О. А. Усенко

Приложения теории информации к задачам радиотехники

учебное пособие

ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ЮФУ
РОСТОВ-НА-ДОНУ–ТАГАНРОГ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное
образовательное учреждение высшего образования
«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Инженерно-технологическая академия

О. А. УСЕНКО

ПРИЛОЖЕНИЯ ТЕОРИИ ИНФОРМАЦИИ К ЗАДАЧАМ РАДИОТЕХНИКИ

Учебное пособие

Ростов-на-Дону – Таганрог
Издательство Южного федерального университета
2021

УДК 621.391.2(075.8)+621.37(075.8)
ББК 32.811я73+32.84я73
У745

*Печатается по решению кафедры радиотехнических
и телекоммуникационных систем
Института радиотехнических систем и управления
Южного федерального университета
(протокол № 15 от 30 июня 2021 г.)*

Рецензенты:

кандидат технических наук, доцент кафедры информатики
Таганрогского института имени А. П. Чехова (филиал) «Ростовского
государственного экономического университета (РИНХ)» *С. Г. Буланов*

кандидат технических наук, доцент кафедры радиотехнических
и телекоммуникационных систем Института радиотехнических систем
и управления *В. Т. Лобач*

Усенко, О. А.

У745 Приложения теории информации к задачам радиотехники : учебное пособие / О. А. Усенко ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2021. – 154 с.

ISBN 978-5-9275-3964-2

В учебном пособии изложены теоретические основы прикладной теории информации. Особое внимание уделено методам и приемам решения практических задач, наиболее часто встречающихся в радиотехнике. В данное пособие включены разделы для самоконтроля в виде теоретических вопросов и практических заданий. Учебное пособие предназначено для бакалавров, специалистов и магистрантов технических вузов.

УДК 621.391.2(075.8)+621.37(075.8)
ББК 32.811я73+32.84я73

ISBN 978-5-9275-3964-2

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Современные достижения техники и технологий открывают широкие возможности для создания эффективных и конкурентоспособных информационных радиотехнических систем, предназначенных для передачи, преобразования и хранения различной информации с помощью электромагнитных колебаний. Понимание происходящих процессов, а также умение количественно оценивать основные характеристики и параметры проектируемых или эксплуатируемых систем является необходимым требованием при подготовке специалистов радиотехнических специальностей.

В учебном пособии представлены основные понятия и определения, связанные с организацией процессов передачи и преобразования информации, проектированием кодов, удовлетворяющих заданным требованиям, а также рассмотрены некоторые вопросы схемотехнической реализации отдельных блоков. Большое внимание уделяется формированию практических навыков у будущих специалистов по решению задач, количественной оценке показателей, освоению методик и алгоритмов расчета в типовых задачах. В пособии по каждой теме предлагаются примеры решений, которые должны стать методологическим руководством для самостоятельного решения задач, также приводимым по каждой теме. Список контрольных вопросов позволит оценить качество усвоения теоретической части материала.

Материал, содержащийся в учебном пособии, может быть рекомендован к изучению как в рамках читаемых учебных дисциплин, так и для самостоятельного освоения с целью применения в профессиональной деятельности.

1. КОЛИЧЕСТВЕННАЯ ОЦЕНКА ИНФОРМАЦИИ

1.1. Понятие информации

В информационных радиотехнических системах (РТС) в том или ином объеме непременно реализованы функции приема, передачи, переработки и хранения информации, в отличие от неинформационных (энергетических) РТС, которые предназначены только для создания высокочастотной энергии.

В зависимости от основного назначения информационные РТС классифицируют на системы:

– передачи информации;
– извлечения информации;
– разрушения информации;
– комбинированные информационные системы.

Для таких систем следует, в первую очередь, конкретизировать понятие информации, которое в различных приложениях может определяться исходя из разных аспектов. И это неудивительно, поскольку информация представляет собой общенаучное понятие, каждая наука или ее направление исследует свои процессы и явления, и для них вводится свое, наиболее подходящее определение информации. Но несмотря на различающиеся по формулировкам определения информации, они не противоречат друг другу, и их следует рассматривать как взаимодополняющие, более глубоко раскрывающие особенности и свойства такого сложного понятия, как информация.

Приведем основные наиболее распространенные определения информации, а далее конкретизируем и введем рабочее определение понятия информации для нашей предметной области, связанной с информационными РТС.

В самом широком смысле [10] информацию (от лат. «infomatio» – «разъяснение, осведомление, изложение») рассматривают наряду с материей и энергией как первичное понятие нашего мира, и потому в строгом смысле полагают, что информация не может быть определена, а только выявлены ее свойства. Понятие информации относится к фундаментальным понятиям, связано с процессами обмена сведениями первоначально между людьми устным, письменным или каким-либо другим способом. Позже с середины XX в. рассматривался обмен сведениями между человеком и устройством, обмен сигналами между живой и неживой природой.

В [6] информация определяется как отражение реального мира, информация есть отраженное разнообразие, т.е. нарушение однообразия.

В настоящее время понятие информации вводят, используя различные подходы. Преимущественными считаются два основных подхода: функциональный (кибернетический подход) и атрибутивный (философско-методологический подход). *Функциональный подход* постулирует, что информация – это «кровь», функция самоорганизующихся систем. В *атрибутивном подходе* считается, что информация – это фундаментальное, неотъемлемое и универсальное свойство материи, ее атрибут (наряду с энтропией и временем). В настоящее время наблюдается тенденция признания атрибутивного подхода. В его рамках выделяется структурная и оперативная информация [12].

Структурная информация содержится в структуре системы (она отражается, например, в структуре кристаллических решеток). Высшей формой такой информации является генетический код. Структурная информация под воздействием окружающей среды может искажаться или пропадать. И только живые организмы могут целенаправленно ее использовать для сохранения своей целостности в условиях меняющейся окружающей среды [12]. При этом такая информация оказывается неотрывной от категории: цель, принятие решения, лицо, принимающее решение (ЛПР), управление, язык. С информационной точки зрения жизнь – это высокоустойчивое состояние вещества, использующее информацию для выработки сохраняющихся реакций.

Оперативная информация – это информация о текущем состоянии среды. Она может объективизироваться путем перехода в структурную информацию. Живые существа не только используют структурную информацию, они способны конструировать новую структурную информации на основе оперативной; но лишь человек способен изготавливать на основе информации различные орудия.

Сторонники функционального подхода сигналом называют лишь то явление, тот процесс, который способен изменять внутреннюю модель внешнего мира некоторой самоорганизующейся системы. Из этого следует, что для одной и той же самоорганизующейся системы различные физические явления внешнего мира могут являться сигналами или нет в зависимости от того, способны ли они изменять внутреннюю модель системы. Одно и то же явление для разных самоорганизующихся систем может быть или не быть сигналом.

Таким образом, функциональный подход предусматривает наличие в самоорганизующейся системе особого вида отражающей подсистемы – *информационной системы.* Важнейшей способностью самоорганизующейся системы является возможность изменяться под воздействием некоторых факторов внешнего мира, которые значимы для данной самоорганизующейся системы. Именно в этом суть выделения сигналов самоорганизующейся системой.

Зачастую, как, например, в работах Р. Карнапа, И. Бар-Хилле, А. Н. Колмогорова, информация рассматривается как некая абстракция, не существующая в физической реальности, подобно тому, как не существует мнимое число или не имеющая линейных размеров точка. С кибернетической точки зрения информация (информационные процессы) есть во всех самоуправляемых системах (технических, биологических, социальных). Она характеризует меру сложности структур, меру их организации. К. Шеннон определил понятие информации как коммуникацию, связь, в процессе которой устраняется неопределенность [14]. Основоположник кибернетики Н. Винер рассматривал информацию как обозначение содержания, полученное нами из внешнего мира в процессе приспосабливания к нему нас и наших чувств. В социальной информатике «Информация (Information) – содержание сообщения или сигнала; сведения, рассматриваемые в процессе их передачи или восприятия, позволяющие расширить знания об интересующем объекте» [4].

И, конечно, в каждом научном направлении, дисциплине специалисты и ученые стремятся конкретизировать понятие информации, адаптировать к своей предметной области, в результате чего можно обнаружить такое многообразие формулировок этого сложного понятия [12]:

- информация – это обозначенное отражение состояния мира;
- информация – это отрицание энтропии;
- информация – это коммуникация и связь, в процессе которой устраняются неопределенности;

- информация – это оригинальность, новизна;
- информация – это мера сложности структур;
- информация – это данные, определенным образом организованные, имеющие смысл, значение и ценность для своего потребителя, необходимую для принятия им решений, а также для реализации других функций и действий;
- информация – это совокупность знаний о фактических данных и зависимостях между ними, являющихся одним из видов ресурсов, используемых человеком в трудовой деятельности и быту;
- информация – это сведения о лицах, предметах, фактах, событиях, явлениях и процессах независимо от формы представления;
- информация – это средство и форма передачи знаний и опыта, сокращающая неопределенность и случайность и неосведомленность;
- информация – это обобщенный термин, относящийся к любым сигналам, звукам, знакам и т.д., которые могут передаваться, приниматься, записываться и/или храниться.

Теория информации в приложении к системам передачи информации связывает понятие информации с функционированием тройки (рис. 1).

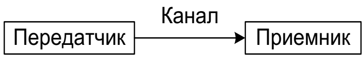


Рис. 1. Общая схема связи (схема К. Шеннона)

Информация всегда возникает в результате передачи сведений от передатчика, являющегося источником информации, через канал связи к приемнику, являющемуся получателем информации.

В самом общем случае *источник информации* – это субъект или объект, порождающий информацию и представляющий ее в виде сообщения, а *приемник информации* – это субъект или объект, принимающий сообщение и способный правильно его интерпретировать.

При реализации этой схемы важно учитывать важнейшую характерную особенность информации, что это категория *нематериальная*. Следовательно, для существования и распространения в материальном мире она должна быть обязательно связана с какой-либо материальной основой – без нее информация не может проявиться, передаваться и сохраняться.

Материальный объект или среду, которые служат для представления или передачи информации, будем называть *материальным носителем.*

В процессе исторического развития человечество использовало различные материальные носители для сохранения, обработки и передачи информации, начиная от камня (наскальная живопись, пирамиды и т.п.), глины (глиняные дощечки для записи и т.п.) и до материалов, используемых в современных технологиях, – бумага, лазерный диск, флэш-носители, а также различные среды – воздух, электромагнитное поле и пр. Однако можно отметить одно отличие, связанное с хранением и передачей информации. *Хранение* не относится к информационным процессам, поскольку любой процесс должен быть связан с изменением некоторой характеристики материального носителя, а хранение связано как раз с обеспечением неизменности состояния, причем как можно более долгое время. В зависимости от способности материала сохранять свои свойства будет зависеть качество и долговременность хранения информации (символы и рисунки на камнях могут насчитывать несколько тысячелетий, современные электромагнитные носители рассчитаны не более чем на несколько лет). При организации *передачи* информации, наоборот, необходимо использовать характеристики, которые способны изменяться с течением времени, например, амплитуда колебаний звуковой волны или напряжение в проводах. Состояния и процессы могут иметь физическую, химическую, биологическую или иную основу – главное, что они материальны.

Однако не с любым процессом можно связать информацию. В частности, *стационарный* процесс, т.е. процесс с неизменными в течение времени характеристиками, информацию не переносит. Примером может служить постоянный электрический ток, ровное горение лампы или равномерный гул – они содержат лишь ту информацию, что процесс идет, т.е. что-то функционирует. Иное дело, если мы будем лампу включать и выключать, т.е. изменять ее яркость, – чередованием вспышек и пауз можно представить и передать информацию (например, посредством азбуки Морзе). Таким образом, для передачи необходим *нестационарный процесс*, т.е. процесс, характеристики которого могут изменяться; при этом информация связывается не с существованием процесса, а именно с изменением какой-либо его характеристики.

Изменение характеристики носителя, которое используется для представления информации, называется *сигналом*, а значение этой характеристики, отнесенное к некоторой шкале измерений, называется *параметром сигнала*.

Сигнал – это материальный носитель информации (предмет, явление, процесс). Сигнал представляет собой динамический процесс, т.е. процесс, изменяющийся со временем, или колебания величины любой природы: напряжения, давления, электромагнитного поля и др.

В табл. 1 приведены примеры процессов, используемых для передачи информации, и связанных с ними сигналов [3].

Таблица 1

Процессы, используемые для передачи информации

Способ передачи	Процесс	Параметры сигнала
Звук	Звуковые волны	Высота и громкость звука
Радио, телевидение	Радиоволны	Частота, амплитуда или фаза радиоволны
Изображение	Световые волны	Частота и амплитуда световых волн
Телефон, компьютерная сеть	Электрический ток	Частота и амплитуда электрических колебаний в линии связи

Однако одиночный сигнал, как известно, не может содержать много информации. Поэтому для передачи информации используется ряд следующих друг за другом сигналов.

Последовательность сигналов или знаков называется *сообщением.* Знаками называют реальные различимые получателем материальные объекты: буквы, цифры, предметы. В теории информации понятия «знак» и «сигнал» часто взаимозаменяемы без ущерба однозначности, однако знаки чаще используют при описании хранения сообщений (информации), а сигналы – при описании передачи, переноса сообщений. В последнем случае знакам однозначно сопоставляются сигналы. Знаки можно рассматривать как особый вид сигналов. Можно отметить, что знаки в некоторых культурах могут нести несравненно большую семантическую нагрузку, чем обычный сигнал, имеют собственную историю и ценность. В теории систем вводится вполне однозначный критерий для разграничения сигналов и знаков. Полагают, что сигналы имеют естественное происхождение, первичны, а знаки – есть результат функционирования и преобразования различными сложными, самоорганизующимися системами.

Множество всех знаков или сигналов называют *алфавитом*, из которого строятся сообщения. Число различных знаков в алфавите называют его *объемом*.

Таким образом, от источника к приемнику информация передается в виде сообщений. Можно сказать, что сообщение выступает в качестве *материальной оболочки* для представления информации при передаче. Следовательно, *сообщение служит переносчиком информации, а информация является содержанием сообщения*.

Соответствие между сообщением и содержащейся в нем информацией называется *правилом интерпретации сообщения.*

Правило интерпретации разделяют на *однозначные* и *неоднозначные*. При создании радиотехнических, телекоммуникационных систем используют однозначное соответствие, которое предполагает единственный способ интерпретации получаемых сообщений. Неоднозначные соответствия встречаются, как правило, в иных, нетехнических сферах. Неоднозначность соответствия между сообщением и информацией возможна в двух вариантах [3, 7]:

- одна и та же информация может передаваться различными сообщениями (например, сообщение может быть получено по радио, телевизору, телефону и пр.);
- одно и то же сообщение может содержать различную информацию для разных приемников (примером может служить передача в 1936 г. по радио фразы «*Над всей Испанией безоблачное небо*», которое для непосвященных людей имело смысл прогноза погоды, а для знакомых с правилом интерпретации – сигналом к началу военных действий).

Информационный процесс – это изменение с течением времени содержания информации или представляющего его сообщения.

Различных видов информационных процессов оказывается немного:

- порождение (создание) новой информации;
- преобразование информации (т.е. порождение новой информации в результате обработки имеющейся);
- уничтожение информации;
- передача информации (распространение в пространстве).

На самом деле все перечисленные события происходят не непосредственно с самой информацией, а с сообщением, т.е. ее материальной оболочкой. И с этих позиций возможны лишь два типа процессов: изменение сообщения с сохранением содержащейся в нем информации и изменение сообщения, сопровождающееся преобразованием информации. К процессам первого типа относится передача информации без потерь и обратимая перекодировка; к процессам второго типа – создание-уничтожение, необратимая перекодировка, передача с потерями, обработка с появлением новой информации.

Отдельно следует остановиться на хранении информации. Как уже было сказано, хранение связывается с фиксацией параметра материального носителя, который далее с течением времени не меняется. Следовательно, запись информации на носитель (непосредственно момент фиксации параметра) и ее последующее считывание подпадают под определение информационного процесса, но само хранение – нет. Хранение следовало бы назвать *информационным состоянием*, однако такое понятие в науке не используется.

Совокупность технических средств, обеспечивающих в радиотехнических системах получение, кодирование, передачу, восстановление, обработку и хранение информации, составляет средства информационной подсистемы [13].

Рассмотрим обобщенную структурную схему подсистемы передачи информации (рис. 2).

Источник сообщений в общем случае образует совокупность источника информации (исследуемого или наблюдаемого объекта) и первичного преобразователя (датчика, человека-оператора и т.д.), воспринимающего информацию о протекающем в нем процессе.

Различают дискретные и непрерывные сообщения.

В радиотехнических, телемеханических, телекоммуникационных системах могут использоваться как дискретные, так и непрерывные (аналоговые) сигналы. Дискретные сообщения формируются по определенным правилам из символов абстрактного алфавита. Непрерывные сообщения связаны с передачей аналоговых сигналов, которые принимают любые значения в заданном континууме и изменяются в произвольные моменты времени. Они описываются непрерывными функциями времени, принимающими непрерывное множество значений (речь, телевизионное изображение).

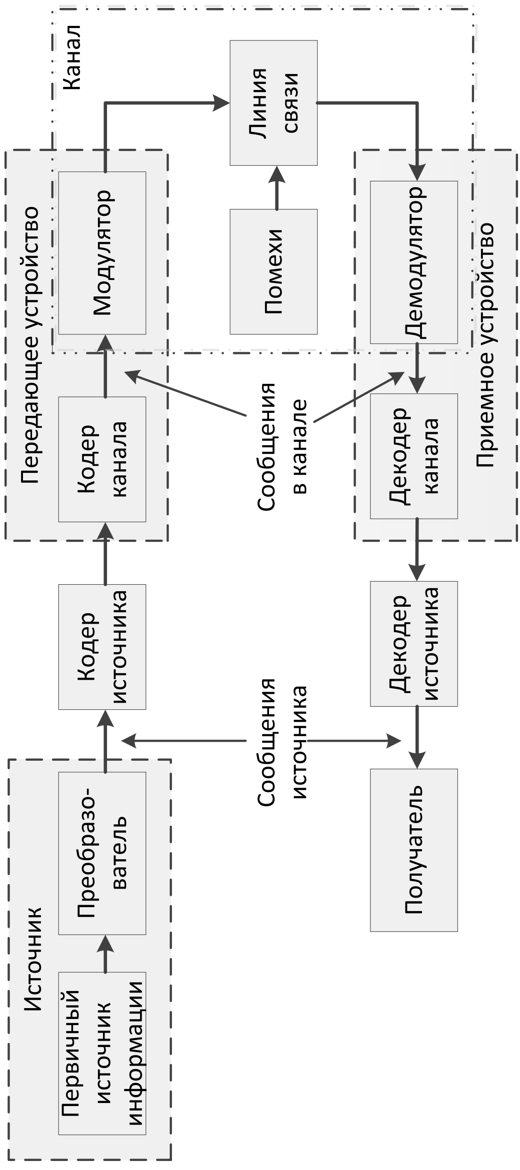


Рис. 2. Структурная схема подсистемы передачи информации

Для передачи сообщения по каналу связи ему ставят в соответствие определенный сигнал. Под сигналом понимают физический процесс, отображающий (несущий) сообщение.

Преобразование сообщения в сигнал, удобный для передачи по данному каналу связи, называют *кодированием в широком смысле слова.* Операцию восстановления сообщения по принятому сигналу называют *декодированием* [2, 8].

Зачастую при решении ряда практических задач при переходе от первичного алфавита к некоторому вторичному стремятся уменьшить количество используемых символов. Такое кодирование принято называть в узком смысле.

Устройство, выполняющее такую операцию, называют кодирующим или *кодером.* Так как алфавит символов меньше алфавита знаков, то каждому знаку соответствует некоторая последовательность символов, которую называют *кодовой комбинацией*.

Число символов в кодовой комбинации называют ее *значностью*, число ненулевых символов – *весом*.

Для операции сопоставления символов со знаками исходного алфавита используют термин “*декодирование*”. Техническая реализация этой операции осуществляется декодирующим устройством или *декодером* [3, 9, 13].

Передающее устройство осуществляет преобразование непрерывных сообщений или знаков в сигналы, удобные для прохождения по линии связи. При этом один или несколько параметров выбранного сигнала изменяют в соответствии с передаваемой информацией. Такой процесс называют *модуляцией*. Он осуществляется *модулятором*. Обратное преобразование сигналов в символы производится *демодулятором*.

Под *линией связи* понимают среду (воздух, металл, магнитную ленту и т.д.), обеспечивающую поступление сигналов от передающего устройства к приемному устройству.

Сигналы на выходе линии связи могут отличаться от сигналов на ее входе (переданных) вследствие затухания, искажения и воздействия помех.

Помехами называют любые мешающие возмущения как внешние, так и внутренние, вызывающие отклонение принятых сигналов от переданных сигналов. Из смеси сигнала с помехой *приемное устройство* выделяет сигнал и посредством декодера восстанавливает сообщение, которое в общем случае может отличаться от посланного.

Меру соответствия принятого сообщения посланному сообщению называют *верностью передачи* [2, 8].

Принятое сообщение с выхода системы связи поступает к абоненту-получателю, которому была адресована исходная информация.

Совокупность средств, предназначенных для передачи сообщений, называют *каналом связи* [2, 8].

В заключение данного подраздела отметим важные свойства информации, которые удалось установить:

- информация приносит знания об окружающем мире, которых у получателя не было до получения этой информации;
- информация не материальна, но она проявляется в форме материальных носителей – дискретных знаков и сигналов или в форме функций времени;
- в процессе передачи информации от источника к приемнику ее носитель (сообщение) в канале связи может быть подвергнут искажению из-за действия различного рода помех;
- знаки и сигналы несут информацию только для получателя, способного распознать их.

1.2. Оценка количества информации при равновероятных состояниях элементов сообщений

Фундаментальным вопросом для теории информации является вопрос о количественной мере информации. Необходимо отметить, что всякая информация получается потребителем после принятия сообщения, проведения опыта и т.п. Сообщение, получаемое на приемной стороне, несет полезную информацию лишь в том случае, если имеется неопределенность относительно состояния источника сообщений.

В теории информации исследуются информационные системы при четко сформулированных условиях (постулатах) [4, 7, 8].

1. Источник сообщения осуществляет выбор сообщения из некоторого множества с определенной вероятностью.

2. Сообщения могут передаваться по каналу связи в закодированном виде. Кодированные сообщения образуют множество, являющееся взаимно однозначным отображением множества сообщений. Правило декодирования известно декодеру (записано в его программе).

3. Сообщения следуют друг за другом, причем число сообщений может быть сколь угодно большим.

4. Сообщение считается принятым верно, если в результате декодирования оно может быть в точности восстановлено. При этом не учитывается, сколько времени прошло с момента передачи сообщения до момента окончания декодирования, и какова сложность операций кодирования и декодирования.

5. Количество информации не зависит от смыслового содержания сообщения, от его эмоционального воздействия, полезности и даже от его отношения к реальной действительности.

Формулу для определения количества информации в сообщении получим при следующих условиях [4, 11, 12]:

1) осуществляется передача дискретных сообщений;
2) сообщения являются равновероятными и взаимно независимыми;
3) символы, выдаваемые источником, взаимно независимы;
4) система счисления конечна.

Определим количество различных сообщений, которые можно составить из n элементов, принимающих любое из m различных состояний. Пусть сообщение состоит из двух элементов ($n = 2$) и каждый из элементов может находиться в одном из m различных состояний. Первому состоянию первого элемента сообщения соответствует m состояний второго элемента и т.д. Всего же m различным состояниям первого элемента и m различным состояниям второго элемента соответствует $N = m^2$ различных сообщений. Если количество сообщений равно трем ($n = 3$), то всего различных сообщений может быть $N = m^3$ и т.д. Очевидно, описанная схема соответствует формированию выборки по принципу размещений с повторениями. Тогда в общем случае если в сообщении содержится n элементов, то количество возможных различных сообщений равно

$$N = m^n. \tag{1}$$

Чем больше N, тем больше потенциально возможных различных сообщений может сгенерировать источник сообщений, тем большим источником неопределенности он является. Действительно, если источник может порождать 3 различных сообщения, то выбор невелик, и можно с высокой долей вероятности «угадать» очередное сообщение. Совершенно иная ситуация складывается, если источник потенциально может порождать несколько тысяч различных сообщений. Предсказать появление очередного сообщения практически невозможно, особенно случае, когда вероятности появления сообщений равны или дифференцированы незначительно. В случае равных вероятностей получение конкретного сообщения равносильно случайному выбору одного из N сообщений с вероятностью $1/N$. Отсюда следует, *что чем больше N, тем большая степень неопределенности характеризует этот выбор и тем более информативным можно считать сообщение*. Поэтому величина N при равновероятных состояниях элементов сообщений может быть принята за меру количества информации, содержащейся в этих сообщениях. Назовем ее квантовой мерой количества информации [1, 7, 11].

Использование N в качестве меры количества информации имеет смысловые и логические противоречия. Так, при $N = 1$, количество информации представляет собой ненулевую величину. Однако по смыслу если опыт имеет лишь один исход и не содержит никакой неопределенности, то наблюдатель заранее будет знать исход этого опыта. Это соответствует источнику, который может выдавать только одно сообщение (или знак). В результате наблюдатель не будет получать никакой информации, т.е. количественная оценка информации должна принимать в данном случае нулевое значение.

Кроме того, количественная мера информации в виде (1) не обладает свойством аддитивности, т.е. при сложении количества информации нескольких независимых источников сообщений не выполняется условие простого линейного сложения количества информации. Пусть первый источник сообщений может генерировать N_1 различных сообщений, а второй – N_2 различных сообщений. Тогда общее число различных сообщений для двух источников равно $N = N_1 \cdot N_2$; для k различных источников сообщений общее число возможных сообщений по принципу произведения из теории вероятности равно $N = \prod_{i=1}^{k} N_i$. Однако логично предположить, что наблюдатель, фиксируя результаты двух независимых экспериментов или получая сообщения от разных источников, суммирует полученную информацию, а не преумножает ее.

В 1928 г. американский ученый Р. В. Л. Хартли предложил в качестве меры количества информации принять *логарифм числа возможных сообщений* [4, 9, 13]:

$$I = log\ N = log\ m^n = n\ log\ m. \tag{2}$$

Эта логарифмическая функция характеризует *количество информации.*

Доказывается, что логарифмическая функция является единственной функцией аргументов N_1 и N_2, удовлетворяющих условиям:

1) $f(N_1, N_2) = f(N_1) + f(N_2)$ – условие аддитивности;
2) $f(1) = 0$;
3) $f(N_1) > f(N_2)$ при $N_1 > N_2$.

Для k различных источников сообщений общее количество информации от всех источников по логарифмической мере количества информации составляет

$$I = \log N = \log\left(\prod_{i=1}^{k} N_i\right) = \sum_{i=1}^{k} \log N_i = \sum_{i=1}^{k} I_i = \sum_{i=1}^{k} n_i \log m_i\ . \tag{3}$$

Таким образом, логарифмическая мера количества информации обладает аддитивностью как в отношении количества элементов в каждом сообщении, так и в отношении суммы различных сообщений.

Количество информации I, содержащееся в сообщениях, является логарифмической мерой числа различных состояний $I = n \log m$ и в принципе безразлично, при каком основании вычислять этот логарифм, так как в силу соотношения $\log_a m = \log_a b \log_b m$ переход от одного основания логарифма к другому сводится лишь к изменению единицы измерения. Основание логарифма влияет лишь на удобство вычисления:

– в двоичных единицах $I = n \log_2 m$ бит;
– в троичных единицах $I = n \log_3 m$ трит;
– в десятичных единицах $I = n \lg m$ дит;
– в натуральных единицах $I = n \ln m$ нат.

Из соотношений логарифмов:

1 дит = $\log_{10}2$ бит = 3,32193 бит;

1 трит = $\log_3 2$ бит = 1,58496 бит;

1 нат = $\log_e 2$ бит = 1,44269 бит.

Так как современная цифровая техника базируется на элементах, имеющих два устойчивых состояния, то обычно выбирают основание логарифма, равное двум.

Единицу количества информации на один элемент сообщения называют *двоичной единицей* или *битом*. При этом единица неопределенности (двоичная единица или бит) представляет собой неопределенность выбора из двух равновероятных событий (*bit* – сокращение от англ. *binary digit* – двоичная единица).

Так как из $\log_2 m = 1$ следует $m = 2$, то ясно, что 1 бит – это количество информации, которым характеризуется один двоичный элемент при равновероятных состояниях 0 и 1. Двоичное сообщение длины n содержит n бит информации.

Аналогично трит представляет собой неопределенность выбора из трех равновероятных событий.

Количество информации, приходящееся на один элемент сообщения (знак, букву), называется *энтропией* [14]:

$$H = \frac{I}{n} = \frac{n \log m}{n} = \log m . \qquad (4)$$

Преимуществами логарифмической меры количества информации являются следующие.

1. Практически более пригодна, так как параметры, имеющие техническое значение (время, триггерные элементы, ширина полосы частот и др.), линейно зависят от логарифма числа возможностей. Например, добавление одного триггера удваивает число возможных состояний схемы.
2. Логарифмическая мера ближе к интуитивному представлению о подходящей мере. Например, два канала связи должны обладать удвоенной пропускной способностью.
3. Она более удобная с математической точки зрения.

Пример 1. Определить количество информации, которое содержится в телевизионном сигнале, соответствующем одному кадру развертки. Пусть в кадре 625 строк, а сигнал, соответствующий одной строке, представляет собой последовательность из 600 случайных по амплитуде импульсов, причем амплитуда импульса может принять любое из 8 значений с шагом в 1 В.

Решение. В рассматриваемом случае длина сообщения, соответствующая одной строке, равна числу случайных по амплитуде импульсов в ней: $n = 600$.

Количество элементов сообщения (знаков) в одной строке равно числу значений, которое может принять амплитуда импульсов в строке: $m = 8$.

Количество информации в одной строке: $I = n \log m = 600 \log 8$, а количество информации в кадре: $I' = 625\, I = 625\; 600 \log 8 = 1{,}125 \cdot 10^6$ бит.

Пример 2. Какое количество информации требуется, чтобы узнать исход броска монеты? Шестигранного кубика? Икосаэдра?

В случае бросания монеты имеет два возможных исхода, $n = 2$, и события равновероятны, т.е. $p_1 = p_2 = 0{,}5$, тогда

$$I = -\,0{,}5\; log_2\, 0{,}5 - 0{,}5\; log_2\, 0{,}5 = 1\; \textit{бит}.$$

В случае бросания шестигранного кубика, n=6, события равновероятны, т.е. $p_1 = p_2 = \ldots\; p_6 = 1/6$, тогда

$$I = -\,(1/6 \log_2 1/6 - 1/6 \log_2 1/6 \ldots - 1/6 \log_2 1/6) =$$
$$= -\,6\cdot(1/6 \log_2 1/6) = 2{,}585\; \textit{бит}.$$

В случае бросания икосаэдра, $n = 20$ (напомним, что икосаэдр – многогранник, имеющий 20 граней), события равновероятны, т.е. $p_1 = p_2 = \ldots\; p_{20} = 1/20$, тогда

$$I = -\,20\cdot(1/20\; log_2\, 1/20) = 4{,}322\; \textit{бит}.$$

Заметим, что с увеличением числа различных исходов, увеличивается неопределенность, а следовательно, увеличивается количество получаемой информации.

Пример 3. Игра «*Угадай-ка*–4». Некто задумал целое число в интервале от 0 до 3. Наш опыт состоит в угадывании этого числа. На наши вопросы Некто может отвечать лишь «*Да*» или «*Нет*». Какое количество информации будет получено в результате угадывания? Как правильно построить процесс угадывания, чтобы за минимальное число вопросов полностью снять неопределенность?

Исходами в данном случае являются: A_1 – «*задуман* 0», A_2 – «*задумана* 1», A_3 – «*задумана* 2», A_4 – «*задумана* 3». Конечно, предполагается, что вероятности быть задуманными у всех чисел одинаковы. Поскольку $n = 4$, следовательно, $p(A_i) = 1/4$, $log_2 p(A_i) = -2$ и $I = 2$ *бит*. Таким образом, для полного снятия неопределенности опыта (угадывания задуманного числа) необходимо 2 бит информации.

Далее определим стратегию угадывания задуманного числа, однако случайный процесс не всегда будет приводить к оптимальной процедуре (и чем больше диапазон, в котором могут загадываться числа, тем это будет более очевидно). Оптимальной в данном случае будем считать такую стратегию, которая за минимальное число вопросов приводит к отгадыванию. Известен метод дихотомии (половинного деления), в соответствии с которым весь интервал неопределенности делится пополам и выясняется, к какому подынтервалу принадлежит задуманное число. Половина интервала исключается из рассмотрения, а интервал, в котором лежит искомое число, вновь разбивается пополам, и процедура повторяется до тех пор, пока не будет получен единичный подынтервал. Описанная процедура наглядно может быть представлена в виде *выборочного каскада*, который показан на рис. 3.

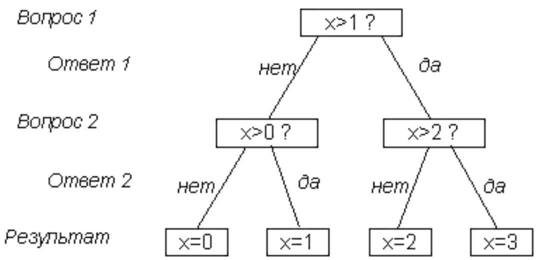


Рис. 3. Выборочный каскад для примера 3

Таким образом, для решения задачи оказалось достаточно 2-х вопросов независимо от того, какое число было задумано. Совпадение между количеством информации и числом вопросов с бинарными ответами неслучайно.

Количество информации численно равно числу вопросов с равновероятными бинарными вариантами ответов, которые необходимо задать, чтобы полностью снять неопределенность задачи.

В рассмотренном примере достаточно оказалось два ответа, чтобы полностью снять начальную неопределенность. *Подобная процедура позволяет определить количество информации в любой задаче, интерпретация которой может быть сведена к парному выбору*. Например, определение символа некоторого алфавита, использованного для представления сообщения. Приведенное утверждение перестает быть справедливым в том случае, если каждый из двух возможных ответов имеет разную вероятность – такая ситуация будет рассмотрена позднее.

Пример 4. Определить минимальное число взвешиваний, которое необходимо произвести на равноплечих весах, чтобы среди 27 внешне неотличимых монет найти одну фальшивую, более легкую.

Решение. Так как монеты внешне не отличимые, то они представляют источник с равновероятными состояниями, а общая неопределенность ансамбля, характеризующая его энтропию, поэтому составляет $H_1 = \log_2 27$ бит.

Одно взвешивание способно прояснить неопределенность ансамбля насчитывающего три возможных исхода (левая чаша весов легче, правая чаша весов легче, весы находятся в равновесии). Так как все исходы равновероятны (нельзя заранее отдать предпочтение одному из них), то результат одного взвешивания представляет источник с равновероятными состояниями, а его энтропия составляет $H_2 = \log_2 3$ бит.

Так как энтропия отвечает требованию аддитивности и при этом $H_1 = 3H_2 = 3\ \log_2 3$, то для определения фальшивой монеты достаточно произвести три взвешивания.

Чтобы построить оптимальную процедуру эксперимента, выявляющего фальшивую монету, необходимо сохранить условие равновероятности появления событий. В нашем случае общее количество из 27 монет следует разделить поровну (для обеспечения равновероятности) на 3 группы по 9 монет в каждой. Выбрать произвольным образом 2 из 3 монеты, и осуществить взвешивание. Если фальшивая монета была в одной из выбранных групп, то весы выявят более легкую, с которой и следует дальше проводить эксперимент. Если чаши весов будут находиться в равновесии, то искомая более легкая фальшивая монета находится в отложенной группе, и далее работать с этой группой. На следующем шаге 9 монет делится на 3 части по 3 монеты, 2 части взвешиваются, определяются 3 монеты, имеющие суммарно меньший вес. Окончательно для третьего взвешивания группы состоят из одной монеты, по результатам третьего взвешивания и выявится фальшивая монета. Таким образом, в отличие от рассмотренного в предыдущем примере выборочного каскада, где вопрос предполагал только два возможных ответа (да, нет), в данном случае по результатам эксперимента возможны три исхода. Если отобразить выборочный каскад, то он должен содержать по 3 исходящие дуги. Его предлагается построить самостоятельно. Очевидно, что количество информации, измеряемое в битах, $H_1 = \log_2 27$ бит, в данном случае не будет отражать оптимальную поисковую процедуру. В качестве единицы измерения информации следовало бы выбрать трит, очевидно, что общее количество информации, которое будет получено в результате эксперимента, составит $H_1 = \log_3 27 = 3$ трит, а количество информации, получаемое при каждом взвешивании, $H_2 = \log_3 3 = 1$ трит.

