МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Белгородский государственный технологический университет
им. В. Г. Шухова

И. В. Иванов

Теория информационных процессов и систем

Учебное пособие

2-е издание, переработанное и дополненное

Рекомендовано Федеральным государственным бюджетным образовательным учреждением высшего профессионального образования «Московский государственный технический университет имени Н.Э.Баумана» в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлению «Информационные системы и технологии»

Регистрационный номер рецензии 2714 от 17 февраля 2014 г. МГУП

Белгород
2014

УДК 32.971
ББК 681.3
И 20

Рецензенты:

Заслуженный деятель науки Российской Федерации, доктор технических наук, профессор, заведующий кафедрой технической кибернетики Белгородского государственного технологического университета им. В.Г. Шухова *В.Г. Рубанов*

Заслуженный деятель науки Российской Федерации, доктор технических наук, профессор, заведующий кафедрой вычислительной техники Юго-Западного государственного университета *В.С. Титов*

Иванов, И.В.
И 20 Теория информационных процессов и систем: учебное пособие / И.В. Иванов. – 2-е изд., перераб. и доп. – Белгород: БГТУ, 2014. – 203 с.
ISBN 978-5-361-00222-1

В издании изложены теоретические основы информационных процессов и систем. Рассмотрены вопросы количественной оценки информации, передачи информации по каналам связи, а также методы противодействия помехам в процессе транспортировки информационных сообщений.

Предыдущее издание пособия вышло в свет в 2007 г.

Учебное пособие предназначено для студентов, обучающихся по укрупненной группе направлений «Информатика и вычислительная техника».

УДК 32.971
ББК 681.3

ISBN 978-5-361-00222-1

Введение

Теория информационных процессов и систем является базовой теоретической дисциплиной направления «Информационные системы и технологии». Ее задача – сформировать представление об окружающем мире, как мире, в котором населяющие его объекты взаимодействуют с помощью процессов приема-передачи, переработки и хранения информации. А сами эти объекты представляют собой иерархические системы, составные части которых также объединены информационными связями. Таким образом, основополагающими понятиями этой дисциплины являются категории «система», «процесс», «информация».

Начинается дисциплина с изучения понятия «система» (от греческого συστημα – состав). В первой главе даются сведения об общей теории систем и закономерностях поведения систем различной физической природы, а также способах формального описания систем с точки зрения их информационного взаимодействия. Вторая глава содержит материал по основам теории информации, включая количественные характеристики информации. В третьей главе рассматриваются процессы передачи информации по каналам связи. Содержанием четвертой главы являются теоретические основы и алгоритмы помехоустойчивого кодирования информации. Пятая глава представляет собой пособие по выполнению вычислительного практикума по дисциплине «Теория информационных процессов и систем».

Каждая глава содержит основной текст, а также вопросы и задачи по материалам главы. Многие задачи сопровождаются решениями. Материал учебного пособия содержит врезки дополнительной информации, которая не обязательна для изучения и служит для расширения кругозора студентов. Такие врезки набраны более мелким шрифтом. Кроме них по ходу изложения материала приводятся примеры, которые также выделены специальным шрифтом. Теоремы и наиболее важные определения, которые рекомендуется запомнить, отмечены двойной вертикальной линией по левому краю.

В основу учебного пособия положены материалы лекций, которые читаются автором на кафедре информационных технологий Белгородского государственного технологического университета им. В.Г. Шухова для студентов направления «Информационные системы и технологии».

Глава 1
Основные понятия теории информационных систем Модели и методы описания систем

1.1. Основные понятия и определения теории систем

Понятие системы. Признаки системности

Потребность в использовании понятия «система» возникала для объектов различной физической природы с древних времен: еще Аристотель обратил внимание на то, что целое (то есть система) не сводится к сумме частей, его образующих. В современной теории систем это свойство называется «принципом эмерджентности».

> Принцип эмерджентности заключается в том, что свойства целого не сводятся к простой сумме свойств составляющих его частей, а при объединении частей в целое образуется новое качество, не присущее отдельным частям.

Представьте себе группу не знакомых друг с другом людей, оружие в штабелях, обмундирование, хранящееся на складе, технику на стоянке. Простая совокупность таких разрозненных объектов не способна к ведению боевых действий. А если тех же людей вооружить, обмундировать, придать им боевую технику, организовать в подразделения, назначив командиров, установив подчиненность, определить способы связи, то возникает новое качество – «боеспособность».

Термин «система» и связанные с ним понятия комплексного, системного подхода исследуются и подвергаются осмыслению философами, биологами, психологами, кибернетиками, физиками, математиками, экономистами, инженерами различных специальностей. Потребность в использовании этого термина возникает в тех случаях, когда невозможно что-то продемонстрировать, изобразить, представить одним выражением и нужно подчеркнуть, что это будет большим, сложным, не полностью сразу понятным (то есть, имеющим некоторую степень неопределенности) и, в то же время, целым, единым. Например – «солнечная система», «система управления станком», «система организационного управления предприятием (городом, регионом и т. п.)», «экономическая система», «система кровообращения» и т.д.

В математике термин система используется для отображения совокупности математических выражений или правил – «система уравнений», «система счисления», «система мер» и т. п. Казалось бы, в этих случаях можно было воспользоваться терминами «множество» или «совокупность». Однако понятие системы подчеркивает упорядоченность, целостность, наличие определенных закономерностей.

Если попытаться дать общее определение для любых систем, то оно будет очень абстрактным и не удобным для практических целей, однако у всех сис-

тем, независимо от их физической природы, есть некоторые общие признаки, свойства и законы поведения.

> Система – множество элементов, находящихся в отношениях и связях друг с другом, которое образует определенную целостность, единство.

Из этого короткого определения вытекают частные признаки системности:

структурированность, то есть возможность расчленения системы на составляющие компоненты. С одной стороны, система – это целостное образование и представляет целостную совокупность элементов, а с другой стороны, в системе четко можно выделить ее элементы (компоненты, неделимые объекты);

взаимосвязанность отдельных частей, то есть наличие более или менее устойчивых связей (отношений) между элементами системы, превосходящих по своей силе (мощности) связи (отношения) этих элементов с элементами, не входящими в данную систему. В системах любой природы между элементами существуют те или иные связи (отношения). При этом с системных позиций определяющими являются не любые связи, а только лишь существенные и важные для функционирования системы связи (отношения), которые определяют интегративные свойства системы;

интегративность системы, то есть наличие единых целей, свойств, качеств, присущих системе в целом, но не присущих ее элементам в отдельности. Интегративные свойства системы обуславливает тот факт, что свойство системы, несмотря на зависимость от свойств элементов, не определяется ими полностью. Из этого следует, что простая совокупность элементов и связей между ними еще не система, и поэтому, расчленяя систему на отдельные части (элементы) и изучая каждую из них в отдельности, нельзя познать все свойства нормально (хорошо) организованной системы в целом.

Понятие «система» широко использовалось в различных областях знаний, имея конкретное содержание в каждой предметной области. Такие разделы науки, как системотехника, методы проектирования, инженерное творчество (в инженерно-технической сфере), исследование операций (в экономике и военном деле), политология, футурология (в социально-административной деятельности) кибернетика, моделирование (в научно-прикладной сфере) по сути исследуют закономерности системного развития. На определенной стадии развития научного знания исследования по теории систем оформились в самостоятельную науку.

Краткая история развития системных представлений

Как уже было отмечено, системные представления зародились еще в древности. Из более поздних научных достижений можно назвать, во-первых, работы А. Ампера.

Андре Мари Ампер (1775-1836) – знаменитый французский физик, математик и естествоиспытатель, член Парижской Академии наук, иностранный член Петербургской Академии наук, профессор Нормальной школы в Париже. Руководил кафедрами физики и математике в Политехнической школе. Известен своими работами в области электродинамики.

А. Ампер написал серию работ «Опыт о философии наук, или аналитическое изложение классификации человеческих знаний» (1834-1843 гг.). Среди других наук он выделил науку об управлении государством и назвал ее «кибернетикой» (от греческого "κιβερ" - кормчий, рулевой, управляющий чем-то). С этой поры термин «кибернетика» вошел в научный обиход.

Его идеи развил польский философ Болеслав Трентовский, который в 1843 г. опубликовал книгу «Отношение философии к кибернетике, как искусству управления народом». Болеслав Трентовский в лекциях по философии кибернетики сформулировал **принцип мудрого кормчего**. Управляющий — кибернет — в своих действиях должен быть подобен кормчему. Он не должен плыть наперекор стихии, он должен уметь использовать ее для своих целей. Для этого ему надо знать все капризы течений и направлений ветров и умело применять их силу для достижения того берега, к которому он стремится. Сила кормчего ничтожна по сравнению с силой стихии. Но зато она в руках у кормчего. И используя ее правильно, накапливая эффект, кормчий сможет достичь цели.

Важный вклад в становление системных представлений внес в начале XIX века наш соотечественник А.А.Богданов.

Александр Александрович Богданов (настоящая фамилия - Малиновский) (1873-1928) – русский ученый, биолог, философ, экономист, политический деятель, работавший в конце 19 – начале 20-го веков. Был членом социал-демократической партии, но подвергался резкой критике В.Ленина за свои философские взгляды. После революции 1917 года был одним из научных советников властей. Возглавлял в Москве Институт по переливанию крови. Умер в 1928 г. в результате опыта по переливанию крови, который проводил на себе.

А. Богданов написал фундаментальный труд – трехтомник «Всеобщая организационная наука (тектология)»[1] (1911-1925 гг.). По Богданову все явления являются процессами организации или дезорганизации. Наука тектология призвана изучать общие закономерности организации. Все наблюдаемые объекты имеют определенную степень организованности. Что такое организованность? Это можно пояснить так: чем уровень организованности выше, тем сильнее свойства целого отличаются от суммы свойств его частей. Богданов рассмотрел

[1] Богданов А.А. Всеобщая организационная наука (тектология). В 2-х кн. – М.: Наука, 1989.

соотношение понятий устойчивость и изменчивость, показал значение обратных связей, роль открытых систем.

Так, философия явилась источником возникновения обобщающего направления, названного теорией систем. И хотя истоки теории уходят в глубь веков, основоположником этого направления принято считать биолога Л. фон Берталанфи[1].

Людвиг фон Берталанфи (1901-1972) – австрийский биолог. Постоянно проживал в Канаде и США. Первооснователь обобщённой системной концепции под названием «Общая теория систем», о чем написал одноименную книгу (1950 г., издание на русском языке – 1969 г.).

В 60-е годы при постановке и исследовании сложных проблем проектирования и управления довольно широкое распространение получил термин системотехника. Сейчас этот раздел науки принято трактовать как совокупность прикладных результатов теории систем.

Применительно к задачам управления в определенный период более широкое распространение получил термин кибернетика, введенный А.Ампером, принятый для названия новой «науки об управлении в живых организмах и машинах» Н.Винером.

Норберт Винер (1894-1964) – американский ученый, «отец кибернетики». В 14 лет получил высшее математическое образование, в 18 лет стал доктором философии Гарвардского университета (по математической логике). Учился в Кембридже (Англия) и Геттингене (Германия). Преподавал математику в ряде американских университетов. С 1919 г. – профессор Массачусетского технологического института – одного из крупнейших вузов США. Знал 10 языков. Написал две автобиографические книги «Я – вундеркинд» и «Я – математик».

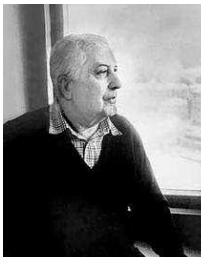
В книге «Кибернетика как наука об управлении и связи в животном и машине» (1948 г.)[2] Н. Винер заложил общие принципы науки управления. Он показал формальное единство процессов, протекающих в системах различной природы. Немного позже издал книгу «Человеческое использование человеческих существ, или Кибернетика и общество», где распространил полученные закономерности на социальные процессы. Взгляды автора, изложенные в ней, противоречили марксистской общественной теории, доминировавшей тогда в странах социалистического лагеря. В связи с этим кибернетика была объявлена советскими властями лженаукой, что предопределило отставание отечественной научной мысли в этой области. Термин «кибернетика» в настоящее время используется в более узком смысле как одно из направлений теории систем, за-

[1] Л. фон Берталанфи. Общая теория систем: критический обзор // Исследования по общей теории систем. М.: Прогресс, 1969.

[2] Винер Н. Кибернетика или Управление и связь в животном и машине. – М.: Наука, 1983. – 344 с.

нимающееся процессами управления техническими объектами. А для обобщения дисциплин, связанных с исследованием и проектированием сложных систем, используется термин «системные исследования», иногда – «системный подход».

Из более поздних достижений следует выделить работу бельгийского ученого российского происхождения Ильи Пригожина «Порядок из хаоса» (1979 г.)[1].



Илья Романович Пригожин (1917–2003) — бельгийский и американский физико-химик, лауреат Нобелевской премии по химии 1977 года, иностранный член Академии наук СССР.

И. Пригожин утверждал, что материя – не пассивная субстанция, ей присуща спонтанная активность, вызванная наличием неравновесных состояний. Было доказано существование неравновесных термодинамических систем, которые при определённых условиях, поглощая массу и энергию из окружающего пространства, могут совершать качественный скачок к усложнению (диссипативные структуры). Причём такой скачок не может быть предсказан, исходя из классических законов статистики. Благодаря таким представлениям возникла так называемая «теория катастроф» (теория бифуркаций).

Описание системы в виде «черного ящика»

Одна из задач общей теории систем – это разработка принципов формального описания функционирования систем, независимо от их физической природы. Описание системы в виде «черного ящика» опирается на следующие интуитивно понятные положения.

Система обособлена и целостна. Это означает, что при описании системы исследователь должен четко указать пространственные границы системы, отметить, что он включает в состав системы, а что относит к внешней среде.

Среда воздействует на систему. Это значит, что существует множество **U** входных воздействий на систему со стороны внешней среды, иначе называемых причинами.

Система воздействует на среду. Это означает, что у системы существует множество **Y** выходов (следствий), которые можно наблюдать извне.

Выходы (следствия) зависят от входов (причин). Это означает, что исследователь может обнаружить и каким-то образом описать причинно-

[1] Пригожин И.Р., Стенгерс И. Порядок из хаоса. Новый диалог человека с природой. М., 1986.

следственные связи между входами и выходами, иначе говоря, отображение множества **U** на множество **Y**: **Y = F(U)** (см. рис. 1.1).

Таким образом, описание системы в виде «черного ящика» включает указание границ системы, описание множеств входов и выходов, а также зависимости выходов от входов. Важно, что исследователь может наблюдать только входы и выходы системы и не имеет информации о внутренних процессах, происходящих в системе. Собственно, поэтому такое описание и называется моделью «черного ящика».

Рис. 1.1. Представление системы в виде «черного ящика»

Описание системы в виде «белого ящика»

Модель «черного ящика», приведенная выше, – это достаточно грубое, приближенное описание системы, которое можно использовать только на начальном этапе исследований. Дело в том, что, если не принимать во внимание внутреннюю структуру системы, то можно столкнуться с ситуацией, когда одни и те же входы порождают разные выходные сигналы.

> Возьмите обычную авторучку, у которой пишущий узел выдвигается из корпуса при нажатии на колпачок и, не глядя на нее, нажмите несколько раз. Если вы не видите, в каком состоянии находится ручка, то не сможете определить, какова будет ее реакция на очередное нажатие: выдвинется пишущий узел из стержня или, наоборот, втянется. То есть, одна и та же причина (нажатие на колпачок) может привести к разным следствиям.

Для более полного описания системы необходимо учесть внутренние параметры, называемые **состояниями**. Если в описание по принципу «черного ящика» вводятся состояния системы, то приходим к модели «белого ящика», которая включает в себя еще и дополнительные функциональные составляющие.

> **Состояния** – некоторые внутренние объекты системы, которые связывают собой всю предысторию входов.
>
> Связь между входом и состоянием называется **переходным отображением**.
>
> Связь между состоянием и выходом называется **отображением выхода** или **функцией наблюдения**.

Введем обозначения составных элементов модели «белого ящика»:

T – индексирующее множество, отражающее последовательность протекания процессов (попросту говоря, время);

U, Y, X – множество значений входа, выхода и состояний;

η – отображение выхода (функция наблюдения), которое можно описать следующим образом:

$\eta: T \times X \to Y$ или $y(t) = \eta(t, x(t))$, где $t \in T, y \in Y, x \in X$;

σ - переходное отображение, имеющее следующее описание:

$\sigma: (T \times T)^{+} \times X \times U \to X$ или $x(t) = \sigma(t, \tau, x(\tau), u)$, где

$(T \times T)^{+}$ - упорядоченное множество: $(T \times T)^{+}: \{(t, \tau), \tau < t\}, \tau \in T, t \in T, u \in U$;

Используя введенные обозначение, сформулируем некоторые аксиомы теории систем:

1. **Аксиома согласованности**.

За нулевой промежуток времени система не может перейти в другое состояние,

или

в один и тот же момент времени система не может находиться в двух разных состояниях.

Формально аксиому согласованности можно записать так:

$x(t) = \sigma(t, t, x(t), u)$.

2. **Аксиома детерминизма**.

Состояние системы в момент времени $t_2 > t_0$ однозначно определяется состоянием в момент времени t_0 и входом на отрезке $[t_0, t_2]$.

Формальная запись:

$\forall t_0 \le t_1 \le t_2: \sigma(t_2, t_0, x(t_0), u) = \sigma(t_2, t_1, \sigma(t_1, t_0, x(t_0), u), u)$

3. **Аксиома причинности**.

Одна и та же причина вызывает одно и то же следствие.

В математическом виде:

$\forall u_1, u_2 \in U: (u_1 = u_2) \Rightarrow \sigma(t, \tau, x(\tau), u_1) = \sigma(t, \tau, x(\tau), u_2)$.

Теперь можно дать формальное определение системы с точки зрения общей теории систем.

> Говорят, что некоторая система $\sum$ определена, если заданы индексирующее множество **T**, множество значений входов **U**, выходов **Y** и состояний **X**, переходное отображение **σ** и отображение выхода **η**, удовлетворяющие аксиомам согласованности, причинности и детерминизма, такие, что:

$\forall\, y \in Y\ \exists\, x \in X,\ u \in U$, для которых при любых $(t, \tau) \in (T \times T)^{+}$ выполняется соотношение $y(t) = \eta(t, \sigma(t, \tau, x(\tau), u))$.

1.2. Вторичные понятия и определения теории систем

Определения системы

Понятие «система» настолько многогранно, что в научной литературе имеется множество альтернативных определений системы, раскрывающие те или иные стороны этой субстанции. Приведем некоторые из них.

D0. Система есть множество входов, множество выходов, множество состояний, связанных оператором переходов (переходным отображением) и оператором выходов (функцией наблюдения):

$$S=(T, U, Y, X, \sigma, \eta),$$

где **T** - время **U** - входы, **Y** - выходы, **X** - состояния, σ - оператор переходов, η - оператор выходов. Это определение учитывает все основные компоненты, рассматриваемые в кибернетике.

В зависимости от количества учитываемых факторов и степени абстрактности определение понятия «система» можно представить и в других формах.

D1. Система есть нечто целое:

$$S=A\{1,0\}.$$

Это определение выражает факт существования и целостность. Двоичное суждение **A{1,0}** отображает наличие или отсутствие этих качеств.

D2. Система есть организованное множество[1]:

$$S=(ORG, M),$$

где **ORG** - оператор организации; **M** - множество.

D3. Система есть множество вещей, свойств и отношений[2]:

$$S=(\{m\},\{n\},\{r\}),$$

где **m** - вещи, **n** - свойства, **r** - отношения.

D4. Система есть множество элементов, образующих структуру и обеспечивающих определенное поведение в условиях окружающей среды:[3]

$$S=(\varepsilon, ST, BE, E),$$

где ε - элементы, **ST** - структура, **BE** - поведение, **E** - среда.

---

[1] Темников Ф.Е., Славинский В.Л. Математические развертывающие системы**.** – М.: Энергия, 1970. 122 с.
[2] А.И.Уемов. Системный подход и общая теория систем. – М.: Мысль, 1978. 272 с.
[3] Основы системного подхода и их приложение к разработке территориальных АСУ / Под ред. Ф.И.Перегудова. Томск: Изд-во Томского ун-та, 1976.

D5. Это определение для биологических и иных саморазвивающихся систем и учитывает генетическое (родовое) начало **GN**, условия существования **KD**, обменные явления **MB**, развитие **EV**, функционирование **FC** и репродукцию (воспроизведения) **RP**:

S=(GN, KD, MB, EV, FC, RP).

D6. Это определение для нейрокибернетических систем, оперирует понятиями модели **F**, связи **SC**, пересчета **R**, самообучения **FL**, самоорганизации **FQ**, проводимости связей **CO** и возбуждения моделей **JN**:

S=(F, SC, R, FL, FO, CO, JN).

D7. Для организационных систем удобно в определении системы учитывать следующее:

S=(PL, RO, RJ, EX, PR, DT, SV, RD, EF),

где **PL** - цели и планы, **RO** - внешние ресурсы, **RJ** - внутренние ресурсы, **EX** - исполнители, **PR** - процесс, **DT** - помехи, **SV** - контроль, **RD** - управление, **EF** - эффект.

Под системой понимается объект, свойства которого не сводятся без остатка к свойствам составляющих его дискретных элементов (неаддитивность свойств). Интегративное свойство системы обеспечивает ее целостность, качественно новое образование по сравнению с составляющими ее частями.

Вторичные понятия теории систем

Элемент. Под элементом принято понимать простейшую неделимую часть системы. Ответ на вопрос, что является такой частью, может быть неоднозначным и зависит от цели рассмотрения объекта как системы, от точки зрения на него или от аспекта его изучения. Таким образом, элемент - это предел деления системы с точки зрения решения конкретной задачи и поставленной цели. Систему можно расчленить на элементы различными способами в зависимости от формулировки цели и ее уточнения в процессе исследования.

Подсистема. Система может быть разделена на элементы не сразу, а последовательным расчленением на подсистемы, которые представляют собой компоненты более крупные, чем элементы, и в то же время более детальные, чем система в целом. Возможность деления системы на подсистемы связана с вычленением совокупностей взаимосвязанных элементов, способных выполнять относительно независимые функции, подцели, направленные на достижение общей цели системы. Названием "подсистема" подчеркивается, что такая часть должна обладать свойствами системы (в частности, свойством эмерджентности). Этим подсистема отличается от простой группы элементов, для которой не сформулирована подцель и не выполняются свойства эмерджентности (для такой группы используется название "компоненты"). Использование этого понятия оказывается особенно плодотворным в тех случаях, когда в качестве

подсистем фигурируют некоторые более или менее самостоятельно функционирующие части системы.

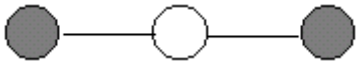
Структура. Это понятие происходит от латинского слова structure, означающего строение, расположение, порядок. Структура отражает наиболее существенные взаимоотношения между элементами и их группами (компонентами, подсистемами), которые мало меняются при изменениях в системе и обеспечивают существование системы и ее основных свойств.

Структура - это совокупность элементов и связей между ними.

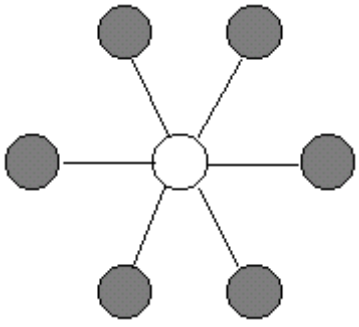
Структура может быть представлена графически, в виде теоретико-множественных описаний, матриц, графов и других языков моделирования структур.

Структуру часто представляют в виде иерархии. Иерархия - это упорядоченность компонентов по степени важности (многоступенчатость, служебная лестница). Между уровнями иерархической структуры могут существовать взаимоотношения строгого подчинения компонентов (узлов) нижележащего уровня одному из компонентов вышележащего уровня, т. е. отношения так называемого древовидного порядка. Такие иерархии называют сильными или иерархиями типа «дерева».

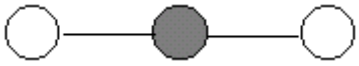
Централистические и скелетные структуры. А.А.Богданов в своих трудах по тектологии выделяет два вида структур: централистическая и скелетная. Первые характеризуются наличием центра, к которому тяготеют и с которым тесно связаны, подчиняясь ему, все кроме части систем. Центр концентрирует активности всех частей. “Скелетная” структура состоит из двух частей: пластичной и скелетной. Пластичность означает подвижный, гибкий характер связей системы, легкость перегруппировки ее элементов. Чем пластичнее система, тем больше в ней образуется комбинаций при изменяющихся к этим условиям. Централистический тип структуры “всего более концентрирует активности, создает возможности максимального их накопления в одной системе”. Скелетный “по преимуществу фиксирует активности, закрепляет их в данной форме, обуславливает максимальную прочность системы”. Основные организационные структуры представлены в тектологии следующим образом:



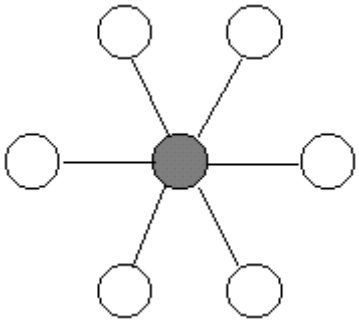
1) простейшая централистическая структура



2) сложная централистическая структура



3) простейшая скелетная структура



4) сложная скелетная структура

Условные обозначения: высшая организованность ○, низшая организованность ●

Связь. Понятие "связь" входит в любое определение системы наряду с понятием "элемент" и обеспечивает возникновение и сохранение структуры и целостных свойств системы. Это понятие характеризует одновременно и строение (статику), и функционирование (динамику) системы.

Связь характеризуется направлением, силой и характером (или видом). По первым двум признакам связи можно разделить на направленные и ненаправленные, сильные и слабые, а по характеру - на связи подчинения, генетические, равноправные (или безразличные), связи управления. Связи можно разделить также по месту приложения (внутренние и внешние), по направленности процессов в системе в целом или в отдельных ее подсистемах (прямые и обратные). Связи в конкретных системах могут быть одновременно охарактеризованы несколькими из названных признаков.

Важную роль в системах играет понятие "**обратной связи**".

Обратные связи бывают положительные и отрицательные, гибкие и жесткие.

> Положительная обратная связь – при увеличении выходного сигнала механизм обратной связи срабатывает так, чтобы выходной сигнал продолжал увеличиваться (а при уменьшении выходного сигнала – продолжал уменьшаться). Отрицательная обратная связь – при увеличении выходного сигнала механизм обратной связи срабатывает так, чтобы выходной сигнал уменьшился (а при уменьшении выходного сигнала – увеличился).

Отрицательная обратная связь способствует устойчивости систем.

> Жесткая обратная связь – когда поправка, подаваемая на вход через канал обратной связи, пропорциональна выходному сигналу. Гибкая обратная связь – поправка пропорциональна производной выходного сигнала.

Обратная связь является основой саморегулирования и развития систем, приспособления их к изменяющимся условиям существования.

Состояние. Понятием "состояние" обычно характеризуют мгновенную фотографию, "срез" системы, остановку в ее развитии. Его определяют либо через входные воздействия и выходные сигналы (результаты), либо через макропараметры, макросвойства системы (например, давление, скорость, ускорение - для физических систем; производительность, себестоимость продукции, прибыль - для экономических систем).

Поведение. Если система способна переходить из одного состояния в другое (например, $\mathbf{S_1 \rightarrow S_2 \rightarrow S_3}$), то говорят, что она обладает поведением. Этим понятием пользуются, когда неизвестны закономерности переходов из одного состояния в другое. Тогда говорят, что система обладает каким-то поведением и выясняют его закономерности. С учетом введенных выше обозначений поведение можно представить как функцию $\mathbf{x(t)=f(x(t-1), u(t))}$.

Внешняя среда. Под внешней средой понимается множество элементов, которые не входят в систему, но взаимодействуют с системой, что вызывает изменение поведения системы.

Модель. Под моделью системы понимается описание системы, отображающее определенную группу ее свойств.

Равновесие - это способность системы в отсутствие внешних возмущающих воздействий (или при постоянных воздействиях) сохранить свое состояние сколь угодно долго.

Устойчивость. Под устойчивостью понимается способность системы возвращаться в состояние равновесия после того, как она была из этого состояния выведена под влиянием внешних возмущающих воздействий.

Состояние равновесия, в которое система способна возвращаться, по аналогии с техническими устройствами называют устойчивым состоянием равновесия. Равновесие и устойчивость в экономических и организационных системах - гораздо более сложные понятия, чем в технике, и до недавнего времени ими пользовались только для некоторого предварительного описательного представления о системе. В последнее время появились попытки формализованного отображения этих процессов и в сложных организационных системах, помогающие выявлять параметры, влияющие на их протекание и взаимосвязь.

Развитие. Это процесс повышения уровня организованности системы. Уровень организованности определяется свойством эмерджентности: $\mathbf{Q(S) > \Sigma q_i}$, где $\mathbf{Q(S)}$ – свойства системы в целом, а $\mathbf{q_i}$ – свойство $\mathbf{i}$-го элемента системы.

Нулевой уровень организованности диагностируется в случае, когда объединение элементов в систему не приводит к возникновению нового качества: $\mathbf{Q(S)=\Sigma q_i}$

Для описания степени организованности вводятся величины α - коэффициент целостности (т.е. степень интегрированности элементов в систему) и β - коэффициент использования элементов (т.е. степень самостоятельности элементов). Считается, что при нулевой организованности $\alpha=0$, $\beta=1$, с ростом организованности α растет в пределе до **1**, β снижается в пределе до **0**.

$$\alpha = 1 - \frac{\sum q_i}{Q(S)},$$

$$\beta = 1 - \alpha.$$

Цель. Применение понятия "цель" и связанных с ним понятий целенаправленности, целеустремленности, целесообразности сдерживается трудностью их однозначного толкования в конкретных условиях. Это связано с тем, что процесс целеобразования и соответствующий ему процесс обоснования целей в организационных системах весьма сложен и не до конца изучен. Его исследованию большое внимание уделяется в психологии, философии, кибернетике. Как правило, цель определяется как «заранее мыслимый результат сознательной деятельности человека». В практических применениях цель - это идеальное устремление, которое позволяет коллективу увидеть перспективы или реальные возможности, обеспечивающие своевременность завершения очередного этапа на пути к идеальным устремлениям.

Состояние системы – это вектор объектов $\mathbf{X=\{x_1...x_n\}}$. Цель состоит в экстремизации (максимизации или минимизации) одной или нескольких переменных. Если можно выразить зависимость этой переменной x_i, от входов **U** и других состояний, то такая зависимость называется целевой функцией $\mathbf{x_i=F_i(U,X,t)}$. Целью в такой постановке является получение $\mathbf{\max F_i(U,X,t)}$.

Изучением задач, связанных с назначением и достижением целей, занимается **системный анализ**. Говорят, что «системный анализ – формализованный здравый смысл».

1.3. Классификация систем

Основные проблемы теории систем

Основные проблемы теории систем легко формализовать в терминах модели «белого ящика».

Идентификация – описание связей между входами, выходами и состояниями (нахождение функции наблюдения η, переходного отображения σ).

Задача идентификации, таким образом, это не что иное, как переход от модели «черного ящика» к модели «белого ящика».

Прогнозирование – предсказание выхода Y по входу U.

Задачей прогнозирования, например, является предположение того, как поведет себя космический зонд в необычных климатических условиях марсианской среды, которые нельзя точно воспроизвести на Земле.

Управление – определение **U** для известного значения **X.**

Задача управления, например, возникает, когда ясно, в какой порт должен прибыть корабль, и требуется выработать маршрут для достижения цели.

Диагностирование – определение состояния **X** по известным входу **U** и выходу **Y.**

Диагностированием знаний учащегося занимается, к примеру, экзаменатор, задающий студенту вопросы и принимающий от него ответы.

Распознавание – определение **U** по известному выходу **Y.**

Эта ситуация может быть проиллюстрирована игрой «крокодил», когда участники пытаются угадать слово, заданное ведущим, наблюдая за мимикой и жестами игрока.

Основная классификация систем

Система называется **дискретной**, если дискретно множество **T**, т.е. **T={tk, k∈ Z** (мн-во целых чисел)}

Система называется **непрерывной**, если **T=R** (совпадает с множеством действительных чисел)

Система называется **конечным автоматом**, если она является дискретной, а множества **U, X, Y** имеют конечной число элементов

Система называется **конечномерной**, если множества **U, X, Y** являются конечномерными линейными пространствами

Система называется **стационарной**, если выполняется условие инвариантности переходного отображения к сдвигу по времени:

$$\sigma(t, \tau, x(\tau), u(t,\tau)) = \sigma(t+\Delta t, \tau+\Delta t, x(\tau+\Delta t), u(t+\Delta t,\tau+\Delta t))$$

Система называется **гладкой**, если переходное отображение σ является общим решением дифференциального уравнения $dx/dt=f(t, x, u)$ или конечно-разностного уравнения $x(tk+1)=f(tk, x(tk), u)$

Система называется **линейной**, если σ и η - линейные функции.

Другие способы классификации информационных систем

Системы разделяются на классы по различным признакам, и в зависимости от решаемой задачи можно выбрать разные принципы классификации. При этом систему можно охарактеризовать одним или несколькими признаками. Системы классифицируются следующим образом:

по виду отображаемого объекта – технические, биологические и др.;

по виду научного направления – математические, физические, химические и т. п.;

по виду формализованного аппарата представления системы – детерминированные и стохастические;

по типу целеустремленности – открытые и закрытые;

по степени организованности – хорошо организованные, плохо организованные (диффузные), самоорганизующиеся системы;

по сложности структуры и поведения – простые и сложные;

по величине – большие и малые.

Классификации всегда относительны. Так в детерминированной системе можно найти элементы стохастических систем. Цель любой классификации ограничить выбор подходов к отображению системы и дать рекомендации по выбору методов.

Технические системы. Параметрами технических объектов являются движущие объекты, объекты энергетики, объекты химической промышленности, объекты машиностроения, бытовая техника и многие другие.

Экономические объекты. Экономическими объектами являются: цех, завод, предприятия различных отраслей. В качестве одной из переменных в них выступают экономические показатели, например, прибыль.

Биологические системы. Живые системы поддерживают свою жизнедеятельность благодаря заложенным в них механизмам управления.

Детерминированные и стохастические системы. Если внешние воздействия, приложенные к системе, являются определенными известными функциями времени **u=f(t),** то в этом случае состоянии системы, описываемой обыкновенными дифференциальными или конечно-разностными уравнениями, в любой момент времени **t** может быть однозначно описано по состоянию системы в предшествующий момент времени.

Системы, для которых состояние системы однозначно определяется начальными значениями и может быть предсказано для любого момента времени называются **детерминированными**.

Стохастические системы – системы, изменения в которых носят случайный характер. Например, воздействие на энергосистему различных пользователей. При случайных воздействиях данных о состоянии системы недостаточно для предсказания в последующий момент времени.

Случайные воздействия могут прикладываться к системе извне, или возникать внутри некоторых элементов (внутренние шумы). Статистические свойства случайной величины определяют по ее функции распределения или плотности вероятности. При этом в основном в технике преобладает нормальный (гауссовский) закон распределения случайных величин.

Открытые и закрытые системы. Понятие открытой системы ввел Л. фон Берталанфи. Основные отличительные черты открытых систем - способность обмениваться с внешней средой энергией и информацией. Закрытые (замкнутые) системы изолированы от внешней среды (с точностью, принятой в модели).

Хорошо и плохо организованные системы. Представить анализируемый объект или процесс в виде «**хорошо организованной системы**» означает определить элементы системы, их взаимосвязь, правила объединения в более крупные компоненты, т. е. определить связи между всеми компонентами и целями системы, с точки зрения которых рассматривается объект или ради достижения которых создается система. Проблемная ситуация может быть описана в виде математического выражения, связывающего цель со средствами, т. е. в виде критерия эффективности, критерия функционирования системы, который может быть представлен сложным уравнением или системой уравнений. Решение задачи при представлении ее в виде хорошо организованной системы осуществляется аналитическими методами формализованного представления системы.

Примеры хорошо организованных систем: солнечная система, описывающая наиболее существенные закономерности движения планет вокруг Солнца; отображение атома в виде планетарной системы, состоящей из ядра и электронов; описание работы сложного электронного устройства с помощью системы уравнений, учитывающей особенности условий его работы (наличие шумов, нестабильности источников питания и т. п.).

Для отображения объекта в виде хорошо организованной системы необходимо выделять существенные и не учитывать относительно несущественные для данной цели рассмотрения компоненты: например, при рассмотрении солнечной системы не учитывать метеориты, астероиды и другие, мелкие по сравнению с планетами, элементы межпланетного пространства.

Описание объекта в виде хорошо организованной системы применяется в тех случаях, когда можно предложить детерминированное описание и экспериментально доказать правомерность его применения, адекватность модели реальному процессу. Попытки применить класс хорошо организованных систем для представления сложных многокомпонентных объектов или многокритериальных задач плохо удаются: они требуют недопустимо больших затрат времени, практически нереализуемы и неадекватны применяемым моделям.

При представлении объекта в виде «**плохо организованной или диффузной системы**» не ставится задача определить все учитываемые компоненты, их свойства и связи между ними и целями системы. Система характеризуется некоторым набором макропараметров и закономерностями, которые находятся на основе исследования не всего объекта или класса явлений, а на основе определенней с помощью некоторых правил выборки компонентов, характеризующих исследуемый объект или процесс. На основе такого выборочного ис-

следования получают статистические характеристики или закономерности и распространяют их на всю систему в целом. При этом делаются соответствующие оговорки. Например, при получении статистических закономерностей их распространяют на поведение всей системы с некоторой доверительной вероятностью.

Подход к отображению объектов в виде диффузных систем широко применяется при: описании систем массового обслуживания, определении численности штатов на предприятиях и учреждениях, исследовании документальных потоков информации в системах управления и т. д.

Самоорганизующиеся системы. Отображение объекта в виде самоорганизующейся системы — это подход, позволяющий исследовать наименее изученные объекты и процессы. Самоорганизующиеся системы обладают признаками диффузных систем: стохастичностью поведения, нестационарностью отдельных параметров и процессов. К этому добавляются такие признаки, как непредсказуемость поведения, способность адаптироваться к изменяющимся условиям среды, изменять структуру при взаимодействии системы со средой, сохраняя при этом свойства целостности, способность формировать возможные варианты поведения и выбирать из них наилучший и др. Иногда этот класс разбивают на подклассы, выделяя адаптивные или самоприспосабливающиеся системы, самовосстанавливающиеся, самовоспроизводящиеся и другие подклассы, соответствующие различным свойствам развивающихся систем.

Примеры: биологические организации, коллективное поведение людей, организация управления на уровне предприятия, отрасли, государства в целом, т. е. в тех системах, где обязательно имеется человеческий фактор.

Классификация систем по сложности

Часто различают понятия «большая» и «сложная» система. Большая система или малая определяется количеством входящих в нее элементов, например:

малые системы ($10...10^3$ элементов),

большие ($10^4...10^7$ элементов),

очень большие ($10^7. ..10^{30}$ элементов),

супербольшие ($10^{30}.. .10^{200}$ элементов).

Является ли система сложной или простой определяется связями между элементами (см. табл. 1.1).

Четкой границы, отделяющей простые системы от сложных, нет. Деление это условное и возникло из-за появления систем, имеющих в своем составе совокупность подсистем с наличием функциональной избыточности. Простая система может находиться только в двух состояниях: состоянии работоспособности (исправном) и состоянии отказа (неисправном). При отказе элемента простая система либо полностью прекращает выполнение своей функции, либо

продолжает ее выполнение в полном объеме, если отказавший элемент резервирован. Сложная система при отказе отдельных элементов и даже целых подсистем не всегда теряет работоспособность, зачастую только снижаются характеристики ее эффективности. Это свойство сложных систем обусловлено их функциональной избыточностью и, в свою очередь, затрудняет формулировку понятия «отказ» системы.

Таблица 1.1. Примеры больших, малых, простых и сложных систем

	Сложные	Простые
Большие	Мозг, экономика, живой организм	Телефонный справочник, словарь, шифрозамок для вора
Малые	Неисправный бытовой прибор для пользователя	Исправный бытовой прибор, шифрозамок для хозяина

1.4. Методы описания систем

Методы описания систем классифицируются в порядке возрастания формализованности - от качественных методов, с которыми в основном и связан был первоначально системный анализ, до количественного моделирования с применением ЭВМ. Разделение методов на качественные и количественные носит, конечно, условный характер.

В **качественных методах** основное внимание уделяется организации постановки задачи, новому этапу ее формализации, формированию вариантов, выбору подхода к оценке вариантов, использованию опыта человека, его предпочтений, которые не всегда могут быть выражены в количественных оценках.

Количественные методы связаны с анализом вариантов, с их количественными характеристиками корректности, точности и т. п. Для постановки задачи эти методы не имеют средств, почти полностью оставляя осуществление этого этапа за человеком.

Между этими крайними классами имеются методы, которые стремятся охватить оба этапа — этап постановки задачи, разработки вариантов и этап оценки и количественного анализа вариантов,— но делают это с разной степенью формализованности.

Качественные методы описания систем

Качественные методы системного анализа применяются, когда отсутствуют описания закономерностей систем в виде аналитических зависимостей.

Методы типа мозговой атаки. Концепция «мозговой атаки» получила широкое распространение с начала 50-х годов прошлого века как метод систематической тренировки творческого мышления, нацеленный на открытие новых идей и достижение согласия группы людей на основе интуитивного мышления. Методы этого типа известны также под названиями «мозговой штурм»,

«конференция идей», а в последнее время наибольшее распространение получил термин «коллективная генерация идей» (КГИ).

Обычно при проведении мозговой атаки или сессий КГИ стараются выполнять определенные правила, суть которых:

- обеспечить как можно большую свободу мышления участников КГИ и высказывания ими новых идей;
- приветствуются любые идеи, находящиеся на стыке отраслей науки и техники;
- не допускается критика, не объявляется ложной и не прекращается обсуждение ни одной идеи;
- желательно высказывать как можно больше идей, особенно нетривиальных.

Подобием сессий КГИ можно считать разного рода совещания — **конструкторат**ы, заседания научных советов по проблемам, заседания специально создаваемых временных комиссий и другие собрания компетентных специалистов.

Методы типа сценариев. Методы подготовки и согласования представлений о проблеме или анализируемом объекте, изложенные в письменном виде, получили название сценария. Первоначально этот метод предполагал подготовку текста, содержащего логическую последовательность событий или возможные варианты решения проблемы, развернутые во времени. Однако позднее обязательное требование явно выраженных временных координат было снято, и сценарием стали называть любой документ, содержащий анализ рассматриваемой проблемы или предложения по ее решению, по развитию системы независимо от того, в какой форме он представлен. Примерами являются послания президента, предвыборные программы, аналитические записки.

Методы экспертных оценок. Термин «эксперт» происходит от латинского слова, означающего «опытный».

При использовании экспертных оценок обычно предполагается, что мнение группы экспертов надежнее, чем мнение отдельного эксперта, хотя это предположение не является очевидным.

При обработке материалов коллективной экспертной оценки используются методы корреляционного анализа. Для количественной оценки степени согласованности мнений экспертов применяется **коэффициент конкордации**

$$\mathbf{W = \frac{12d}{m^2(n^3 - n)}},\ \text{где}\quad \mathbf{d = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n}\left[\sum_{j=2}^{m} r_{ij} - 0.5m(n+1)\right]^2},$$

где **m** – количество экспертов, $j=\overline{\mathbf{1,m}}$, **n** – количество рассматриваемых объектов (или свойств объекта), $\mathbf{i}=\overline{1,\mathbf{n}}$ $\mathbf{r_{ij}}$ – место, которое занял **i**-й объект в ранжи-

ровке **j**-м экспертом; $\mathbf{d_i}$ – отклонение суммы рангов по **i**-му объекту от среднего арифметического сумм рангов по **n** объектам.

Коэффициент конкордации **W** позволяет оценить, насколько согласованы между собой ряды предпочтительности, построенные каждым экспертом. Его значение находится в пределах $\mathbf{0 \le W \le 1}$. $\mathbf{W = 0}$ означает полную противоположность, а $\mathbf{W = 1}$ – полное совпадение ранжировок. Практически достоверность считается хорошей, если $\mathbf{W = 0{,}7...0{,}8}$.

Методы типа «Дельфи». Характерный для середины XX века бурный рост науки и техники вызвал большие перемены в отношении к оценкам будущего развития систем. Одним из результатов этого периода в развитии методов анализа сложных систем явилась разработка методов экспертной оценки, известных в литературе как «методы Дельфи». Название этих методов связано с древнегреческим городом Дельфи, где при храме Аполлона с IX века до н.э. до IV века н.э. по преданиям существовал Дельфийский оракул.

Суть метода Дельфи заключается в следующем. В отличие от традиционного подхода к достижению согласованности мнений экспертов путем открытой дискуссии метод Дельфи предполагает полный отказ от коллективных обсуждений. Это делается для того, чтобы уменьшить влияние таких психологических факторов, как присоединение к мнению наиболее авторитетного специалиста, нежелание отказаться от публично выраженного мнения, следование за мнением большинства. В методе Дельфи прямые дебаты заменены тщательно разработанной программой последовательных индивидуальных опросов, проводимых обычно в форме анкетирования. Ответы экспертов обобщаются и вместе с новой дополнительной информацией поступают в распоряжение экспертов, после чего они уточняют свои первоначальные ответы. Такая процедура повторяется несколько раз до достижения приемлемой сходимости совокупности высказанных мнений. Результаты эксперимента показали приемлемую сходимость оценок экспертов после пяти туров опроса.

Метод Дельфи первоначально был предложен О. Хелмером как итеративная процедура при проведении мозговой атаки, которая должна помочь снизить влияние психологических факторов при проведении повторных заседаний и повысить объективность результатов. Однако почти одновременно Дельфи-процедуры стали основным средством повышения объективности экспертных опросов с использованием количественных оценок при оценке деревьев цели и при разработке сценариев.

Методы типа дерева целей (дерева задач). Идея метода дерева целей впервые была предложена Черчменом в связи с проблемами принятия решений в промышленности. Термин «дерево целей» подразумевает использование иерархической структуры, полученной путей разделения общей цели на подцели, а их, в свою очередь, на более детальные составляющие — новые подцели, функции и т. д. Как правило, этот термин используется для структур, имеющих отношение строгого древесного порядка, но метод дерева целей используется

иногда и применительно к «слабым» иерархиям в которых одна и та же вершина нижележащего уровня может быть одновременно подчинена двум или нескольким вершинам вышележащего уровня.

Древовидные иерархические структуры используются и при исследовании и совершенствовании организационных структур. Не всегда разрабатываемое даже для анализа целей дерево может быть представлено в терминах целей. Иногда, например, при анализе целей научных исследований удобнее говорить о дереве направлений прогнозирования. В. М. Глушковым, например, был предложен и в настоящее время широко используется термин «прогнозный граф». При использовании этого понятия появляется возможность более точно определить понятие дерева как связного ориентированного графа, не содержащего петель, каждая пара вершин которого соединяется единственной цепью.

Морфологические методы. Основная идея морфологических методов — систематически находить все «мыслимые» варианты решения проблемы или реализации системы путем комбинирования выделенных элементов или их признаков. Идеи морфологического образа мышления восходят к Аристотелю, Платону, к известной средневековой модели механизации мышления Р. Луллия. В систематизированном виде морфологический подход был разработан и применен впервые швейцарским астрономом Ф. Цвикки и долгое время был известен как метод Цвикки.

Цвикки предложил три метода морфологического исследования.

