: Месси Дж. Пороговое декодирование. «Мир», М., 1964.)

Месси (Massey J. L.), 1965, Implementation of Burst Correcting Convolutional Codes. IEEE Trans. Inform. Theory, **IT-11**, 416—422.

Месси (Messey J. L.), 1969, Feedback Shift Register Synthesis and BCH Decoding. IEEE Trans. Inform. Theory, **IT-15**, 122—128.

Месси, Лью (Massey J. L., Liu R. W.), 1964, Application of Lyapunov's direct method to the Error Propagation Effect in Convolutional Codes. IEEE Trans. Inform. Theory, **IT-10**, 248—250.

* Мешковский К. А., Кириллов Н. Е., 1966, Кодирование в технике связи. «Связь», М.

* Молчанов С. А., 1967. О передаче конечного числа сообщений по двоичному несимметрическому каналу. Пробл. перед. инф., **3**, 1, 10—17.

* Морозов В. А., 1967, О последовательном декодировании, оптимальном по вероятности ошибки. Изв. Акад. наук СССР, Техн. кибернетика, **3**, 164—167.

* Морозов В. А., 1969, Двухэтапная процедура последовательного декодирования. Изв. Акад. наук СССР, Техн. кибернетика, 4, 118—127.

* Нгуен Дан Те, 1970, Вероятность ошибки при передаче по гауссовскому каналу с бесшумной обратной связью. Пробл. перед. инф., **6**, 1, 25—32.

* Овсеевич И. А., 1963, Пропускная способность многопутевой системы. Пробл. перед. инф., вып. 14, Изд. Акад. наук СССР, 43—58.

* Овсеевич И. А., 1970, Оптимальная передача гауссовского сообщения по каналу с белым гауссовским шумом при наличии обратной связи. Пробл. перед. инф., **6**, 3, 3—14.

* Овсеевич И. А., Пинскер М. С., 1959, Скорость передачи информации, пропускная способность многопутевой системы и прием по методу линейно операторного преобразования. Радиотехника, **14**, 3, 9—21.

* Овсеевич И. А., Пинскер М. С., 1960, Пропускная способность каналов с общим и селективным замиранием. Радиотехника, **15**, 12, 3—9.

* Овсеевич И. А., Пинскер М. С., 1961, О пропускной способности многопутевой системы. Изв. Акад. наук СССР, Энергетика и автоматика, 4, 208—210.

* Овсеевич И. А., Пинскер М. С., 1965, Согласование источника сообщений с каналом методом перестановки спектров. Изв. Акад. наук СССР, Техн. кибернетика, 2, 81—87.

* Оганесян С. Ш., Ягджян В. Г., 1970, Весовой спектр для некоторых классов корректирующих циклических кодов. Пробл. перед. инф., **6**, 3, 31—37.

Отт (Ott G.), 1967, Compact Encoding of Stationary Markov Sources IEEE Trans. Inform. Theory, **IT-13**, 82—86.

* Овчинников В. В., 1972, Моделирование последовательного декодирования в двоичном симметричном канале с памятью. Пробл. перед. инф., **8**, 2, 103—105.

Пилк (Pilc R. J.), 1967, Coding Theorems for Discrete Source—Channel Pairs. Ph. D. Thesis, Dept. of E. E., MIT, Cambridge, Mass.

Пинкстон (Pinkston J. T.), 1967, Encoding Independent Sample Information Sources. MIT Research. Lab. of Electronics, Tech. Rept. 462.

* Пинскер М. С., 1954, Количество информации о гауссовском случайном стационарном процессе, содержащейся во втором процессе стационарно с ним связанном. Докл. Акад. наук СССР, **99**, 2, 213—216.

* Пинскер М. С., 1956, Вычисление скорости создания сообщений стационарным случайным процессом и пропускной способности стационарного канала. Докл. Акад. наук СССР, **111**, 4, 753—756.