Пример 5. Чему равно количество информации при получении 8сообщений равномерного четырехзначного троичного кода?

Решение. Число качественных признаков m = 3. В коде они комбинируются по 4, т.е. n = 4. Число сообщений такого кода N = mn = 34. Энтропия на одно сообщение $H = \log_2 N = 4 \cdot \log_2 3$. Количество информации в 8 сообщениях

$$I = 8 \cdot 4 \cdot \log_2 3 = 50.72 \ \text{бит}.$$

Можно считать количество информации, определив энтропию на букву, блок, страницу и т.д. Количество информации в определенном объеме определяется умножением полученного значения энтропии соответственно на число букв, блоков, страниц.

Рассмотрим еще один пример, иллюстрирующий возможные различные способы передачи информации. Несмотря на то, что пример приводится для шахматной доски, военный специалист напрямую может использовать приведенные ниже рассуждения в своей профессиональной деятельности.

Пример 6. Сколькими способами можно передать положение фигур на шахматной доске? Чему равно количество информации в каждом случае [2, 5]?

Решение. Отметим, что приводимые ниже рассуждения можно использовать и в задачах с любой произвольной геометрической интерпретацией, где области могут быть разбиты на площади с равной вероятностью попадания в них.

Наиболее очевидный способ решения состоит в том, чтобы закодировать клетки шахматной доски, например, путем простой их нумерации в любой последовательности. Тогда достаточно формировать сообщение, содержащее номер клетки. Количество передаваемой информации будет равно $I = \log_2 64 = 6$ *бит*.

Кроме этого, можно предложить немного другой подход, в соответствии с которым ввести плоскую систему координат, т.е. клетку задавать парой чисел – одно число задает ее номер по вертикали, второе – по горизонтали. Для каждой из координат достаточно восьми качественных признаков (восемь по вертикали и восемь по горизонтали), но передавать в этом случае нужно уже два сообщения. При этом количество информации

$$I = 2 \cdot \log_2 8 = 2 \cdot 3 = 6 \text{ } \textit{бит}.$$

Пример 7. Определить энтропию полной многоуровневой иерархической системы, количество элементов которой на каждом уровне определяется зависимостью $l_n = K^n$, где K – основание системы, а n – номер иерархического уровня. При этом считается, что корень графа, представляющего иерархическое дерево системы, расположен на нулевом уровне. Каждый элемент системы может находиться с равной вероятностью в m состояниях.

Решение. Количество элементов N-уровневой системы можно определить по формуле

$$L = \sum_{n=0}^{N} K^n,$$

тогда энтропия системы может быть определена по формуле

$$H = \log m^L = \sum_{n=0}^{N} K^n \log m.$$

По сути полученная формула представляет собой формулу Хартли, в которой сомножитель $\sum\limits_{n=0}^{N} K^n$ позволяет оценить количество вершин иерархической системы.

В частности, на рис. 4 представлена полная 3-уровневая иерархическая система, основание системы $K = 2$, поскольку из каждой вершины исходит ровно по 2 дуги.

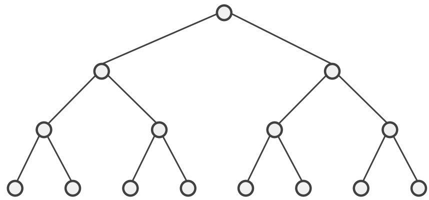


Рис. 4. Полная 3-уровневая иерархическая система

Оценим энропию системы, если каждый элемент может находиться с равной вероятностью в пяти состояниях, т.е. $m = 5$.

$$H = \sum_{n=0}^{N} K^n \log m = \sum_{n=0}^{3} 2^n \log_2 5 = \left(2^0 + 2^1 + 2^2 + 2^3\right)\log_2 5 = 15 \cdot \log_2 5 = 34{,}829\ \textit{бит}.$$

1.3. Оценка количества информации при разновероятных состояниях элементов сообщений

Рассмотренная выше оценка информации основана на предположении о равновероятности всех знаков алфавита. Формула Хартли учитывает лишь число различных возможных состояний системы, количество различных сообщений, которые может выдавать источник, но не учитывает, как часто появляется каждое из них. Существует множество ситуаций, когда возможные события имеют различные вероятности реализации. Формулу для вычисления количества информации в случае различных вероятностей событий предложил К. Шеннон в 1948 г. Так же, как и формула Хартли, формула Шеннона дает синтаксическую оценку количества информации, отвлеченно от ее смысла, семантического содержания [1, 13].

Количество информации в неравновероятных дискретных сообщениях может быть получено с учетом аддитивности количественной меры информации. Пусть имеется источник сообщений, в котором вероятности появления отдельных сообщений произвольны и в общем случае не подчиняются равномерному закону распределения, т.е. появление сообщений неравновероятно.

Пусть на основании статистического анализа известно, что в сообщении, состоящем из n элементов, каждый из которых является независимым и может принимать любое из m состояний $x_1, x_2, \ldots, x_m$, знак x_i появляется n_i раз, т.е. вероятность появления знаков равна соответственно: $p_1, p_2, \ldots, p_m$, где $p_i = \frac{n_i}{n}$.

Пусть для некоторого одиночного сообщения число элементов, принявших состояние x_1, равно n_1, число элементов, принявших состояние x_2, равно n_2 и т.д. Такое сообщение может быть представлено в виде табл. 2.

Таблица 2

Распределение вероятностей появления символов в сообщении

Состояния элементов	x_1	x_2	…	x_k	…	x_{m-1}	x_m
Вероятности состояний	p_1	p_2	…	p_k	…	p_{m-1}	p_m
Число элементов	n_1	n_2	…	n_k	…	n_{m-1}	n_m

Общее число элементов сообщения $n = \sum_{i=1}^{m} n_i$. Для рассматриваемого сообщения вероятность того, что состояние x_i примут n_i элементов, равна $p_i^{n_i}$. На основании теоремы о совмещении независимых событий вероятность того, что состояния $x_1, x_2, \ldots, x_m$ примут $n_1, n_2, \ldots, n_m$ элементов, определяется как произведение вероятностей вида $p_i^{n_i}$

$$p^* = \prod_{i=1}^{m} p_i^{n_i}. \tag{5}$$

При больших n вполне допустимо следующее приближение:

$$\frac{n_i}{n} = p_i; \;\; n_i = np_i, \tag{6}$$

которое математически обосновано теоремой Бернулли и известно как закон больших чисел. Подставив значение n_k в выражение (5), получим

$$p^* = \prod_{i=1}^{m} p_i^{np_i}. \tag{7}$$

Данная формула получена в предположении, что число элементов в передаваемом сообщении достаточно велико. Реально передаваемые сообщения не всегда удовлетворяют этому требованию и тогда выражение (7) вследствие приближенности (6) может оказаться неточным.

Однако можно рассматривать не одно отдельное сообщение, а некоторую совокупность однотипных сообщений, общее число элементов в которых достаточно велико. Тогда можно трактовать числа $n_1, n_2, \ldots, n_m$ и соответствующие им вероятности как количественные характеристики для совокупности сообщений. В этом случае выражение (7) можно рассматривать как среднюю вероятность p передачи одного сообщения из числа всех возможных сообщений совокупности [8, 12, 14]:

$$p = \prod_{i=1}^{m} p_i^{np_i} . \qquad (8)$$

По средней вероятности p передачи одного сообщения вычислим среднее число N всех возможных сообщений

$$N = \frac{1}{p} = \frac{1}{\prod_{i=1}^{m} p_i^{np_i}} . \qquad (9)$$

Зная среднее число N всех возможных сообщений, определим и среднее количество информации I, содержащееся в одном сообщении:

$$I = \log N = -\log \prod_{i=1}^{m} p_i^{np_i} , \qquad (10)$$

или после преобразования

$$I = -n \sum_{i=1}^{m} p_i \log p_i . \qquad (11)$$

Соотношение (11) было получено Шенноном для определения среднего количества информации в сообщении с произвольными вероятностями появления состояний (значений символов). При равновероятных состояниях элементов сообщений, т.е. при $p_i = \frac{1}{m}$, формула Шеннона (11) переходит в формулу Хартли (2):

$$I = -n \sum_{i=1}^{m} p_i \log p_i = -n \sum_{i=1}^{m} \frac{1}{m} \log \frac{1}{m} = n \log m . \qquad (12)$$

Зная, что энтропия – количество информации, приходящейся на один элемент сообщения, запишем формулу Шеннона для оценки количества энтропии:

$$H = \frac{I}{n} = -\sum_{i=1}^{m} p_i \log p_i \quad . \qquad (13)$$

Величину $p_i \log p_i$, характеризующую неопределенность отдельного i-го сообщения, принято называть *частной энтропией*.

Так же, как и в формуле Хартли (2), основание логарифма в формулах Шеннона (11) и (13) может быть выбрано произвольно. Естественно, что при анализе информационных процессов в радиотехнических системах и других устройствах, функционирующих на основе двоичной системы счисления, удобно пользоваться двоичными единицами. При анализе процессов в приборах, работающих в десятичной (или двоично-десятичной) системе счисления, целесообразно пользоваться дитами. В математических выкладках часто удобно пользоваться натуральными единицами – натами.

Свойства энтропии

Подобие формул информации и энтропии можно рассматривать не только с формальной точки зрения, когда определении энтропии было введено, как количество информации, приходящееся на один элемент сообщения. Здесь следует указать еще на глубинный, философский смысл такого подобия, свидетельствующий о том, что количество получаемой информации равно численно (на знак сообщения) энтропии, которая имела место относительно источника сообщений. Понятие «информация» следует рассматривать в связи с процессом снятия неопределенности, которая существовала до получения сообщения. И чем большая неопределенность была до передачи сообщения, тем большее количество информации получают в принимаемом сообщении.

В этой взаимосвязи количества информации и энтропии проявляется известный диалектический закон единства и борьбы противоположностей, так как информация рассматривается в связи со своей противоположностью – энтропией и, с другой стороны, рассматривается как мера уничтожения, снятия энтропии [8, 9].

Продемонстрируем сказанное на несложных классических примерах.

Предположим, в урне находятся 100 шаров: 50 черных и 50 белых. Вынимаем шар, сообщаем по каналу связи в какой-то условный пункт цвет шара, кладем его обратно в урну. Затем перемешиваем шары и повторяем процедуру. Вероятности получения сообщения, что вынут черный шар или белый соответственно равны $p_1 = 0{,}5$, $p_2 = 0{,}5$. Поскольку имеем дело с равновероятными состояниями, то количество информации, получаемое после извлечения одного шара, равно (2)

$$I = 1 \cdot \log_2 2 = 1 \ (\textit{бит}) \ (n = 1;\ m = 2).$$

Можно было бы использовать и формулу Шеннона (11):

$$I = -\left(p_1 \log_2 p_1 + p_2 \log_2 p_2\right) = -\left(0{,}5 \log_2 0{,}5 + 0{,}5 \log_2 0{,}5\right) = -\log_2 2^{-1} = 1 \ \textit{бит}.$$

Теперь предположим, что в урне находится 99 черных и 1 белый шар. Повторяем опыт. Вероятности узнать о том, что вынут черный или белый шар, соответственно равны $p_1 = 0{,}99$, $p_2 = 0{,}01$. Количество информации, получаемое при извлечении одного шара, равно (2.10)

$$I = -\left(p_1 \log_2 p_1 + p_2 \log_2 p_2\right) = -\left(0{,}99 \log_2 0{,}99 + 0{,}01 \log_2 0{,}01\right) =$$
$$= 0{,}0143 + 0{,}0664 = 0{,}0807 \ \textit{бит}.$$

В данном случае заранее предсказать, какой будет вынут шар, не представляет особого труда, так как мы почти уверены в результате опыта. Количество получаемой информации мало. В первом случае, наоборот, трудно предсказать исход опыта, и неопределенность при равных вероятностях извлечения белого и черного шара является максимальной.

Итак, приходим к выводу, что чем больше априорная неопределенность, тем большее количество информации мы получаем при снятии ее. В этом смысле величина, определяющая степень неопределенности, является мерой оценки количества информации и удобна при исследовании свойств информации.

В теории информации мерой неопределенности является энтропия – удельное количество информации, приходящееся на один элемент сообщения. Для сообщения из n элементов эта величина равна [14]

$$H = \frac{I}{n} = -\sum_{i=1}^{m} p_i \log p_i = \sum_{i=1}^{m} p_i \log \frac{1}{p_i}. \qquad (14)$$

Величина $\log \frac{1}{p_i}$ называется частной энтропией, характеризующей информативность лишь *i*-го состояния. Как видно из выражения (14), энтропия есть среднее значение частных энтропий.

Количество информации *I*, содержащееся в сообщении, зависит от числа элементов сообщения *n*, числа возможных состояний *m* и вероятностей состояний p_k. Наибольший интерес представляет зависимость количества информации от *m* и p_i, поскольку зависимость от *n* является линейной. Поэтому более подробно рассмотрим свойства энтропии – количества информации, приходящейся на один элемент сообщения. При этом считаем, что энтропия характеризует меру неопределенности совокупности событий, составляющих полную группу, т.е. $\sum_{i=1}^{m} p_i = 1$. В противном случае теряет свою справедливость выражение для энтропии (14).

К основным свойствам энтропии следует отнести следующие.

1. При равновероятности знаков алфавита $p_i = 1/m$ из формулы Шеннона получают:

$$H = -\sum_{i=1}^{m} p_i \log p_i = -\sum_{i=1}^{m} \frac{1}{m} \log \frac{1}{m} = -\left(m \frac{1}{m}\right)\left(-\log \frac{1}{m}\right) = \log m .$$

Из этого следует, что при равновероятности знаков алфавита энтропия определяется исключительно числом знаков *m* алфавита и по существу является характеристикой только алфавита.

Следующие два свойства связаны с двумя предельными случаями и имеют некоторые вычислительные особенности, о которых необходимо упомянуть отдельно.

2. Если $m = 1$, то передача иных элементов в принципе не предусмотрена, следовательно, вероятность его появления $p_i = 1$. Формула для вычисления количества энтропии преобразовывается к тривиальному виду $H = -1 \log 1 = 0$. Этот результат согласуется с эмпирическим представлением об информации и энтропии. Источник, который может генерировать единственное сообщение, неопределенностью не обладает. Получатель и без сообщения осведомлен о возможной генерации единственного сообщения, а потому никакой информации не получит.

Таким образом, энтропия детерминированных сообщений равна нулю.

3. Если по результатам приема и передачи сообщений установлено, что некоторые сообщения не появляются ни разу, то можно поставить вопрос о практической целесообразности сохранять такое сообщение во множестве потенциально возможных сообщений для источника. Скорее всего, его следует исключить. Однако с математической точки зрения можно проанализировать выражение, которое получится при $p_i = 0$. Слагаемое в формуле Шеннона (14) принимает вид неопределенности типа $(0 \cdot \infty)$. Действительно, $p_i \log \frac{1}{p_i} = 0 \cdot \infty$. Раскроем эту неопределенность:

$$\lim_{p_i \to 0} \left\{ p_i \log \frac{1}{p_i} \right\} = \lim_{p_i \to 0} \left\{ \log \frac{1/p_i}{1/p_i} \right\}.$$

Обозначим $1/p_i = \alpha$; при $p_i \to 0$, $\alpha \to \infty$, тогда согласно правилу Лопиталя

$$\lim_{p_i \to 0} \left\{ \log \frac{1/p_i}{1/p_i} \right\} = \lim_{\alpha \to \infty} \left\{ \log \frac{\alpha}{\alpha} \right\} = \lim_{\alpha \to \infty} \left\{ \frac{1}{\alpha} \log e \right\} = 0.$$

Таким образом, были определены два предельных случая, когда выражение, стоящее под знаком суммы энтропий (14), минимально и равно нулю.

Исследуем теперь выражение под знаком суммы на экстремум. Для этого, как известно, необходимо найти точку, в которой первая производная исследуемой функции меняет свой знак, т.е.

$$\frac{d}{dp_i}\left(-p_i \log p_i\right) = -\log p_i - \log e = -\log p_i e = 0.$$

Откуда получаем, что функция $p_i \log p_i$ при $p_i = 1/e = 0{,}37$ имеет максимум $-p_i \log p_i = \frac{1}{e} \log e = 0{,}531$, т.е. координаты максимума (0,37; 0,531).

4. Вышесказанное позволяет сделать вывод о том, что энтропия есть величина вещественная, положительная и имеет экстремум. График на рис. 5 наглядно демонстрирует названные свойства. Кроме того, можно отметить еще одну особенность энтропии, связанную с вкладом каждого

слагаемого в общую величину энтропии. Как видно из графика, при значении $p_i < 0{,}1$ величина $p_k \log \frac{1}{p_k}$ растет круто. Это означает, что на этом участке даже незначительное уменьшение вероятности ведет к резкому уменьшению слагаемого $p_i \log \frac{1}{p_i}$, т.е. при малых значениях вероятности p_i члены, содержащие p_i, не играют существенной роли в выражении энтропии. Наибольшие значения слагаемые вида $p_i \log \frac{1}{p_i}$ достигают при значениях вероятности появления элемента с i-м состоянием, лежащих в области от 0,2 до 0,6. Так как при малых вероятностях появления i-го состояния легко предсказать его отсутствие в сообщении и, наоборот, при больших вероятностях i-го состояния легко предсказать его наличие в сообщении, то в обоих случаях величина неопределенности, существующей до получения сообщения, будет мала. Невелико и количество информации при снятии этой неопределенности [3, 4, 7].

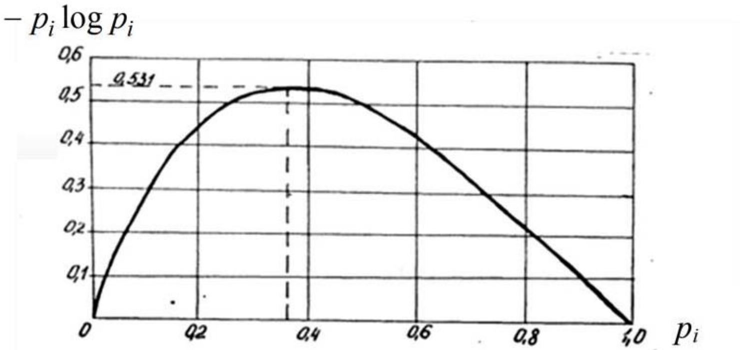


Рис. 5. Зависимость энтропии от вероятности

5. Рассмотрим случай, когда $m = 2$. Тогда в отличие от рассмотренного выше в п. 2 свойства, детерминированность исчезает. Предположим, что одно состояние имеет вероятность появления p, тогда вероятность появления второго составит $1 - p$, и энтропия будет равна

$$H = -\sum_{i=1}^{2} p_i \log p_i = -\{p \log p + (1-p)\log(1-p)\}\cdot$$

Формулу для двух состояний удобно проанализировать на свойства энтропии, соответствующие графики представлены на рис. 6. Из графика видно, что энтропия бинарного сообщения изменяется от 0 до 1 и достигает максимального значения, равного 1 биту, при равных вероятностях ($p_1 = p_2 = 0{,}5$) появления в сообщении обоих состояний. Если графики отдельных составляющих асимметричны, то график суммарной энтропии является симметричным.

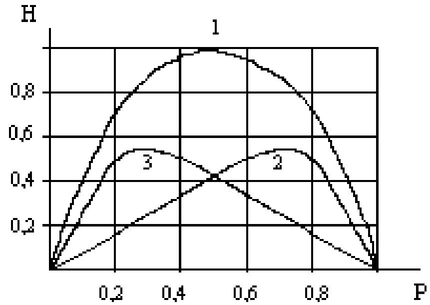


Рис. 6. График зависимости энтропии H двоичных сообщений (1) и ее составляющих (2, 3): $-(1-p)\log(1-p)$ и $-- p\log p$ от p

Доказано [1, 4, 6], что для сообщений, в которых количество состояний элементов более двух ($m > 2$), максимальная энтропия также будет при соблюдении условия равной вероятности появления состояний в сообщении.

Если знаки алфавита неравновероятны, то алфавит можно рассматривать как дискретную случайную величину, заданную статистическим распределением частот n_i появления знаков x_i (или вероятностей $p_i = n_i / n$) (см. табл. 2). Такие распределения получают обычно на основе статистического анализа конкретных типов сообщений от источника. Поэтому, если знаки алфавита неравновероятны и хотя формально в выражение для энтропии входят только характеристики алфавита (вероятности появления его знаков), то энтропия отражает статистические свойства некоторой совокупности сообщений.

Энтропия максимальна, если все знаки алфавита равновероятны, т.е. $H_{max} = \log m$.

6. Одной из причин отказа в ряде практических случаев от квантовой меры информации является отсутствие свойства аддитивности, если имеется более одного источника информации. Как известно, этот недостаток был успешно преодолен в формуле Хартли, свойство аддитивности для независимых источников сообщений в полной мере присуще и формуле Шеннона, которая является более общей.

Рассмотрим случай двух независимых сообщения A и B, которые имеют энтропии $H(A)$ и $H(B)$ соответственно. Как известно из теории вероятности, вероятность совместного события AB равна произведению вероятностей событий A и B: $p_{ij} = p_i p_j = p(A)p(B)$. Тогда [4, 7]

$$H(AB) = -\sum_i \sum_j p_{ij} \log p_{ij} = -\sum_i \sum_j p_i p_j \log\left(p_i p_j\right) =$$

$$= -\sum_i \sum_j p_i p_j \left(\log p_i + \log p_j\right) = -\sum_i p_i \log p_i \sum_j p_j - \sum_j p_j \log p_j \sum_i p_i .$$

Так как $\sum_i p_i = 1$ и $\sum_j p_j = 1$, то

$$H(AB) = -\sum_i p_i \log p_i - \sum_j p_j \log p_j = H(A) + H(B) \cdot \qquad (15)$$

Приведенное соотношение (15) может быть обобщено на любое количество сообщений ($n > 2$). Таким образом, если имеется несколько независимых источников сообщений, то общее количество информации (энтропии), получаемое (снимаемое) получателем, следует определять по правилу сложения. Это свойство (15).

Пример 8. Приемопередающей подсистемой телекоммуникационной системы были зафиксированы сообщения от некоторого источника, в результате чего установлено, что в абстрактном алфавите имеется 4 буквы – a_1, a_2, a_3, a_4, кроме того, частота появления букв оказалась различной. После набора статистики достаточного объема стало возможным оценить вероятности появления символов на выходе этого источника, которые представлены в виде следующего ряда:

$$A = \begin{vmatrix} a_1 & a_2 & a_3 & a_4 \\ 0{,}1 & 0{,}3 & 0{,}2 & 0{,}4 \end{vmatrix}.$$

Оценить энтропийный объем получаемых сообщений.

Решение. Если бы вероятности появления символов были одинаковыми, то можно было бы воспользоваться формулой Хартли, и энтропия была бы равна $H = \log_2 4 = 2$ *бит*. Однако вероятности дифференцированы, поэтому используем формулу Шеннона (13):

$$H = -\sum_{i=1}^{4} p_i \log p_i = -(0{,}1 \cdot \log_2 0{,}1 + 0{,}3 \cdot \log_2 0{,}3 + 0{,}2 \cdot \log_2 0{,}2 + 0{,}4 \cdot \log_2 0{,}4) =$$
$$= 1{,}84644 \text{ } \textit{бит}.$$

Как видно, дифференциация вероятностей закономерно привела к уменьшению энтропии.

Пример 9. Определить энтропию системы, состоящей из двух подсистем. Первая подсистема состоит из трех элементов, каждый из которых может находиться в двух состояниях с вероятностями $p_1 = 0{,}6$ и $p_2 = 0{,}4$. Вторая подсистема состоит из двух элементов, каждый из которых может находиться в трех состояниях с вероятностями: $p_1 = 0{,}1$, $p_2 = 0{,}4$, $p_2 = 0{,}5$.

Решение. При решении данной задачи следует помнить об аддитивности информации и энтропии. Определим энтропию на элемент первой подсистемы:

$$H_1 = -\sum_{i=1}^{2} p_i \log p_i = -(0{,}6 \cdot \log_2 0{,}6 + 0{,}4 \cdot \log_2 0{,}4) = 0{,}971 \text{ } \textit{бит}.$$

Энтропия первой подсистемы с учетом количества элементов равна $3H_1$.

Определим энтропию на элемент второй подсистемы:

$$H_2 = -\sum_{i=1}^{3} p_i \log p_i = -(0{,}1 \cdot \log_2 0{,}1 + 0{,}4 \cdot \log_2 0{,}4 + 0{,}5 \cdot \log_2 0{,}5) = 1{,}361 \text{ } \textit{бит}.$$

Энтропия второй подсистемы с учетом количества элементов равна $2H_2$.

Окончательно энтропия всей системы будет равна

$$H = 3H_1 + 2H_2 = 3 \cdot 0{,}971 + 2 \cdot 1{,}361 = 5{,}635 \text{ } \textit{бит}.$$

Пример 10. Приемник инфокоммуникационной системы фиксирует сообщения, содержащие 4 различных символов абстрактного алфавита. Вероятности появления этих символов оказались равны следующим значениям: $p(x_1) = 0{,}1$ $p(x_2) = 0{,}1$ $p(x_3) = 0{,}1$ $p(x_4) = 0{,}7$. Определить минимальное количество символов вторичного алфавита, способного передавать такой же объем энтропии.

Решение. Определим энтропию источника с разновероятными знаками алфавита, воспользовавшись формулой Шеннона:

$$H(\alpha)=-p(x_1)\log p(x_1)-p(x_2)\log p(x_2)-p(x_3)\log p(x_3)-p(x_4)\log p(x_4)=$$
$$=-3(0{,}1\ \log 0{,}1)-0{,}7\log 0{,}7=1{,}36.$$

Полученное значение энтропии подставим в формулу Хартли, откуда и определим мощность алфавита с равновероятным появлением знаков:

$$H(\beta)=1{,}357=\log_2 m,$$
$$m=2^{1{,}357}=2{,}561.$$

Поскольку число знаков алфавита должно быть целым числом, то округляем полученный результат до 3, т.е. источник с равновероятным появлением знаков в сообщении должен иметь 3 знака.

1.4. Представление сообщений и определение энтропии при непрерывном распределении состояний элементов

В предыдущих разделах была рассмотрена мера неопределенности выбора для дискретного источника информации. Радиотехнические системы помимо дискретных подсистем имеют также и аналоговые блоки, и если они являются источниками информации, то множество их возможных состояний составляет континуум. Такие источники называют *непрерывными* источниками информации [4, 7].

Во многих случаях они преобразуются в дискретные посредством использования устройств дискретизации и квантования. Вместе с тем существует немало и таких систем, в которых информация передается и преобразуется непосредственно в форме непрерывных сигналов. Примерами могут служить системы телефонной связи, телевидения и некоторых подсистем радиотехнических систем.

Оценка неопределенности выбора для непрерывного источника информации имеет некоторую специфику. Во-первых, значения, реализуемые источником, математически отображаются случайной непрерывной величиной. Во-вторых, вероятности значений этой случайной величины не могут использоваться для оценки неопределенности, поскольку в данном случае вероятность любого конкретного значения равна нулю. Естественно, связывать неопределенность выбора значения случайной непрерывной величины с плотностью распределения вероятностей этих значений. Учитывая, что для совокупности значений, относящихся к любому сколь угодно малому интервалу случайной непрерывной величины, вероятность конечна, попытаемся найти формулу для энтропии непрерывного источника информации, используя операции квантования и последующего предельного перехода при уменьшении кванта до нуля.

Пусть имеются сообщения, элементы которых могут принимать любые состояния на некотором интервале состояний элементов. Подобные сообщения будем называть непрерывными по состоянию элементов.

Рассмотрим диаграмму состояний такого сообщений (рис. 7) [4, 7]. Через x_{min} и x_{max} обозначим границы интервала состояний элементов. Интервал состояний может быть ограничен сверху и снизу либо иметь неограниченные границы с одной или с двух сторон. Через x_1, x_2,…, x_i,…, x_n обозначены текущие значения состояний элементов, характерные для данного частного сообщения, состоящего из n элементов. Статистической характеристикой состояний элементов является функция плотности распределения вероятностей состояний $\omega(x)$ (рис. 8).

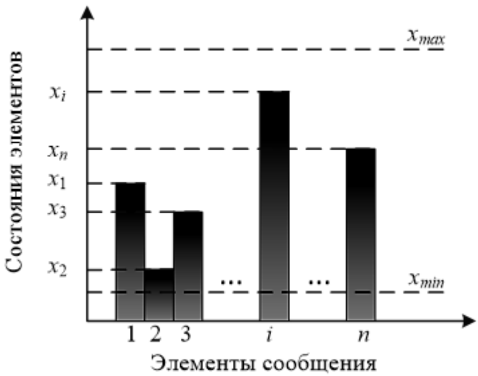


Рис. 7. Плотность вероятности дискретного распределения

Кривая распределения характеризует вероятность попадания случайной непрерывной величины x на некоторый элементарный участок dx (с точностью до бесконечно малых высшего порядка) либо вероятность пребывания случайной величины x в пределах от α до β [11]:

$$p(\alpha < x < \beta) = \int_{\alpha}^{\beta} \omega(x)\,dx \cdot$$

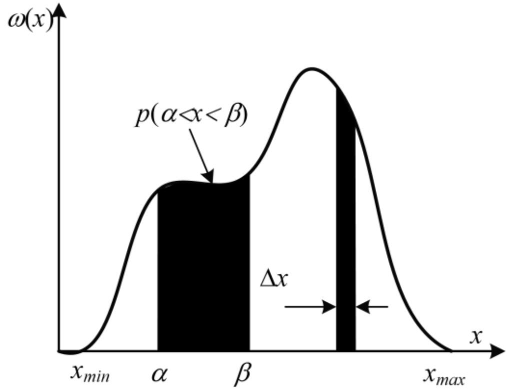


Рис. 8. Плотность вероятности непрерывного распределения

На необходимость анализа сообщений с непрерывным распределением состояний указывает тот факт, что многие физические величины, передаваемые по каналам связи, имеют непрерывные распределения состояний. Такими распределениями обладают, например, аналоговые блоки и подсистемы радиотехнических систем.

Обобщим понятие энтропии, введенное для сообщений, дискретных по состояниям элементов, на случай непрерывных сообщений. Пусть задана плотность распределения вероятностей состояний элементов непрерывного сообщения. Характер процесса существенно не изменится, если непрерывные состояния заменить дискретными, отстоящими друг от друга на расстоянии Δx и имеющими вероятности $p_k = \omega(x_k)\Delta x$. Такая замена будет тем более точной, чем меньше Δx (см. рис. 8).

Энтропия эквивалентного сообщения [8, 13]

$$H(X) = -\sum_{k=1}^{m} p_k \log p_k = -\sum_{k=1}^{m} \omega(x_k)\Delta x \log\{\omega(x_k)\Delta x\} =$$

$$= -\sum_{k=1}^{m} \omega(x_k)\Delta x \log\{\omega(x_k)\} - \sum_{k=1}^{m} \omega(x_k)\Delta x \log \Delta x.$$

При $\Delta x \to 0$ имеем

$$H(X) = -\int_{-\infty}^{+\infty} \omega(x) \log \omega(x) dx - \log \Delta x \cdot \tag{16}$$

Обозначим

$$H^*(X) = -\int_{-\infty}^{+\infty} \omega(x) \log \omega(x) dx, \tag{17}$$

тогда

$$H(X) = H^*(X) - \log \Delta x \cdot \tag{18}$$

Первое слагаемое в правой части соотношения имеет конечное значение, которое зависит только от закона распределения непрерывной случайной величины X и не зависит от шага квантования. Оно имеет точно такую же структуру, как энтропия дискретного источника. Первое слагаемое в этой сумме, называемое также ***приведенной энтропией***, целиком определяет информативность сообщений, обусловленных статистикой состояний их элементов. Из выражений (17) и (18) кроме того видно, что информативность сообщений, обусловленная статистикой состояний элементов сообщения, целиком определяется величиной $H^*(x)$.

Поскольку для определения этой величины используется только функция плотности вероятности, т.е. дифференциальный закон распределения, она получила название относительной дифференциальной энтропии или просто ***дифференциальной энтропии*** непрерывного источника информации (непрерывного распределения случайной величины X).

Величина $\log\Delta x$ зависит только от выбранного интервала Δx**,** определяющего точность квантования состояний, и при Δx = const она постоянна. Очень часто величину $\log\Delta x$ исключают из рассмотрения, полагая [10, 14]

$$H(X) = -\int_{-\infty}^{+\infty} \omega(x) \log \omega(x) dx, \tag{19}$$

что не является достаточно строгим, так как для многих задач при определении информативности сообщений необходимо учитывать не только статистику состояний элементов сообщения, но и величину квантования состояния Δx.

Как видно из (16) или (19), энтропия и количество информации зависят от распределения плотности вероятностей $p(x)$.

В теории информации большое значение имеет решение вопроса о том, при каком распределении обеспечивается максимальная энтропия $H(x)$. Несколько слов о виде функции плотности распределения вероятностей состояний элементов непрерывных сообщений, обладающих максимальной энтропией. Решение задачи нахождения такой функции имеет большое значение для оценки максимальной скорости передачи информации, для оценки наибольшей эффективности помех и др. При этом возможны дополнительные условия, накладываемые на свойства сообщений. Особый интерес представляют два случая [8, 9].

1. Дисперсия состояний элементов сообщения – величина конечная и постоянная:

$$\int_{-\infty}^{\infty}(x-m_x)^2\,\omega(x)dx=\sigma^2=Dx=const.$$

Если, например, под состоянием понимать напряжение (амплитудная модуляция), то дисперсия во многих частных случаях – пропорциональна мощности сообщений. Доказано, что в этом случае сообщение обладает наибольшей информативностью (максимальной энтропией), когда состояния элементов распределены по нормальному закону:

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

Так как дисперсия определяет среднюю мощность сигнала, то отсюда следуют практически важные выводы: а) если задана средняя мощность помехи, то помеха является наиболее эффективной (энтропия помехи максимальна), когда состояния составляющих помеху элементов распределены по нормальному закону; б) передача наибольшего количества информации при заданной мощности сигнала (или наиболее экономичная передача информации) достигается при такой обработке сигнала, которая приближает распределение плотности вероятности его элементов к нормальному распределению.

В то же время, обеспечивая нормальное распределение плотности вероятности элементам помехи, обеспечивают ее наибольшую “ информативность”, т.е. наибольшее пагубное воздействие на прохождение сигнала. Найдем значение энтропии, когда состояния элементов источника сообщений распределены по *нормальному закону*:

$$H = -\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \log\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}\right] dx - \log \Delta x =$$

$$= \log\left(\sigma\sqrt{2\pi}\,\frac{1}{\sigma\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx + \frac{\log e}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx - \log \Delta x = \log\left(\frac{\sigma}{\Delta x}\sqrt{2\pi e}\right).$$

Выражение (16) для энтропии сообщений, состояния элементов которых распределены по нормальному закону, после преобразования принимает вид [10]

$$H(X) = \log\left\{\frac{\sigma}{\Delta x}\sqrt{2\pi e}\right\}. \tag{20}$$

2. Дисперсия состояний элементов сообщения не ограничена. В этом случае сообщения обладают наибольшей информативностью (максимальной энтропией), когда состояния элементов распределены внутри интервала по закону *равномерной* плотности:

$$\omega(x) = \begin{cases} \dfrac{1}{b-a} & \textit{при}\ \ a \le x \le b; \\ 0 & \textit{при}\ \ x < a,\ x > b. \end{cases}$$

Выражение (16) для энтропии сообщений, состояния элементов которых распределены по закону равномерной плотности, после преобразования принимает вид

$$H(X) = -\frac{1}{b-a}\int_a^b \log\left(\frac{1}{b-a} - \Delta x\right) dx = \log\frac{b-a}{\Delta x}. \tag{21}$$

Дисперсия равномерного распределения $\sigma_p^2 = \dfrac{(b-a)^2}{12}$, поэтому $(b - a) = 2\sqrt{3}\sigma_p$. С учетом этого можно записать

$$H_p(x) = \log\left(\frac{\sigma_p}{\Delta x} 2\sqrt{3}\right).$$

Сравнивая между собой сообщения с равномерным и нормальным распределением вероятностей при условии $H_н(x) = H_р(x)$, получаем

$$\sigma_p^2 = \frac{e}{6}\sigma^2 \approx 1{,}42\sigma^2.$$

Пример 11. РЛС в некотором секторе обнаруживает воздушные цели и фиксирует их курс. Определить энтропию, связанную с фиксацией курса каждой воздушной цели при условии, что цели могут двигаться в произвольном направлении с равной вероятностью.