Первый метод — метод *систематического покрытия поля*, основанный на выделении так называемых опорных пунктов знания в любой исследуемой области и использовании для заполнения поля некоторых сформулированных принципов мышления. Второй — *метод отрицания и конструирования*, базирующийся на идее Цвикки, заключающейся в том, что на пути конструктивного прогресса стоят догмы и компромиссные ограничения, которые есть смысл отрицать, и, следовательно, сформулировав некоторые предложения, полезно заменить их затем на противоположные и использовать при проведении анализа.

Третий — *метод морфологического ящика*, нашедший наиболее широкое распространение. Идея морфологического ящика состоит в определении всех «мыслимых» параметров, от которых может зависеть решение проблемы, и представлении их в виде матриц-строк, а затем в определении в этом морфологическом матрицеящике всех возможных сочетаний параметров по одному из каждой строки. Полученные таким образом варианты могут затем подвергаться оценке и анализу с целью выбора наилучшего. Морфологический ящик может быть не только двумерным. Например, А. Холл использовал для исследования структуры систем трехмерный ящик.

Морфологические ящики Цвикки нашли широкое применение для анализа и разработки прогноза в технике. Для организационных же систем, систем управления такой ящик, который, по-видимому, был бы многомерным, практически невозможно построить.

Методика системного анализа. Методики, реализующие принципы системного анализа в конкретных условиях, направлены на то, чтобы формализовать процесс исследования системы, процесс постановки и решения проблемы. Методика системного анализа разрабатывается и применяется в тех случаях, когда у исследователя нет достаточных сведений о системе, которые позволили бы выбрать адекватный метод формализованного представления системы.

Общим для всех методик системного анализа является формирование вариантов представления системы (процесса решения задачи), выбор наилучшего варианта, корректировка. Положив в основу методики системного анализа эти три этапа, их затем можно разделить на подэтапы. Например, этапы проектирования систем представлены на рис. 1.2.

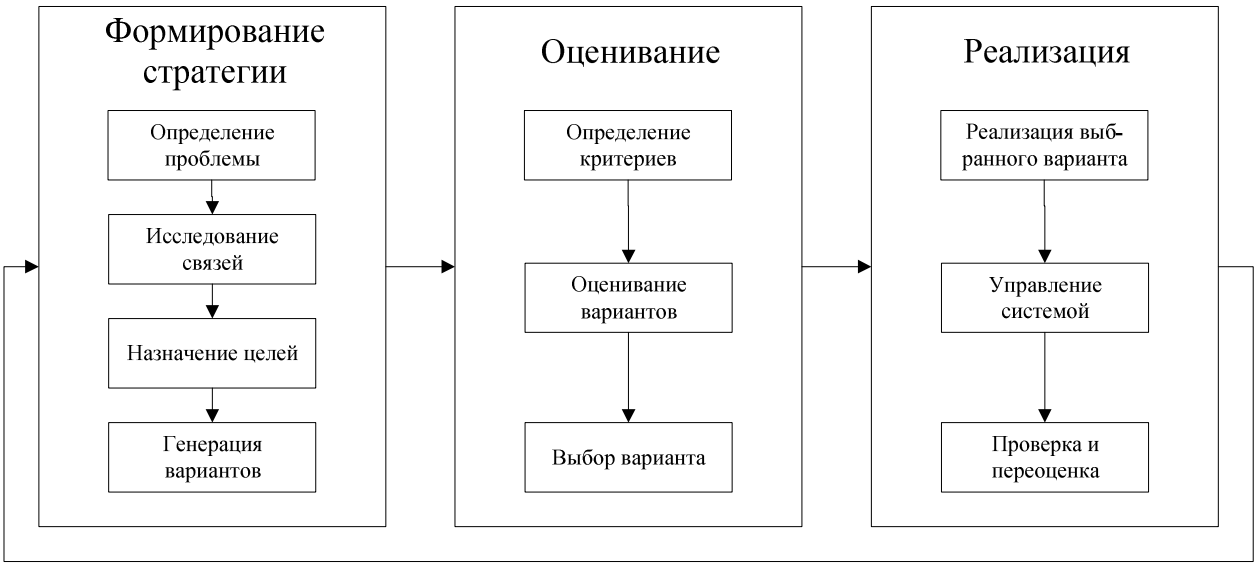


Рис. 1.2. Этапы проектирования систем

В настоящее время трудно привести примеры методик, в которых все этапы были бы проработаны равноценно.

Количественные методы описания систем

Уровни описания систем. При создании и эксплуатации сложных систем требуется проводить многочисленные исследования и расчеты, связанные с:

- оценкой показателей, характеризующих различные свойства систем;
- выбором оптимальной структуры системы;
- выбором оптимальных значений ее параметров.

Выполнение таких исследований возможно лишь при наличии математического описания процесса функционирования системы, т. е. ее математической модели.

Сложность реальных систем не позволяет строить для них «абсолютно» адекватные модели. Математическая модель описывает некоторый упрощенный процесс, в котором представлены лишь основные явления, входящие в реальный процесс, и лишь главные факторы, действующие на реальную систему.

Какие явления считать основными и какие факторы главными — существенно зависит от назначения модели, от того, какие исследования с ее помощью предполагается проводить. Поэтому процесс функционирования одного и того

же реального объекта может получить различные математические описания в зависимости от поставленной задачи.

Так как математических моделей сложной системы может быть сколько угодно много, и все они определяются принятым уровнем абстрагирования, то рассмотрение задач на каком-либо одном уровне абстракции позволяет дать ответы на определенную группу вопросов, а для получения ответов на другие вопросы необходимо провести исследование уже на другом уровне абстракции. Каждый из возможных уровней абстрагирования обладает ограниченными, присущими только данному уровню абстрагирования возможностями. Для достижения максимально возможной полноты сведений необходимо изучить одну и ту же систему на всех целей сообразных для данного случая уровнях абстракции.

Наиболее пригодными являются следующие подходы абстрактного описания систем:

- символический или, иначе, лингвистический;
- теоретико-множественный;
- топологический;
- логико-математический;
- теоретико-информационный;
- кибернетический;
- эвристический.

Лингвистический подход описания — наиболее высокий уровень абстрагирования. Из него как частные случаи можно получить другие уровни абстрактного описания систем более низкого ранга. Процесс формализации в математике обычно понимают как отвлечение от изменчивости рассматриваемого объекта. Поэтому формальные построения наиболее успешно используются, когда удается с предметами или процессами действительности каким-то образом сопоставлять некоторые стабильные, неизменные понятия. Используются инструменты формальной лингвистики: предикаты, термы, функторы.

При **теоретико-множественном подходе** абстракции можно получить только общие сведения о реальных системах, а для более конкретных целей необходимы другие модели, которые позволили бы производить более тонкий анализ различных свойств реальных систем. Эти более низкие уровни абстрагирования, в свою очередь, являются уже частными случаями по отношению к теоретико-множественному уровню формального описания систем.

Если же на элементах рассматриваемых множеств определены некоторые пространственные структуры, то в этом случае приходим к **топологическому подходу** абстрактного описания систем. При этом может быть использован язык общей топологии или ее ветвей, именуемых гомологической топологией, алгебраической топологией и т. д.

Логико-математический подход описания систем нашел широкое применение для: формализации функционирования автоматов; задания условий функционирования автоматов; изучения вычислительной способности автоматов.

При любом процессе управления или регулирования, осуществляемом живым организмом или автоматически действующей машиной либо устройством, происходит переработка входной информации в выходную. Поэтому при **теоретико-информационном подходе** абстрактного описания систем информация выступает как способ взаимодействия объектов и явлений, которые посредством отражения передаются от одного объекта к другому и запечатлеваются в его структуре (возможно, в измененном виде).

Отображение множества состояний источника во множество состояний носителя информации называется кодированием, а образ состояния при выбранном способе кодирования — кодом этого состояния.

Абстрагируясь от физической сущности носителей информации и рассматривая их как элементы некоторого абстрактного множества, а способ их расположения как отношение в этом множестве, приходят к абстрактному понятию кода информации как способа ее представления. При таком подходе любая система представляется в виде кодирующего устройства. Т.е. входы в систему представляются в виде исходной информации, а выходы – в виде информации, преобразованной по формальным правилам.

Кибернетический подход абстрактного описания систем связан с представлением системы как некоторого объекта, куда в определенные моменты времени можно вводить вещество, энергию и информацию, а в другие моменты времени — выводить их. Система представляется как объект управления, двигающийся к определенной цели.

Эвристический подход абстрактного описания систем предусматривает поиски удовлетворительного решения задач управления в связи с наличием в сложной системе человека. Эврика — это догадка, основанная на общем опыте решения родственных задач. Изучение интеллектуальной деятельности человека в процессе управления имеет очень большое значение. Моделируется человеческое мышление, интуиция.

Таким образом, обзор уровней абстрактного описания систем показывает, что выбор подходящего метода формального описания при изучении той или иной реальной системы является всегда наиболее ответственным и трудным шагом в теоретико-системных построениях. Эта часть исследования почти не поддастся формализации и во многом зависит от эрудиции исследователя, его профессиональной принадлежности, целей исследования и т. д. Наибольшее значение в настоящее время в абстрактной теории систем придается теоретико-множественному, теоретико-информационному и кибернетическому подходам описания систем.

1.5. Теоретико-множественный подход к описанию систем

Для получения математической модели процесса функционирования системы, чтобы она охватывала широкий класс реальных объектов, в общей теории систем исходят из общих предположений о характере функционирования системы:

1) система функционирует во времени; в каждый момент времени система может находиться в одном из возможных состояний;
2) на вход системы могут поступать входные сигналы;
3) система способна выдавать выходные сигналы;
4) состояние системы в данный момент времени определяется предыдущими состояниями и входными сигналами, поступившими в данный момент времени и ранее;
5) выходной сигнал в данный момент времени определяется состояниями системы, относящимися к данному и предшествующим моментам времени.

Первое предположение отражает динамический характер процесса функционирования в пространстве и времени. При этом процесс функционирования протекает как последовательная смена состояний системы под действием внешних и внутренних причин. 2-е и 3-е – отражают взаимодействие системы с внешней средой. В 4-м и 5-м предложениях отражается реакция системы на внутренние факторы и воздействия внешней среды: последействие и принцип физической реализуемости.

> **Последействие** – это тенденции, определяющие поведение системы в будущем, зависят не только от того, в каком состоянии находится система в настоящий момент времени, но и в той или иной степени от ее поведения в предыдущие моменты времени.
>
> **Принцип физической реализуемости**: система не реагирует в данный момент времени на «будущие» факторы и воздействия внешней среды.

Для описания систем ее подсистемы (или элементы) перечисляются с помощью некоторых множеств $\mathbf{V_i}$ и устанавливается характер связей между ними.

$\mathbf{S \subset \otimes \{V_i, i \in I\}}$, где

$\mathbf{V_i}$ – **i**-тая компонента декартова произведения $\otimes \mathbf{V_i}$, называемая объектом системы **S**, I-множество индексов. Или иначе:

$$\mathbf{S \subset V_1 \times V_2 \times V_3 \times ... \times V_m}$$

Абстрактно-алгебраические модели описывают связи как семейство отношений (унарных, бинарных ... **n**-арных)

$$\mathbf{R = \{R_1, R_2, ..., R_n\}}$$

> Под отношением, введенным на множестве **A,** понимается подмножество декартового произведения конечной степени $\mathbf{A^n = A \times A \times .... \times A}$

данного множества **A**, т.е. подмножество кортежей $(a_1, a_2, \ldots a_n)$ из **n** элементов множества **A**.

Подмножество $R \subset A^n$ называется **n**-местным или **n**-арным отношением на множестве **A**. Число **n** называется рангом или типом отношения **R**. Множество всех **n**-арных отношений на множестве **A** относительно операций $\cup$ и $\cap$ является булевой алгеброй.

Примеры отношений на множестве **V** «Люди»**:**

унарное отношение **M** «мужчины»: $M=\{v \in V \mid \text{пол}(v)=\text{мужской}\}$;

бинарное отношение **O** «старше»:

$O=\{(v_1,v_2) \subset V^2 \mid \text{возраст}(v_1)>\text{возраст}(v_2)\}$;

трехместное отношение **P** «являются родителями»: $P=\{(v_1,v_2,v_3) \subset V^3$ **, где** v_1 **– отец,** v_2 **– мать,** v_3 **– ребенок}**.

Функциональные модели определят связи как множество отображений.

Если множество индексов **I** конечно, то разобьем его два подмножества I_u и I_y. В общем случае пересечение этих подмножеств может быть не пусто. $I_u \subset I$ и $I_y \subset I$. Множество $U=\otimes\{V_i \mid i \subset I_u\}$ назовем **причинами**, а множество $Y=\otimes\{V_i \mid i \subset I_y\}$ назовем **следствиями**. Тогда система $S \subset U \times Y$. Система **S** называется функциональной, если она представляется в виде отображения $S: U \to Y$.

Временны*е* модели в качестве одного из объектов системы **S** вводят множество моментов времени **T.**

Если элементы одного из объектов системы есть функции, например $\nu: T_\nu \to V_\nu$, то этот объект называют **функциональным**. В случае, когда области определения всех функций для данного объекта **V** одинаковы, т.е. каждая функция отображает **T** в **V**, $\nu: T \to V$, то **T** называется **индексирующим множеством** для ν. Если индексирующее множество линейно-упорядочено, то его называют **множеством моментов времени**. Функции, определенные на множестве моментов времени, принято называть функциями времени. Объект, элементами которого являются временные функции, называют **временным объектом**, а системы определенные на временных объектах – **временными системами**.

1.6. Кибернетический подход к описанию систем

Управление как процесс. Кибернетический подход к описанию систем состоит в том, что всякое целенаправленное поведение рассматривается как управление. Управление — в широком, кибернетическом смысле — это обобщение приемов и методов, накопленных разными науками об управлении искусственными объектами и живыми организмами. Язык управления — это использование понятий «объект», «среда», «обратная связь», «алгоритм» и т.д. Основы современной кибернетики заложил Н. Винер.

Под **управлением** будем понимать процесс организации такого целенаправленного воздействия на некоторую часть среды, называемую объектом управления, в результате которого удовлетворяются потребности субъекта, взаимодействующего с этим объектом.

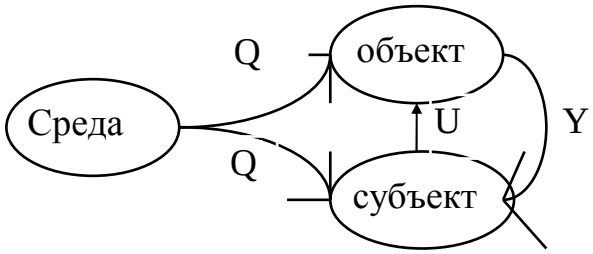


Рис. 1.3. Кибернетический подход к процессу управления

Анализ управления заставляет выделить тройку — среду, объект и субъект, внутри которой разыгрывается процесс управления (рис. 1.3). В данном случае субъект ощущает на себе воздействие среды **Q** и объекта **Y**. Причем, если состояние среды **Q** он изменить не может, то состоянием объекта **Y** он может управлять с помощью специально организованного воздействия **U**. Это воздействие и есть управление.

Состояние объекта **Y** влияет на состояние потребностей субъекта. Потребности субъекта $\mathbf{A = (a_1,\ldots,a_k)}$, где $\mathbf{a_i}$ — состояние **i**-й потребности субъекта, которая выражается неотрицательным числом, характеризующим насущность, актуальность этой потребности. Свое поведение субъект строит так, чтобы минимизировать насущность своих потребностей, т. е. решает задачу многокритериальной оптимизации:

$$\mathbf{a_i \rightarrow min\ \{R\},\ \forall (i = \overline{1,k})}, \tag{1.1}$$

где **R** — ресурсы субъекта. Эта зависимость выражает неизвестную, но существующую связь потребностей с состоянием среды **Q** и поведением **U** субъекта.

Пусть $\mathbf{U_q^*}$ — решение задачи (1.1), т. е. оптимальное поведение субъекта, минимизирующее его потребности **A**. Способ решения задачи (1.1), позволяющий определить $\mathbf{U_q^*}$, называется алгоритмом управления

$$\mathbf{U_q^* = \varphi(A_t, Q)}, \tag{1.2}$$

где $\boldsymbol{\varphi}$ — алгоритм, позволяющий синтезировать управление по состоянию среды **Q** и потребностей $\mathbf{A_t}$. Потребности субъекта изменяются не только под влиянием среды или объекта, но и самостоятельно, отражая жизнедеятельность субъекта, что отмечается индексом **t**.

Алгоритм управления **φ,** которым располагает субъект, и определяет эффективность его функционирования в данной среде. Обычно алгоритм имеет рекуррентный характер:

$$\mathbf{U}_{N+1} = \varphi(\mathbf{U}_N, \mathbf{A}_t, \mathbf{Q}),$$

т. е. позволяет на каждом шаге улучшать управление. Например, в смысле $\mathbf{A}_t(\mathbf{Q}, \mathbf{U}_{N+1}) < \mathbf{A}_t(\mathbf{Q}, \mathbf{U}_N)$, т. е. уменьшения уровня своих потребностей. Впрочем, потребности еще надо осознать, и это составляет пока неформализуемую стадию управления.

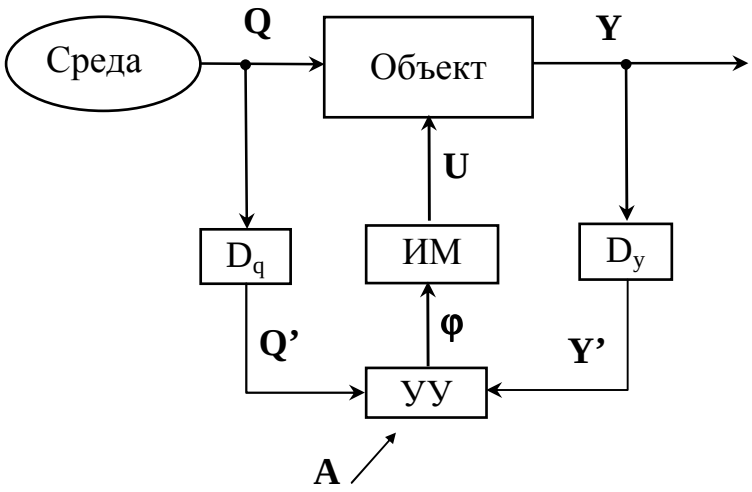


Рис. 1.4. Структурная схема системы управления

Структурная схема системы управления приведена на рис. 1.4. Здесь $\mathbf{D_q}$ и $\mathbf{D_y}$ — датчики, измеряющие состояние среды и объекта соответственно. Результаты измерений $\mathbf{Q'=D_q(Q)}$ и $\mathbf{Y'=D_y(Y)}$ образуют исходную информацию **{Q', Y'}** для устройства управления (УУ), которое на ее основе вырабатывает команду управления **φ**. Исполнительный механизм (ИМ), реализуя команду управления, вырабатывает входное воздействие **U** на управляемые входы объекта.

Управление — целенаправленная организация того или иного процесса, протекающего в системе. В общем случае процесс управления состоит из следующих четырех элементов:

- получение информации о задачах управления (**A**),
- получение информации о результатах управления (т. е. о поведении объекта управления **Y'**);
- анализ полученной информации и выработка решения **(φ=φ(A,Q',У'))**,
- исполнение решения (т. е. осуществление управляющих воздействий U).

Процесс управления — это информационный процесс, заключающийся в сборе информации о ходе процесса, передаче ее в пункты накопления и переработки, анализе поступающей, накопленной и справочной информации, принятии решения на основе выполненного анализа, выработке соответствующего управляющего воздействия и доведении его до объекта управления.

Каждая фаза процесса управления протекает во взаимодействии с окружающей средой при воздействии различного рода помех. Цели, принципы и границы управления зависят от сущности решаемой задачи.

Система управления — совокупность взаимодействующих между собой объекта управления и органа управления, деятельность которых направлена заданной цели управления.

В системах управления решаются четыре основные задачи управления: стабилизация, выполнение программы, слежение, оптимизация.

Задачами **стабилизации** системы являются задачи поддержания выходной величины **Y** вблизи некоторого заранее заданного значения, несмотря на действие помех.

Например, стабилизация напряжения и частоты тока в электрической сети вне зависимости от изменения потребления энергии.

Задача **выполнения программы** возникает в случаях, когда требуется изменять величину **U** во времени заранее известным образом.

Например, вывод космического аппарата на орбиту, который выполняется по заранее заданному алгоритму.

В системах **оптимального управления** требуется формировать входное воздействие **U** так, чтобы экстремизировать выход объекта управления **Y**. при заданных реальных условиях и ограничениях. Понятие оптимальности должно быть конкретизировано для каждого отдельного случая.

Например, переезд транспортного средства из одной точки в другую с минимальными затратами топлива, управление корпорацией с целью достижения максимальной прибыли.

Задачи **слежения** означают, что величину **Y** требуется поддерживать на уровне, определяемом неконтролируемым воздействием среды **Q**.

Например, слежение радаров за передвижением летательных аппаратов, наведение ракеты на цель.

Системы управления делятся на два больших класса: системы автоматического управления (САУ) и автоматизированные системы управления (АСУ). В САУ управление объектом или системой осуществляется без непосредственного участия человека автоматическими устройствами. Это замкнутые системы. Основные функции САУ: автоматический контроль и измерения, автоматическая сигнализация, автоматическая защита, автоматические пуск и остановка различных двигателей и приводов, автоматическое поддержание заданных режимов работы оборудования, автоматическое регулирование. В отличие от САУ в АСУ в контур управле-

ния включен человек, на которого возлагаются функции принятия наиболее важных решений и ответственности за принятые решения. Под АСУ обычно понимают человеко-машинные системы, использующие современные экономико-математические методы, средства электронно-вычислительной техники и связи, а также новые организационные принципы для отыскания и реализации на практике наиболее эффективного управления объектом (системой).

1.7. Агрегативное описание систем

Понятие «агрегат» в теории систем

Пусть **T** – фиксированное подмножество действительных чисел (множество рассматриваемых моментов времени), **Z, U, Y, X** – множества любой природы. Элементы указанных множеств назовем, как и ранее: $t \in T$ – моментом времени; $u \in U$; входным сигналом; $z \in Z$ - управляющим сигналом; $y \in Y$ – выходным сигналом; $x \in X$ состоянием. Разница между входным сигналом и управляющим состоит в следующем – управляющий сигнал влияет на выходной непосредственно, а входной сигнал – опосредованно через состояние. Состояния, входные, управляющие и выходные сигналы, рассматриваемые как функции времени, обозначим **x(t), u(t), z(t)** и **y(t).**

> Под **агрегатом** будем понимать объект **<T,U,Z,Y,X,H,G>**, где **H,G** – операторы (вообще говоря, случайные). Операторы переходов и выходов **H** и **G** реализуют функции **x(t)** и **y(t)** и представляют собой обобщение переходного отображения σ и функции наблюдения η. Структура этих операторов собственно и выделяет агрегаты среди прочих систем.

Предположение 1. Будем предполагать, что за конечный интервал времени в агрегат поступает конечное число входных и управляющих сигналов и вырабатывается конечное число выходных сигналов.

Операторы переходов и выходов агрегата

Операторы переходов. Наряду с состоянием **x(t)** будем рассматривать также точки **x(t+0).** Договоримся считать, что для любого $t_1 > t$ момент $(t+0) \in (t,t_1]$. Аналогично: **x(t-0)** означает, что $\forall\ t_0 < t: (t-0) \in [t_0,t)$. Вид оператора **H** зависит от того, содержит ли рассматриваемый интервал времени моменты т.н. **особых состояний** агрегата или не содержит.

> Под **особыми состояниями** будем понимать его состояния в момент получения входного либо управляющего сигналов или выдачи выходного сигнала. Все остальные состояния агрегата будем называть **неособыми**.

Предположение 2. Из особых состояний агрегат может переходить в новое состояние скачком.

Пусть **x(t*)** – некоторое особое состояние агрегата, а z_s – последний управляющий сигнал $z_s \in Z$. Примем следующие обозначения для операторов,

являющихся частными видами оператора **H** и определяющих состояние агрегата в момент **t*+0**. Если **t*** - момент поступления входного сигнала **u**, то

$$\mathbf{x(t^*+0)=V'[x(t^*), u, z_s]} \ . \qquad (1.3)$$

Аналогично, если **t*** - момент поступления управляющего сигнала **z**, то

$$\mathbf{x(t^*+0)=V''[x(t^*), z]} \ . \qquad (1.4)$$

При одновременном поступлении **u** и **z**

$$\mathbf{x(t^*+0)=V\,[x(t^*), u, z]} \qquad (1.5)$$

Наконец, если **t*** - момент выдачи выходного сигнала **y**, то

$$\mathbf{x(t^*+0)=W[x(t^*), z_s]} \ . \qquad (1.6)$$

В интервале между особыми состояниями, значение **x(t)** определяется при помощи операторов **Q**, вид которых в общем случае зависит от особого состояния, являющегося для данного интервала времени начальным состоянием:

$$\mathbf{x(t+0)=Q_{t^*}[t,x(t), z_s].} \qquad (1.7)$$

Здесь **t*** - момент особого состояния, являющегося исходным для данного интервала времени. То есть **H** является общим обозначением операторов **Q,V′,V″,V** и **W**.

Оператор выходов. Во множестве **X** состояний **x(t)** агрегата выделим подмножество $\mathbf{X_y}$ (подмножество состояний, влекущих за собой необходимость выдачи выходного сигнала), обладающее следующими свойствами. Выходной сигнал **y** выдается в момент **t′** в двух случаях, когда:

1) $\mathbf{x(t')\in X_y;\ x(t'-0)\notin X_y}$

2) $\mathbf{x(t'+0)\in X_y}$**, но** $\mathbf{x(t')\notin X_y}$.

Тогда, оператор **G** можно представить в виде совокупности двух операторов: функционального оператора **G′**, вырабатывающего выходной сигнал

$$\mathbf{y(t')=G'[x(t'),z_s]} \qquad (1.8)$$

и логического оператора **G″**, проверяющего для каждого **t** принадлежность **x(t)** к подмножеству $\mathbf{X_y}$. Заметим, что в общем случае, оператор **G′** является случайным оператором. Это значит, что данным **t, x(t), z** ставится в соответствие не одно определенное значение выходного сигнала, а некоторое множество значений **y** с распределением вероятностей, задаваемых оператором **G′**.

Например, в качестве одной составляющих вектора **x(t)** (например $\mathbf{x_1(t)}$) может быть время, оставшееся до выдачи выходного сигнала. Тогда оператор **G″** проверяет неравенство $\mathbf{x_1(t)>0}$.

Процесс функционирования агрегата. Агрегат функционирует следующим образом. В начальный момент времени $\mathbf{t_0}$ заданы начальное состояние агрегата $\mathbf{x_0}$ и начальное значение управляющего сигнала $\mathbf{z_0}$.

Пусть $\mathbf{t_1}$ и $\mathbf{t_2}$ – моменты поступления первого $\mathbf{u_1}$ и второго $\mathbf{u_2}$ входных сигналов, τ_1 – момент поступления первого управляющего сигнала $\mathbf{z_1}$ и, для определенности $\mathbf{t_1<\tau_1<t_2}$, $\mathbf{t'}$ – момент выдачи первого выходного сигнала, причем пусть $\mathbf{t'<t_1}$ (см. рис. 1.5).

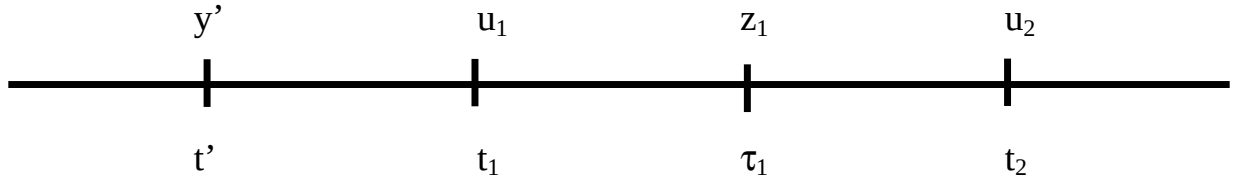


Рис. 1.5. Пример наступления событий в агрегате

Рассмотрим полуинтервал $\mathbf{(t_0,t']}$. Состояния агрегата изменяются с течением времени по закону (1.7):

$$\mathbf{x(t)=Q_{t_0}[t,x_0,z_0],\ (t_0<t\le t')}$$

до тех пор (оператор $\mathbf{G''}$), пока в момент $\mathbf{t'}$ состояние $\mathbf{x(t')}$ не окажется принадлежащим подмножеству $\mathbf{X_y}$, хотя состояние $\mathbf{x(t'-0)}$ не принадлежало подмножеству $\mathbf{X_y}$. В этом случае в момент $\mathbf{t'}$ выдается выходной сигнал $\mathbf{y'}$, вырабатываемый оператором $\mathbf{G'}$. Вместе с тем закон изменения состояний (1.7) нарушается и переключается на закон (1.6):

$$\mathbf{x(t'+0)=W[x(t'),z_0].}$$

Пусть теперь в момент $\mathbf{t_1}$ поступает входной сигнал $\mathbf{u_1}$. Проследим поведение агрегата в момент $\mathbf{t_1}$ (см. закон (1.3)). Тогда в силу действия входного сигнала $\mathbf{u_1}$ состояние агрегата имеет вид

$$\mathbf{x(t_1+0)=V'[x(t_1),u_1,z_0],} \qquad (1.9)$$

а в дальнейшем, если состояние (1.9) не соответствует выдаче выходного сигнала, то:

$$\mathbf{x(t)=Q_{t_1}[t,V'[x(t_1),u_1,z_0],z_0]\ \ t_1<t\le\tau_1}$$

Пусть в момент $\boldsymbol{\tau_1}$ в агрегат поступает управляющий сигнал $\mathbf{z_1}$. Тогда состояние агрегата имеет вид (см. закон (1.4)).

$$\mathbf{x(\tau_1+0)=V''[x(\tau_1),z_1],}$$

Необходимо отметить, что управляющий сигнал $\mathbf{z}$ в общем случае является параметром, определяющим операторы $\mathbf{V', V'', W, Q, G', G''.}$ Поэтому в дальнейшем вместо начального значения управляющего сигнала $\mathbf{z_0}$ в этих операторах должно использоваться значение $\mathbf{z_1}$ до тех пор, пока не поступит сле-

дующий управляющий сигнал $\mathbf{z}_2$. Например, в полуинтервале $(\tau_1, t_2]$, если нет оснований для выдачи выходного сигнала

$$\mathbf{x(t)= Q}_{\tau_1}\mathbf{[t, x(\tau_1+0), z_1]} \quad \tau_1<t\le t_2$$

В частном случае, операторы **H** и **G** могут оставаться неизменными при поступлении очередного управляющего сигнала. Аналогично, оператор **Q** может быть одним и тем же при любых выходных сигналах (при попадании **x(t)** в любые подмножества $\mathbf{X}_y$).

Агрегат представляет собой математическую схему весьма общего вида, частными случаями которой являются функции алгебры логики, релейно-контактные схемы, конечные автоматы, всевозможные классы систем массового обслуживания, динамические системы, описываемые обыкновенными дифференциальными уравнениями и некоторые другие объекты. С точки зрения моделирования агрегат выступает как достаточно универсальный переработчик информации – он воспринимает входные и управляющие сигналы и выдает выходные сигналы.

Кусочно-линейные агрегаты

Как показывает анализ моделирующих алгоритмов, можно добиться их существенных упрощений, если рассматривать объекты более частные, чем агрегат общего вида, но сохраняющие возможность описания достаточно широкого класса реальных систем. Практически удобным для формализации широкой совокупности разнообразных процессов и явлений материального мира являются так называемые **кусочно-линейные агрегаты** (КЛА).

Понятие о кусочно-линейном агрегате. Для поставленных здесь задач достаточно считать, что на агрегат не поступают управляющие сигналы **z**, а поступают лишь входные сигналы **u** (это допущение не ограничивает общности, так как в качестве **u** можно рассматривать входной сигнал в широком смысле, в том числе и управляющий). Итак, мы рассматриваем агрегат как объект, который в каждый момент времени **t** характеризуется внутренним состоянием **x(t),** имеет вход и выход. На вход агрегат в изолированные моменты времени могут поступать сигналы, с выхода могут сниматься выходные сигналы. Класс кусочно-линейных агрегатов выделяется с помощью конкретизации структуры множеств **X, U, Y** а также операторов **H** и **G**, которые представляют собой линейные пространства. Опишем данную конкретизацию.

Рассмотрим некоторое конечное или счетное множество **I**. Для определенности предположим, что **I = {0,1,2,…},** хотя в конкретных задачах **I** может иметь и другой вид. Назовем **I** множеством особых состояний, а элементы $\nu\in\mathbf{I}$ – особыми состояниями. Каждому особому состоянию $\nu\in\mathbf{I}$ поставим в соответствие некоторое целое неотрицательное число $||\nu||$, которое назовем рангом основного состояния (смысл этой величины – размерность вектора ν-го состояния). Кроме того, каждому $\nu\in\mathbf{I}$ поставим в соответствие выпуклый многогран-

ник $\mathbf{X^{(\nu)}}$ (множество допустимых значений для состояния ν) в евклидовом пространстве размерности $||\nu||$. Будем считать, что $\mathbf{X = \cup X^{(\nu)}}$, т. е. пространство состояний $\mathbf{X}$ можно представить состоящим из всевозможных пар вида $(\nu, \mathbf{x}^{(\nu)})$, где $\nu \in \mathbf{I}$, а $\mathbf{x^{(\nu)}}$ является вектором размерности $||\nu||$ и принимает значения из многогранника $\mathbf{X^{(\nu)}}$. Вектор $\mathbf{x^{(\nu)}}$ будем называть вектором координат. Если $||\nu|| = 0$ для некоторого ν, то это означает, что в данном состоянии ν координаты не определяются.

Процесс функционирования КЛА. Опишем сначала динамику КЛА, т.е. процесс изменения внутренних состояний во времени, в предположении отсутствия поступления $\mathbf{u}$. В предыдущей терминологии, определим действия оператора $\mathbf{Q}$. Пусть в начальный момент времени $\mathbf{t_0}$ агрегат находится в состоянии $\mathbf{x(t_0)=(\nu, x^{(\nu)}(0))}$, где $\mathbf{x^{(\nu)}(0)}$ - внутренняя точка многогранника $\mathbf{X^{(\nu)}}$. Тогда при $\mathbf{t>t_0}$ точка $\mathbf{x^{(\nu)}(t)}$ перемещается внутри многогранника $\mathbf{X^{(\nu)}}$ до тех пор, пока не достигнет его границы. Пусть это произойдет в момент $\mathbf{t_1}$, который назовем «особым». Тогда при $\mathbf{t_0<t\le t_1}$ «движение» агрегата описывается следующими законами:

$$\mathbf{\nu(t)=\nu=const} \tag{1.10}$$

$$\mathbf{x^{(\nu)}(t)=x^{(\nu)}(0)+ (t-t_0)\cdot\alpha^{(\nu)}.} \tag{1.11}$$

Данному значению ν соответствует вектор $\boldsymbol{\alpha}^{(\nu)}$ (скорость изменения координат) размерности $||\nu||$ (ср. (1.7)).

Значение особого момента $\mathbf{t_1}$ определяется траекторией $\mathbf{x(t)}$, вернее её некоторыми параметрами и может быть найдено из соотношения

$$\mathbf{t_1=\inf\{t:\ x^{(\nu)}(0)+(t-t_0)\alpha^{(\nu)} \notin X^{(\nu)},\ t>t_0\}.} \tag{1.12}$$

Поскольку $\mathbf{X^{(\nu)}}$ – многогранник, то нахождение $\mathbf{t_1}$ по правилу (1.12) сводится к следующему. Предположим, что $\mathbf{X^{(\nu)}}$ содержит $\mathbf{m}$ граней. Эти грани могут быть заданы линейными уравнениями:

$$\sum_{i=1}^{||\nu||} \mathbf{d_{ji}^{(\nu)} x_i^{(\nu)} + d_{j0}^{(\nu)} = 0}, \quad \mathbf{j=1,\ldots, m},$$

где $\mathbf{x_i^{(\nu)}}$ – компоненты вектора $\mathbf{x^{(\nu)}}$, $\mathbf{i=1..\ ||\nu||}$. Легко понять, что (1.12) может быть записано в виде

$$\mathbf{t_1 = \min_j\{t : t > t_0, \sum_{i=1}^{||\nu||} d_{ji}^{(\nu)}\ (x_i^{(\nu)}(0) + (t - t_0)\alpha_i^{(\nu)}) + d_{jo}^{(\nu)} = 0\}} \tag{1.13}$$

Обозначим $\boldsymbol{\tau_j} = -\dfrac{\mathbf{d_{jo}^{(\nu)} + \sum_{i=1}^{||\nu||} d_{ji}^{(\nu)} x_i^{(\nu)}(0)}}{\mathbf{\sum_{i=1}^{||\nu||} d_{ji}^{(\nu)} \alpha_i^{(\nu)}}}$, $\mathbf{j=1,\ldots m}$ (1.14)

Пусть $\tau=\min\{\tau_j;\tau_j>0\}$ (1.15)

Тогда из (1.13)-(1.15) следует, что $t_1=t_0+\tau$.

То есть τ – это время, за которое агрегат может достичь ближайшей грани многогранника и прийти к смене состояния, а t_1 – ближайший особый момент времени.

В момент t_1 состояние рассматриваемого кусочно-линейного агрегата изменяется скачкообразно. Значение $x(t_1+0)$ является, вообще говоря, случайным. В момент t_1 может выдаваться выходной сигнал **y** (см. оператор **G**). Содержание (и необходимость выдачи) **y** зависит от состояния $x(t_1)$. Подмножество X_y, введенное в общем определении агрегата, в данном случае совпадает с $\bigcup_{\nu=0}^{\infty}\bigcup_{j=1}^{m} X_j^{(\nu)}$. Важно указать, что множество **Y** имеет структуру, аналогичную **X**, т.е. выходные сигналы **y** представляются как $y=(\lambda, y^{(\lambda)})$, где λ – элемент некоторого счетного множества, $y^{(\lambda)}$ – вектор, принимающий значения из евклидова пространства размерностью, зависящей от λ. При $t>t_1$ движение агрегата вновь происходит в соответствии с формулами (1.10) и (1.11) до очередного особого момента t_2, где вместо t_0 теперь нужно понимать t_1 и т.д.

Обратимся теперь к случаю поступления входного сигнала. Подчеркнем, что для КЛА множество **U** структурно аналогично множествам **X** и **Y**, т.е. $u=(\mu, u^{(\mu)})$, где μ – элемент конечного или счетного множества, а $u^{(\mu)}$ – вектор, размерность которого зависит от μ. Следующее описание поведения КЛА можно рассматривать как раскрытие действия оператора V'.

Пусть в рассматриваемый момент **t** состояние агрегата $x(t)=(\nu, x^{(\nu)})$ и пусть в этот момент поступает входной сигнал $u=(\mu, u^{(\mu)})$. При этом состояние агрегата меняется скачкообразно. Значение $x(t+0)$ является случайным, задаваемым распределением, которое, вообще говоря, зависит от $x(t)$ и **u**. Будем считать, что в рассматриваемый момент может выдаваться выходной сигнал, содержание и необходимость выдачи которого зависит не только от состояния $x(t)$ (и, быть может, $x(t+0)$), но и от содержания поступившего входного сигнала **u**. После рассматриваемого момента времени **t** движение агрегата происходит в соответствии с формулами (1.10) и (1.11) до следующего момента поступления входного сигнала или выхода вектора состояния на границу допустимых значений.

В виде КЛА могут быть формализованы многие реальные процессы: процессы передачи и обмена данными в сетях связи, системы массового обслуживания и материально-технического снабжения, процессы автомобильного движения на дорогах, разнообразные дискретные производственные процессы, вычислительные системы и т.д. При этом всюду основные состояния агрегата указывают на качественно различные состояния моделируемых объектов. Дополнительные же координаты характеризуют происходящие количественные изме-

нения и часто носят сугубо вспомогательный характер, «вбирая» в себя необходимую информацию о предыстории модели. Следует отметить, что представление реальных систем в форме КЛА неоднозначно, поскольку неоднозначно могут быть выбраны состояния агрегатов. Выбор же состояний определяется как целями исследования, так и стремлением уменьшить размерность задачи. При этом всегда приходится идти на компромисс между точностью описания и полнотой получаемой информации с одной стороны и простотой модели – с другой.

Вероятностный автомат Мура. Этот автомат не имеет «жесткой» тактности, а изменяет свое состояние всякий раз, когда поступает входной сигнал. Пусть **X** – конечное множество внутренних состояний автомата, **U** – его входной (конечный) алфавит, **Y** – его выходной (конечный) алфавит. Для определенности будем считать, что

X={1,2,…N}, U={1,2,..K}, Y={1,2,..M}.

Динамика автомата описывается следующим образом. Если в момент **t** состояние автомата **x(t)=i** и поступил входной сигнал, **u(t)=k**, то состояние **x(t+0)=j** выбирается случайно с вероятностью $\mathbf{p_{ij}^{(k)}}$, $\mathbf{p_{ij}^{(k)} \geq 0}$, $\sum_{\mathbf{j=1}}^{\mathbf{N}} \mathbf{p_{ij}^{(k)}} = \mathbf{1}$

при любом **k**, **1≤k≤K**. Выходной сигнал **y∈Y**, выдаваемый в этот момент, является однозначной функцией «нового» состояния **j: y=m=Φ(j),** где **Φ** - некоторая детерминированная функция с областью определения **X** и множеством значений **Y**. Представим этот автомат в виде кусочно-линейного агрегата.

Состояние агрегата **x(t) ∈ X,** не имеет дополнительных координат и определяется только его номером **ν**. Следовательно, **||ν|| = 0** при всех **ν∈X**. При таком выборе состояния КЛА не определяются многогранники $\mathbf{X^{(\nu)}}$**,** отпадают вопросы о движении внутри многогранника, выходе агрегата на границу.

Все движение рассматриваемого КЛА состоит из скачков состояния при поступлении входных сигналов, причем ввиду отсутствия вектора координат речь идет лишь о скачках основного состояния **ν**. Если в момент времени **t*** состояние агрегата было **x(t*) = i** и поступил входной сигнал **u = k**, то в следующий момент времени состояние изменилось: **x(t*+0)=j** с вероятностью $\mathbf{p_{ij}^{(k)}}$.

Таким образом, требуется задать лишь распределение $\mathbf{p_{ij}^{(k)}}$. Содержание же выходного сигнала, выдаваемого в момент поступления входного, определяется только функцией **Φ: y(t*+0)=Φ(j)**. Вообще, если предположить, что **||ν||=0, ||λ||=0, ||μ||=0** ∀ **ν, α, μ,** то легко видеть, что КЛА превращается в вероятностный автомат весьма общего вида.

Система массового обслуживания. Пусть на обслуживающий прибор поступает ординарный поток требований, причем **i**-е требование характеризуется параметром θ_i, который представляет собой предельно до-

пустимое время ожидания **i**-м требованием начала обслуживания. Заявки, для которых реальное время ожидания превышает допустимое θ_i , получают отказ. Время обслуживания **i**-того требования равно ζ_i, причем $\{\zeta_i\}$ – последовательность независимых одинаково распределенных случайных величин. Искомыми являются вероятностные характеристики длины очереди и времени ожидания.

Представим данный процесс в виде КЛА. За состояние агрегата выберем вектор $\mathbf{x=(\nu, x^{(\nu)})}$, где ν - число требований, находящихся в системе в текущий момент времени, $\|\nu\| = \nu$, а $\mathbf{x^{(\nu)}}$ – вектор, координаты которого определяются только при $\nu>0$ и имеют следующий смысл: $\mathbf{x_1^{(\nu)}}$ – время, оставшееся до окончания обслуживания требования, находящегося на приборе, а $\mathbf{x_i^{(\nu)}}$ $\mathbf{(1<i\le\nu)}$ – длительности обслуживания требований, которые стоят в очереди и будут впоследствии обслужены. В соответствие с этим вектор $\alpha^{(\nu)}$ скоростей изменения дополнительных координат имеет компоненты $\alpha_1^{(\nu)}=-1$, $\alpha_i^{(\nu)}=0$ $\mathbf{(1<i\le\nu)}$, многогранник $\mathbf{X^{(\nu)}}$ при каждом $\nu>0$ совпадает с первым октантом евклидова пространства размерности ν: $\mathbf{X^{(\nu)}=\{x^{(\nu)}: x_i^{(\nu)}\ge 0, i=1,2,..\ \nu\}}$

В качестве входного сигнала рассмотрим пару $\mathbf{(1,\theta)}$, где символ **1** просто указывает на факт поступления требования (и, таким образом, дискретная компонента μ принимает лишь одно значение **1**), а величина θ равна допустимому времени ожидания поступающего требования (т.е. $\|\mu\|=1$).

Рассмотрим динамику данного агрегата (рис.1.6). Между особыми моментами времени гладко изменяется лишь первая компонента вектора координат, остальные – неизменны. $\boldsymbol{x_1(t) = x_1(t_0) - (t-t_0)}$ (см. формулу (1.11), отсюда $\alpha_1^{(\nu)}=-1$). На рис. 1.6 – это движение от точки **t** до **t***.

Пусть момент **t*** является моментом окончания обслуживания. Тогда с необходимостью $\boldsymbol{x_1(t^*) = 0,\ t^* = x_1(t_0) + t_0}$ (на рис. 1.6 – точка **t***), а $\mathbf{x(t^*) = (\nu, 0, x_2^{(\nu)},..x_\nu^{(\nu)}),\ где\ \nu>0}$.

Обслуженное требование должно покинуть систему, а его место занимает требование стоящее первым в очереди (если очередь не пуста). Следовательно, из состояния $\mathbf{x(t^*)}$ агрегат скачкообразно перейдет в состояние $\mathbf{x(t^*+0) = (\nu-1, x_2^{(\nu)},..x_\nu^{(\nu)}),}$ (на рис.1.6 – точка **t*+0**). При этом размерность вектора **x** уменьшилась на **1**, а компонента $\mathbf{x_2^{(\nu)}}$ заняла место компоненты $\mathbf{x_1^{(\nu)}}$. Будем считать, что в рассматриваемый момент **t*** выходной сигнал не выдается. Далее происходит обслуживание очередной заявки, что выражается в непрерывном уменьшении компоненты $\mathbf{x_1^{(\nu)}}$, а на рис. 1.6 отображается движением от точки **t*+0** до точки **t****.