Пинскер М. С., 1957, Вычисление и оценка количества информации пропускной способности канала и скорости создания сообщений по вторым моментам распределений. Канд. диссертация, М.

Пинскер М. С., 1960, Информация и информационная устойчивость случайных величин и процессов. Пробл. перед. инф., вып. 7, Изд. Акад. наук СССР.

* Пинскер М. С., 1963 а, Источники сообщений. Пробл. перед. инф., вып. 14, Изд. Акад. наук СССР, 5—20.

* Пинскер М. С., 1963 б, Гауссовские источники. Пробл. перед. инф., вып. 14, Изд. Акад. наук СССР, 59—100.

* Пинскер М. С., 1965, О сложности декодирования. Пробл. перед. инф., **1**, 1, 113—117.

* Пинскер М. С., Шевердяев А. Ю., 1970, Пропускная способность с нулевой ошибкой и стиранием. Пробл. перед. инф., **6**, 1, 20—24.

Пирс (Pierce J. N.), 1961, Theoretical Limitations on Frequency and Time Diversity for Fading Binary Transmission. IRE Trans. on Communication Systems, **CS-9**, 186—187.

Питерсон (Peterson W. W.), 1960, Encoding and Error—Correction Procedures for the Bose—Chaudhuri Codes. IRE Trans. Inform. Theory, **IT-6**, 459—470. (Русский перевод: Питерсон У. У., Кодирование и исправление ошибок для кодов Боуза—Чоудхури. Кибернетический сб., вып. 6, ИЛ, М., 25—54, 1963).

Питерсон (Peterson W. W.), 1961, Error Correcting Codes. MIT Press Cambridge, Mass. and Wiley, New York. (Русский перевод: Питерсон У. У., Коды, исправляющие ошибки. ИЛ., М., 1964.)

Питерсон, Месси (Peterson W. W., Massey J. L.), 1963, Report on Progress in Information Theory in USA, 1960—1963, Coding Theory. IEEE Trans. Inform. Theory, **IT-9**, 223—229. (Русский перевод: Питерсон У. У., Обсуждение состояния теории кодирования и вопросов практического использования кодов. Кибернетический сб., вып. 9, «Мир», М., 1964, 90—108).

Плоткин (Plotkin M.), 1951, Binary Codes with Specified Minimum Distance. IRE Trans. Inform. Theory, **IT-6**, 445—450, 1960. (Русский перевод: Плоткин М., Двоичные коды с заданным минимальным расстоянием. Кибернетический сб., вып. 7, ИЛ, М., 1963.)

* Попов О. В., Турин В. Я., Игельник Б. М., 1967, Об оценке верности передачи дискретных сообщений по каналам с пакетами ошибок или стираний. Пробл. перед. инф., **3**, 4, 37—48.

Прейндж (Prange E.), 1957, Cyclic Error-Correcting Codes in Two Symbols. ARCRC-TN-57-103, Air Force Cambridge Res. Center, Cambridge, Mass.

Прейндж (Prange E.), 1958, Some Cyclic Error—Correcting Codes with Simple Decoding Algorithms. AFCRC-TN-58-156, Air Force Cambridge Res. Center, Bedford, Mass.

* Прелов В. В., 1966, Об асимптотике пропускной способности некоторых каналов связи. Пробл. перед. инф., **2**, 1, 14—27.

* Прелов В. В., 1967, *Об асимптотике пропускной способности* каналов со счетным алфавитом с приложением к асинхронным каналам. Пробл. перед. инф., **3**, 2, 22—36.

* Прелов В. В., 1969, Асимптотика пропускной способности канала с малым аддитивным шумом. Пробл. перед. инф., **5**, 2, 31—36.

* Прелов В. В., 1970, Асимптотика *пропускной способности* непрерывного канала с большим шумом. Пробл. перед. инф., **6**, 2, 40—57.

* Преснякова Г. В., Мирончиков Е. Т., 1969, О циклических $M(n, k)$-кодах. Пробл. перед. инф., **5**, 2, 19—22.