Решение. Рассмотрим каждую обнаруженную воздушную цель как случайную величину. Поскольку цели могут двигаться в произвольном направлении с равной вероятностью, то имеет место равномерное распределение вероятностей углов от 0 до 360°. Таким образом, функция плотности имеет вид

$$\omega(x) = \frac{1}{360} \cdot$$

По формуле (19) имеем

$$H(x) = -\int_{-\infty}^{\infty} \omega(x) \log \omega(x) dx = -\int_{0}^{360} \frac{1}{360} \log \frac{1}{360} dx = -\frac{1}{360} \log \frac{1}{360} \cdot x \Big|_{0}^{360} =$$

$$= -\log \frac{1}{360} = \log 360 = 8{,}49 \ \textit{бит}.$$

Пример 12. Определить энтропию непрерывного источника сообщений, если функция плотности распределения вероятностей состояний элементов непрерывных сообщений имеет вид, показанный на рис. 9.

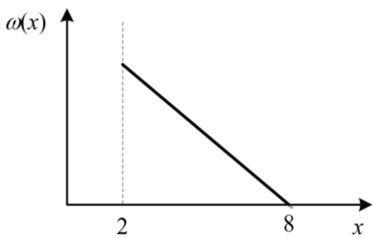


Рис. 9. Функция плотности распределения вероятностей состояний элементов непрерывных сообщений

Решение. Чтобы воспользоваться формулой (19), необходимо получить аналитическое выражение для функции плотности распределения вероятностей состояний элементов непрерывных сообщений ω(x), которая в нашем случае представляет собой прямую. Уравнение прямой ω(x), проходящей через две различные точки на плоскости (x_1, y_1) и (x_2, y_2), получим из известного соотношения

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}, \tag{22}$$

Координаты одной из точек известны из графика (см. рис. 9) – это (8, 0). Ординату второй точки можно найти, если учесть, что функция плотности любого распределения удовлетворяет условию

$$\int_{-\infty}^{\infty} \omega(x)dx = 1,$$

т.е. площадь под прямой равна 1. Фигура под прямой представляет собой треугольник, один из катетов равен 8 – 2 = 6, тогда ординату найдем из уравнения

$$1 = \frac{1}{2} \cdot 6 \cdot y_2,$$

$$y_2 = \frac{1}{\frac{1}{2} \cdot 6} = \frac{1}{3}.$$

Координаты второй точки $\left(2, \frac{1}{3}\right)$, подставим координаты найденных точек в (22):

$$\frac{x-8}{2-8} = \frac{y-0}{1/3-0},$$

$$x - 8 = -6 \cdot 3 \cdot y,$$

$$y = -\frac{x}{18} + \frac{4}{9}.$$

Полученную функцию плотности распределения $\omega(x) = -\frac{x}{18} + \frac{4}{9}$ подставим в (19), для упрощения расчетов вычислим энтропию в нитах, пределы интегрирования определяются областью определения ω(x) (см. рис. 9):

$$H(x) = -\int_{2}^{8} \left(-\frac{x}{18} + \frac{4}{9}\right) \cdot \ln\left(-\frac{x}{18} + \frac{4}{9}\right) dx,$$

введем замену переменной $u = \frac{4}{9} - \frac{x}{18}$, откуда $x = 8 - 18u$. Тогда $dx = -18udu$. В результате подстановки имеем

$$-\int_{2}^{8} u \cdot \ln(u) \cdot (-18)du = 18\int_{2}^{8} u \cdot \ln(u)du.$$

Далее воспользуемся таблицей трансцендентных интегралов, в соответствии с которой

$$\int x^n \cdot \ln(x)dx = x^{n+1}\left[\frac{\ln x}{n+1} - \frac{1}{(n+1)^2}\right] + C \cdot$$

Для нашего интеграла

$$18\int_2^8 u \cdot \ln(u)du = 18u^2\left[\frac{\ln u}{2} - \frac{1}{2^2}\right]\Bigg|_2^8 = 9u^2 \cdot \ln u - \frac{9u^2}{2}\Bigg|_2^8 \cdot$$

Осуществляем обратную замену $u = \frac{4}{9} - \frac{x}{18}$:

$$9u^2 \cdot \ln u - \frac{9u^2}{2}\Bigg|_2^8 = 9\left(\frac{4}{9} - \frac{x}{18}\right)^2 \cdot \ln\left(\frac{4}{9} - \frac{x}{18}\right) - \frac{9\left(\frac{4}{9} - \frac{x}{18}\right)^2}{2}\Bigg|_2^8 =$$

$$= 9\left(\frac{4}{9} - \frac{8}{18}\right)^2 \cdot \ln\left(\frac{4}{9} - \frac{8}{18}\right) - \frac{9\left(\frac{4}{9} - \frac{8}{18}\right)^2}{2} - \left(9\left(\frac{4}{9} - \frac{2}{18}\right)^2 \cdot \ln\left(\frac{4}{9} - \frac{2}{18}\right) - \frac{9\left(\frac{4}{9} - \frac{2}{18}\right)^2}{2}\right) =$$

$$= 0 - 9\left(\frac{1}{3}\right)^2 \cdot \ln\left(\frac{1}{3}\right) + \frac{9\left(\frac{1}{3}\right)^2}{2} = \ln(3) + \frac{1}{2} = 1.59861 \textit{ нит.}$$

1.5. Условная энтропия и ее свойства

До сих пор предполагалось, что все элементы сообщения независимы, т.е. появление каждого данного элемента никак не связано с предшествующими элементами.

Согласно правилу сложения энтропий количество информации, содержащееся в двух различных по содержанию (независимых) книгах, равна сумме количеств информации, содержащихся в отдельных книгах. Однако если одна книга содержит часть другой, то количество информации от двух таких книг не будет равно сумме количеств информации от книг в отдельности, а будет несколько меньше. Большое значение для определения количества информации, содержащейся в статистически зависимых событиях, имеет понятие условной энтропии [8, 9].

Понятие условной энтропии в теории информации используется при определении взаимозависимости между символами кодируемого алфавита, для определения потерь при передаче информации по каналам связи, при вычислении энтропии объединения.

Пусть имеется два ансамбля сообщений. Ансамблем принято называть множество с известным распределением вероятностей его элементов. Элементы первого ансамбля сообщений принимают состояния $x_1, x_2, \ldots, x_m$ с вероятностями $p(x_1), p(x_2),\ldots, p(x_m)$, и элементы второго ансамбля принимают состояния $y_1, y_2, \ldots, y_n$ с вероятностями $p(y_1), p(y_2),\ldots, p(y_n)$. Вследствие зависимости сообщений от заданного состояния x_k вероятности появления состояния ансамбля Y определяется условными вероятностями:

$$p\left(y_1/x_k\right), p\left(y_2/x_k\right),\ldots, p\left(y_j/x_k\right),\ldots, p\left(y_n/x_k\right).$$

Здесь $p(y_j/x_k)$ – условная вероятность появления символа y_j при условии, что появился символ x_k. Чем большая статистическая связь существует между ансамблями X и Y, тем большая разница будет между наибольшим и наименьшим значениями условных вероятностей. В частном случае, когда статистическая зависимость наиболее сильная, определенному состоянию x_k соответствует одно состояние из множества Y (например, y_j) и тогда одна условная вероятность примет наибольшее значение, равное единице $p\left(y_j/x_k\right)=1$, а остальные условные вероятности примут наименьшее значение. В общем случае полная система условных вероятностей определяется матрицей $n \times m$ [7, 9]:

$$\left\| p\left(y_j/x_k\right) \mid j=\overline{1,n};\ k=\overline{1,m} \right\|. \tag{23}$$

Систему двух случайных величин (сообщений) X, Y можно изобразить случайной точкой на плоскости. Событие, состоящее в попадании случайной точки (X, Y) в область D, принято обозначать в виде $(X, Y) \subset$ D.

Закон распределения системы двух случайных дискретных величин может быть задан с помощью табл. 3.

В табл. 3 p_{ij} – вероятность события, заключающегося в одновременном выполнении равенства $X = x_i$, $Y = y_j$. При этом $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1$.

Таблица 3

Закон распределения двумерной случайной величины

Y / X	y_1	y_2		y_n
x_1	p_{11}	p_{12}	…	p_{1n}
x_2	p_{21}	p_{22}	…	p_{2n}
…	:	:	:	:
x_m	p_{m1}	p_{m2}	…	p_{mn}

Закон распределения системы случайных непрерывных величин (X, Y) задают при помощи функции плотности вероятности $p(x, y)$. Вероятность попадания случайной точки (X, Y) в область D определяется равенством $P[(X,Y)\subset D]=\iint\limits_D p(x,y)dxdy$.

Напомним известные из курса теории вероятностей важнейшие свойства функция плотности распределения вероятности:

1) $p(x, y) \geq 0$;

2) $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p(x,y)dxdy=1$.

Если все случайные точки (X, Y) принадлежат области D, то $\iint\limits_D p(x,y)dxdy=1$.

Условным распределением составляющей X при $Y = y_j$ (y_j сохраняет одно и то же значение при всех возможных значениях X) называют совокупность условных вероятностей $p(y_j/x_1), p(y_j/x_2),..., p(y_j/x_m)$. Аналогично определяется условное распределение составляющей Y.

Для фиксированного состояния x_k совокупность условных вероятностей определяет частную условную энтропию [8, 11]

$$H(Y/x_k)=-\sum_{j=1}^{n} p(y_j/x_k)\log p(y_j/x_k), \tag{24}$$

которая характеризует информативность сообщений Y после того, как стало известно состояние x_k. Если частную условную энтропию усреднить по всем состояниям x_k с учетом вероятности появления каждого из состояний $p(x_k)$, то найдем общую условную энтропию сообщений Y относительно сообщений X:

$$H(Y/X)=\sum_{k=1}^{m}p(x_k)H(Y/x_k)$$

или с учетом выражения (24)

$$H(Y/X)=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k)\,p(y_j/x_k)\log p(y_j/x_k)\cdot \qquad (25)$$

Известно, что вероятность совместного появления двух статистически зависимых состояний $x_k\ y_j$ определяется равенством

$$p(x_k,y_j)=p(x_k)\,p(y_j/x_k) \qquad (26)$$

или

$$p(x_k,y_j)=p(y_j)\,p(x_k/y_j). \qquad (27)$$

Условные вероятности составляющих X и Y вычисляют соответственно по формулам:

$$p(y_j/x_k)=\frac{p(x_k,y_j)}{p(x_k)};\ \ p(x_k/y_j)=\frac{p(x_k,y_j)}{p(y_j)}.$$

Для контроля вычислений целесообразно убедиться, что сумма вероятностей условного распределения равна единице.

Поэтому, преобразуя (25) согласно (26), имеем [8, 11]

$$H(Y/X)=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k,y_j)\log p(y_j/x_k)\cdot \qquad (28)$$

Основной смысл средней условной энтропии (или просто условной энтропии) состоит в том, что она показывает, какую энтропию дает сообщение Y, когда уже известна энтропия сообщений X.

Различают понятия частной и общей условной энтропии. Выражения (24) представляют собой частные условные энтропии. Общая условная энтропия сообщения Y относительно сообщения X характеризует количество информации, содержащееся в любом символе алфавита, и определяется усреднением по всем символам, т.е. по всем состояниям с учетом вероятности появления каждого из состояний, и равна сумме вероятностей появления символов алфавита на неопределенность, которая остается после того, как адресат принял сигнал. Выражения (25) и (28) являются общим выражением для определения количества информации на один символ сообщения для случая неравномерных и взаимозависимых символов.

Основные свойства условной энтропии

Перечислим наиболее важные для практических случаев свойства условной энтропии [3, 8, 9].

1. Если сообщения Y и X статистически независимы, то условная энтропия сообщений Y относительно сообщений X равна безусловной энтропии сообщений $H(Y)$:

$$H(Y/X) = H(Y).$$

В этом случае вся информация, которую содержат сообщения Y, является новой по отношению к информации, содержащейся в сообщениях X.

Действительно, если сообщения Y и X статистически независимы, то $p\left(y_j / x_k\right) = p\left(y_j\right)$, и условная энтропия может быть записана в виде

$$H\left(Y / X\right) = -\sum_{k=1}^{m}\sum_{j=1}^{n} p\left(x_k\right) p\left(y_j\right)\log p\left(y_j\right) = \sum_{j=1}^{n} p\left(y_j\right)\log p\left(y_j\right)\sum_{k=1}^{m} p\left(x_k\right).$$

Так как сумма вероятностей всех сообщений ансамбля X равна единице $\sum_{k=1}^{m} p\left(x_k\right) = 1$, то

$$H\left(Y / X\right) = \sum_{j=1}^{n} p\left(y_j\right)\log p\left(y_j\right) = H\left(Y\right).$$

2. Если сообщения X и Y являются статистически жестко связанными, т.е. появление одного из них непременно подразумевает появление другого, то условная энтропия сообщений Y относительно сообщений X равна нулю:

$$H(Y/X) = 0.$$

Это означает, что сообщения Y не содержат никакой новой информации сверх той, которая содержится в сообщениях X.

Справедливость данного утверждения вытекает из следующего свойства условных вероятностей: при полной статистической зависимости $p\left(y_j / x_k\right) = 1$ либо $p\left(y_j / x_k\right) = 0$. Тогда все слагаемые выражения (25) обращаются в ноль и $H(Y/X) = 0$.

3. Энтропия при взаимно зависимых элементах всегда меньше, чем при независимых, т.е. $H' < H$.

Это можно количественно показать на несложном примере. Пусть алфавит состоит из двух элементов 0 и 1. Если эти элементы равновероятны, то количество информации, приходящееся на один элемент сообщения: $H_0 = \log m = \log_2 2 = 1$ бит. Если же, например, P(0)= 3/4, а P(1) = 1/4, то

$$H = -\sum_{i=1}^{m} p_i \log p_i = -P(0)\log P(0) - P(1)\log P(1) = -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right) = 0{,}815.$$

В случае же взаимной зависимости элементов, определяемой, например, условными вероятностями $P_0(0) = 2/3$; $P_0(1) = 1/3$; $P_1(0) = 1$; $P_1(1) = 0$, условная энтропия равна

$$H' = -P(0)\left[P_0(0)\log P_0(0) + P_0(1)\log P_0(1)\right] - P(1)\left[P_1(0)\log P_1(0) + P_1(1)\log P_1(1)\right] =$$
$$= -\frac{3}{4}\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) - \frac{1}{4}\left(0 \cdot \log 0 + 0 \cdot \log 00\right) = 0{,}685.$$

Пример 13. В результате статистических испытаний установлено, что при передаче каждых 100 сообщений длиной по 5 символов в сообщении символ K встречается 50 раз, а символ T – 30 раз. Вместе с символом K символ T встречается 10 раз. Определить условные энтропии $H(K/T)$ и $H(T/K)$.

Решение. Общее количество переданных символов $n = 100 \cdot 5 = 500$. Вероятности появления символов K и T соответственно равны:

$$p(K) = \frac{50}{500} = 0{,}1, \quad p(T) = \frac{30}{500} = 0{,}06.$$

Вероятность совместного появления символов K и T

$$p(KT) = \frac{10}{500} = 0{,}02.$$

Поскольку $p(KT) = p(T)p(K/T) = p(K)p(T/K)$, то условные вероятности можно вычислить по формулам:

$$p(K/T) = \frac{p(KT)}{p(T)} = \frac{0{,}02}{0{,}06} = 0{,}33, \quad p(T/K) = \frac{p(KT)}{p(K)} = \frac{0{,}02}{0{,}1} = 0{,}2.$$

Условные энтропии определим по формуле (24)

$$H(K/T) = -\{p(K/T)\log_2 p(K/T) + (1 - p(K/T))\log_2(1 - p(K/T))\} =$$
$$= -(0{,}33\log_2 0{,}33 + 0{,}67\log_2 0{,}67) = 0{,}9149 \ \textit{бит/символ}.$$

Пример 14. Определить общую условную энтропию сообщений, составленных из алфавита A, B, если вероятности появления символов в сообщении равны $p_A = 0{,}6$, $p_B = 0{,}4$. Условные вероятности переходов одного символа в другой, связанные с искажениями при передаче по каналам связи, равны $p(B/A) = 0{,}15$, $p(A/B) = 0{,}1$.

Решение. Используем формулу (25):

Понятие условной энтропии используется при определении взаимозависимости между символами кодируемого алфавита, для определения потерь при передаче информации по каналам связи, при вычислении энтропии объединения.

При вычислении информационных потерь в каналах связи с шумами удобно использовать канальную матрицу, которая при передаче m сигналов от источника A к приемнику B имеет следующий вид [3, 8, 9]:

$A \backslash B$	b_1	b_2	…	b_j	…	b_m
a_1	$p(b_1/a_1)$	$p(b_2/a_1)$	…	$p(b_j/a_1)$	…	$p(b_m/a_1)$
a_2	$p(b_1/a_2)$	$p(b_2/a_2)$	…	$p(b_j/a_2)$	…	$p(b_m/a_2)$
…	…	…	…	…	…	…
a_i	$p(b_1/a_i)$	$p(b_2/a_i)$	…	$p(b_j/a_i)$	…	$p(b_1/a_i)$
…	…	…	…	…	…	…
a_m	$p(b_1/a_m)$	$p(b_2/a_m)$	…	$p(b_j/a_m)$	…	$p(b_m/a_m)$

Как видно из структуры матрицы, вероятности, которые расположены по главной диагонали, определяют правильный прием сообщений без искажений, остальные значения – ложный, обусловленный помехами в канале связи. Значения цифр, заполняющих элементы канальной матрицы, обычно уменьшаются по мере удаления от главной диагонали и при полном отсутствии помех все элементы, кроме главной диагонали, равны нулю.

Прохождение каждого сигнала через канал связи описывается со стороны источника сообщений распределением условных вероятностей $p(b_j/a_i)$, суммы условных вероятностей по каждой строке равны 1. Потери информации, которые приходятся на долю сигнала a_i, описываются при помощи частной условной энтропии вида

$$H\left(b_j / a_i\right) = -\sum_{i=1}^{m} p\left(b_j / a_i\right)\log p\left(b_j / a_i\right).$$

Тогда общая условная энтропия описывает передачу всех m сигналов через канал связи. В случае равновероятных появлений сигналов на выходе источника сообщений [3, 8, 9]:

$$H(B/A) = -\frac{1}{m}\sum_{j=1}^{m} p(b_j/a_i)\log p(b_j/a_i). \quad (29)$$

Здесь необходимо разделить на m, так как энтропия есть неопределенность на один символ.

В случае неравновероятного появления символов источника сообщений следует учесть вероятность появления каждого символа, умножив на нее соответствующую частную условную энтропию

$$H(B/A) = -\sum_{i=1}^{m} p(a_i)\sum_{j=1}^{m} p(b_j/a_i)\log p(b_j/a_i). \quad (30)$$

При исследовании канала связи со стороны приемника сообщений канальная матрица будет иметь следующий вид:

$A \backslash B$	b_1	b_2	…	b_j	…	b_m
a_1	$p(a_1/b_1)$	$p(a_1/b_2)$	…	$p(a_1/b_j)$	…	$p(a_1/b_m)$
a_2	$p(a_2/b_1)$	$p(a_2/b_2)$	…	$p(a_2/b_j)$	…	$p(a_2/b_m)$
…	…	…	…	…	…	…
a_i	$p(a_i/b_1)$	$p(a_i/b_2)$	…	$p(a_i/b_j)$	…	$p(a_i/b_1)$
…	…	…	…	…	…	…
a_m	$p(a_m/b_1)$	$p(a_m/b_2)$	…	$p(a_m/b_j)$	…	$p(a_m/b_m)$

Соответственно потери информации в случае равновероятного и неравновероятного появления символов источника сообщений могут быть определены по аналогичным формулам частной и общей энтропии соответственно [13]:

$$H(B/A) = -\frac{1}{m}\sum_{i=1}^{m} p(a_i/b_j)\log p(a_i/b_j), \quad (31)$$

$$H(B/A) = -\sum_{j=1}^{m} p(b_j)\sum_{i=1}^{m} p(a_i/b_j)\log p(a_i/b_j). \quad (32)$$

На основе канальной матрицы $p(b_j/a_i)$, если заданы безусловные вероятности приемника $p(a_i)$, можно определить безусловные вероятности приемника $p(b_j) = \sum_{i=1}^{m} p(a_i)\log p(b_j/a_i)$. Тогда можно вычислить энтропию приемника сообщений $H(B) = -\sum_{j=1}^{m} p(b_j)\log p(b_j)$. Аналогично на основе канальной матрицы $p(a_i/b_j)$, если заданы безусловные вероятности приемника $p(b_j)$, можно определить безусловные вероятности приемника $p(a_i)=\sum_{j=1}^{m}p(b_j)\log p(a_i/b_j)$. Тогда можно вычислить энтропию источника сообщений $H(A)=-\sum_{i=1}^{m}p(a_i)\log p(a_i)$.

Пример 15. Телекоммуникационная система осуществляет передачу сообщений по незащищенному каналу связи, подверженному влиянию помех. Влияние помех задано в виде канальной матрицы следующего вида:

$$p(b/a)=\begin{vmatrix}0{,}98 & 0{,}01 & 0{,}01\\ 0{,}15 & 0{,}75 & 0{,}1\\ 0{,}3 & 0{,}2 & 0{,}5\end{vmatrix}.$$

Оценить количество энтропии сообщений, связанное с данным каналом связи:

а) определить полную условную энтропию при равновероятном появлении символов в сообщении;

б) в случае дифференцированных вероятностей с распределением: $p(a_1)=0{,}7$, $p(a_2)=0{,}2$, $p(a_3)=0{,}1$.

Решение. При ответе на первый вопрос в случае равновероятного появления символов в сообщении энтропия источника может быть определена по формуле Хартли

$$H(A)=\log_2 3=1{,}58\ \textit{бит/символ}.$$

Для определения полной условной энтропии следует использовать формулу (25), в которой благодаря равновероятности появления символов можно сумму по i заменить множителем $1/m$:

$$H(B/A)=-\sum_{i=1}^{3}\sum_{j=1}^{3}p(a_i)p(b_j/a_i)\log p(b_j/a_i)=-\frac{1}{3}\sum_{j=1}^{3}p(b_j/a_i)\log p(b_j/a_i)=$$
$$=-\frac{1}{3}(0{,}98\log_2 0{,}98+2\cdot 0{,}01\log_2 0{,}01+0{,}15\log_2 0{,}15+0{,}75\log_2 0{,}75+0{,}1\log_2 0{,}1+$$
$$+0{,}3\log_2 0{,}3+0{,}2\log_2 0{,}2+0{,}5\log_2 0{,}5)=0{,}9\ \textit{бит/символ.}$$

При ответе на второй вопрос в случае неравновероятного появления символов в сообщении энтропия источника должна определяться по формуле Шеннона

$$H(A) = -\sum_{j=1}^{3} p(a_i)\log p(a_i) = -(0{,}7\log_2 0{,}7 + 0{,}2\log_2 0{,}2 + 0{,}1\log_2 0{,}1) =$$
$$= 1{,}156 \textit{ бит/символ.}$$

Полная условная энтропия равна

$$H(B/A) = -\sum_{i=1}^{3}\sum_{j=1}^{3} p(a_i)p(b_j/a_i)\log p(b_j/a_i) =$$
$$= -[0{,}7(0{,}98\log_2 0{,}98 + 2\cdot 0{,}01\log_2 0{,}01) +$$
$$+ 0{,}2\cdot(0{,}15\log_2 0{,}15 + 0{,}75\log_2 0{,}75 + 0{,}1\log_2 0{,}1) +$$
$$+ 0{,}1\cdot(0{,}3\log_2 0{,}3 + 0{,}2\log_2 0{,}2 + 0{,}5\log_2 0{,}5)] = 0{,}463 \textit{ бит/символ.}$$

Пример 16. Определить энтропию приемника сообщений, если канальная матрица имеет вид

$$p(b/a) = \begin{vmatrix} 0{,}97 & 0{,}03 & 0 \\ 0{,}01 & 0{,}98 & 0{,}01 \\ 0 & 0{,}04 & 0{,}96 \end{vmatrix},$$

а вероятности появления символов на выходе источника сообщений равны: $p(a_1) = 0{,}5$, $p(a_2) = 0{,}2$, $p(a_3) = 0{,}2$.

Решение. Учитывая, что $p(a_i, b_j) = p(a_i)\log p(b_j/a_i)$, можно сказать, что вероятность приема символа b_1 равна вероятности передачи символа a_1, умноженной на условную вероятность правильного приема символа b_1 при передаче a_1, плюс сумма произведений условных вероятностей перехода любого из передаваемых символов в b_1 на вероятность появления этих символов в сообщении, т.е.

$$p(b_1) = \sum_{i=1}^{3} p(a_i)p(b_1/a_i) = p(a_1)p(b_1/a_1) + p(a_2)p(b_1/a_2) + p(a_3)p(b_1/a_3) =$$
$$= 0{,}5\cdot 0{,}97 + 0{,}3\cdot 0{,}01 + 0{,}2\cdot 0 = 0{,}488,$$
$$p(b_2) = 0{,}5\cdot 0{,}03 + 0{,}3\cdot 0{,}98 + 0{,}2\cdot 0{,}04 = 0{,}314,$$
$$p(b_3) = 0{,}5\cdot 0 + 0{,}3\cdot 0{,}01 + 0{,}2\cdot 0{,}96 = 0{,}195.$$

$$H(B) = -\sum_{i=1}^{3} p(b_j)\log_2 p(b_j) = -(0{,}488\log_2 0{,}488 + 0{,}317\log_2 0{,}317 +$$
$$+ 0{,}195\log_2 0{,}195 = 0{,}505103 + 0{,}525410 + 0{,}459893 = 1{,}490412 \textit{ бит/символ.}$$

1.6. Энтропия объединения. Свойства энтропии объединения

Энтропия объединения используется для вычисления энтропии совместного объединения статистически зависимых сообщений. Общую энтропию зависимых ансамблей *X* и *Y* определяют по формуле Шеннона [14]:

$$H(X,Y)=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k,y_j)\log p(x_k,y_j)=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k)p(x_k/y_j)\log p(x_k)p(x_k/y_j)=$$
$$=-\sum_{k=1}^{m}p(x_k)\log p(x_k)\sum_{j=1}^{n}p(x_k/y_j)-\sum_{k=1}^{m}p(x_k)\sum_{j=1}^{n}p(x_k/y_j)\log p(x_k/y_j).$$

С учетом условия нормировки, справедливого также и для совокупностей условных вероятностей $\sum_{j=1}^{n}p(x_k/y_j)=1$, получают

$$H(X,Y)=H(X)+H(Y/X),$$

где $H(X)$ – энтропия ансамбля X; $H(Y/X)$ – условная энтропия ансамбля Y при условии, что сообщение ансамбля X известны:

$$H(Y/X)=-\sum_{k=1}^{m}p(x_k)\sum_{j=1}^{n}p(x_k/y_j)\log p(x_k/y_j). \tag{33}$$

Формула (33) удобна при использовании также и для функционально зависимых элементов ансамбля X, для этого достаточно принять $Y=X$, откуда получим

$$H'=-\sum_{k=1}^{m}p(x_k)\sum_{j=1}^{n}p(x_k/y_j)\log p(x_k/y_j).$$

Энтропия объединения двух ансамблей сообщений определяется суммой слагаемых вида по всем возможным состояниям объединения:

$$H(X,Y)=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k,y_j)\log p(x_k,y_j), \tag{34}$$

где $p(x_k, y_j)$ – вероятность совместного появления двух статистически зависимых состояний x_k и y_j.

Поскольку вероятность совместного появления двух состояний может быть выражена через вероятность появления одного из состояний и условную вероятность появления другого состояния равенствами (26) и (27), преобразуем выражение для энтропии объединения к следующему виду [3, 8]:

$$\begin{aligned}H(X,Y)&=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k,y_j)\log p(x_k,y_j)=\\&=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k)p(y_j/x_k)\log\{p(x_k)p(y_j/x_k)\}=\\&=-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k)p(y_j/x_k)\log p(x_k)-\sum_{k=1}^{m}\sum_{j=1}^{n}p(x_k)p(y_j/x_k)\log p(y_j/x_k).\end{aligned} \tag{35}$$

Первая сумма в правой части равенства (35) представляет собой энтропию сообщений X (поскольку $\sum_{j=1}^{n} p\left(y_j / x_k\right) = 1$), а вторая сумма – условную энтропию $H(Y/X)$. Тогда окончательно имеем

$$H(X, Y) = H(X) + H(Y/X). \tag{36}$$

Аналогичное преобразование выражения (34) для энтропии объединения с использованием формулы для условной вероятности дает результат

$$H(X, Y) = H(Y) + H(X/Y), \tag{37}$$

который совместно с (36) более полно раскрывает смысл условной энтропии, рассмотренной в предыдущем разделе.

Основные свойства энтропии объединения

1. $H(X,Y) = H(Y,X) = H(X) + H(Y/X) = H(Y) + H(X/Y)$ (38)

2. Если сообщения X и Y статистически независимы, то энтропия объединения равна сумме энтропий сообщений X и Y. Это утверждение (правило сложения энтропий) следует из свойств условной энтропии. В рассматриваемом случае $H(Y/X) = H(Y)$, $H(X/Y) = H(X)$ и энтропия объединения, выражаемая (36) либо (37), равна

$$H(X,Y) = H(X) + H(Y). \tag{39}$$

3. Если сообщения X и Y являются полностью статистически зависимыми, то, как следует из свойств условной энтропии, $H(Y/X) = 0$ и энтропия объединения равна

$$H(X,Y) = H(X) = H(Y), \tag{40}$$

т.е. при полной статистической зависимости энтропии сообщений X и Y равны между собой.

1.7. Взаимная энтропия

Пусть ансамбли X и Y относятся соответственно к передаваемому и принимаемому сообщениям. Различия между X и Y обуславливаются искажениями в процессе передачи сообщений под воздействием помех.

При отсутствии помех различий между ансамблями X и Y не будет, а энтропии передаваемого и принимаемого сообщений будут равны: $H(X) = H(Y)$.

Воздействие помех оценивают условной энтропией $H_Y(X)$. Поэтому получаемое потребителем количество информации на один элемент сообщения равно [11, 13]

$$E(X, Y) = H(X) - H(X/Y). \tag{41}$$

Величину $E(X, Y)$ называют *взаимной энтропией*.

Если ансамбли X и Y независимы, то это означает, что помехи в канале привели к полному искажению сообщения, т.е. $H_Y(X) = H(X)$, а получаемое потребителем количество информации на один элемент сообщения: $E(X, Y) = 0$.

Если X и Y полностью зависимы, т.е. помехи в канале отсутствуют, то $H_Y(X) = 0$ и $E(X, Y) = H(Y)$.

Так как $H_Y(X) = H(X, Y) - H(Y)$, то $E(X, Y) = H(X) + H(Y) - H(X, Y)$, или $E(X,Y) = -\sum_{k=1}^{m}\sum_{j=1}^{n} p(x_k, y_j)\log\frac{p(x_k / y_j)}{p(x_k)}$.

Поясним понятия условной и полной энтропии на следующем простом, но наглядном примере, используя графическую интерпретацию. Пусть имеются два зависимых источника сообщений X и Y, энтропии которых были определены по отдельности, и они соответственно равны $H(X) = 5$ бит; $H(Y) = 10$ бит.

При отсутствии взаимосвязи между источниками информации количество информации можно отобразить кругами, как это показано на рис. 10.

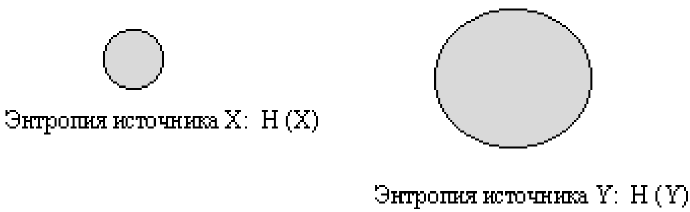


Рис. 10. Графическая интерпретация понятия энтропии независимых источников сообщений

Как это было указано в первом свойстве условной энтропии, она будет равна безусловной энтропии $H(Y/X) = H(Y) = 10$ бит, а $H(X/Y) =$

$= H(X) = 5$ бит. Каждый из источников будет нести новую информацию независимо от того, что сообщил другой. Общее количество информации и, следовательно, $H(X,Y) = H(X) + H(Y) = 5 + 10 = 15$ бит. Таким образом, когда источники независимы, условные энтропии принимают максимальные значения.

По мере увеличения взаимосвязи источников $H(Y/X)$ и $H(X/Y)$ будут уменьшаться (рис. 11).

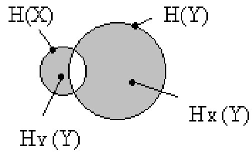


Рис. 11. Графическая интерпретация понятия энтропии зависимых источников сообщений

При полной зависимости двух источников один из них не вносит никакой информации относительного другого, так как при появлении, например, некоторого сообщения x_i неизбежно вызывает появление сообщения y_j, т.е. $p(x_i, y_j) = 1$ при $i = j$ и нулю при $i \neq j$. Поэтому $H(X/Y) = 0$ и, следовательно, $H(X, Y) = H(Y/X)$.

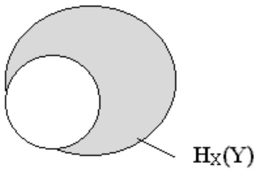


Рис. 12. Графическая интерпретация понятия энтропии при полной зависимости источников сообщений

При этом $H(Y/X) = H(Y) - H(X) = 10 - 5 = 5$ бит. Поэтому условная энтропия $H(Y/X)$ будет изменяться от 10 до 5 бит при максимально возможном изменении $H(X/Y)$ от 5 до 0 бит.

Пример 17. Имеется два зависимых источника информации X и Y, каждый из которых генерирует по три различных сообщения. Зависимость источников выражается в том, что появление сообщений от одного из них влияет на вероятность появления сообщения на другом. Условные вероятности появления представлены в следующем виде:

$X \backslash Y$	y_1	y_2	y_3
x_1	0,4	0,1	0
x_2	0	0,2	0,1
x_3	0	0	0,2

Вычислить энтропии $H(X)$, $H(Y)$, условные энтропии $H(Y/X)$, $H(X/Y)$, а также полную энтропию $H(X, Y)$ и взаимную энтропию $E(X, Y)$.

Решение. Для определения энтропий $H(X)$ и $H(Y)$ по формуле Шеннона необходимы безусловные вероятности появления сообщений. Чтобы вычислить безусловные вероятности $p(x_i)$ и $p(y_j)$ системы, необходимо сложить вероятности по строкам и столбцам исходной таблицы соответственно:

$p(x_1) = 0,4 + 0,1 + 0 = 0,5;$

$p(x_2) = 0 + 0,2 + 0,1 = 0,3;$

$p(x_3) = 0 + 0 + 0,2 = 0,2;$

$p(y_1) = 0,4 + 0 + 0 = 0,4;$

$p(y_2) = 0,1 + 0,2 + 0 = 0,3;$

$p(y_3) = 0 + 0,1 + 0,2 = 0,3.$

Теперь можно определить энтропии источников информации X и Y:

$$H(X) = -\sum_{i=1}^{3} p(x_i)\log p(x_i) = -(0,5\log 0,5 + 0,3\log 0,3 + 0,2\log 0,2) = 1,485 \text{ бит.}$$

$$H(Y) = -\sum_{j=1}^{3} p(y_j)\log p(y_j) = -(0,4\log 0,4 + 0,3\log 0,3 + 0,3\log 0,3) = 1,57 \text{ бит.}$$

Для определения условной энтропии источника информации Y при условии, что сообщения источника X известны, воспользуемся формулой (33). Здесь нам понадобятся условные вероятности $p(y_j/x_i)$, которые вычислим, используя соотношение $p(x_i, y_j) = p(x_i)p(y_j / x_i)$, откуда $p(y_j / x_i) = \dfrac{p(x_i, y_j)}{p(x_i)}$.