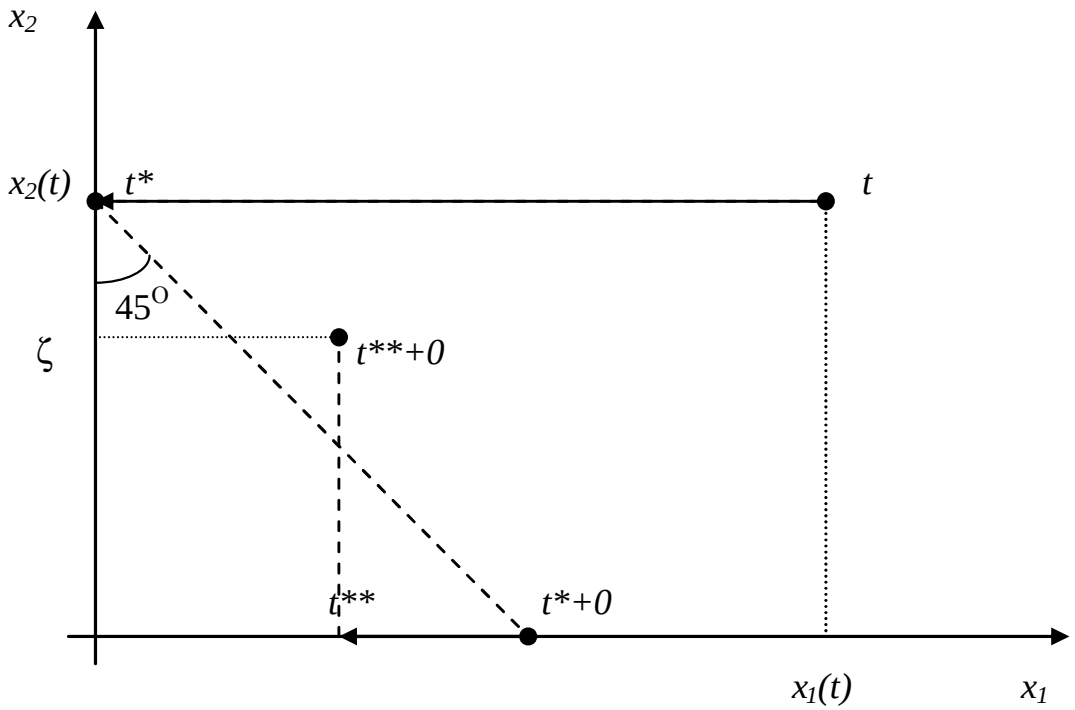


Рис. 1.6. Изменение координат КЛА во времени при ν=2

Рассмотрим теперь случай поступления входного сигнала **(1,θ)** в момент $\mathbf{t^{**}}$. Пусть при этом состояние агрегата было $\mathbf{x(t^{**})=(\nu,x^{(\nu)})}$, где **ν≥0**. Поступление рассматриваемого входного сигнала означает приход в систему обслуживания требования, обладающего временем обслуживания **ζ** и предельным временем ожидания **θ**. (на рис. 1.6 – скачок агрегата из точки **t**** в точку **t**+0**). Рассмотрим два случая: **ν=0** и **ν>0**. В первом из них требование поступает в пустую систему, а во втором – в занятую обслуживанием. При **ν=0** состояние $\mathbf{x(t^{**}+0)}$ должно отражать тот факт, что поступившее требование сразу принято к обслуживанию и ему случайным образом назначено время обслуживания, т.е. $\mathbf{\nu=0,\ x(t^{**})=(0),\ x(t^{**}+0)=(1,\zeta),}$ где **ζ** - случайная величина, имеющая заданное распределение. При **ν>0** состояние $\mathbf{x(t^{**}+0)}$ должно отражать тот факт, что требование будет принято к обслуживанию тогда и только тогда, когда его предельное время ожидания **θ** превосходит реальное (т.е. $\mathbf{\theta>\sum_{i=1}^{(\nu)} x_i^{(\nu)}}$) и в случае, если оно принято, ему назначается случайное время обслуживания.

ν>0

$\mathbf{x(t^{**})= (\nu,\ x_1(t^{**}),\ldots,x_\nu(t^{**}))}$

$$\mathbf{x(t^{**}+0)=}\begin{cases}\mathbf{(\nu,\ x_1(t^{**}),\ldots,x_\nu(t^{**}))}, \text{ если } \theta\le\sum_{i=1}^{\nu} x_i^{(\nu)}\\ \mathbf{(\nu+1,\ x_1(t^{**}),\ldots,x_\nu(t^{**}),\zeta)}, \text{ если } \theta>\sum_{i=1}^{\nu} x_i^{(\nu)}\end{cases}$$

Будем считать, что в момент t^{**} выдается входной сигнал, фиксирующий имеющуюся длину очереди, время ожидания и факт принятия или непринятия поступающего требования. В соответствии с этим предположим, что $y=(\lambda,y)$, где $\lambda=(\lambda_1,\lambda_2)$, λ_1 – число требований, находящихся в системе в момент t^{**}, $\lambda_2=0$ или 1, если поступающее требование не принимается или принимается соответственно к обслуживанию, $||\lambda||=\lambda_2$, а компонента y равна времени ожидания принятым требованиям начала обслуживания.

В данном случае λ представляет собой вектор размерности 2 с целочисленными компонентами. Из сказанного следует, что имеют место зависимости:

$\lambda_1=\nu$

$$\lambda_2=\begin{cases} 0, \text{ если } \theta\le\sum_{i=1}^{\nu}x_i \\ 1, \text{ в противном случае} \end{cases}$$

$y=\sum_{i=1}^{\nu}x_i$ (координата определяется лишь при $\lambda_2=1$)

Выходной сигнал накапливается в системе сбора статистики. λ_1 используется для вычисления средней длины очереди, y – для вычисления среднего времени ожидания обслуживания.

1.8. Марковские цепи

Функционирование многих объектов представляет собой последовательность переходов их из одного состояния в другое (ЭВМ, каналы передачи информации, и т.п.). Ввиду особой важности таких объектов для современных средств обработки и передачи информации рассмотрим специфический способ описания систем – так называемые **марковские цепи**.

> Система называется **системой с дискретными состояниями**, если множество ее состояний конечно, а переходы из одного состояния в другое осуществляются скачком.
>
> Последовательность состояний такой системы называется **цепью**.

Допустим, некоторая система может находиться в одном из N состояний. Цепью будет являться случайная последовательность состояний системы. Простейшей характеристикой случайного процесса, являющегося цепью, служит набор вероятностей состояний $p_1(t), p_2(t),\ldots, p_n(t)$, где $p_i(t)$ – вероятность того, что в момент t система находится в состоянии i. $p_1(t)+p_2(t)+\ldots+p_N(t)=1$

> Случайный процесс, протекающий в системе, называется **марковским**, если соблюдается принцип **отсутствия последействия**, то есть, для любого момента времени t_0 вероятность любого состояния системы в будущем (при $t > t_0$) зависит только от ее состояния в настоящем (при

$t = t_0$) и не зависит от того, каким образом система пришла в это состояние.

Марков (старший) Андрей Андреевич (1856-1922) – русский математик, доктор физико-математических наук, профессор, академик Петербургской академии наук. Работал профессором Петербургского университета, опубликовал около 70 работ по теории чисел, теории приближенных функций, дифференциальных уравнений, теории вероятностей. В цикле работ, опубликованных в 1906-1912 годах, заложил основы одной из общих схем естественных процессов, которая была названа цепями Маркова. А.А. Марков пользовался большим авторитетом у студентов. Был материалистом и убежденным атеистом.

Переход системы из одного состояния в другое является в общем случае случайным событием. Последовательность смены состояний является **потоком событий**.

Поток событий является **ординарным**, если события происходят поодиночке (нет двух одновременных событий). Поток называется **стационарным**, если его вероятностные характеристики не изменяются во времени. Чаще всего применяются **пуассоновские** потоки событий, то есть имеющие неизменную интенсивность (плотность) – среднее число событий в единицу времени постоянна. $\lambda = \mathbf{const.}$

Дискретные марковские цепи

Рассмотрим случайный марковский процесс с дискретными состояниями и дискретным временем. Такой процесс описывает систему **S** с конечным числом состояний, причем переходы возможны только в фиксированные моменты времени $t_1, t_2, \ldots, t_K$. Процесс функционирования представим в виде цепи $S_0\ (0) \rightarrow S_i\ (1) \rightarrow S_j\ (2) \rightarrow \ldots \rightarrow S_m\ (K)$.

Случайная последовательность является **дискретной марковской цепью**, если смена состояний происходит в дискретные моменты времени и соблюдается принцип отсутствия последействия (т.е. для каждого шага вероятность перехода из любого состояния S_i в любое состояние S_j ($i, j = 1, 2, \ldots N$) не зависит от того, как система пришла в состояние S_i.

Каждому переходу системы из состояния S_i в состояние S_j в момент времени t_k соответствует переходная вероятность $p_{ij}\ (t_k)$. Это условная вероятность $p_{ij}\ (t_k) = P(S_j(t_k) \mid S_i(t_{k-1}))$. Очевидно, для каждого номера шага **k** возможные переходы образуют полную группу событий, т.е. $\sum_{j=1}^{N} p_{ij}(t_k) = 1 \quad \forall k$. Следует обратить внимание, что $\sum_{i=1}^{N} p_{ij}(t_k) \neq 1$.

Дискретная марковская цепь называется **однородной**, если переходные вероятности не зависят от номера шага: $p_{ij}\ (t_k) = p_{ij}$.

Полным описанием однородной марковской цепи могут служить квадратная матрица переходных вероятностей $\mathbf{P_П}$:

$$\mathbf{P_П}=[\mathbf{p_{ij}}]=\begin{bmatrix} \mathbf{p_{11}} & \mathbf{p_{12}} & \cdots & \mathbf{p_{1N}} \\ \mathbf{p_{21}} & \mathbf{p_{22}} & \cdots & \mathbf{p_{2N}} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{p_{N1}} & \mathbf{p_{N2}} & \cdots & \mathbf{p_{NN}} \end{bmatrix}$$

и вектор вероятностей всех начальных состояний **P(0)**

$\mathbf{P(0)} = [\mathbf{p_i(0)}] = [\mathbf{p_1(0)}, \mathbf{p_2(0)}\ldots \mathbf{p_N(0)}]$

Переходные вероятности, соответствующие невозможным переходам, равны нулю, вероятности, расположенные на главной диагонали, соответствуют тому факту, что система не изменила своего состояния.

Для однородной марковской цепи найдем вектор вероятностей всех состояний для любого **k**-го шага. В соответствии с формулой полной вероятности вероятность **i**-го состояния на первом шаге равна:

$\mathbf{p_j(1)}=\sum_{\mathbf{i=1}}^{\mathbf{N}}\mathbf{p_i(0)p_{ij}}$, $\mathbf{j}=\overline{\mathbf{1,N}}$, или в матричной форме: $\mathbf{P(1) = P(0)\cdot P_П}$.

Аналогично, для второго шага: $\mathbf{P(2) = P(1)\cdot P_П = P(0)\cdot P_П \cdot P_П}$

Соответственно, для **k**-го шага: $\mathbf{P(k) = P(0)P_П^k}$

Обозначим элемент матрицы $\mathbf{P_П^k}$ таким образом: $\mathbf{p_{ij}(k)}$.

Если возможен переход из состояния $\mathbf{S_i}$ в состояние $\mathbf{S_j}$ за **k** шагов, то $\mathbf{p_{ij}(k)>0}$. Если при этом возможен и обратный переход за произвольное число шагов, то состояния $\mathbf{S_i}$ и $\mathbf{S_j}$ называются **сообщающимися**. Состояние $\mathbf{S_i}$ называется **возвратным**, если вероятность того, что система, выйдя из этого состояния, вернется в него за конечное число шагов хотя бы один раз, равна единице, и **невозвратным**, если вероятность возврата за конечное число шагов меньше единицы.

Эргодические и поглощающие марковские цепи

Динамику переходов системы из состояния в состояние с течением времени можно представить в виде графа. Нередко это иерархический граф, дерево.

> Сообщающиеся состояния, находящиеся на последней ступени иерархии, называются **эргодическим подмножеством** состояний. В частном случае, эргодическое множество может состоять из одного элемента, который называется **поглощающим**. Если все эргодические подмножества цепи состоят только из одного поглощающего состояния, такая цепь называется **поглощающей**.

Из поглощающего состояния нельзя перейти ни в какое другое. В матрице переходных вероятностей поглощающему состоянию соответствует строка, в которой все переходные вероятности $\mathbf{p_{ij}=0}$, кроме одной (диагональной) $\mathbf{p_{ii}=1}$.

Для эргодических цепей характерно то, что при достаточно большом шаге **k** наступает **стационарный** (**установившийся**) режим, при котором вероятности состояний $\mathbf{p_i(k)}$ практически не изменяются с течением времени, т.е. $\lim\limits_{\mathbf{k}\to\infty}\mathbf{p_i(k)}=\mathbf{p_{ci}}$. Вектор $\mathbf{P_C} = \mathbf{[p_{ci}] = [p_{c1}, p_{c2}, \ldots, p_{cN}]}$ называется вектором стационарных вероятностей. До наступления стационарного режима система находится в переходном режиме. Переходный режим заканчивается, когда $\mathbf{||P_C - P(k)||<\varepsilon}$, где $\boldsymbol{\varepsilon}$ - малая положительная величина. Каждая компонента $\mathbf{p_{ci}}$ вектора стационарных вероятностей характеризует среднюю долю времени, в течение которого система находится в состоянии $\mathbf{S_i}$. Если все состояния цепи являются сообщающимися, то вся цепь является эргодической. В такой системе из любого состояния можно попасть в любое за конечное число шагов, в матрице стационарных переходных вероятностей нет нулевых элементов, а граф системы является сильно связанным.

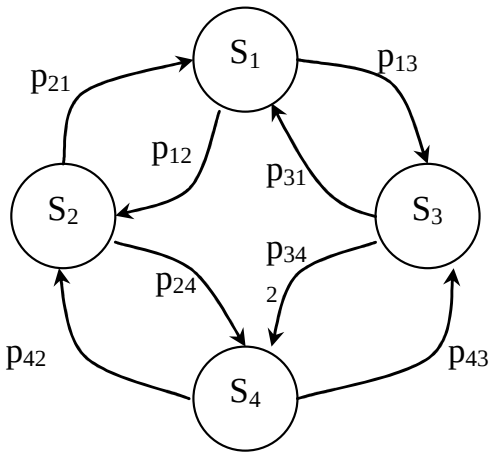


Рис. 1.7. Пример эргодической марковской цепи

Для определения стационарных вероятностей $\mathbf{p_{ci}}$ нахождения системы в состоянии $\mathbf{S_i}$ нужно составить систему **N** линейных алгебраических уравнений с **N** неизвестными: $\mathbf{p_{Cj}} = \sum\limits_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{p_{Ci}p_{ij}}$, $\mathbf{j}=\overline{\mathbf{1,N}}$, причем искомые вероятности удовлетворяют условию нормировки $\sum\limits_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{p_{Ci}} = \mathbf{1}$.

Для системы на рис. 1.7 выполняется:

$$\begin{cases} p_{C1}=p_{C2}p_{21}+p_{C3}p_{31} \\ p_{C2}=p_{C1}p_{12}+p_{C4}p_{42} \\ p_{C3}=p_{C1}p_{13}+p_{C4}p_{43} \\ p_{C4}=p_{C2}p_{24}+p_{C3}p_{34} \\ \text{Условие нормировки: } p_{C1}+p_{C2}+p_{C3}+p_{C4}=1 \end{cases}$$

Эта система уравнений переопределена, в ней есть линейно-зависимые уравнения.

На практике при исследовании эргодических марковских цепей часто ограничиваются рассмотрением стационарных режимов.

Поглощающие марковские цепи характеризуются тем, что эргодическое подмножество состоит из единственного элемента, который и является поглощающим (см. рис. 1.8).

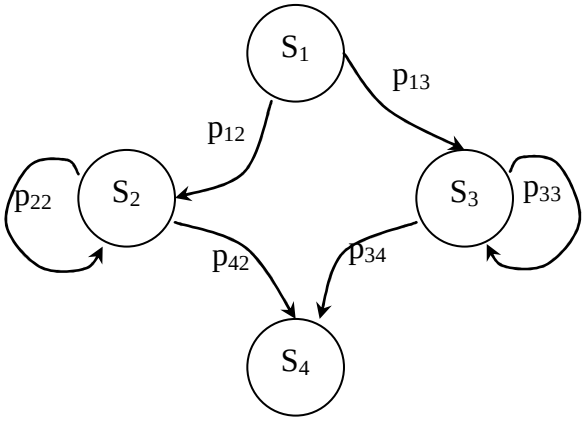


Рис. 1.8. Пример поглощающей цепи

В установившемся режиме, независимо от начального состояния, вероятность нахождения поглощающей марковской цепи в поглощающем состоянии близка к единице, а вероятности остальных состояний близки к нулю. Для примера на рисунке $p_{C1}= p_{C2} = p_{C3}= 0, p_{C4}=1.$ В связи с этим для исследования интересен только переходный режим.

Пример[1]**.** Требуется построить математическую модель и определить основные характеристики системы обработки информации, содержащей канал передачи информации **К**, буфер **Б**, специализированное вычислительное устройство **В**. На вход системы поступает информация, генерируемая источником входной информации **И** (рис. 1.9).

[1] Пример заимствован из кн. Вероятностные методы в вычислительной технике/ А.В.Крайников, Б.А.Курдиков, А.Н.Лебедев и др. – М.: Высш. шк., 1986.

Схема работы вычислительной системы следующая. Источник **И** выдает один пакет сразу, как только освобождается канал **К**. Канал **К** передает пакет в течение двух тактов. В канале может находиться только один пакет. Если буфер занят, канал хранит пакет, но блокируется по входу. Буфер **Б** может хранить от **0** до **N** пакетов. Если буфер занят полностью, он не принимает новый пакет. Буфер выдает пакет вычислителю **В** сразу, как только вычислитель освобождается и тут же может принять новый пакет. Вычислитель **В** обрабатывает пакет в течение одного такта и выводит его из системы. С вероятностью π происходит сбой, и обработка пакета повторяется.

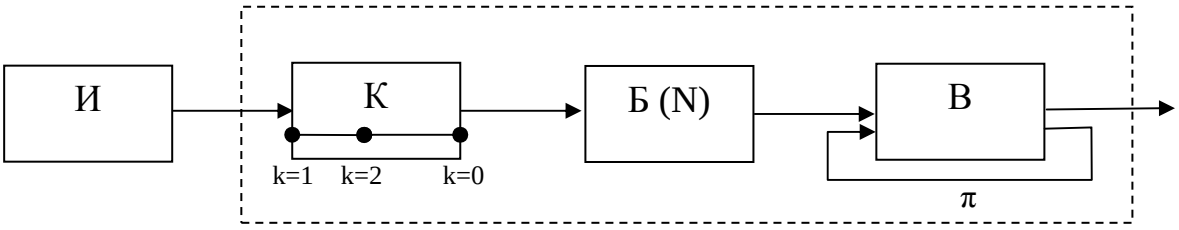


Рис. 1.9. Пример системы обработки информации

Введем множество состояний.

Канал **К** может находиться в одном из трех состояний:

k=1 – пакет принят и находится в начале передачи (передастся через 2 такта);

k=2 – пакет находится в середине передачи (передастся через 1 такт);

k=0 – передача пакета завершена, но буфер занят и канал заблокирован.

Буфер **Б** может находиться в одном из **N+1** состоянии:

n=0 – буфер свободен;

n=1, 2,…, N – в буфере находятся **1, 2,…, N** пакетов.

Вычислитель **В** может находиться в одном из двух состояний:

i=0 – вычислитель свободен; **i=1** – вычислитель занят.

Состояние системы **S** на каждом такте описывается вектором трех компонентов **(k, n, i)**. **||S|| = 3·(N+1)·2**, хотя некоторые состояния являются невозможными.

Построим граф марковской цепи и определим переходные вероятности (рис. 1.10).

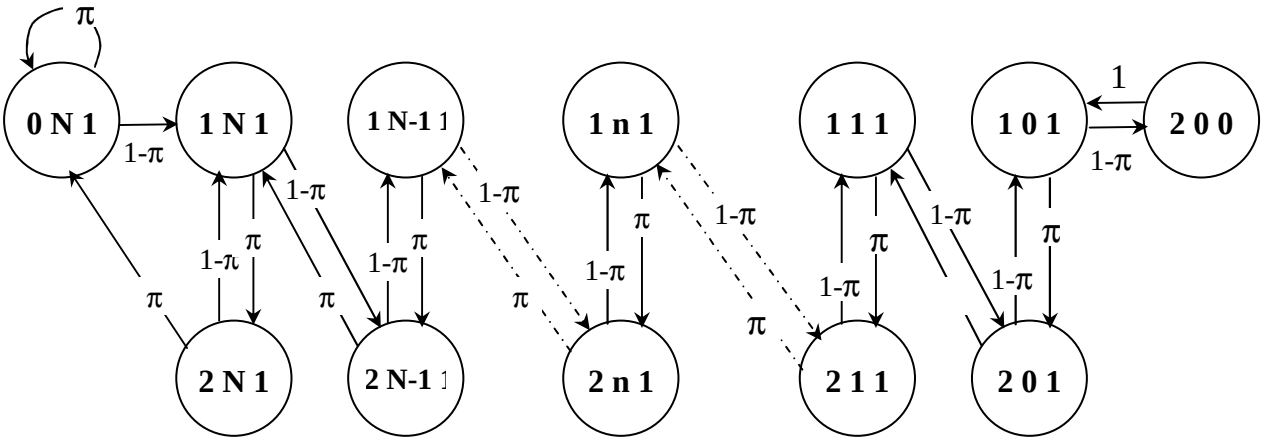


Рис. 1.10. Граф марковской цепи, описывающей работу вычислительной системы

Начнем с состояния **(0,N,1)** – вычислитель занят, буфер занят, канал заблокирован и содержит один пакет, готовый к передаче в буфер. Из этого состояния возможны два перехода:

Состояние **(0,N,1)** –

с вероятностью **π** произойдет сбой и произойдет возврат в то же состояние (на рис. 1.10 этому случаю соответствует дуга)

с вероятностью **1-π** пакет будет обработан, **В** освободится и тут же будет занят пакетом из буфера (**i=1**), **Б** уменьшится и тут же будет вновь заполнен пакетом из **К** (**n=N**), канал освободится и тут же получит новый пакет, который будет находиться в начале передачи (**k=1**). То есть произойдет переход в новое состояние **(1,N,1)**.

Состояние **(1,N,1)** –

с вероятностью **π** произойдет сбой, **В** по-прежнему будет занят (**i=1**), **Б** по-прежнему будет заполнен (**n=N**), пакет по каналу перейдет на следующую стадию (**k=2**). Новое состояние **(2,N,1).**

с вероятностью **1-π** пакет будет обработан, **В** освободится и тут же будет занят пакетом из буфера (**i=1**), в **Б** станет на один пакет меньше (**n=N-1**), пакет по каналу перейдет на следующую стадию (**k=2**). Новое состояние **(2,N-1,1)**.

Состояние **(2,N,1)** –

с вероятностью **π** произойдет сбой, **В** по-прежнему будет занят (**i=1**), **Б** по-прежнему будет заполнен (**n=N**), пакет по каналу завершит передачу и канал заблокируется (**k=0**). Новое состояние **(0,N,1).**

с вероятностью **1-π** пакет будет обработан, **В** освободится и тут же будет занят пакетом из буфера (**i=1**), **Б** уменьшится и тут же будет вновь заполнен пакетом из **К** (**n=N**), канал освободится и тут же получит новый

пакет, который будет находиться в начале передачи (**k=1**). Новое состояние **(1,N,1).**

Состояние **(2,N-1,1)** –

с вероятностью **π** произойдет сбой, **В** по-прежнему будет занят (**i=1**), в **Б** поступит новый пакет (**n=N**), канал освободится и тут же получит новый пакет, который будет находиться в начале передачи (**k=1**). Новое состояние (**1,N,1**).

с вероятностью **1-π** пакет будет обработан, **В** освободится и тут же будет занят пакетом из буфера (**i=1**), **Б** уменьшится и тут же будет вновь заполнен пакетом из **К** (**n=N-1**), канал освободится и тут же получит новый пакет, который будет находиться в начале передачи (**k=1**). Новое состояние **(1,N-1,1).**

И так далее… Перейдем к рассмотрению завершающих состояний.

Состояние **(1,0,1)** –

с вероятностью **π** произойдет сбой, **В** по-прежнему будет занят (**i=1**), **Б** по-прежнему будет свободен (**n=0**), пакет по каналу перейдет на следующую стадию (**k=2**). Новое состояние **(2,0,1).**

с вероятностью **1-π** пакет будет обработан, **В** освободится и поскольку в буфере пакетов нет, останется свободным (**i=0**), **Б** по-прежнему будет свободным (**n=0**), пакет по каналу перейдет на следующую стадию (**k=2**). Новое состояние **(2,0,0).**

Состояние **(2,0,0)** – сбой произойти не может, так как вычислитель не работает, пакет из канала завершит передачу и поступит в **Б**, а затем сразу в **В** (**i=1, n=0**), канал получит новый пакет из источника (**k=1**). Новое состояние **(1,0,1)** с вероятностью **1**.

Матрица переходных вероятностей содержит **6·(N+1)** строк и столбцов. В матрице есть нулевые строки (для невозможных состояний). В каждой ненулевой строке один элемент равен **π**, другой **1-π,** остальные – нулевые, так как из каждого состояния возможен переход только в два других состояния. В строке, соответствующей состоянию **(2,0,0)**, один элемент равен **1**, остальные – **0**.

Составим систему уравнений для нахождения стационарных вероятностей.

$p(0,N,1) = \pi \cdot p(0,N,1) + \pi \cdot p(2,N,1)$,

$p(1,N,1) = (1-\pi) \cdot p(0,N,1) + (1-\pi) \cdot p(2,N,1) + \pi \cdot p(2,N-1,1)$,

$p(2,N,1) = \pi \cdot p(1,N,1)$,

……

для n=N-1, N-2,…, 1:

$p(1,n,1) = \pi \cdot p(2,n-1,1) + (1-\pi) \cdot p(2,n,1)$,

$p(2,n,1) = \pi \cdot p(1,n,1) + (1-\pi) \cdot p(1,n+1,1)$,

……

для n = 0:

p(2,0,1) = π·p(1,0,1) + (1-π)·p (1,1,1),

p(1,0,1) = p(2,0,0) + (1-π)·p(2,0,1),

p(2,0,0) = (1-π)·p(1,0,1).

Кроме того, выполняется условие нормировки: $\sum_k \sum_n \sum_i \mathbf{p(k,n,i) = 1}$.

Обозначим p(2,0,0) = p; π/(1-π) = **ω**.

Из последнего уравнения: $\mathbf{p(1,0,1) = \frac{1}{1-\pi} p}$.

Из предпоследнего: $\mathbf{p(2,0,1) = (\frac{1}{1-\pi} p - p)\frac{1}{1-\pi} = \frac{\pi}{(1-\pi)^2} p = \frac{\varpi}{1-\pi} p}$.

Из предпредпоследнего:

$$\mathbf{p(1,1,1)} = \frac{1}{1-\pi}\left(\frac{\varpi}{1-\pi} p - \frac{\pi}{1-\pi} p\right) = \frac{1}{1-\pi} p \frac{\pi^2}{(1-\pi)^2} = \frac{\varpi^2}{1-\pi} p .$$

и так далее…

Для **n = 0, 1, 2, …, N-1** можно записать:

$$\mathbf{p(1,n,1)} = \frac{\varpi^{2n}}{1-\pi} p ,$$

$$\mathbf{p(2,n,1)} = \frac{\varpi^{2n+1}}{1-\pi} p .$$

Из первых трех уравнений вытекает:

из первого

$$\mathbf{p(0,N,1) = \frac{\pi}{1-\pi} p(2,N,1) = \varpi \cdot p(2,N,1)} ;$$

из второго (с учетом, что **p(2,N,1) = π·p(1,N,1)**):

$$\mathbf{p(1,N,1) = (1-\pi)\,\varpi\,\pi\, p(1,N,1) + (1-\pi)\,\pi\, p(1,N,1) + \pi \frac{\varpi^{2N-1}}{1-\pi} p .}$$

Преобразуем далее:

$$\mathbf{p(1,N,1) = p(1,N,1)((1-\pi)\varpi\pi + (1-\pi)\pi) + \varpi^{2N} p.}$$

Отсюда:

$$\mathbf{p(1,N,1) = \frac{\varpi^{2N}}{1-\pi} p .}$$

$p(2,N,1) = \varpi^{2N+1} p$

$p(0,N,1) = \varpi^{2N+2} p$

из условия нормировки:

$$\sum_k \sum_n \sum_i p(k,n,i) = 1 = p \cdot \left(\sum_{j=0}^{2N} \frac{\varpi^j}{1-\pi} + \varpi^{2N+1} + \varpi^{2N+2}\right)$$

отсюда $p = \left(\sum_{j=0}^{2N} \frac{\varpi^j}{1-\pi} + \varpi^{2N+1} + \varpi^{2N+2}\right)^{-1}$

Можно определить стационарные вероятностные характеристики системы. Например, вероятность того, что вычислитель простаивает $p(2,0,0) = p$, вероятность того, что канал простаивает $p(0,N,1) = \varpi^{2N+2} p$.

Самостоятельно определите вероятность того, что в буфере находятся **i** пакетов; больше, чем **i** пакетов.

Непрерывные марковские цепи

На практике часто встречаются ситуации, когда система имеет конечное число дискретных состояний, а переходы между состояниями происходят в произвольные моменты времени. Например, отказы аппаратуры могут произойти в любой случайный момент времени.

> Случайный процесс с непрерывным временем называется **непрерывной марковской цепью**, если смена состояний системы происходит в непрерывные моменты времени и поведение системы после произвольного момента времени **t** зависит только от состояния в этот момент времени и не зависит от истории процесса, предшествующей моменту **t** (соблюдается принцип отсутствия последействия).

Очень часто на практике интервалы времени **x** между двумя сменами состояний системы подчинены показательному закону распределения $f(x) = \lambda e^{-\lambda x}$, где параметр закона распределения λ имеет смысл интенсивности переходов (среднее количество переходов в единицу времени).

Для непрерывной марковской цепи определим вероятности всех состояний системы для любого момента времени $p_i(t)$, $i=1, 2,\ldots, N$. Так как для любого момента **t** все состояния образуют полную группу событий, то

$$\sum_{i=1}^{N} p_i(t) = 1$$

Пусть система в момент времени **t** находится в состоянии S_i. Рассмотрим элементарный промежуток времени Δt, примыкающий к моменту времени **t**. Вероятность перехода из состояния S_i в состояние S_j за промежуток Δt обозначим $p_{ij}(\Delta t)$. Назовем **плотностью вероятности перехода** величину λ_{ij}, определяемую так:

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t},$$

то есть, при малых Δt вероятность перехода $p_{ij}(\Delta t) \approx \lambda_{ij}(t)\,\Delta t$.

Если все плотности вероятностей перехода не зависят от t, то такой марковский процесс называется **однородным**: $\lambda_{ij}(t) = \lambda_{ij} = \text{const}$ (и **неоднородным** – в противном случае). При этом для случая, когда интервал времени между переходами системы распределен по показательному закону, плотность вероятности перехода равна параметру λ показательного закона.

Пусть известны плотности вероятностей переходов λ_{ij} для всех пар состояний S_i и S_j. Определим вероятности состояний системы $p_i(t)$. Для момента времени $t+\Delta t$ справедливо соотношение:

$$p_i(t+\Delta t) = \sum_{j=1}^{N} p_j(t) p_{ji}(\Delta t) = p_i(t) p_{ii}(\Delta t) + \sum_{\substack{j=1 \\ j \neq i}}^{N} p_j(t) p_{ji}(\Delta t).$$

Из свойства матрицы переходных вероятностей (сумма вероятностей по строке равна 1) следует:

$$p_{ii}(\Delta t) = 1 - \sum_{\substack{j=1 \\ j \neq i}}^{N} p_{ij}(\Delta t).$$

Подставив это в предыдущее выражение, получим:

$$p_i(t+\Delta t) = p_i(t) \cdot \left(1 - \sum_{\substack{j=1 \\ j \neq i}}^{N} p_{ij}(\Delta t) \right) + \sum_{\substack{j=1 \\ j \neq i}}^{N} p_j(t) p_{ji}(\Delta t),$$

$$p_i(t+\Delta t) - p_i(t) = -p_i(t) \cdot \sum_{\substack{j=1 \\ j \neq i}}^{N} p_{ij}(\Delta t) + \sum_{\substack{j=1 \\ j \neq i}}^{N} p_j(t) p_{ji}(\Delta t).$$

Разделим обе части равенства на Δt и устремим его к нулю, получаем:

$$\lim_{\Delta t \to 0} \frac{p_i(t+\Delta t) - p_i(t)}{\Delta t} = -p_i(t) \cdot \sum_{\substack{j=1 \\ j \neq i}}^{N} \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t} + \sum_{\substack{j=1 \\ j \neq i}}^{N} p_j(t) \lim_{\Delta t \to 0} \frac{p_{ji}(\Delta t)}{\Delta t}.$$

Получили систему дифференциальных уравнений А.Н.Колмогорова:

$$\frac{dp_i(t)}{dt} = -p_i(t) \cdot \sum_{\substack{j=1 \\ j \neq i}}^{N} \lambda_{ij} + \sum_{\substack{j=1 \\ j \neq i}}^{N} p_j(t) \lambda_{ji}, \quad i = \overline{1, N}.$$

Интегрирование этой системы по времени позволит вычислить функции $p_i(t)$. При этом должно соблюдаться условие нормировки.

Колмогоров Андрей Николаевич (1903 - 1987). Великий русский ученый, один из крупнейших математиков XX столетия, член Национальной Академии наук США и американской Академии искусств и наук, член Нидерландской Королевской академии наук, Академии наук Финляндии, Академии наук Франции, Германской академии естествоиспытателей "Леопольдина", Международной академии истории наук и национальных академий Румынии, Венгрии и Польши, почетный член Королевского статистического общества Великобритании, Лондонского математического общества, Международного статистического института и Математического общества Индии, иностранный член Американского философского общества, Американского метеорологического общества, лауреат самых почетных научных премий: премии П.Л.Чебышева и Н.И.Лобачевского АН СССР, Международной премии фонда Больцано и Международной премии фонда Вольфа, а также государственной и Ленинской премии, награжденный 7-ю орденами Ленина, медалью "Золотая Звезда" Герой Социалистического труда академик Андрей Николаевич Колмогоров сам себя всегда называл "просто профессор Московского университета". На протяжении почти полувека А.Н.Колмогоров был общепризнанным лидером в теории вероятностей. Вместе с А.Я. Хинчиным и многими своими учениками он внес значительный вклад в развитие теории информации. К середине 50-х гг. именно А.Н.Колмогоров предложил наиболее общее определение количества информации в вероятностном смысле, а в дальнейшем развил и другой подход, так называемую алгоритмическую теорию информации, в котором под энтропией понималась сложность объекта, равная сложности алгоритма, описывающего объект.

В эргодических марковских цепях существует установившийся режим, при котором вероятности не меняются с течением времени. Следовательно, $\mathbf{p_i(t)=p_i}$. Производные в системе дифференциальных уравнений Колмогорова при этом равны **0** и система Колмогорова становится сходной с системой для нахождения стационарных вероятностей для дискретных систем.

Вопросы и задачи к главе 1

1. Опишите в виде «черного ящика» утюг.

Решение. Система «УТЮГ». Границы системы: собственно утюг с присоединенным резервуаром для воды, а также электрошнур с вилкой. Множество входов $\mathbf{U = \{u_1, u_2, u_3, u_4, u_5\}}$, где: $\mathbf{u_1}$ - электропитание, $\mathbf{u_2}$ - вода, $\mathbf{u_3}$ - перемещение регулятора температуры, $\mathbf{u_4}$ - нажатие кнопки пуска пара, $\mathbf{u_5}$ – перемещение руки гладильщика. Множество выходов $\mathbf{Y = \{y_1, y_2, y_3\}}$, где $\mathbf{y_1}$ - температура подошвы утюга, $\mathbf{y_2}$ - пар, $\mathbf{y_3}$ - перемещение утюга по поверхности изделия. Очевидна зависимость выхода $\mathbf{y_1}$ от входов $\mathbf{u_1}$ и $\mathbf{u_3}$, выхода $\mathbf{y_2}$ от входов $\mathbf{u_2}$ и $\mathbf{u_4}$, выхода $\mathbf{y_3}$ от входа $\mathbf{u_5}$. Таким образом, отображение множества входов на множество выходов представляет трехкомпонентную вектор-функцию

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}_1(\mathbf{u}_1, \mathbf{u}_3) \\ \mathbf{f}_2(\mathbf{u}_2, \mathbf{u}_4) \\ \mathbf{f}_3(\mathbf{u}_5) \end{pmatrix}$$

2. Опишите в виде черного ящика другой известный бытовой прибор: телевизор, пылесос.

3. Упорядочьте в порядке возрастания организованности системы: Комплект полупроводниковых элементов (ПЭ), Система управления станка с ЧПУ (СУ), Комплект электронных блоков (ЭБ), Комплект электронных плат (ЭП).

Решение. ПЭ – ЭП – ЭБ – СУ

4. Упорядочьте в порядке возрастания организованности системы: Набор предопределенных выражений языка программирования (ВЯ), Набор операторов языка программирования (ОЯ), Программа для ЭВМ (П), Библиотека процедур и функций (ПФ), Набор допустимых символов языка программирования (СЯ).

5. Что такое «Эмерджентность»?

6. Что такое «Элемент»?

7. Чем отличается модель «белого ящика» от модели «черного ящика»?

8. Назовите противоположные по смыслу классы систем:

линейная – нелинейная

детерминированная – …

большая – …

динамическая – …

стационарная – …

закрытая – …

9. Дайте определения понятиям «входы», «выходы» и «состояния» системы.

10. На телеграфе установлен автоматизированный пункт передачи сообщений[1], работающий в режиме самообслуживания. Отправителю предоставляется рабочее место с микроЭВМ, позволяющей набрать нужный текст, отобразить его на экране и отредактировать. Набор текста начинается с клавиши «Ввод». Если операции ввода и редактирования отправителем выполнены, то по нажатию специальной клавиши «Завершение» текст запоминается и начинает обрабатываться машиной (распознается адрес, подсчитывается количество символов, определяется наличие признаков срочности телеграммы и т.п.). После завершения обработки на экране появляется символ готовности к передаче. Пользователь может нажать клавишу «Передача» или клавишу «Сброс». В первом случае состоится передача текста, во втором – текст будет удален из памяти ЭВМ и система придет в исходное состояние. Дать теоретико-множественное описание системы.

Решение. Определим входы, выходы, состояния системы и связи между этими множествами. В качестве входов можно определить вектор из четырех булев-

[1] Сюжет примера взят из кн.: Дегтярев, Ю.И. Системный анализ и исследование операций. – М.: Высш. шк., 1996.

ских переменных: $\mathbf{U} = \{\mathbf{u_1}, \mathbf{u_2}, \mathbf{u_3}, \mathbf{u_4}\}$, принимающих значение **Истина**, если нажаты клавиши «Ввод», «Завершение», «Передача» и «Сброс» соответственно. Состояниями в этом случае будут две булевских величины $\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}\}$. Первая принимает значение **Истина**, если аппарат находится в режиме редактирования текста, и **Ложь**, если аппарат находится в режиме простоя. Вторая принимает **истинное** значение, если аппарат занят, и **ложное** – в случае простоя аппарата. Выходом системы будет логическая величина **y**: происходит (**Истина**) или не происходит (**Ложь**) передача телеграммы. Функционирование такой системы можно описать таблицей истинности.

u_1	u_2	u_3	u_4	x_1	x_2	y
0	**0**	**0**	**0**	**0**	**0**	**0**
1	**0**	**0**	**0**	**1**	**1**	**0**
0	**1**	**0**	**0**	**0**	**1**	**0**
0	**0**	**1**	**0**	**0**	**1**	**1**
0	**0**	**0**	**1**	**0**	**0**	**0**

Эту же систему можно описать, используя функции алгебры логики: $\mathbf{x_1 = u_1}$; $\mathbf{x_2 = u_1 \vee u_2 \vee u_3}$; $\mathbf{y = x_2 \wedge u_3}$.

11. «Министерский» телефонный аппарат имеет телефонную трубку, **N** кнопок для вызова абонентов, память на один телефонный номер, кнопку повторного вызова последнего набранного абонента и кнопку стирания номера абонента из памяти. Вызов абонента осуществляется при поднятой телефонной трубке нажатием одной из **N** клавиш, соответствующей требуемому абоненту. При соединении с абонентом его номер автоматически заносится в память аппарата. Дать теоретико-множественное описание системы.

12. Игра в большой теннис. Вероятность выигрыша подачи игроком **A** – 0.6, вероятность выигрыша подачи игроком **B** – 0.4. Описать один гейм матча в виде марковской цепи.

13. Укажите положительные и отрицательные стороны применения качественных и количественных методов описания систем.

14. Какие из подходов к описанию систем применены, по-вашему, в следующих случаях: принципиальная электрическая схема прибора; рецепт приготовления борща; уравнения, описывающие движение планет Солнечной системы; запись логической функции; схема плана боевой операции; нотация шахматной партии; нотная запись мелодии; организационная схема управления предприятием; медицинская карта - история болезни человека; конституция государства; блок-схема алгоритма.

15. Чем непрерывная марковская цепь отличается от дискретной?

16. Что значит «описать систему в виде цепи Маркова»?

17. Какую роль играют обратные связи в управлении системами?

18. Чем оптимальное управление отличается от программного управления?

19. Какие допущения вносятся при описании системы в виде кусочно-линейного агрегата?

Глава 2
Основы количественной теории информации

2.1. Основные понятия и определения теории информации

В теории информации изучаются количественные закономерности процессов передачи, хранения и обработки информации. Существует несколько подходов к изучению информации, положенные в основу соответствующих теорий. Основные из них – комбинаторный, вероятностный и алгоритмический. Важный вклад в исследование информационных процессов внесли выдающиеся ученые и инженеры: Р. Хартли, К. Шеннон, А.Н. Колмогоров, Н. Винер, Р. Фишер и др. Наибольшее развитие и практическое применение получила вероятностная теория информации, основоположником которой является Клод Шеннон. Эта теория считается классической, и ей будет посвящена основная часть следующего учебного материала.

Ключевой термин теории – **информация**. Это понятие нельзя считать лишь техническим и вообще узкодисциплинарным. Информация — это фундаментальная философская категория, о которой давно дискутируют последователи самых разных научных направлений. Концепции и толкования, возникающие в ходе научных споров, порождают многообразные определения информации, охватывающие те или иные грани этой субстанции.

Приведем некоторые из многочисленных определений информации.

Информация - свойство материальных объектов и явлений (процессов) порождать многообразие состояний, которые посредством взаимодействий передаются другим объектам и запечатлеваются в их структуре[1].

Информация - сведения (сообщения, данные) независимо от формы их представления[2].

Информация - это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний[3].

Информация - это обозначение содержания, полученного из внешнего мира в процессе нашего приспособления к нему и приспособления к нему наших чувств[4].

---

[1] В.М. Глушков, Н.М. Амосов и др. «Энциклопедия кибернетики». Киев: Наукова думка. – 1975.
[2] Федеральный закон от 27.07.2006 № 149-ФЗ (ред. от 28.12.2013) "Об информации, информационных технологиях и о защите информации".
[3] Н.В.Макарова. Информатика: учебник. М.: Финансы и статистика. – 2000.
[4] Н. Винер. Кибернетика, или Управление и связь в животном и машине. М.: Советское радио. - 1968.

Наиболее полно отражает сущность информации и, поэтому, выглядит достаточно абстрактно следующее определение информации.

> **Информация** – свойство материи, состоящее в том, что в результате взаимодействия объектов между их состояниями устанавливается определенное соответствие.

Информация не может существовать сама по себе, она обязательно представляется в некоторой объективной форме, причем одна и та же информация может быть представлена в разных формах. Для того, чтобы подчеркнуть многообразие форм представления информации, вводится термин **сообщение**.

> **Сообщение** – совокупность символов конечного алфавита, являющаяся формой выражения информации.

Например, информация о характере изменения функции. Формой представления такой информации, то есть сообщением может быть аналитическое выражение функции, либо графическое изображение кривой, либо таблица, содержащая значения функции и аргумента.

Информация является результатом взаимодействия объектов и, с другой стороны, имеет прагматический смысл только при ее передаче от одного объекта к другому.

> **Передача информации** – процесс перемещения сообщения от источника к получателю посредством вещества или энергии.
>
> **Сигнал** – материальный носитель сообщений, обладающий переменными параметрами и служащий для перемещения сообщений.

Материальным носителем для передачи информации может быть звук, свет, радио сигналы, напряжение, угловое или линейное перемещение, бумага, магнитные среды и тому подобное. Сигнал как носитель информации имеет смысл только тогда, когда он заранее неизвестен для «получателя», то есть случаен. Детерминированный сигнал, сигнал с постоянными и заранее известными параметрами информации не несет. Поэтому при описании информационных систем используется аппарат теории вероятностей. Причем, чем менее вероятно сообщение, тем больше информации оно несет (сообщение «В декабре на Урале выпал снег» несет мало информации, так как является весьма вероятным).

> Сигнал называется **непрерывным** (или аналоговым), если его параметр может принимать любое континуальное значение в пределах некоторого интервала и изменяться в произвольный момент времени.

Если обозначить **Z** – значение параметра сигнала, а **t** – время, то зависимость **Z(t)** будет непрерывной функцией (рис. 2.1 *а*).

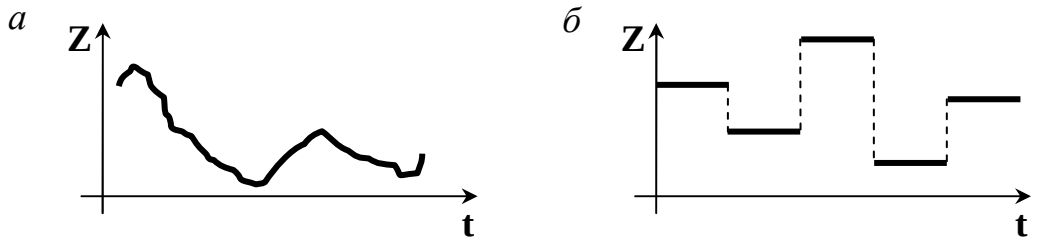


Рис. 2.1. Непрерывный и дискретный сигналы:
а – непрерывный сигнал; *б* – дискретный сигнал

Примерами непрерывных сигналов являются речь и музыка, изображение, показания термометра (параметр сигнала – высота столба спирта или ртути – имеет непрерывный характер значений), стрелочного манометра и прочее.

> Сигнал называется **дискретным**, если его параметр может принимать конечное число значений в пределах некоторого интервала и изменяться в конкретные моменты времени.

Пример дискретных сигналов представлен на рис. 2.1 (б). Подробнее дискретные сигналы будут изучаться в разделе 3.2. Как следует из определения, дискретные сигналы могут быть описаны дискретным и конечным множеством значений параметров. Примерами устройств, использующих дискретные сигналы, являются часы (электронные и механические), цифровые измерительные приборы, книги, табло и прочее.

Значения параметров сигнала определяются сообщением, которое сигнал передает, а значит, символами, из которых сообщение состоит.

> **Символ** – это элемент некоторого конечного множества отличных друг от друга сущностей.

Природа символа может любой – жест, рисунок, буква, сигнал светофора, определенный звук и т.д. Природа знака определяется носителем сообщения и формой представления информации в сообщении.

> Набор символов, в котором установлен порядок их следования, называется **алфавитом**.

Следовательно, алфавит – это упорядоченная совокупность знаков. Порядок следования знаков в алфавите называется лексикографическим. Благодаря этому порядку между знаками устанавливаются бинарные отношения «больше–меньше»: для двух знаков ψ принимается, что $\xi < \psi$, если порядковый номер ξ в алфавите меньше, чем порядковый номер ψ.

Примером алфавита может служить совокупность арабских цифр **0,1…9** – с его помощью можно записать любое целое число в системах счисления от двоичной до десятичной. Если в этот алфавит добавить знаки «+» и «–», то им можно будет записать любое целое число, как положительное, так и

отрицательное. Наконец, если добавить знак разделителя разрядов («**.**» или «**,**»), то такой алфавит позволит записать любое вещественное число.