* Пятошин Ю. П., 1968, Некоторые свойства m-ичных систем связи с кодированием. Пробл. перед. инф., **4**, 1, 45—51.

* Радченко А. Н., Мирончиков Е. Т., 1961, Многотактные методы исправления одиночных и многократных близко расположенных ошибок в групповых кодах. Радиотехника и электроника, 11, 1805—1812.

Рид (Reed I. S.), 1954, A Class of Multiple-Frror-Correcting Codes and the Decoding Scheme. IRE Trans. Inform. Theory, **IT-4**, 38—49. (Русский перевод: Рид И. С., Класс кодов с исправлением нескольких ошибок и схема декодирования. Кибернетический сб., вып. 1, ИЛ, М., 1960, 189—205.)

Рид, Соломон (Reed I. S., Solomon G.), 1960, Polynomial Codes over Certain Finite Fields, J. Soc. Indust. Appl. Math., **8**, 300—304. (Русский перевод: Рид И. С., Соломон Г. Полиномиальные коды над некоторыми конечными полями. Кибернетический сб., вып. 7, ИЛ, М., 1963, 74—79.)

Рейффен (Reiffen B.), 1960, Sequential Encoding and Decoding for the Discrete Memoryless Channel. MIT Research Lab. of Electronics Tech. Rept. 374. (Русский перевод: Рейффен Б. Последовательное декодирование для каналов без памяти с дискретным входом, в кн. Возенкрафт Дж., Рейффен Б., Последовательное декодирование. ИЛ., М., 1963.)

Рейффен (Reiffen B.), 1963, A Note on Very Noisy Channels. Inform. and Control, **6**, 126—130.

Рейффен (Reiffen B.), 1966, A Per Letter Converse to the Channel Coding Theorem. IEEE Trans. Inform. Theory, **IT-12**, 475—480.

Рисс, Надь (Riesz F., Sz-Nagy B.), 1955, Functional Analisis. Ungar, New York. (Русский перевод: Рисс Ф., Надь Б., Лекции по функциональному анализу, ИЛ., М., 1954.)

Ричтерс (Richters J. S.), 1967, Communication over Fading Dispersive Channels. MIT Research Lab. of Electronics. Tech. Rept. 464.

* Сагалович Ю. Л., 1971, Сложность схемной реализаци алгоритма Рида и декодирование в автомате. Пробл. перед. инф., **7**, 1, 23—28.

Сакрисон (Sakrison D.), 1968, Communication Theory: Transmission of Waveforms and Digital Information, Wiley, New York.

* Самойленко С. И., 1966, Помехоустойчивое кодирование. «Наука», М.

Сардинас, Паттерсон (Sardinas A. A., Patterson G. W.), 1953, A Necessary and Sufficient Condition for the Unique Decomposition of Coded Messages. IRE Convention Record, Part 8, 104—108. (Русский перевод: Сардинас А. А., Паттерсон Дж., Необходимое и достаточное условие однозначного разложения закодированных сообщений. Кибернетический сб., вып. 3, ИЛ, М., 1961, 93—102.)

Севэдж (Savage J. E.), 1965, The Computation Problem with Sequential Decoding. MIT Research Lab. of Electronics, Tech. Rept. 439.

* Семаков Н. В., Зиновьев В. А., 1968, Эквидистантные q-ичные коды с максимальным расстоянием и разрешимые уравновешенные неполные блок-схемы. Пробл. перед. инф., **4**, 2, 3—10.

* Семаков Н. В., Зиновьев В. А., 1969 а, Совершенные и квазисовершенные равновесные коды. Пробл. перед. инф., **5**, 2, 14—18.

* Семаков Н. В., Зиновьев В. А., 1969 б, Равновесные коды и тактические конфигурации. Пробл. перед. инф., **5**, 3, 16—36.

* Семаков Н. В., Зайцев Г. В., Зиновьев В. А., 1969, Класс максимальных эквидистантных кодов. Пробл. перед. инф., **5**, 2, 84—87.