Последовательно найдем условные вероятности возможных значений Y при условии, что составляющая X приняла значение x_1:

$$p(y_1 / x_1) = \frac{p(x_1, y_1)}{p(x_1)} = \frac{0,4}{0,5} = 0,8; \quad p(y_2 / x_1) = \frac{p(x_1, y_2)}{p(x_1)} = \frac{0,1}{0,5} = 0,2;$$

$$p(y_3 / x_1) = \frac{p(x_1, y_3)}{p(x_1)} = \frac{0}{0,5} = 0.$$

Далее для x_2:

$$p(y_1 / x_2) = \frac{p(x_2, y_1)}{p(x_2)} = \frac{0}{0,3} = 0; \quad p(y_2 / x_2) = \frac{p(x_2, y_2)}{p(x_2)} = \frac{0,2}{0,3} = 0,67;$$

$$p(y_3 / x_2) = \frac{p(x_2, y_3)}{p(x_2)} = \frac{0,1}{0,5} = 0,33.$$

Для x_3:

$$p(y_1 / x_3) = \frac{p(x_3, y_1)}{p(x_3)} = \frac{0}{0,2} = 0; \quad p(y_2 / x_3) = \frac{p(x_3, y_2)}{p(x_3)} = \frac{0}{0,2} = 0;$$

$$p(y_3 / x_3) = \frac{p(x_3, y_3)}{p(x_3)} = \frac{0,2}{0,2} = 1.$$

Окончательно условная энтропия может быть определена из формулы

$$H(Y / X) = -[0,5 \cdot (0,8 \log_2 0,8 + 0,2 \log_2 0,2) +$$
$$+ 0,3 \cdot (0,67 \log_2 0,67 + 0,33 \log_2 0,33) + 0,2 \cdot (1 \log_2 1)\} = 0,635 \textit{ бит.}$$

Аналогичные расчеты произведем для определения условной энтропии источника информации X при условии, что сообщения источника Y известны:

$$p(x_i / y_j) = \frac{p(x_i, y_j)}{p(y_j)}.$$

Для y_1:

$$p(x_1 / y_1) = \frac{p(x_1, y_1)}{p(y_1)} = \frac{0,4}{0,4} = 1; \quad p(x_2 / y_1) = \frac{p(x_2, y_1)}{p(y_1)} = \frac{0}{0,4} = 0;$$

$$p(x_3 / y_1) = \frac{p(x_3, y_1)}{p(y_1)} = \frac{0}{0,4} = 0.$$

Для y_2:

$$p(x_1 / y_2) = \frac{p(x_1, y_2)}{p(y_2)} = \frac{0,1}{0,3} = 0,33; \quad p(x_2 / y_2) = \frac{p(x_2, y_2)}{p(y_2)} = \frac{0,2}{0,3} = 0,67;$$

$$p(x_3 / y_2) = \frac{p(x_3, y_2)}{p(y_2)} = \frac{0}{0,3} = 0.$$

Для y_3:

$$p(x_1 / y_3) = \frac{p(x_1, y_3)}{p(y_3)} = \frac{0}{0,3} = 0; \quad p(x_2 / y_3) = \frac{p(x_2, y_3)}{p(y_3)} = \frac{0,1}{0,3} = 0,33;$$

$$p(x_3/y_3)=\frac{p(x_3,y_3)}{p(y_3)}=\frac{0{,}2}{0{,}3}=0{,}67\cdot$$

$$H(X/Y)=-[0{,}4\cdot(1\log_2 1)+0{,}3\cdot(0{,}33\log_2 0{,}33+0{,}67\log_2 0{,}67)+$$
$$+0{,}3\cdot(0{,}33\log_2 0{,}33++0{,}67\log_2 0{,}67)]\approx 0{,}55\ \textit{бит.}$$

Общую энтропию зависимых источников информации X и Y определим по формуле (34):

$$H(X,Y)=-\sum_{i=1}^{3}\sum_{j=1}^{3}p(x_i,y_j)\log_2 p(x_i,y_j)=$$
$$=-(0{,}4\log_2 0{,}4+0{,}1\log_2 0{,}1+0{,}2\log_2 0{,}2+0{,}1\log_2 0{,}1+0{,}2\log_2 0{,}2)=$$
$$=0{,}529+0{,}332+0{,}464+0{,}332+0{,}464=2012\ \textit{бит.}$$

Кроме того, можно убедиться, что справедливо свойство, записанное формулой (38):

$H(X, Y) = H(X) + H(Y/X) = 1{,}485 + 0{,}635 = 2{,}12$ бит.

$H(X, Y) = H(Y) + H(X/Y) = 1{,}57 + 0{,}55 = 2{,}12$ бит.

Взаимную энтропию вычислим по формуле (41):

$E(X, Y) = H(X) – H(X/Y) = 1{,}485 – 0{,}55 = 0{,}935$ бит.

Контрольные вопросы

1. Дайте определение информации. В чем заключается сложность определения информации?
2. Каково соотношение информации и сообщения? Приведите примеры.
3. Для чего предназначены знаки и сигналы при организации передачи информации?
4. Назовите основные свойства информации.
5. Что такое информационный процесс? Какие виды информационных процессов вам известны?
6. Каким образом возникает, хранится, обрабатывается и передается информация?
7. Поясните функции и назначение блоков структурной схемы подсистемы передачи информации.
8. Что такое количество информации? Какой принцип положен в основу измерения количества информации?

9. Как определяется квантовая мера информации? Каковы ее недостатки?
10. Запишите формулу Хартли для определения количества информации при равновероятных состояниях элементов сообщений. Какие известны единицы измерения количества информации?
11. Как определяется понятие энтропии? Как она связана с информацией?
12. Как определить количество информации в одном сообщении, если известно максимально возможное количество сообщений?
13. Запишите формулу Шеннона для определения количества информации при разновероятных состояниях элементов сообщений.
14. Как оценивается частная энтропия и что она характеризует?
15. Каковы свойства энтропии?
16. Как связаны формулы Хартли и Шеннона?
17. Чему равна энтропия детерминированных сообщений?
18. При каких условиях энтропия принимает максимальное значение?
19. Чему равна энтропия нескольких независимых источников?
20. В чем особенности определения количества информации непрерывных сообщений?
21. Что такое приведенная и дифференциальная энтропия? Как они определяются?
22. При каком распределении обеспечивается максимальная энтропия H(x), если дисперсия состояний элементов сообщения – величина конечная и постоянная?
23. При каком распределении обеспечивается максимальная энтропия H(x), если дисперсия состояний элементов сообщения не ограничена?
24. От чего зависит величина Δx в формуле (16) для определения количества информации непрерывных сообщений?
25. Дайте определение условной энтропии. Приведите примеры, когда возникает необходимость оценивать условную энтропию.
26. Каким образом задается система двух случайных дискретных величин?
27. Как определяется закон распределения системы случайных непрерывных величин?
28. Запишите формулы для частной условной энтропии и общей условной энтропии. В чем заключается их отличие?

29. Перечислите основные свойства условной энтропии. Для каких практических задач в радиотехнике необходимо вычисление условной энтропии?

30. Каким образом задается канальная матрица? В чем заключается отличие канальных матриц, если задача рассматривается со стороны источника и со стороны приемника сообщений?

31. Что такое энтропия объединения и как она вычисляется?

32. Какую величину называют взаимной энтропией, и что она характеризует при передаче сообщений?

Задачи для самостоятельного решения

1. Определите, какие сообщения несут информацию (в широком смысле), а какие – нет. В каждом случае ответ поясните:

1.1) сибирские леса составляют 25 % всех лесов на планете;

1.2) $f(x) = 4x^5 + 2x^3 + 1$;

1.3) Россия – единственная страна, которая омывается одновременно двенадцатью морями;

1.4) светодиодный индикатор светится;

1.5) остров Пасхи имеет форму треугольника со сторонами 16, 24 и 18 километров;

1.6) из колоды карт вынуты 4 карты, два из которых – тузы;

1.7) теорема Байеса;

1.8) жители Папуа – Новой Гвинеи разговаривают на 820 языках, хотя официальным языком является английский, на котором общаются лишь 2 % населения;

1.9) в результате эксперимента с подбрасыванием монеты она была утрачена;

1.10) теория информации – важная научная дисциплина.

2. Определите, в каком тексте количественная мера синтаксической информации больше. Почему в том или ином из приведенных текстов информации больше или меньше?

2.1) АБ, АБ, АБ, АБ, АБ, АБ, АБ, АБ;

2.2) интерактивная интерпретация инновационного инсайта на сайте;

2.3) электромагнитное колебание;

2.4) радиотехническими системами (РТС) называют системы передачи или извлечения информации с помощью радиосигналов.

3. Какое количество информации приходится на букву алфавита, состоящего из 16, 27, 32 букв?

4. Известно, что одно из равновероятных возможных сообщений несет 3 бита информации. Из скольких качественных признаков состоит алфавит, если N = 64?

5. Чему равно количество информации в сообщении, переданном в двоичном коде пятизначной комбинацией и двумя пятизначными комбинациями, если символы кодируемого алфавита равновероятны? Как изменится количество информации, если сообщения будут передаваться в троичном коде? В десятичном коде?

6. Чему равна вероятность появления комбинации 10110 при передаче пятизначных двоичных кодов? Чему равно среднее количество информации, приходящейся на одну комбинацию?

7. Чему равно количество информации при получении сообщения о выходе из строя одного из 16 радиолокационных комплексов, одновременно поступивших на вооружение от одного и того же производителя?

8. При угадывании целого числа в некотором диапазоне было получено 6 бит информации. Сколько чисел содержит этот диапазон?

9. Сообщение о том, что отказала одна из радиолокационных станций несет 4 трита информации. Сколько всего радиолокационных станций в военной группировке?

10. Определите энтропию иерархической системы, заданной графом (рис. 13), если каждый элемент системы может с равной вероятностью находиться в четырех состояниях.

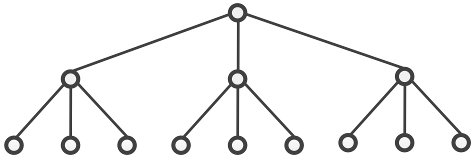


Рис. 13. Граф иерархической системы

11. Чему равно количество информации в сообщении о неисправности 3 радиолокационных комплексов из 12 одновременно введенных в эксплуатацию?

12. Определить энтропию физической системы, которая может находиться в одном из 7 состояний. Вероятности системы заданы рядом распределения

$$A = \begin{vmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \\ 0{,}05 & 0{,}3 & 0{,}15 & 0{,}07 & 0{,}11 & 0{,}08 & 0{,}24 \end{vmatrix}.$$

13. Имеется три независимых источника сообщений, из которых первый выдает 2 сообщения с равной вероятностью 0,5, второй – 5 сообщений с равной вероятностью 0,2, третий – 2 сообщения с вероятностями 0,25 и 0,75 соответственно. Определите, какой источник обладает большей энтропией.

14. Телекоммуникационная система ведет прием сообщений от двух источников, характеризующихся различными распределениями появления абстрактных символов алфавита сообщений. Эти распределения представлены в виде рядов:

Случайные величины x_i	0,5	0,7	0,9	0,3
Вероятности их появления	0,25	0,25	0,25	0,25

Случайные величины y_j	5	10	15	8
Вероятности их появления	0,25	0,25	0,25	0,25

Проанализируйте, с каким из источников связано больше неопределенности, и сообщения от какого из них будут более информативными.

15. Бортовая РЛС предназначена для обнаружения воздушных и морских объектов. Сообщение о том, что РЛС обнаружила морской объект, несет 4 бита информации. Вероятность появления в контролируемом секторе воздушного объекта в два раза меньше, чем вероятность появления морского. Сколько информации несет сообщение о том, что РЛС обнаружила воздушный объект?

16. Состояние летательного аппарата характеризуется двумя случайными величинами: высотой полета и модулем скорости. Высота лета-

тельного аппарата распределена с равномерной плотностью на участке (h_1, h_2), где h_1 = 9 км, h_2 = 12 км, скорость – по нормальному закону с математическим ожиданием m = 800 км/ч и средним квадратическим отклонением σ = 300 км/ч. Вычислите энтропию, связанную с определением высоты полета и энтропию, связанную с определением скорости летательного аппарата.

17. Непрерывная случайная величина распределена по экспоненциальному закону, функция плотности которого имеет вид

$$\omega(x) = \begin{cases} \lambda e^{-\lambda x} & при \;\; x \geq 0, \\ 0 & при \;\; x < 0. \end{cases}$$

Найдите дифференциальную энтропию этой случайной величины.

18. Радиотехническая система использует линию связи, по которой передаются амплитудо-модулированные сигналы, имеющие нормальную плотность распределения. Параметры нормального распределения: математическое ожидание m = 0, среднее квадратическое отклонение σ = 8. Точность измерения сигнала Δx = 0,2. Найдите энтропию непрерывного сигнала.

19. Определите энтропию непрерывного источника сообщений, если функция плотности распределения вероятностей состояний элементов непрерывных сообщений подчиняется закону Симпсона с параметрами a = 3, b = 8:

$$\omega(x) = \begin{cases} 0 \;\; при \;\; x \leq a, \\ \dfrac{4(x-a)}{(b-a)^2} \;\; при \;\; a < x \leq \dfrac{b+a}{2}, \\ -\dfrac{4(x-b)}{(b-a)^2} \;\; при \;\; \dfrac{b+a}{2} < x \leq b, \\ 0 \;\; при \;\; x > b, \end{cases}$$

а Δx = 0,1.

20. При передаче текстовых сообщений статистические наблюдения показали, что для слов со средней длиной в 6 букв на каждые 100 сообщений буква A встречается 80 раз, буква B – 50 раз, буквы A и B вместе встречаются 10 раз. Определить условную энтропию появления A, если в слове присутствует B, и условную энтропию B, если в слове присутствует A.

21. Сообщения передаются двоичным кодом. В первом случае вероятности появления 0 и 1 равны соответственно $p_0 = 0{,}8$ и $p_1 = 0{,}2$. Помехи в канале связи отсутствуют, т.е. условные вероятности переходов 0 в 1 и 1 в 0 равны 0. Во втором случае символы передаются с равными вероятностями $p_0 = p_1 = 0{,}5$, однако в результате действия помех условные вероятности переходов равны: $p(1/1) = 0{,}8$, $p(1/0) = 0{,}2$, $p(0/0) = 0{,}8$, $p(0/1) = 0{,}2$. Чему равна энтропия сообщений в первом и втором случаях?

22. Используя свойства безусловной и условной энтропии, охарактеризуйте особенности передачи сообщений по каналу связи, канальная матрица которого имеет следующий частный вид:

$$p(b/a) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

23. Определить общую условную энтропию сообщений, передаваемых по каналу связи, который описывается следующей канальной матрицей:

$$p(a/b) = \begin{vmatrix} 0{,}9 & 0{,}1 & 0 & 0 \\ 0{,}05 & 0{,}84 & 0{,}01 & 0 \\ 0{,}03 & 0{,}06 & 0{,}98 & 0{,}1 \\ 0{,}02 & 0 & 0{,}01 & 0{,}9 \end{vmatrix},$$

символы алфавита, из которого составлены сообщения, – равновероятны.

24. Определить энтропию источника сообщений, если вероятность появления сигналов на входе приемника $p(b_1) = 0{,}1$, $p(b_2) = 0{,}3$, $p(b_3) = 0{,}4$, $p(b_4) = 0{,}2$, а канальная матрица имеет вид

$$p(a/b) = \begin{vmatrix} 0{,}99 & 0{,}02 & 0 & 0 \\ 0{,}01 & 0{,}98 & 0{,}01 & 0{,}01 \\ 0 & 0 & 0{,}98 & 0{,}02 \\ 0 & 0 & 0{,}01 & 0{,}97 \end{vmatrix}.$$

25. Определите энтропию квантованного сигнала с распределением $f(x)$, заданным графиком (рис. 14), с точностью 1В.

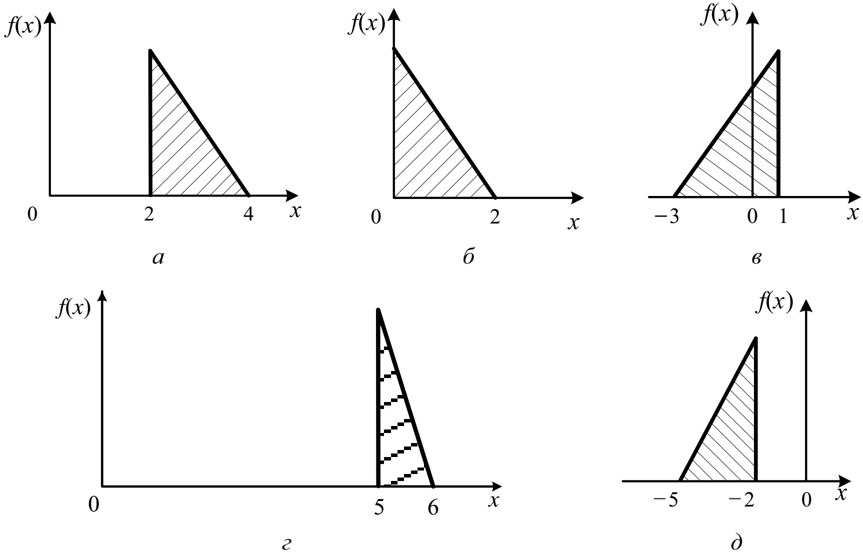


Рис. 14. Исходные данные для задачи 25

26. При передаче сообщений по каналу связи с шумами была получена следующая статистика: частота f_1 из 100 раз была принята 97 раз, 2 раза была принята частота f_2 и 1 раз – частота f_3; при передаче f_2 – 98 раз принята f_2, два раза – f_1; при передаче f_3 96 раз принята f_3, два раза – f_2 и два раза – f_4; при передаче f_4 99 раз принята f_4 и один раз – f_3:

а) составить канальную матрицу, описывающую данный канал связи с точки зрения условий прохождения частот $f_1 \div f_4$;

б) определить общую условную энтропию сообщений, алфавитом которых являются частоты $f_1 \div f_4$, если вероятности появления этих частот в передаваемых сообщениях соответственно равны:

$$p(f_1) = 0{,}37,\ p(f_2) = 0{,}3,\ p(f_3) = 0{,}23,\ p(f_4) = 0{,}1;$$

в) определить энтропию принятых сообщений.

27. Определите энтропии

$$H(X),\ H(Y),\ H(X,\ Y),\ H(Y/X),\ H(X/Y),\ E(X,\ Y),$$

если

$$p(x_1,\ y_1) = 0{,}3;\ p(x_1,\ y_2) = 0{,}2;\ p(x_2,\ y_3) = 0{,}1;\ p(x_3,\ y_2) = 0{,}1;\ p(x_3,\ y_3) = 0{,}25.$$

28. Радиотехническая система осуществляет обработку сообщений, поступающих из канала связи, подверженному воздействию помех, канальная матрица данного канала связи имеет вид:

$$p(y_j / x_i) = \begin{vmatrix} 0.98 & 0.01 & 0.01 \\ 0.15 & 0.7 & 0.15 \\ 0.1 & 0.1 & 0.8 \end{vmatrix}$$

Кроме того, известно, что $p(x_1) = 0{,}8$, $p(x_2) = 0{,}1$, $p(x_3) = 0{,}1$. Требуется оценить, какое количество информации несут сообщения от такого источника, каковы информационные потери из-за наличия помех в канале связи в общем случае. В частности, определить потери при передаче сообщения из 500 символов алфавита x_1, x_2, x_3? Изменится ли количество принятой информации, если помех в канале связи будут отсутствовать?

29. Канал передачи данных осуществляет передачу двоичных символов, т.е. абстрактный алфавит A = {0, 1} и описывается графовой моделью, показанной на рис. 15. Скорость появления элементов на выходе канала составляет 800 эл/с. Определите количество информации, передаваемое по такому каналу, и скорость передачи информации.

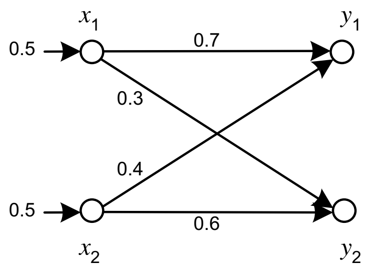


Рис. 15. Модель канала передачи данных для задачи 29

30. Канал передачи данных осуществляет передачу двоичных символов, т.е. абстрактный алфавит A = {0, 1} и описывается графовой моделью, показанной на рис. 16. Скорость появления элементов на выходе канала составляет 500 эл/с. Определите количество информации, передаваемое по такому каналу, и скорость передачи информации.

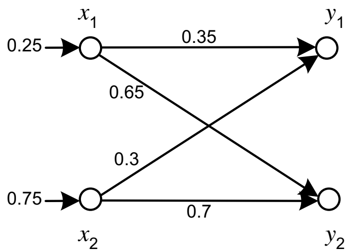


Рис. 16. Модель канала передачи данных для задачи 30

2. КОДИРОВАНИЕ ИНФОРМАЦИИ

После количественной оценки информации обычно следует важнейший этап преобразования информации. В самом широком смысле под кодированием понимают процесс формирования определенного представления информации. Исходная информация в радиотехнических системах, как правило, имеет непрерывный характер. Как показывает практика, при передаче, обработке и хранении информации значительные преимущества дает дискретная форма представления сигналов. Поэтому в тех случаях, когда первичные сигналы радиотехнических систем являются непрерывными, их обычно предварительно преобразовывают в дискретные сигналы. В связи с этим в данном разделе будем рассматривать процессы кодирования, относящиеся к дискретной информации. Кодирование здесь понимается в более узком смысле, а именно как отображение дискретных сообщений сигналами в виде определенных сочетаний символов. Происходит перевод информации, представленной сообщением в первичном алфавите, в некоторый вторичный алфавит в виде последовательности кодов. Совокупность правил, в соответствии с которыми производятся эти операции, называют кодом.

Кодирование преследует несколько целей. Первая из них заключается в том, чтобы представить сообщения в такой системе символов, которая обеспечивала бы простоту и надежность аппаратной реализации подсистем и блоков радиотехнических систем и их необходимую эффективность. Вторая цель кодирования состоит в том, чтобы обеспечить наилучшее согласование свойств источника сообщений со свойствами канала связи. Путем такого согласования добиваются выигрыша во времени передачи, т.е. повышения эффективности системы. Наконец, при наличии помех кодирование может обеспечить достаточно высокую достоверность передачи или обработки информации.

Однако при этом важно, чтобы кодирование, обеспечивающее изменение структуры первичных сигналов, не изменяло количество информации, заключенной в первоначальном сообщении.

Кодирование информации распадается на следующие этапы [2, 8, 9].

1. Определение объёма информации, подлежащей кодированию.

2. Классификация и систематизация информации.

3. Выбор системы кодирования и разработка кодовых обозначений.
4. Непосредственное кодирование.

Приведем необходимые для дальнейшего изучения определения основных понятий и терминов. Поскольку кодирование в цифровых системах рассматривается как процесс сопоставления элементов алфавита цифрам, тогда само правило сопоставления следует понимать как *код*.

Цифры, сопоставленные элементам алфавита, называют *кодовыми словами*. Следует отметить, что шифрование – разновидность кодирования и представляет его частный случай. *Шифр* – это код, значение и правила использования которого известно ограниченному кругу лиц.

Общее число символов, составляющих кодовую комбинацию, называется *значностью* кода, или длиной кода n. Если значность одинакова для всех элементов алфавита, то код называется *равномерным*, в противном случае – *неравномерным*.

Характерным примером равномерного кода является пятизначный двоичный (бинарный) телеграфный код Бодо. Для равномерного кода число всех возможных кодовых комбинаций выражается формулой [2, 8, 9]

$$N = m^n.$$

Неравномерные коды требуют либо специальных разделительных знаков, указывающих конец одной или начало другой кодовой комбинации, либо должны строиться так, чтобы никакая кодовая комбинация не являлась начальной частью («префиксом») другой кодовой комбинации. Коды, удовлетворяющие последнему условию, называются неприводимыми (префиксными) кодами.

Количество значений кодовых признаков, используемых в кодовых комбинациях, называется *основанием* кода m.

При передаче большого объема информации применяется комбинирование, т.е. передача сообщений не одним сигналом, а комбинацией из нескольких элементарных сигналов.

Совокупность элементарных сигналов, используемая для передачи одного сообщения, называется кодовой комбинацией. Под кодом понимается набор кодовых комбинаций, поставленных в соответствие набору сообщений. Число возможных значений комбинаций информационного параметра элементарного сигнала называется основанием кода. При математическом описании процесса комбинирования значения информационного параметра записываются с помощью символов 0, 1, ..., m–1, которые называются кодовыми символами. Множество возможных кодовых символов 0, 1, ..., m–1, используемых при кодировании, называется кодовым алфавитом.

2.1. Простые безызбыточные коды

По числу различных символов в кодовых комбинациях различают коды [2, 8, 9]:

– единичные, $m = 1$,
– двоичные, $m = 2$,
– многопозиционные, $m > 2$.

В единичном коде используются одинаковые символы. Кодовые комбинации отличаются друг от друга лишь количеством символов (импульсов). Такие коды также называются числоимпульсными, они применяются в телефонии. Код с данным признаком называется блочным, неравномерным числовым кодом. Такой код отличается своей простотой. Однако он, очевидно, является неравномерным, что создает ряд трудностей при технической реализации передачи большого количества сообщений, а также для обеспечения помехозащищенности. В связи с этим его, как правило, не используют для передачи по каналам связи, а прибегают к нему лишь в промежуточных преобразованиях сигналов на передающей и приемной сторонах.

Двоичные коды получили наибольшее распространение ввиду того, что большинство современных устройств реализовано на элементной базе, имеющей число устойчивых состояний, равное степени двойки (к такого рода устройствам принадлежат большинство электромагнитных, электронных, магнитных и других типов бесконтактных реле). Кроме того, для двоичных кодов хорошо проработан математический аппарат для выполнения арифметических и логических операций.

Многопозиционные коды не находят широкого применения из-за сложности технического обеспечения большого количества устойчивых состояний, сложностью выполнения арифметических и логических операций.

Если код представляет собой многоэлементный сигнал с постоянным числом n единиц, то этот код называют блочным, равномерным кодом. Для данных кодов $m \geq 2$.

Правило равномерного кодирования в алфавите m основано на представлении сообщения в m-ичной системе счисления. Слева к числам дописываются недостающие до числа m нули.

Максимально возможное число кодовых комбинаций (мощность кода) $N_0 = m^n$**,** где m называют длиной кода.

При выборе параметров кода определяют требуемое число N кодовых комбинаций по правилу $N \leq N_0 = m^n$. Тогда зная требуемое число N кодовых комбинаций и основание проектируемого кода m, длину кода определяют по формуле [2, 8, 9]

$$n =]log_m N[, \tag{42}$$

где].[– операция округления до ближайшего большего целого числа; N – число сообщений, которое необходимо кодировать. В частности, если $m = 2$, то длина кода определяется из соотношения

$$n =]log_2 N[. \tag{43}$$

Код для числа N, построенный с параметрами, выбранными по правилу (42), называется простым (первичным), безызбыточным кодом. В телемеханике данный код называется кодом на все сочетания, так как

$$N = 2^m = \sum_{i=0}^{m} C_m^i = C_m^0 + C_m^1 + .. + C_m^m \cdot$$

В технических информационных системах используются преимущественно арифметические (цифровые) коды, которые основаны на известных системах счисления. Системы счисления подразделяют на позиционные и непозиционные.

Общепризнанным в настоящее время является позиционный принцип образования системы счисления. Значение каждого символа (цифры) зависит от его положения – позиции в ряду символов, представляющих число.

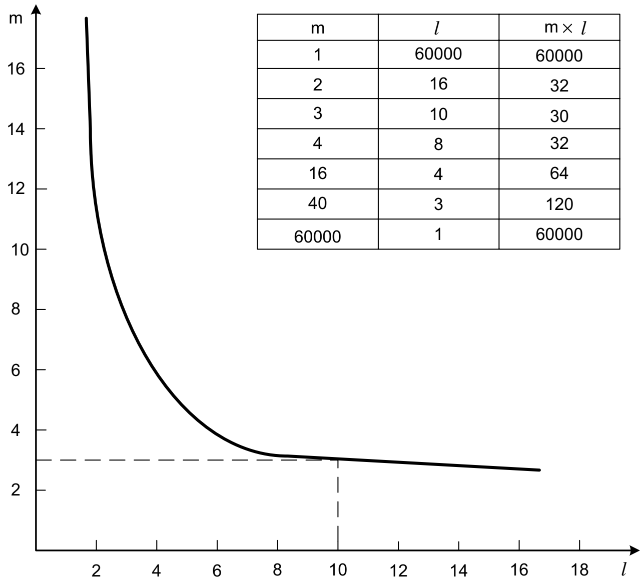
Единица каждого следующего разряда больше единицы предыдущего разряда в m раз, где m – основание системы счисления. Полное число получаем, суммируя значения по разрядам,

$$Q = \sum_{i=1}^{l} a_i m^{i-1} = a_l m^{l-1} + a_{l-1} m^{l-2} + \ldots + a_2 m^1 + a_1 m^0, \qquad (44)$$

где i – номер разряда данного числа; l – количество разрядов; a_i – множитель, принимающий любые целочисленные значения в пределах от 0 до m – 1 и показывающий, сколько единиц i-го разряда содержится в числе.

Чем больше основание системы счисления, тем меньшее число разрядов требуется для представления данного числа, а следовательно, и меньшее время для его передачи.

Однако с ростом основания существенно повышается требование к линии связи и аппаратуре создания и распознавания сигналов, соответствующих различным символам. Логические элементы вычислительных устройств в этом случае должны иметь большее число устойчивых состояний.



m	l	m × l
1	60000	60000
2	16	32
3	10	30
4	8	32
16	4	64
40	3	120
60000	1	60000

Рис. 17. Зависимость количества разрядов в числе от основания системы счисления

Учитывая оба обстоятельства, целесообразно выбрать систему, обеспечивающую минимум произведения количества различных символов m на количество разрядов l для выражения любого числа. Исследуем и найдем этот минимум по графику на рис. 17, где показана связь между величинами m и l при воспроизведении определенного достаточно большого числа Q ($Q \approx 60\,000$) [2, 8, 9].

Из графика следует, что наиболее эффективной системой является троичная. Незначительно уступают ей двоичная и четверичная. Системы с основанием 10 и более существенно менее эффективны. Сравнивая эти системы с точки зрения удобства физической реализации соответствующих им логических элементов и простоты выполнения в них арифметических и логических действий, предпочтение необходимо отдать двоичной системе. Действительно, логические элементы, соответствующие этой системе, должны иметь всего два устойчивых состояния. Задача различения сигналов сводится в этом случае к задаче обнаружения (есть импульс или нет импульса), что значительно проще.

Основные арифметические и логические действия в двоичной системе счисления определяются по правилам, приведенным ниже.

Арифметические операции:

Правила сложения	Правила вычитания	Правила умножения
$0 + 0 = 0$	$0 - 0 = 0$	$0 \cdot 0 = 0$
$0 + 1 = 1$	$1 - 0 = 1$	$0 \cdot 1 = 0$
$1 + 0 = 1$	$1 - 1 = 0$	$1 \cdot 0 = 0$
$1 + 1 = 10$	$10 - 1 = 1$	$1 \cdot 1 = 1$

Логические операции:

Правила сложения по модулю 2	Правила логического сложения	Правила логического умножения
$0 \oplus 0 = 0$	$0 \vee 0 = 0$	$0 \,\&\, 0 = 0$
$0 \oplus 1 = 1$	$0 \vee 1 = 1$	$0 \,\&\, 1 = 0$
$1 \oplus 0 = 1$	$1 \vee 0 = 1$	$1 \,\&\, 0 = 0$
$1 \oplus 1 = 0$	$1 \vee 1 = 1$	$1 \,\&\, 1 = 1$

Особенности технической реализации большинства цифровых устройств, составляющих базис современных средств приема, обработки и передачи информации, а также математические свойства двоичных кодов привели к тому, что они получили наибольшее распространение. Однако запись чисел в двоичной системе счисления оказывается наиболее длинной и не всегда удобной при первичной обработке, а также для восприятия человеком, если он составляет необходимое звено в более сложных радиотехнических и инфокоммуникационных системах. Отсюда вытекает практический интерес в рассмотрении кодов, которые обладают преимуществами двоичных кодов, но имеют более компактную запись, это так называемые двоично-десятичные коды, которые предлагается рассмотреть в следующем разделе.

2.2. Составные коды

Составные коды базируются на составных системах счисления, имеющих два и более оснований. К таким кодам относятся двоично-десятичные, которые, как было указано в предыдущем разделе, имеют особое значение, поскольку создавались, чтобы сохранить преимущества двоичной системы и удобство десятичной системы. В таком коде каждая цифра десятичного числа записывается в виде четырехразрядного двоичного числа (тетрады). С помощью четырех разрядов можно образовать 16 различных комбинаций, из которых любые 10 могут составить двоично-десятичный код. Для фиксации цифр десятичного числа наибольшее пратическое применение нашли четырехэлементные двоичные весовые коды 8-4-2-1, 7-4-2-1, 5-1-2-1 и 2-4-2-1, три из которых представлены в табл. 4. Эти коды называют весовыми (или взвешенными), поскольку цифры в названии кода означают вес единиц в соответствующих двоичных разрядах [2, 8, 9].

Таблица 4

Примеры весовых кодов

Число в десятичном коде	Двоично-десятичный код с весами 8-4-2-1		Двоично-десятичный код с весами 5-1-2-1		Двоично-десятичный код с весами 2-4-2-1	
0	0000	0000	0000	0000	0000	0000
1	0000	0001	0000	0001	0000	0001
2	0000	0010	0000	0010	0000	0010
3	0000	0011	0000	0011	0000	0011
4	0000	0100	0000	0111	0000	0100

Окончание табл. 4

Число в десятичном коде	Двоично-десятичный код с весами 8-4-2-1		Двоично-десятичный код с весами 5-1-2-1		Двоично-десятичный код с весами 2-4-2-1	
5	0000	0101	0000	1000	0000	1011
6	0000	0110	0000	1001	0000	1100
7	0000	0111	0000	1010	0000	1101
8	0000	1000	0000	1011	0000	1110
9	0000	1001	0000	1111	0000	1111
10	0001	0000	0001	0000	0001	0000

Отметим, что отображение числа во взвешенных кодах может быть различным. Например, в коде 5-1-2-1 цифру 7 можно представить как 1010 ($1 \cdot 5 + 0 \cdot 1 + + 1 \cdot 2 + 0 \cdot 1 = 7$) или 1101 ($1 \cdot 5 + 1 \cdot 1 + 0 \cdot 2 + 1 \cdot 1 = 7$). Предпочтение отдается той или иной форме, исходя из дополнительных соображений простоты или надежности технической реализации, согласования электромагнитных параметров и т.д.

Двоично-десятичные коды обычно используются как промежуточный при введении в вычислительную машину данных, представленных в десятичном коде, или при выводе из машины информации для регистрации в десятичном коде. В частности, коды с весами 5-1-2-1 и 2-4-2-1 широко используются при поразрядном уравновешивании в цифровых измерительных приборах.

В простом двоичном или взвешенном коде при переходе от изображения одного числа к изображению соседнего старшего или соседнего младшего числа может происходить одновременное изменение цифр в нескольких разрядах (например, в двоичном коде переход от 7 к 8 или в 5-1-2-1-коде от 4 к 5 потребует изменения во всех 4 разрядах – это преобразование от 0111 к 1000). Это может явиться источником значительных ошибок при некоторых способах кодирования непрерывных сообщений.

Для устранения этого явления используют специальные двоичные коды, у которых при переходе от изображения одного числа к изображению следующего соседнего числа изменяется значение цифры только одного разряда. К числу таких кодов относится код Грея, часто называемый циклическим или рефлексно-двоичным (отраженным).

Код Грея является непозиционным кодом, вес единицы не определяется номером разряда. В этом коде модно выделить оси симметрии (оси «отражения»), относительно которых наблюдается идентичность элементов в некоторых разрядах. Так, например, имеет место симметрия относительно оси, проведенной между числами 7 и 8. В комбинациях, симметричных относительно этой оси, идентичны три символа младших разрядов. Отмеченная особенность послужила основанием для введения термина «рефлексный код» (от англ. to reflect – отражать).