Поскольку при передаче сообщения параметр сигнала должен меняться, очевидно, что минимальное количество различных его значений равно двум и, следовательно, алфавит должен содержать минимум два знака – такой алфавит называется **двоичным** (синоним – **бинарным**). Верхней границы количества знаков в алфавите не существует; примером могут служить иероглифы, каждый из которых обозначает целое понятие, и общее их количество исчисляется десятками тысяч. Количество знаков в алфавите называют **мощностью** алфавита.

Верность передачи информации – мера соответствия принятого сообщения переданному.

Переработка информации – выполнение формальных операций над входными величинами, над параметрами сигнала в соответствии с заданным алгоритмом.

Хранение информации – фиксация параметров носителя информации.

Помехоустойчивость – способность системы передачи информации противостоять воздействию помех.

Скорость передачи информации – количество информации, переданное в единицу времени.

Скорость передачи информации **J** определяется формулой

$$J=I(S)/T,$$

где **I** – это информация, содержащаяся в сообщении **S**, **T** – это время передачи сообщения **S**.

Пропускная способность – наибольшая достижимая скорость передачи информации для данной информационной системы.

Формула для пропускной способности **C** может выглядеть следующим образом:

$$C=I(S)_{max}/T.$$

2.2. Количество информации

Количество информации – мера неопределённости, устраненной («снятой») при получении сообщения.

В соответствии с принятой в вероятностной теории парадигмой, количество информации в сообщении о некотором событии существенно зависит от вероятности этого события.

Пусть имеется **m** качественных признаков сообщения (количество символов алфавита, количество уровней квантования). **m** называется **мощностью алфавита**. Пусть **n** - число элементов сообщения. Тогда существует $\mathbf{m^n}$ различных сообщений длиной **n**. Первый подход к измерению количества информации как раз связан с этими соображениями.

Первая количественная мера информации – комбинаторная

Первая (комбинаторная) количественная мера информации определяется формулой:

$$\mathbf{I=m^n} \qquad (2.1)$$

Пример: Пусть требуется позвонить по городскому 6-значному номеру или по внутреннему 4-значному номеру телефона. Если нужный номер телефона вам не известен, то в первом случае неопределенность больше. Очевидно, что при использовании 10-значного алфавита (цифры от 0 до 9) неопределенность относительно городского номера в 100 раз ($100=10^6/10^4$) больше, чем неопределенность по поводу внутреннего номера телефона. Поэтому поступившее сообщение (номер нужного телефона) снимает бóльшую неопределенность и, соответственно, несет в 100 раз бóльшую информацию.

Недостаток этой меры – неадитивность (непропорциональность количества информации и длины сообщения). Нам было бы интуитивно проще осознать меру количества информации, если бы сообщение, имеющее, например, в два раза большую длину, несло в два раза большую информацию.

Вторая количественная мера информации – мера Р.Хартли

Ральф Винтон Лайон Хартли (1888-1970) – американский инженер, работал в Западной Электрической компании, которая обеспечивала передачу трансатлантических телефонных и телеграфных сообщений. Рассуждая о количестве информации, содержащемся в телеграфном тексте, заложил основы теории информации, определив логарифмическую меру количества информации (1928 г.). Хартли принадлежат больше чем 70 патентов на изобретения.

С целью обеспечения аддитивности Р.Хартли произвел логарифмирование, вообще говоря, по произвольному основанию правой части формулы (2.1):

$$\mathbf{I = n \cdot \log m}. \qquad (2.2)$$

Понятно, что сообщение должно иметь минимум один символ: $\mathbf{n_{min}=1}$. Алфавит языка должен иметь минимум два элемента: $\mathbf{m_{min}=2}$. Вот в таком сообщении и содержится минимально возможное количество информации.

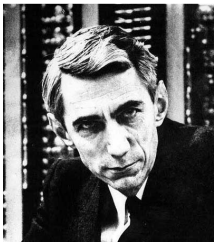
$$\mathbf{I_{min} = n_{min} \cdot \log m_{min} = 1 \cdot \log 2}$$

Удобно использовать двоичный логарифм, так как в этом случае $\mathbf{I_{min} = 1.}$ Отсюда возникла единица измерения информации – бит (от англ. **bi**nary dig**it** – двоичная единица). Поэтому в теории информации используется логарифм по основанию **2**.

$$\mathbf{I = n \cdot \log_2 m.} \qquad (2.3)$$

Недостаток меры Хартли (логарифмической меры) (2.2) состоит в том, что она не учитывает вероятностный характер сообщения (ранее мы уже отмечали, что сообщение имеет смысл, когда оно неизвестно заранее).

Третья количественная мера информации – мера К.Шеннона



Клод Элвуд Шеннон (1916-2001) – американский инженер, член Национальной АН США, профессор Массачусетского технологического института, в 1948 г. издал книгу «Математическая теория связи» с изложением основ теории информации. Будучи студентом Массачусетского технологического института, который он окончил в 1936 году, Шеннон специализировался одновременно и в математике, и в электротехнике. В 1940 году он защитил диссертацию, в которой доказал, что работу обычных переключателей и реле в электрических схемах можно представить посредством булевой алгебры. Сейчас булева алгебра лежит в основе современной цифровой схемотехники, но тогда применение к технике методов английского ученого Джорджа Буля было делом революционным. В 1941 году 25-летний Клод Шеннон поступил на работу в Bell Laboratories, где, помимо всего прочего, прославился тем, что катался на одноколесном велосипеде по коридорам лаборатории, одновременно жонглируя четырьмя мячиками. В годы войны он занимался разработкой криптографических систем, и позже это помогло ему открыть методы кодирования с коррекцией ошибок. Группа инженеров под руководством К.Шеннона сумела декодировать германскую систему шифрования Энигма, что помогло союзникам быть в курсе гитлеровских планов. Кстати, в те же сороковые годы Шеннон занимался конструированием летающего диска на ракетном двигателе. В работах 1957-61 годов Шеннон построил свою теорию пропускной способности каналов связи.

Количественная мера Шеннона получена из следующих соображений. Если символ появляется в сообщении с вероятностью **1** (на определенном месте), то такой символ, очевидно, никакой информации не несет. В случае если все **m** символов алфавита равновероятны, вероятность каждого символа **p=1/m**. Количество информации, содержащееся в сообщении из одного элемента определяется так:

$$\mathbf{I_0 = 1 \cdot \log m = \left| \begin{matrix} p = \frac{1}{m} \\ m = \frac{1}{p} \end{matrix} \right| = \log \frac{1}{p} = -\log p}$$

Если **p** – вероятность символа, то $\mathbf{I_0}$ – информация в сообщении из одного символа. Попробуем разобраться, что такое «один символ».

Сколько, например, символов в сообщениях «МАМА МЫЛА РАМУ» или «Над всей Испанией безоблачное небо»? В последнем случае сообщение несет только **1** бит информации, так как оно является

кодовой фразой для начала фашистского мятежа в Испании, и алфавит франкистов имел всего два «символа» - кодовая фраза либо произносится, либо не произносится.

Очевидно, что любое сообщение можно интерпретировать как один символ соответствующего алфавита. Количество информации не должно зависеть от способа выбора алфавита. Соответственно, вышеуказанную формулу можно распространить на произвольные сообщения (произвольных алфавитов и произвольной длины), с учетом того, что под **P** будем понимать вероятность сообщения.

$$\mathbf{I = - \log_2 P} \qquad (2.4)$$

Формула (2.4) является «великой формулой Шеннона», имеющей универсальный характер и обладающей свойством аддитивности.

Допустим, поступило **n** независимых сообщений (ансамбль сообщений): $\mathbf{a_1, a_2, \ldots a_n}$. Совместная вероятность ансамбля $\mathbf{P(a_1, a_2, \ldots, a_n) = P(a_1) \cdot P(a_2) \cdot \ldots \cdot P(a_n)}$.

Количество информации в этом ансамбле равно:

$$\mathbf{I(a_1, a_2, \ldots, a_n) = - \log_2 P(a_1, a_2, \ldots, a_n) = -\sum \log_2 P(a_i) = -\sum I(a_i)}$$

Таким образом, мы убедились, что мера Шеннона (2.4) обладает свойством аддитивности.

Далее рассмотрим несколько частных случаев, вытекающих из формулы Шеннона.

Количество информации для равновероятных символов в сообщении

Пусть в сообщении из **n** элементов вероятность каждого символа равна **p**. Вероятность всего сообщения, то есть вероятность появления **n** символов равна $\mathbf{P = p^n}$. По формуле Шеннона (2.4):

$$\mathbf{I = -\log_2 P = -\log_2 p^n = -n \log_2 p}$$

С учетом того, что для равновероятных символов алфавита **p=1/m** имеем

$$\mathbf{I = -n \log_2 \frac{1}{m} = n \log_2 m} \qquad (2.5)$$

Как видим, мера Хартли (2.3) является частным случаем меры Шеннона (2.4) для случая равновероятных символов алфавита.

Количество информации для неравновероятных независимых символов в сообщении

Рассмотрим ситуацию, когда символы алфавита имеют разные вероятности.

Пусть в алфавите **A** имеется **m** элементов, и получено сообщение из **n** символов. Пусть каждый **i**-й символ встречался $\mathbf{n}_i$ раз, а вероятность появления **i**-го символа $\mathbf{p}_i$.

Сведем в таблицу статистику сообщения:

$\mathbf{a}_1$	$\mathbf{a}_2$	$\mathbf{a}_3$	…	$\mathbf{a}_m$	символы
$\mathbf{p}_1$	$\mathbf{p}_2$	$\mathbf{p}_3$	…	$\mathbf{p}_m$	вероятности
$\mathbf{n}_1$	$\mathbf{n}_2$	$\mathbf{n}_3$	…	$\mathbf{n}_m$	количество появлений

Тогда вероятность появления $\mathbf{n_i}$ раз символа $\mathbf{a_i}$ будет равна $\mathbf{p_i^{n_i}}$, а вероятность появления всего сообщения (ввиду независимости символов) будет следующей:

$$\mathbf{P} = \prod_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{p_i}^{\mathbf{n_i}} .$$

С другой стороны, вероятность $\mathbf{p_i}$ можно определить апостериорно, как частоту появления символа $\mathbf{a_i}$. Если сообщение достаточно длинное, то можно записать:

$$\mathbf{p_i} \approx \frac{\mathbf{n_i}}{\mathbf{n}} \Rightarrow \mathbf{n_i} = \mathbf{n} \cdot \mathbf{p_i}$$

Тогда $\mathbf{P} = \prod_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{p_i}^{\mathbf{n_i}} = \prod_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{p_i}^{\mathbf{np_i}}$, откуда вытекает

$$\mathbf{I} = -\log_2 \mathbf{P} = -\log_2 \prod_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{p_i^{np_i}} = -\sum_{\mathbf{i=1}}^{\mathbf{m}} \log_2 \mathbf{p_i^{np_i}} = -\mathbf{n} \sum_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{p_i} \log_2 \mathbf{p_i} \quad (2.6)$$

Формула (2.6) является частным случаем формулы Шеннона (2.4) для случая неравновероятных и независимых символов алфавита.

Рассмотрим два примера, сходные по сюжету, но принципиально отличающиеся по способу решения.

Пример 2.1. Студент Вася сообщил, что у него день рождения 25 октября. Какое количество информации он сообщил?

Решение. Вероятность этого сообщения **P** (примем, что люди рождаются равновероятно в течение года) равна **1/365≈0,0027**

По формуле (2.4): $\mathbf{I = -\log_2 P = -\log_2 0.0027 \approx 8.5}$ (бит)

Пример 2.2. У студента Васи спросили «У тебя день рождения 25 октября?» Какое количество информации содержит ответ?

Решение. Ответ представляет собой сообщение, состоящее из одного символа двоичного алфавита: «да» или «нет». Вероятность символа «да» - **1/365**, вероятность символа «нет» - **364/365**. По формуле (2.6):

$\mathbf{I = -1/365 \cdot \log_2 1/365 - 364/365 \cdot \log_2 364/365 \approx 0.027}$ (бит)

Количество информации в случае неравновероятных зависимых символов в сообщении

Предположение о независимости символов в сообщении не всегда допустимо. В реальных условиях символы и отсчеты, образующие сообщения, взаимосвязаны, например:

- снимается квантованный по уровню и времени электронный сигнал (см. рис. 2.1(б));
- количество заглавных букв в тексте связано с количеством точек;
- при передаче данных по сети количество заголовков пакета связано с количеством контрольных сумм и т.д.

Поэтому при вычислении вероятности **P** сообщения надо использовать совместные вероятности символов. Если учитывать взаимосвязь между парами символов $\mathbf{(a_i, a_j)}$, то следует использовать совместную вероятность появления пары $\mathbf{p(a_i, a_j) = p_{ij}}$. В этом случае количество информации определяется формулой

$$\mathbf{I} = -\frac{\mathbf{1}}{\mathbf{2}}\mathbf{n}\sum_{\mathbf{i}}\sum_{\mathbf{j}}\mathbf{p_{ij}}\log\mathbf{p_{ij}}, \mathbf{при}\sum_{\mathbf{i}}\sum_{\mathbf{j}}\mathbf{p_{ij}} = \mathbf{1}, \qquad (2.7)$$

а если учитывать взаимосвязь между тремя отсчетами (символами) $\mathbf{(a_i, a_j\ a_k)}$, то

$$\mathbf{I} = -\frac{\mathbf{1}}{\mathbf{3}}\mathbf{n}\sum_{\mathbf{i}}\sum_{\mathbf{j}}\sum_{\mathbf{k}}\mathbf{p_{ijk}}\,\mathbf{log}\,\mathbf{p_{ijk}}$$ и т.д.

2.3. Энтропия и ее свойства

Энтропия – мера неопределенности случайного состояния некоторой системы. Мы рассматриваем информационные системы, то есть системы, воспринимающие, хранящие, перерабатывающие и использующие информацию. Нормальное функционирование подобных систем – это прием-передача информационных сообщений. При получении сообщения неопределенность, то есть мера «незнания», уменьшается или вовсе устраняется. Таким образом, энтропия может служить информационной характеристикой количества информации, устраненной при получении сообщения.

> Для целей теории информации мы определим **энтропию** как среднее количество информации, приходящееся на одно сообщение в ансамбле сообщений (или на один символ в отдельном сообщении). Иначе говоря, **энтропия** – это математическое ожидание количества информации в сообщении.

Пусть информационная система может порождать ансамбль (алфавит) сообщений $\mathbf{a_1, a_2,\ldots,a_m}$. Вероятности каждого сообщения следующие: $\mathbf{P(a_1), P(a_2), \ldots,P(a_m)}$. Так как вероятности сообщений не одинаковы, то они несут разное количество информации, определяемое формулой Шеннона:

$$\mathbf{I(a_i) = - \log_2 P(a_i)}.$$

Среднее количество информации (математическое ожидание количества информации) ансамбля сообщений вычисляется по известной формуле:

$$\mathbf{H(a) = M[I(a)] = \sum_{i=1}^{m} P(a_i) \cdot I(a_i) = -\sum_{i=1}^{m} P(a_i) \cdot \log_2 P(a_i)}$$

Совершенно аналогично вводится энтропия сообщений:

$$\mathbf{H = -\sum_{i=1}^{m} p_i \log_2 p_i} \qquad (2.8)$$

Энтропия не зависит от конкретного сообщения. Это характеристика информационной системы (источника сообщений или канала передачи сообщений). Энтропия в таком виде является априорной характеристикой и может быть вычислена до эксперимента, если известны вероятностные характеристики сообщений. Энтропия характеризует неопределенность ситуации до передачи сообщения, поскольку заранее не известно, какое сообщение из ансамбля будет передано. Чем больше энтропия, тем сильнее неопределенность и тем большую информацию в среднем несет одно сообщение источника. Сравнивая формулы (2.8) и (2.6) видим, что $\mathbf{I = n \cdot H}$.

Свойства энтропии

1) **Энтропия принимает значение, равное 0, только в случае детерминированного источника сообщений системы**.

Детерминированность источника означает, что один из возможных символов генерируется источником постоянно (с единичной вероятностью), а остальные – не производятся вовсе. Предположим для определенности, что генерируется **k**-й символ.

Пусть $\mathbf{P(a_k)=1}$, а $\mathbf{P(a_i)=0}$ для всех $\mathbf{i=1,\ldots,k-1,k+1,\ldots,m}$, то есть, $\mathbf{i \neq k}$

Тогда, обозначив **i**-й элемент суммы в формуле (2.8) через $\mathbf{h_i}$, получим

$$\mathbf{h_k = -P(a_k) \log_2 P(a_k) = 0}$$

$$\mathbf{h_i = -0 \cdot \log_2 0 = 0 \cdot \infty \ (?)}$$

Раскроем неопределенность вида $\mathbf{0 \cdot \infty}$ по правилу Лопиталя.

$$\lim_{x\to 0}(-x\log_2 x)=\lim_{x\to 0}x\log_2\frac{1}{x}=\lim_{x\to 0}\frac{\log_2\frac{1}{x}}{\frac{1}{x}}=\left|\frac{1}{x}=z\right|=$$

$$=\lim_{z\to\infty}\frac{\log_2 z}{z}=\lim_{z\to\infty}\frac{(\log_2 z)'}{z'}=\left|(\log_2 z)'=\frac{1}{z\ln 2}\right|=\lim_{z\to\infty}\frac{1}{z\ln 2}=0$$

Следовательно, для детерминированного источника $\mathbf{h_i=0}$ для всех **i**. С другой стороны, если ни одна $\mathbf{p(a_i)\neq 1}$, то ни одно слагаемое $\mathbf{h_i}$ не обращается в **0**. Что и требовалось доказать.

2) Энтропия - величина неотрицательная и ограниченная.

Если каждое слагаемое $\mathbf{h_i=-p(a_i)\log_2 p(a_i)}$ неотрицательно и ограниченно, то и их сумма также будет неотрицательна и ограниченна.

Докажем неотрицательность:

$$\mathbf{p(a)\geq 0;\quad p(a)\leq 1 \Rightarrow \log p(a)\leq 0 \Rightarrow -p(a)\log p(a)\geq 0}$$

Докажем ограниченность, найдя экстремум, для чего продифференцируем $\mathbf{h_i}$ по **p**:

$$\frac{\partial h_i}{\partial p_i}=-\frac{\partial p_i}{\partial p_i}\log p_i-p_i\frac{\partial \log p_i}{\partial p_i}=-\log p_i-p_i\frac{1}{p_i\ln 2}=-\log p_i-\log e=0,$$

отсюда $\mathbf{p_i = 1/e}$.

Следовательно, величина $\mathbf{h_i}$ имеет экстремум (можно доказать, что это максимум), а значит это величина ограниченная (см. рис. 2.2)

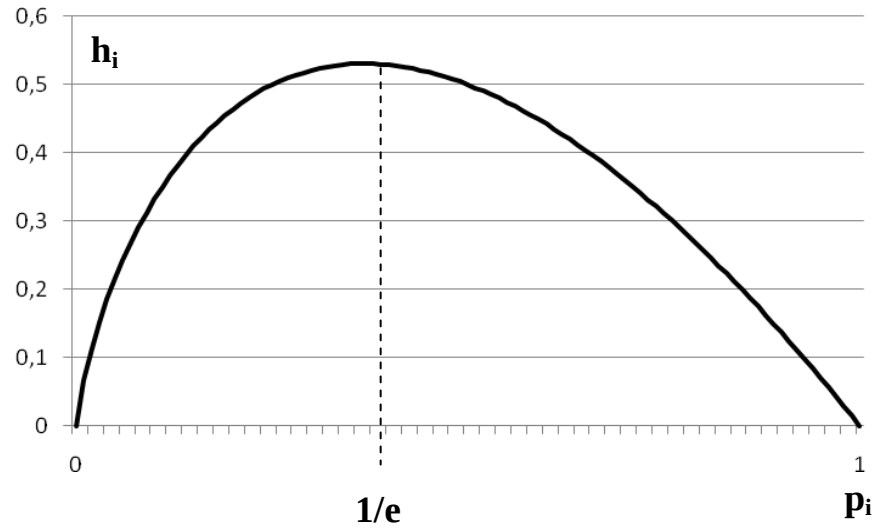


Рис. 2.2. К свойству ограниченности энтропии

3) Энтропия дискретной системы, имеющей m равновероятных состояний, максимальна и равна $\log_2 m$.

Докажем максимальность для случая **m=2.**

Пусть **m=2**, тогда $p_1+p_2=1$

$H = -p_1 \cdot \log p_1 - p_2 \cdot \log p_2 = -p_1 \cdot \log p_1 - (1-p_1) \cdot \log(1-p_1)$.

Чтобы найти максимум энтропии, определим и приравняем нулю частную производную.

$$\frac{\partial H}{\partial p_1} = -\log p_1 - \log e - \frac{\partial(1-p_1)}{\partial p_1}\log(1-p_1) - (1-p_1)\frac{\partial \log(1-p_1)}{\partial p_1} =$$

$$= -\log p_1 - \log e + \log(1-p_1) - (1-p_1)\frac{-1}{(1-p_1)\ln 2} =$$

$$= -\log p_1 - \log e + \log(1-p_1) + \log e = 0.$$

$-\log p_1 + \log(1-p_1) = 0$

$p_1 = 1/2 = p_2$

Следовательно, при двух символах в алфавите максимум энтропии достигается в случае равновероятных символов. То же можно доказать и для большего числа символов в алфавите.

Найдем значение максимальной энтропии. Пусть все символы равновероятны: $p_i = 1/m$.

$$H_{max} = -\sum_{i=1}^{m}\frac{1}{m}\log_2\frac{1}{m} = \sum_{i=1}^{m}\frac{1}{m}\log_2 m = \log_2 m$$

4) Совместная энтропия независимых источников сообщений равна сумме энтропий.

Пусть источник **A** порождает ансамбль **Ma** сообщений $(a_1, a_2, \ldots, a_{Ma})$, а источник **B** порождает ансамбль **Mb** сообщений $(b_1, b_2, \ldots, b_{Mb})$, и источники **независимы**. Общий алфавит источников представляет собой множество пар вида (a_i, b_j), общая мощность алфавита равна **Ma×Mb**.

Совместная энтропия композиции двух источников равна

$$H(A,B) = -\sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i, b_j)\log P(a_i, b_j)$$

Поскольку **A** и **B** независимы, то $P(a_i,b_j) = P(a_i)\cdot P(b_j)$,
а $\log P(a_i,b_j) = \log P(a_i) + \log P(b_j)$. Отсюда вытекает:

$$\mathbf{H(A,B) = -\sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i)P(b_j)(\log P(a_i) + \log P(b_j)) =}$$

$$\mathbf{= -\sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i)P(b_j)\log P(a_i) - \sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i)P(b_j)\log P(b_j)}$$

Изменим порядок суммирования

$$\mathbf{H(A,B) = -\sum_{i=1}^{Ma} P(a_i)\log P(a_i)\sum_{j=1}^{Mb} P(b_j) - \sum_{j=1}^{Mb} P(b_j)\log P(b_j)\sum_{i=1}^{Ma} P(a_i)}$$

учитывая, что $\sum_{i=1}^{Ma} P(a_i) = 1$ и $\sum_{j=1}^{Mb} P(b_j) = 1$, получим

$$\mathbf{H(A,B) = -\sum_{i=1}^{Ma} P(a_i)\log P(a_i) - \sum_{j=1}^{Mb} P(b_j)\log P(b_j) = H(A) + H(B)} \qquad (2.9)$$

Вывод можно распространить и на большее количество независимых источников

Условная энтропия

Найдем совместную энтропию сложной информационной системы (композиции **A**, **B**) в том случае, если их сообщения не являются независимыми, то есть если на содержание сообщения **B** оказывает влияние сообщение **A**.

Например, сообщение о матче футбольных команд Комета и Ракета «Комета выиграла» полностью снимает неопределенность о том, как сыграла Ракета.

Другой пример: сообщение **A** содержит информацию о женщине (фамилию, имя, отчество, год рождения, место рождения, образование, домашние адрес и телефон), а сообщение **B** содержит аналогичную информацию о мужчине – супруге упомянутой женщины. Очевидно, что сообщение **B** частично содержит в себе информацию **A**, а именно: фамилию жены, ее домашний адрес и телефон, скорее всего совпадающие с фамилией, домашним адресом и телефоном мужа, а также вероятностную оценку ее года рождения, который скорее всего близок к году рождения мужа. Таким образом, сообщение **B** несет для нас меньше информации, чем сообщение **A**, и совместная информация двух сообщений не является простой суммой информаций отдельных сообщений.

Пусть источник **A** порождает ансамбль **Ma** сообщений $\mathbf{(a_1, a_2,\ldots, a_{Ma})}$, источник **B** порождает ансамбль **Mb** сообщений $\mathbf{(b_1, b_2,\ldots, b_{Mb})}$ и источники **зависимы**. Общий алфавит источников представляет собой множество пар вида $\mathbf{(a_i, b_j)}$, общая мощность алфавита: **Ma×Mb**.

Энтропия сложной информационной системы (из двух источников) равна

$$\mathbf{H(A,B) = -\sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i,b_j)\log P(a_i,b_j)}$$

Поскольку **A** и **B** зависимы, то $\mathbf{P(a_i,b_j) = P(a_i)\cdot P(b_j|a_i)}$,

а $\mathbf{\log P(a_i,b_j) = \log P(a_i) + \log P(b_j|a_i)}$. Подставив это в выражение для энтропии сложной системы, получаем:

$$\mathbf{H(A,B) = -\sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i)P(b_j \mid a_i)(\log P(a_i) + \log P(b_j \mid a_i)) =}$$

$$\mathbf{= -\sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i)P(b_j \mid a_i)\log P(a_i) - \sum_{i=1}^{Ma}\sum_{j=1}^{Mb} P(a_i)P(b_j \mid a_i)\log P(b_j \mid a_i)}$$

В первом слагаемом индекс **j** имеется только у **B,** изменив порядок суммирования, получим член вида: $\mathbf{\sum_{j=1}^{Mb} P(b_j \mid a_i)}$, который равен **1** поскольку характеризует достоверное событие (какое-либо сообщений $\mathbf{b_j}$ в любом случае реализуется). Следовательно, первое слагаемое оказывается равным:

$$\mathbf{-\sum_{i=1}^{Ma} P(a_i)\log P(a_i) = H(A)}$$

Во втором слагаемом члены вида $\mathbf{-\sum_{j=1}^{Mb} P(b_j \mid a_i)\log P(b_j \mid a_i) = H(B \mid a_i)}$ имеют смысл энтропии источника **B** при условии, что реализовалось сообщение $\mathbf{a_i}$ – будем называть ее **частной условной энтропией**. Если ввести данное понятие и использовать его обозначение, то второе слагаемое будет иметь вид:

$$\mathbf{\sum_{i=1}^{Ma} P(a_i)H(B \mid a_i) = H(B \mid A)},$$

или подробнее

$$\mathbf{H(B \mid A) = -\sum_{i=1}^{Ma} P(a_i)\sum_{j=1}^{Mb} P(b_j \mid a_i)\log P(b_j \mid a_i)} \qquad (2.10)$$

где **H(B|A)** есть **общая условная энтропия** источника **B** относительно источника **A**. Окончательно получаем для энтропии сложной системы:

$$\mathbf{H(A, B) = H(A) + H(B|A)} \qquad (2.11)$$

Полученное выражение представляет собой общее правило нахождения энтропии сложной системы. Совершенно очевидно, что выражение (2.9) является частным случаем (2.11) при условии независимости источников **A** и **B**.

Относительно условной энтропии можно высказать следующие утверждения:

1. Условная энтропия является величиной **неотрицательной**. Причем **H(B|A) = 0** только в том случае, если **любое** сообщение **A** полностью определяет сообщение **B**, т.е.

$$H(B|a_1) = H(B|a_2) = \ldots = H(B|a_N) = 0$$

В этом случае **H(A,B) = H(A)**.

2. Если источники **A** и **B** независимы, то **H(B|A) = H(B)**, причем это оказывается **наибольшим** значением условной энтропии. Другими словами, сообщение источника **A** не может повысить неопределенность сообщения источника **B**; оно может либо не оказать никакого влияния (если источники независимы), либо понизить энтропию **B**.

Приведенные утверждения можно объединить одним неравенством:

$$0 \le H(B|A) \le H(B), \qquad (2.12)$$

т.е. **условная энтропия не превосходит безусловную**.

3. Из соотношений (2.11) и (2.12) следует, что

$$H(A, B) \le H(A) + H(B),$$

причем равенство реализуется только в том случае, если источники **A** и **B** независимы.

Энтропия источника непрерывных сообщений

Рассмотрим систему, где качественные признаки состояния изменяются непрерывно (непрерывный сигнал). Вероятность нахождения системы в состоянии **x** (т.е. сигнал принимает значение **x**) характеризуется плотностью вероятности **f(x).** Чтобы найти энтропию такого сообщения, разбиваем диапазон возможного изменения сигнала на дискреты размером Δx.

Вероятность нахождения системы в **i**-й дискрете равна

$$P(x_i) = f(x_i)\cdot\Delta x$$

Тогда энтропия системы вычисляется так:

$$H = -\sum_i f(x_i)\Delta x \log(f(x_i)\Delta x) = -\sum_i f(x_i)\Delta x \cdot (\log f(x_i) + \log \Delta x) =$$

$$= -\sum_i f(x_i)\log f(x_i)\Delta x - \log \Delta x \cdot \sum_i f(x_i)\Delta x$$

при малых Δx:

$$\sum_i f(x_i)\log f(x_i)\Delta x \approx \int_{-\infty}^{\infty} f(x)\log f(x)dx$$

А также $\sum_i f(x_i)\Delta x \approx \int_{-\infty}^{\infty} f(x)dx = 1$

Таким образом

$$\mathbf{H = -\int_{\infty}^{\infty} f(x)\log f(x)dx - \log \Delta x} \qquad (2.13)$$

Положим $\Delta\mathbf{x=1}$ (это возможно сделать, выбрав соответствующий масштаб и единицу измерения), тогда $\mathbf{H = H^* = -\int_{-\infty}^{\infty} f(x)\log f(x)dx}$

Величина **H*** называется **приведенной** или **дифференциальной** энтропией.

При уменьшении $\Delta\mathbf{x}$ приведенная энтропия **H** стремится к ∞. Это естественно, так как чем точнее мы хотим задать состояние системы, тем большую степень неопределенности мы должны устранить. Дифференциальная энтропия не является мерой количества информации, хотя и характеризует степень неопределенности, присущую источнику.

2.4. Количественные характеристики источника сообщений

Относительная энтропия

Идеальные источники сообщений, имеющие максимальную энтропию, оптимальны в том смысле, что в них на один символ (элемент, уровень квантования) приходится наибольшее количество информации.

В реальных сообщениях символы всегда коррелированны (после запятой не появляется точка, после гласной – мягкий знак), вследствие чего количество информации, приходящееся на один символ, будет меньше, чем в идеальных. Соотношение реальных и оптимальных сообщений выражается посредством **коэффициента сжатия μ(s)** (иное название – относительная энтропия)/

$$\mathbf{\mu(s) = H_p(s) / H_0(s) = n_0 / n_p}$$

где $\mathbf{H_p(s)}$ и $\mathbf{H_0(s)}$ – энтропия реального и идеального источника сообщений соответственно, $\mathbf{n_0}$ и $\mathbf{n_p}$ – количество символов оптимального и реального сообщения.

Одно и то же количество информации **I(s)** может содержаться в сообщении, состоящим из $\mathbf{n_p}$ символов с энтропией $\mathbf{H_p(s)}$ или из $\mathbf{n_0}$ символов с энтропией $\mathbf{H_0(s)}$

$\mathbf{I(s) = n_p \cdot H_p(s) = n_0 \cdot H_0(s)}$, а так как $\mathbf{H_p(s) \le H_0(s)}$, то $\mathbf{n_p \ge n_0}$.

Избыточность источника сообщений

Поскольку реальные источники информации имеют энтропию, меньшую оптимальной, то сообщения таких источников содержат избыточные символы. Коэффициент избыточности **φ** выражается так:

$$\varphi(s) = \frac{n_p - n_0}{n_p} = 1 - \frac{n_0}{n_p} = 1 - \mu(s)$$

$$\varphi(s) = \frac{H_0 - H_p}{H_0} = 1 - \frac{H_p}{H_0}$$

Коэффициент избыточности показывает, какая часть реального сообщения является излишней и могла бы не передаваться, если бы источник сообщений был организован оптимально.

Экономичность источников информации

Для сокращения длины сообщений без потери информации, очевидно, надо увеличивать энтропию источника. Энтропию можно увеличивать за счет обеспечения равновероятности символов алфавита, а также за счет увеличения мощности алфавита. Однако увеличение мощности алфавита приводит к сложностям приема-передачи информации (непрерывный сигнал передается и воспринимается с погрешностями, китайские иероглифы трудны для освоения, для них не хватает клавиш на клавиатуре), к увеличению избыточности сообщений (в языках программирования ряд команд применяется редко). Существует теоретический оптимум для мощности алфавита. Найдем его.

Пусть имеется источник с алфавитом мощности **m**. Тот же алфавит можно получить, используя два источника с алфавитами **m/2** или три источника с алфавитами **m/3** и т.д. При какой мощности алфавита **m** общая энтропия будет максимальной, если **k·m = const**, где **k** – количество независимых источников, а **m** – это мощность алфавита каждого источника? (Под независимыми источниками можно понимать и независимые сигналы одного источника.)

Пусть **k·m = a**. Энтропия композиции независимых источников равна

$$H = \sum_{i=1}^{k} H_i = k \cdot H_i = k \cdot \log m$$

$$k = a/m$$

$$H = \frac{a}{m} \log_2 m$$

Найдем максимум энтропии, для чего продифференцируем по **m**

$$\frac{\partial H}{\partial m} = \frac{a}{m} \frac{1}{m \ln 2} - \frac{a}{m^2} \log m = \frac{a}{m^2} \log_2 e - \frac{a}{m^2} \log_2 m = 0$$

$$\frac{a}{m^2} \log_2 e = \frac{a}{m^2} \log_2 m$$

$$m = e$$

Оптимальная мощность алфавита теоретически равна основанию натуральных логарифмов **e** (**2.718281828459045**…), а практически – трем, так как невозможно придумать физически реализуемый объект с нецелым количеством состояний.

> Очевидно, что троичный алфавит является более экономичным, чем двоичный. Именно поэтому в истории развития вычислительной техники были случаи создания компьютеров, использующих троичный алфавит. В 1958 году группа советских инженеров под руководством конструктора Н.П. Брусенцова представила электронно-вычислительную машину «Сетунь», работающую на принципах троичной логики. Элементной базой такого компьютера были магнитные усилители на ферритовых сердечниках. Они допускали три устойчивых состояния: ток в прямом направлении (логическая «единица»), ток в обратном направлении (логическая «минус единица») и отсутствие тока (логический «ноль»). Машины этой серии выпускались с 1962 по 1964 год и отличались исключительной надежностью. Архитектурно они были совершеннее «двоичных» полупроводниковых аналогов. Помешали их массовому распространению миниатюризация, удешевление и повышение надежности полупроводниковых элементов. «Сетунь» стала экономически невыгодной.

Производительность источника сообщений

Производительностью источника называется количество информации, порождаемое источником в среднем за единицу времени

Пусть **H** – энтропия источника, **m** – мощность алфавита, $\mathbf{p_i}$ **(i=1, 2,…, m)** – вероятность появления **i**-го символа, $\boldsymbol{\theta_i}$ – длительность генерации **i**–го символа. Рассмотрим процесс генерации **n** символов. В среднем, один символ генерируется за время $\mathbf{M[\theta]} = \sum_{i=1}^{m} \boldsymbol{\theta_i} \mathbf{p_i}$. На генерацию **n** символов будет затрачено время **T=n·M[θ]**.

Количество информации, порожденное источником за это время, равно: **I = n·H**. Производительность источника будет вычислена следующим образом:

$$\mathbf{R} = \frac{\mathbf{I}}{\mathbf{T}} = \frac{\mathbf{n} \cdot \mathbf{H}}{\mathbf{n} \cdot \mathbf{M[\theta]}} = \frac{-\sum_{i=1}^{m} \mathbf{p_i} \log \mathbf{p_i}}{\sum_{i=1}^{m} \boldsymbol{\theta_i} \mathbf{p_i}}$$

Если все символы алфавита генерируются за одно и то же время **θ**, то

$$\mathbf{R} = \frac{-\sum_{i=1}^{m} \mathbf{p_i} \log \mathbf{p_i}}{\boldsymbol{\theta}}.$$

Максимальной производительностью обладает источник с максимальной энтропией, которая в соответствии с третьим свойством энтропии равна $\log_2 \mathbf{m}$:

$$\mathbf{R_{max}} = \frac{\log_2 \mathbf{m}}{\boldsymbol{\theta}}. \qquad (2.14)$$

2.5. Краткий обзор иных теорий информации

Определение количества информации, безусловно, имеет важное значение в практических целях организации передачи и хранения информации. Однако, количество информации совсем не обязательно соответствует интуитивно понимаемой важности информации (вспомним пример с кодовой фразой «Над всей Испанией безоблачное небо», несущей всего **1** бит информации). Существует ряд подходов, позволяющих оценить смысловую сторону сообщения. Такие методы составляют **семантическую теорию информации**. Представителями этой теории являются, например, Р.Карнап[1] и И.Бар-Хиллель. Они предложили измерять величину семантической информации с помощью так называемой логической вероятности, которая является мерой подтверждения той или иной гипотезы. Количество семантической информации, содержащейся в сообщении, тем больше, чем меньше степень подтверждения априорной гипотезы. В предельном случае, если вся гипотеза основана на предварительных данных, которые целиком повторяются в сообщении, такое сообщение семантической информации не несет.

Советский ученый А.А.Харкевич высказался в том смысле, что ценность информации связана с целями получателя сообщения[2]. Если полученное сообщение способствует достижению цели, повышает вероятность ее достижения, то такая информация имеет положительную ценность для потребителя. Он ввел меру информации следующим образом:

$$\mathbf{I = \log p_1 - \log p_2},$$

где $\mathbf{p_1}$ и $\mathbf{p_2}$ – вероятности достижения цели соответственно после и до получения сообщения. Семантическая теория информации определяет меру нового знания с точки зрения ее полезности для достижения целей получателя. Поэтому такую трактовку еще называют **прагматической теорией** или **теорией полезности информации**. Семантическая теория определяет не меру полезности информации вообще, а ее полезность относительно данного приемника и тем самым тесно связана с информационным тезаурусом адресата.

Еще один подход, названный **теорией сложности** или **алгоритмической теорией информации**, продемонстрировал в 1960-х годах академик А.Н.Колмогоров[3]. Он высказал предположение, что вместо, например, длинной строки двоичных символов можно было бы передавать код программы, способной генерировать эту строку. Исходя из этих соображений, А.Н.Колмогоров определил сложность информации, содержащейся в сообщении, как длину кратчайшей программы для универсального компьютера,

---

[1] Карнап, Р. Значение и необходимость. М.,1959.
[2] Харкевич, А.А. Теоретические основы радиосвязи. М., 1957
[3] Колмогоров, А. Н. «Три подхода к определению понятия «количество информации»».// Проблемы передачи информации, т.1. 1964.

генерирующую это сообщение. В качестве универсального компьютера можно использовать машину Тьюринга, моделирующую функционирование любого другого универсального компьютера. При некоторых допущениях сложность по Колмогорову есть не что иное, как шенноновская энтропия, так как в среднем длина программы, способной воспроизвести случайный объект, равна энтропии случайной системы, из которой извлечен этот объект.

Если буквой **X** обозначить входные данные для программы **P**, а буквой **Y** – выходные, то относительной сложностью сообщения **Y** относительно **X** считается минимальная длина программы, преобразующая **X** в **Y**. Дав этой сложности обозначение K(Y|X), можно записать формулу Колмогорова для оценки алгоритмического количества информации в сообщении **Y** относительно сообщения **X**:

$$\mathbf{I(Y,X) = K(Y|1) – K(Y|X).}$$

Алгоритмическая информация может принимать как положительные, так и отрицательные значения, но не меньше некоторой отрицательной константы, зависящей от выбранного метода программирования. А.Н. Колмогоров сам указывал на недостатки такого подхода, связанные с тем, что алгоритмическая оценка зависит от выбранного метода программирования, то есть имеет, по сути, субъективный характер.

Вопросы и задачи к главе 2

1. В студенческой группе **24** человека: **21** юноша и **3** девушки. Определить количество информации, содержащееся в сообщении, что староста группы – девушка.

Решение. Вероятность того, что староста группы – девушка, равна $\mathbf{P = 3/24 = 1/8}$ (считаем, что решение о назначении старосты не зависит от пола студента). По формуле (2.4) количество информации в таком сообщении равно $\mathbf{- \log_2 P = - \log_2 1/8 = 3}$ (бита).

2. Известно, что в студенческой группе **2** отличника, **16** хорошистов, **6** троечников и **4** двоечника. Какое количество информации содержится в сообщениях: а) «Коля хорошист»; б) «Коля не двоечник»; в) «Коля учится на 4 и 5»?

3. У Коли на втором курсе было **8** предметов и **4** аттестации. Нам сообщили, что у него не было двоек и троек. Какое количество информации нам сообщили?

4. В лотерее **N** билетов, из них **k** выигрышных. Студент Вася купил **M** билетов и после розыгрыша сообщил вам, что выиграл (но, возможно, и не на один билет). Какое количество информации вы получили?

5. «Граждане встречающие! Поезд 26 прибывает на 7-й путь 4-й платформы в 19:35» Какое количество информации содержится в этом сообщении? Всего на вокзале **6** платформ, на каждой по **2** пути, ежедневно прибывает **120** поездов.

6. Вася выучил стихотворение, в котором **32** слова. Оцените информацию, которую он запомнил. Считается, что в русском языке **10000** слов.

7. У Васи есть игральная карта. Известно, что это карта трефовой масти. Какое количество информации нам известно?

8. Нам сказали, что сумма чисел у кости домино равна **4**. Какое количество информации нам сообщили?

9. Нам сообщили, что у кости домино числа разные. Какое количество информации нам сообщили? Какое количество осталось сообщить, чтобы однозначно определить кость домино?

10. Мы сказали продавцу, что хотим купить коньки. Коньки бывают обычные и роликовые, белого, черного и зеленого цветов, есть размеры 38, 39, 41 и 42. Какое количество информации нужно сообщить продавцу?

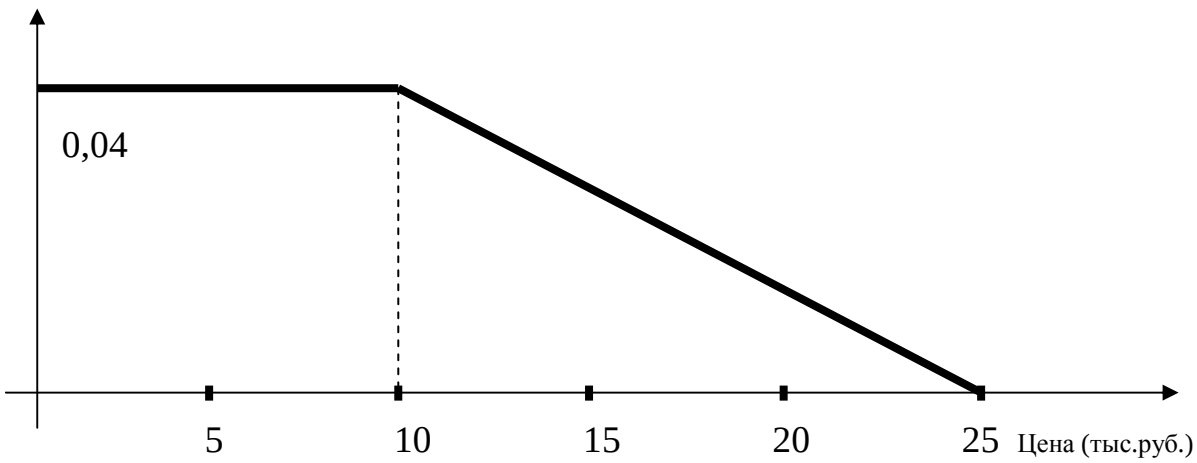
11. Бросаются одновременно две игральные кости. Определить количество информации, содержащееся в сообщении о том, что произведение числа выпавших очков четно.

12. Бросаются одновременно две игральные кости. Определить количество информации, содержащееся в сообщении о том, что сумма числа выпавших очков больше **4**.

13. Сообщение содержит **32 Кбайта** информации. Оно состоит из **64** страниц, на каждой странице по **16** строк, разбитых на **4** абзаца, в каждой строке **64** символа. Найти мощность алфавита такого сообщения.

14. Два стрелка, для которых вероятности попадания в мишень равны соответственно **0.6** и **0.7**, производят по одному выстрелу. В результате оказалось, что мишень поражена. Какое количество информации содержится в этом сообщении?

15. Магазин торгует бытовой техникой ценой до **25** тыс. руб. Плотность распределения вероятности покупки в зависимости от цены товара представлена на рисунке.



30% посетителей покидают магазин без покупки. На выходе из магазина покупателей опрашивает маркетинговая группа, которая выясняет стоимость покупки и заполняет анкету. В анкете **4** позиции: 1) без покупки; 2) покупка на сумму до **5** тыс. р.; 3) покупка на сумму от **5** до **10** тыс. руб.; 4) покупка на сумму свыше **10** тыс. руб. Маркетологи опросили **100** посетителей магазина. Какое количество информации получили они?

16. Имеются два ящика, в каждом из которых по 12 шаров. В первом – 3 белых, 3 черных и 6 красных; во втором – по 4 шара каждого цвета. Опыты состоят в вытаскивании по одному шару из каждого ящика. Каково количество информации, содержащееся в сообщении об исходе опыта?

Решение. Вероятности исходов опыта для первого ящика: $\mathbf{p_{бел}=3/12}$, $\mathbf{p_{чер}=3/12}$, $\mathbf{p_{крас}=6/12}$. Сообщение об исходе опыта содержит один символ, всего в алфавите источника сообщений **3** символа («белый», «черный», «красный»). Вероятности каждого символа равны вероятностям соответствующих исходов опыта. По формуле (2.6):

$\mathbf{I_1 = -3/12 \cdot \log_2 3/12 - 3/12 \cdot \log_2 3/12 - 6/12 \cdot \log_2 6/12 = 1.5}$ (бит)

Вероятности исходов опыта для второго ящика: $\mathbf{p_{бел}=4/12}$, $\mathbf{p_{чер}=4/12}$, $\mathbf{p_{крас}=4/12}$.

$\mathbf{I_2 = -4/12 \cdot \log_2 4/12 - 4/12 \cdot \log_2 4/12 - 4/12 \cdot \log_2 4/12 \approx 1.58}$ (бит)

17. Перед нами **5** черных ящиков. В каждом из них находится либо черный, либо белый шарик. Нам говорят, что три шарика белые. Какое количество информации мы получили?

18. Перед нами **N** черных ящиков. В каждом из них находится либо черный, либо белый шарик. Нам говорят, что **60%** шариков – белые. Какое количество информации мы получили?

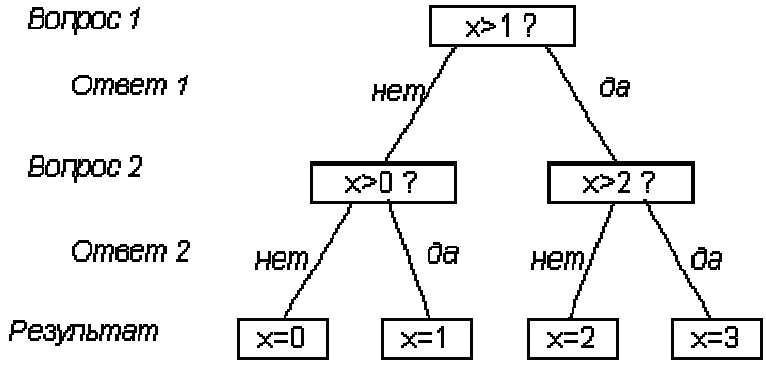
19. Какое количество информации требуется, чтобы узнать исход броска монеты?