* Сидельников В. М., 1969, О некоторых k-значных псевдослучайных последовательностях и кодах, близких к оптимальным. Пробл. перед. инф., **5**. 1, 16—22.

* Сидельников В. М., 1971, О спектре весов двоичных кодов Боуза—Чоудхури—Хоквингема. Пробл. перед. инф., 7, 1, 14—22.

* Синай, 1959, Наименьшая ошибка и наилучший способ передачи стационарных сообщений при линейном кодировании декодировании в случае гауссовских каналов связи. Пробл. перед. инф., вып. 2, Изд. Акад. наук СССР, 40—48.

* Сифоров В. И., 1956, К теории идеального кодирования бинарной передачи. Радиотехника и электроника, **1**, 4, 407—417.

* Сифоров В. И., 1956, О помехоустойчивости систем с корректирующими кодами. Радиотехника и электроника, **1**, 2, 131—142.

* Сифоров В. И., 1958, О пропускной способности каналов связи с медленными случайными изменениями параметров. Научн. докл. высш. школы, Радиотехника и электроника, 1, 3—8.

Слепян (Slepian D.), 1956, A Class of Binary Signaling Alphabets. Bell System Tech. J., 35, 203—234. (Русский перевод: Слепян Д., Класс двоичных сигнальных алфавитов. Сб. Теория передачи сообщений, ИЛ, М., 1957, 82—114.)

Слепян (Slepian D.), 1963 Bounds of Communication. Bell System Tech. J., **42**, 681—707.

Слепян (Slepian D.), 1965, Some Asymptotic Expansions for Prolate Spheroidal Wave Functions. J. Math. and Phys., **44**, 99—140.

Слепян, Поллак (Slepian D., Pollak H. O.), 1961, Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty-I. Bell System Tech. J., **40**, 43—64.

Стиглитц (Stiglitz I. G.), 1966, Coding for a Class of Unknown Channels. IEEE Trans. Inform. Theory, **IT-12**, 189—195.

Стиглитц (Staglitz I. G.), 1967, A Coding Theorem for a Class of Unknown Channels. IEEE Trans. Inform. Theory, **IT-13**, 217—220.

* Стратанович Р. Л., 1967, Количество информации и энтропия отрезков стационарных гауссовских процессов. Пробл. перед. инф., **3**, 2, 3—21.

Титчмарш (Titchmarsh E. C.), 1937, Introduction to the Theory of Fourier Integrals. Oxford Univ. Press, London. (Русский перевод. Титчмарш Е., Введение в теорию интегралов Фурье. Гостехиздат, 1948.)

Томасян (Thomasian A. J.), 1960, An Elementary Proof of the AEP of Information Theory. Ann Math. Stat, **31**, 452—456.

Томасян (Thomasian A. J.), 1963, A Finite Criterion for Indecomposable Channels. Ann. Math. Stat, **34**, 337—338.

* Тылкин М. Е., 1960, О геометрии Хэмминга единичных кубов. Докл. Акад. наук СССР, **134**, 3, 1037—1040.

* Тылкин М. Е. 1962, О реализуемости матриц расстояний в единичных кубах. Пробл. кибернетики, вып. 7, Физматгиз, М., 31—42.

* Фадеев Д. К., 1956, К понятию энтропии конечной вероятностной схемы. Успехи матем. наук, **11**, 1, 227—231.

Файнстейн (Feinstein A.), 1954, A New Basic Theorem of Information Theory. IRE Trans. Inform. Theory, **PGIT-4**, 2—22.

Файнстейн (Feinstein A.), 1955, Error Bounds in Noisy Channels without Memory. IRE Trans. Information Theory, **IT-1**, 13—14.

Файнстейн (Feinstein A.), 1958, Foundations of Information Theory. McGraw-Hill, New York. (Русский перевод: Файнстейн А., Основы теории информации, ИЛ, М., 1960.)