Код Грея используется в технике аналого-кодового преобразования, где он позволяет свести к единице младшего разряда ошибку неоднозначности при считывании. Комбинации кода Грея, соответствующие десятичным числам от 0 до 15, приведены в табл. 5 [2, 8, 9].

Таблица 5

Таблица соответствия кодовых комбинаций в десятичном коде и коде Грея

Число в десятичном коде	Код Грея	Число в десятичном коде	Код Грея	Число в десятичном коде	Код Грея
0	0000	6	0101	11	1110
1	0001	7	0100	12	1010
2	0011	8	1100	13	1011
3	0010	9	1101	14	1001
4	0110	10	1111	15	1000
5	0111				

Правила перевода числа из кода Грея в обычный двоичный сводится к следующему: первая единица со стороны старших разрядов остается без изменения, последующие цифры (0 и 1) остаются без изменения, если число единиц, им предшествовавших, четно, и инвертируются, если число единиц нечетно.

Выразим одно из чисел кода Грея, например 1010, в обычном двоичном коде. Первая единица слева переписывается. Следующая цифра будет единицей, так как в этом разряде кода Грея стоит 0 и впереди только одна единица. Далее следует записать нуль, так как в следующем разряде исходного числа стоит 1 и впереди снова имеется только одна единица. Поскольку перед последней цифрой числа в коде Грея стоят две единицы, то она должна остаться неизменной, то есть нулем. Таким образом, числу 1010 в коде Грея соответствует обычное двоичное число 1100.

Итак, в зависимости от системы счисления, положенной в основу преобразования цифр, различают кодирования: бинарное, триарное, тетрарное и т.д. в первом случае буквы исходного алфавита сопоставляются бинарным числом, составленным из единиц и нулей.

Необходимо отметить, что «строение» кода никак не связано с физическим осуществлением его различных элементов. Бинарный код характеризуется наличием двух разных элементов; физический характер их различия не играет никакой роли. Два элемента бинарного кода (нуль и единица) могут представлять собой два импульса, различающихся между собой амплитудой либо фазой, либо частотой, либо длительностью, либо, наконец, формой в самом общем смысле.

Физическое различие элементов кода – это вопрос вида модуляции, а не строения кода.

Строение кода удобно представлять в виде кодового дерева. Из каждого узла исходит число ветвей, равное основанию кода. Переход из узла по левой ветви соответствует присоединению справа к формируемой кодовой комбинации нуля, переход вправо – единицы.

Для примера рассмотрим кодовые деревья для трех различных двоичных кодов A, B и C (рис. 18). Здесь общее число кодируемых сообщений кодовых комбинаций равно 8; сообщения показаны буквами от *a* до *h* [12].

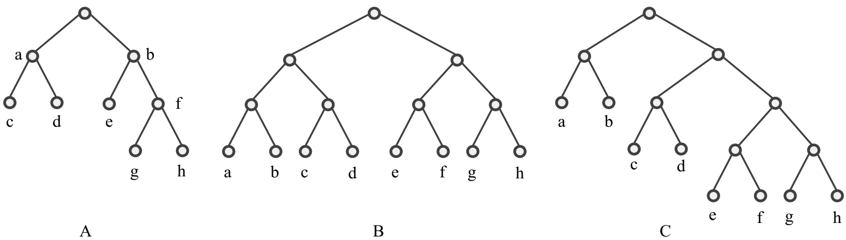


Рис. 18. Примеры кодовых деревьев

В табл. 6 приведены кодовые комбинации для всех трех кодов. Код A – неравномерный, он требует разделительных знаков между комбинациями. Код B – равномерный. Код C – неравномерный префиксный код, не требующий разделительных знаков. Любая непрерывная последовательность комбинаций кода C легко расшифровывается при помощи кодового дерева. Очевидно, что задание кода в виде кодового дерева и таблицы соответствия является эквивалентным.

Таблица 6

Коды, полученные по кодовым деревьям на рис. 18

Код \ Символ	a	b	c	d	e	f	g	h
A	0	1	00	01	10	11	110	111
B	000	001	010	011	100	101	110	111
C	00	01	100	101	1100	1101	1110	1111

Следует заметить, что равномерный код также является префиксным. Поэтому при работе с ним не требуется разделительных знаков. Необходимо лишь знать начало последовательности кодовых комбинаций. Предполагается, что в передаваемой непрерывной последовательности ни один знак не выпал и не был принят ошибочно. В противном случае это приводит к сквозной ошибке (треку ошибки) и неправильному декодированию последующих кодовых комбинаций. С этой точки зрения введение разделительных знаков, обрамляющих каждую кодовую комбинацию (стартстопный режим), может рассматриваться как средство повышения надежности. Это относится к любому префиксному коду.

2.3. Коды по законам комбинаторики

Способы комбинирования позволяют строить комбинаторные коды. Различают коды, использующие все возможные комбинации и с их частичным использованием.

Коды по закону размещений

Закон соединений предусматривает, что кодовая комбинация включает n символов из их общего числа m. Длина кодовой комбинации может быть $2 \le n \le m - 1$. Комбинации различаются либо составом символов, либо порядком их следования [2, 8, 9].

Мощность кода определится по формуле

$$N = A_m^n = \frac{m!}{(m-n)!}.$$

Пример 18. Пусть алфавит содержит кодовые символы {0,1,2,3}, мощность кода $m = 4$. Требуется построить код по закону размещений с кодовыми словами длины 2, т.е. $n = 2$.

Решение. Количество различных кодовых комбинаций равно

$$N = A_4^2 = \frac{m!}{(m-n)!} = \frac{4!}{(4-2)!} = 3 \cdot 4 = 12 \cdot$$

Множество кодовых комбинаций $\{01, 01, 03, 12, 13, 23, 10, 20, 30, 21, 31, 32\}$.

Коды по закону сочетаний

Закон соединений определяет построение кодовых комбинаций с включением в них n символов из m в алфавите. Кодовые комбинации различаются только составом символов [2, 8, 9].

Мощность кода определится по формуле

$$N = C_m^n = \frac{m!}{n!(m-n)!} \cdot$$

Пример 19. Пусть алфавит, как и ранее, содержит кодовые символы $\{0,1,2,3\}$, мощность кода $m = 4$. Требуется построить код по закону сочетаний с кодовыми словами длины 2, т.е. $n = 2$.

Решение. Количество различных кодовых комбинаций равно

$$N = C_4^2 = \frac{m!}{n!(m-n)!} = \frac{4!}{2!(4-2)!} = \frac{3 \cdot 4}{2} = 6 \cdot$$

Множество кодовых комбинаций $\{01, 02, 03, 12, 13, 23\}$ или $\{10, 20, 30, 21, 31, 32\}$. При этом нельзя включать в код, например, 01 и 10 одновременно, поскольку их состав одинаков, порядок следования символов в кодовом сообщении не имеет значения и, следовательно, такие кодовые комбинации будут неразличимы.

Коды по закону перестановок

Закон перестановок характеризуется тем, что все m символов алфавита однократно входят в каждую кодовую комбинацию, т.е. $n = m$. Кодовые комбинации различаются только порядком следования символов. Мощность кода определится по формуле [2, 8, 9]

$$N=P_m=m!$$

Пример 20. Пусть алфавит содержит кодовые символы $\{0,1,2\}$, мощность кода $m = 3$. Требуется построить код по закону перестановок.

Решение. Количество различных кодовых комбинаций равно

$$N = P_3 = 3! = 6.$$

Множество кодовых комбинаций {012, 021, 102, 120, 201, 210}.

Можно отметить, что с ростом m количество различных кодовых комбинаций растет достаточно быстро, при $m = 5$, N уже составляет 120, а при $m = 10$, $N = 3\,628\,800$.

В общем случае для всех кодов по законам комбинаторики для увеличения мощности кода необходимо увеличивать либо мощность алфавита (количество различных используемых символов), что не всегда оказывается оправданным с точки зрения технической реализации, либо длину кодовой комбинации.

Уменьшить алфавит кода при заданном числе N можно за счет увеличения длины кодовой комбинации, т.е. r-кратного повторения одного или нескольких символов в кодовой комбинации.

Пусть каждая кодовая комбинация содержит r_i число символов i, тогда мощность кода определится по формуле

$$N = \frac{n!}{r_0!\,r_1!\,r_2!\ldots r_{m-1}!},$$

где длина кодовой комбинации n будет определяться как $n = r_0 + r_1 + r_2 + \ldots + r_{m-1}$.

Данный код называется кодом по закону перестановок с повторением.

Пример 21. Пусть $m = 3$, $m = \{0,1,2\}$, используем повторение символов, первый пусть повторяется 2 раза, а второй и третий – без повторений, т.е. $r_0 = 2$, $r_1 = r_2 = 1$. Требуется построить код по закону перестановок с повторениями.

Решение. Длина кодовой комбинации составит $n = r_0 + r_1 + r_2 = 2 +$ $+ 1 + 1 = 4$. Тогда мощность кода равна

$$N = \frac{n!}{r_0!\,r_1!\,r_2!} = \frac{4!}{2!\cdot 1!\cdot 1!} = 3 \cdot 4 = 12.$$

Множество кодовых комбинаций {0012, 0021, 0102, 0120, 0201, 0210, 1002, 1020, 1200, 2001, 2010, 2100}.

Сменно-качественные коды

Это разновидность кодов на соединения. В данных кодах соседние символы не должны быть одинаковыми. Основание должно быть $m > 2$.

Если $m = 2$, то получим при $n = 4$ только две кодовые комбинации 0101 и 1010. Мощность кода определится по формуле [2, 8, 9]

$$N=m(m-1)^{n-1}.$$

Пример 22. Пусть алфавит содержит кодовые символы {0,1,2}, мощность кода $m = 3$. Требуется построить сменно-качественный код с кодовыми словами длины 4, т.е. $n = 4$.

Решение. Мощность кода равна $N=\boldsymbol{m(m-1)^{n-1}} = 3\cdot(3-1)^{4-1} =$
$= 3\cdot 2^3 = 24$.

Множество кодовых комбинаций
{010, 012, 020, 021, 101, 102, 120, 121, 201, 202, 212, 210}.
{1010, 1012, 1020, 1021, 0101, 0102, 0120, 0121,
0201, 0202, 0210, 0212, 2010, 2012, 2020, 2021,
1201, 1202, 1210, 1212, 2101, 2102, 2121, 2120}.

2.4. Эффективное кодирование

Задачи кодирования при отсутствии помех и при наличии помех в канале связи существенно различны. В первом случае ставится задача добиться представления элементов алфавита источника при минимальной средней значности кода, т.е. минимальным числом элементов кода в среднем на букву алфавита. Это достигается путем по возможности полной ликвидации избыточности сообщения. Во втором случае ставится задача снижения вероятности ошибок в передаче элементов сообщения. Это достигается, наоборот, введением избыточности («излишних» элементов) в кодовые слова.

Эффективным называется кодирование, при котором достигается решение первой задачи.

В большинстве случаев буквы сообщений преобразуются в последовательности двоичных символов. Учитывая статистические свойства источника сообщения, можно минимизировать среднее число двоичных символов, требующихся для выражения одной буквы сообщения, что при отсутствии шума позволяет уменьшить время передачи или объем запоминающего устройства.

Эффективное кодирование как раз и направлено на уменьшение времени передачи или объема запоминающих устройств при хранении информационных массивов и может быть достигнуто за счет минимизации среднего числа символов, требующихся для представления одной буквы сообщения. Такое эффективное кодирование базируется на теореме Шеннона о кодировании для каналов без шума. Шеннон доказал, что минимальная средняя длина кодовых слов [14]

$$\bar{N}_{\min} = \frac{H}{\log m}, \tag{45}$$

где H – энтропия источника сообщений; m – основание кода. Для бинарного кода очевидно, что

$$\bar{N}_{\min} = \frac{H}{\log_2 2} = H. \tag{46}$$

Шеннон доказал, что сообщения, составленные из букв некоторого алфавита, можно закодировать так, что среднее число двоичных символов на букву будет сколь угодно близко к энтропии источника этих сообщений, но не меньше этой величины.

Отношение $\bar{N}_{\min}$ к реально достигнутой в данном коде средней длине кодовых слов $\bar{N}$

$$\chi = \frac{\bar{N}_{\min}}{\bar{N}} = \frac{H}{\bar{N}\log m} \tag{47}$$

называется *эффективностью кода*. Качество кода можно охарактеризовать также его избыточностью

$$r = 1 - \frac{n_{\min}}{n}, \tag{48}$$

где n – использованное число элементов кода для передачи, а $n_{\min}$ – теоретически возможное число элементов.

К пределу, определяемому отношением (44), практически можно приблизиться в том смысле, что средняя длительность кодовых комбинаций будет удовлетворять условию [2, 8, 9]

$$\frac{H}{\log m} \le \bar{N} \le \frac{H}{\log m} + 1. \tag{49}$$

Средняя длительность кодовых комбинаций может быть подсчитана как

$$\bar{N} = \sum_{i=1}^{M} p\left(z_i\right) n_i \cdot \qquad (50)$$

где $p(z_i)$ – вероятность элементов из алфавита M; n_i – длительность кодовой комбинации, соответствующей элементу z_i. Если все кодовые комбинации равновероятны, то, очевидно, все n_i одинаковы, и тогда $\bar{N}_{\min} = n_i = n \cdot$

Отношение максимальной энтропии к средней длине кодовых комбинаций спроектированного кода называется коэффициентом статистического сжатия:

$$K_{c.c.} = \frac{H_{\max}}{\bar{N}} = \frac{\log m}{\sum_{i=1}^{M} p(z_i) n_i} \cdot \qquad (51)$$

Теорема Шеннона, к сожалению, не указывает конкретного способа кодирования, и общее правило получения эффективного кода неизвестно. Однако из теоремы следует, что при выборе каждого символа кодовой комбинации необходимо стараться, чтобы он нес максимальную информацию. Следовательно, для бинарного кода каждый символ должен принимать значения 0 и 1 по возможности с равными вероятностями, и каждый выбор должен быть независим от значений предыдущих символов.

Принципы близкого к эффективному кодирования заключаются в следующем [2, 8, 9]:

– длина кодового слова должна быть обратно пропорциональна вероятности появления в сообщении соответствующего элемента алфавита;

– начало более длинного кодового слова не должно совпадать с более коротким (условие префиксности, необходимое для однозначного декодирования без использования дополнительных разделительных знаков);

– в длинной последовательности элементы кода должны быть независимы и равновероятны.

Алгоритм кодирования по методу Шеннона – Фано

Для случая отсутствия статистической взаимосвязи между буквами конструктивные методы построения эффективных кодов были даны впервые Шенноном и Фано. Их методики существенно не отличаются, и поэтому соответствующий код получил название кода Шеннона – Фано.

Согласно методике Шеннона – Фано построение оптимального кода сводится к следующему [2, 8, 9].

1. Буквы алфавита сообщений располагаются в порядке убывания вероятностей.

2. Алфавит кодируемых букв разбивается на две группы таким образом, чтобы суммы вероятностей букв обеих групп были бы по возможности равны. Если абсолютно равной вероятности в подгруппах нельзя достичь, то их делят так, чтобы в верхней части (в верхней подгруппе) оставались буквы, суммарная вероятность которых меньше суммарной вероятности букв в нижней части (в нижней полугруппе).

3. Первой группе (верхней полугруппе) в качестве первого символа приписывается символ 1, второй группе – 0.

4. Каждая из полученных групп в свою очередь разбивается на две части таким образом, чтобы суммарные вероятности вновь образованных подгрупп были по возможности равны.

5. Первым группам каждой из подгрупп вновь приписывается 1, а вторым – 0. Таким образом, получают вторые цифры кода. Затем каждая из четырех групп вновь делится на равные (с точки зрения суммарной вероятности) части и т.д. Процесс повторяется до тех пор, пока в каждой подгруппе не останется по одной букве.

Алгоритм Шеннона – Фано применим и при основании кода $m > 2$. В этом случае алфавит разбивается на m частей примерно одинаковой суммарной вероятности.

Рассмотрим алфавит из восьми букв, вероятности появления которых равны $p(z_1) = 1/2$, $p(z_2) = 1/4$, $p(z_3) = 1/8$, $p(z_4) = 1/16$, $p(z_5) = 1/32$, $p(z_6) = 1/64$, $p(z_7) = 1/128$, $p(z_8) = 1/128$. В этом примере первая подгруппа при каждом разделении оказывается состоящей из одного элемента. Ясно, что при обычном (не учитывающем статистических характеристик) кодировании для представления каждой буквы требуется три символа.

Характеристики такого ансамбля и коды букв представлены в табл. 7.

При кодировании по методике Шеннон – Фано наибольший эффект сжатия получается в случае, когда вероятности букв представляют собой целочисленные отрицательные степени двойки. Среднее число символов на букву в этом случае точно равно энтропии. Убедимся в этом, вычислив энтропию

$$H(z) = -\sum_{i=1}^{8} p(z_i) \log p(z_i) = 1\frac{63}{64},$$

а среднее число символов на букву

$$\bar{N} = \sum_{i=1}^{8} p(z_i) n(z_i) = 1\frac{63}{64},$$

где $n(z_i)$ – число символов в кодовой комбинации, соответствующей букве z_i (число разрядов в кодовой комбинации здесь и в дальнейшем обозначено через n для того, чтобы избежать расхождения с общепринятой терминологией в области эффективного и помехоустойчивого кодирования).

Таблица 7

Результат кодирования по методу Шеннона – Фано

Буквы	Вероятности	Кодовые комбинации	Номер разбиения
z_1	1/2	1	I
z_2	1/4	01	II
z_3	1/8	001	III
z_4	1/16	0001	IV
z_5	1/32	00001	V
z_6	1/64	000001	VI
z_7	1/128	0000001	VII
z_8	1/128	0000000	

В более общем случае для алфавита из восьми букв среднее число символов на букву будет меньше трех, но больше энтропии алфавита $H(Z)$.

Для ансамбля, приведенного в табл. 8, энтропия равна 2,76, а среднее число символов на букву равно 2,84.

Таблица 8

Результат кодирования по методу Шеннона – Фано, 1-й вариант построения

Буквы	Вероятности	Кодовые комбинации	Номер разбиения
z_1	0,22	11	II
z_2	0,20	101	III
z_3	0,16	100	I
z_4	0,16	01	IV
z_5	0,10	001	V

Окончание табл. 8

Буквы	Вероятности	Кодовые комбинации	Номер разбиения
z_6	0,10	0001	VI
z_7	0,04	00001	VII
z_8	0,02	00000	

Следовательно, некоторая избыточность в последовательностях символов осталась. Из теоремы Шеннона следует, что эту избыточность также можно устранить, если перейти к кодированию достаточно большими блоками, о чем более подробно будет изложено ниже в разделе о повышении эффективности кодирования.

Рассмотренная нами методика Шеннона – Фано не всегда приводит к однозначному построению кода. Ведь при разбиении на подгруппы можно сделать большей по вероятности как верхнюю, так и нижнюю подгруппы.

В табл. 8 разбиение можно было произвести другим способом, например, так, как это показано в табл. 9 [12].

Таблица 9

Результат кодирования по методу Шеннона – Фано
2-й вариант построения

Буквы	Вероятности	Кодовые комбинации	Номер разбиения
z_1	0,22	1 1	II
z_2	0,20	1 0	I
z_3	0,16	0 1 1	IV
z_4	0,16	0 1 0	III
z_5	0,10	0 0 1	V
z_6	0,10	0 0 0 1	VI
z_7	0,04	0 0 0 0 1	VII
z_8	0,02	0 0 0 0 0	

При этом среднее число символов на букву оказывается равным 2,80. Таким образом, построенный код может оказаться не самым лучшим. При построении эффективных кодов с основанием $m > 2$ неопределенность становится еще больше.

Для преодоления указанного недостатка можно использовать алгоритм Хаффмена.

Алгоритм кодирования по методу Хаффмена

Метод Хаффмена гарантирует однозначное построение кода с наименьшим для данного распределения вероятностей средним числом символов на букву.

Для двоичного кода методика сводится к следующему [2, 8, 9].

1. Буквы алфавита сообщений выписываются в основной столбец в порядке убывания вероятностей.

2. Две последние буквы объединяются в одну вспомогательную букву, которой приписывается суммарная вероятность.

3. Вероятности букв, не участвовавших в объединении, и полученная суммарная вероятность снова располагаются в порядке убывания вероятностей в дополнительном столбце, а две последние объединяются. Процесс продолжается до тех пор, пока не получим единственную вспомогательную букву с вероятностью, равной единице.

Рассмотрим процедуру формирования кода по алгоритму Хаффмена на примере.

Пример 23. Используем исходные данные, представленные в табл. 8. Процедура формирования кода представлена в табл. 10.

Таблица 10

Построение кода по методу Хаффмена

Буквы	Вероятности	Вспомогательные столбцы						
		1	2	3	4	5	6	7
z_1	0,22	0,22	0,22	0,26	0,32	0,42	0,58	1
z_2	0,20	0,20	0,20	0,22	0,26	0,32	0,42	
z_3	0,16	0,16	0,16	0,20	0,22	0,26		
z_4	0,16	0,16	0,16	0,16	0,20			
z_5	0,10	0,10	0,16	0,16				
z_6	0,10	0,10	0,10					
z_7	0,04	0,06						
z_8	0,02							

Для составления кодовой комбинации, соответствующей данному сообщению, необходимо проследить путь перехода сообщения по строкам и столбцам таблицы.

Для наглядности строится кодовое дерево. Из точки, соответствующей вероятности 1, направляются две ветви, причем ветви с большей вероятностью присваивается символ 1, а с меньшей 0. Такое последовательное ветвление продолжаем до тех пор, пока не дойдем до вероятности каждой буквы. Кодовое дерево для алфавита букв, рассматриваемого в примере табл. 10, приведено на рис. 19.

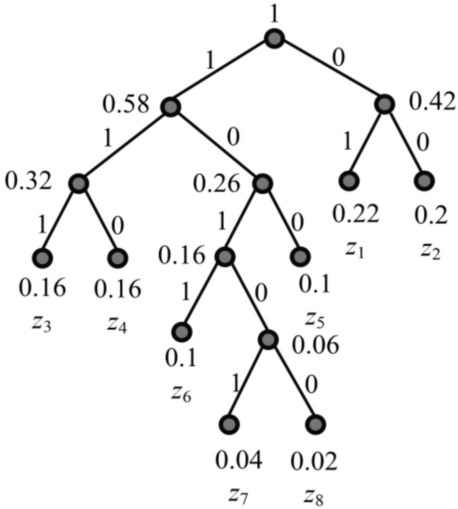


Рис. 19. Кодовое дерево, построенное по табл. 10
в соответствии с методом Хаффмена

Теперь, двигаясь по кодовому дереву сверху вниз, можно записать для каждой буквы соответствующую ей кодовую комбинацию:

z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
01	00	111	110	100	1011	10101	10100

Рассмотрев методики построения эффективных кодов, нетрудно убедиться в том, что эффект достигается благодаря присвоению более коротких кодовых комбинаций более вероятным буквам и более длинных – менее вероятным буквам. Таким образом, эффект связан с различием в числе символов кодовых комбинаций. А это приводит к трудностям при декодировании. Конечно, для различения кодовых комбинаций можно ставить специальный разделительный символ, но при этом значительно снижается эффект, которого мы добивались, т.к. средняя длина кодовой комбинации по существу увеличивается на символ.

Более целесообразно обеспечить однозначное декодирование без введения дополнительных символов. Для этого эффективный код необходимо строить так, чтобы ни одна комбинация кода не совпадала с началом более длинной комбинации. Коды, удовлетворяющие этому условию, называются префиксными кодами. Последовательность комбинаций префиксного кода, например кода

z_1	z_2	z_3	z_4
00	01	101	100

,

декодируется однозначно

100	00	01	101	101	101	00
z_4	z_1	z_2	z_3	z_3	z_3	z_1

.

Последовательность комбинаций непрефиксного кода, например кода

z_1	z_2	z_3	z_4
00	01	101	010

(комбинация 01 является началом комбинации 010), может быть декодирована по-разному:

00	01	01	01	010	101
z_1	z_2	z_2	z_2	z_4	z_3

,

00	010	101	010	101
z_1	z_4	z_3	z_4	z_3

,

или

00	01	010	101	01	01
z_1	z_2	z_4	z_3	z_2	z_2

.

Нетрудно убедиться, что коды, получаемые в результате применения методики Шеннона – Фано или Хаффмена, являются префиксными.

Пример 24. Построить код по алгоритму Хаффмена для передачи символов, имеющих следующие вероятности: $p(z_1) = 0{,}35$, $p(z_2) = 0{,}06$, $p(z_3) = 0{,}1$, $p(z_4) = 0{,}24$, $p(z_5) = 0{,}04$, $p(z_6) = 0{,}05$, $p(z_7) = 0{,}01$, $p(z_8) = 0{,}15$. Построить кодовое дерево. Определить среднюю длину кодовых сообщений.

Решение. Предварительно расположим вероятности в убывающем порядке $p(z_1) = 0{,}35$, $p(z_4) = 0{,}24$, $p(z_8) = 0{,}15$, $p(z_3) = 0{,}10$, $p(z_2) = 0{,}06$, $p(z_6) = 0{,}05$, $p(z_5) = 0{,}04$, $p(z_7) = 0{,}01$. Далее заполним табл. 11, где представлен процесс суммирования вероятностей двух последних значений в ранжированном ряду с последующей пересортировкой вероятностей.

Таблица 11

Построение кода по методу Хаффмена

Буквы	Вероятности	Вспомогательные столбцы						
		1	2	3	4	5	6	7
z_1	0,35	0,35	0,35	0,35	0,35	0,40	0,60	1
z_4	0,24	0,24	0,24	0,24	0,25	0,35	0,40	
z_8	0,15	0,15	0,15	0,16	0,24	0,25		
z_3	0,10	0,10	0,10	0,15	0,16			
z_2	0,06	0,06	0,10	0,10				
z_6	0,05	0,05	0,06					
z_5	0,04	0,05						
z_7	0,01							

Далее построим кодовое дерево. Двигаясь от правого столбца к началу таблицы, будем отображать ветвление на дереве (рис. 20).

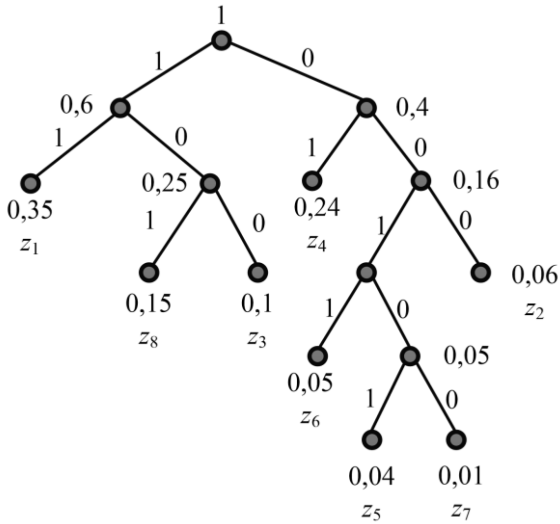


Рис. 20. Кодовое дерево, построенное по табл. 11
в соответствии с методом Хаффмена

Последовательный переход от корня дерева к листьям позволяет определить кодовые комбинации для исходных символов сообщений. Сразу добавим строку с вероятностью и длиной получившейся кодовой комбинации:

Буква	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
Код	11	000	100	01	00101	0011	00100	101
n_i	2	3	3	2	5	4	5	3
$p(z_i)$	0,35	0,06	0,1	0,24	0,04	0,05	0,01	0,15

Определим среднюю длину кодовых сообщений по формуле (50)

$$\overline{N} = 0{,}35 \cdot 2 + 0{,}06 \cdot 3 + 0{,}1 \cdot 3 + 0{,}24 \cdot 2 + 0{,}04 \cdot 5 + 0{,}05 \cdot 4 + 0{,}01 \cdot 5 + 0{,}15 \cdot 3 = 2{,}56 \cdot$$

Пример 25. Построить оптимальный код для передачи сообщений, имеющих следующие вероятности появления: $p(z_1)$ = 0,5, $p(z_2)$ = 0,25, $p(z_3)$ = 0,098, $p(z_4)$ = 0,052, $p(z_5)$ = 0,04, $p(z_6)$ = 0,03, $p(z_7)$ = 0,019, $p(z_8)$ = = 0,011. Определить коэффициент статистического сжатия и коэффициент относительной эффективности.

Решение. Заметим, что буквы первичного алфавита уже упорядочены по убыванию вероятностей их появления в сообщении, в противном случае их потребовалось бы предварительно упорядочить. Используем алгоритм Шеннона – Фано. Результаты представлены в табл. 12.

Таблица 12

Результат кодирования по методу Шеннона – Фано

Буква, z_i	Вероятности появления, $p(z_i)$	Кодовое слово	Количество знаков в кодовом слове, n_i	$p(z_i)\cdot n_i$
z_1	0,5	0	1	0,5
z_2	0,25	1 0	2	0,5
z_3	0,098	1 1 0 0	4	0,392
z_4	0,052	1 1 0 1	4	0,208
z_5	0,04	1 1 1 0	4	0,16
z_6	0,03	1 1 1 1 0	5	0,15
z_7	0,019	1 1 1 1 1 0	6	0,114
z_8	0,011	1 1 1 1 1 1	6	0,066

Определим энтропию сообщений по формуле Шеннона (13)

$$H = \frac{I}{n} = -\sum_{i=1}^{8} p_i \log p_i = -(0{,}5\log_2 0{,}5 + 0{,}25\log_2 0{,}25 + ... + 0{,}011\log_2 0{,}011) =$$
$$= 0{,}5 + 0{,}5 + 0{,}32841 + 0{,}2218 + 0{,}18575 + 0{,}15177 + 0{,}10864 + 0{,}07187 =$$
$$= 2{,}06793\,\textit{бит}/\textit{символ}.$$

Коэффициент статистического сжатия вычислим по формуле (51)

$$K_{с.с.} = \frac{H_{\max}}{\overline{N}} = \frac{\log_2 8}{\sum_{i=1}^{8} p(z_i) n_i} = \frac{\log_2 8}{0{,}5 + 0{,}5 + ... + 0{,}066} = \frac{3}{2{,}09} = 1{,}4354.$$

Коэффициент эффективности вычислим по формуле (50)

$$\chi = \frac{H}{\overline{N}\log m} = \frac{2{,}06793}{2{,}09 \cdot 3} = 0{,}32981.$$

Коэффициент относительной эффективности равен

$$K_{о.э.} = \frac{H}{\overline{N}} = \frac{2{,}06793}{2{,}09} = 0{,}98944.$$

Пример 26. Распределение вероятностей появления букв в сообщении равны: $p(z_1) = 0{,}15$, $p(z_2) = 0{,}01$, $p(z_3) = 0{,}02$, $p(z_4) = 0{,}05$, $p(z_5) = 0{,}06$, $p(z_6) = 0{,}011$, $p(z_7) = 0{,}016$, $p(z_8) = 0{,}004$, $p(z_9) = 0{,}019$, $p(z_{10}) = 0{,}007$, $p(z_{11}) = 0{,}11$, $p(z_{12}) = 0{,}08$, $p(z_{13}) = 0{,}041$, $p(z_{14}) = 0{,}023$, $p(z_{15}) = 0{,}018$, $p(z_{16}) = 0{,}015$, $p(z_{17}) = 0{,}087$, $p(z_{18}) = 0{,}063$, $p(z_{19}) = 0{,}072$, $p(z_{20}) = 0{,}014$, $p(z_{21}) = 0{,}055$, $p(z_{22}) = 0{,}075$.

Построить оптимальный код в четверичном алфавите для передачи сообщений, используя алгоритм Шеннона – Фано.

Решение. Предварительно запишем символы в порядке убывания вероятностей их появления.

$p(z_1) = 0{,}15$, $p(z_{11}) = 0{,}11$, $p(z_{17}) = 0{,}087$, $p(z_{12}) = 0{,}08$, $p(z_{22}) = 0{,}075$, $p(z_{19}) = 0{,}072$, , $p(z_{18}) = 0{,}063$, $p(z_5) = 0{,}06$, $p(z_{21}) = 0{,}055$, $p(z_4) = 0{,}05$, $p(z_{13}) = 0{,}041$, $p(z_{14}) = 0{,}023$, $p(z_3) = 0{,}02$, $p(z_9) = 0{,}019$, $p(z_{15}) = 0{,}018$, $p(z_7) = 0{,}016$, $p(z_{16}) = 0{,}015$, $p(z_{20}) = 0{,}014$, $p(z_6) = 0{,}011$, $p(z_2) = 0{,}01$, $p(z_{10}) = 0{,}007$, $p(z_8) = 0{,}004$.

Поскольку вторичный алфавит должен быть четверичным, то алгоритм Шеннона – Фано требует модификации. На каждом шаге будем разбивать множество первичных символов на 4 группы (по мощности вторичного алфавита) таким образом, чтобы суммарные вероятности появления букв были по возможности равны, а получающимся группам и подгруппам будем присваивать кодовые символы 0, 1, 2 и 3.

Результат применения такой процедуры представлен в табл. 13.

Средняя длина кодовых сообщений при этом составит

$$\overline{N} = 0{,}15 \cdot 2 + 0{,}11 \cdot 2 + ... + 0{,}004 \cdot 4 = 2{,}244.$$

Таблица 13

Результат кодирования по методу Шеннона – Фано в четверичном коде

Буква, z_i	Вероятности появления, $p(z_i)$	Кодовое слово	Количество знаков в кодовом слове, n_i	$p(z_i)\cdot n_i$
z_1	0,150	0 0	2	0,300
z_{11}	0,110	0 1	2	0,220
z_{17}	0,087	1 0	2	0,174
z_{12}	0,080	1 1	2	0,160
z_{22}	0,075	1 2	2	0,150
z_{19}	0,072	2 0	2	0,144
z_{18}	0,063	2 1	2	0,126
z_5	0,060	2 2	2	0,120
z_{21}	0,055	2 3	2	0,110
z_4	0,050	3 0	2	0,100
z_{13}	0,041	3 1 0	3	0,123
z_{14}	0,023	3 1 1	3	0,069
z_3	0,020	3 2 0	3	0,060
z_9	0,019	3 2 1	3	0,057
z_{15}	0,018	3 2 2	3	0,054
z_7	0,016	3 3 0	3	0,048
z_{16}	0,015	3 3 1	3	0,045
z_{20}	0,014	3 3 2 0	4	0,056
z_6	0,011	3 3 2 1	4	0,044
z_2	0,010	3 3 3 0	4	0,040
z_{10}	0,007	3 3 3 1	4	0,028
z_8	0,004	3 3 3 2	4	0,016

Пример 27. Распределение вероятностей появления букв в сообщении равны: $p(z_1) = 0{,}007$, $p(z_2) = 0{,}013$, $p(z_3) = 0{,}22$, $p(z_4) = 0{,}022$, $p(z_5) = 0{,}062$, $p(z_6) = 0{,}002$, $p(z_7) = 0{,}001$, $p(z_8) = 0{,}033$, $p(z_9) = 0{,}017$, $p(z_{10}) = 0{,}048$, $p(z_{11}) = 0{,}032$, $p(z_{12}) = 0{,}12$, $p(z_{13}) = 0{,}026$, $p(z_{14}) = 0{,}051$, $p(z_{15}) = 0{,}003$, $p(z_{16}) = 0{,}016$, $p(z_{17}) = 0{,}015$, $p(z_{18}) = 0{,}087$, $p(z_{19}) = 0{,}092$, $p(z_{20}) = 0{,}075$, $p(z_{21}) = 0{,}028$, $p(z_{22}) = 0{,}009$, $p(z_{23}) = 0{,}021$.

Построить оптимальный код в троичном алфавите для передачи сообщений, используя алгоритм Хаффмена.

Решение. Упорядочиваем буквы первичного алфавита по убыванию вероятностей их появления.