Решение. В данном случае **n=2** и события равновероятны, т.е. $\mathbf{p_1=p_2=0,5}$. Согласно (2.6): $\mathbf{I = -0,5 \cdot \log_2 0,5 - 0,5 \cdot \log_2 0,5 = 1}$ бит

20. 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 мы должны получить, чтобы узнать задуманное число, т.е. полностью снять начальную неопределенность? Как правильно построить процесс угадывания?

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

Теперь выясним, какие вопросы необходимо задать, чтобы процесс угадывания был оптимальным, т.е. содержал минимальное их число. Здесь удобно воспользоваться так называемым выборочным каскадом:



Таким образом, для решения задачи оказалось достаточно двух 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.

21. В Белгороде **340000** жителей. Какое минимальное количество вопросов, требующих ответа "**да**" или "**нет**", необходимо, чтобы однозначно найти одного жителя?

Решение. Каждый житель – элемент случайной системы, количество элементов **m = 340000**. Ответ на один вопрос дает **1** бит информации, следовательно, количество информации равно количеству вопросов. По формуле Хартли $\mathbf{I=n\cdot\log_2 m}$. $\mathbf{I=\log_2 340000\approx 18.4}$ (бит). Округляя в большую сторону, получаем ответ – **19** вопросов.

22. Коля съел на перерыве шоколадку, яблоко и кекс. Сколько бинарных вопросов надо задать, чтобы узнать, в какой последовательности он их съел?

23. Случайным образом вынимается карта из колоды в **32** карты. Какое количество информации требуется, чтобы угадать, что это за карта? Как построить угадывание?

24. В БГТУ **12000** студентов. Какое минимальное количество вопросов, требующих ответа "да" или "нет", необходимо, чтобы однозначно найти одного студента.

25. Имеются три города (**А**, **Б** и **В**), причем жители **А** во всех случаях говорят правду, жители **Б** - только неправду, а жители **В** через раз отвечают на вопросы верно и неверно. Наблюдатель **Н** хочет выяснить, в каком городе он находится и в каком городе живет встреченный им человек. Какое минимальное количество вопросов ему потребуется задать этому встречному, если на все вопросы последний отвечает лишь "да" или "нет"? Все предположения равновероятны.

26. По данным социологических опросов, проведенных накануне президентских выборов на Украине, В.Янукович имел рейтинг **35%**, Ю.Тимошенко – **25%**, С.Тигипко – **10%**, А.Яценюк – **8%**, остальные кандидаты делят оставшиеся голоса. Какое количество информации мы получим после объявления победителя выборов?

27. Орудие стреляет по удаленной цели. При каждом выстреле она поражается с вероятностью **p = 0.1**. Разведка может только один раз проверить, поражается ли цель хоть один раз. Через какое количество выстрелов **k** следует провести проверку, чтобы она дала максимальное количество информации?

28. Датчик технологического процесса имеет погрешность **1%**. Датчик опрашивается с интенсивностью **5 с**. Определить общее количество информации, поступившее от датчика за **2 мин**, энтропию и скорость передачи информации датчиком.

29. АСУТП посредством АЦП опрашивает потенциометрические датчики **Д1** и **Д2**, имеющие погрешность δ_1**=0,1%** и δ_2**=0,2%** соответственно. Датчик **Д1** опрашивается с интервалом **0,25** с, датчик **Д2** – с интервалом **0,2** с. Определить общее количество информации, поступившее от датчиков за **2** с. и скорость передачи информации каждым датчиком.

Решение. Число различных уровней квантования (число символов в алфавитах датчиков):

$$\mathbf{m_1 = \frac{100\%}{\delta_1} = 1000, m_2 = \frac{100\%}{\delta_2} = 500}$$

Количество символов сообщения (отсчетов), поступивших за **2** с:

$$\mathbf{n_1 = \frac{2}{0.25} = 8, n_2 = \frac{2}{0.2} = 10}.$$

По формуле (2.5)

$\mathbf{I = n_1 \cdot \log_2 m_1 + n_2 \cdot \log_2 m_2 = 8 \cdot \log_2 1000 + 10 \cdot \log_2 500 \approx 170}$ (бит)

скорость передачи **J=170** бит/ **2** с **= 85** (бит/с)

Для сравнения: скорость обычной речи – примерно **20** бит/с, муравьи обмениваются информацией со скоростью **0.1** бит/с.

30. Символы азбуки Морзе могут появиться в сообщении с вероятностями: для точки - **0.51**, для тире - **0.31**, для промежутка между буквами - **0.12**, между словами - **0.06**. Определить среднее количество информации в сообщении из **500** символов данного алфавита, считая, что связь между последовательными символами отсутствует.

31. Для передачи сообщений используется алфавит из **32** прописных русских букв[1] (не используется **«Ё»**). Все передаваемые слова содержат ровно по **8** букв. Каждое передаваемое слово начинается с одной из четырех букв (**К, Л, М, Н**). Остальные буквы в каждом слове могут быть любыми из используемого алфавита. Какое количество информации (в битах) несет произвольная фраза из **10** слов, если для ее кодирования использовалось минимальное количество бит в рамках описанных выше правил. В ответе укажите целое число.

32. Имеется **12** монет одного достоинства; **11** из них имеют одинаковый вес, а одна - фальшивая - отличается по весу от остальных (причем неизвестно, легче она или тяжелее настоящих). Каково наименьшее число взвешиваний на чашечных весах без гирь, которое позволяет обнаружить фальшивую монету и выяснить, легче она, чем остальные монеты, или тяжелее?

33. Радиотехническое устройство состоит из **5** блоков (**А, Б, В, Г,** Д). Блок **А** в среднем выходит из строя **1** раз в **100** дней, блок **Б - 1** раз в **25** дней, **В - 1** раз в **5** дней, **Г - 1** раз в **4** дня и **Д - 1** раз в **2** дня. Контрольный прибор позволяет за одно измерение проверить работоспособность в целом любой комбинации блоков. Как нужно проводить контроль, чтобы затратить на поиски неисправного блока в среднем минимальное количество проверок? Найти это среднее значение.

34. Полиция ищет преступника. На допрос вызван его предполагаемый сообщник, которому, чтобы обезопасить себя от уголовного преследования, надо назвать три правдивых приметы разыскиваемого. Какие приметы он должен сообщить полиции, чтобы дать минимальное количество информации? Рассчитайте это количество информации.

Информация к размышлению:

Истинные приметы разыскиваемого	Статистика примет потенциальных подозреваемых
1	2
Нос прямой	Прямой нос – 40%, курносый нос – 30%, мясистый нос – 30%
Лицо круглое	Овальное лицо – 40%, круглое лицо – 30%, треугольно лицо – 20%, квадратное лицо – 10%
Волосы темные	Темные волосы – 60%, светлые волосы – 20%, русые волосы – 20%
Рост 180-185 см	Рост случайно распределен по закону $\mathbf{f(P)=\frac{1}{10\sqrt{2\pi}}exp\left(-\frac{(P-170)^2}{200}\right)}$

[1] Сюжет задачи взят с олимпиады СПбИТМО

Окончание

1	2
Возраст 20-25 лет	Возраст равномерно распределен в интервале от 16 до 66
Родинка на правой щеке	10% имеют родинку на одной щеке
Татуировка на предплечье	80% имеют татуировку на предплечье

35. Шпион имеет возможность похитить только три документа, содержащие различную информацию. Какие документы он должен похитить, чтобы количество похищенной информации было максимальным. Рассчитайте это количество информации.

Информация к размышлению (размышлять быстро ☺)

Содержание документов	Вспомогательная информация
Фамилия «крота» в контрразведке	В контрразведке работают 50 человек
Московский телефон резидента	В Москве 7-значные номера
Город главного удара	В списке потенциальных направлений главного удара – 20 городов
Диапазон частот для секретной радиосвязи 100-105 МГц	Диапазон возможных частот случайно распределен по закону $\mathbf{f(Д)=\frac{1}{5\sqrt{2\pi}}exp\left(-\frac{(Д-100)^2}{50}\right)}$
Толщина лобовой брони нового танка не менее 125 мм	Толщина брони танков соответствующего класса равномерно распределена в интервале от 95 до 170 мм
Шифровальная комната контрразведки находится на 7-м этаже	Контрразведка располагается в 10-этажном здании
Главный аналитик – майор	70% контрразведчиков – майоры

36. Служебная таможенная собака различает **50** запахов запрещенных веществ. На обнюхивание одного объекта она затрачивает **1** мин. Какое количество информации поступает собаке за 2-часовую смену?

37. Эллочка-Людоедка знает **20** слов. В обычном состоянии она произносит в среднем **50** слов в минуту. Причем слова «мрак», «жуть», «хо-хо» и «парниша» она произносит в четыре раза чаще других слов. Какое количество информации получит от нее Остап Бендер в течение получасового общения?

Решение. Мощность алфавита Эллочки $\mathbf{m = 20}$. Символы алфавита неравновероятны. Найдем вероятности символов. Пусть вероятности слов

«мрак», «жуть», «хо-хо» и «парниша» - $\mathbf{p_1}$, вероятности остальных шестнадцати слов – $\mathbf{p_2}$. По условию задачи $\mathbf{p_1 = 4 \cdot p_2}$. По условию нормировки $\mathbf{\Sigma p_i = 1}$.

$\mathbf{4 \cdot p_1 + 16 \cdot p_2 = 1;\ 16 \cdot p_2 + 16 \cdot p_2 = 1;\ p_2 = 1/32;\ p_1 = 1/8}$.

За полчаса Эллочка произнесет **50·30 = 1500** слов. Длина сообщения **n=1500**.

По формуле (2.6) количество информации $\mathbf{I = -n \cdot \Sigma p_i \cdot \log_2 p_i =}$
$\mathbf{= -1500 \cdot (4 \cdot p_1 \cdot \log_2 p_1 + 16 \cdot p_2 \cdot \log_2 p_2) = -1500 \cdot (-4/8 \cdot \log_2 8 - 16/32 \cdot \log_2 32) =}$
$\mathbf{= 1500 \cdot (3 \cdot 4/8 + 5 \cdot 16/32) = 1500 \cdot 4 = 6000}$ (бит)

38. Индеец Чуткое Ухо различает голоса **100** видов птиц. Из них **20** видов встречаются часто – составляют **70%** птиц, обитающих в прериях, а **80** видов – редкие птицы. Найти среднее количество информации, содержащееся для индейца в птичьем крике.

39. Индеец Меткий Глаз видит бизона на расстоянии до **1800** м и может определить расстояние до него с точностью до **20** м. Найти общее количество информации, которое получает Меткий Глаз о расположении стада бизонов, если в нем находится **20** особей.

40. На шахматной доске в одной из клеток поставлена фигура. Вероятность нахождения фигуры на любой клетке одинакова. Определить информацию, получаемую от сообщения с координатами фигуры на доске.

41. Во время пожара температура вблизи датчика пожарной сигнализации распределена по закону $\mathbf{f(t) = \frac{1}{10\sqrt{2\pi}} \exp\left(-\frac{(t-300)^2}{200}\right)}$. При нахождении температуры в диапазоне от **280** до **320** градусов датчик срабатывает с вероятностью $\mathbf{p_1 = 0.8}$, при температуре выше **320** градусов датчик срабатывает с вероятностью $\mathbf{p_2 = 0.95}$. Найти количество информации, которое поступает на пульт пожарной охраны во время пожара.

42. Алфавит источника сообщений состоит из двух элементов: **a** и **b**. Рассмотрим случаи передачи информации с помощью этих элементов:

1) Элементы независимы и равновероятны. Количество информации, которое несет каждый символ сообщения (см. формулу (2.5)):

$\mathbf{I_1 = \log_2 m = \log_2 2 = 1}$ (бит), где **m** – мощность алфавита.

2) Элементы независимы и неравновероятны. Пусть $\mathbf{P(a) = \frac{3}{4}}$, а $\mathbf{P(b) = \frac{1}{4}}$. Тогда количество информации на один символ составит по формуле (2.6):

$$\mathbf{I_2 = -\sum_{i=1}^{2} p_i \log p_i = -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right) \approx 0.815} \text{ (бит).}$$

3) Элементы взаимозависимы и неравновероятны. Пусть, например, **P(a) = 3/4, P(b) = 1/4.**

Пусть **P(a|a) = 2/3**, **P(b|b) = 0** – вероятности повторения символов

P(a|b) = 1, **P(b|a) = 1/3** - вероятности чередования символов

Чтобы применить формулу (2.7) воспользуемся правилом **P(X,Y) = P(X)·P(Y|X)**.

P(a,a) = P(a)·P(a|a) = 3/4·2/3 = 1/2;

P(a,b) = P(a)·P(b|a) = 3/4·1/3 = 1/4;

P(b,a) = P(b)·P(a|b) = 1/4·1 = 1/4;

P(b,b) = P(b)·P(b|b) = 1/4·0 = 0.

По формуле (2.7): $\mathbf{I_3 = -1/2\cdot(1/2\cdot log1/2+1/4\cdot log1/4+1/4\cdot log1/4+0) = 0.75}$ (бит).

43. Система **X** имеет восемь равновероятных состояний. Определить энтропию. ***Решение***. n = 8, $\mathbf{H(x) = \log_2 8 = 3}$ (бит/символ).

44. Источник генерирует знак $\mathbf{z_1}$ с вероятностью **0.8** и $\mathbf{z_2}$ с вероятностью **0.2**. Какова энтропия источника?

***Решение*.** $\mathbf{H = -(0.8\cdot log_2 0.8 + 0.2\cdot log_2 0.2) = 0.72}$ (бит/символ)

45. Определить энтропию системы, состояние которой описывается таблицей.

x_i	x_1	x_2	x_3	x_4	x_5
p_1	**0.01**	**0.01**	**0.01**	**0.01**	**0.96**

46. Алфавит состоит из букв **a, b, c, d**. Даны вероятности: $\mathbf{p_a = p_b = 0{,}25;\ p_c = 0{,}34}$. Найти энтропию источника сообщений.

47. Найти число значений **m** случайной величины **Y**, все значения которой одинаково вероятны. При этом необходимо, чтобы энтропия **Y** была равна энтропии случайной величины **X**, вероятности значений которой заданы таблицей:

x_j	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(x_j)$	**1/2**	**1/4**	**1/8**	**1/16**	**1/32**	**1/64**	**1/128**	**1/128**

48. В буфере информационной системы ожидают обработки **6** заданий. **2** из них запрашивают дополнительный ресурс (например, принтер), **4** – не требуют ресурса. Последовательно в случайном порядке отправляются на обработку два задания. Найти энтропию запроса дополнительного ресурса.

Решение. Будем считать сообщением **A** – посылку первого задания на выполнение. Ресурс потребуется с вероятностью $\mathbf{P(A_1)=2/6=1/3}$. Ресурс не потребуется с вероятностью $\mathbf{P(A_2)=2/3}$. Энтропия сообщения **A** равна:

$\mathbf{H(A)= -P(A_1)logP(A_1) - P(A_2)logP(A_2)= -1/3\ log1/3 - 2/3\ log\ 2/3 = 0.918}$ бит

Сообщение **B** – посылка второго задания на выполнение. Вероятность того, что потребуется ресурс $\mathbf{P(B_1)}$, зависит от того, какое задание было послано первым.

при $\mathbf{A_1}$: $\mathbf{P(B_1|A_1)=1/5,\ P(B_2|\ A_1)=4/5;\ H(B|A_1)= -1/5\cdot log1/5 - 4/5\cdot log4/5 = 0.722}$

при A_2: $P(B_1|A_2)=2/5$, $P(B_2|A_2)=3/5$; $H(B|A_2)= -2/5\cdot\log 2/5 – 3/5\cdot\log 3/5 = 0.971$

Следовательно, энтропия сообщения **B** равна

$H(B|A)= P(A_1)\cdot H(B|A_1)+P(A_2)\cdot H(B|A_2) = 1/3\cdot 0.722 + 2/3\cdot 0.971 = 0.888$ бит

Обратим внимание, что энтропия сообщения **B** оказалась меньше, чем сообщения **A**. Это естественно, так как, получив информацию об исходе **A**, у нас уменьшилась неопределенность относительно исхода **B**. Полная совместная энтропия получается по формуле (2.11):

$H(A,B)= 0.918 + 0.888 = 1.806$ бит.

49. В ящике имеются **2** белых шара и **4** черных. Из ящика извлекают последовательно два шара без возврата. Найти энтропию, связанную с первым и вторым извлечениями, а также энтропию обоих извлечений.

50. Имеется три тела с одинаковыми внешними размерами, но с разными массами x_1, x_2 и x_3. Необходимо определить энтропию, связанную с нахождением наиболее тяжелого из них, если сравнивать веса тел можно только попарно.

Решение. Последовательность действий достаточно очевидна: сравниваем вес двух любых тел, определяем из них более тяжелое, затем с ним сравниваем вес третьего тела и выбираем наибольший из них. Поскольку внешне тела неразличимы, выбор номеров тел при взвешивании будет случаен, однако общий результат от этого выбора не зависит. Пусть опыт **A** состоит в сравнении веса двух тел, например, первого и второго. Этот опыт, очевидно, может иметь два исхода: A_1: $x_1 > x_2$, его вероятность $p(A_1) = 1/2$; исход A_2: $x_1 < x_2$; также его вероятность $p(A_2)=1/2$.

$H(A) = –1/2 \log_2 1/2 – 1/2 \log_2 1/2 = 1$ (бит)

Опыт **B** – сравнение весов тела, выбранного в опыте **A**, и третьего – имеет четыре исхода: B_1: $x_1> x_3$, B_2: $x_1< x_3$, B_3: $x_2> x_3$, B_4: $x_2< x_3$; вероятности исходов зависят от реализовавшегося исхода **A** – для удобства представим их в виде таблицы:

	B_1	B_2	B_3	B_4
A_1	1/2	1/2	0	0
A_2	0	0	1/2	1/2

Вновь, воспользовавшись формулами (2.10) и (2.11), находим:

$$H(B \mid A_1) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H(B \mid A_2) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$H(B|A)=p(A1)\cdot H(B|A1)+p(A2)\cdot H(B|A2)=1/2\cdot 1+1/2\cdot 1=1$ бит

Следовательно, энтропия сложного опыта, т.е. всей процедуры испытаний:

H(A,B)=H(A)+H(B|A)=2 (бит)

51. Известно, что из **100** изготовленных деталей в среднем **10** деталей имеют дефекты. Для выявления брака используется метод, дающий всегда отрицательный эффект, если деталь изготовлена с браком. Если брак отсутствует, то деталь признается годной лишь в **80%** случаев. Какое количество информации о качестве детали можно получить в среднем по результату такого метода отбраковки?

52. Установленное на предприятии оборудование в результате эксплуатации может оказаться в одном из трех состояний износа:

C1 – оборудование работоспособно, но требует небольшого ремонта;

C2 – большая часть деталей изношена, требуется серьезный ремонт;

C3 – дальнейшая эксплуатация оборудования невозможна.

Предыдущая практика показывает, что вероятность состояния **C1** равна **20%**, **p(C2) = 50%, p(C3) = 30%**. Найти неопределенность (энтропию) состояния оборудования.

Решение. По формуле (2.8) $\mathbf{H(C) = -(0.2 \cdot \log_2 0.2 + 0.5 \cdot \log_2 0.5 + 0.3 \cdot \log_2 0.3) = 1.48}$

53. Для уточнения состояния оборудования из предыдущей задачи на предприятии проведены испытания оборудования. Недостаточная квалификация персонала и отсутствие необходимой контрольно-измерительной аппаратуры привели к тому, что результаты испытаний не достоверно отражают истинное состояние оборудования. В результате испытаний возможны 4 исхода:

Z1 – оборудование исправно;

Z2 – требуется регулировка;

Z3 – требуется замена отдельных деталей;

Z4 – оборудование не пригодно к эксплуатации.

Условные априорные вероятности каждого исхода в зависимости от истинного состояния оборудования сведены в таблицу:

p(Z|C)	Z1	Z2	Z2	Z4
C1	0.5	0.5	0	0
C2	0	0.5	0.5	0
C3	0	0	0.25	0.75

Насколько уменьшилась неопределенность о состоянии оборудования в результате испытаний?

Решение. Необходимо найти общую условную энтропию **C** при условии получения сообщения **Z**: **H(C|Z)**.

Найдем вероятность каждого исхода **Zi**.

p(Z1) = 0.2·0.5+0.5·0+0.3·0 = 0.1

p(Z2) = 0.2·0.5+0.5·0.5+0.3·0 = 0.35

p(Z3) = 0.2·0+0.5·0.5+0.3·0.25 = 0.325

p(Z4) = 0.2·0+0.5·0+0.3·0.75 = 0.225

Найдем апостериорную вероятность состояний **Cj** по формуле Байеса

p(Cj|Zi) = p(Zi|Cj)·p(Cj)/p(Zi). Результаты сведем в таблицу

p(C|Z)	Z1	Z2	Z3	Z4
C1	1	0.29	0	0
C2	0	0.71	0.77	0
C3	0	0	0.23	1

Наконец, вычислим общую условную энтропию **H(C|Z)**. По формуле (2.10)

$$\mathbf{H(C|Z)} = \sum_{i=1}^{4} \mathbf{P(Zi)} \sum_{j=1}^{3} \mathbf{P(Cj|Zi)} \log_2 \mathbf{P(Cj|Zi)} = \mathbf{0+0.35\cdot 0.87+0.325\cdot 0.78+0 = 0.56}$$

Видно, что в результате испытаний неопределенность уменьшилась.

54. Найти энтропию непрерывной системы **X**, все состояния которой на участке от **a** до **b** равновероятны.

Решение. Плотность вероятности состояний системы **X** определяется функцией:

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{b-a}}, \mathbf{a \le x \le b} \\ \mathbf{0}, \quad \mathbf{x < a, x > b} \end{cases}$$

По формуле (2.13) $\mathbf{H} = -\int_{\infty}^{\infty} \mathbf{f(x)} \log_2 \mathbf{f(x)} \mathbf{dx} - \log \Delta \mathbf{x} = -\int_{a}^{b} \frac{\mathbf{1}}{\mathbf{b-a}} \log \frac{\mathbf{1}}{\mathbf{b-a}} \mathbf{dx} - \log \Delta \mathbf{x}$

$$= \log(\mathbf{b-a}) - \log \Delta \mathbf{x} = \log \frac{\mathbf{b-a}}{\Delta \mathbf{x}}$$

Приведенная энтропия **H* = log(b-a)**.

55. Найти энтропию непрерывной системы **X**, вероятности состояний которой подчинены нормальному закону распределения.

56. Устройство управления (УУ) ЭВМ вырабатывает **100** команд, которые могут быть разбиты на 2 группы. **80** команд используются редко и составляют **1%** от общего числа используемых команд. **20** команд используются часто – в **99%** случаев. Определить избыточность, содержащуюся в командах УУ.

Решение. Вероятность появления команды из первой группы – 1/100; вероятность появления конкретной команды из первой группы 1/8000;

$P(S_1) = 1/8000$

аналогично:

$P(S_2) = 1/20 \cdot 99/100 = 99/2000$

Энтропия разных команд:

$H(S) = -n_1 \cdot P(S_1) \cdot \log_2 P(S_1) - n_2 \cdot P(S_2) \cdot \log_2 P(S_2)$ =

$$= -80 \cdot \frac{1}{8000} \log_2 \frac{1}{8000} - 20 \cdot \frac{99}{2000} \approx 4.4$$ (бит на команду)

Оптимальная энтропия $H_0 = \log_2 100 \approx 6,7$ (бит на команду);

избыточность $\varphi = 1 - 4,4/6,7 = 0,34$.

57. В студенческой группе зачет по теории информационных процессов и систем до 1 января не сдали 8 человек, среди которых – один студент из Камбоджи и один студент из Таджикистана. Преподаватель перенес зачет на весенний семестр. Во время повторной сдачи зачета преподаватель спросил у одного из студентов: «Вы сегодня отмечаете Новый год?». Какое количество информации содержит ответ студента? Дополнительные сведения к задаче: в Камбодже Новый год отмечается 3 дня с 14 по 16 февраля; в Таджикистане Новый год (Навруз) отмечается 3 дня с 21 по 23 марта; весенний семестр длится 120 дней.

58. В скачках на ипподроме участвуют 4 лошади: жеребцы Гигабайт (18% побед в скачках), Мегабит (29% побед), Сигнал (31% побед) и молодая кобыла Энтропия. Найти количество информации, содержащееся в сообщении о победителе двух забегов.

59. Какими свойствами обладает логарифмическая мера информации?

60. Из каких соображений получена шенноновская мера информации?

61. Что называется энтропией?

62. Что такое приведенная энтропия?

63. Что такое условная энтропия?

64. Докажите, что для независимых источников сообщений **A** и **B** выполняется **H(B|A)=H(B)**

65. Каковы причины появления избыточности в сообщении?

Глава 3
Информационные процессы и сигналы

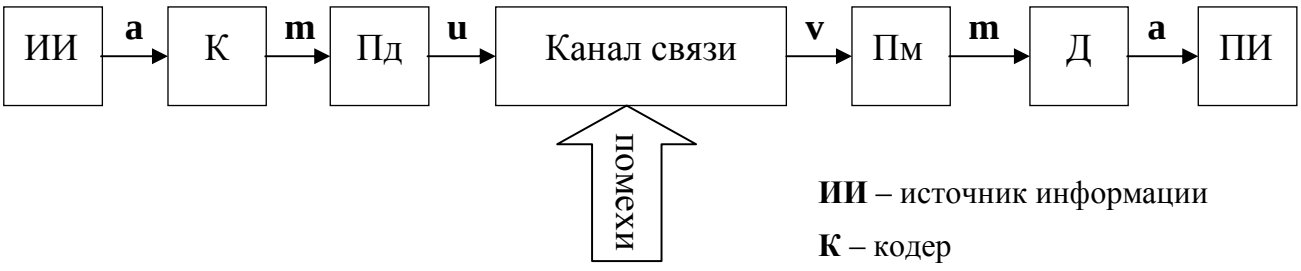
Использование информации для решения каких-либо задач, безусловно, сопряжено с необходимостью ее распространения, то есть осуществления процессов передачи и приема. При этом приходится решать проблему согласования метода кодирования с характеристиками канала связи, а также обеспечить защиту передаваемой информации от возможных искажений.

> Под **информационными процессами** будем понимать процессы сбора, передачи и приема информации посредством специальных каналов связи.

3.1. Общая схема передачи информации в линии связи

Источник информации определяется как объект или субъект, порождающий информацию и представляющий ее в виде сообщения, т.е. последовательности символов. При этом человек в информационном взаимодействии с окружающей средой ограничен возможностями собственных органов чувств. Однако спектр процессов, на основе которых производится передача информации, может быть расширен за счет использования средств связи:

> **Средства связи** – совокупность устройств, обеспечивающих преобразование первичного сообщения от источника информации в сигналы заданной физической природы, их передачу, прием и представление в форме удобной потребителю.



ИИ – источник информации

К – кодер

Пд – передатчик

Пм – приемник

Д – декодер

ПИ – получатель информации

Рис. 3.1. Общая схема передачи информации в каналах связи

Источник информации (**ИИ**) выдает ее в виде первичного сообщения **a**, представленного последовательностью символов первичного алфавита (см. рис. 3.1). Для дальнейшей передачи это сообщение преобразуется (кодируются) кодером (**К**) в сообщения такого характера, которые могут храниться или распространяться в заданном материальном носителе – формируется вторичное сообщение **m**. Для задач, ориентированных на применение вычислительной техники, в качестве вторичного сообщения используется бинарная последовательность из **0** и **1**. Таким образом, под кодером, как правило (но не обязательно), понимается устройство, переводящее исходное сообщение к двоичному алфавиту. В связи с этим в дальнейшем мы будем специально рассматривать двоичное кодирующее устройство и под вторичным алфавитом понимать именно двоичный.

Примерами преобразователей являются: мегафон или телефонный аппарат, преобразующие голосовые сигналы в электрические; радиопередатчик, преобразующие голосовые сигналы в радиоволны; телекамера, преобразующая изображение в последовательность электрических импульсов; модем, переводящий высокочастотные компьютерные сигналы в аналоговые низкочастотные и обратно, и пр. В общем случае при преобразовании выходные сигналы не полностью воспроизводят все особенности первичного сообщения, а лишь его существенные стороны, т.е. при преобразовании часть информации теряется. Например, полоса пропускания частот при телефонной связи от 300 до 3400 Гц, в то время как частоты, воспринимаемые человечески ухом, лежат в интервале 16 – 20000 Гц (т.е. телефонные линии «обрезают» высокие частоты, что приводит к искажениям звука); в черно-белом телевидении при преобразовании теряется цвет изображения. Именно в связи с возможными потерями встает задача выработки таких способов представления и последующего преобразования первичного сообщения, которые обеспечивали бы возможно более полную сохранность исходной информации и, одновременно, согласование со скоростью передачи информации по данной линии связи.

Непосредственная передача осуществляется передатчиком вторичного сообщения (**Пд**). Он инициирует некоторый нестационарный процесс **x**, подготавливает в соответствии со вторичным сообщением материальный носитель, который может быть физически перемещен в другую точку пространства посредством **канала связи**.

Канал связи – это материальная среда, а также физический или иной процесс, посредством которого осуществляется перемещение сигнала, т.е. его распространение в пространстве с течением времени.

Ниже приведены примеры некоторых каналов связи.

Канал связи	Среда	Носитель сообщения	Процесс, используемый для передачи сообщения
1	2	3	4
Почта, курьеры	Среда обитания человека	Бумага	Механическое перемещение носителя
Телефон, компьютерные сети	Проводник	Электрический ток	Перемещение электрических зарядов
Радио, телевидение	Электромагнитное поле	Электромагнитные волны	Распространение электромагнитных волн

Окончание

1	2	3	4
Зрение	Электромагнитное поле	Световые волны	Распространение световых волн
Слух	Воздух	Звуковые волны	Распространение звуковых волн
Обоняние, вкус	Воздух, пища	Химические вещества	Химические реакции
Осязание	Поверхность кожи	Объект, воздействующий на органы осязания	Теплопередача, давление

Любой реальный канал связи подвержен внешним воздействиям, а также в нем могут происходить внутренние процессы, в результате которых искажаются передаваемые сигналы и, следовательно, связанная с ними информация. Такие воздействия называются шумами (помехами).

Источники помех могут быть внешними, например, так называемые «наводки» от мощных потребителей электричества или атмосферных явлений, приводящие к появлению нарушений в радиосвязи; одновременное действие нескольких близко расположенных однотипных источников (одновременный разговор нескольких человек). К помехам могут приводить и внутренние особенности данного канала, например, физические неоднородности носителя; паразитные явления в шинах; процессы затухания сигнала в линии связи из-за большой удаленности, эффект «гонок» в вычислительных системах. Если уровень помех оказывается соизмерим с интенсивностью несущего сигнала, то передача информации по данному каналу оказывается существенно затрудненной. Однако и при относительно низких уровнях шумов они могут вызывать искажения передаваемых сигналов и, следовательно, частичную потерю связанной с ними информации. Существуют и применяются методы защиты от помех, например, экранирование электрических линий связей; улучшение избирательности приемного устройства и т.д. Другим способом защиты от помех является использование специальных методов кодирования информации, о чем речь пойдет ниже.

После прохождения сигнала по каналу связи он попадает в приемное устройство (**Пм**), где одновременно преобразуется в форму, необходимую для дальнейшей интерпретации. Если перед передачей применялось кодирование, после приема сигнал **v** направляется в декодер (**Д**) и лишь затем – к получателю (потребителю) информации (**ПИ**). При этом декодер может быть совмещен с преобразователем (например, телеграфный аппарат или компьютер) или с приемником информации (радист, принимающий сигналы азбуки Морзе и интерпретирующий их).

3.2. Формирование сигналов в канале связи

Для передачи сообщения сигнал должен изменять свои физические параметры в соответствии с передаваемым сообщением (говорят, что сигнал модулируется сообщением). Параметры сигнала обязательно должны меняться, причем случайным образом. Сигнал с постоянными характеристиками не может передавать информацию. Например, звук одной частоты и громкости не несет информации. Сигналы могут иметь непрерывный (синоним – аналоговый) или дискретный (синоним – цифровой) характер. Непрерывные

сигналы могут изменять параметры в любой момент времени, а дискретные – только в определенные моменты времени (см. рис. 2.1). Обычно дискретные моменты времени изменения параметров сигнала отстоят друг от друга на фиксированный интервал, причем заранее известный как передатчику, так и приемнику сигнала.

Модуляция сигналов

Для того, чтобы сигнал передавал информационное сообщение, в технических системах передачи информации передатчик содержит в своей структуре генератор, который производит периодический сигнал с фиксированными параметрами, называемый несущим сигналом. А другая функциональная подсистема передатчика – **модулятор** – воспринимает информационное сообщение от кодера и изменяет параметры несущего сигнала в соответствии с информационным сообщением. В зависимости от того, какой характер (непрерывный или дискретный) имеют несущий и информационный сигнал различают различные виды модуляции: аналоговая модуляция, цифровая модуляции (иначе – манипуляция), импульсная модуляция и некоторые другие. Обратный процесс – выделение информационного сигнала из модулированного – называется демодуляций, а соответствующее устройство – **демодулятором**.

Аналоговая модуляция сигналов

При аналоговой модуляции несущий сигнал представляет собой непрерывные гармонические колебания высокой частоты. Сообщение также представляется в виде непрерывной функции времени.

Любой гармонический сигнал описывается формулой $\mathbf{x(t)=A\sin(\omega t+\varphi).}$ Параметрами такого сигнала являются амплитуда **A**, круговая частота **ω** и сдвиг фазы **φ**. Изменяя эти параметры, можно превратить гармонический сигнал в носитель информации (модулировать сигнал информацией). Возможны три основных типа аналоговой модуляции.

Амплитудная модуляция (рис. 3.2) предполагает изменение амплитуды сигнала во времени: $\mathbf{x_{AM}(t)=A(t)\cdot\sin(\omega t+\varphi),}$ причем закон изменения амплитуды определяется информационным сообщением и часто является также гармоническим: $\mathbf{A(t)=A_0 + Aм\cdot\sin(\Omega t+\Phi),}$

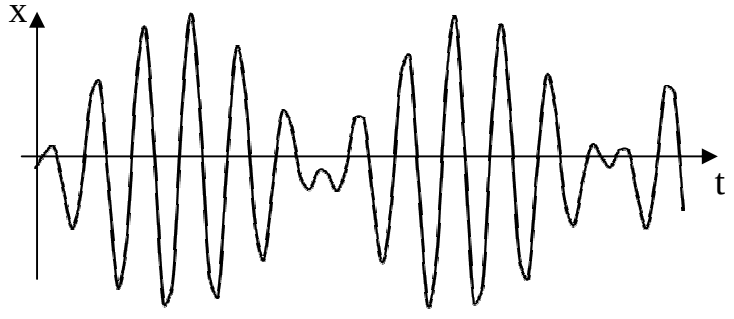


Рис.3.2. Пример амплитудной модуляции непрерывного сигнала

Частотная модуляция (рис. 3.3) – это вид аналоговой модуляции, при котором информационный сигнал управляет частотой несущего колебания. При этом амплитуда несущего сигнала, как правило, остаётся постоянной. Вид частотно-модулированного сигнала можно представить функцией времени: $\mathbf{x_{ЧМ}(t)=A\cdot\sin(\omega(t)t+\varphi),}$ причем частота сама тоже нередко изменяется по гармоническому закону: $\mathbf{\omega(t) = \omega_0 + \omega_0 \cos \Omega t,}$ где $\mathbf{\omega_0}$ называется частотной девиацией.

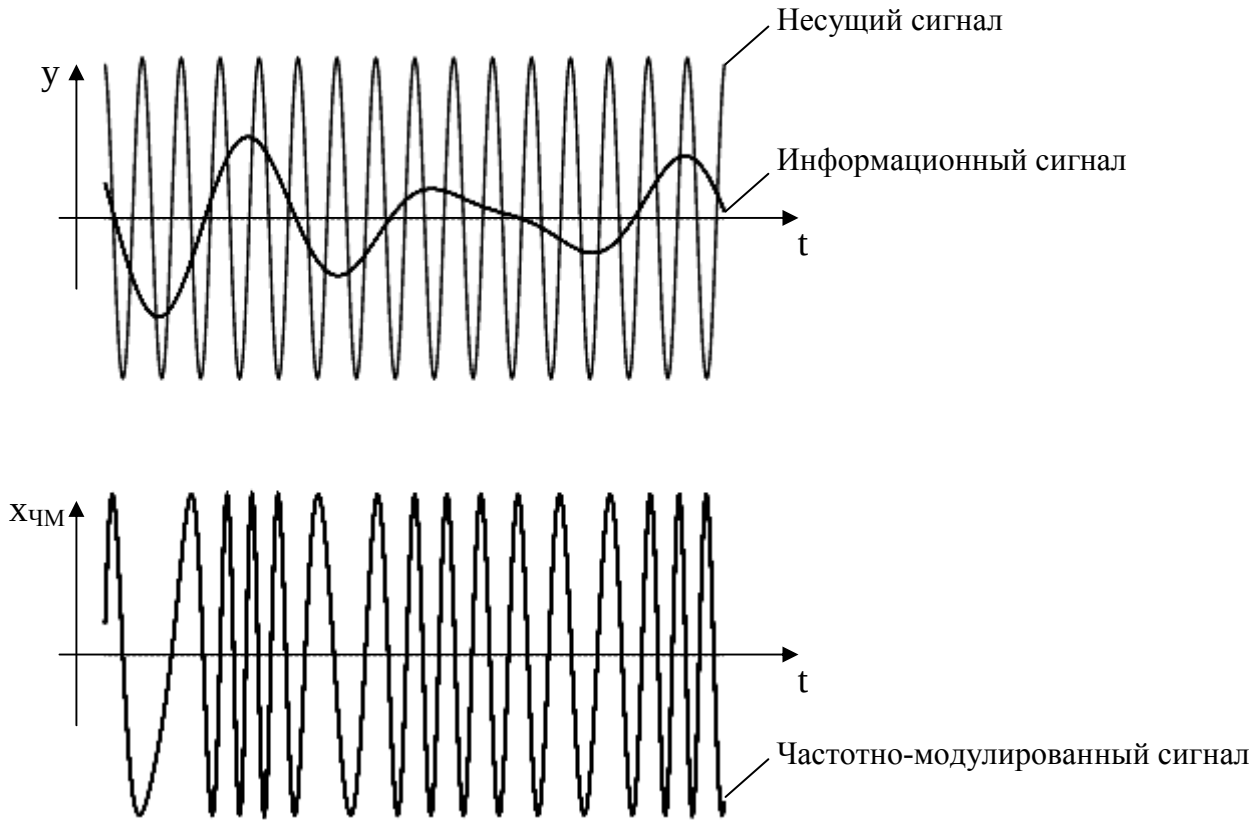


Рис.3.3. Пример частотной модуляции непрерывного сигнала

Фазовая модуляция предполагает, что модулирующим параметром сигнала является фазовый сдвиг: $\mathbf{x_{ФМ}(t)=A\cdot\sin(\omega t+\varphi(t)).}$

Можно показать, что $\mathbf{\omega(t)t+\varphi = \omega_0 t + \theta(t) + \theta_0}$, таким образом, частотная и фазовая модуляция – это два варианта технической реализации одного вида модуляции, называемого **угловой модуляцией**.

Цифровая модуляция сигналов (манипуляция)

Цифровая модуляция сигналов или манипуляция наблюдается, когда высокочастотный несущий сигнал скачкообразно меняет свои параметры под

воздействием дискретного информационного сообщения. Параметры, которые могут модулироваться информационным сообщением, те же самые – амплитуда, частота и фаза. Таким образом можно построить различные виды цифровой манипуляции.

При **амплитудной манипуляции** (рис. 3.4) в процессе формирования сигнала изменяется амплитуда несущего сигнала: $\mathbf{x_{AM}(t)=A(t)\cdot\sin(\omega t+\varphi),}$ где **A(t)** – последовательность дискретных чисел, соответствующих символам вторичного алфавита, в бинарном случае – **0** или **1**.

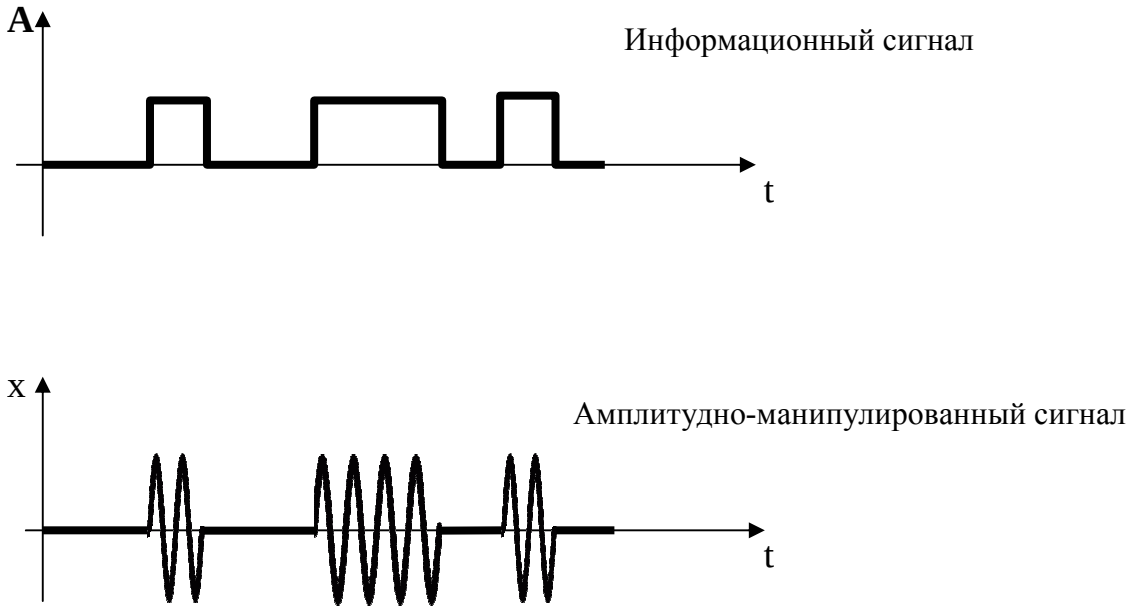


Рис.3.4. Пример амплитудной манипуляции при бинарном информационном сообщении

Частотная манипуляция (рис. 3.5) предполагает, что изменяется частота несущего сигнала при неизменной амплитуде. При получении бинарного информационного сообщения модулятор генерирует результирующий сигнал, характеризующийся двумя существенно различными значениями частоты. Высокая частота, например, соответствует единичному символу сообщения, а низкая частота – нулевому символу. $\mathbf{x_{ЧМ}(t)=A_0(t)\cdot\sin(k(t)\omega_0 t+\varphi),}$ где **k(t) –** индекс модуляции, принимающий значения **1**, если в момент **t** получен единичный символ информационного сообщения и фиксированное значение меньше **1** (например, **0.5**), если получен нулевой символ информационного сообщения. Частотная манипуляция обладает высокой помехоустойчивостью, так как шумы искажают, как правило, амплитуду, а не частоту сигнала.

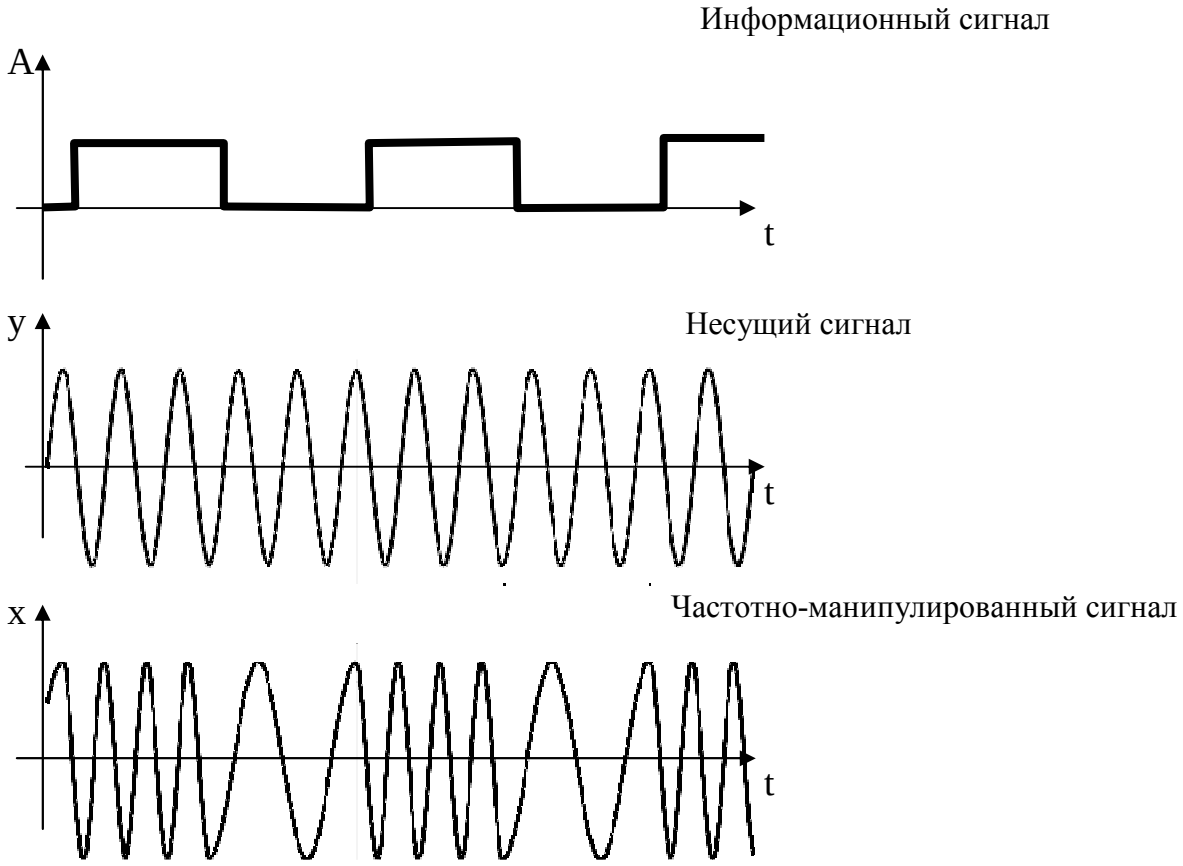


Рис.3.5. Пример частотной манипуляции при бинарном информационном сообщении

При **фазовой манипуляции** скачкообразно изменяется фаза несущего колебания в зависимости от поступившего информационного разряда. В бинарном случае изменение разряда информационного сообщения сдвигает фазу несущего сигнала на $\boldsymbol{\pi}$ по сравнению с предыдущим значением, то есть знак сигнала меняется на противоположный (рис. 3.6).

Имеются и некоторые другие способы цифровой модуляции сигналов: квадратурная манипуляция (одновременное изменение амплитуды и фазы сигнала), гауссовская манипуляция, мультиплексирование и т.д.