Файр (Fire P.), 1959, A Class of Multiple-Frror-Correcting Binary Codes for Non-Independent Errors. Sylvania Report RSL-E-2, Sylvania Reconnaissance Systems Laboratory, Mountain View, Calif.

Фано (Fano R. M.), 1952, Class Notes for Transmission of Information. Course 6.574, MIT, Cambridge, Mass.

Фано (Fano R. M.), 1961, Transmission of Information. MIT Press, Cambridge, Mass. and Wiley, New York. (Русский перевод: Фано Р., Передача информации. Статистическая теория связи. «Мир», М., 1965.)

Фано (Fano R. M.), 1963, A Heuristic Discussion of Probabilistic Decoding. IEEE Trans. Inform. Theory, **IT-9**, 64—74. (Русский перевод: Фано Р. М., Эвристическое обсуждение вероятностного декодирования. Сб. Теория кодирования, «Мир», М., 1964, 166—198.)

Феллер (Feller W.), 1950, An Introduction to Probability Theory and its Applications, Vol 1. Wiley, New York (3 rd. ed, 1968). (Русский перевод: Феллер В., Введение в теорию вероятностей и ее приложения, т. 1., «Мир», М., 1967.)

Феллер (Feller W.), 1966, An Introduction to Probability Theory and its Applications, Vol. 2. Wiley, New York. (Русский перевод: Феллер В., Введение в теорию вероятностей и ее приложения. т. 2. «Мир», М., 1967.)

* Финк Л. М., 1970, Теория передачи дискретных сообщений. «Советское радио», М.

* Фитингоф Б. М., 1966, Оптимальное кодирование при неизвестной и меняющейся статистике сообщений. Пробл. перед. инф., **2**, 2, 3—11.

* Фитингоф Б. М., 1967, Сжатие дискретной информации. Пробл. перед. инф., **3**, 3, 28—36.

* Флейшман Б. С., 1963. Конструктивные методы оптимального кодирования для каналов с шумами. Изд. Акад. наук СССР.

Форни (Forney G. D.), 1965, On Decoding BCH Codes. IEEE Trans. Inform. Theory **IT-11**, 549—557.

Форни (Forney G. D.), 1967, Concatenated Codes. MIT Press, Cambridge, Mass. (Русский перевод: Форни Г. Д. Каскадные коды. «Мир», М., 1970.)

Форни (Forney G. D.), 1968, Exponential Error Bounds for Erasure, List, and Decision Feedback Schemes. IEEE Trans. Inform. Theory, **IT-14**, 206—220. (Русский перевод: Форни Г. Д., Экспоненциальные границы для ошибки в системах со стиранием, декодированием списком и решающей обратной связью. Сб. Некоторые вопросы теории кодирования, «Мир», М., 1970, 166—204.)

Фэлконер (Falconer D. D.), 1966, A Hybrid Sequential and Algebraic Decoding Scheme. Ph D. Thesis, Dept. of E. E., MIT, Cambridge, Mass.

Халмош (Halmos P. R.), 1950, Measure Theory. Van Nostrand, Princeton, N. J. (Русский перевод: Халмош П., Теория меры. ИЛ., М., 1953.)

Харди, Литтльвуд, Полиа (Hardy G. H., Littlewood J. E., Polya G.), 1934, Inequalities. Cambridge, Univ. Press, London. (Русский перевод: Харди Г., Литтльвуд Дж., Полиа Г., Неравенства. ИЛ, М., 1948.)

* Харкевич А. А., 1965, Борьба с помехами. Физматгиз, М.

Хаффман (Haffman D. A.), 1962, A Method for the Construction of Minimum Redundancy Codes. Proc. IRE, **40**, 1098—1101. (Русский перевод: Хаффман Д. А., Метод построения кодов с минимальной избыточностью. Кибернетический сб., вып. 3, ИЛ, М., 1961, 79—87.)

Хегельбергер (Hagelbarger D. W.), 1959, Recurrent Codes: Easily Mechanized, Burst—Correcting, Binary Codes. Bell System Tech. J., **38**, 969—984.