$p(z_3) = 0{,}22$, $p(z_{12}) = 0{,}12$, $p(z_{19}) = 0{,}092$, $p(z_{18}) = 0{,}087$, $p(z_{20}) = 0{,}075$, $p(z_5) = 0{,}062$, $p(z_{14}) = 0{,}051$, $p(z_{10}) = 0{,}048$, $p(z_8) = 0{,}033$, $p(z_{11}) = 0{,}032$, $p(z_{21}) = 0{,}028$, $p(z_{13}) = 0{,}026$, $p(z_4) = 0{,}022$, $p(z_{23}) = 0{,}021$, $p(z_9) = 0{,}017$, $p(z_{16}) = 0{,}016$, $p(z_{17}) = 0{,}015$, $p(z_2) = 0{,}013$, $p(z_{22}) = 0{,}009$, $p(z_1) = 0{,}007$, $p(z_{15}) = 0{,}003$, $p(z_6) = 0{,}002$, $p(z_7) = 0{,}001$.

Аналогично предыдущему примеру для троичного кодирующего алфавита алгоритм Хаффмена требует некоторых модификаций. На каждом шаге необходимо объединять не две, а три последние буквы в ранжированном ряду, и далее пересортировывать последовательность. Описанный процесс представлен в табл. 14.

Таблица 14

Построение кода по методу Хаффмена в троичном коде

z_i	$p(z_i)$	Вспомогательные столбцы										
		1	2	3	4	5	6	7	8	9	10	11
z_3	0,220	0,220	0,220	0,220	0,220	0,220	0,220	0,220	0,220	0,288	0,492	1
z_{12}	0,120	0,120	0,120	0,120	0,120	0,120	0,120	0,159	0,213	0,220	0,288	
z_{19}	0,092	0,092	0,092	0,092	0,092	0,092	0,109	0,120	0,159	0,213	0,220	
z_{18}	0,087	0,087	0,087	0,087	0,087	0,087	0,092	0,109	0,120	0,159		
z_{20}	0,075	0,075	0,075	0,075	0,075	0,076	0,087	0,092	0,109	0,120		
z_5	0,062	0,062	0,062	0,062	0,062	0,075	0,076	0,087	0,092			
z_{14}	0,051	0,051	0,051	0,051	0,060	0,062	0,075	0,076	0,087			
z_{10}	0,048	0,048	0,048	0,048	0,051	0,060	0,062	0,075				
z_8	0,033	0,033	0,033	0,044	0,048	0,051	0,060	0,062				
z_{11}	0,032	0,032	0,032	0,033	0,044	0,048	0,051					
z_{21}	0,028	0,028	0,028	0,032	0,033	0,044	0,048					
z_{13}	0,026	0,026	0,026	0,028	0,032	0,033						
z_4	0,022	0,022	0,022	0,026	0,028	0,032						
z_{23}	0,021	0,021	0,022	0,022	0,026							
z_9	0,017	0,017	0,021	0,022	0,022							
z_{16}	0,016	0,016	0,017	0,021								
z_{17}	0,015	0,015	0,016	0,017								
z_2	0,013	0,013	0,015									
z_{22}	0,009	0,009	0,013									
z_1	0,007	0,007										
z_{15}	0,003	0,006										
z_6	0,002											
z_7	0,001											

Далее построим кодовое дерево. Двигаясь от правого столбца к началу таблицы, будем производить ветвление уже по трем дугам, присваивая кодовые символы 2, 1 и 0 переходам слева направо соответственно. Результат построения кодового дерева представлен на рис. 21.

Формирование кодовых комбинаций осуществляется по общему правилу и приведены ниже:

Буква	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}	z_{11}	z_{12}
Код	21221	1220	0	2220	220	212201	212200	121	2120	210	120	20
n_i	5	4	1	4	3	6	6	3	4	3	3	2

Буква	z_{13}	z_{14}	z_{15}	z_{16}	z_{17}	z_{18}	z_{19}	z_{20}	z_{21}	z_{22}	z_{23}
Код	2221	211	212202	1222	1221	10	11	221	2222	21222	2121
n_i	4	3	6	4	4	2	2	3	4	5	4

Средняя длина кодовых сообщений составит:

$$\overline{N} = 0{,}007 \cdot 5 + 0{,}013 \cdot 4 + \ldots + 0{,}021 \cdot 4 = 2{,}469 .$$

Алгоритмы Шеннона – Фано и Хаффмена предназначены для кодирования независимых сообщений. Если источник характеризуется зависимостью элементов, то использование этих алгоритмов возможно по отношению к парам элементов, однако предварительно должны быть вычислены условные вероятности их совместного появления.

Принципы оптимального кодирования нашли применение в задачах технической диагностики сложных объектов (в том числе и радиотехнических, и телекоммуникационных систем), для которых известны статистические или логические связи между результатами проверок и видами неисправностей. Поиск неисправностей организуется таким образом, чтобы каждая проверка обеспечивала максимальную информационную нагрузку, т.е. снимала бы наибольшую неопределенность на каждом шаге.

Повышение эффективности кодирования

Поскольку в соответствии с методами Шеннона – Фано и Хаффмена буквам, имеющим большую вероятность, присваиваются более короткие кодовые комбинации, чем сравнительно маловероятным буквам, в результате, хотя некоторые кодовые комбинации здесь и могут иметь весьма значительную длину, среднее значение длины такой комбинации все же оказывается лишь немногим большим значения энтропии источника сообщений.

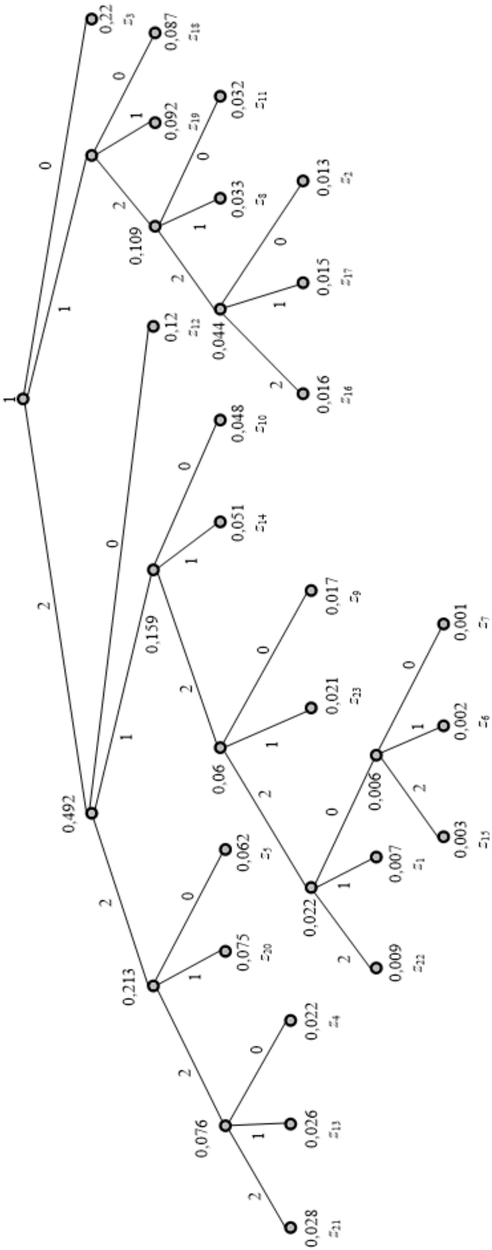


Рис. 21. Кодовое дерево, построенное по табл. 14 в соответствии с методом Хаффмена

Тем не менее, достигнутая в рассмотренных примерах степень близости среднего числа двоичных символов, приходящихся на одну букву сообщения, к значению H может быть еще сколь угодно увеличена при помощи перехода к кодированию все более и более длинных блоков по методам Шеннона – Фано и Хаффмена.

Повысить эффективность кодирования можно, строя код не для символа, а для блоков из n символов, причем частота блока рассчитывается как произведение частот символов, входящих в блок [2, 8, 9].

В случае поблочного кодирования, где каждый из блоков состоит из M независимых букв $a_1, a_2, \ldots, a_M$, минимальная средняя длина кодового блока лежит в пределах:

$$\frac{MH}{\log m} \le \bar{N} \le \frac{MH}{\log m} + 1 \cdot \tag{52}$$

Общее выражение среднего числа элементарных символов на букву сообщения при поблочном кодировании

$$\frac{H}{\log m} \le \bar{N} \le \frac{H}{\log m} + \frac{1}{M} \cdot \tag{53}$$

С точки зрения информационной нагрузки на символ сообщения поблочное кодирование всегда выгоднее, чем побуквенное. В общем случае «выгода» блочного кодирования получается за счет того, что в блоках происходит выравнивание вероятностей отдельных букв, что ведет к повышению информационной нагрузки на символ.

Подчеркнем в заключение особенности систем эффективного кодирования. Одна из них обусловлена различием в длине кодовых комбинаций. Если моменты снятия информации с источника не управляемы (например, при непрерывном съеме информации с запоминающего устройства па магнитной ленте), кодирующее устройство через равные промежутки времени выдает комбинации различной длины. Поскольку линия связи используется эффективно только в том случае, когда символы поступают в нее с постоянной скоростью, то на выходе кодирующего устройства должно быть предусмотрено буферное устройство («упругая» задержка). Оно запасает символы по мере поступления и выдает их в линию связи с постоянной скоростью. Аналогичное устройство необходимо и на приемной стороне.

Вторая особенность связана с возникновением задержки в передаче информации.

Наибольший эффект достигается при кодировании длинными блоками, а это приводит к необходимости накапливать буквы, прежде чем поставить им в соответствие определенную последовательность символов. При декодировании задержка возникает снова. Общее время задержки может быть велико, особенно при появлении блока, вероятность которого мала. Это следует учитывать при выборе длины кодируемого блока.

Еще одна особенность заключается в специфическом влиянии помех на достоверность приема. Одиночная ошибка может перевести передаваемую кодовую комбинацию в другую, не равную ей по длительности. Это повлечет за собой неправильное декодирование целого ряда последующих комбинаций, который называют треком ошибки.

Специальными методами построения эффективного кода трек ошибки стараются свести к минимуму.

Покажем особенности блочного кодирования на конкретном примере. Рассмотрим сообщения, образованные с помощью алфавита, состоящего всего из двух букв z_1 и z_2 с вероятностями появления соответственно $p(z_1) = 0{,}9$ и $p(z_2) = 0{,}1$.

Поскольку вероятности не равны, то последовательность таких букв будет обладать избыточностью. Однако при буквенном кодировании мы никакого эффекта не получим.

Действительно, на передачу каждой буквы требуется символ либо 1, либо 0, в то время как энтропия равна 0,47.

Осуществим кодирование блоками по две буквы. Для этого необходимо предусмотреть все возможные комбинации сочетаний появления букв, в данном случае их будет всего 4: $z_1\ z_1$, z_1z_2, $z_2\ z_1$, $z_2\ z_2$. Вероятности появления таких блоков определим на основании теоремы умножения вероятностей, после чего можно применять любой из алгоритмов оптимального кодирования.

При кодировании блоков, содержащих по две буквы, использовали алгоритм Шеннона – Фано, в результате получим коды, показанные в табл. 15.

Определим эффективность построенного кода. Так как буквы статистически не связаны, вероятности блоков определяются как произведение вероятностей составляющих букв. Среднее число символов на блок получается равным

$$\overline{N}_{блока} = 0{,}81 \cdot 1 + 0{,}09 \cdot 2 + 0{,}09 \cdot 3 + 0{,}01 \cdot 3 = 1{,}29 \cdot$$

Таблица 15

Результат кодирования двухбуквенных блоков

Буквы	Вероятности	Кодовые комбинации	Номер разбиения
$z_1\ z_1$	0,81	1	I
$z_1 z_2$	0,09	01	II
$z_2\ z_1$	0,09	100	III
$z_2\ z_2$	0,01	000	IV

Поскольку в блоке два символа ($n = 2$), то для одного символа длина кодовой комбинации составит $\overline{N} = \frac{\overline{N}_{блока}}{n} = \frac{1{,}29}{2} = 0{,}645$. При посимвольном кодировании для эффективного кода потребуется по одному двоичному разряду. Применение метода Шеннона – Фано даёт результат, представленный в таблице.

Исходные символы	Частоты символов	Построение кода
a	0,9	1
b	0,1	0

Таким образом, при блочном кодировании выигрыш составил 1 – 0,645 = 0,355 двоичных разрядов на один кодируемый символ в среднем.

Эффективность блочного кодирования тем выше, чем больше символов включается в блок.

Кодирование блоков, содержащих по три буквы, дает еще больший эффект. Соответствующий ансамбль и коды приведены в табл. 16.

Таблица 16

Результат кодирования трехбуквенных блоков

Буквы	Вероятности	Кодовые комбинации	Номер разбиения
$z_1\ z_1\ z_1$	0,729	1	I
$z_2\ z_1\ z_1$	0,081	011	III
$z_1\ z_2\ z_1$	0,081	010	II
$z_1\ z_1\ z_2$	0,081	01	IV
$z_2\ z_2\ z_1$	0,009	00011	VI
$z_2\ z_1\ z_2$	0,009	00010	V
$z_1\ z_2\ z_2$	0,009	00001	VII
$z_2\ z_2\ z_2$	0,001	00000	

Среднее число символов на блок равно 1,59, а на букву 0,53, что всего на 12 % больше энтропии. Теоретический минимум $H(Z)$ = 0,47 может быть достигнут при кодировании блоков, включающих бесконечное число букв:

$$\lim_{n\to\infty} \bar{N} = H(Z).$$

Следует подчеркнуть, что увеличение эффективности кодирования при укрупнении блоков не связано с учетом все более далеких статистических связей, так как нами рассматривались алфавиты с некоррелированными буквами. Повышение эффективности определяется лишь тем, что набор вероятностей, получающийся при укрупнении блоков, можно делить на более близкие по суммарным вероятностям подгруппы.

Пример 28. Построить оптимальный неравномерный код для передачи сообщений, алфавит которых состоит из двух букв z_1 и z_2 с вероятностями $p(z_1)$ = 0,89, $p(z_2)$ = 0,11, при кодировании по одному, по два и по три символа в блоке. Оценить эффективность полученных кодов.

Таблица 17

Результат кодирования методом Шеннона – Фано блоками различной длины

Длина блока n	Блок	Вероятность появления блока, p_i	Кодовое слово	Число знаков в кодовом слове, l_i	$p_i \cdot l_i$
1	z_1	0,89	0	1	0,89
	z_2	0,11	1	1	0,11
2	$z_1\ z_1$	0,792	0	1	0,792
	$z_1\ z_2$	0,098	1 0	2	0,196
	$z_2\ z_1$	0,098	1 1 0	3	0,294
	$z_2\ z_2$	0,012	1 1 1	3	0,036
3	$z_1\ z_1\ z_1$	0,705	0	1	0,705
	$z_1\ z_1\ z_2$	0,087	1 0 0	3	0,261
	$z_1\ z_2\ z_1$	0,087	1 0 1	3	0,261
	$z_1\ z_2\ z_2$	0,087	1 1 0	3	0,261
	$z_2\ z_1\ z_1$	0,011	1 1 1 0 0	5	0,055
	$z_2\ z_1\ z_2$	0,011	1 1 1 0 1	5	0,055
	$z_2\ z_2\ z_1$	0,011	1 1 1 1 0	5	0,055
	$z_2\ z_2\ z_2$	0,001	1 1 1 1 1	5	0,055

Решение. Результаты кодирования с использованием алгоритма Шеннона – Фано представлены в табл. 17.

Среднее число символов на блок в каждом случае составляет:

$$\overline{N}_1 = 0{,}89 \cdot 1 + 0{,}11 \cdot 1 = 1,$$

$$\overline{N}_2 = 0{,}792 \cdot 1 + 0{,}098 \cdot 2 + 0{,}098 \cdot 3 + 0{,}012 \cdot 3 = 1{,}318,$$

$$\overline{N}_2 = 0{,}705 \cdot 1 + 3(0{,}087 \cdot 3) + 3(0{,}011 \cdot 5) + 0{,}001 \cdot 5 = 1{,}658.$$

Энтропия блоковых сообщений в каждом случае составляет:

$$H_1 = -\sum_{i=1}^{2} p_i \log p_i = -(0{,}89 \cdot \log_2 0{,}89 + 0{,}11 \cdot \log_2 0{,}11) =$$
$$= 0{,}3503 + 0{,}1496 = 0{,}499 \; \textit{бит/символ},$$

$$H_2 = -\sum_{i=1}^{4} p_i \log p_i = -(0{,}792 \cdot \log_2 0{,}792 + 2 \cdot 0{,}098 \cdot \log_2 0{,}098 + 0{,}012 \cdot \log_2 0{,}012) =$$
$$= 0{,}2664 + 2 \cdot 0{,}3284 + 0{,}076 = 0{,}9948 \; \textit{бит/символ},$$

$$H_3 = -\sum_{i=1}^{8} p_i \log p_i = -(0{,}705 \cdot \log_2 0{,}705 + 3 \cdot 0{,}087 \cdot \log_2 0{,}087 + 3 \cdot 0{,}011 \cdot \log_2 0{,}011 +$$
$$+ 0{,}001 \cdot \log_2 0{,}001) = 0{,}3555 + 3 \cdot 0{,}9065 + 3 \cdot 0{,}0716 + 0{,}01 = 1{,}4998 \; \textit{бит/символ}.$$

Коэффициент статистического сжатия в каждом случае составляет:

$$K_{с.с.1} = \frac{H_{\max 1}}{\overline{N}_1} = \frac{\log_2 2}{1} = 1,$$

$$K_{с.с.2} = \frac{H_{\max 2}}{\overline{N}_2} = \frac{\log_2 4}{1{,}318} = 1{,}517,$$

$$K_{с.с.3} = \frac{H_{\max 3}}{\overline{N}_3} = \frac{\log_2 8}{1{,}658} = 1{,}809.$$

Коэффициент относительной эффективности в каждом случае составляет:

$$K_{о.э.1} = \frac{H_1}{\overline{N}_1} = \frac{0{,}499}{1} = 0{,}499,$$

$$K_{о.э.2} = \frac{H_2}{\overline{N}_1} = \frac{0{,}9948}{1{,}318} = 0{,}7548,$$

$$K_{о.э.1} = \frac{H_1}{\overline{N}_1} = \frac{1{,}4998}{1{,}658} = 0{,}9046.$$

С увеличением числа символов в блоке эффективность кода быстро растет до определенного предела. Затем рост эффективности постепенно уменьшается. Практика показывает, что с увеличением символов в блоке (более 4) сложность кодирующих устройств растет быстрее, чем эффективность.

Декодирование эффективных кодов

Особенностью эффективных кодов является переменное число двоичных разрядов в получаемых кодовых комбинациях. Это затрудняет процесс декодирования.

Рассмотрим вначале как происходит декодирование сообщения, если использовались коды постоянной длины. Пусть кодовая таблица имеет следующий вид:

Исходные символы	Двоичные коды
a	00
b	01
c	10
d	11

,

а закодированное сообщение – 001000011101.

Поскольку длина кода равна двум символам, в этом сообщении слева направо выделяются по два двоичных символа и сопоставляются с кодовой таблицей.

Тогда имеем

00	10	00	10	11	01
a	*c*	*a*	*b*	*d*	*b*

.

Таким образом, в исходном сообщении содержится текст *acabdb*. Декодирование выполнено. Для декодирования кодов переменной длины рассмотренный подход не годится. Но закодированные сообщения могут декодироваться благодаря свойству *префиксности* эффективных кодов: ни одна более короткая кодовая комбинация не является началом более длинной кодовой комбинации. Для раскрытия данного тезиса воспользуемся, например, следующим эффективным кодом, обладающим свойством префиксности:

$$a – 1;\ b – 01;\ c – 001;\ d – 000.$$

Здесь самым коротким кодом является код для символа *a* со значением 1. Как видно, ни один другой код (более длинный) не имеет в начале символ 1. Второй по длине код для символа *b* имеет значение 01 и, как показывает анализ, не является началом ни для кода 001, ни для кода 000. Таким образом, данный код является префиксным. Свойство префиксности позволяет декодировать сообщения, закодированные эффективными кодами. Пусть получено сообщение 1010010001.

Выполним его декодирование. В сообщении слева направо выделяется по одному двоичному символу и делается попытка декодирования в соответствии с заданной таблицей кодов. Если попытка успешна, то двоичный символ (или символы) исключается из исходной цепочки и заменяется соответствующим исходным символом. Если попытка не удается, то во входной цепочке выделяется следующий двоичный символ и уже с двумя двоичными символами делается попытка их декодирования по таблице кодов. Если попытка и тогда неудачна, то выделяют следующий третий и т.д.

На рис. 22 представлен процесс декодирования сообщения 1010010001.

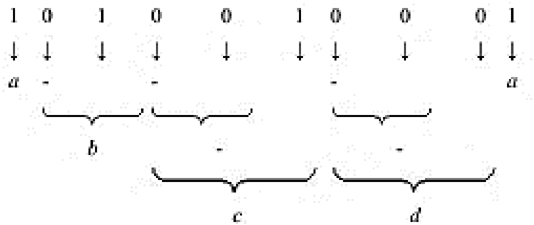


Рис. 22. Пример декодирования сообщения в префиксном коде

Здесь знак «–» означает, что попытка декодирования не удалась. Таким образом, при декодировании получили строку *abcda*.

Как это показано на примере, неравномерный код может быть однозначно декодирован, если никакой из кодов не совпадает с началом (префиксом) какого-либо иного более длинного кода. Например, если имеется код 110, то уже не могут использоваться коды 1, 11, 1101, 110101 и пр. Если условие Фано выполняется, то при прочтении (расшифровке) закодированного сообщения путем сопоставления со списком кодов всегда можно точно указать, где заканчивается один код и начинается другой.

Описанную процедуру декодирования сообщений в префиксном коде представим более формально в виде следующих шагов [3, 8, 9].

1. Выделить в текущем сообщении крайний левый символ.

2. Присоединить его к рабочему кодовому слову.

3. Сравнить рабочее кодовое слово с кодовой таблицей; если совпадения нет, перейти к (1).

4. Декодировать рабочее кодовое слово, очистить его.

5. Проверить, имеются ли еще знаки в сообщении; если "да", перейти к (1).

Пример 29. Пусть задана следующая таблица кодировки символов первичного алфавита (табл. 18).

Таблица 18

Таблица кодов

а	о	и	е	ы	в	т	ж	ч	щ	с	,
11	01	100	001	0000	1010	00011	000101	10111	000100	101100	101101

Требуется декодировать сообщение:
00011011011011011100011011010000010000010000100011001101101000110100010100110000010000100011101011101100.

Решение. Очевидно, что представленный в табл. 18 код является неравномерным. Проанализируем все коды, убеждаемся, что ни один из более коротких кодов не является началом более длинного, следовательно, код обладает свойством префиксности, и заданное сообщение можно однозначно декодировать. Применим процедуру декодирования для префиксного кода. Процесс декодирования представлен в табл. 19.

Таблица 19

Процедура декодирования сообщения

Шаг	Рабочее слово	Текущее сообщение	Распознанный знак	Декодированное сообщение
0	пусто		-	-
1	0	0011011011011011100011011010000010000010000100011001101101000110100010100110000010000100011101011101100	нет	-
2	00	011011011011011100011011010000010000010000100011001101101000110100010100110000010000100011101011101100	нет	-
3	000	11011011011011100011011010000010000010000100011001101101000110100010100110000010000100011101011101100	нет	-

Продолжение табл. 19

Шаг	Рабочее слово	Текущее сообщение	Распо-знанный знак	Декодирован-ное сообщение
4	0001	101101101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	-
5	00011	01101101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	т	т
6	0	1101101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	т
7	01	101101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	о	то
8	1	01101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то
9	10	1101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то
10	101	101101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то
11	1011	01101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то
12	10110	1101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то

Окончание табл. 19

Шаг	Рабочее слово	Текущее сообщение	Распо-знанный знак	Декодирован-ное сообщение
13	101101	101110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	,	то,
14	1	01110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то,
15	10	1110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то,
16	101	110001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то,
17	1011	10001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	нет	то,
18	10111	0001101 10100000100000100001000110 01101101000110100010100110 000010000100011101011101100	ч	то, ч
…	…	…	…	…

Доведя процедуру до конца, получим сообщение: «То, что вы ищете, тоже ищет вас».

Таким образом, использование префиксного кодирования позволяет делать сообщение более коротким, поскольку нет необходимости передавать разделители знаков. Однако условие Фано не устанавливает способа формирования префиксного кода и, в частности, наилучшего из возможных.

2.5. Помехоустойчивое кодирование

Теорема Шеннона о кодировании при наличии помех

Теория помехоустойчивого кодирования базируется на результатах исследований, проведенных Шенноном и сформулированных им в виде основной теоремы для дискретного канала с шумом: при любой скорости передачи двоичных символов, меньшей, чем пропускная способность канала, существует такой код, при котором вероятность ошибочного декодирования будет сколь угодно мала; вероятность ошибки не может быть сделана произвольно малой, если скорость передачи больше пропускной способности канала.

Шеннон доказал, что для любой линии связи с помехами всегда можно подобрать специальный код, позволяющий передавать сообщения по этой линии с заданной скоростью, сколь угодно близкой к [3, 8, 9]

$$v = \frac{c^*}{H} \frac{\text{букв}}{\text{ед.времени}} \tag{54}$$

(но не обязательно все же несколько меньшей, чем эта величина) так, чтобы вероятность ошибки в определении каждой переданной буквы оказалась меньше любого заранее заданного числа ε (например, меньшей чем 0,001 или чем 0,0001, или чем 0,000001). Здесь c^* – пропускная способность линии связи; H – энтропия одной буквы передаваемого сообщения.

Теорема Шеннона не указывает путей построения кода, обеспечивающего указанную идеальную передачу. Значение теоремы огромно, поскольку обосновывает принципиальную возможность такого кодирования.

Кодирование должно осуществляться так, чтобы сигнал, соответствующий принятой последовательности символов, после воздействия на него предполагаемой в канале помехи, оставался ближе к сигналу, соответствующему данной переданной последовательности символов, чем к сигналам, соответствующим другим возможным последовательностям. Это достигается ценой введения при кодировании избыточности, которая позволяет наложить на передаваемые последовательности символов дополнительные условия, проверка которых на приемной стороне дает возможность обнаружить и исправить ошибки.

Коды, обладающие таким свойством, получили название помехоустойчивых, или корректирующих.

Анализ библиографических источников показывает, что основные усилия направлены на проектирование помехоустойчивых кодов, способных обнаруживать и исправлять взаимно независимые ошибки с заранее определенной кратностью [3, 8, 9].

Взаимно независимыми ошибками называются такие искажения в передаваемой последовательности символов, при которых вероятность появления любой комбинации искаженных символов зависит только от числа искаженных символов r и вероятности искажения одного символа p.

Количество искаженных символов в кодовой комбинации называется *кратностью ошибки*.

При взаимно независимых ошибках вероятность искажения любых r символов в n-разрядной кодовой комбинации определяется по формуле биномиального закона:

$$p_r = C_n^r p^r \left(1 - p\right)^{n-r}. \tag{55}$$

Если учесть, что $p << 1$, то в этом случае наиболее вероятны ошибки низшей кратности. Их и следует обнаруживать, и исправлять в первую очередь.

Вектором ошибки e_p называется двоичный вектор той же значности, что и кодовые комбинации. Нули в e_p означают, что ошибки в соответствующих позициях a_k не происходят; единицы стоят на позициях, которые искажаются при передаче, так что число единиц в e_p равно кратности ошибок q.

В системах передачи информации дискретный канал связи представляется в виде модели, показанной на рис. 23.

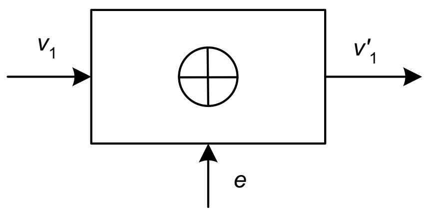


Рис. 23. Дискретный канал

Передаваемый кодовый вектор v_1 складывается в дискретном канале поразрядно по модулю 2 с вектором ошибки e, и в общем случае принимается уже другой, искаженный кодовый вектор v'_1. Например, если v_1 = 11111 и e = 01000, то v'_1 = 11111+01000 = 10111. Во втором разряде результирующей кодовой комбинации v'_1, как видим, произошла ошибка.

Таким образом, в разрядах передаваемой кодовой комбинации, соответствующих единичным разрядам вектора e, возникают ошибки.

При теоретических исследованиях процесса возникновения ошибок в дискретном канале используют математические модели ошибок. Под *математической моделью ошибки* понимается распределение вероятностей по всем возможным векторам ошибки.

В качестве примера рассмотрим одну из наиболее часто встречающихся моделей ошибки, которая основана на следующей статистической гипотезе: в каждом разряде вектора ошибки единица появляется с вероятностью p независимо от того, какие значения получили остальные разряды вектора ошибки. Назовем величину, равную числу единиц в векторе ошибки, *кратностью ошибки* и обозначим символом q. В теории вероятностей доказывается, что выдвинутой статистической гипотезе отвечает биномиальный закон распределения кратности ошибки. Таким образом, для рассматриваемого примера математической моделью ошибки может служить выражение $P_{n,q} = C_n^q p^q \left(1-p\right)^{n-q}$. Здесь $P_{n,q}$ – вероятность того, что при передаче по дискретному каналу в кодовой комбинации бинарного кода длины n возникнет ошибка кратности q.

Пример 30. Определить вероятности возникновения ошибок кратности q = 0, 1, 2, 3, 4, 5 в кодовой комбинации длины 5 бинарного кода, если вероятность ошибочного приема разряда равна 0,1. Определить вероятность ошибочного приема кодовой комбинации.

Используем модель ошибки $P_{5,q} = C_5^q 0{,}1^q 0{,}9^{5-q}$. Вычисления по данной формуле сведем в табл. 20.

Таблица 20

Вероятности ошибочного приема сообщений при кратности ошибок *q*

q	0	1	2	3	4	5
$P_{5,q}$	0,56	0,33	0,07	0,008	0,0004	0,00001

Из табл. 20 видно, что вероятность появления ошибок большой кратности мала. Наиболее часто появляются ошибки кратности 1.

Вероятность ошибочного приема кодовой комбинации определим по формуле $P_{\text{ош}} = 1 - P_{5,q} = 1 - 0{,}56 = 0{,}44$.

Математические модели ошибок должны отражать реальные процессы, происходящие в канале связи, и строиться на статистике помех. Чем точнее математическая модель описывает действительность, тем точнее можно получить оценки относительно спроектированного кода.

Следует учесть, что эффективность того или иного помехоустойчивого кода всегда зависит от вида помех, действующих в канале связи. Код может быть весьма эффективным (в том смысле, что число необнаруженных ошибок при его применении будет очень мало) при одной статистике помех и очень плохим – при другой. Поэтому при проектировании помехоустойчивых кодов необходимо ориентироваться на определенный вид помех и в соответствии с этим в качестве исходной иметь определенную модель ошибок.

Корректирующие способности кода напрямую зависят от кодового расстояния в кодовом пространстве. При взаимно независимых ошибках наиболее вероятен переход в кодовую комбинацию, отличающуюся от передаваемой в наименьшем числе символов. Расстояние между кодовыми комбинациями показывает различие между ними.

При анализе воздействия ошибок на кодовые векторы в кодовом пространстве вводится метрика. Наибольший практический интерес представляют метрики Ли и Хэмминга.

Для определения метрики Ли используется понятие *веса вектора* $v = a_1a_2...a_n$, где $a_i \in B$ $\{0, 1, 2, ..., m - 1\}$: $|v| = \sum_{i=1}^{n} |a_i|$ [3, 8, 9].

При этом величина $|a_i|$ определяется по правилу

$$|a_i| = \begin{cases} a_i, \ \text{если} \ 0 \le a_i \le m/2, \\ m - a_i, \ \text{если} \ m/2 \le a_i \le m - 1. \end{cases} \tag{56}$$

Расстоянием Ли $d(v_j, v_k)$ между векторами $v_j = a_1a_2...a_n$ и $v_k = b_1b_2...b_n$ называется вес разности векторов v_j и v_k:

$$d\left(v_j, v_k\right) = \sum_{i=1}^{n} |a_i - b_i| \cdot$$

Для двоичных кодов используется метрика Хэмминга, которая имеет наглядную геометрическую интерпретацию в кодовом пространстве. Если заданы две кодовые комбинации $x_i = \{x_{i1}, x_{i2},\ldots, x_{ik},\ldots, x_{in}\}$ и $x_j = \{x_{j1}, x_{j2},\ldots, x_{jk},\ldots, x_{jn}\}$, то для оценки расстояния между ними принято пользоваться метрикой

$$d_{ij} = \sum_{k=1}^{n} | x_{ik} - x_{jk} |. \quad (57)$$

Для двоичного кода расстояние просто равно числу знаков, в которых одна комбинация отличается от другой. *Вес вектора v* (*по Хэммингу*) равен числу ненулевых разрядов этого вектора, а *расстояние Хэмминга* между векторами v_j и v_k определяется как вес разности векторов v_j и v_k. Следует отметить, что при $m = 2$ и $m = 3$ расстояния Хэмминга и Ли совпадают.

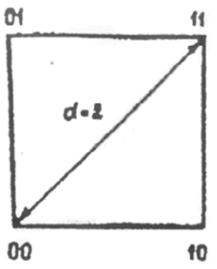
Величина $\min d(v_i, v_j)$, представляющая собой минимальное расстояние между парой векторов набора V_1, называется *кодовым расстоянием* и обозначается символом d.

Понятие кодового расстояния легко иллюстрируется, если использовать геометрическую интерпретацию. Любая n-разрядная двоичная кодовая комбинация может быть интерпретирована как вершина n-мерного единичного куба, т.е. куба с длиной ребра, равной 1.

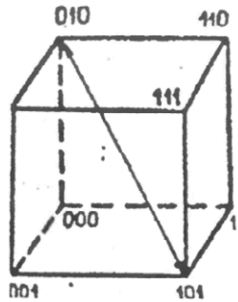
При $n = 2$ кодовые комбинации располагаются в вершинах квадрата (рис. 24, а); при $n = 3$ в вершинах единичного куба (рис. 24, б); при $n = 4$ в вершинах четырехмерного куба (рис. 24, в).

В общем случае n-мерный единичный куб имеет 2^n вершин, что равно наибольшему возможному числу кодовых комбинаций. Такая модель дает простую геометрическую интерпретацию и кодовому расстоянию между отдельными кодовыми комбинациями. Оно соответствует наименьшему числу ребер единичного куба, которые необходимо пройти, чтобы попасть от одной комбинации к другой.

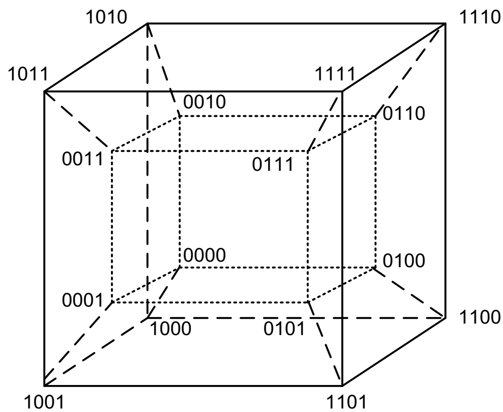
Теперь метод декодирования при исправлении одиночных независимых ошибок можно пояснить следующим образом. В подмножество каждой разрешенной комбинации относят все вершины, лежащие в сфере радиусом $(d - 1)/2$ и центром в вершине, соответствующей данной разрешенной кодовой комбинации. Если в результате действия шума комбинация переходит в точку, находящуюся внутри сферы (d – 1)/2, то такая ошибка может быть исправлена.



а



б



в

Рис. 24. Геометрическая интерпретация:
а – двухразрядного; б – трехразрядного; в – четырехразрядного двоичного кода

Если помеха смещает точку разрешенной комбинации на границу двух сфер (расстояние $d/2$) или дальше (но не в точку, соответствующую другой разрешенной комбинации), то такое искажение может быть обнаружено. Для кодов с независимым искажением символов лучшие корректирующие коды – это такие, у которых точки, соответствующие разрешенным кодовым комбинациям, расположены в пространстве равномерно [3, 8, 9].

Пример 31. Определить кодовое расстояние системы векторов $V_2 = \{11001, 01101, 01110, 00000\}$.

Прежде всего заметим, что для бинарных кодов выражение (57), определяющее метрику, может быть заменено выражением $d\left(v_j, v_k\right) = \sum_{i=1}^{n}(a_i + b_i)$, в котором под операцией, обозначенной символом «+», понимается сложение по модулю 2. Отсюда следует, что для определения расстояния между двумя кодовыми векторами бинарного кода достаточно сложить поразрядно по модулю 2 соответствующие кодовые комбинации и подсчитать число единиц в результирующей кодовой комбинации.