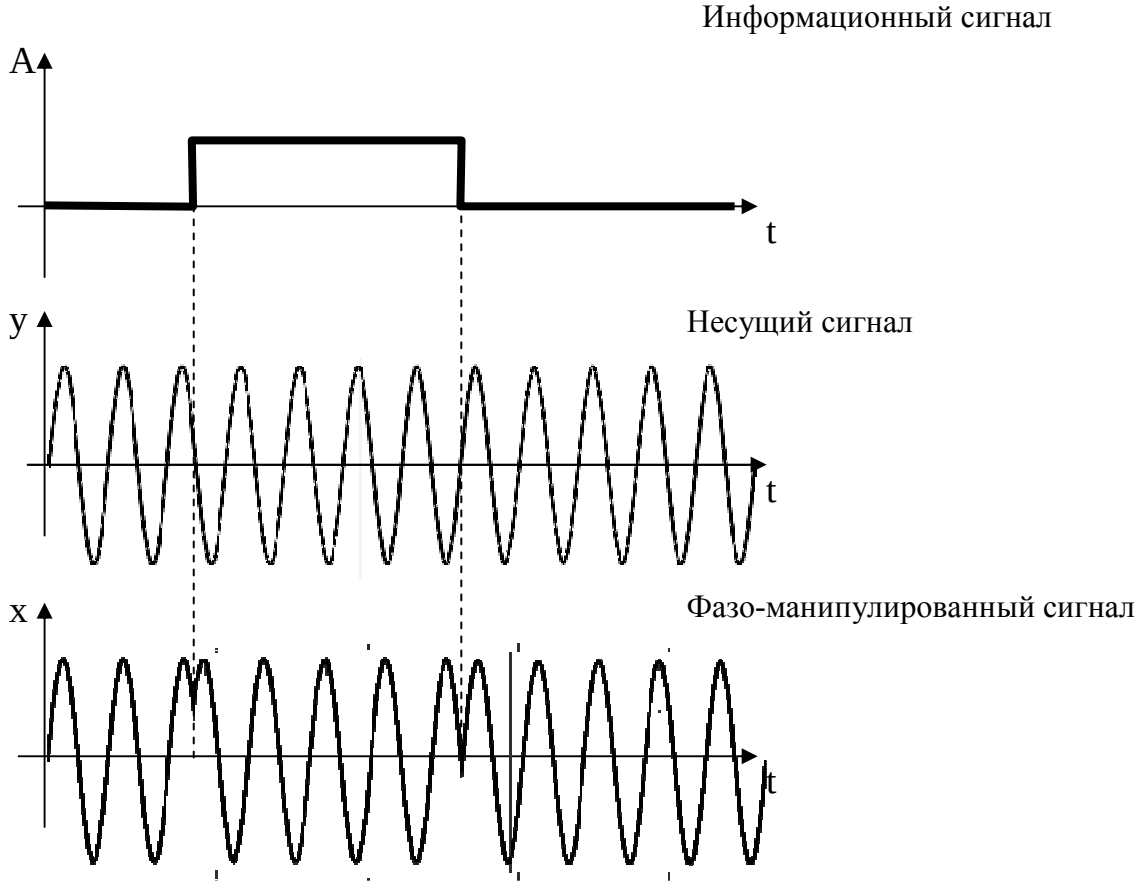


Рис.3.6. Пример фазовой манипуляции при бинарном информационном сообщении

Импульсная модуляция сигналов

При импульсной модуляции несущий сигнал представляет собой последовательность импульсов. Амплитуда, форма и частота импульсов могут быть различными. Чаще всего используются импульсы прямоугольной, треугольной и пилообразной формы. При использовании импульсов прямоугольной формы могут меняться такие параметры несущего сигнала, как амплитуда, частота (или обратная ей величина – длительность импульса), фаза, ширина импульса, скважность (отношение длительности импульса к его ширине) (рис. 3.7). Соответственно, существуют различные виды модуляции импульсного сигнала информационным сообщением, основанные на изменении тех или иных параметров несущего сигнала, а так же их комбинаций.

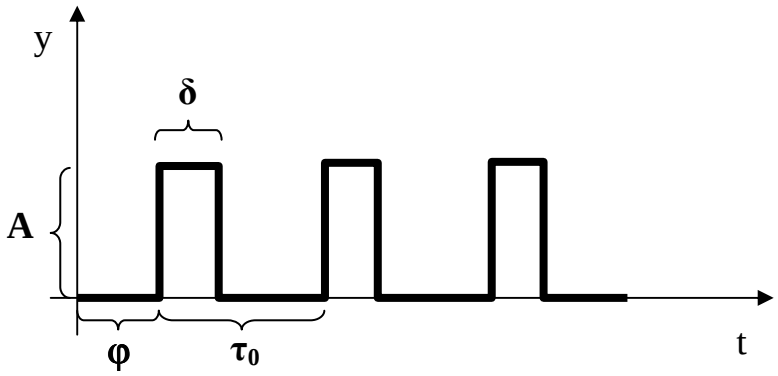


Рис.3.7. Основные параметры импульсного сигнала прямоугольной формы:
A – амплитуда; **τ_0** – длительность; **φ** – фаза; **δ** – ширина

Модулятор, принимая информационное бинарное сообщение, модифицирует несущий сигнал, скачкообразно изменяя параметры импульсов: амплитуду (**амплитудно-импульсная модуляция**), длительность (**частотно-импульсная модуляция**), ширину (**широтно-импульсная модуляция**), фазу (**фазово-импульсная модуляция**), скважность (**скважностно-импульсная модуляция**) или их комбинации (рис. 3.8).

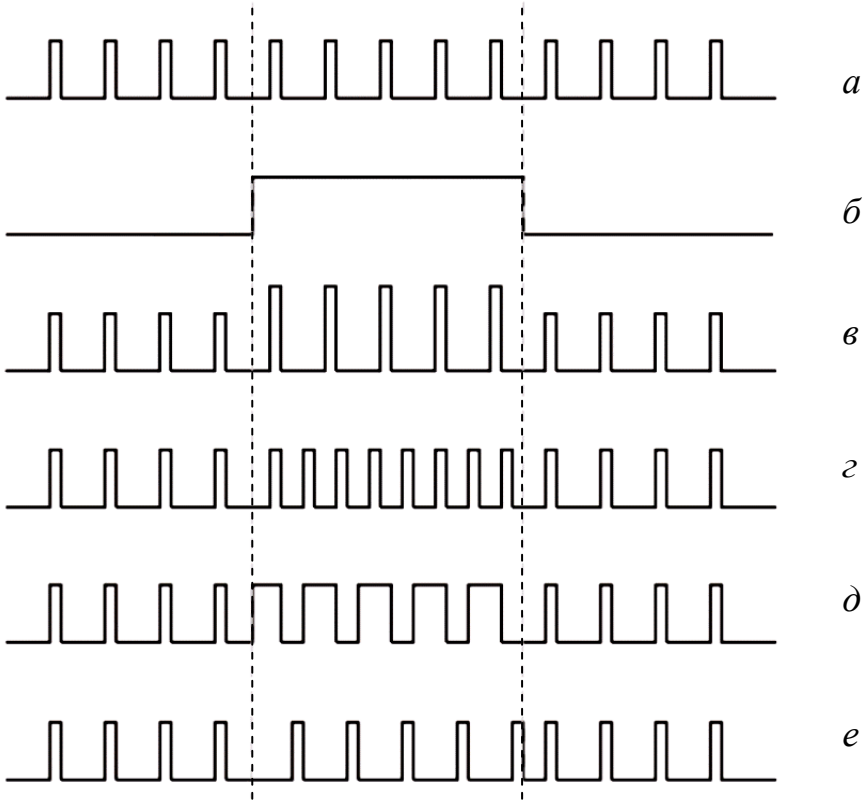


Рис.3.8. Примеры импульсной модуляции:

а – несущий сигнал; *б* – информационный сигнал; *в* – амплитудно-импульсная модуляция; *г* – частотно-импульсная модуляция; *д* – широтно-импульсная модуляция; *е* – фазово-импульсная модуляция

Дискретизация сигнала по уровню (квантование)

Непрерывный сигнал в целях его обработки и/или обеспечения помехоустойчивости можно представлять его дискретными значениями. Сигнал будет считаться квантованным, если амплитуда сигнала может принимать определенные значения из дискретного множества (рис. 3.9). Уравнение квантованного по уровню сигнала можно представить так:

$$\mathbf{x(t)} = \sum_{\mathbf{k}} \mathbf{A_k} \cdot \mathbf{1(t - t_k)},$$

где **1(·)** – это единичная функция (единичный скачок):

$$\mathbf{1}(\tau) = \begin{cases} \mathbf{0, \text{если } } \tau \le \mathbf{0} \\ \mathbf{1, \text{если } } \tau > \mathbf{0} \end{cases},$$

а $\mathbf{t_k}$ – это моменты времени, когда скачкообразно изменяется амплитуда сигнала.

Уровень амплитуды $\mathbf{A_k}$ может принимать значения из некоторого дискретного множества.

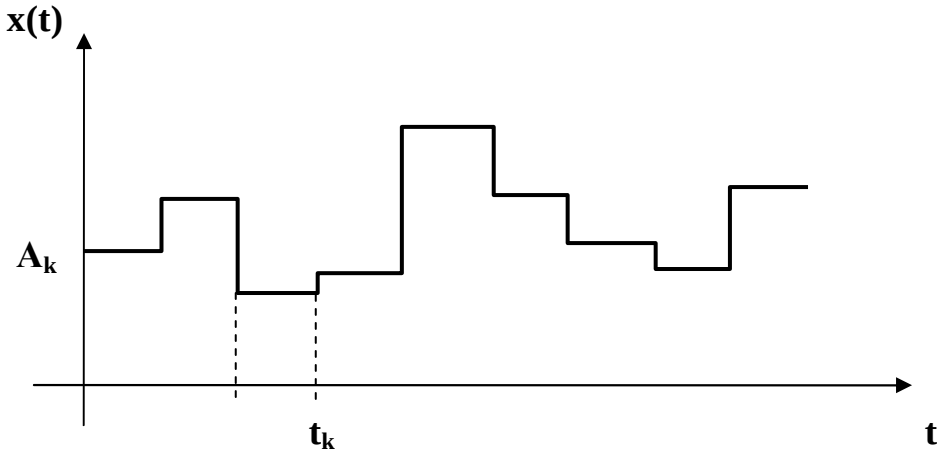


Рис. 3.9. Дискретный сигнал, квантованный по уровню

Если известны вероятностные характеристики непрерывного сигнала, например плотность вероятности амплитуды **f(x)**, то вероятность того, что при квантовании будет получен уровень сигнала $\mathbf{A_k}$ может быть определена так:

$$\mathbf{p}(\mathbf{A_k}) = \int_{\mathbf{A_{k-1}}}^{\mathbf{A_k}} \mathbf{f(x)dx}.$$

Дискретизация сигнала по времени

Сигнал считается дискретизированным по времени, если моменты времени, когда сигнал изменяется, могут принимать значения из некоторого дискретного множества, как правило, отстоящие друг от друга на фиксированный интервал. Дискретизация по времени часто называется

собственно дискретизацией, а интервал между соседними моментами времени – интервалом дискретизации (рис. 3.10). Аналитически дискретизированный сигнал можно записать следующим образом:

$$\mathbf{x(t) = \sum_k a(t_k) \cdot \delta(t - t_k)},$$

где $\mathbf{t_k = k \cdot \Delta t}$, а $\mathbf{\Delta t}$ в свою очередь – это интервал дискретизации непрерывного сигнала, $\mathbf{a(t_k)}$ – амплитуда непрерывного сигнала **x** в моменты времени $\mathbf{t_k}$, $\mathbf{\delta(\cdot)}$ – ограниченная дельта-функция:

$$\mathbf{\delta(\tau) = \begin{cases} 0, \text{ если } \tau \neq 0 \\ 1, \text{ если } \tau = 0 \end{cases}}.$$

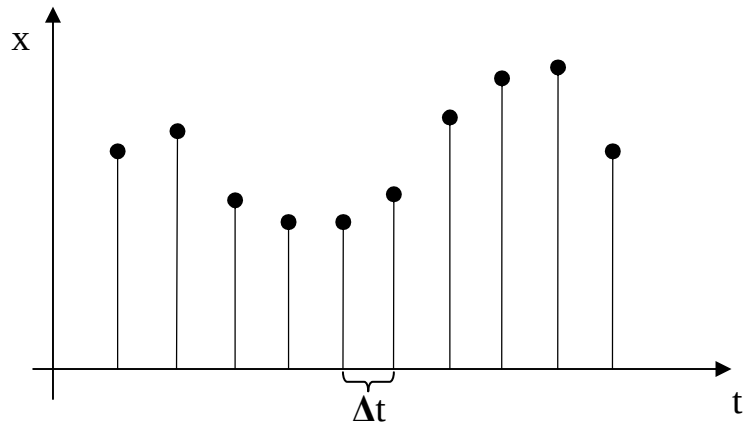


Рис. 3.10. Дискретизированный сигнал, не квантованный по уровню

Дискретный сигнал менее подвержен помехам, чем непрерывный. Очевидно, что в целях помехоустойчивости непрерывные информационные сигналы имеет смысл заменять дискретными. Но не будет ли при этом теряться информация?

Теорема В.А. Котельникова

Котельников Владимир Александрович (1908-2005) – советский инженер-радиотехник, вице-президент Академии наук СССР. В 1930-е годы руководил научно-технической службой, обеспечивавшей секретность телефонной связи Советского правительства. В 1933 г. вышла статья Котельникова «Пропускная способность эфира и проволоки», где он сформулировал нижеприведенную теорему. Спустя 15 лет, не зная о работе Котельникова, аналогичный результат получил К.Шеннон, в связи с чем ныне теорему часто называют по имени обоих ученых, иногда добавляя к ним фамилию Найквиста. В 1947 году В.А. Котельниковым была разработана фундаментальная теория потенциальной помехоустойчивости. Эта теория дала инженерам инструмент для синтеза оптимальных устройств обработки принимаемых сигналов на фоне шумов и помех, на ее основе были разработаны методы оценки качества приема аналоговых и цифровых сигналов в различных каналах связи.

В.А.Котельников доказал следующую теорему, определяющую возможность дискретизации непрерывных сигналов.

> Сигнал **x(t)**, имеющий ограниченный спектр частот в диапазоне от **0** до $\omega_{гр}$ может быть передан с любой степенью точности при помощи своих дискретных значений, следующих с интервалом $\Delta t < \frac{\pi}{\omega_{гр}}$

Любую функцию, удовлетворяющую условию Дирихле (конечное число минимумов, максимумов и точек разрыва на любом конечном отрезке), можно представить рядом Фурье:

$$f(t) = C + \sum_{k=0}^{\infty}(a_k \cos k\omega_0 t + b_k \sin k\omega_0 t).$$

Таким образом, максимальная величина $k\omega_0$, для которой $a_к \neq 0$ и $b_к \neq 0$, является верхней границей спектра частот. Хотя теоретически в непрерывной функции могут содержаться гармоники неограниченной частоты, на практике мы можем ограничить спектр частот разумными пределами. Например, слуховой анализатор человека не воспринимает звуки с частотой выше 20 кГц, в стандартном телефонном канале верхней границей частоты передаваемого электрического сигнала принято считать 3400 Гц.

Котельников доказал, что любой процесс, содержащий гармонические колебания с частотами не выше $\omega_{гр}$, с любой степенью точности представим в виде суммы его дискретных значений, умноженных на функцию отсчета.

$$x(t) = \sum_{k=0}^{\infty} x(k\Delta t)\frac{\sin \omega_{гр}(t - k\Delta t)}{\omega_{гр}(t - k\Delta t)} \qquad (3.1)$$

Таким образом, установлено условие, при котором дискретизация непрерывных сигналов не ведет к потере информации. Именно теорема Котельникова определяет фундаментальную возможность импульсной модуляции сигнала.

3.3. Передача информации по каналу связи без учета помех

Пропускная способность дискретного канала связи без помех

Пусть **M** – мощность первичного алфавита, которым оперирует источник информации, а **m** – мощность вторичного алфавита, используемого при передаче сообщения по каналу связи. В этом случае пропускная способность канала связи определяется формулой:

$$C = \frac{I_{max}}{T} = \frac{n \log_2 m}{T} = \frac{n H_{max}}{T}$$

Введем величину **V** – частота снятия отсчетов (т.е. сколько элементарных сигналов пройдет в единицу времени). Физически эта величина определяется частотой тактового генератора канала связи.

$V=n/T=1/\tau$, где **τ** – время передачи одного символа первичного алфавита.

Тогда $\mathbf{C=V\cdot H_{max} = V\cdot \log_2 m.}$ (3.2)

Отсюда $\mathbf{C = \frac{\log_2 m}{\tau}}$ (бит/с). (3.3)

Для двоичных сигналов **m = 2**, а $\tau = \tau_0$ следовательно $\mathbf{C = \frac{1}{\tau_0} = V}$

Когда речь идет о дискретизации непрерывных сигналов, стараются, чтобы длительность элементарного импульса τ_0 определялась в соответствии с теоремой Котельникова, т.е. $\tau_0 \le \Delta t$. Тогда $\mathbf{C = \frac{1}{\Delta t} = \frac{\varpi_{гр}}{\pi} = \frac{2\pi F}{\pi} = 2F}$. Величину **F** называют **частотой манипуляции** или **полосой пропускания**, а выражение **C = 2F** – **пределом Найквиста**. Величина **C** является характеристикой канала связи, определяется его конструктивными особенностями.

Гарри Найквист (1989-1976) – шведский ученый, радиотехник, работал инженером-исследователем в Bell Laboratories, где занимался разработками в области телеграфии, факсимильной передачи, телевидения и другими телекоммуникационными проблемами. Ранние работы Найквиста по определению ширины частотного диапазона, требуемого для передачи информации, опубликованные в статье «Определённые проблемы теории телеграфной передачи» (1928), заложили основы для последующих успехов Клода Шеннона в разработке теории информации.

Скорость передачи информации по дискретному каналу без помех

Скорость передачи информации – это количество информации, передаваемое по каналу в единицу времени. Если тактовый генератор канала выдает **V** элементарных импульсов в единицу времени, каждый из которых интерпретируется как один бинарный разряд, а средняя длина кода одного символа первичного алфавита составляет **K** разрядов бинарного алфавита, то, очевидно, отношение **V/K** будет выражать число символов первичного алфавита, транслируемых по каналу за единицу времени. Если с каждым из символов первичного алфавита связано среднее количество информации **H** (энтропия источника сообщений), то можно найти общее количество информации, передаваемой по каналу связи за единицу времени – эта величина называется скоростью передачи или (будем обозначать ее **J**).

$$\mathbf{J} = \frac{\mathbf{V}}{\mathbf{K}}\mathbf{H} = \frac{\mathbf{H}}{\tau_0 \cdot \mathbf{K}},$$

где **H** – энтропия источника информации, определяемая известной формулой (2.8): $\mathbf{H} = -\sum_{i=1}^{M} \mathbf{p_i} \log_2 \mathbf{p_i}$.

Эту же формулу можно получить из других соображений. Пусть один символ первичного алфавита передается за время **τ** и несет количество информации, равное **H**. Тогда скорость передачи информации **J** определяется отношением $\frac{\mathbf{H}}{\tau}$. С учетом того, что каждый символ первичного алфавита кодируется **K** символами вторичного (бинарного) алфавита, а время передачи одного бинарного символа равно τ_0, имеем $\tau = \tau_0 \cdot \mathbf{K}$. Отсюда $\mathbf{J} = \frac{\mathbf{H}}{\tau_0 \cdot \mathbf{K}}$.

Размерностью скорости **J**, как и пропускной способности **C**, является бит/с. Каково же соотношение этих характеристик? Выразим **V** из (3.2) и получим:

$$\mathbf{J} = \frac{\mathbf{V} \cdot \mathbf{H}}{\mathbf{K}} = \frac{\mathbf{C} \cdot \mathbf{H}}{\mathbf{K} \cdot \log_2 \mathbf{m}}. \qquad (3.4)$$

Согласно теории информации Шеннона при любом способе кодирования длина кода подчинена соотношению $\mathbf{K} \geq \frac{\mathbf{H}}{\log_2 \mathbf{m}}$, хотя может быть сколь угодно близкой к этому значению. Следовательно, всегда $\mathbf{J} \leq \mathbf{C}$, т.е. скорость передачи информации по каналу связи не может превысить его пропускной способности.

Пример 3.1. Первичный алфавит состоит из трех знаков с вероятностями $\mathbf{p_1 = 0{,}2}$; $\mathbf{p_2 = 0{,}7}$; $\mathbf{p_3 = 0{,}1}$. Для передачи по каналу без помех используются равномерный двоичный код. Частота тактового генератора **500 Гц**. Какова пропускная способность канала и скорость передачи?

Решение. Поскольку код двоичный, **m = 2**; **C =V = 500 бит/с**.

Энтропия источника: $\mathbf{H = -\,0{,}2 \cdot \log_2 0{,}2 - 0{,}7 \cdot \log_2 0{,}7 - 0{,}1 \cdot \log_2 0{,}1 = 1{,}16}$ бит/символ.

Поскольку код равномерный $\mathbf{K \geq H/\log_2 2 = 2}$ (т.е. для кодирования каждого знака первичного алфавита используется 2 элементарных разряда).

Следовательно, в силу (3.4), получаем:

$$\mathbf{J} = \frac{\mathbf{C} \cdot \mathbf{H}}{\mathbf{K} \log_2 \mathbf{m}} = \frac{\mathbf{500 \cdot 1.16}}{\mathbf{2}} = \mathbf{290}$$ (бит в секунду)

Видно, что реальная скорость передачи информации меньше пропускной способности. Так получается вследствие того, что каждый символ первичного алфавита, занимая два разряда, несет информации меньше двух бит. Если приблизить длину кода **K** к значению реальной энтропии, можно увеличить скорость передачи информации.

Эффективное статистическое кодирование сообщений. Теорема Шеннона для каналов без помех

Для дискретных каналов без помех К. Шенноном была доказана следующая теорема (**первая теорема Шеннона**):

> Если производительность источника $\mathbf{R \le C - \varepsilon}$, где $\boldsymbol{\varepsilon}$ – сколь угодно малая величина, то всегда существует способ кодирования, позволяющий передавать по каналу все сообщения источника. Передачу всех сообщений при $\mathbf{R > C}$ осуществить невозможно.

Как бы ни была велика избыточность источника, все его сообщения могут быть переданы по каналу, если пропускная способность информационного канала хоть немного больше производительности источника: $\mathbf{C > R}$.

Для рационального использования пропускной способности канала необходимо применять соответствующие способы кодирования.

> **Эффективным статистическим кодированием** называется кодирование, при котором статистические характеристики источника информации согласуются с характеристиками канала связи.

При эффективном кодировании фактическая скорость передачи информации приближается к пропускной способности канала.

Рассмотрим основные принципы оптимального кодирования для двоичного канала без помех. Пусть источник оперирует алфавитом символов $\mathbf{a_i}$, $\mathbf{i=1…M}$. Вероятность каждого символа $\mathbf{P(a_i)}$. Кодер преобразует символ $\mathbf{a_i}$ в двоичную последовательность длиной $\mathbf{n_i}$. Средняя длительность передачи кодовой комбинации одного символа первичного алфавита вычисляется так:

$\boldsymbol{\tau = \tau_0 \sum_{i=1}^{M} n_i P(a_i)}$, где $\boldsymbol{\tau_0}$ – длительность передачи одного двоичного разряда кода.

При этом средняя длина кода определяется как $\mathbf{K = \sum_{i=1}^{M} n_i P(a_i)}$. (3.5)

Соответственно, тогда $\boldsymbol{\tau = K \cdot \tau_0}$.

Величина **V/K** определена ранее как среднее число знаков первичного алфавита, транслируемых по каналу в единицу времени. Соответственно, величина **K/V** – это средняя длительность трансляции одного знака первичного алфавита, т.е. $\boldsymbol{\tau = K/V}$.

Значит, в соответствии с формулой (3.4) скорость передачи в канале $\mathbf{J = \frac{H}{\tau}}$. Подставляя выражения для средней длительности и энтропии, получим:

$$\mathbf{J = \frac{-\sum_{i=1}^{M} P(a_i)\log P(a_i)}{\tau_0 \sum_{i=1}^{M} n_i P(a_i)}}. \qquad (3.6)$$

В выражении (3.6) значение числителя определяется исключительно статистическими свойствами источника, а $\boldsymbol{\tau_0}$ – свойствами канала связи. Возникает вопрос: можно ли так закодировать сообщение, чтобы скорость передачи достигала своего максимального значения? Максимальная скорость передачи определяется пропускной способностью канала связи. Для двоичного канала: $\mathbf{C = \frac{1}{\tau_0}}$. Если в выражении (3.6) положить $\mathbf{n_i = - \log_2 P(a_i)}$, то **J** станет равным $\mathbf{1/\tau_0}$, то есть **C**. Применение неравномерного кодирования (например, кода Шеннона-Фано) может приблизить длину кода $\mathbf{n_i}$ к величине $\mathbf{- \log_2 P(a_i)}$.

Пример 3.2. Первичный алфавит состоит из трех знаков **A**, **B**, **C** с вероятностями $\mathbf{p_A = 0{,}2}$; $\mathbf{p_B = 0{,}7}$; $\mathbf{p_C = 0{,}1}$. Для передачи по каналу без помех используются код Шеннона-Фано. Частота тактового генератора **500** Гц. Какова пропускная способность канала и скорость передачи?

Решение. Поскольку код Шеннона-Фано – двоичный, то **m = 2**; **C =V = 500** бит/с.

Энтропия источника: $\mathbf{H = - 0{,}2 \cdot \log_2 0{,}2 - 0{,}7 \cdot \log_2 0{,}7 - 0{,}1 \cdot \log_2 0{,}1 = 1{,}16}$ бит

Длительность одного бинарного разряда в канале $\boldsymbol{\tau_0 = 1/V = 0.002}$ с.

Закодируем первичный алфавит кодом Шеннона-Фано: **A→10**, **B→0**, **C→11**, длины кодов будут равны: $\mathbf{n_A = 2}$; $\mathbf{n_B = 1}$; $\mathbf{n_C = 2}$

Средняя длина кода $\mathbf{K = 0.2 \cdot 2 + 0.7 \cdot 1 + 0.1 \cdot 2 = 1.3}$

Следовательно, получаем:

$$\mathbf{J = \frac{H}{\tau_0 (n_A p_A + n_B p_B + n_C p_C)} = \frac{1.16}{0.002 \cdot (2 \cdot 0.2 + 1 \cdot 0.7 + 2 \cdot 0.1)} = 446} \text{ (бит/с).}$$

По сравнению с равномерным двоичным кодом (см. пример 3.1) скорость передачи возросла на 54% и приблизилась к пропускной способности.

Пример 3.3. Можно ли с помощью кодирования еще больше увеличить скорость передачи?

Решение. Первичный алфавит из примера 3.2 будем кодировать по парам символов (это так называемое **укрупнение кодов**). Пары

символов, их вероятности, коды Шеннона-Фано и длины кодовых последовательностей приведены в таблице:

Символ	Вероятность	Код	длина
BB	0.7·0.7=0.49	0	1
AB	0.2·0.7=0.14	100	3
BA	0.14	101	3
BC	0.07	1100	4
CB	0.07	1101	4
AA	0.04	1110	4
CA	0.02	11110	5
AC	0.02	111110	6
CC	0.01	111111	6

Средняя длина кодового слова для пары (см. формулу (3.5)) равна **2.42**, следовательно, для одного символа **K**= **1.21**.

Скорость передачи $\mathbf{J} = \frac{1.16}{0.002 \cdot 1.21} \approx 479$ (бит/с).

Скорость передачи еще больше приблизилась к своему пределу – пропускной способности канала.

Эффективность кода определяется соотношениями средней длины кода **K**, энтропии источника **H** и оптимальной энтропии $\mathbf{H_O}$. Коэффициент общей эффективности кода показывает, насколько выбранный код соответствует статистическим характеристикам источника $\mathbf{K_{оэ}} = \frac{\mathbf{H}}{\mathbf{K}}$. Коэффициент статического сжатия показывает соответствие кода идеальному (оптимальному) источнику $\mathbf{K_{сс}} = \frac{\mathbf{K}}{\mathbf{H_o}}$.

Теоремы побуквенного неравномерного двоичного кодирования

Прямая теорема

> Для алфавита **X ={x, p(x)}** с энтропией **H** существует побуквенный неравномерный префиксный двоичный код со средней длиной кодовых слов $\mathbf{K \le H + 1}$.

Обратная теорема

> Для любого однозначно декодируемого двоичного кода алфавита **X={x, p(x)}** с энтропией **H** средняя длина кодовых слов **K** удовлетворяет неравенству **K ≥ H**.

С помощью теорем побуквенного кодирования можно дать оценку возможной средней длины неравномерного кода.

$$\mathbf{H \le K \le H+1} \qquad (3.7)$$

Первая теорема и, соответственно, правая часть неравенства (3.7) гарантирует, что при любых самых неблагоприятных статистических характеристиках источника сообщений можно построить неравномерный код средней длины не больше чем **H+1**. Вторая теорема и левая часть указанного неравенства говорит о том, что даже при самых «удачных» вероятностях символов первичного алфавита нельзя построить код средней длиной меньше **H**. «Удачными» вероятностями в данном контексте будут такие, для которых двоичный логарифм является целым числом, то есть совпадающими с отрицательными степенями числа **2**. Из этих же теорем вытекает оценка общей эффективности кода: $\mathbf{K_{ОЭ} \le 1}$.

3.4. Передача информации по каналу с помехами

До сих пор мы предполагали, что информация, поступившая от кодера/передатчика в канал связи в точности соответствует информации, принятой приемником/декодером из канала. Наличие помех в канале связи приводит к тому, что часть информации при перемещении по каналу теряется, искажается, зашумляется. Информация, принятая приемником, не полностью снимает неопределенность относительно переданной источником, хотя и уменьшает ее. Если на вход канала связи поступил сигнал **u**, а с выхода канала принят сигнал **v**, то говорят о **взаимной информации I(u,v)**.

Этот термин используется, когда при передаче сообщений на них воздействуют помехи. Помехи в канале связи, в свою очередь, характеризуются условной энтропией.

> **Взаимной (полезной) информацией** между сообщениями **u** и **v** называется величина **I(u,v)**, определяемая соотношением
>
> **I(u,v) = H(u) – H(u|v)**, в котором **H(u)** является энтропией источника информации, а **H(u|v)** представляет собой потерю информации, принимаемой от источника, обусловленную воздействием помех на передаваемое сообщение.

Полезная информация в указанном виде имеет смысл удельной информации в расчете на один передаваемый символ первичного алфавита. Используя зависимость (2.11), можно записать иначе:

$$\mathbf{I(u,v) = H(u) – H(u|v) = H(u) – (H(u,v) – H(v)) = H(u) + H(v) – H(u,v) =}$$

= H(u) + H(v) – (H(u) + H(v|u)) = H(v) – H(v|u)

То есть, формула для взаимной информации симметрична:

I(u,v) = H(u) – H(u|v) = H(v) – H(v|u) (3.8)

В формуле (3.8) использованы следующие обозначения:

H(u) – априорная энтропия источника сообщения;

H(u|v) – апостериорная энтропия, которая учитывает утечку информации при передаче из-за разрушения ее помехами. Иначе называется **ненадежность канала**;

H(v) – энтропия приемника (выхода) канала;

H(v|u) – характеризует постороннюю информацию, вносимую помехами. Называется **энтропия шума**.

Формулу (3.8) можно проиллюстрировать схемой на рис. 3.11.

Пусть передатчик сигнала оперирует алфавитом **Mu**, порождая сигналы $\mathbf{u_i}$, а приемник сигнала обладает алфавитом **Mv** и воспринимает сигналы $\mathbf{v_j}$. Тогда по формуле (2.8):

$$\mathbf{H(v) = -\sum_{j=1}^{Mv} P(v_j) \log P(v_j)}, \quad (3.9)$$

а по формуле (2.10):

$$\mathbf{H(v \mid u) = -\sum_{i=1}^{Mu} P(u_i) \sum_{j=1}^{Mv} P(v_j \mid u_i) \log P(v_j \mid u_i)} \quad (3.10)$$

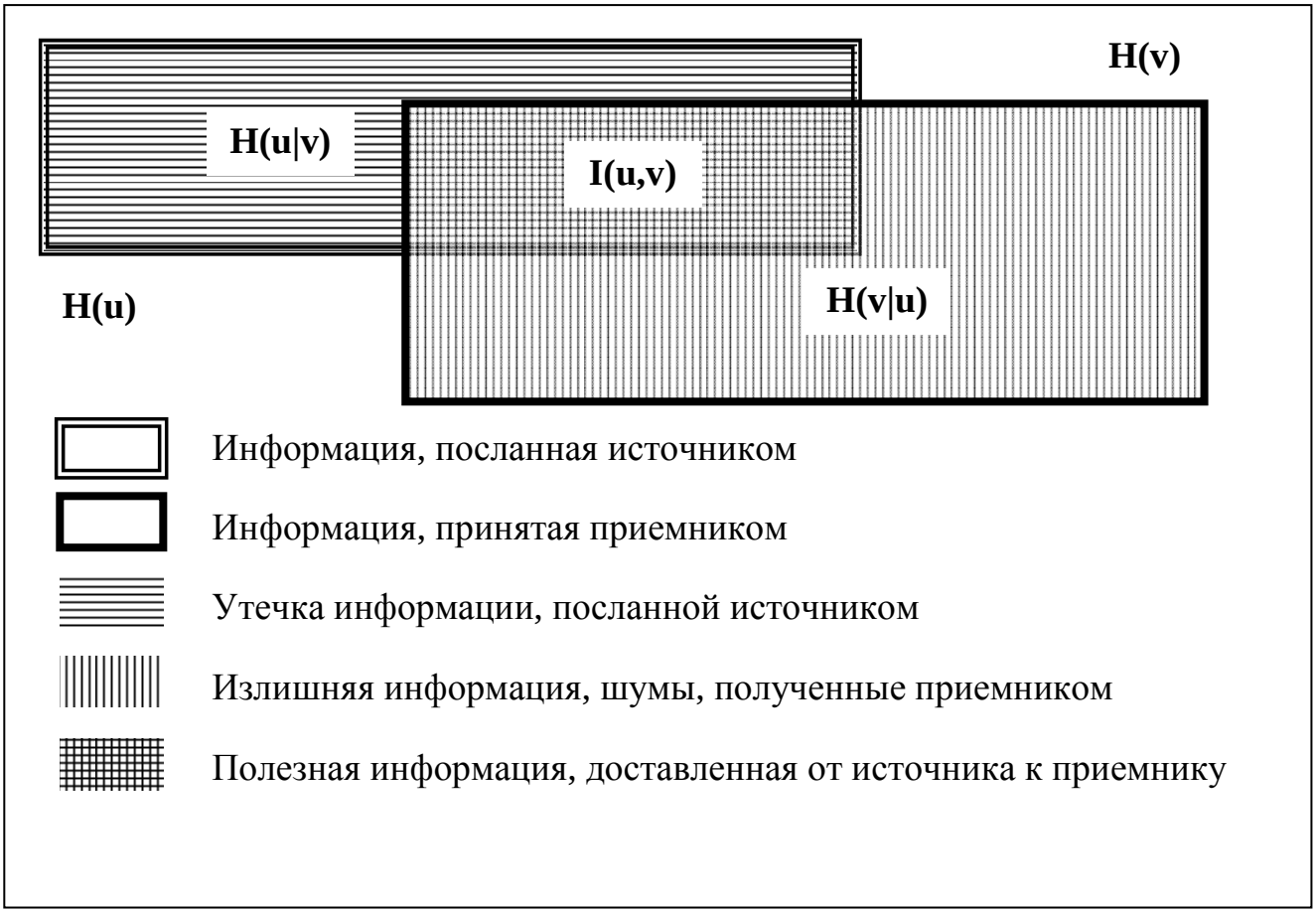


Рис. 3.11. Соотношение характеристик канала связи с помехами

Информация, перемещаемая по каналу связи, определяется в соответствии с формулой (3.8):

$$\mathbf{I(u,v) = -\sum_{j=1}^{Mv} P(v_j)\log P(v_j) + \sum_{i=1}^{Mu} P(u_i)\sum_{j=1}^{Mv} P(v_j \mid u_i)\log P(v_j \mid u_i)} \qquad (3.11)$$

Рассмотрим два крайних случая. Первый случай – абсолютно зашумленный канал, то есть выходной сигнал абсолютно не зависит от входного (обрыв связи). При этом в силу независимости сигналов $\mathbf{P(v_j|u_i)=P(v_j).}$ Подставив это в формулу (3.11), поменяв порядок суммирования и учтя, что $\sum_{i=1}^{M_V} \mathbf{P(u_i) = 1}$, получим:

$$\mathbf{I(u,v) = -\sum_{j=1}^{Mv} P(v_j)\log P(v_j) + \sum_{j=1}^{Mv} P(v_j)\log P(v_j) = 0}$$

То есть в случае обрыва связи полезная информация отсутствует.

Второй случай – отсутствие помех. При этом наблюдается жесткая статистическая связь между входом и выходом: $\mathbf{P(v_j|u_i)=\{1, 0\}.}$ Если $\mathbf{P(v_j|u_i)=1}$

то $\log P(v_j|u_i)=0$. Если $P(v_j|u_i)=0$, то $P(v_j|u_i)\cdot\log P(v_j|u_i)=0$ (см. доказательство первого свойства энтропии в разделе 3.3). В любом случае $H(v|u)=0$, а значит $I(u,v) = H(v) = H(u)$. Как видим, в этом случае информация источника доходит до приемника без изменений.

Скорость передачи информации в канале с помехами определяется как количество полезной информации, передаваемое по каналу в единицу времени. С учетом того, что полезная информация, определяемая по формуле (3.8), передается одним символом первичного алфавита за время τ, скорость передачи вычислим следующим образом:

$$J = \frac{I(u,v)}{\tau}, \text{ отсюда:}$$

$$J = \frac{1}{\tau}(H(u) - H(u|v)) = \frac{1}{\tau}(H(v) - H(v|u)), \qquad (3.12)$$

где τ – средняя длительность передачи одного символа первичного алфавита.

Пропускная способность также определяется по аналогии с каналом без помех, с учетом потерь информации.

$$C = \max J = \frac{1}{\tau}\max\{H(u) - H(u|v)\} = \frac{1}{\tau}\max\{H(v) - H(v|u)\} \qquad (3.13)$$

Понятие о канальной матрице

Ситуацию с передачей информации по каналу связи с помехами можно описать, используя аппарат марковских цепей.

Посылку в канал в момент времени t_k одного из символов u_i представим как нахождение системы в состоянии i. Всего таких состояний – Mu. Прием из канала в следующий момент времени t_{k+1} одного из символов v_j представим как переход системы из состояния i в состояние j. Интервал времени t_{k+1}-t_k равен средней длительности передачи одного символа τ. Вероятности нахождения системы в состоянии i равны вероятности генерации источником символа u_i. Вероятности перехода за время из состояния i в состояние j равны условной вероятности $p(v_j|u_i)$ принятия символа v_j при условии, что послан символ u_i. Таким образом, функционирование системы на одном шаге описывается вектором начальных вероятностей $p(u_i)$, $(i=\overline{1,M_u})$, представляющим собой распределение вероятностей символов первичного алфавита, и матрицей переходных вероятностей $p(v_j|u_i)$ размером $Mu \times Mv$.

$$\begin{bmatrix} p(u_1) \\ p(u_2) \\ \dots \\ p(u_{Nu}) \end{bmatrix} \begin{bmatrix} p(v_1|u_1) & p(v_2|u_1) & \dots & p(v_{Nv}|u_1) \\ p(v_1|u_2) & p(v_2|u_2) & \dots & p(v_{Nv}|u_2) \\ \dots & \dots & \dots & \dots \\ p(v_1|u_{Nu}) & p(v_2|u_{Nu}) & \dots & p(v_{Nv}|u_{Nu}) \end{bmatrix}$$

Матрица переходных вероятностей, используемая для описания ситуации передачи сообщений по каналу с помехами, называется канальной матрицей.

Пропускная способность бинарного симметричного канала с помехами типа «инверсия»

Рассмотрим работу достаточно типичной системы связи, в которой информация передается двоичными сигналами «**0**» и «**1**», имеющими разные уровни квантования. Приемное устройство (демодулятор) анализирует выход канала в течение промежутка времени, соответствующего длительности элементарного сигнала, для которой мы ранее приняли обозначение τ_0. Затем вычисляет некоторую скалярную величину μ - средний за время τ_0 уровень принятого сигнала. Решение принимается сравнением величины μ с некоторым порогом ρ. При $\mu > \rho$ принимается решение в пользу «**1**», а в противном случае, при $\mu < \rho$, решением будет «**0**». При правильном выборе порога ρ вероятности ошибок при передаче сигналов «**0**» и «**1**» будут одинаковыми, и мы приходим к модели **бинарного симметричного канала с инверсией** (Б.С.К.И.). Недостаток такой простейшей схемы приема состоит в том, что демодулятор теряет информацию о надежности принимаемых сигналов. Очевидно, значениям $\mu \approx \rho$ соответствуют ненадежные решения, и эти сведения могли бы быть полезными при последующей обработке информации.

Сформулируем модель Б.С.К.И. Пусть на вход канала подаются сигналы двух типов (u_1 и u_2 – например, импульс и пауза или же сигнал высокого уровня и сигнал низкого уровня) и они же принимаются на выходе, т.е. **{u} = {v}**, **Mu = Mv**. Безошибочный прием сигнала означает, что при посылке u_1 принимается v_1, а при посылке u_2 принимается v_2. Пусть, далее, вероятность ошибки передачи для обоих типов сигналов одинакова и равна **p**. Тогда вероятность безошибочной передачи равна **1 - p**. То есть можно записать:

$$P(v_1|u_1) = P(v_2|u_2) = 1 - p;\ P(v_2|u_1) = P(v_1|u_2) = p,$$

В виде графа такой канал можно представить следующим образом (рис. 3.12):

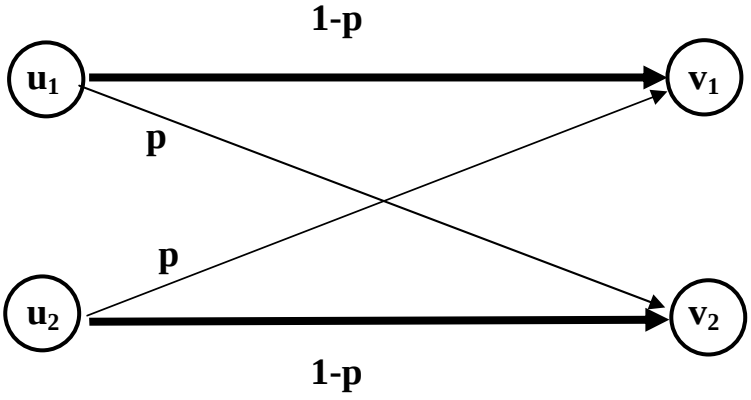


Рис. 3.12. Граф передачи сигнала
в бинарном симметричном канале с инверсией

Линии со стрелками указывают, в какие принимаемые сигналы могут перейти те, что отправлены на входе; рядом со стрелками указаны вероятности соответствующих переходов. Толстыми стрелками указаны безошибочные переходы, тонкими – ошибочная инверсия сигналов. Эту же систему можно представить в виде матрицы переходов марковской цепи с переходными вероятностями:

P(v|u)	**v_1**	**v_2**
u_1	**1-p**	**p**
u_2	**p**	**1-p**

Такой канал называется двоичным симметричным.

Найдем пропускную способность двоичного симметричного канала с инверсией. Для этого потребуется вычислить полезную информацию **I(u,v)**, найти скорость передачи информации и установить ее максимум как функции от вероятности ошибки **p**.

$$\mathbf{I(u,v) = H(v) - H(v|u)}$$

Вычислим энтропию принятого сигнала:

$$\mathbf{H(v) = - P(v_1)\cdot\log_2 P(v_1) - P(v_2)\cdot\log_2 P(v_2)}$$

и энтропию шума:

$$\mathbf{H(v|u) = -P(u_1)\cdot(P(v_1|u_1)\cdot\log_2 P(v_1|u_1) + P(v_2|u_1)\cdot\log_2 P(v_2|u_1)) - P(u_2)\cdot(P(v_1|u_2)\cdot\log_2 P(v_1|u_2) + P(v_2|u_2)\cdot\log_2 P(v_2|u_2)).}$$

Подставляя вероятности из матрицы перехода, получим:

$$\mathbf{H(v|u) = -P(u_1)\cdot((1-p)\cdot\log_2(1-p) + p\cdot\log_2 p) - P(u_2)\cdot(p\cdot\log_2 p + (1-p)\cdot\log_2(1-p)) =}$$

$$\mathbf{= -(P(u_1) + P(u_2))\cdot((1-p)\cdot\log_2(1-p) + p\log_2 p)) = -(1-p)\cdot\log_2(1-p) - p\cdot\log_2 p)}$$

При заданных вероятностях ошибок энтропия **H(v|u) –** величина постоянная. Максимум скорости передачи информации можно искать, варьируя вероятности **$P(v_i)$.** Известно, что для бинарной системы энтропия будет максимальна, если сигналы равновероятны, и равна при этом **1**, то есть **max H(v) = 1.** Значит, пропускная способность будет равна

$$\mathbf{C = \frac{1}{\tau_0}\max(H(v) - H(v\,|\,u)) = \frac{1}{\tau_0}(1 + p\log_2 p + (1-p)\log_2(1-p))} \qquad (3.14)$$

График функции **C(p)** изображен на рисунке 3.13.

Максимального значения, равного $\mathbf{1/\tau_0}$, функция **C** достигает в двух точках: при **p = 0**, что, очевидно, означает отсутствие помех, и при **p = 1**, что соответствует ситуации, когда канал всегда инвертирует входные сигналы (то есть заменяет **0** на **1**, а **1** на **0**). Детерминированная инверсия входных сигналов не служит препятствием для однозначной идентификации посланного сигнала по принятому и, следовательно, не снижает пропускной способности канала. Во всех остальных ситуациях (т.е. при $\mathbf{0 < p < 1}$) верно неравенство $\mathbf{C < 1/\tau_0}$. Судя по графику функции, с ростом вероятности инверсии двоичного разряда **p** на интервале от **0** до **0.5** пропускная способность канала с помехами резко уменьшается, не пропорционально росту **p**. Например, при **p=0.01** пропускная способность уменьшается на **8%**, а при **p=0.1** – уже на **43%**. Наконец, при **p = 0.5** пропускная способность становится равной **0** – это вполне естественно, поскольку вероятность искажения **0.5** означает, что независимо от того, какой сигнал был послан, на приемном конце с равной вероятностью может появиться любой из двух допустимых сигналов. Ясно, что передача в таких условиях оказывается невозможной.