Хинчин А. Я., 1956, Об основных теоремах теории информации. Успехи матем. наук, **11**, 1, 17—75.

Хоквингем (Hocquenghem A.), 1959, Codes Correcteurs D'erreurs. Chiffres (Paris), **2**, 147—156.

Холзингер (Holsinger J. L.), 1964, Digital Communication over fixed Time—continuous Channels with Memory, with Special Application

to Telephone Channels. MIT Research Lab. of Electronics, Tech. Rept. 430 (also MIT Lincoln Lab. T. R. 366).

Хэмминг (Hamming R. W.), 1950, Error Detecting and Error Correcting Codes. Bell System Tech. J., **29**, 147—160. (Русский перевод: Хэмминг Р. Сб. Коды с обнаружением и исправлением ошибок, ИЛ, М., 1956.)

* Цареградский И. П., 1958, Замечание о пропускной способности стационарного канала с конечной памятью. Успехи матем. наук, **13**, 6, 49—61.

Цирлер (Zierler N.), 1960, A Class of Cyclic Linear Error-Correcting Codes in p^m Symbols. MIT Lincoln Lab. Group Report 55—19, Lexington, Mass.

* Цыбаков Б. С., 1959 а, О пропускной способности однолучевого канала со случайными изменениями поглощения. Научн. докл. высш. школы, Радиотехника и электроника, 2, 44—51.

* Цыбаков Б. С., 1959 б, Пропускная способность некоторых многолучевых каналов связи. Радиотехника и электроника, **4**, 10, 1602—1608.

* Цыбаков Б. С., 1965, Пропускная способность векторного гауссовского канала без памяти. Пробл. перед. инф., **1**, 1, 26—40.

* Цыбаков Б. С., 1966, Об асинхронных каналах с синхросимволом. Пробл. перед. инф., **2**, 1, 28—36.

* Цыбаков Б. С., 1969, Эпсилон энтропия векторного сообщения. Пробл. перед. инф., **5**, 1, 96—97.

Чень (Chien R. T.), 1964, Cyclic Decoding Procedures for Bose-Chaudhuri—Hocquenghem Codes. IEEE Trans. Inform. Theory, **IT-10**, 357—363.

Чернов (Chernoff H.), 1952, A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on a Sum of Observations. Ann. Math. Stat, **23**, 493—507.

Шелквийк (Schalkwijk J. P. M.), 1966, A Coding Scheme for Additive Noise Channels with Feedback, Part 2. IEEE Trans. Inform. Theory, **IT-12**, 183—189.

Шелквийк (Schalkwijk J. P. M.), 1968, Center of Gravity Information Feedback. IEEE Trans. Inform. Theory, **IT-14**, 324—331.

Шелквийк, Кайлат (Schalkwijk J. P. M., Kailath T.), 1966, A Coding Scheme for Additive Noise Channels with Feedback Part 1. IEEE Trans. Inform. Theory, **IT-12**, 172—182.

Шеннон (Shannon C. E.), 1948, A Mathematical Theory of Communication. Bell System Tech. J., **27**, 379—423, (Part I), 623—656 (Part II). (Русский перевод: Шеннон К., Математическая теория связи. Сб. Работы по теории информации и кибернетике, ИЛ., М., 1963, 243—332.)

Шеннон (Shannon C. E.), 1949, Communication in the Presence of Noise. Proc. IRE, **37**, 10—21. (Русский перевод: Шеннон К., Связь при наличии шума. Сб. работы по теории информации и кибернетике, ИЛ, М., 1963, 433—460.)

Шеннон (Shannon C. E.), 1956, The Zero Error Capacity of a Noisy Channel. IRE Trans. Inform. Theory, **IT-2**, 8—19. (Русский перевод: Шеннон К., Пропускная способность канала с шумом при нулевой ошибке. Сб. Работы по теории информации и кибернетике, ИЛ, М., 1963, 664—487.)