Чтобы определить кодовое расстояние, необходимо между каждой парой двоичных векторов произвести операцию поразрядного суммирования по модулю 2.

$$\begin{array}{r} 11001 \\ +01101 \\ \hline 10100 \end{array} \qquad \begin{array}{r} 11001 \\ +01110 \\ \hline 10111 \end{array} \qquad \begin{array}{r} 01101 \\ +01110 \\ \hline 00011 \end{array}$$

Кодовое расстояние будет равно минимальному весу результирующей двоичной комбинации. В рассмотренном нами случае $d = 2$.

Естественно предположить, что в результате действия помех будут появляться наиболее вероятные ошибки, следовательно, принятую кодовую комбинация следует отождествлять с той разрешенной, которая находится от нее на наименьшем кодовом расстоянии. При этом исправляется наиболее вероятная ошибка. Очевидно, что при $d = 1$ все кодовые комбинации являются разрешенными, и код не обладает свойствами обнаружения и исправления ошибок, а с увеличением кодового расстояния корректирующие способности кода должны увеличиваться.

Ниже приведены основные соотношения по выбору параметров кода, который должен обладать заданными корректирующими свойствами [3, 8, 9].

В общем случае при необходимости обнаруживать ошибки кратности r минимальное хэммингово расстояние между разрешенными кодовыми комбинациями должно быть по крайней мере на единицу больше r, т.е.

$$d_{\min} \geq r + 1. \tag{58}$$

Для исправления ошибок кратности s минимальное хэммингово расстояние между разрешенными комбинациями должно удовлетворять соотношению

$$d_{\min} \geq 2s + 1. \tag{59}$$

Для исправления всех ошибок кратности s и одновременного обнаружения всех ошибок кратности r минимальное хэммингово расстояние нужно выбирать из условия

$$d_{\min} \geq r + s + 1. \tag{60}$$

Формулы (58)–(60) рекомендуется использовать в случае взаимно независимых ошибок. Если имеются какие-либо корреляции помехи и сигнала, то эти оценки являются завышенными. Наиболее распространенными коррелированными искажениями являются ошибки, связанные с тем, что длительность импульсов помехи превышает длительность символа. Это означает, что один импульс помехи искажает сразу несколько подряд идущих кодовых символов. Ошибки такого рода получили название пачек ошибок или пакетов ошибок. Длиной пачки ошибок называется число следующих друг за другом символов, левее и правее которых в кодовой комбинации искаженных символов не содержится. Таким образом, для пачек ошибок и асимметричного канала при той же корректирующей способности минимальное хэммингово расстояние между разрешенными комбинациями может быть меньше.

Подчеркнем, что каждый конкретный корректирующий код не гарантирует исправления любой комбинации ошибок. Коды предназначены для исправления комбинаций ошибок, наиболее вероятных для заданного канала.

Если характер и уровень помехи будет отличаться от предполагаемых, то эффективность применения кода резко снизится. Применение корректирующего кода не может гарантировать безошибочность приема, но дает возможность повысить вероятность получения на выходе правильного результата.

2.6. Коды для обнаружения одиночных ошибок

Среди кодов, обладающих свойствами обнаружения ошибок, можно привести следующие [3, 8, 9].

1. Код с контролем на четность (нечетность). Мощность кода с контролем на четность (нечетность) определяется по формуле: $N = 2^{n-1}$. Построение данного кода осуществляется путем добавления одного контрольного разряда (справа) к простому коду, поэтому длина кода $n = m + 1$, где m – число информационных разрядов. Символы контрольного разряда выбираются так, чтобы число единиц в получаемой кодовой комбинации было четным (нечетным). Так как место в кодовой комбинации контрольного разряда известно, то данный код называется разделимым. Данный код обнаруживает ошибки нечетной кратности.

2. Код с постоянным весом. Этот код предполагает формирование кодовых комбинаций таким образом, чтобы количество единиц было во всех кодовых комбинациях одинаковым. В самом простом случае кодовые комбинации в коде на одно сочетание будут содержать одну единицу, при этом мощность кода равна $N = C_n^1$. Этот код способен обнаруживать ошибки нечетной кратности. Так как место в кодовой комбинации контрольного разряда неизвестно, то данный код относится к неразделимым.

Пример 32. Пусть длина кодовых сообщений $n = 5$, $N = C_5^1 = 5$.

Множество кодовых комбинаций {00001, 00010, 00100, 01000, 10000}.

Этот код еще называется распределительным в телемеханических устройствах.

Заметим, что коды по законам комбинаторики, рассмотренные ранее, также обладают свойствами помехоустойчивости. Например, коды по законам сочетаний C_n^2, C_n^3 и т.д.

Пример 33. Пусть длина кодовых сообщений $n = 5$, построим код на сочетания из 5 по 2, мощность кода $N = C_5^2 = 10$.

Множество кодовых комбинаций {00011, 00101, 00110, 01001, 01010, 01100, 10001, 10010, 10100, 11000}.

3. Корреляционный код**.** При формировании кодовых сообщений в данном коде происходит увеличение длины передаваемых сообщений вдвое, поскольку правило построения состоит в замене одного исходного сообщения на пару символов 01 и 0 – на 10. Например, комбинация простого кода 001 в корреляционном коде примет вид 101001. Такой принцип построения приводит к тому, что в кодовой комбинации не может быть

больше трех рядом стоящих одинаковых символов, благодаря чему код способен обнаруживать все одиночные ошибки, а также ошибки двойной кратности, не связанные с трансформацией элементов, т.е. $1\to11$ и $0\to00$.

4. Код с инверсным дополнением. Код с инверсным дополнением строится по правилу дополнения к исходной комбинации простого кода инверсной последовательности, например, $1001 \to 10010110$. Код позволяет обнаруживать практически все ошибки в комбинации. Ошибки не будут обнаружены лишь тогда, когда одновременно исказятся два, четыре и т.д. элемента в исходной комбинации и соответствующие два, четыре и т.д. элемента инверсной части кодовой комбинации. Это так называемые ошибки, связанные с равнопозиционностью, например, $10\boxed{0}101\boxed{1}0$.

Из рассмотренных кодов инверсный код обладает наибольшей помехоустойчивостью.

Все коды, обнаруживающие одиночные ошибки, имеют кодовое расстояние $d = 1$.

2.7. Построение группового кода

Самый большой класс разделимых кодов составляют систематические коды, у которых значения проверочных символов определяются в результате проведения линейных операций над определенными информационными символами. Для случая двоичных кодов каждый проверочный символ выбирается таким, чтобы его сумма с определенными информационными символами стала равной нулю. Символ проверочной позиции имеет значение 1, если число единиц информационных разрядов, входящих в данное проверочное равенство, нечетно, и нулю, если оно четно. Число проверочных равенств (а следовательно, и число проверочных символов) и номера конкретных информационных разрядов, входящих в каждое из равенств, определяются тем, какие и сколько ошибок должен исправлять или обнаруживать данный код. Проверочные символы могут располагаться на любом месте кодовой комбинации.

При декодировании определяется справедливость проверочных равенств. В случае двоичных кодов такое определение сводится к проверкам на четность числа единиц среди символов, входящих в каждое из равенств (включая проверочный). Совокупность проверок дает информацию о том, имеется ли ошибка, а в случае необходимости и о том, на каких позициях символы искажены.

Любой двоичный систематический код является групповым кодом, так как совокупность входящих в него кодовых комбинаций образует группу.

Построение конкретного корректирующего кода производится исходя из требуемого объема кода Q, т.е. необходимого числа передаваемых команд (или дискретных значений измеряемой величины) и статистических данных о наиболее вероятных векторах ошибок в используемом канале связи. Вектором ошибки будем называть кодовую комбинацию, имеющую единицы в разрядах, подвергшихся искажению, и нули во всех остальных разрядах. Любую искаженную кодовую комбинацию можно рассматривать теперь как сумму (или разность) по модулю 2 разрешенной кодовой комбинации и вектора ошибки.

При рассмотрении систематических кодов вместо термина «кодовая комбинация» часть употребляют термин «кодовый вектор». Вектор, состоящий из нулей, называется нулевым вектором. Число единиц в кодовом векторе называется его весом. Таким образом, расстояние любого вектора от нулевого измеряется его весом.

Систематическим называется n-значный код, состоящий из $N = 2^m$ кодовых векторов. Из n символов, образующих кодовый вектор, m символов являются информационными, а остальные $k = n - m$ избыточными [3, 8, 9].

Наличие избыточных (контрольных, проверочных) символов определяет корректирующую способность кода. Во всех кодовых векторах систематического кода проверочные символы занимают одни и те же позиции. Для систематического кода применяется обозначение (n, m)-код.

В систематическом коде значения проверочных символов подбираются так, чтобы сумма по модулю 2 всех символов (включая проверочный), входящих в каждое из равенств, равнялась нулю. В таком случае число единиц среди этих символов четное. Поэтому операции определения символов опознавателя называют проверками на четность. При отсутствии ошибок в результате всех проверок на четность образуется опознаватель, состоящий из одних нулей. Если проверочное равенство не удовлетворяется, то в соответствующем разряде опознавателя появляется единица. Исправление ошибок возможно лишь при наличии взаимно однозначного соответствия между множеством опознавателей и множеством смежных классов, а следовательно, и множеством подлежащих исправлению векторов ошибок.

Таким образом, количество подлежащих исправлению ошибок является определяющим для выбора числа избыточных символов $k = n - m$. Их должно быть достаточно для того, чтобы обеспечить необходимое число опознавателей.

Опознавателем называется некоторая контрольная последовательность символов, которая ставится в однозначное соответствие каждой разрешенной группе кодовых комбинаций.

Исходя из требуемых корректирующих свойств проектируемого кода, определяют кодовое расстояние d по формулам (58)–(60). Зная d, на следующем этапе определяют число контрольных элементов k по формулам Хэмминга [3, 8, 9]:

$$k \ge \log_2 \sum_{i=0}^{\frac{d-1}{2}} C_n^i, \; d - \textit{нечетное число;} \tag{61}$$

$$k \ge 1 + \log_2 \sum_{i=0}^{\frac{d-2}{2}} C_{n-1}^i, \; d - \textit{четное число.} \tag{62}$$

Обратим внимание на особенность приведенных формул (61)–(62) – неизвестная величина k содержится в левой и правой части соотношения, в правой части в неявном виде, поскольку $n = m + k$. Выразить ее в явном виде не представляется возможным, поэтому вычисление по формулам (61)–(62) необходимо производить итерационно, последовательно увеличивая значение k до тех пор, пока неравенство не станет справедливым. Естественно выбирать наименьшее число контрольных символов k, обеспечивающих заданные корректирующие свойства проектируемого кода, чтобы не увеличивать общую длину кодовой комбинации.

Зная основные параметры кода m, n и k, можно перечислить все разрешенные кодовые комбинации. Если последовательно располагать их друг под другом, то получим матрицу, насчитывающую n столбцов и $2^m - 1$ строк. Например,

$$\begin{array}{cccccccc} a_5 & a_8 & a_1 & a_2 & a_3 & a_4 & a_6 & a_7 \end{array}$$

$$\left\|\begin{array}{cccccccc} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right\|.$$

Очевидно, что при больших n и m матрица оказывается слишком громоздкой, поэтому для сокращения записи целесообразно использовать одно важное свойство группового кода. Исходная матрица, содержащая все кодовые комбинации кода, будет состоять из строк, которые будут линейно зависимыми. Для полного же определения кода достаточно записать только линейно независимые строки. Среди $2^m - 1$ ненулевых комбинаций кода их только m. Действительно, если рассматривать m информационных разрядов корректирующего (n, m) кода, как m-разрядный неизбыточный код, то все $2^m - 1$ ненулевых комбинаций этого кода можно записать компактно, приняв за линейно независимые строки единичной матрицы.

Суммируя строки единичной матрицы в различных сочетаниях, можно получить любую из $2^m - 1$ ненулевых комбинаций m-разрядного неизбыточного кода. Поэтому эту матрицу называют образующей (порождающей, производящей) по отношению к данному коду.

Комбинацию β группового кода запишем в виде последовательности $\beta = b_1b_2,\ldots,b_m,c_1,c_2,\ldots,c_k$, где $b_1b_2,\ldots,b_m$ – информационные разряды, а $c_1,c_2,\ldots,c_k$ – контрольные разряды.

По определению для групповых кодов $\beta_i \oplus \beta_j$, также комбинация группового кода, т.е. результат поразрядного суммирования комбинации β_i с комбинацией β_j даст другую комбинацию группового кода. Из этого следует, что кодовое расстояние группового кода определяется $(d(\beta_i\beta_j) = \omega\beta_i \oplus \beta_j)$ весом кодовой комбинации с минимальным числом единиц.

Именно благодаря этому свойству замкнутости группового кода возможно упростить его описание. Можно задать групповой код, указывая не все кодовые комбинации, а только их часть, полагая, что остальные кодовые комбинации могут быть определены через них. Это происходит следующим образом.

Совокупность β_1, β_2, …, β_n кодовых комбинаций называется линейно зависимой, если существует набор элементов α_l, α_2, …, α_n $(\alpha_i \in \{0,1\})$, среди которых хотя бы один отличен от нуля, и выполняется условие $\alpha_1\beta_1\oplus\alpha_2\beta_2\oplus\ldots\oplus\alpha_n\beta_n = 0$.

Если это равенство возможно при всех α_i=0, то кодовые комбинации $\beta_1, \beta_2, \ldots, \beta_n$ называются линейно независимыми.

Если среди 2^n кодовых комбинаций группового кода выбрано m линейно независимых кодовых комбинаций $\beta_1, \beta_2, \ldots, \beta_m$, то для любого набора $\boldsymbol{\alpha_1, \alpha_2, \ldots, \alpha_m}$ (одновременно не равных нулю) получим комбинацию группового кода по правилу [3, 8, 9]

$$\beta_r = \alpha_1\beta_1\oplus\alpha_2\beta_2\oplus\ldots\oplus\alpha_m\beta_m\neq 0. \qquad (63)$$

Составляя всевозможные наборы элементов $\alpha_1, \alpha_2, \ldots, \alpha_m$, можно получить 2^m кодовых комбинаций группового кода по правилу (62).

Любой набор из m линейно независимых кодовых комбинаций порождает групповой код (n,m). Такой набор записывается в виде матрицы $G_{m,n}$, которая называется образующей или порождающей. Эта матрица содержит информацию о способе построения кода.

Полное множество разрешенных комбинаций каждого из кодов можно получить, суммируя по модулю 2 строки порождающей матрицы во всех возможных сочетаниях.

Матрицу $G_{m,n}$ можно привести к канонической форме $G_{m,n}=|I_k,R_{m,k}|$, где I_m – единичная матрица; $R_{m,k}$ – матрица контрольных элементов.

Образующая матрица в каноническом виде имеет блочную структуру и состоит из двух матриц [3, 8, 9]:

1) единичной транспонированной в канонической форме I_m^T, которая соответствует m информационным разрядам;

2) дописываемой справа дополнительной матрицы $C_{n-m,m}$, которая соответствует $n-m$ проверочным разрядам.

$$G_{m,n} = \begin{vmatrix} 1 & 0 & \ldots & 0 & q_{11} & q_{12} & \ldots & q_{1k} \\ 0 & 1 & \ldots & 0 & q_{21} & q_{22} & \ldots & q_{2k} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 & q_{m1} & q_{m2} & \ldots & q_{mk} \end{vmatrix}.$$

Если $B_{1,m}$ – матрица-строка безызбыточного двоичного кода, то кодовая комбинация группового кода определится в виде произведения

$$\beta = B_{1,m} \times G_{m,n} = \| b_1 b_2, \ldots, b_m, c_1, c_2, \ldots, c_k \|, \tag{64}$$

причем контрольные элементы определятся по формуле

$$c_j = \sum_{i=1}^{m} b_i q_{ij} \,. \tag{65}$$

Из формулы (64) следует, что первые m элементов комбинации β группового кода определяются комбинацией безызбыточного кода, а остальные k = n – m элементов определяются как комбинацией безызбыточного кода, так и элементами матрицы $R_{m,k}$. Поэтому первые m элементов комбинации β группового кода называются информационными элементами, а остальные k – контрольными.

В зависимости от выбранного базиса k-мерного подпространства n-мерного кодового пространства кодовое расстояние совокупности 2^k векторов k-мерного подпространства может быть различным. При проектировании (n, k)-кода ставится задача оптимального размещения кодовых векторов в n-мерном кодовом пространстве в соответствии с заданной статистикой ошибок и, в частности, обеспечения максимально возможного кодового расстояния.

При проектировании группового кода матрицу $R_{m,k}$ контрольных элементов формируют исходя из следующих условий [3, 8, 9]:

а) вес каждой строки должен быть не менее d – 1;

б) две любые строки должны отличаться друг от друга не менее, чем в d – 2 разрядах.

Уравнение (65) задает преобразование m-разрядной кодовой комбинации безызбыточного кода в n-разрядную кодовую комбинацию группового кода. Так как контрольные символы получаем в результате линейных операций над информационными элементами, то групповой код называют еще линейным.

Пример 34. Задана образующая матрица

$$G_{3,5} = \begin{vmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{vmatrix}.$$

Требуется определить контрольные символы для сообщения $B_{1,3}$ = 001.

Решение. Используем соотношение (65) для определения контрольных элементов:

$$c_1=0\times1\oplus0\times1\oplus1\times0=0,$$
$$c_2=0\times0\oplus0\times0\oplus1\times1=1,$$

следовательно, кодовая комбинация группового кода β = 00101.

Обнаружение ошибок

В безызбыточных равномерных кодах длины k все 2^k возможных кодовых комбинаций используются, т.е. любой из 2^k кодовой комбинации сопоставляется какой-либо символ внешнего алфавита. Такие коды получили название *первичных кодов.* Ошибка любой кратности в какой-либо кодовой комбинации всегда приводит к ошибочному декодированию этой кодовой комбинации.

Обнаружение ошибок помехоустойчивым кодом возможно благодаря тому, что для передачи информации используются не все 2^n n-разрядные кодовые комбинации равномерного кода, а лишь часть из них. Для разделимых кодов эта часть составляет 2^k кодовых комбинаций, получивших название *разрешенных кодовых* комбинаций. Оставшаяся часть $2^n - 2^k$ кодовых комбинаций, составляющая *запрещенные кодовые комбинации,* при передаче информации не применяется.

Использование при кодировании символов внешнего алфавита лишь части кодовых комбинаций позволяет разнести разрешенные кодовые комбинации в кодовом пространстве на расстояние, превышающее 1. Нетрудно видеть, что если расстояние $d > 1$, то все одиночные ошибки будут переводить разрешенные кодовые комбинации в запрещенные, а появление запрещенной кодовой комбинации на приемной стороне может служить индикатором того, что произошла ошибка [3, 8, 9].

При разработке реальных кодов учитываются статистика ошибок и требование верности передачи информации. *Верность* передачи оценивается часто как среднее число верно принятых кодовых комбинаций, приходящихся на одну ошибочно принятую кодовую комбинацию, или как вероятность верного приема кодовой комбинации. Так, при выполнении статистической гипотезы о том, что ошибки меньшей кратности появляются чаще ошибок большей кратности, исходя из требования верности передачи, определяют максимальную кратность ошибки, начиная с которой все ошибки меньшей кратности должен обнаруживать помехоустойчивый код. По максимальной кратности ошибки q_m выбирают такое минимальное кодовое расстояние, при котором все разрешенные кодовые комбинации при действии на них ошибок кратностью, не превышающей q_m, переходят в подмножество запрещенных кодовых комбинаций и, следовательно, могут быть обнаружены на приемной стороне системы передачи информации.

Результатом действия ошибки кратности q на разрешенную кодовую комбинацию является новая кодовая комбинация, удаленная от первоначальной на расстояние q. Отсюда ясно, что если кодовое расстояние $d \le q$, то при действии ошибки кратности q на какую-либо разрешенную кодовую комбинацию последняя может перейти в другую, но тоже разрешенную, кодовую комбинацию и такая ошибка уже не может быть обнаружена. Поэтому для обнаружения всех ошибок, кратность которых не превышает q, кодовое расстояние должно быть больше q ($d > q$). Для обнаружения всех ошибок кратности, не превышающей q_m, кодовое расстояние должно на единицу превышать максимальную кратность ошибки: $d = q_m + 1$.

Наличие в принятой кодовой комбинации возможно на основе проверочной матрицы, которую можно получить в результате преобразования соотношения (65) к виду

$$\sum_{i=1}^{m} b_i q_{ij} \oplus c_j = 0, \quad \mathrm{j} = \overline{1,\mathrm{k}} \cdot \tag{66}$$

Соотношениям (65) и (66) должны удовлетворять все символы кодовых комбинаций, поэтому эти соотношения называют проверочными.

Если записать правило (66) формирования каждого контрольного элемента в виде последовательностей из нулей и единиц, где единицы на позициях, соответствующих информационным элементам, указывают, какие информационные разряды участвуют в образовании того контрольного элемента, на позиции которого в последовательности стоит единица, то получим k последовательностей. Запишем эти последовательности в прямоугольную таблицу, размерности $k \times n$, называемую контрольной или проверочной матрицей:

$$H_{k,n} = \begin{vmatrix} q_{11} & q_{21} & \dots & q_{m1} & 1 & 0 & \dots & 0 \\ q_{12} & q_{22} & \dots & q_{m2} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ q_{1k} & q_{2k} & \dots & q_{mk} & 0 & 0 & \dots & 1 \end{vmatrix} = \left| R_{k,m}, I_k \right|, \quad R_{k,m} = R_{m,k}^{T} \cdot$$

В первой строке матрицы $H_{k,n}$ записано уравнение для формирования первого контрольного элемента

$$b_1q_{11}\oplus b_2q_{21}\oplus b_3q_{31}\oplus\ldots\oplus b_mq_{m1}\oplus c_1=0.$$

Во второй строке матрицы $H_{k,n}$ записано уравнение для формирования второго контрольного элемента

$$b_1q_{12}\oplus b_2q_{22}\oplus b_3q_{32}\oplus\ldots\oplus b_mq_{m2}\oplus c_2=0$$

и т.д.

В k-й строке матрицы $H_{k,n}$ записано уравнение для формирования k-го контрольного элемента

$$b_1q_{1k}\oplus b_2q_{2k}\oplus b_3q_{3k}\oplus\ldots\oplus b_mq_{mk}\oplus c_k=0.$$

Пример 35. Пусть код (6, 3) задан в виде образующей матрицы

$$G_{3,6}=\begin{vmatrix}1&0&0&0&1&1\\0&1&0&1&0&1\\0&0&1&1&1&0\end{vmatrix}.$$

Построить проверочную матрицу.

Решение. Матрица $\mathbf{R_{k,m}=R^T_{m,k}}$ имеет вид

$$R_{3,3}=\begin{vmatrix}0&1&1\\1&0&1\\1&1&0\end{vmatrix}.$$

Проверочная матрица также имеет блочную структуру. В качестве первого блока запишем транспонированную проверочную матрицу, вторая матрица представляет собой единичную диагональную матрицу

$$H_{3,6}=\begin{vmatrix}0&1&1&1&0&0\\1&0&1&0&1&0\\1&1&0&0&0&1\end{vmatrix}.$$

Уравнения формирования контрольных элементов запишем на основании формулы (65):

$$b_2\oplus b_3\oplus c_1=0,\ b_1\oplus b_3\oplus c_2=0,\ b_1\oplus b_2\oplus c_3=0.$$

Если информационная последовательность имеет вид 101, то $\mathbf{c_1}=1$, $\mathbf{c_2}=0$, $\mathbf{c_3}=1$, а кодовая комбинация – 101101.

Таким образом, задание проверочной матрицы $H_{k,n}$ является одним из способов описания группового кода (можно формировать кодовые комбинации).

Образующая и проверочная матрицы связаны проверочным соотношением

$$G_{m,n} \times H^T_{k,n} = 0 .$$

Если принимаемая кодовая комбинация $\boldsymbol{\beta}^*$ принадлежит кодовому множеству, то для нее выполняется соотношение (65), а матричное произведение

$$\beta^* \times H^T_{k,n} = B_{1,m} \times G_{m,n} \times H^T_{k,n} = 0 \cdot \qquad (67)$$

Выполнение условия (67) свидетельствует об отсутствии ошибки в принятой кодовой комбинации β^*.

Если при передаче возникла ошибка **e**, то $\beta^*=\beta\oplus e$ и тогда

$$\beta^* \times H^T_{k,n} = (\beta \oplus e) \times H^T_{k,n} = \beta \times H^T_{k,n} \oplus e \times H^T_{k,n} = e \times H^T_{k,n} \cdot \qquad (68)$$

Двоичная последовательность $\mathbf{D = e \times H^T_{k,n}}$ называется *опознавателем* (*корректором* или *синдромом*) ошибки. Опознаватель ошибки представляет собой k-разрядное двоичное число $d_1 d_2, \ldots, d_k$. Опознаватель ошибки можно также получить, если применить к принятой кодовой комбинации β^* систему проверочных соотношений (66):

$$d_j = \sum_{i=1}^{m} b_i q_{ij} \oplus c_j , \quad \mathrm{j} = \overline{1,\mathrm{k}} \cdot \qquad (69)$$

Если все элементы $d_i = 0$ (D = 0), то в принятой кодовой комбинации ошибок нет.

Пример 36. Групповой (6, 3)-код задан проверочной матрицей

$$H_{3,6} = \begin{vmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{vmatrix} .$$

Определить, имелись ли ошибки при передаче кодовых сообщений $\boldsymbol{\beta}^*$=110110 и $\boldsymbol{\beta}^*$=110010.

Решение. Уравнения формирования контрольных элементов:

$$b_2 \oplus b_3 \oplus c_1 = 0,\ b_1 \oplus b_3 \oplus c_2 = 0,\ b_1 \oplus b_2 \oplus c_3 = 0.$$

Результаты проверок сообщения $\boldsymbol{\beta}^*$=110110 следующие:

$$d_1=1\oplus0\oplus1=0,\ d_2=1\oplus0\oplus1=0,\ d_3=1\oplus1\oplus0=0.$$

Аналогичные результаты можно получить в результате умножения

$$\beta^*\times H_{k,n}^T=\beta^*\times H_{3,6}^T=\begin{vmatrix}1\times0\oplus1\times1\oplus0\times1\oplus1\times1\oplus1\times0\oplus0\times0\\1\times1\oplus1\times0\oplus0\times1\oplus1\times0\oplus1\times1\oplus0\times0\\1\times1\oplus1\times1\oplus0\times0\oplus1\times0\oplus1\times0\oplus0\times1\end{vmatrix}=(000)\cdot$$

Результаты проверок сообщения $\boldsymbol{\beta}^*$=110010 следующие:

$$d_1=1\oplus0\oplus0=1,\ d_2=1\oplus0\oplus1=0,\ d_3=1\oplus1\oplus0=0.$$

Таким образом, **D**=100 фиксирует наличие ошибки в принятой кодовой комбинации.

Условия обнаружения и исправления ошибок

Для исправления ошибок необходимо, чтобы различным ошибкам соответствовали различные значения синдрома, т.е. [3, 8, 9]

$$e_i\times H_{k_1n}^T\neq e_j\times H_{k_1n}^T\cdot$$

Отметим еще одно свойство проверочной матрицы. Групповой код имеет минимальное кодовое расстояние d, если любые d – 1 или менее столбцов проверочной матрицы линейно независимы. Произведение (68) представим в виде

$$e\times H_{k_1n}^T=e_1\times h_1\oplus e_2\times h_2\oplus\ldots.\oplus e_n\times h_n, \tag{70}$$

где e_i – компонент вектора ошибки; h_i – i-й столбец проверочной матрицы; $e_i\times h_i$ – произведение скаляра e_i на матрицу-строку h_i.

Следовательно, для обнаружения ошибок кратности d – 1 и менее необходимо и достаточно, чтобы d – 1 или менее столбцов проверочной матрицы были линейно независимыми.

Из условия (70) следует, что опознаватель ошибок можно получить поразрядным сложением тех столбцов h_i проверочной матрицы, которым соответствуют единицы на позициях комбинации ошибок.

Поскольку опознаватель одиночных ошибок содержит только одну единицу, то состав столбцов проверочной матрицы приобретает следующий смысл. В столбцах h_i проверочной матрицы записаны опознаватели одиночных ошибок, имеющих место в i-м разряде кодовой комбинации.

Применяя условия (68), (69) и (70) можно построить все возможные формы опознавателей обнаруживаемых и исправляемых кодом ошибок, которые сводятся в таблицу, называемую таблицей декодирования.

Поскольку получателю выдаются информационные разряды, то исправление ошибок в контрольных разрядах не производится.

Показатели качества корректирующего кода

Одной из основных характеристик корректирующего кода является избыточность кода, указывающая степень удлинения кодовой комбинации для достижения определенной корректирующей способности. Если на каждые n символов выходной последовательности кодера канала приходится k информационных и $n - k$ проверочных, то относительная избыточность кода может быть выражена одним из соотношений [3, 8, 9]:

$$R_n = \frac{n-k}{n} \tag{71}$$

или

$$R_k = \frac{n-k}{k}. \tag{72}$$

Величина R_k предпочтительнее, так как лучше отвечает смыслу понятия избыточности.

Коды, обеспечивающие заданную корректирующую способность при минимально возможной избыточности, называются *оптимальными* (иногда – *плотноупакованными*).

Однако не всегда целесообразно стремиться к использованию кодов, близких к оптимальным. Необходимо учитывать другой, не менее важный показатель качества корректирующего кода – сложность технической реализации процессов кодирования и декодирования.

Если информация должна передаваться по медленно действующей, ненадежной и дорогостоящей линии связи, а кодирующее и декодирующее устройства предполагается выполнить на высоконадежных и быстродействующих элементах, то сложность этих устройств не играет существенной роли. Решающим фактором в таком случае является повышение эффективности использования линии связи и поэтому желательно применение корректирующих кодов с минимальной избыточностью.

Если же корректирующий код должен быть применен в системе, выполненной на элементах, надежность и быстродействие которых равны или близки надежности и быстродействию элементов кодирующей и декодирующей аппаратуры (например, для повышения достоверности воспроизведения информации с запоминающего устройства цифровой вычислительной машины), то критерием качества корректирующего кода является надежность системы в целом, т.е. с учетом возможных искажений и отказов в устройствах кодирования и декодирования. В этом случае более целесообразными часто оказываются коды с большей избыточностью, но обладающие преимуществом простоты технической реализации.

Приведем несколько практических примеров проектирования групповых кодов и решения частных задач, возникающих при их использовании.

Пример 37. Определить эффективность 5-разрядного кода с проверкой на четность при условии, что справедлива модель ошибки $P_{5,q} = C_q^5 0{,}1^q \cdot 0{,}9^{5-q}$.

Решение. Используем модель ошибки $P_{5,q} = C_5^q 0{,}1^q 0{,}9^{5-q}$. Вычисления по данной формуле сведем в таблицу.

q	0	1	2	3	4	5
$P_{5,q}$	0,59	0,33	0,07	0,008	0,0004	0,00001

Из таблицы видно, что вероятность появления ошибок большой кратности мала. Наиболее часто появляются ошибки кратности 1.

Вероятность ошибочного приема кодовой комбинации определим по формуле $P_{\text{ош}} = 1 - P_{5,q} = 1 - 0{,}56 = 0{,}44$.

По таблице находим, что вероятность одиночных обнаруживаемых ошибок $P_{5,1} = 0{,}33$. Кроме того, обнаруживаются все ошибки нечетной кратности. Таким образом, вероятность появления необнаруженной ошибки $P_{\text{н.ош}} = P_{5,2} + P_{5,4} = 0{,}07+0{,}0004 = 0{,}0704$.

Пользуясь полученными данными, определим коэффициент повышения верности $K_{\text{пв}} = 0{,}44/0{,}0704 = 6{,}25$. Результат показывает, что при данной статистике помех проверка на четность позволяет в 6,25 раз уменьшить число необнаруживаемых ошибок.

Пример 38. Привести образующую матрицу к каноническому виду.

$$G_{4,7} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

Решение. Используем свойство группового кода, заключающееся в том, что если складывать строки матрицы по 2, 3,…, то вновь будем получать кодовые комбинации, принадлежащие этому коду. Проделывая эту процедуру многократно с исходной матрицей, получим

$$G_{4,7} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Пример 39. Пусть задана образующая матрица

$$G_{4,7} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

и кодовая комбинация β = 0110. Определить контрольные элементы и сформировать кодовую комбинацию группового кода. Построить проверочную матрицу.

Решение. Определяем контрольные элементы, используя подматрицу контрольных элементов:

$$c_1 = 0\times1 \oplus 1\times0 \oplus 1\times1 \oplus 0\times1 = 1,$$
$$c_2 = 0\times0 \oplus 1\times1 \oplus 1\times1 \oplus 0\times1 = 0,$$
$$c_3 = 0\times1 \oplus 1\times1 \oplus 1\times1 \oplus 0\times0 = 0.$$

Тогда кодовая комбинация группового кода β = 0110100.

Поскольку $H_{m,n} = | Q^T_{m,n}, I \,|$, то в нашем случае проверочная матрица имеет вид

$$H_{4,7} = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Пример 40. Построить корректирующий код, который будет передавать 34 сообщения и обнаруживать две ошибки. Закодировать в спроектированном коде сообщение, соответствующее десятичному числу 12.

Решение. Число информационных символов кода определим по формуле

$$m = \left]\log_2 34\right[= 6.$$

Поскольку код должен обладать свойствами только обнаружения двух ошибок (т.е. $r = 2$), то определения кодового расстояния используем формулу (61)

$$d_{\min} \geq r + 1 = 2 + 3 = 3.$$

$d = 3$ – нечетное, поэтому для определения числа контрольных символов используем формулу (61). Поиск осуществляем в соответствии с итерационной процедурой, начиная с $k = 1$:

– при $k = 1$, $n = m + k = 6 + 1 = 7$:

$$1 \geq \log_2 \sum_{i=0}^{\frac{3-1}{2}} C_7^i = \log_2 \sum_{i=0}^{1} C_7^i = \log_2\left(C_7^0 + C_7^1\right) = \log_2(1+7) = 3,$$

$1 \geq 3$ – неравенство не выполняется;

– при $k = 2$, $n = m + k = 6 + 2 = 8$:

$$2 \geq \log_2 \sum_{i=0}^{\frac{3-1}{2}} C_8^i = \log_2 \sum_{i=0}^{1} C_7^i = \log_2\left(C_8^0 + C_8^1\right) = \log_2(1+8) = 3{,}17,$$

$2 \geq 3{,}17$ – неравенство не выполняется;

– при $k = 3$, $n = m + k = 6 + 3 = 9$:

$$3 \geq \log_2 \sum_{i=0}^{\frac{3-1}{2}} C_9^i = \log_2 \sum_{i=0}^{1} C_9^i = \log_2\left(C_9^0 + C_9^1\right) = \log_2(1+9) = 3{,}32,$$

$3 \geq 3{,}32$ – неравенство не выполняется;

– при $k = 4$, $n = m + k = 6 + 4 = 10$:

$$4 \geq \log_2 \sum_{i=0}^{\frac{3-1}{2}} C_9^i = \log_2 \sum_{i=0}^{1} C_{10}^i = \log_2\left(C_{10}^0 + C_{10}^1\right) = \log_2(1+10) = 3{,}46,$$

$4 \geq 3{,}46$ – неравенство выполняется.

Итерационную процедуру прекращаем, следовательно, число контрольных символов $k = 4$. Образующая матрица должна содержать 6 строк (по числу информационных символов) и 10 столбцов (по общему количеству символов передаваемого сообщения $n = m + k = 6 + 4 = 10$). Матрицу запишем в каноническом виде, она будет содержать единичную квадратную матрицу размерности 6×6 и подматрицу контрольных элементов размерности 6×4. Построим шестнадцать комбинаций простого кода ($2^k = 2^4 = 16$), выберем любые шесть, удовлетворяющие условиям построения матрицы $R_{6,4}$ контрольных элементов, а именно вес кодовой комбинации должен быть не менее чем $d - 1 = 3 - 1 = 2$, а отличаться между собой кодовые комбинации должны не менее чем в $d - 2 = 3 - 2 = 1$ разрядах.