Поскольку канал двоичный, $\mathbf{1/\tau_0 = V = C_0}$ (так обозначим **идеальную пропускную способность**, то есть пропускную способность двоичного канала без помех). Произведя соответствующую замену, получим:

$$\mathbf{C = C_0(1 + p\log p + (1-p)\log(1-p))} \qquad (3.15)$$

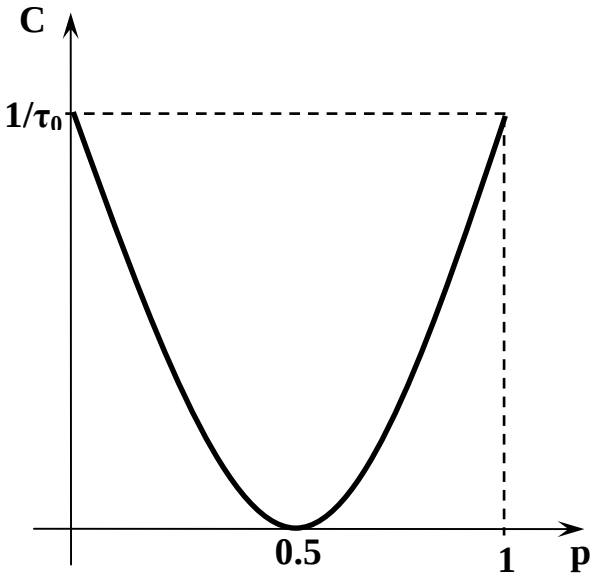


Рис. 3.13. Зависимость пропускной способности бинарного симметричного канала с инверсией от вероятности ошибок

Идеальная (или, иначе, конструктивная или техническая) пропускная способность $\mathbf{C_0}$ зависит только от аппаратуры канала, в частности, от частоты тактового генератора, определяющей длительность элементарного импульса. Выражение в скобках не превышает **1**, следовательно, справедливо соотношение: $\mathbf{C \le C_0}$, т.е. можно считать доказанным, что наличие помех снижает пропускную способность (и даже может сделать ее равной **0**).

Пропускная способность симметричного канала со стиранием

Рассмотренный выше способ принятия решения относительно поданного в канал сигнала можно усовершенствовать, введя «защитный интервал» или «зону стирания» величиной $\mathbf{2\Delta}$. Алгоритм модифицируется следующим образом. При $\mathbf{\mu > \rho + \Delta}$ решение принимается в пользу сигнала «**1**», а при $\mathbf{\mu < \rho - \Delta}$ принимаются решение в пользу «**0**». А при попадании в зону стирания, то есть при $\mathbf{\rho - \Delta < \mu < \rho + \Delta}$ будем считать, что сигнал становится «нераспознаваемым». Получаем модель бинарного симметричного канала со стиранием символа (Б.С.К.С.). Это небольшое изменение заметно повышает эффективность системы, поскольку задача исправления стираний проще задачи исправления ошибок. Один и тот же корректирующий код позволяет исправить примерно в два раза больше стираний, чем инверсий.

Перейдем к моделированию Б.С.К.С. Рассмотрим двоичный канал (на входе сигналы $\mathbf{u_1}$ и $\mathbf{u_2}$ с вероятностями появления $\mathbf{p(u_1)}$ и $\mathbf{p(u_2)}$, соответственно). На приемном конце канала связи любой из них с вероятностью **p** может быть интерпретирован как противоположный (смотри предыдущий раздел), но, помимо этого, с вероятностью **q** искажения в канале оказываются такими, что принятый знак не идентифицируется ни с одним из поступающих на вход. В таком случае можно считать, что принят новый сигнал $\mathbf{v_3}$, появление которого можно интерпретировать как пропажу (стирание) входного сигнала – по этой причине канал назван двоичным симметричным со стиранием (рис. 3.14). Символ $\mathbf{v_3}$ не входит в состав алфавита источника. Тогда

$$\mathbf{P(v_1|u_1) = P(v_2|u_2) = 1 - p - q,\ P(v_2|u_1) = P(v_1|u_2) = p,\ P(v_3|u_1) = P(v_3|u_2) = q,}$$

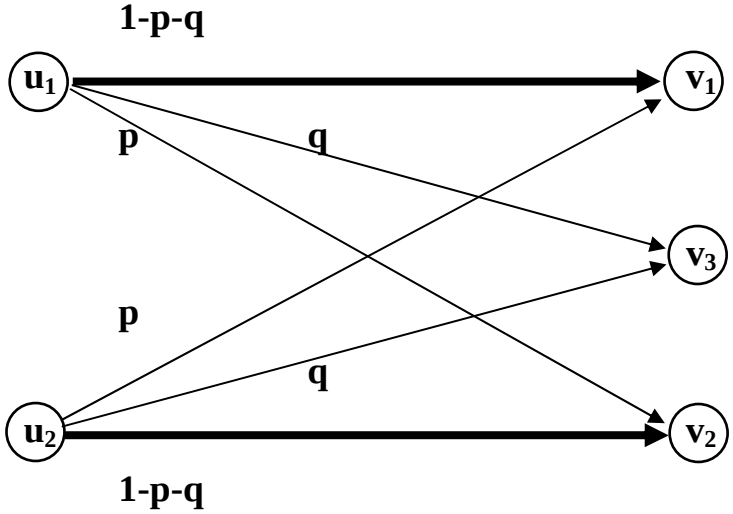


Рис. 3.14 - Граф передачи сигнала

Эту же систему можно представить в виде марковской цепи со следующей матрицей переходных вероятностей.

P(v|u)	$\mathbf{v_1}$	$\mathbf{v_2}$	$\mathbf{v_3}$
$\mathbf{u_1}$	**1-p-q**	**p**	**q**
$\mathbf{u_2}$	**p**	**1-p-q**	**q**

Расчет условной энтропии шума в соответствии с формулой (3.10) дает:

$$\mathbf{H(v|u) = -(u_1)\cdot(P(v_1|u_1)\cdot\log_2 P(v_1|u_1) + P(v_2|u_1)\cdot\log_2 P(v_2|u_1) + P(v_3|u_1)\cdot\log_2 P(v_3|u_1)) - -P(u_2)\cdot(P(v_1|u_2)\cdot\log_2 P(v_1|u_2) + P(v_2|u_2)\cdot\log_2 P(v_2|u_2) + P(v_3|u_2)\cdot\log_2 P(v_3|u_2)).}$$

Подставляя вероятности из матрицы переходных вероятностей, получим:

$$\mathbf{H(v|u) = -P(u_1)\cdot((1-p-q)\cdot\log_2(1-p-q) + p\cdot\log_2 p + q\cdot\log_2 q) -}$$

$$\mathbf{- P(u_2)\cdot(p\cdot\log_2 p + (1-p-q)\cdot\log_2(1-p-q) + q\cdot\log_2 q) =}$$

$$\mathbf{-(P(u_1) + P(u_2))\cdot((1-p-q)\cdot\log_2(1-p) + p\cdot\log_2 p + q\cdot\log_2 q)) =}$$ |в силу того, что $\mathbf{P(u_1)+P(u_2)=1}$| $\mathbf{= -(1-p-q)\cdot\log_2(1-p-q) - p\cdot\log_2 p - q\cdot\log_2 q.}$

Таким образом, в силу (3.8)

$$\mathbf{I(u,v) = H(v) + (1-p-q)\cdot\log_2(1-p-q) + p\cdot\log_2 p + q\cdot\log_2 q.}$$

Поскольку **H(v|u)** не зависит от значений априорных вероятностей, взаимная информация **I(u,v)** достигает максимума при таких вероятностях, когда наибольшее значение приобретает энтропия **H(v)**. Для нахождения **H(v)** необходимо знать вероятности всех сигналов, появляющихся на выходе из канала (обозначим эти вероятности $\mathbf{q_j\ (j = 1,2,3)}$).

Вероятность появления $\mathbf{v_3}$ (то есть стирания символа) уже установлена: $\mathbf{q_3 = q}$. Для $\mathbf{v_1}$ вероятность $\mathbf{q_1 = p(u_1)\cdot(1 - p - q) + p(u_2)\cdot p}$; аналогично для $\mathbf{v_2}$ находим $\mathbf{q_2 = p(u_2)\cdot p + p(u_2)\cdot(1 - p - q)}$. Тогда по формуле (3.9)

$$\mathbf{H(v) = -\sum_{i=1}^{3} q_i \log q_i = -q_1 \log q_1 - q_2 \log q_2 - q \log q}$$

Поскольку **q** определяется конструктивными особенностями канала и не зависит от априорных вероятностей сигналов на входе, наибольшая энтропия

выхода **H(v)** будет при максимальном значении выражения $-\mathbf{q_1 \cdot \log_2 q_1} - \mathbf{q_2 \cdot \log_2 q_2}$, причем, при любых $\mathbf{p(u_1)}$ и $\mathbf{p(u_2)}$ справедливо $\mathbf{q_1 + q_2 = 1 - q}$ (так как $\mathbf{\Sigma\, q_i = 1}$) Можно показать (аналогично доказательству третьего свойства энтропии), что указанное выражение достигает максимума при условии $\mathbf{q_1 = q_2 = 0{,}5 \cdot (1 - q).}$ Тогда

$$\max \mathbf{H(v)} = -\frac{\mathbf{1-q}}{\mathbf{2}} \log \frac{\mathbf{1-q}}{\mathbf{2}} - \frac{\mathbf{1-q}}{\mathbf{2}} \log \frac{\mathbf{1-q}}{\mathbf{2}} - \mathbf{q} \log \mathbf{q} = \mathbf{1-q-(1-q)\log(1-q) - q\log q}$$

$$\max \mathbf{I(u,v) = 1-q-(1-q)\log(1-q) - q\log q + (1-p-q)\log(1-p-q) + p\log p + q\log q}$$

приведя подобные, получим:

$$\max \mathbf{I(u,v) = (1-q)(1-\log(1-q)) + (1-p-q)\log(1-p-q) + p\log p}$$

Окончательно для пропускной способности двоичного симметричного канала со стиранием имеем:

$$\mathbf{C = C_0((1-q)\cdot(1-\log_2(1-q)) + (1-p-q)\cdot\log_2(1-p-q) + p\cdot\log_2 p)} \qquad (3.16)$$

Проанализируем полученный результат. **C = C(p,q),** причем, **C** будет уменьшаться при увеличении как **p**, так и **q**. Если вероятности **p** и **q** отличны от **0**, то, как видно из полученного выражения, $\mathbf{C < C_0}$. В реальных двоичных каналах со стиранием $\mathbf{p < q}$, т.е. вероятность такого искажения входного сигнала, при котором его невозможно распознать, выше вероятности такого искажения, при котором сигнал становится похожим на второй из используемых сигналов. В тех ситуациях, когда **p** пренебрежимо мала и единственным искажением оказывается стирание сигнала, пропускная способность оказывается равной: $\mathbf{C = C_0 \cdot (1 - q).}$ График этой функции представлен на рис. 3.15.

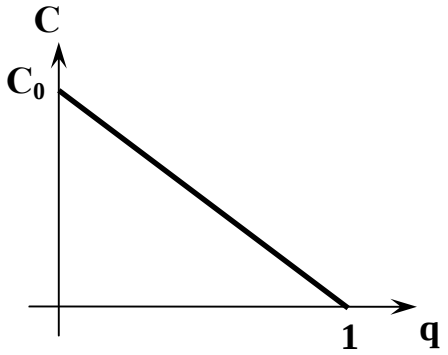


Рис. 3.15 - Зависимость пропускной способности от вероятности стирания

Полученный результат представляется вполне закономерным: при **p = 0** из **V** двоичных сигналов, передаваемых по каналу за единицу времени, в среднем $\mathbf{V \cdot q}$ будет «стираться», но при этом остальные $\mathbf{V \cdot (1- q)}$ сигналов будут

на приемном конце расшифровываться без потерь, и с каждым из них связан ровно **1** бит информации.

Заканчивая рассмотрение характеристик реального дискретного канала передачи информации, мы можем сделать следующие заключения.

Помехи, существующие в реальном канале связи, приводят к снижению его пропускной способности (по сравнению с аналогичным каналом без помех).

Пропускная способность реального канала может быть рассчитана по известным априорным вероятностям. Для их определения требуются статистические исследования передачи информации в канале.

Теорема Шеннона для дискретного канала с помехами

Для дискретного канала с помехами К.Шенноном была доказана следующая теорема (**вторая теорема Шеннона**):

> Если производительность источника $\mathbf{R \leq C - \varepsilon}$, где $\boldsymbol{\varepsilon}$ – сколь угодно малая положительная величина, то существует способ кодирования, позволяющий передать все сообщения источника со сколь угодно малой вероятностью ошибки.

Если производительность информационной системы меньше пропускной способности канала, то сообщение от этого источника можно преобразовать так, чтобы передавать их по каналу с помехами с любой степенью точности, то есть за счет существования избыточности в сообщениях, вводимой специальным образом, можно уменьшить вероятность ошибки до сколь угодно малой величины.

С точки зрения технической реализации эта теорема означает, что существует способ кодирования и декодирования, при котором вероятность ошибочного декодирования может быть сколь угодно малой. Если $\mathbf{R \geq C}$, то таких способов не существует.

Вторая теорема Шеннона является идеологической основой для существования помехоустойчивого (корректирующего) кодирования в каналах связи.

Пропускная способность непрерывного канала связи с помехами

Выше мы обсуждали передачу информации в канале связи посредством дискретных сигналов. Однако при этом непосредственно сам канал связи - проводники, электромагнитное поле, звук, оптоволоконные линии и прочее – свойствами дискретности не обладает. Другими словами, по тем же каналам может передаваться и аналоговая информация – характер передаваемых сигналов определяется передатчиком. Линии связи, основанные на использовании аналоговых сигналов, имеют весьма широкую область практического применения - это радио- и телевизионная связь, телефон и модем, различные телеметрические каналы и тому подобное.

Непрерывным называется канал, который обеспечивает передачу непрерывных (аналоговых) сигналов.

Непрерывные сигналы, поступающие в канал связи из передатчика (Пд) описываются некоторой непрерывной функцией времени **U(t).**

Ограничения на значения этой функции задаются величиной *средней мощности* передаваемых сигналов P_U. Другой характеристикой непрерывного канала, как и канала дискретного, является полоса пропускания – интервал частот сигналов, которые могут распространяться в данном канале $\omega_{min} - \omega_{max}$. Если по своему физическому смыслу U является напряжением или силой электрического тока, то при неизменном электрическом сопротивлении канала связи $P_U \sim U^2$, т.е. мощность сигнала определяет его амплитуду и средний квадрат значения параметра сигнала.

Сигнал на выходе канала **V(t),** поступающий в приемник (Пм), также является аналоговым и формируется он в результате наложения помех, которые можно описать некоторой непрерывной функцией времени **δ(t)**, на входной сигнал; в результате: **V(t) = U(t) ¤ δ(t)** (рис. 3.16). Под символом «¤» понимается какая-либо композиция полезного сигнала и помехи. Чаще применяется аддитивная модель помех, когда информационный сигнал складывается с помехой, реже – мультипликативная модель, когда сигналы перемножаются.

Явный вид функции помех заранее неизвестен. Поэтому для количественного описания прохождения сигналов по непрерывному каналу приходится принимать ту или иную модель помех и модель канала. Наиболее распространенной является модель гауссовского канала: принимается, что помехи, будучи непрерывными случайными величинами, подчиняются нормальному (гауссовскому) статистическому распределению с математическим ожиданием (средним значением) равным нулю (**m[δ] = 0**):

$$\mathbf{f(\delta) = \frac{1}{\sqrt{2\pi\sigma_\delta^2}} \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right)}$$

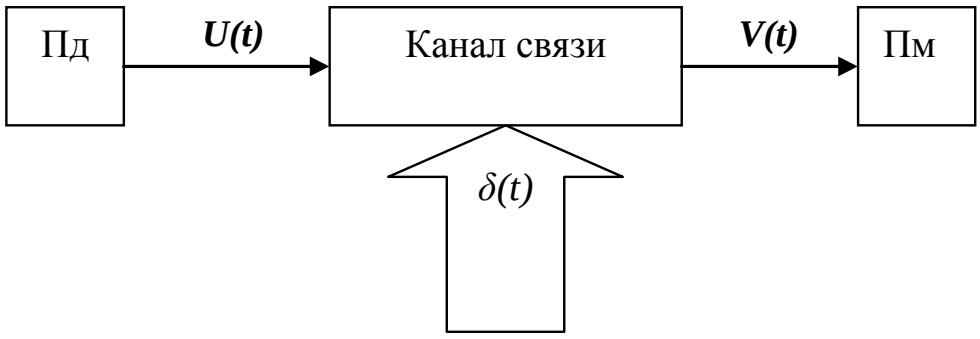


Рис. 3.16 - Аналоговый канал связи с помехами

Эта функция имеет единственный параметр **σ**, квадрат которого называется дисперсией и имеет смысл средней мощности помех.

> Если при этом выполняется условие, что в пределах полосы пропускания средняя мощность помех оказывается одинаковой на всех частотах, а вне этой полосы она равна нулю, то такие помехи называются ***белым шумом***.

Определим количество информации, передаваемое по непрерывному каналу с помехами. Для того чтобы нести информацию, сигнал должен быть случаен, заранее неизвестен для приемника. Поэтому, чтобы описать входной сигнал, надо определить его вероятностные характеристики. Доказано, что нормальное распределения имеет наибольшую энтропию среди всех законов с фиксированной дисперсией. Поэтому в качестве входного сигнала рассмотрим **U(t)**, который представляет собой гауссовский процесс с математическим ожиданием **m[u] = a** и среднеквадратичным отклонением σ_u.

$$\mathbf{f(u)} = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{(u-a)^2}{2\sigma_u^2}\right)$$

В случае применения аддитивной модели помех выходной сигнал **V(t)** тоже имеет гауссовское распределение.

$$\mathbf{f(v)} = \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right)$$

где

$\sigma_v^2 = \sigma_u^2 + \sigma_\delta^2$, **m[v] = m[u] + m[δ] = a.**

Взаимная информация входного и выходного сигнала равна:

I(V,U) = H(V) – H(V|U)

В качестве значений энтропий **H(V)** и **H(V|U)** можно применять приведенные энтропии, так как величины **log ΔV** (см. формулу (2.13)) у них одинаковы и при вычитании компенсируются.

Найдем приведенную энтропию для гауссовского сигнала:

$$\mathbf{H(V)} = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right) \cdot \log_2\left(\frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right)\right) dv =$$

$$= -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(\frac{-(v-a)^2}{2\sigma_v^2}\right) \cdot \log_2\left(\frac{1}{\sqrt{2\pi\sigma_v^2}}\right) dv - \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(\frac{-(v-a)^2}{2\sigma_v^2}\right) \cdot \log_2\left(\exp\left(\frac{-(v-a)^2}{2\sigma_v^2}\right)\right) dz$$

$$= \frac{1}{2}\log_2(2\pi\sigma_v^2) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right) dv + \log_2(e) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right)\left(\frac{(v-a)^2}{2\sigma_v^2}\right) dv$$

$$= \frac{1}{2}\log_2(2\pi\sigma_v^2) + \log_2(e) \cdot \frac{1}{2\sigma_v^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right)(v-a)^2 d(v-a)\,.$$

Дисперсия непрерывных величин для любых законов распределения вводится следующим образом: $\sigma_X^2 = \int_{-\infty}^{+\infty} f(x)x^2 dx$.

Значит, выражение $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right)(v-a)^2 d(v-a)$ представляет собой дисперсию нормальной случайной величины **(v-a)** и равно σ_v^2**.** То есть

$$\log_2(e)\cdot\frac{1}{2\sigma_v^2}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma_v^2}}\exp\left(-\frac{(v-a)^2}{2\sigma_v^2}\right)(v-a)^2 d(v-a) = \frac{1}{2}\log_2(e)$$

Окончательно получаем:

$$H(V) = \frac{1}{2}\log_2(2\pi\sigma_V^2) + \frac{1}{2}\log_2 e = \frac{1}{2}\log_2(2\pi\cdot e\cdot\sigma_V^2) \qquad (3.17)$$

H(V|U) – энтропия шума, определяемая помехами **δ(t).** Для нее получим аналогично:

$$H(V \mid U) = \frac{1}{2}\log_2(2\pi\cdot e\cdot\sigma_\delta^2) \qquad (3.18)$$

Таким образом, информация, передаваемая по непрерывному каналу в условиях гауссовых аддитивных помех, равна:

$$I(V,U) = H(V) - H(V|U) = \frac{1}{2}\log_2(2\pi\cdot e\cdot\sigma_V^2) - \frac{1}{2}\log_2(2\pi\cdot e\cdot\sigma_\delta^2) =$$

$$= \frac{1}{2}\log_2\left(\frac{\sigma_V^2}{\sigma_\delta^2}\right) = \frac{1}{2}\log_2\left(\frac{\sigma_U^2+\sigma_\delta^2}{\sigma_\delta^2}\right) = \frac{1}{2}\log_2\left(1+\frac{\sigma_U^2}{\sigma_\delta^2}\right) \qquad (3.19)$$

Обратим внимание, что формула (3.19) имеет смысл среднего по времени количества полезной информации или энтропии. В системах связи с дискретным характером источника/приемника его можно принять как количество информации в расчете на один импульс (отсчет) тактового генератора канала связи. Для получения пропускной способности количество информации на один отсчет нужно умножить на частоту снятия отсчетов. С учетом правила Найквиста (**C = 2FH**) получим:

$$C = 2F\cdot H = F\cdot\log_2\left(1+\frac{\sigma_U^2}{\sigma_\delta^2}\right) \qquad (3.20)$$

Учитывая, что дисперсия сигнала пропорциональна его мощности, получим

$$C = 2F\cdot H = F\cdot\log_2\left(1+\frac{P_U}{P_\delta}\right) \qquad (3.21)$$

Это формула Шеннона для непрерывного канала с аддитивными гауссовскими помехами.

Для реального канала с ограниченной мощностью сигнала P_X пропускная способность оказывается несколько иной, чем по формуле Шеннона. В этом случае пропускная способность канала может быть рассчитана по формуле:

$$\mathbf{C = F \cdot \log_2(1 + \alpha \frac{P_U}{P_\delta})},$$

где α – коэффициент, учитывающий ухудшение информационных свойств применяемого класса сигналов по сравнению с идеальным гауссовским сигналом, причем $0 \leq \alpha \leq 1$. Как показывают расчеты, $\alpha \approx 0.3$ для экспоненциального сигнала. Для импульсных сигналов $\alpha \approx 0.03$. Для идеального гауссовского сигнала α = 1, и применяется классическая формула Шеннона.

Пропускная способность определяется отношением мощностей сигнала и помех, а также шириной спектра полезного сигнала. Ограничение пропускной способности непрерывного канала связано с тем, что любые используемые для связи сигналы имеют конечную мощность.

C = 0 только при $\mathbf{P_U = 0}$**.** То есть непрерывный канал обеспечивает передачу информации даже в том случае, если уровень шумов превышают уровень сигнала. Это используется для скрытой (неперехватываемой) передачи.

Приведем характеристики некоторых каналов связи.

Вид связи	*F (Гц)*	P_U / P_δ	*C (бит/с)*
1	2	3	4
Телеграф	120	2^6	640
Телефон	$3 \cdot 10^3$	2^{17}	$5 \cdot 10^4$

Окончание

1	2	3	4
Телевидение	$7 \cdot 10^6$	2^{17}	$130 \cdot 10^6$
Компьютерная сеть			до 10^{10}
Слух человека	$20 \cdot 10^3$		$5 \cdot 10^4$
Зрение человека			$5 \cdot 10^6$

Из сопоставления данных видно, что пропускная способность телефонного канала связи совпадает с пропускной способностью органов слуха человека. Однако она существенно выше скорости обработки информации человеком, которая составляет не более 50 бит/с. Другими словами, человеческие каналы связи допускают значительную избыточность информации, поступающей в мозг.

Мощность шума можно представить так: $\mathbf{P_\delta = F \cdot N_0}$, где $\mathbf{N_0}$ – это мощность белого шума. Тогда формула Шеннона может быть переписана в следующем виде:

$$\mathbf{C = F \log_2 (1 + \frac{P_U}{N_0 F}) = \log_2 (1 + \frac{P_U}{N_0 F})^F = \log_2 \left[(1 + \frac{P_U}{N_0 F})^{\frac{N_0 F}{P_U}} \right]^{\frac{P_U}{N_0}}}.$$

Рассмотрим предел $\mathbf{\lim_{F \to \infty} C = \log_2 e^{\frac{P_U}{N_0}} = \frac{P_U}{N_0} \log_2 e \approx 1.443 \frac{P_U}{N_0}}$.

Из последнего соотношения следует, что для передачи одного бита в секунду необходимо обеспечить мощность полезного сигнала $\mathbf{P_U \geq N_0 / \log_2 e \approx 0.69 N_0}$.

Теорема Шеннона для непрерывных каналов с помехами

При описании передачи непрерывных сигналов возникают проблемы, связанные с невозможностью однозначно дать определение тому, возникла ли ошибка в исходном сообщении. Такой проблемы нет для дискретных сообщений, так как в дискретном случае ошибка понимается как разность между входящим в канал и выходящим сообщениями. Дискретная ошибка может быть нулевой, что интерпретируется как отсутствие помех в канале.

В непрерывном случае даже при безошибочной передаче сигнала по каналу связи приемник воспринимает поступивший сигнал с какой-то погрешностью. То есть всегда есть ненулевая разница между входным и выходным сигналами, но не всякое различие может быть признаком помех и ошибочной передачи информации. Исходя из этих соображений, вводится понятие **эквивалентности** сообщений.

Принятое сообщение **V(t)** и переданное **U(t)** называются эквивалентными, если различие между ними несущественно в смысле выбранного критерия (обычно это критерий среднеквадратичного отклонения). Для оценки существенности отклонения $\mathbf{\varepsilon(t) = u(t) - v(t)}$ задается предельная погрешность

ε_0^2. Если среднее за некоторый временной период отклонение не превосходит допустимого предела: $\overline{\varepsilon^2}(t) < \varepsilon_0^2$, то отклонение признается несущественным, и реализации входного и выходного непрерывных сигналов считаются эквивалентными.

> **Эпсилон-энтропией $H_\varepsilon(U)$** называется минимальное среднее количество информации, содержащееся в одном отсчете сообщения **V(t)** относительно сообщения **U(t)**, при котором эти сообщения еще эквивалентны.

В соответствии с соотношением $H_\varepsilon(U) = \min \{I(U,V) \mid \varepsilon_0^2 \geq \varepsilon^2\}$ эпсилон-энтропия определяет количество существенной информации, содержащейся в одном отсчете непрерывного сообщения.

Производительность источника непрерывных сообщений пропорциональна энтропии источника $R = \frac{H(u)}{\Delta t}$. Поскольку энтропия источника непрерывных сигналов – бесконечно большая, то и производительность такого источника получается бесконечно большой.

Для корректировки понятия производительности источника непрерывных сообщений применяется эпсилон-энтропия. Производительность источника определяется как $R = \frac{H_\varepsilon(U)}{\Delta t}$. По теореме Котельникова $\frac{1}{\Delta t} = 2F$, где **F –** полоса пропускания, значит $R = 2F \cdot H_\varepsilon(U)$.

Теорема Шеннона для непрерывного канала с помехами (**третья теорема**):

> Если при заданном критерии эквивалентности сообщений ε_0^2 производительность источника информации меньше пропускной способности канала, то есть $R < C$, то существует такой способ кодирования и декодирования в обобщенном смысле (т. е. преобразование сообщения в сигнал и обратно), при котором неточность воспроизведения сообщения сколь угодно близка к ε_0^2. При $R > C$ такого способа не существует.

Задачи и вопросы к главе 3

1. Источник генерирует знак $\mathbf{z_1}$ с вероятностью **0.8** и $\mathbf{z_2}$ с вероятностью **0.2**. Постройте эффективные коды для однобуквенного кодирования, кодирования последовательности из двух знаков, трех знаков. Каково среднее число символов на знак? Сравните с энтропией источника.

2. В информационном канале используется алфавит с четырьмя различными символами. Длительности всех символов одинаковы и равны **t = 1** мкс. Определить пропускную способность канала при отсутствии шумов.

3. Источник генерирует символы **A** и **B** с вероятностями P(A)=0.7, P(B)=0.3. Построить эффективный бинарный код для однобуквенного, двухбуквенного и трехбуквенного кодирования. Оценить среднюю длину кодового слова и эффективность каждого кода.

Решение.

1) Составим однобуквенный код. **A → 0**, **B → 1**. Энтропия источника H = 0,88 бит/символ. Средняя длина кода $\mathbf{К_1 = 1}$. Коэффициент эффективности кода $\mathbf{К_{ОЭ} = 0.88/1 = 0.88}$.

2) Составим двухбуквенный код.

Пара	Вероятность пары	Код пары	Длина кода
AA	0.7·0.7=0.49	1	1
AB	0.7·0.3=0.21	01	2
BA	0.3·0.7=0.21	001	3
BB	0.3·0.3=0.09	000	3

Средняя длина кода $\mathbf{К_2=1·0.49+2·0.21+3·0.21+3·0.09=1.81}$; в расчете на один символ $\mathbf{К_1=К_2/2=0.905}$. Коэффициент эффективности кода $\mathbf{К_{ОЭ} = 0.88/0.905 = 0.972}$.

2) Составим алфавит из трёхбуквенных комбинаций.

Пара	Вероятность пары	Код пары	Длина кода
AAA	0.343	11	2
AAB	0.147	10	2
ABA	0.147	011	3
BAA	0.147	010	3
ABB	0.063	0011	4
BAB	0.063	0010	4
BBA	0.063	0001	4
BBB	0.027	0000	4

Средняя длина кода $\mathbf{К_З=2.686}$; в расчете на один символ $\mathbf{К_1=К_З/3=0.895}$.

Коэффициент эффективности кода $\mathbf{К_{ОЭ}}$ = **0.88/0.895 = 0.983**.

4. На вход информационной системы поступает непрерывный сигнал с предельной круговой частотой $\boldsymbol{\omega}$ **= 610** Гц. Сигнал оцифровывается и пропускается далее по двоичному каналу связи. Пропускная способность канала составляет **200** бит в секунду. Возможна ли такая дискретизация входного сигнала, чтобы его можно было передать по каналу связи без потери информации?

5. Источник генерирует три символа первичного алфавита: **A**, **B** и **C** с вероятностями **0.5**, **0.25** и **0.25** соответственно. В кодере символы кодируются равномерным бинарным кодом. При передаче сообщения по каналу связи возможны искажения: знак **0** искажается в **1** с вероятностью **0.1**, знак 1 не искажается. Построить канальную матрицу. Рассчитать энтропию шума, утечку, полезную информацию.

Решение. Закодируем символ **A** словом **00**, символ **B** словом **01**, символ **C** словом **10**. С учетом того, что **P(0→1)=0.1**, **P(0→0)=0.9**, **P(1→0)=0**, **P(1→1)=1**, получим **P(A|A)=P(0|0)·P(0|0)=0.81**, **P(B|A)=P(0|0)·P(1|0)=0.09**, **P(C|A)=P(1|0)·P(0|0)=0.09**, **P(A|B)=P(0|0)·P(0|1)=0**,… Кроме того, приемник будет получать код **11**, не соотнесенный ни с одним символом первичного алфавита. Обозначим этот код буквой **Z** и получим канальную матрицу:

	A	**B**	**C**	**Z**
A	0.81	0.09	0.09	0.01
B	0	0.9	0	0.1
C	0	0	0.9	0.1

6. Источник каждую **0.1** с генерирует один из 8 знаков первичного алфавита с вероятностями:

$\mathbf{x_j}$	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$	$\mathbf{x_4}$	$\mathbf{x_5}$	$\mathbf{x_6}$	$\mathbf{x_7}$	$\mathbf{x_8}$
$\mathbf{p(x_j)}$	**1/2**	**1/4**	**1/8**	**1/16**	**1/32**	**1/64**	**1/128**	**1/128**

Длительность передачи одного элементарного символа по бинарному каналу – **0.01** с.

Найти пропускную способность такого канала связи.

С целью защиты от помех при передаче информации по каналу использован принцип «голосования по большинству из трех».

Можно ли построить 1) равномерный, 2) неравномерный однобуквенный бинарный код, обеспечивающий пропускание такого сигнала?

7. Насколько снижается пропускная способность канала, если средняя частота появления ошибки при передаче сообщения в двоичном симметричном канале составляет **1** ошибочный сигнал на **100** переданных?

Решение. Очевидно, вероятность появления ошибки передачи $\mathbf{p = 0.01}$. Следовательно, по формуле (3.15) получаем:

$$\frac{\mathbf{C}}{\mathbf{C_0}} = \mathbf{1} + \mathbf{p}\log_2 \mathbf{p} + (\mathbf{1}-\mathbf{p})\log_2(\mathbf{1}-\mathbf{p}) \approx \mathbf{0.9192}$$

т.е. пропускная способность канала снизилась приблизительно на **8%**.

8. Во сколько раз средняя мощность сигнала с равномерным распределением значений отсчетов должна быть больше мощности сигнала с нормальным распределением отсчетов при условии, что оба сигнала имеют одинаковые дифференциальные энтропии?

9. По каналу в одну секунду передается $\mathbf{10^6}$ символов (скорость передачи $\mathbf{10^6}$ бод). Символы "**0**" и "**1**" поступают на вход канала с равной вероятностью. Определите пропускную способность канала при следующих условиях:
1) символ «**1»** воспринимается как «**1»** с вероятностью **0.9**, и как «**0»** с вероятности **0.1**, так же искажается и символ «**0»**;
2) в пакете из 4-х символов с вероятностью **0.1** искажается один символ.

10. По каналу в одну секунду передается **1024** символа. Символы **"0"** и **"1"** поступают на вход канала с равной вероятностью. Определите пропускную способность канала при условии: символ **"1"** воспринимается как **1** с вероятностью **0.8**, и как **0** с вероятность **0.2**, символ **"0"** ошибочно инвертируется с вероятностью **0.1** и теряется с вероятностью **0.3**.

11. Источник создает последовательность из алфавита в **16** равновероятных и статистически независимых букв. При передаче по каналу с шумом буквы искажаются так, что четверть всех букв принимается неправильно, причем все ошибки одинаково вероятны. Определить среднюю информацию в принятой букве относительно переданной.

12. Двоичный стирающий канал является одним из наиболее простых типов канала с шумом. В нем переданные символы могут быть "стертыми", но никогда не могут быть приняты ошибочно. Найти среднее количество информации, переносимое одним символом в таком канале, если вероятность стирания равна **0.1** и не зависит от переданного символа; вероятности символов на входе одинаковы.

13. Сообщения источника с производительностью **850** бит/с поступают на вход двоичного симметричного канала с вероятностью инверсии разряда $\mathbf{p = 0.05}$. Длительность символов сигнала в канале $\mathbf{t = 10^{-3}}$ с. Достаточна ли пропускная способность канала для передачи всей информации, поступающей от источника?

14. Сообщения источника с производительностью **850 бит/с** поступают на вход двоичного симметричного канала со стиранием. Длительность символов сигнала в канале $\mathbf{t = 10^{-3}}$ **с**. Какова должна быть вероятность стирания, чтобы

пропускная способность канала была достаточна для передачи всей информации, поступающей от источника?

15. Источник генерирует непрерывный радиосигнал частотой от **6** до **66** мГц. Спектр частот равномерный. Радиодетектор воспринимает сигнал источника и определяет, к какому из пяти предопределенных диапазонов относится сигнал. Затем номер диапазона в равномерном двоичном коде с интервалом **2 с**. пересылается по локальной сети. При передаче возможна инверсия двоичного символа с вероятностью **0.05**. Определить скорость передачи полезной информации о номере диапазона.

16. Источник порождает двоичные символы **0** и **1** с равной вероятностью. Передатчик преобразует сообщения источника в электрический сигнал: символ **1** преобразуется в сигнал напряжением **5 В**, а символ **0** – в сигнал напряжением **0 В**. Сигнал проходит по каналу связи, где на него воздействует аддитивная гауссовская помеха с матожиданием **1 В**, и среднеквадратичным отклонением **1 В**. Приемник дискретизирует сигнал с интервалом **0.1 с** и преобразует его в двоичный код по следующему правилу. Значения сигнала, большие **3 В**, преобразуются в двоичную **1**; значения сигнала, меньшие **2 В**, преобразуются в двоичный **0**; значения сигнала в интервале от **2** до **3 В** не распознаются (считаются стертыми). Найти скорость передачи информации по каналу с помехами.

17. На флоте при передаче сообщений прожектором используется два световых сигнала: длинный (Д) и короткий (К). Оба сигнала в морской азбуке равновероятны. Длительность длинного сигнала в среднем **0.5** с, а короткого – **0.25** с. В тумане каждую восьмую вспышку не удается рассмотреть (теряется), а каждый десятый длинный сигнал воспринимается как короткий. Короткие сигналы распознаются безошибочно (если не теряются). Определить скорость передачи такого канала связи.

Решение. Составим матрицу переходных вероятностей:

	априорные вероятности	Длинный	Короткий	Потеря
Д	0.5	1-0.1-0.125=0,775	0.1	1/8 = 0.125
К	0.5	0	1-0.125=0.875	1/8 = 0.125

Вычислим энтропию сигнала:

H(v)=-p(Д)·log p(Д)-p(К)·log p(К)-p(П)·log p(П) =-0.5·0,775·log(0.5·0.775)-(0.5·0.1+0.5·0,875)·log(0.5·0.1+0.5·0.875)-(0.5·0.125+0.5·0.125)·log(0.5·0.125+0,5·0,125)=

-0,3875·log0.3875-0.4875·log0.4875-0.125·log0.125=1.41 (бит)

Вычислим энтропию шума:

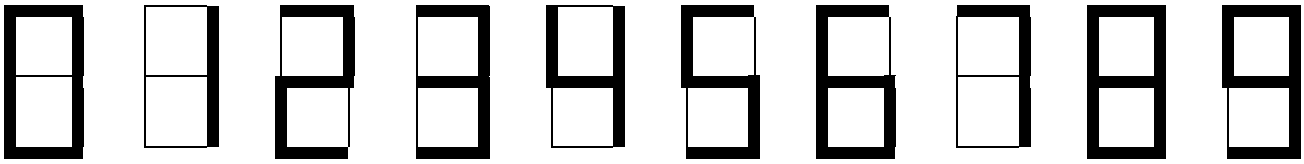
H(v/u)=-0.5·(0.775·log0.775+0.1·log0.1+0.125·log0.125)-0.5(0.875·log0.875+0.125·log0.125)=0.77 (бит)

I(u,v)=H(v)-H(v/u)=0.64 (бит)

Средняя длительность одного сигнала: **0.5·0.5**(с)+**0.5·0.25**(с)=**0.375**(с)

Скорость передачи**: 0,64/0,375=1,71** (бит/с)

18. Цифровое табло показывает время в 24-часовом формате: две цифры для часов и две цифры для минут. Изображение каждой цифры формируется из 7 секций, которые могут находиться в состоянии «включено» или «выключено». Цифры изображаются так:



При отображении цифры возможны неполадки, связанные с тем, что секция может не включиться, в то время как она должна быть включена. Вероятность того, что секция ошибочно не включится, для каждой секции равна **0.1**. Обратное невозможно, то есть, если секция не должна быть включена, то ошибочное включение не произойдет. Рассчитать помехи, связанные с передачей информации на табло.

19. Человек в спокойном темпе произносит фразу: «***не имей сто рублей а имей сто друзей»***. Оценить техническую пропускную способность канала связи, достаточную для синхронной передачи этой фразы в двоичном побуквенном коде с учетом возможных **10%**-х искажений.

20. По сюжету задачи №37 из главы 2 найти пропускную способность канала связи при условии, что Остап пропускает мимо ушей каждое второе слово Эллочки.

21. На вход информационной системы поступает непрерывный сигнал с частотой манипуляции (полосой пропускания) **F = 300** Гц. Сигнал оцифровывается и передается далее по двоичному каналу связи. Техническая пропускная способность канала (не учитывающая искажения) составляет **1** мегабит в секунду. Возможна ли такая дискретизация входного сигнала, чтобы его можно было передать по каналу связи с учетом возможных **10%**-х искажений?

22. Источник генерирует один из пяти знаков первичного алфавита каждую 0.1 с. с вероятностью:

x_j	x_1	x_2	x_3	x_4	x_5
$P(x_j)$	1/2	1/4	1/8	1/16	1/16

Длительность передачи одного элементарного символа по бинарному каналу – **0.02** с.

Найти пропускную способность канала связи без помех.

Найти пропускную способность канала связи с помехами типа «инверсия», вероятность помех – **0.2**

Найти пропускную способность канала связи с помехами типа «стирание», вероятность помех – **0.4**

Достаточна ли пропускная способность во всех трех случаях при использовании а) равномерного, б) эффективного кодирования?

23. По каналу связи передаются сообщения, которые представляют последовательность десятичных цифр, вероятности которых соответственно равны:

p(0)=0,4; p(1)=0,2; p(2)=0,05; p(3)=0,03; p(4)=0,02;

p(5)=0,15; p(6)=0,08; p(7)=0,01; p(8)=0,04; p(9)=0,02.

Канальная матрица, определяющая потери информации в канале связи имеет вид:

P(y|x)=

0.97	**0,02**	**0,01**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0,02	**0,95**	**0,02**	**0,01**	**0**	**0**	**0**	**0**	**0**	**0**
0	**0,01**	**0,98**	**0,01**	**0**	**0**	**0**	**0**	**0**	**0**
0	**0,01**	**0,02**	**0,94**	**0,02**	**0,01**	**0**	**0**	**0**	**0**
0	**0**	**0**	**0,01**	**0,98**	**0,01**	**0**	**0**	**0**	**0**
0	**0**	**0**	**0,01**	**0,01**	**0,96**	**0,0**	**0,0**	**0**	**0**
0	**0**	**0**	**0**	**0,01**	**0,02**	**0,94**	**0,02**	**0,01**	**0**
0	**0**	**0**	**0**	**0**	**0**	**0,02**	**0,96**	**0,02**	**0**
0	**0**	**0**	**0**	**0**	**0**	**0**	**0,01**	**0,98**	**0,0**
0	**0**	**0**	**0**	**0**	**0**	**0**	**0,01**	**0,01**	**0,98**

Определить:

- энтропию источника информации- H(X);
- безусловную энтропию приемника информации- H(Y);
- общую условную энтропию- H(Y/X);
- скорость передачи информации, если время передачи одного символа первичного алфавита τ=0.1мс;
- определить потери информации в канале связи при передаче 500 символов алфавита;
- количество принятой информации.

Построить эффективный код по методу Шеннона - Фано для передачи сообщений.

Оценить эффективность построенного кода.

24. Сообщения передаются в двоичном коде. Длительность передачи нулевого бинарного символа – $\boldsymbol{\tau_0 = 1}$ с, длительность передачи единичного бинарного символа $\boldsymbol{\tau_1 = 5}$ секунд. Определить скорость передачи информации для случаев равновероятных символов и неравновероятных символов с вероятностями **p(0) = 0.37**, **p(1) = 0.63**.

Решение.

1) Символы равновероятны и независимы.

$$\mathbf{J = \frac{H}{\tau_{ср}} = \frac{\log_2 2}{\frac{1}{2}(\tau_0 + \tau_1)} = \frac{1}{3} = 0{,}33\ (бит/с)}$$

2) Символы неравновероятны: $\mathbf{p_0 = 0.37}$, $\mathbf{p_1 = 0.63}$.

$$\mathbf{J = \frac{H}{\tau_{ср}} = \frac{-\sum_{i=0}^{1} p_i \log p_i}{\sum_{i=0}^{1} \tau_i\, p_i} = \frac{-0.37 \cdot \log 0.37 - 0.63 \cdot \log 0.63}{1 \cdot 0.37 + 5 \cdot 0.63} = 0{,}27\ (бит/с)}$$

25. Источник генерирует один из пяти знаков первичного алфавита каждые **0.2 сек**. с вероятностью:

x_j	x_1	x_2	x_3	x_4	x_5
$p(x_j)$	1/2	1/16	1/8	1/16	1/4

Затем каждый знак передается по бинарному каналу со следующими параметрами: частота тактового генератора – **500 Гц**, вероятность ошибочной инверсии сигнала – **0.2**, вероятность потери сигнала – **0.4**. Существует ли способ кодирования, обеспечивающий передачу информации по бинарному каналу?

Решение:

Для того, чтобы ответить на вопрос задачи, необходимо воспользоваться 2-й теоремой Шеннона (для дискретного канала с помехами), т.е. сравнить производительность источника с пропускной способностью канала.

1) Рассчитаем пропускную способность канала. Речь в задаче идет о бинарном симметричном канале со стиранием. Поскольку канал бинарный, то

$\mathbf{C_0 = 1/\tau_0 = V = 500}$ **бит/с,**

где $\mathbf{C_0}$ **–** идеальная (техническая) пропускная способность для бинарного канала без помех.

Используем формулу (3.16) для пропускной способности двоичного симметричного канала со стиранием. Подставляя в (3.16) $\mathbf{p = 0.2}$, $\mathbf{q = 0.4}$ и $\mathbf{C_0 = 500}$, получаем $\mathbf{C \approx 24{,}5}$ **бит/с**.

2) Рассчитаем производительность источника **R**. Поскольку все символы генерируются за одно и то же время, то **R** найдем по формуле (2.14):

$$\mathbf{R = \frac{-\sum_{i=1}^{m} p_i \log p_i}{\theta} \approx 9.4}\ \text{бит/с}.$$

3) Т.к. **R ≈ 9.4 < C ≈ 24.5**, то способ кодирования, обеспечивающий передачу информации по бинарному каналу, существует.

26. Энтропия источника информации равна **6** бит/символ, энтропия приемника равна **8** бит/символ, энтропия шума равна **3** бита/символ. Вычислить утечку информации из канала связи.

27. Студент Вася может получить на экзамене оценку «5» с вероятностью **30%**, оценку «4» с вероятностью **40%**, оценку «3» с вероятностью **20%**, и даже иногда оценку «2». О своей оценке Вася обязательно сообщает родителям. Причем, о получении отличной оценки Вася всегда говорит правду (молодец!), при других результатах экзамена в **10%** случаев происходит ложное завышение оценки на один балл. Сколько полезной информации получают в среднем родители Васи, узнавая от него о результатах сессии (5 экзаменов)? На сколько увеличилось бы количество получаемой ими информации, если бы Вася всегда говорил правду?

28. Студент Вася вместо того, чтобы готовиться к зачету по теории информационных процессов и систем, идет в ночной клуб, где встречает одну из своих однокурсниц. В шуме он может не расслышать ее имя (тем более что на занятиях они встречались редко) и с вероятностью **0.4** решить, что это Саша, с вероятностью **0.4** - что это Маша и с вероятностью **0.2** - что это Даша. Наутро Вася рассказывает другу Пете о ночной встрече. При этом с вероятностью **0.5** Вася может ошибиться в имени девушки, а с вероятностью **0.1** - соврать, сказав, что провел ночь в библиотеке. Верный друг Петя докладывает о похождениях Васи куратору группы. При этом Петя совершенно точно сообщает о посещении Васей библиотеки и с вероятностью **0.4** ошибается в имени девушки. Рассчитать шумы, утечку и полезную информацию на каждом шаге передачи информации. Оценить количество полезной информации, полученное куратором группы по поводу Васиной подготовки к зачету.

29. Два друга — Петя и Вася — совместно используют канал доступа в Интернет[1] с пропускной способностью **4 Кбайт/сек**. Система балансировки нагрузки настроена таким образом, что если в данный момент времени канал использует только один человек, то скачивание файла происходит со скоростью равной пропускной способности канала, а если канал используют оба друга — пропускная способность канала поровну делится между пользователями. Петя начал скачивать музыкальную композицию. Через **8 секунд** Вася начал скачивать графический файл. Петя закончил скачивать музыкальную композицию через **34 секунды** от начала скачивания своего файла. Музыкальная композиция была оцифрована в режиме «моно» с частотой дискретизации **1024 Гц** и **65536** уровнями квантования. Графический файл содержал **8192 (64×128) пикселей**, кодированных с использованием палитры из **256** цветов. И в файле с музыкальной композицией и в графическом файле не

[1] Сюжет задачи взят с олимпиады СПбИТМО

использовалось сжатие данных. Кроме упомянутых скачиваемых файлов другой нагрузки на канал доступа в Интернет не было. Сколько секунд длится музыкальная композиция, которую скачал Петя? В ответе укажите число.