Шеннон (Shannon C. E.), 1957, Certain Results in Coding Theory for Noisy Channels. Inform. and Control, **1**, 6—25. (Русский перевод: Шеннон К., Некоторые результаты теории кодирования для каналов с шумами. Сб. Работы по теории информации и кибернетике, ИЛ, М., 1963, 509—531.)

Шеннон (Shannon C. E.), 1959, Coding Theorems for a Discrete Source with Fidelity Criterion. IRE Nat. Con. Record, Part 4, 142—163. (Русский перевод: Шеннон К., Теоремы кодирования для дискретного канала при заданном критерии точности. Сб. Работы по теории информации, ИЛ, М., 1963, 587—621.)

Шеннон (Shannon C. E.), 1959, Probability of Error for Optimal Codes in a Gaussian Channel. Bell System Tech. J., **38**, 611—656. (Русский перевод: Шеннон К., Вероятность ошибки для оптимальных кодов в гауссовском канале. Сб. Работы по теории информации и кибернетике, ИЛ, М., 1963, 540—586.)

Шеннон (Shannon C. E.), 1961, Two-Way Communication Channels. Proc. Fourth Berkeley Symp. on Prob. and Stat., **1**, 611—644. University of California Press, Berkeley Calif. (Русский перевод: Шеннон К., Двусторонние каналы связи. Сб. Работы по теории информации и кибернетике, *ИЛ*, М., 1963, 622—663.)

Шеннон, Галлагер, Берлекэмп (Shannon C. E., Gallager R. G., Berlekamp E. R.), 1967, Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels. Inform. and Control, **10**, 65—103 (Part I), 522—552 (Part II). (Русский перевод: Шеннон К., Галлагер Р., Берлекэмп Е., Нижние границы вероятности ошибки для кодирования в дискретном канале без запоминания. Зарубежная радиоэлектроника, 2 и 6, 1968, 52—81 и 41—64.)

Шольтц (Scholtz R. A.), 1966, Codes with Synchronization Capability. IEEE Trans. Inform. Theory, **IT-12**, 135—140.

Эберт (Ebert R. M.), 1966, Error Bounds for Paralles Communication Channels. MIT Research Lab. of Electronics, Tech. Rept. 448.

Эбрамсон (Abramson N.), 1963, Information Theory and Coding McGraw-Hill, New York.

Элайс (Elias P.), 1954, Error Free Coding. IRE Trans. Inform. Theory, **IT-4**, 29—37. (Русский перевод: Элайс П., Безошибочное кодирование. Сб. Коды с обнаружением и исправлением ошибок, ИЛ, М., 1956, 59 — 71).

Элайс (Elias P.), 1955, Coding for Noisy Channels. IRE Convention Record, Part 4, 37—46.

Элайс (Elias P.), 1956, Coding for Two Noisy Channels. Cherry C (Ed.), Inform. Theory, Butterworth, London, 61—74. (Русский перевод: Элайс П., Кодирование для двух каналов с шумами. Сб. Теория передачи сообщений. *ИЛ*., М., 114—138.)

Элайс (Elias P.), 1957, List Decoding for Noisy Channels, MIT Research Lab. of Electronics Tech. Rept. 335.

Элайс (Elias P.), 1961, Channel Capacity without Coding, Lectures on Communication System Theory. Baghdady E. J. (Ed.), McGraw-Hill, New York. (Русский перевод: Грин П., Системы с обратной связью, приложение. Сб. Лекции по теории систем связи, «Мир», 1964, 339—344.)

Элайс (Elias P.), 1967, Networks of Gaussian Channels with Applications to Feedback Systems. IEEE Trans. Inform. Theory, **IT-13**, 493—501. (Русский перевод: Элайс П., Сети гауссовских каналов и их применение к системам с обратной связью. Сб. Некоторые вопросы теории кодирования. «Мир», М., 1970, 205—229.)