Образующая матрица группового кода (10, 6) имеет следующий вид:

$$G_{6,10} = \begin{vmatrix} 100000 & 0011 \\ 010000 & 1101 \\ 001000 & 0101 \\ 000100 & 0110 \\ 000010 & 1001 \\ 000001 & 1010 \end{vmatrix}.$$

Образующая матрица полностью задает групповой код.

Требуется сформировать кодовую комбинацию, соответствующую десятичному числу 12 в разработанном коде. Переведем 12 в двоичный код – это 1100. Поскольку в коде 6 информационных разрядов, то добавим слева нули, т.е. информационная часть кодовой комбинации будет $\beta = 001100$. Контрольные символы сформируем на основании соотношения (65), т.е. перемножим информационные разряды на элементы контрольной подматрицы образующей матрицы:

$\mathbf{c_1}=0\times0\oplus0\times1\oplus1\times0\oplus1\times0\oplus0\times1\oplus0\times1= 0$,

$\mathbf{c_2}=0\times0\oplus0\times1\oplus1\times1\oplus1\times1\oplus0\times0\oplus0\times0= 0$,

$\mathbf{c_3}=0\times1\oplus0\times0\oplus1\times0\oplus1\times1\oplus0\times0\oplus0\times1= 1$,

$\mathbf{c_4}=0\times1\oplus0\times1\oplus1\times1\oplus1\times0\oplus0\times1\oplus0\times0= 1$.

В разработанном групповом коде окончательно кодовая комбинация для передачи по каналу связи будет следующая: $\beta^* = 0011000011$.

Пример 41. Построить матрицу группового кода, способного исправлять одиночную ошибку при передаче 16 символов первичного алфавита.

Решение. Поскольку необходимо передавать 16 различных сообщений, соответствующих 16 символам первичного алфавита, поэтому количество информационных разрядов равно $m =]\log_2 16[= 4$. Зная, что код должен исправлять одиночную ошибку, для определения кодового расстояния используем формулу (59)

$$d_{\min} \geq 2s + 1 = 2 \cdot 1 + 1 = 3.$$

d = 3 – нечетное, поэтому для определения числа контрольных символов используем формулу (61). Поиск осуществляем в соответствии с итерационной процедурой, начиная с $k = 1$:

– при $k = 1$, $n = m + k = 4 + 1 = 5$:

$$1 \geq \log_2 \sum_{i=0}^{\frac{3-1}{2}} C_5^i = \log_2 \sum_{i=0}^{1} C_5^i = \log_2\left(C_5^0 + C_5^1\right) = \log_2(1+5) = 2{,}58,$$

$1 \geq 2{,}58$ – неравенство не выполняется;

– при $k = 2$, $n = m + k = 4 + 2 = 6$:

$$2 \geq \log_2 \sum_{i=0}^{1} C_6^i = \log_2\left(C_6^0 + C_6^1\right) = \log_2(1+6) = 2{,}81,$$

$2 \geq 2{,}81$ – неравенство не выполняется;

– при $k = 3$, $n = m + k = 4 + 3 = 7$:

$$3 \geq \log_2 \sum_{i=0}^{1} C_7^i = \log_2\left(C_7^0 + C_7^1\right) = \log_2(1+7) = 3,$$

$3 \geq 3$ – неравенство выполняется.

Итерационную процедуру прекращаем, следовательно, число контрольных символов $k = 6$. Образующая матрица должна содержать 4 строки и 7 столбцов. Матрицу запишем в каноническом виде, в качестве строк контрольной подматрицы будем брать трехзначные двоичные комбинации с весом не менее чем $d - 1 = 3 - 1 = 2$, и отличающихся между собой не менее чем в $d - 2 = 3 - 2 = 1$ разрядах. Благодаря таким параметрам кода можно построить несколько вариантов образующей матрицы:

$$G_{4,7} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{Vmatrix} \text{ или } G_{4,7} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{Vmatrix},$$

или

$$G_{4,7}=\begin{Vmatrix}1&0&0&0&0&1&1\\0&1&0&0&1&0&1\\0&0&1&0&1&1&0\\0&0&0&1&1&1&1\end{Vmatrix}.$$

Пример 42. Постройте корректирующий код (образующую и проверочную матрицу), который будет передавать 29 сообщений, обнаруживать 2 ошибки и исправлять 1 ошибку.

Решение. Определяем число информационных символов кода $m=\left]\log_2 29\right[=5$. Из формулы кодового расстояния определяем, что $d\ge r+s+1=2+1+1=4$. Поскольку d – четное, то для определения числа контрольных символов используем формулу

$$k\ge 1+\log_2\sum_{i=0}^{\frac{d-2}{2}}C_{n-1}^{i}=1+\log_2\sum_{i=0}^{1}C_{n-1}^{i}=1+\log_2\left(C_{n-1}^{0}+C_{n-1}^{1}\right).$$

– при k = 1 – условие

$$1\ge 1+\log_2\left(C_5^0+C_5^1\right)=1+\log_2(1+5)=1+\log_2 6=1+2.58=3.58$$

не выполняется,

– при k = 2 – условие

$$2\ge 1+\log_2\left(C_6^0+C_6^1\right)=1+\log_2(1+6)=1+\log_2 7=3.81$$

не выполняется,

– при k = 3 – условие

$$3\ge 1+\log_2\left(C_7^0+C_7^1\right)=1+\log_2(1+7)=1+\log_2 8=4$$

не выполняется,

– при k = 4 – условие

$$4\ge 1+\log_2\left(C_8^0+C_8^1\right)=1+\log_2(1+8)=1+\log_2 9=4.17$$

не выполняется,

– при k = 5 – условие

$$5\ge 1+\log_2\left(C_9^0+C_9^1\right)=1+\log_2(1+9)=1+\log_2 10=4.32$$

выполняется.

Следовательно, получили код (10, 5). Построим образующую матрицу, в которой подматрица контрольных элементов должна быть выбрана таким образом, чтобы вес каждой строки был не менее d – 1 = = 4 – 1 = 3, а две любые строки отличались друг от друга не менее чем в d – 2 = 4 – 2 = 2 разрядах.

$$G_{5,10} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Пример 43. Построить код для передачи 25 сообщений, который будет обнаруживать и исправлять одну ошибку.

Решение. Поскольку методика построения групповых кодов детально продемонстрирована в примерах 40–42, то приведем в данном случае только результаты, а промежуточные расчеты предлагается выполнить самостоятельно. Таким образом при **M** = 25, $m = 5$, $s = 1$, $r = 1$, $d = 3$. Число контрольных разрядов $k = 4$.

Образующая и проверочная матрицы имеют вид:

$$G_{5,9} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{vmatrix}, \quad H_{4,9} = \begin{vmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1` & 0 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

Пример 44. Сформировать все кодовые комбинации группового кода, заданного производящей матрицей

$$G_{4,7} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{Vmatrix}.$$

Решение. В порождающей матрице четыре строки, следовательно, кодовые комбинации будут содержать $m = 4$ информационных разрядов и $k = 7 - 4 = 3$ контрольных разрядов. При $m = 4$ число возможных информационных комбинаций равно $2^4 = 16$:

1) 0000	5) 0010	9) 0001	13) 0011
2) 1000	6) 1010	10) 1001	14) 1011
3) 0100	7) 0110	11) 0101	15) 0111
4) 1100	8) 1110	12) 1101	16) 1111.

Находим последовательно корректирующие разряды всех информационных комбинаций путем суммирования по модулю 2 тех строк подматрицы контрольных символов, номера которых совпадают с номерами разрядов, содержащих единицы в информационной части кода. Приведенная ниже процедура показывает процесс формирования контрольных символов и иллюстрирует особенности структуры групповых кодов:

1) 000

2) 111

3) 110

4) $\begin{array}{r} \oplus 111 \\ 110 \\ \hline 001 \end{array}$

5) 101

6) $\begin{array}{r} \oplus 111 \\ 101 \\ \hline 010 \end{array}$

7) $\begin{array}{r} \oplus 110 \\ 101 \\ \hline 011 \end{array}$

8) $\begin{array}{r} \oplus 111 \\ \oplus 110 \\ 101 \\ \hline 100 \end{array}$

9) 011

10) $\begin{array}{r} \oplus 111 \\ 011 \\ \hline 100 \end{array}$

11) 110

12) $\begin{array}{r} \oplus 111 \\ 001 \\ \hline 100 \end{array}$

13) $\begin{array}{r} \oplus 101 \\ 011 \\ \hline 110 \end{array}$

14) $\begin{array}{r} \oplus 111 \\ \oplus 101 \\ 011 \\ \hline 011 \end{array}$

15) $\begin{array}{r} \oplus 110 \\ \oplus 101 \\ 011 \\ \hline 000 \end{array}$

16) $\begin{array}{r} \oplus 111 \\ \oplus 110 \\ \oplus 101 \\ 011 \\ \hline 111 \end{array}$

Окончательно все кодовые комбинации заданного группового кода имеют следующий вид:

1) 0000000	5) 0010101	9) 0001011	13) 0011110
2) 1000111	6) 1010010	10) 1001100	14) 1011001
3) 0100110	7) 0110011	11) 0101110	15) 0111000
4) 1100001	8) 1110100	12) 1101100	16) 1111111.

Пример 45. Групповой код задан проверочной матрицей

$$H_{4,9} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Записать уравнения для формирования контрольных элементов. Применить к принятой кодовой комбинации β* систему проверочных соотношений и определить, имеется ли в β* ошибка:

1) β* = 11010 1001; 2) β* = 00111 0011.

Решение. Уравнения для формирования контрольных элементов:

$$b_4 \oplus b_5 \oplus c_1 = 0,$$
$$b_2 \oplus b_3 \oplus b_5 \oplus c_2 = 0,$$
$$b_1 \oplus b_2 \oplus b_3 \oplus c_3 = 0,$$
$$b_1 \oplus b_3 \oplus b_4 \oplus b_5 \oplus c_4 = 0.$$

Результаты проверок следующие:

1) для β* = 11010 1001

$$d_1 = 1 \oplus 0 \oplus 1 = 1,$$
$$d_2 = 1 \oplus 0 \oplus 0 \oplus 0 = 1,$$
$$d_3 = 1 \oplus 1 \oplus 0 \oplus 0 = 0,$$
$$d_4 = 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 = 1;$$

2) для β* = 00111 0011

$$d_1 = 1 \oplus 1 \oplus 0 = 0,$$
$$d_2 = 0 \oplus 1 \oplus 1 \oplus 0 = 0,$$
$$d_3 = 0 \oplus 0 \oplus 1 \oplus 1 = 0,$$
$$d_4 = 0 \oplus 1 \oplus 1 \oplus 1 \oplus 1 = 0.$$

В первом случае D = 1101, код фиксирует наличие ошибки, во втором случае D = 0000 – ошибки нет.

Тот же результат можно получить, если умножить β* на транспонированную проверочную матрицу.

Пример 46. Известно, что в кодовой комбинации β*, принадлежащей групповому коду, заданному проверочной матрицей $H_{2,3}$, содержится ошибка. Определить, достаточно ли корректирующей способности кода, чтобы исправить ошибку.

$$H_{4,9} = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

1) β* = 110011101;

2) β* = 001100111.

Решение. Для этого необходимо умножить принятую кодовую комбинацию на транспонированную проверочную матрицу и проверить совпадение синдрома ошибки со строкой полученной матрицы. Если совпадения есть, то определяем номер разряда, в котором произошла ошибка, если нет, то код не может исправить ошибку.

Транспонируем матрицу
$$H_{4,9}^T = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

1)
$$|110011101| \times \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = |1101|$$
, есть совпадение с 5-й строкой транспонированной матрицы, следовательно, код может исправить ошибку (ошибка произошла в 5-м разряде);

2)
$$|011100101| \times \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = |0101|$$
, совпадений нет, следовательно, код не может исправить такую ошибку.

Пример 47. Групповой (7, 4)-код задан порождающей матрицей:

$$G_{7,4} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{Vmatrix}.$$

Учитывая связь элементов порождающей и проверочной матриц, представленных в канонической форме, имеем

$$H_{7,4} = \begin{Vmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{Vmatrix}.$$

Проиллюстрируем способность разработанного (7, 4)-кода исправлять одиночные ошибки. Пусть передаваемая кодовая комбинация v (7, 4)-кода образована путем сложения первой, второй и четвертой строк матрицы $G_{7,4}$, тогда $v = 1101001$.

Предположим, что при передаче по каналу произошла ошибка в третьем разряде кодовой комбинации v. В этом случае на приемной стороне канала будет получена кодовая комбинация $v' = 1111001$. Умножим принятую кодовую комбинацию на транспонированную проверочную матрицу $H_{7,4}^T$:

$$\begin{Vmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 1 \end{Vmatrix} \times \begin{Vmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Vmatrix} = \begin{Vmatrix} 1 & 1 & 0 \end{Vmatrix}.$$

Отличие от нуля синдрома ошибки свидетельствует о том, что произошла ошибка. Чтобы определить, в каком разряде кодовой комбинации произошла ошибка, проверим, с какой строкой матрицы $H_{7,4}^T$ совпадает синдром. Как видим, он совпадает с третьей строкой. Отсюда следует, что ошибка произошла в третьем разряде кодовой комбинации. В декодирующих устройствах систем передачи информации результат перемножения принятой кодовой комбинации на проверочную матрицу (синдром ошибки) используется для автоматического исправления соответствующего разряда. Однако напомним, что если код не предназначен для обнаружения определенного типа ошибок, то синдром также будет нулевым.

Контрольные вопросы

1. Что следует понимать под кодированием в широком и узком смысле? Приведите примеры кодирования в природе, обществе и технических системах.
2. Назовите основные цели кодирования в радиотехнических системах.
3. Рассмотрите типовую структуру радиотехнической или телекоммуникационной системы. Выделите в ней блоки, структуры и подсистемы, в функции которых входит кодирование информации. Определите, какого рода кодовые преобразования они осуществляют и каковы их особенности?
4. Приведите классификацию известных вам способов кодирования.
5. Поясните суть двоичного позиционного кодирования. Чем обусловлено широкое распространение двоичных кодов?
6. Приведите примеры непозиционных кодов.
7. Для чего предназначены двоично-десятичные коды? Как формируются взвешенные коды?
8. Для каких целей используются рефлексные коды. Поясните правило перехода от кода Грея к двоичным кодам. Приведите пример.
9. Какие коды по законам комбинаторики вам известны? Приведите формулы для определения мощности кодов и правила формирования кодовых комбинаций по законам комбинаторики.
10. Какое кодирование называют эффективным? Какие цели преследуются при эффективном кодировании?
11. Какие коды называются равномерными и неравномерными? В чем их преимущества и недостатки?
12. В чем заключается условие префиксности? Какими преимуществами обладают префиксные коды?
13. Как строятся кодовые деревья? Как по строению кодового дерева определить равномерные, неравномерные и префиксные коды?
14. В чем суть теоремы Шеннона о кодировании при отсутствии помех?
15. Какие показатели характеризуют качество кода?

16. Перечислите принципы построения кодов, близких к эффективным.

17. Сформулируйте основные этапы построения оптимального неравномерного кода по алгоритму Шеннона – Фано.

18. В чем основной недостаток алгоритма Шеннона – Фано?

19. Сформулируйте основные этапы построения оптимального неравномерного кода по алгоритму Хаффмена.

20. Каким образом следует модифицировать алгоритмы Шеннона – Фано и Хаффмена, если необходимо построить код в троичном или четверичном алфавите?

21. Сформулируйте условие Фано. Какое значение имеет его выполнение при проектировании кодов?

22. Как осуществляется декодирование сообщений в префиксном коде?

23. Каким образом осуществляется повышение эффективности кодирования? В чем суть блочного кодирования?

24. Какие коды называются помехоустойчивыми?

25. Что утверждает теорема Шеннона о кодировании при наличии помех?

26. Что такое вектор ошибки? Опишите модель канала связи с помехами.

27. Что такое избыточность кода? Каким образом может быть использована избыточность для обнаружения и исправления ошибок?

28. Что такое кодовое расстояние и как оно определяется? Поясните с использованием геометрической интерпретации.

29. Как связаны корректирующая способность кода и кодовое расстояние? Приведите соотношения для определения кодового расстояния по заданным корректирующим свойствам кода.

30. Каким образом строятся коды для обнаружения одиночных ошибок: код с контролем на четность (нечетность), код с постоянным весом, корреляционный код, код с инверсным дополнением?

31. Какими свойствами обладают групповые коды?

32. Приведите соотношения для определения числа избыточных символов при проектировании групповых кодов.

33. Какое кодовое расстояние должно быть у кодов, способных обнаруживать одиночные ошибки?

34. Охарактеризуйте код с контролем на четность (нечетность), коды с постоянным весом, корреляционный код и код с инверсным дополнением. Приведите примеры и формулы для определения мощности кода в каждом случае.

35. Какой код называется систематическим? Каким требованиям должны удовлетворять исходные комбинации систематического кода?

36. Приведите формулы для определения кодового расстояния для обеспечения заданных свойств проектируемого систематического кода. В чем особенность вычислений по этим формулам?

37. Как строится производящая (образующая) матрица группового кода?

38. Каким образом определяются контрольные символы по информационным символам кодовых комбинаций с помощью образующей матрицы?

39. Сформулируйте правило записи уравнений для формирования контрольных элементов.

40. Как записывается образующая матрица группового кода в каноническом виде?

41. Каким условиям должна удовлетворять подматрица контрольных элементов образующей матрицы группового кода в каноническом виде?

42. Что такое проверочная матрица, для чего она предназначена и каким образом строится?

43. Запишите основные соотношения для обнаружения и исправления ошибок в принятой кодовой комбинации.

44. Что называется *опознавателем* (*корректором* или *синдромом*) ошибки и как он определяется?

45. Какими показателями можно оценивать качество корректирующих кодов? В качестве примера оцените качество кодов, обнаруживающих одиночные ошибки.

Задачи для самостоятельного решения

1. Можно ли построить двоично-десятичный код со следующими весами:

   1.1) 4-2-1-1,
   1.2) 7-4-2-1,
   1.3) 3-5-1-2,
   1.4) 3-4-2-2.

Если нет, то объясните почему, если да, то постройте его.

2. Расшифруйте кодовую комбинацию 000111110011100010100011111100, если код задан кодовым деревом на рис. 18, С.

3. Постройте кодовые деревья для кодов, представленных следующими таблицами соответствий:

3.1)

a	b	c	d	e	f	g
10	110	010	0110	111	0111	000

3.2)

а	б	в	г	д	е
000	001	010	11	10	011

3.3)

%	№	&	@	*	!	#	^	⊗	®
1100	100	1011	1110	1010	1101	1111	011	010	00

3.4)

а	о	и	е	у	ю	ё	я	э	м	н	п	р
0100	1000	0101	1111	0111	1110	0000	1001	1011	0001	1100	0011	1101

4. Переведите число из кода Грея в двоичное, а затем в десятичное:

4.1) 0011011101011100;

4.2) 110011011000100;

4.3) 111110001100000;

4.4) 101110111001100.

5. Представьте десятичные числа 125, 8974, 451 в коде Грея.

6. Постройте безызбыточные коды по законам комбинаторики, если мощность алфавита равна m, а кодовая комбинация должна содержать n символов:

6.1) по закону размещений, $m = 5$, $n = 2$;

6.2) по закону размещений, если $m = 8$, $n = 3$;

6.3) по закону размещений, если $m = 12$, $n = 1$. В чем особенность построенного кода?

6.4) по закону размещений, если $m = 6$, $n = 6$. В чем особенность построенного кода?

6.5) по закону сочетаний, если $m = 6$, $n = 4$;

6.6) по закону сочетаний, если $m = 8$, $n = 3$;

6.7) по закону сочетаний, если $m = 9$, $n = 9$. Целесообразно ли использовать такой код на практике?

6.8) по закону перестановок, если $m = 4$;

6.9) по закону перестановок, если $m = 7$;

6.10) по закону перестановок с повторениями, если $m = 3$, $r_1 = 1$, $r_2 = 1$, $r_3 = 2$;

6.11) по закону перестановок с повторениями, если $m = 2$, $r_1 = 3$, $r_2 = 3$;

6.12) по закону перестановок с повторениями, если $m = 5$, $r_1 = 2$, $r_2 = 1$, $r_3 = 2$, $r_4 = 4$, $r_5 = 2$;

6.13) сменно-качественный код, если $m = 3$, $n = 5$;

6.14) сменно-качественный код, если $m = 5$, $n = 3$;

6.15) сменно-качественный код, если $m = 2$, $n = 6$. В чем особенность построенного кода?

6.16) сменно-качественный код, если $m = 3$, $n = 2$. В чем особенность построенного кода?

В каждом случае определите мощность кода по заданным параметрам и перечислите все возможные кодовые комбинации.

Замечание. Если мощность кода составляет более 20, можно рекомендовать использовать автоматизированные средства или составить алгоритм формирования кода и написать программу на любом языке программирования.

7. Чему равна длина кодовых комбинаций оптимального кода для передачи сообщений, составленных из 16, 32 и 28 равновероятных двоичных символов?

8. Чему равна средняя длина кодового слова оптимального кода для первичного алфавита со следующим распределением вероятностей: $p(z_1) = 0{,}13$, $p(z_2) = 0{,}16$, $p(z_3) = 0{,}02$, $p(z_4) = 0{,}03$, $p(z_5) = 0{,}6$, $p(z_6) = 0{,}01$, $p(z_7) = 0{,}05$?

9. Постройте код по алгоритму Шеннона – Фано, если известны вероятности букв:

9.1) $p(x_1) = 0{,}5$; $p(x_2) = 0{,}18$; $p(x_3) = 0{,}08$; $p(x_4) = 0{,}10$; $p(x_5) = 0{,}07$; $p(x_6) = 0{,}03$; $p(x_7) = 0{,}04$;

9.2) $p(x_1) = 0,3$; $p(x_2) = 0,12$; $p(x_3) = 0,15$; $p(x_4) = 0,10$; $p(x_5) = 0,09$; $p(x_6) = 0,08$; $p(x_7) = 0,16$;

9.3) $p(x_1) = 0,05$; $p(x_2) = 0,13$; $p(x_3) = 0,01$; $p(x_4) = 0,04$; $p(x_5) = 0,11$; $p(x_6) = 0,16$; $p(x_7) = 0,02$; $p(x_8) = 0,09$; $p(x_9) = 0,03$; $p(x_{10}) = 0,12$; $p(x_{11}) =$ $= 0,14$; $p(x_{12}) = 0,04$; $p(x_{13}) = 0,03$; $p(x_{14}) = 0,03$;

9.4) $p(x_1) = 0,08$; $p(x_2) = 0,12$; $p(x_3) = 0,04$; $p(x_4) = 0,1$; $p(x_5) = 0,05$; $p(x_6) = 0,24$; $p(x_7) = 0,06$; $p(x_8) = 0,09$; $p(x_9) = 0,03$; $p(x_{10}) = 0,05$. Определите вероятности появления букв x_{11} и x_{12}, если известно, что они равны.

В каждом случае запишите кодовую таблицу. Оцените избыточность полученного кода.

10. Постройте код по алгоритму Хаффмена, если известны вероятности букв:

10.1) $p(x_1) = 0,45$; $p(x_2) = 0,08$; $p(x_3) = 0,11$; $p(x_4) = 0,09$; $p(x_5) = 0,12$; $p(x_6) = 0,08$; $p(x_7) = 0,07$.

10.2) $p(x_1) = 0,25$; $p(x_2) = 0,16$; $p(x_3) = 0,08$; $p(x_4) = 0,12$; $p(x_5) = 0,14$; $p(x_6) = 0,15$; $p(x_7) = 0,1$;

10.3) $p(x_1) = 0,01$; $p(x_2) = 0,07$; $p(x_3) = 0,05$; $p(x_4) = 0,15$; $p(x_5) = 0,19$; $p(x_6) = 0,06$; $p(x_7) = 0,08$; $p(x_8) = 0,03$; $p(x_9) = 0,04$; $p(x_{10}) = 0,05$; $p(x_{11}) =$ $= 0,09$; $p(x_{12}) = 0,05$; $p(x_{13}) = 0,08$; $p(x_{14}) = 0,05$;

10.4) $p(x_1) = 0,02$; $p(x_2) = 0,15$; $p(x_3) = 0,03$; $p(x_4) = 0,09$; $p(x_5) = 0,11$; $p(x_6) = 0,08$; $p(x_7) = 0,04$; $p(x_8) = 0,05$; $p(x_9) = 0,02$; $p(x_{10}) = 0,17$. Определите вероятности появления букв x_{11} и x_{12}, если известно, что вероятность появления x_{11} вдвое больше вероятности появления x_{12}.

В каждом случае постройте кодовое дерево и запишите кодовую таблицу. Оцените избыточность полученного кода.

11. Постройте оптимальный неравномерный код для передачи сообщений, составленных из алфавита со следующим распределением вероятностей появления букв в сообщениях: $p(z_1) = 0,49$, $p(z_2) = 0,14$, $p(z_3) = 0,014$, $p(z_4) = 0,07$, $p(z_5) = 0,07$, $p(z_6) = 0,04$, $p(z_7) = 0,02$, $p(z_8) = 0,02$, $p(z_9) = 0,01$. Определите коэффициент статистического сжатия и коэффициент относительной эффективности.

12. Сравните среднюю длину кодового слова оптимального кода и энтропию первичного алфавита со следующим распределением вероятно-

стей: $p(z_1) = 0{,}03$, $p(z_2) = 0{,}02$, $p(z_3) = 0{,}013$, $p(z_4) = 0{,}18$, $p(z_5) = 0{,}13$, $p(z_6) = 0{,}15$, $p(z_7) = 0{,}16$, $p(z_8) = 0{,}11$, $p(z_9) = 0{,}02$, $p(z_{10}) = 0{,}07$.

13. Распределение вероятностей появления букв в сообщении равны: $p(z_1) = 0{,}017$, $p(z_2) = 0{,}023$, $p(z_3) = 0{,}019$, $p(z_4) = 0{,}091$, $p(z_5) = 0{,}073$, $p(z_6) = 0{,}067$, $p(z_7) = 0{,}028$, $p(z_8) = 0{,}07$, $p(z_9) = 0{,}082$, $p(z_{10}) = 0{,}01$, $p(z_{11}) = 0{,}02$, $p(z_{12}) = 0{,}005$, $p(z_{13}) = 0{,}05$, $p(z_{14}) = 0{,}007$, $p(z_{15}) = 0{,}023$, $p(z_{16}) = 0{,}011$, $p(z_{17}) = 0{,}033$, $p(z_{18}) = 0{,}02$, $p(z_{19}) = 0{,}014$, $p(z_{20}) = 0{,}042$, $p(z_{21}) = 0{,}008$, $p(z_{22}) = 0{,}017$.

Постройте оптимальный код в троичном и четверичном алфавитах для передачи сообщений, используя алгоритм Шеннона – Фано. Оцените эффективность полученных кодов.

14. Распределение вероятностей появления букв в сообщении равны: $p(z_1) = 0{,}01$, $p(z_2) = 0{,}07$, $p(z_3) = 0{,}05$, $p(z_4) = 0{,}015$, $p(z_5) = 0{,}2$, $p(z_6) = 0{,}015$, $p(z_7) = 0{,}003$, $p(z_8) = 0{,}04$, $p(z_9) = 0{,}078$, $p(z_{10}) = 0{,}096$, $p(z_{11}) = 0{,}045$, $p(z_{12}) = 0{,}021$, $p(z_{13}) = 0{,}012$, $p(z_{14}) = 0{,}056$, $p(z_{15}) = 0{,}045$, $p(z_{16}) = 0{,}002$, $p(z_{17}) = 0{,}003$, $p(z_{18}) = 0{,}036$, $p(z_{19}) = 0{,}022$, $p(z_{20}) = 0{,}11$, $p(z_{21}) = 0{,}041$, $p(z_{22}) = 0{,}008$, $p(z_{23}) = 0{,}022$.

Постройте оптимальный код в троичном и четверичном алфавитах для передачи сообщений, используя алгоритм Хаффмена. Оцените эффективность полученных кодов.

15. Осуществите блочное кодирование по два, три и четыре символа исходного алфавита сообщений, состоящего из двух букв z_1 и z_2 с вероятностями $p(z_1) = 0{,}72$, $p(z_2) = 0{,}28$. Кодирование произведите неравномерным оптимальным кодом. Оцените эффективность полученных кодов.

16. Разработан префиксный код:

111	110	10	011	010	00
л	к	о	п	т	а

Дешифрируйте следующие сообщения:

16.1) 1100001110010;

16.2) 0101001110010;

16.3) 110001100010;

16.4) 0111011111000;

16.5) 1101011101100110;

16.6) 0111110011000010.

17. Определите кодовое расстояние для вектора кодовых комбинаций:

    17.1) A = {0, 1, 2, 3, 4, 5} 02144, 52114, 32142, 41254, 11242;

    17.2) A = {0, 1, 2, 3, 4} 221431, 234410, 330012, 411243, 322110.

18. Определите кодовое расстояние в системе двоичных векторов: 11001, 00111, 10110, 00101, 01011.

19. Определите, является ли группой следующая совокупность векторов:

    19.1) 00000, 11011, 10110, 00111, 01101, 11100, 10001, 01010;

    19.2) 00000, 00110, 10101, 11001, 11101, 00011.

20. Какой вид должна иметь образующая матрица группового кода, оптимального с точки зрения минимума контрольных разрядов при максимуме информационных разрядов, для использования его в системе телемеханики, проектируемой для передачи не менее 2 000 различных сообщений.

21. Какой вид имеют комбинации группового кода с кодовым расстоянием равным 3, построенного для передачи четырехзначных двоичных комбинаций на все сочетания, если его порождающая матрица имеет вид

$$G_{4,7} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{Vmatrix}.$$

22. По заданной образующей матрице группового кода постройте проверочную матрицу, а также определите контрольные элементы для заданной кодовой комбинации β:

    22.1) $G_{4,7} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$, β = 1010;

    22.2) $G_{5,9} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{pmatrix}$, β = 01110;

22.3) $G_{5,8}=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$, $\beta = 10110$;

22.4) $G_{4,8}=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$, $\beta = 1101$.

23. Приведите порождающую матрицу к каноническому виду

$$G(x)=\left(\begin{array}{cccc|ccc} 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \end{array}\right).$$

24. Сигнальная система формирует кодовое сообщение о состоянии 15 различных объектов в виде двоичного вектора (0 – сигнал отсутствует, 1 – сигнал alarm). Среда передачи подвергается воздействию помех. Спроектируйте помехоустойчивый код в виде образующей и проверочной матриц, который способен обнаруживать до 4 ошибок в сообщении. Запишите все кодовые сообщения, в которых сигнал alarm появляется только от одного объекта, сформируйте для них контрольные символы. Оцените количество возможных различных сообщений, если сигнал alarm поступает от 2, 3 и т.д. объектов одновременно.

25. Постройте корректирующий код (образующую и проверочную матрицу), который будет передавать 25 сообщений и исправлять 3 ошибки. В качестве сообщения запишите число 9 в двоичном виде и определите для него проверочные символы в разработанном коде.

26. Проектируется телемеханическая система, предназначенная для передачи 18 различных сообщений по каналам связи, подвергающихся воздействию помех. Распределение вероятностей возникновения ошибок из-за воздействия помех подчиняется биномиальному закону с парамет-

ром $p = 0{,}8$. Необходимо разработать помехоустойчивый код, который способен обнаруживать 3 и исправлять 2 ошибки. Выпишите кодовые комбинации для всех 18 сообщений с контрольными символами. Опишите процесс исправления кодовой комбинации, соответствующей десятичному числу 11, если вектор ошибки равен 01100.

ЗАКЛЮЧЕНИЕ

В учебном пособии рассмотрены основные теоретические вопросы, связанные с оценкой количества информации, передаваемой или перерабатываемой радиотехнической, телекоммуникационной системой. Изучение данных теоретических вопросов играет важнейшую роль при решении ряда практических задач по проектированию и выбору оптимальных кодов для радиотехнических систем, исходя из различных критериев. Получение подобных практических навыков необходимо для современных инженеров, поскольку связано как с улучшением технических характеристик проектируемого изделия (увеличение быстродействия, повышение верности передачи сообщений по каналам связи с помехами и др.), так и позволяет получить экономическую выгоду, что немаловажно в современных условиях.

Каждый раздел содержит необходимые теоретические сведения, алгоритм решения рассмотренного типа задач и подробные примеры решения, а также задачи для самостоятельного решения и контрольные вопросы для проверки полученных знаний и практических навыков.

В учебном пособии представлен материал разного уровня сложности, что должно способствовать развитию умения систематизировать и обобщать теоретический материал, а также приобретению навыков комплексного системного подхода к решению задач анализа и синтеза радиоэлектронных систем и комплексов.

СПИСОК ЛИТЕРАТУРЫ

1. *Блинова, И. В.* Теория информации [Текст]: учебное пособие / И. В. Блинова, И. Ю. Попов. – Санкт-Петербург: Университет ИТМО, 2018. – 84 с.
2. *Гошин, Е. В.* Практикум по теории информации и кодирования [Текст]: учебное пособие / Е. В. Гошин. – Самара: Изд-во Самарского ун-та, 2018. – 80 с.
3. *Дмитриев, В. И.* Прикладная теория информации [Текст]: учебное пособие / В. И. Дмитриев. – Москва: Высшая школа, 1989. – 320 с.
4. *Духин, А. А.* Теория информации [Текст]: учебное пособие для специальности 090102 «Компьютерная безопасность» и др. / А. А. Духин. – Москва: Гелиос АРВ, 2007. – 248 с.
5. *Зверева, Е. Н.* Сборник примеров и задач по основам теории информации и кодирования сообщений [Текст] /Е. Н. Зверева, Е. Г. Лебедько. – Санкт-Петербург: НИУ ИТМО, 2014. – 76 с.
6. Информатика. Энциклопедический словарь для начинающих [Текст]; под ред. Д. А. Поспелова. – Москва: Педагогика-Пресс, 1994.
7. *Кудряшов, Б. Д.* Теория информации [Текст] / Б. Д. Кудряшов. – Санкт-Петербург: Питер, 2009. – 320 с.
8. *Кузьмин, И. В.* Основы теории информации и кодирования [Текст] / И. В. Кузьмин, В. А. Кедрус. – Киев: Вища школа. Головное изд-во, 1986. – 238 с.
9. *Ломакин, Д. В.* Прикладная теория информации и кодирования [Текст]: конспект лекций / Д. В. Ломакин, А. И. Туркин. – Горький: Горьковский политехнический институт, 1988. – 52 с.
10. *Орлов, В. А.* Теория информации в упражнениях и задачах [Текст]: учебное пособие для втузов / В. А. Орлов, Л. И. Филиппов. – Москва: Высш. школа, 1976. – 136 с.
11. Прикладная теория информации [Текст]: учеб. пособие / А. С. Гуменюк, Н. Н. Поздниченко. – Омск : Изд-во ОмГТУ, 2015.
12. Приложения теории информации и криптографии в радиотехнических системах [Текст]/ О. А. Усенко. – Таганрог: Издательство Южного федерального университета, 2017. – 240 с.

13. *Солопченко, Г. Н.* Теория информации [Текст]: учебное пособие / Г. Н. Солопченко. – Санкт-Петербург, 2014. – 165 с.
14. *Шеннон, К.* Работы по теории информации и кибернетике [Текст] / К. Шеннон. – Москва: Изд-во иностранной литературы, 1963. – 832 с.

Учебное издание

УСЕНКО Ольга Александровна

ПРИЛОЖЕНИЯ ТЕОРИИ ИНФОРМАЦИИ К ЗАДАЧАМ РАДИОТЕХНИКИ

Учебное пособие

Редактор *Н. И. Селезнева*
Корректоры: *З. И. Надточий*, *Н. И. Селезнева*
Компьютерная верстка *И. А. Бобровской*

Подписано в печать 30.12.2021 г.
Бумага офсетная. Формат 60×84 1/16. Усл. печ. лист. 8,95.
Уч. изд. л. 5,38. Тираж 30 экз. Заказ № 8388.

Издательство Южного федерального университета

Отпечатано в отделе полиграфической, корпоративной и сувенирной продукции
Издательско-полиграфического комплекса КИБИ МЕДИА ЦЕНТРА ЮФУ.
344090, г. Ростов-на-Дону, пр. Стачки, 200/1. Тел. (863) 243-41-66.

9 785927 539642