Элспас, Шорт (Elspas B., Short R. A.), 1962, A Note of Optimum Burst-Error-Correcting Codes. IRE Trans. Inform. Theory, **IT-8**, 39—42. (Русский перевод: Элспас В., Шорт Р., Об оптимальных кодах, исправляющих пакеты ошибок. Сб. Теория кодирования, «Мир». М., 1964, 83—96.)

Эш (Ash R. B.), 1965, Information Theory. Interscience Publishers, New York.

Юдкин (Yudkin H. L.), 1964, An Error Bound for Gaussian Signals in Gaussian Noise. MIT Research Lab. of Electronics, QPR, No 73, 149—155, Cambridge, Mass.

Юдкин (Yudkin H. L.), 1964, Channel State Testing in Information Decoding. Sc. D. Thesis, Dept. of E. E., MIT, Cambridge, Mass.

Юдкин (Yudkin H. L.), 1967, On the Exponential Error Bound and Capacity for Finite State Channels. International Symposium on Inform. Theory, San Remo, Italy.

Яглом А. М., 1952, Введение в теорию стационарных случайных функций. Успехи матем. наук, 7, 5 (51), 3—168.

* Яглом А. М., 1960, Явные формулы для экстраполяции, фильтрации и вычисления количества информации в теории гауссовских стохастических процессов. Trans. of the Second Prague Conf. on Inform. Theory, Statist. Decision Funct. and Random Processes. Prague, 251—262.

* Яглом А. М., Яглом И. М., 1973, Вероятность и информация, «Наука», М.

ИМЕННОЙ УКАЗАТЕЛЬ*)

*) См. также список литературы (стр. 695—708). *Прим. ред.*

Галлагер Р.

Теория информации и надежная связь. Перев. с англ. под ред. М. С. Пинскера и Б. С. Цыбакова. М., «Сов. радио», 1974.

720 с. с ил.

В книге собраны, подытожены и заново переосмыслены все основные результаты теории информации.

Книга предназначена для широкого круга инженеров и математиков, специализирующихся по системам связи, системам управления, вычислительным машинам и кибернетическим устройствам. Она также может служить хорошим учебным пособием для аспирантов и студентов.

Information Theory and Reliable Communication

Robert G. Gallager

MASSACHUSETTS INSTITUTE OF TACHNOLOGY

John Wiley and sons, Inc.
New York, London, Sydney, Toronto, 1968

Р. ГАЛЛАГЕР

ТЕОРИЯ ИНФОРМАЦИИ И НАДЕЖНАЯ СВЯЗЬ

*

Перевод с английского под редакцией
М. С. Пинскера и Б. С. Цыбакова

Москва · Советское радио · 1974

Галлагер Р. **Теория информации и надежная связь.** США, 1968 г. Пер. с англ., под ред. М. С. Пинскера и Б. С. Цыбакова, М., «Советское радио», 1974, 720 с.

В книге собраны, подытожены и заново переосмыслены все основные результаты теории информации. Конструкция наиболее перспективных для практического использования кодов, разнообразные методы декодирования, выражения для вероятностей ошибки, пропускная способность реальных каналов связи, методы сокращения избыточности — все это и многое другое изложено с самых современных позиций. Предлагаемые читателю результаты (вместе с изящными и полными их доказательствами) сведены в книге в единую систему. Математические рассуждения удачно сочетаются с инженерными выводами и техническими рекомендациями.

Книга предназначена для широкого круга инженеров и математиков, специализирующихся по системам связи, системам управления, вычислительным машинам и кибернетическим устройствам. Она также может служить хорошим учебным пособием для аспирантов и студентов.

ОГЛАВЛЕНИЕ

ДИСКРЕТНЫЕ КАНАЛЫ БЕЗ ПАМЯТИ И ПРОПУСКНАЯ СПОСОБНОСТЬ 88

5

ТЕОРЕМА КОДИРОВАНИЯ ДЛЯ КАНАЛА С ШУМАМИ 132

6

МЕТОДЫ КОДИРОВАНИЯ И ДЕКОДИРОВАНИЯ 211