Решение: Исходя из данных о кодировании графического файла, вычислим его информационный объем. Зная, что каждый пиксель кодируется с использованием палитры из 256 цветов, можно сделать вывод, что на каждый пиксель приходится 1 байт кода. Следовательно, весь графический файл будет иметь информационный объем, равный 8 Кбайт. По условию задачи, одновременная передача файлов делит пропускную способность канала пополам. Следовательно, файл такого объема был получен за 8/2=4 секунды.

Следовательно, время передачи музыкальной композиции составило 34 секунды, из которых 4 секунды скачивание происходило со скоростью 2 Кбайт в секунду, а 30 секунд – со скоростью 4 Кбайт в секунду. Таким образом, можно сделать вывод, что информационный объем музыкальной композиции составил 30*4+4*2 = 128 Кбайт.

Частота дискретизации при кодировании музыкального файла составила 1024 Гц. Следовательно, каждая секунда файла представлена 1024 отсчетами. Каждый отсчет кодирован с 65536 уровнями квантования. Это значит, что каждый отсчет имеет информационный объем 2 Байт, а секунда звучания музыкальной композиции («моно» обозначает, что мы используем только один канал) будет иметь информационный объем 2 Кбайт. Теперь зная общий информационный объем музыкальной композиции и информационный объем, занимаемой одной секундой этой композиции можно вычислить время звучания композиции: 128/2=64 секунды.

30. Перед Новым годом дети пишут письма с просьбами о подарках. По статистике в **40%** случаев ребята просят прислать им игрушки, в **10%** - подарить домашних животных, в **20%** - рассчитывают на сладости, а в остальных случаях хотят денег. Такие письма в среднем через **5 суток** приходят в специальное почтовое отделение, где сортируются. Письма с просьбами об игрушках и домашних животных направляются поездом Деду Морозу, с просьбами о сладостях – на оленях в резиденцию Йоулупукки, а с просьбами о деньгах – самолетом Санта Клаусу. Письма в Лапландию для Йоулупукки доходят без ошибок за **3 дня**, Санта Клаусу за океан – за **два дня**, причем половина писем по поводу денег якобы теряется в пути, а Деду Морозу – за **4 дня**, причем дедушка в половине случаев вместо просьбы о настоящем животном ошибочно понимает, что его просят подарить игрушку. Рассчитать помехи и скорость передачи информации о подарках для детишек.

31. Локальная компьютерная сеть построена по технологии Ethernet с использованием спецификации 10Base – T. Найти реальную пропускную способность среды передачи данных, приняв вероятность инверсии отдельного разряда равной $\mathbf{10^{-4}}$.

32. Инструктор дрессирует собаку, обучая ее командам «Ко мне!», «Сидеть!», «Лежать!» и «Служить!». Он с равной частотой подает команды «Сидеть!», «Лежать!» и «Служить!», предваряя каждую из них командой «Ко мне!». Команду «Ко мне!» собака понимает и выполняет безошибочно. Команду «Сидеть!» понимает и выполняет в трех случаях из четырех. А команды «Лежать!» и «Служить!» путает, не улавливая разницы между ними. Определить количество полезной информации в командах инструктора, поступившей собаке за время тренировки, если было подано 40 команд.

33. Первый советский луноход управлялся с Земли человеком, пилотом лунохода. Пусть у пилота имеется рычаг управления типа «джойстик», у которого 5 фиксированных позиций: 4 позиции направления («вперед», «назад», «направо», «налево») и позиция «нейтральное положение». Из центра управления луноходом сигналы от джойстика передаются на луноход в виде двоичных кодов: «вперед» имеет код **1100**, «назад» - **0011**, «направо» - **1010**, «налево» - **0101**, позиция «нейтральное положение» кодируется как **0000**. Команды из центра управления передаются на луноход каждые **0.2** с. Вероятность инверсии одного бинарного разряда во время передачи составляет **0.1**. Применяемая при передаче система защиты от шумов обеспечивает то, что в коде каждой команды могут быть ошибки не более чем в двух разрядах. Если луноход не распознал команду, он выполняет команду «нейтральное положение». Рассчитать скорость передачи информации при передаче команд с Земли на луноход.

34. Какой знак следует поставить между левой и правой частями формулы, чтобы она оказалась верной?

$H(v)$ **?** $H(u) + H(v|u) - H(u|v)$

35. От чего зависит скорость передачи информации в канале связи?

36. Что означает свойство симметричности бинарного канала с помехами?

37. Укажите возможные причины, из-за которых скорость передачи информации не достигает значения пропускной способности канала связи

Глава 4
Помехоустойчивое кодирование сообщений

В предыдущей главе были рассмотрены основы эффективного кодирования данных, задача которого – представить подлежащие передаче сообщения в максимально компактной форме. Целью помехоустойчивого кодирования является такое представление сообщений, которое бы предотвращало или парировало искажения, возникающие при передаче. На практике ситуации с естественным или намеренным изменением информации встречаются чрезвычайно часто.

Для начала рассмотрим задачу о разведчиках, которая поможет изучить основные подходы и алгоритмы помехоустойчивого кодирования.

> Задача о разведчиках[1]. Разведывательный отряд в составе командира и восьми бойцов высадился на вражеском острове вблизи перекрестка, откуда исходят дороги по четырем направлениям. Задача отряда – обнаружить секретный объект противника. Об объекте известно, что он расположен на одной из четырех дорог в часе ходьбы от перекрестка. Командиру известно, что двое из его бойцов – предатели, которые будут стараться обмануть командира, чтобы не допустить обнаружения объекта. Необходимо спланировать действия командира по обнаружению объекта и предателей.
>
> К этой задаче мы будет обращаться многократно. Для указания на то, что рассматривается задача о разведчиках, будем использовать символ .

При передаче информации по каналу связи с помехами в принятых данных могут возникать ошибки. Если такие ошибки имеют небольшую величину или возникают достаточно редко, информация может быть использована потребителем. При большом числе ошибок полученной информацией пользоваться нельзя.

Ранее уже говорилось о понятии **помехоустойчивости** (способности информационных систем противостоять воздействию помех). Для реализации принципа помехоустойчивости информационных систем может быть использовано **помехоустойчивое кодирован**ие.

> Помехоустойчивыми (корректирующими) называются **коды,** позволяющие обнаружить и при необходимости исправить ошибки в принятом сообщении.

Возможность использования кодирования для уменьшения числа ошибок в канале была теоретически показана К. Шенноном в 1948 году в его работе

---

[1] Автор сюжета задачи – К.Кноп. См. Константин Кноп. О разведчиках и кодах Хемминга / Компьютера, 1997, №6.

"Математическая теория связи". Теперь это утверждение принято именовать **второй теоремой Шеннона** (см. раздел 3.4).

4.1. Общие принципы помехоустойчивого кодирования

Хотя различные схемы кодирования очень не похожи друг на друга и основаны на различных математических теориях, всем им присущи два общих свойства.

Первое − **использование избыточности**. Помехоустойчивые последовательности всегда содержат избыточность, например, в виде дополнительных символов.

Второе — **свойство усреднения**, означающее, что избыточные символы зависят от нескольких информационных символов, то есть информация, содержащаяся в исходной кодовой последовательности**,** перераспределяется затем также и на избыточные символы.

Пусть **M** – число знаков первичного алфавита. Длина равномерного двоичного кода $\mathbf{k \geq \log_2 M}$, при этом каждый знак получает свою уникальную последовательность знаков вторичного (бинарного) алфавита. Общее число кодовых комбинаций $\mathbf{S_p = 2^k}$, очевидно, $\mathbf{S_p \geq M}$.

Например, мощность алфавита **M = 64**. Длина равномерного бинарного кода $\mathbf{k \geq \log_2 M = 6}$. И наоборот - с помощью **6** бит можно получить $\mathbf{2^6=64}$ кодовых комбинаций. Они будут считаться разрешенными.

В дальнейшем будем называть часть помехоустойчивого кода, составленную из указанных **k** бит, **информационной** (поскольку именно она содержит информацию о передаваемом знаке первичного алфавита). Если пересылать только эти информационные биты, то любое искажение, состоящее в инверсии хотя бы одного бита, приведет к появлению новой разрешенной кодовой комбинации, которая будет без сомнений воспринята получателем информации. Следовательно, такое искажение обнаружено быть не может.

Сколько информационных бит потребуется командиру разведчиков?

Ответ: сообщение от одного разведчика о том, что на дороге обнаружен разыскиваемый объект, будем обозначать **1**, а сообщение об отсутствии объекта на дороге – **0**. Для обнаружения объекта достаточно получить информацию о наличии/отсутствии объекта с трех дорог. Следовательно, количество информационных разрядов равно **3**.

Возможность обнаружения и исправления ошибок в помехоустойчивых кодах достигается тем, что после первичного кодирования (установления соответствия каждому знаку первичного алфавита его кода) осуществляется вторичное кодирование, в ходе которого к **k** информационным битам по определенным правилам добавляются **r проверочных** (**корректирующих**) разрядов. В результате общая длина кодовой комбинации становится равной **n = k + r**. В дальнейшем такие коды будем называть **(n,k)-кодами**, а число

возможных кодовых комбинаций, составленных из **n** разрядов, возрастает до $S = 2^n$. Из них не все оказываются разрешенными – их только S_p, остальные же $S_f = S - S_p$ комбинаций являются запрещенными.

Допустим, помехоустойчивый код содержит **6** информационных и **3** проверочных бита. Общая длина кодового слова будет равна **6+3=9** бит, а общее количество кодов $S=2^9=512$. Из них разрешенных кодов – $2^6=64$, а запрещенных – **448**.

Командир по одной дороге идет сам, по остальным посылает разведчиков. Определить количество проверочных битов.

Ответ: Всего командир получит от бойцов **8** однобитных сообщений, **три** из них – информационные, остальные **5** – проверочные.

Если при передаче возникает ошибка, она проявится в том, что разрешенная кодовая комбинация перейдет в запрещенную – это можно отследить и даже исправить. Такое обнаружение, очевидно, окажется невозможным, если в результате ошибки передачи одна разрешенная кодовая комбинация перейдет в другую разрешенную. В связи с этим возникает проблема поиска таких способов избыточного кодирования, при которых вероятность перехода одной разрешенной кодовой комбинации в другую была бы минимальной.

Классификация помехоустойчивых кодов

Первый классификационный признак – коды бывают **блочными** или **непрерывными** (рис. 4.1). При **блочном кодировании** передаваемые двоичные сообщения сгруппированы в блоки, которыми кодируются знаки (или группы знаков) первичного алфавита. В блоке присутствуют информационные и проверочные биты. Известно, что если все кодовые комбинации имеют одинаковую длину, код называется равномерным; если нет – неравномерным. При декодировании удобнее (проще) иметь дело с равномерным кодом, поэтому именно он, как правило, используется в помехоустойчивом кодировании. **Непрерывные** (синонимы: цепные, сверточные, рекуррентные) коды представляют собой непрерывную последовательность бит, не разделяемую на блоки (информационные и проверочные биты в них чередуются по определенному правилу). Блочное кодирование удобно использовать в тех случаях, когда исходные данные по своей природе уже сгруппированы в какие-либо блоки или массивы. При передаче по радиоканалам чаще используется сверточное кодирование, которое лучше приспособлено к побитовой передаче данных. Кроме этого, при одинаковой избыточности сверточные коды, как правило, обладают лучшей исправляющей способностью.

Рис. 4.1. Классификация помехоустойчивых кодов

Второй классификационные признак, относящийся как к блочным, так и к непрерывным кодам, подразделяет коды на **разделимые** и **неразделимые**. **Разделимыми** называются коды, в которых информационные и проверочные биты располагаются в строго определенных позициях. В **неразделимых** кодах такой определенности нет, что затрудняет их кодирование и декодирование. Поэтому практический интерес представляют в основном разделимые коды, а из неразделимых – только **коды с постоянным весом**. Под кодами **с постоянным весом** понимаются такие, у которых соотношение информационных и проверочных разрядов на изменяется в процессе кодирования. Соответственно, **коды с переменным весом** таким качеством не обладают.

Третий классификационный признак относится только к блочным разделимым кодам – они подразделяются на **систематические** (линейные) и **несистематические**. Двоичный код является **линейным**, если сумма **по модулю 2** двух кодовых слов также является кодовым словом этого кода. В линейных кодах проверочные биты являются результатом линейных операций над информационными разрядами. В **несистематических** (нелинейных) кодах

информационные и проверочные биты либо вообще не имеют связи, либо эта связь нелинейна – такие коды применяются редко.

Наиболее часто в линиях связи используются блочные линейные коды, называемые **(n,k)-коды**, к которым относятся циклические, коды Хемминга, матричные канонические и ряд других.

Примеры простейших кодов

Код с проверкой на четность. Самым простым линейным блочным кодом является (**n,n-1**)-код, построенный с помощью одной общей проверки на четность. Например, кодовое слово (**4,3**)-кода можно записать в виде вектора-столбца:

$$\mathbf{u} = (\mathbf{m_1}, \mathbf{m_2}, \mathbf{m_3}, \mathbf{m_1} \oplus \mathbf{m_2} \oplus \mathbf{m_3}),$$

где $\mathbf{m_i}$ - символы исходной информационной последовательности, принимающие значения **0** и **1**, а суммирование производится по модулю **2** и обозначается символом ⊕.

Основная идея проверки на четность состоит в следующем. Пусть информационная последовательность источника имеет вид **m = (1 0 1)**.

Тогда соответствующая ей кодовая последовательность будет выглядеть так: $\mathbf{u} = (\mathbf{u_1}, \mathbf{u_2}, \mathbf{u_3}, \mathbf{u_4}) = (1\ 0\ 1\ 0)$, где проверочный символ $\mathbf{u_4}$ формируется путем суммирования символов информационной последовательности **m**:

$$\mathbf{u_4} = \mathbf{m_1} \oplus \mathbf{m_2} \oplus \mathbf{m_3} = 1 \oplus 0 \oplus 1 = 0.$$

Если число единиц в последовательности **m** четно, то результатом суммирования будет **0**, если нечетно — **1**, то есть проверочный символ дополняет кодовую последовательность таким образом, чтобы количество единиц в ней было четным.

При передаче по каналам связи в принятой последовательности возможно появление ошибок, то есть символы принятой последовательности могут отличаться от соответствующих символов переданной кодовой последовательности (нуль переходит в единицу, а **1** − в **0**). Если при передаче рассматриваемого (**4,3**)-кода произошла одна ошибка, причем неважно, в какой его позиции, то общее число единиц в принятой последовательности уже не будет четным. Таким образом, признаком отсутствия ошибки в принятой последовательности может служить четность числа единиц. Поэтому такие коды и называются кодами с проверкой на четность. Правда, если в принятой последовательности произошло две ошибки, то общее число единиц в ней снова станет четным и ошибка обнаружена не будет. Однако вероятность двойной ошибки значительно меньше вероятности одиночной, поэтому наиболее вероятные одиночные ошибки таким кодом обнаруживаться все же будут.

Отметим следующий момент. Если посимвольно сложить два кодовых слова, принадлежащих рассматриваемому (**4, 3**)-коду:

a = (a$_1$, a$_2$, a$_3$, a$_1$ ⊕ a$_2$ ⊕ a$_3$), и **b = (b$_1$, b$_2$, b$_3$, b$_1$ ⊕ b$_2$ ⊕ b$_3$)**,

то получим

c = (a$_1$⊕b$_1$, a$_2$⊕b$_2$, a$_3$⊕b$_3$, a$_1$⊕b$_1$⊕a$_2$⊕b$_2$⊕a$_3$⊕b$_3$) = (c$_1$, c$_2$, c$_3$, c$_1$⊕c$_2$⊕c$_3$),

то есть проверочный символ в новом слове **c** определяется по тому же правилу, что и в слагаемых. Поэтому **c** также является кодовым словом данного кода.

Этот пример отражает важное свойство линейных блочных кодов — **замкнутость**, означающее, что сумма двух кодовых слов данного кода также является кодовым словом.

Несмотря на свою простоту и не очень высокую эффективность, коды с проверкой на четность широко используются в системах передачи и хранения информации. Они ценятся за невысокую избыточность: достаточно добавить к передаваемой последовательности всего один избыточный символ − и можно узнать, есть ли в принятой последовательности ошибка. Правда, определить место этой ошибки и, следовательно, исправить ее, пока нельзя. Можно лишь повторить передачу слова, в котором была допущена ошибка, и тем самым ее исправить.

Итеративный код. Еще одна простая схема кодирования, которая также часто используется, может быть построена следующим образом.

Предположим, что нужно передать, к примеру, девять информационных символов **m = (m$_1$, m$_2$, ..., m$_9$).** Эти символы можно расположить в виде квадратной матрицы, как это показано в таблице 4.1, и добавить к каждой строке и каждому столбцу этой таблицы по проверочному символу (проверка на четность).

Таблица 4.1. Формирование проверочных символов итеративного кода

m$_1$	**m$_2$**	**m$_3$**	**m$_1$ ⊕ m$_2$ ⊕ m$_3$**
m$_4$	**m$_5$**	**m$_6$**	**m$_4$ ⊕ m$_5$ ⊕ m$_6$**
m$_7$	**m$_8$**	**m$_9$**	**m$_7$ ⊕ m$_8$ ⊕ m$_9$**
m$_1$ ⊕ m$_4$ ⊕ m$_7$	**m$_2$⊕ m$_5$ ⊕ m$_8$**	**m$_3$ ⊕ m$_6$ ⊕ m$_9$**	**m$_1$ ⊕ m$_2$ ⊕ m$_3$ ⊕ … ⊕ m$_9$**

Таким образом, по строкам и по столбцам этой таблицы будет выполняться правило четности единиц.

Если в процессе передачи по каналу с помехами в этой таблице произойдет одна ошибка (например, в символе **m$_4$**), то проверка на четность в соответствующей строке и столбце (в данном случае – в первом столбце и

второй строке) не будет выполняться. Иными словами, координаты ошибки однозначно определяются номерами столбца и строки, в которых не выполняются проверки на четность. Таким образом, этот код, используя различные проверки на четность (по строкам и по столбцам), способен не только обнаруживать, но и исправлять ошибки (если известны координаты ошибки, то ее исправление состоит просто в инверсии, то есть замене символа на противоположный: если **0**, то на **1**, если **1** – то на **0**).

Описанный метод кодирования, называемый итеративным, оказывается полезным в случае, когда данные естественным образом формируются в виде массивов, например, на шинах ЭВМ, в памяти, имеющей табличную структуру, и т.д. При этом размер таблицы в принципе не имеет значения (**3×3** или **20×20**), однако в первом случае будет исправляться одна ошибка на **3×3=9** символов, а во втором – одна на **20×20=400** символов.

Если в простом коде с проверкой на четность для обнаружения ошибки приходится добавлять к информационной последовательности всего один символ, то для того, чтобы код стал исправлять однократную ошибку, понадобилось к девяти информационным символам добавить еще семь проверочных. Таким образом, избыточность этого кода оказалась очень большой, а исправляющая способность – сравнительно низкой. Поэтому усилия специалистов в области помехоустойчивого кодирования всегда были направлены на поиск таких кодов и методов кодирования, которые при минимальной избыточности обеспечивали бы высокую исправляющую способность.

«Облачный» код. Разместим четырехразрядное слово $\mathbf{m = (m_1,m_2,m_3,m_4)}$, которое требуется закодировать, в областях пересечения трех кругов («облаков»), а в свободных местах «облаков» поместим три дополнительных символа так, чтобы в каждом «облаке» сумма битов была четной, то есть нулевой (рис.4.2).

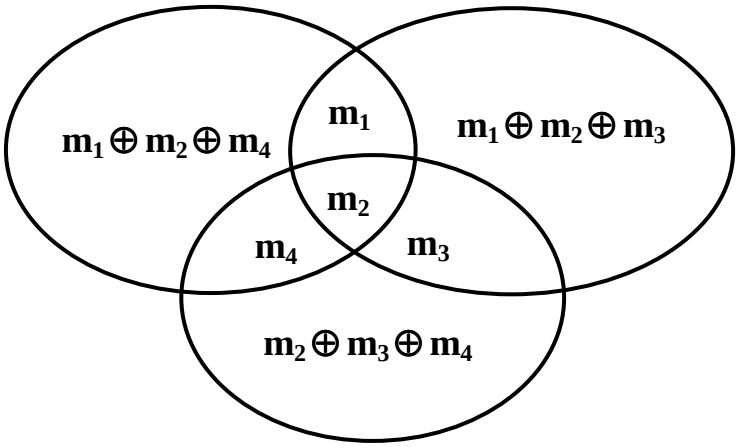


Рис. 4.2. Иллюстрация облачного кода

При искажении любого из разрядов исходной комбинации **m** четность нарушится в двух или трех «облаках». По тому, в каких именно «облаках» обнаружена нечетность, можно определить место ошибки, а, следовательно, исправить ее. Например, если после передачи семиразрядного слова будет обнаружены нечетные суммы в левом и правом верхних «облаках», это свидетельствует об ошибке в разряде $\mathbf{m_1}$. Если же нечетные суммы будут во всех трех «облаках», это говорит об ошибке в разряде $\mathbf{m_2}$. Нечетная сумма только в одном из «облаков» означает ошибку в дополнительном разряде, которую можно не исправлять. Отметим, что этот код можно определить как **(7, 4)**-код и он относится к классу оптимальных, то есть для заданного количества информационных разрядов имеет минимально возможное число проверочных битов.

Порождающие матрицы блочных кодов

Только что в качестве примера были рассмотрены простейшие корректирующие коды - код с простой проверкой на четность, позволяющий обнаруживать однократную ошибку в принятой последовательности, а также блочный итеративный код и «облачный» код, исправляющие одну ошибку с помощью набора проверок на четность. Во всех кодах в процессе помехоустойчивого кодирования формировались дополнительные разряды, присовокупляемые к исходной кодовой комбинации.

Зададим формальные (порождающие) правила, по которым осуществляется кодирование, то есть преобразование информационной последовательности в кодовое слово.

Простейшим способом описания, или задания, корректирующих кодов является **табличный способ**, при котором каждой информационной последовательности просто назначается кодовое слово из таблицы кода. Например, для простейшего кода с проверкой на четность таблица соответствия исходных и кодовых комбинаций будет следующей:

m	**u**
000	**0000**
001	**0011**
010	**0101**
011	**0110**
100	**1001**
101	**1010**
110	**1100**
111	**1111**

Такой способ описания кодов применим для любых, а не только линейных кодов. Однако, при больших **k** размер кодовой таблицы оказывается слишком большим, чтобы им пользоваться на практике.

Другим способом задания линейных блочных кодов является использование так называемой **системы порождающих уравнений**, определяющих правило, по которому символы информационной последовательности преобразуются в кодовые символы. Для того же примера система порождающих уравнений будет выглядеть следующим образом:

$$\begin{cases} u_1 = m_1, \\ u_2 = m_2, \\ u_3 = m_3, \\ u_4 = m_1 \oplus m_2 \oplus m_3. \end{cases}$$

Однако наиболее удобным и наглядным способом описания линейных блочных кодов является их задание с использованием **порождающей матрицы**, являющейся компактной формой представления системы проверочных уравнений.

> **Линейный блочный** (**n,k**)-код полностью определяется матрицей **G** размером **k×n** с двоичными матричными элементами. При этом каждое кодовое слово является линейной комбинацией строк матрицы **G**, а каждая линейная комбинация строк **G –** кодовым словом.

Линейные блочные коды, задаваемые порождающими матрицами, будем называть **матричными кодами**. Обычное (каноническое) представление порождающей матрицы выглядит так:

$$\mathbf{G} = \left[\begin{array}{cccc:ccc} 1 & 0 & \dots & 0 & P_{11} & \dots & P_{1,n-k} \\ 0 & 1 & \dots & 0 & P_{21} & \dots & P_{2,n-k} \\ \hdashline 0 & 0 & \dots & 1 & P_{k1} & \dots & P_{k,n-k} \end{array}\right]$$

Например, для простейшего **(4,3)**-кода с проверкой на четность порождающая матрица будет иметь вид:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Пусть $\mathbf{m} = (m_1, m_2, \dots, m_k)$ будет тем блоком-сообщением, который необходимо закодировать с использованием данного кода.

Тогда соответствующим ему кодовым словом **U** будет

$$\mathbf{u} = \mathbf{m} \cdot \mathbf{G}$$

С учетом структуры матрицы **G** символы кодового слова **u** будут такими:

для $\mathbf{i = 1, 2, \ldots , k}$: $\mathbf{u_i = m_i}$;

для $\mathbf{i = k+1, \ldots , n}$: $\mathbf{u_i = m_1 \cdot P_{1,i-k} \oplus m_2 \cdot P_{2,i-k} \oplus m_3 \cdot P_{3,i-k} \oplus \ldots \oplus m_k \cdot P_{k,i-k}}$.

Иными словами, **k** крайних левых символов кодового слова совпадает с символами кодируемой информационной последовательности, а остальные (**n - к**) символов являются линейными комбинациями символов информационной последовательности.

Например, если входная последовательность кодера **m** = **(1 0 1)**, то с применением порождающей матрицы код будет построен так:

$$\mathbf{u = m \cdot G = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix}}$$

Командир послал трех разведчиков по первой дороге, трех – по второй, двух – по третьей, по четвертой пошел сам. Составить порождающую матрицу такого кода.

Ответ: по сюжету задачи сообщение, полученное командиром лично, не может быть искаженным. Поэтому ограничимся исследованием информации, передаваемой бойцами. В группе, отправившейся по одной из дорог, каждый боец должен доложить командиру либо об обнаружении объекта (обозначим такой доклад «**1**»), либо об отсутствии объекта («**0**»). При отсутствии искажений доклады каждого бойца из одной и той же группы должны совпадать. Следовательно:

$$\mathbf{G = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}}$$

Матрицу можно привести к каноническому виду, перенумеровав бойцов, то есть, по первой дороге послав первого, четвертого и пятого, по второй дороге – второго, шестого и седьмого, по третьей дороге – третьего и восьмого бойцов. Получим следующую порождающую матрицу:

$$\mathbf{G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}}$$

Определенный таким образом код называется **линейным блочным систематическим (n,k)-кодом с обобщенными проверками на четность**, а задающая его матрица **G** называется **порождающей матрицей кода**.

Характеристики блочных линейных кодов

Напомним для начала, что двоичный код – это код, составленный из двоичных цифр – нуля и единицы.

Количество разрядов в каждой кодовой комбинации (блоке) называют **длиной** или **значностью** кода и обозначают **n**.

Символы каждого разряда могут принимать значения **0** или **1**.

Количество единиц в кодовой комбинации называют **весом** и обозначают **ω**.

Например, кодовая комбинация **100101100** имеет значность **n = 9** и вес **ω = 4**.

Степень отличия двух любых кодовых комбинаций характеризуется **кодовым расстоянием d** (**расстоянием Хемминга**), которое определяется как число разрядов, в которых комбинации отличаются одна от другой.

Для определения кодового расстояния надо просуммировать (по модулю **2**) две кодовые комбинации и определить вес суммы.

Пример. Определить кодовое расстояние между комбинациями **100101100** и **110110101**.

Просуммируем:

$$\begin{array}{r} 100101100 \\ \oplus\ 110110101 \\ \hline 010011001 \end{array}$$

Вес полученной суммы (количество единиц) **ω=4**, следовательно, кодовое расстояние **d=4**.

Определить все кодовые расстояния между безошибочными комбинациями сообщений разведчиков.

Ответ: расстояние между первой и второй строкой порождающей матрицы – **6**, между первой и третьей – **5**, между второй и третьей – **5**.

При передаче в кодовых комбинациях возникают ошибки типа «инверсия». Если ошибка произошла в одном разряде блока, она называется **однократной**, при ошибках в двух, трех и т.д. разрядах они называются двукратными, трехкратными и т.д. Для описания возникающих в канале ошибок используют вектор ошибки, обычно обозначаемый как **e** и представляющий собой двоичную последовательность длиной **n** с единицами в тех позициях, в которых произошли ошибки. Вес вектора ошибки равен **кратности** ошибки.

Так, вектор ошибки **e = (0 0 0 1 0 0 0)** означает однократную ошибку в четвертом разряде, вектор ошибки **e = (1 1 0 0 0 0 0)** - двукратную ошибку в первом и втором разрядах.

Допустим, по каналу связи передается кодовое слово **u,** на приемном конце канала декодером принята последовательность **û**, возможно, содержащая ошибки. Если **e** – вектор ошибок, то верны соотношения **û** = **u** ⊕ **e** , а также **e** = **u** ⊕ **û** и **u** = **û** ⊕ **e**.

Помехоустойчивость кодирования обеспечивается за счет введения избыточности. Это значит, что из **n** символов кодовой комбинации для передачи информации используется **k < n** символов.

Коэффициент избыточности кода – это отношение количества проверочных битов к длине кода. $\mathbf{F = \frac{n-k}{n}}$

При длине кодового слова **n** всего существует $\mathbf{2^n}$ кодовых слов, из них допустимыми могут быть $\mathbf{2^k}$. Соответственно, все множество кодовых комбинаций разбивается на две группы: разрешенные комбинации и запрещенные комбинации. Разрешенных комбинаций $\mathbf{S_p = 2^k}$, запрещенных $\mathbf{S_f = S - S_p = 2^n - 2^k}$**.** Все разрешенные комбинации известны как кодеру, так и декодеру, поскольку полностью определяются первичным алфавитом, а точнее, системой первичного кодирования алфавита источника. Если на приемной стороне канала передачи информации установлено, что принятая комбинация относится к разрешенным, то считается, что сообщение прошло без искажений, а если принята запрещенная комбинация, то делается вывод, что произошла ошибка. Однако, если ошибка такова, что посланная комбинация, претерпев искажения, тем не менее, попала во множество разрешенных комбинаций, такая ошибка обнаружена не будет: у декодера нет оснований считать ее ошибочной.

Пусть всего $\mathbf{S_p}$ разрешенных комбинаций. Каждая из них при передаче может трансформироваться в любую из **S** возможных комбинаций, т.е. всего имеется $\mathbf{S \cdot S_p}$ возможных вариантов передачи. Из них $\mathbf{S_p}$ вариантов безошибочной передачи, $\mathbf{S_p \cdot (S_p - 1)}$ вариантов ошибочной трансформации в другие разрешенные комбинации и $\mathbf{S_p \cdot (S - S_p)}$ вариантов трансформации в запрещенные комбинации (см. рис. 4.3).

Только передача в запрещенные варианты может быть обнаружена. **Доля обнаруживаемых ошибок** составляет

$$\frac{\mathbf{S_p(S-S_p)}}{\mathbf{S \cdot S_p}} = \mathbf{1 - \frac{S_p}{S} = 1 - \frac{2^k}{2^n} = 1 - 2^{k-n}}$$

Обнаруженную ошибку можно исправить, если для каждой запрещенной комбинации можно указать единственную исходную, то есть посланную комбинацию. Таким образом, ошибка исправляется в $\mathbf{S - S_p}$ случаях, равных количеству запрещенных комбинаций. **Доля исправляемых ошибочных комбинаций** от числа обнаруживаемых составляет:

$$\frac{S - S_p}{S_p(S - S_p)} = \frac{1}{S_p} = 2^{-k}$$

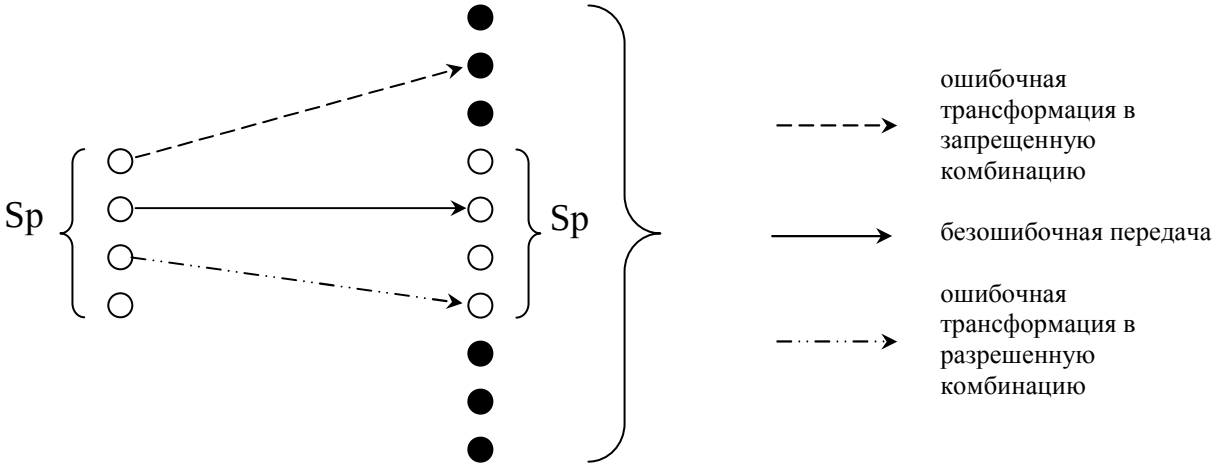


Рис. 4.3. Схема возникновения ошибок при передаче кодовых слов:
○ - разрешенные комбинации; ● - запрещенные комбинации

Эти величины характеризуют корректирующую способность кода. Однако на практике следует учитывать не только количество, но и вероятностные характеристики тех или иных ошибочных комбинаций.

Определим вероятность ошибочного приема сообщения. Пусть **p** – вероятность появления ошибки при передаче отдельного разряда кодовой комбинации. На практике эта вероятность определяется конструктивной защищенностью канала связи от внешних помех. Например, для локальных сетей, где средой передачи данных является витая пара, $\mathbf{p = 10^{-4}…10^{-6}}$. А для оптоволоконных каналов $\mathbf{p = 10^{-9}}$. Для оценки вероятности возникновения ошибки при передаче кодовой комбинации, состоящей из **n** бит, необходимо принять определенные исходные предположения, которые называются **математической моделью ошибок**. Наиболее простая из них (рассмотрим только ее) – считать появление ошибки в каждом отдельном разряде (бите) кода случайным событием, которое не зависит от того, с ошибками или без них были переданы предыдущие биты. Тогда:

1 – p – вероятность безошибочной передачи отдельного разряда;

$\mathbf{(1 - p)^n}$ – вероятность безошибочной передачи цепочки **n** разрядов;

$\mathbf{P_n = 1 - (1 - p)^n}$ – вероятность появления хотя бы одной ошибки в комбинации **n** разрядов. При малых **p** можно воспользоваться соотношением Бернулли, тогда $\mathbf{P_n \approx n \cdot p}$. Так можно оценить суммарную вероятность появления всех ошибок. Если же требуется найти вероятность ошибки с заданной кратностью **t**, то следует воспользоваться формулой биноминального распределения:

$$P_t = C_n^t \cdot p^t \cdot (1-p)^{n-t}$$

Вероятность ошибочной передачи сообщения с учетом всех ошибок кратности от **1** до **t** находится так:

$$P_{1\div t} = \sum_{i=1}^{t} C_n^i \cdot p^i \cdot (1-p)^{n-i} \qquad (4.1)$$

Если помехоустойчивый код рассчитан на парирование ошибок кратностью до **t** включительно, то вероятность исправления ошибки определяется величиной $P_{1\div t}$.

Связь между корректирующей способностью кода и кодовым расстоянием

Важная характеристика помехоустойчивого кода – **наименьшее расстояние между разрешенными кодовыми комбинациями d_{min}**. Оно обеспечивает корректирующую способность кода и влияет на кратность ошибок, которые могут быть обнаружены и исправлены.

Определить минимальное кодовое расстояние между разрешенными комбинациями сообщений разведчиков.

Ответ: $d_{min} = \min\{6, 5, 5\} = 5$.

Проиллюстрируем этот подход для **(3,2)**-кода: количество информационных разрядов **k = 2**, следовательно, число разрешенных кодовых комбинаций $S_p = 2^2 = 4$; общее число кодов $S = 2^3 = 8$. Число проверочных бит **r = n – k = 1** и устанавливать их значение условимся таким образом, чтобы количество «единиц» во всех кодовых комбинациях было бы четным (по этой причине такой проверочный бит называется **битом четности**). Для этого кода $d_{min} = 2$.

u_i	a_1	a_2	b	Σ
u_0	0	0	0	0
u_1	0	1	1	2
u_2	1	0	1	2
u_3	1	1	0	2

Будем обозначать разрешенные кодовые комбинации u_i. Их информационная часть может принимать значения **00, 01, 10** и **11** (обозначим эти биты a_1 и a_2); проверочный бит **b** принимает значения, приведенные в таблице. Последняя колонка содержит суммы (количество) бит со значением "**1**" в каждой кодовой комбинации. Любую из **8** существующих для **n = 3** кодовых комбинаций можно считать вектором в пространстве, построенном на единичных векторах a_1, a_2, **b** – это иллюстрируется рис. 4.4. Отметим разрешенные комбинации ноликами; остальные, очевидно, будут запрещенными -

их отметим крестиками. Видно, что минимальное кодовое расстояние между разрешенными комбинациями равно **2** (расстояние по ребрам куба между разрешенными вершинами). На рисунке однократной ошибке соответствует переход из вершины куба в соседнюю вершину. Любая однократная ошибка переводит разрешенную комбинацию в запрещенную, следовательно, может быть обнаружена. Однако исправить такую ошибку нельзя, поскольку в любую из запрещенных вершин за один шаг можно попасть, по крайней мере, из двух разрешенных.

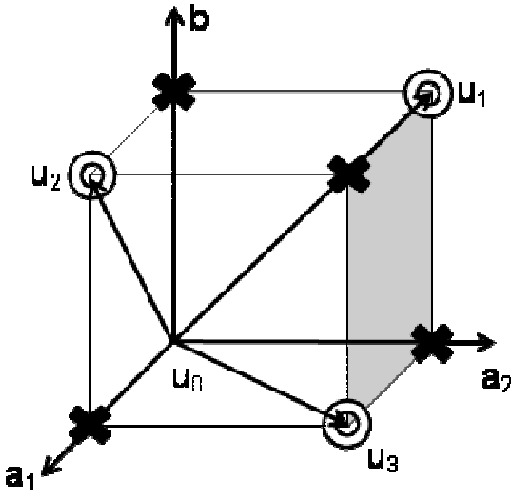


Рис. 4.4. К вопросу о связи кодового расстояния и корректирующей способности кода

Обобщим рассуждения. Пусть необходимо построить код, обнаруживающий все ошибки кратности τ и меньше. Другими словами, мы принимаем, что вес вектора ошибки будет не больше τ. Построить такой код – это значит из множества **S** возможных комбинаций выбрать S_p разрешенных комбинаций так, чтобы сумма (по модулю **2**) любой разрешенной комбинации с любым вектором ошибок с весом $\omega \le \tau$ не дала бы в результате никакой другой разрешенной комбинации. Поскольку расстояние Хемминга d_{min} – это вес суммы двух кодовых комбинаций, то для достижения цели обнаружения ошибки кратности τ необходимо, чтобы наименьшее кодовое расстояние удовлетворяло условию

$$d_{min} \ge \tau + 1 \qquad (4.2)$$

Сможет ли командир обнаружить двукратную ошибку?

Ответ: да, потому что 5 ≥ 2+1.

Рассмотрим код со значностью **n=3**. Все возможные комбинации этого кода приведены в таблице:

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
000	**001**	**010**	**011**	**100**	**101**	**110**	**111**

Составим матрицу кодовых расстояний (табл. 4.2). Для того чтобы код обеспечивал обнаружение однократных ошибок, необходимо из **8** возможных комбинаций выбрать в качестве разрешенных такие, расстояние между которыми было бы не менее **d=2**. Например, определим разрешенными комбинациями следующие: A_2 = **001**, A_3 = **010**, A_5 = **100**, A_8 = **111**. Любая однократная ошибка переводит разрешенную комбинацию в запрещенную.

Таблица 4.2. Матрица кодовых расстояний

	A_1 000	A_2 001	A_3 010	A_4 011	A_5 100	A_6 101	A_7 110	A_8 111
A_1	0	1	1	2	1	2	2	3
A_2	1	0	2	1	2	1	3	2
A_3	1	2	0	1	2	3	1	2
A_4	2	1	1	0	3	2	2	1
A_5	1	2	2	3	0	1	1	2
A_6	2	1	3	2	1	0	2	1
A_7	2	3	1	2	1	2	0	1
A_8	3	2	2	1	2	1	1	0

Для обнаружения двукратных ошибок наименьшее кодовое расстояние должно быть d_{min}**=3**. В качестве примера разрешенных комбинаций в этом случае можно выбрать A_3**=010** и A_6**=101**.

Пусть теперь необходимо построить код, обеспечивающий исправление однократных ошибок. Выберем в качестве первой разрешенной комбинации A_2**=001**. При однократной ошибке комбинация A_2 перейдет в одну из трех комбинаций: A_1**=000**, A_4**=011** или A_6**=101** (рис. 4.5). Эти комбинации объявляются запрещенными комбинациями для A_2. Это означает, что при появлении одной из этих комбинаций на выходе канала связи будет принято решение, что передана комбинация A_2.

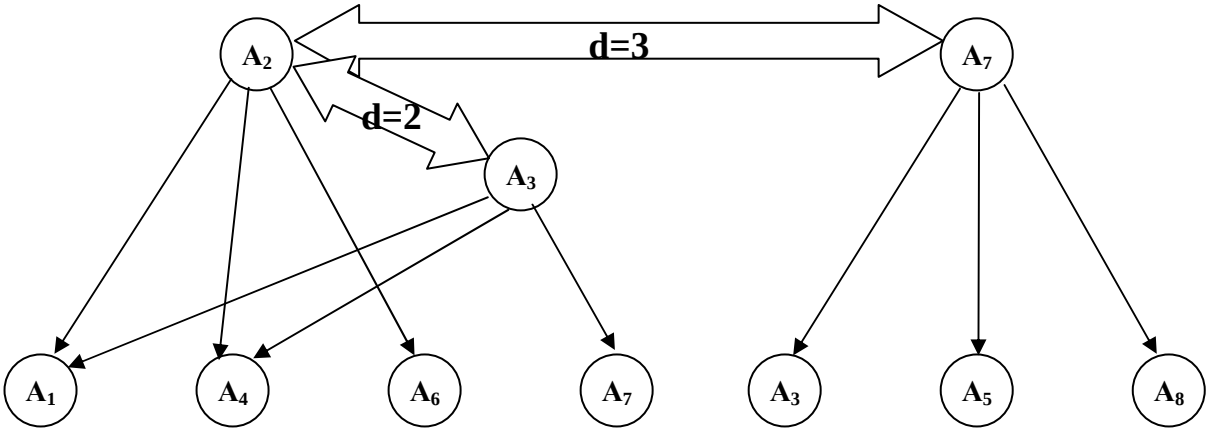


Рис. 4.5. Варианты перехода между кодовыми комбинациями при однократной ошибке

Допустим, в качестве второй разрешенной комбинации выбрана $\mathbf{A_3=010}$, отстоящая на расстоянии **d=2**. При возникновении однократной ошибки ей соответствует подмножество запрещенных комбинаций $\mathbf{A_4=011}$, $\mathbf{A_1=000}$ и $\mathbf{A_7=110}$. Однако получилось, что пересечение подмножеств запрещенных комбинаций не пусто. Следовательно, при приеме запрещенных комбинаций $\mathbf{A_4}$ или $\mathbf{A_1}$ нельзя однозначно установить, какая разрешенная комбинация была послана – $\mathbf{A_2}$ или $\mathbf{A_3}$. Если же в качестве второй разрешенной комбинации принять $\mathbf{A_7=110}$, которой соответствуют запрещенные комбинации $\mathbf{A_8=111}$, $\mathbf{A_5=100}$ и $\mathbf{A_3=010}$, то в этом случае подмножества запрещенных комбинаций не пересекаются и имеется однозначное соответствие принятой и переданной комбинации.

Следовательно, при $\mathbf{d_{min}=3}$ обеспечивается исправление всех однократных ошибок.

В общем случае, для исправления ошибок кратности **t** минимальное кодовое расстояние должно удовлетворять условию

$$\mathbf{d_{min} \geq 2t + 1} \qquad (4.3)$$

Сможет ли командир исправить двукратную ошибку?

Ответ: да, потому что $5 \geq 2 \cdot 2+1$

Аналогично рассуждая, можно установить, что для исправления всех ошибок кратности не более **t** и одновременного обнаружения всех ошибок кратности не более $\boldsymbol{\tau}$ (при $\boldsymbol{\tau \geq t}$) кодовое расстояние должно удовлетворять условию

$$\mathbf{d_{min} \geq t + \tau + 1} \qquad (4.4)$$

Связь между корректирующей способностью кода и длиной кода

Обычная последовательность выбора кода следующая:

- исходя из мощности **M** первичного алфавита, определяется количество информационных разрядов **k**;
- задается возможная кратность ошибок **t**, подлежащих обнаружению и исправлению;
- определяется количество дополнительных проверочных разрядов **r**, которые вместе с информационными определят длину кода **n**.

Пусть известна мощность первичного алфавита **M**. Необходимое количество информационных битов $\mathbf{k \geq \log_2 M}$.

Пусть необходимо исправить ошибки кратности от **1** до **t**. Число возможных однократных ошибок в коде длиной **n** равно $\mathbf{C_n^1 = n}$, число двукратных ошибок равно $\mathbf{C_n^2 = n(n-1)/2}$, число возможных **t**-кратных ошибок равно $\mathbf{C_n^t = \frac{n!}{t!(n-t)!}}$.

Общее число ошибок $\mathbf{E = \sum_{i=1}^{t} C_n^i}$

Эти **E** ошибок могут проявиться в каждой из $\mathbf{2^k}$ возможных входных последовательностей. Полное число ошибочных комбинаций, подлежащих исправлению, равно $\mathbf{E \cdot 2^k}$. Код длиной **n** обеспечивает исправление не более $\mathbf{2^n - 2^k}$ комбинаций, так как существует именно столько запрещенных комбинаций. Следовательно, необходимое условие для возможности исправления ошибок можно записать в виде $\mathbf{E \cdot 2^k \leq 2^n - 2^k}$

Отсюда получим: $\mathbf{\frac{2^n}{2^k} \geq 1 + E}$

Обозначив буквой **r** число проверочных символов, и учитывая, что **r = n – k**, приходим к виду:

$$\mathbf{2^r \geq 1 + \sum_{i=1}^{t} C_n^i}$$

Учитывая, что $\mathbf{1 = C_n^0}$, можно записать $\mathbf{2^r \geq \sum_{i=0}^{t} C_n^i}$ или

$$\mathbf{r \geq \log_2 \sum_{i=0}^{t} C_n^i} \qquad (4.5)$$

Это так называемая **граница Р. Хемминга** (или **условие Хемминга**), связывает число проверочных разрядов **r** и значность кода **n**.

Richard Wesley Hamming (1915–1998) was a mathematician whose had work many implications for computer science and telecommunications. His contributions include the Hamming code (which makes use of a Hamming matrix), the Hamming window, Hamming numbers, Sphere-packing (or hamming bound) and the Hamming distance. He was born in Chicago, Illinois and died in Monterey, California. He received his bachelor's degree from the University of Chicago in 1937, a master's degree from the University of Nebraska in 1939, and finally a Ph.D. from the University of Illinois at Urbana-Champaign in 1942. He was a professor at the University of Louisville during World War II, and left to work on the Manhattan Project in 1945, programming one of the earliest electronic digital computers to calculate the solution to equations provided by the project's physicists. Later 1946-1976 he worked at the Bell Telephone Laboratories, where he collaborated with Claude E. Shannon. In 1976 he moved to the Naval Postgraduate School, where he worked as an Adjunct Professor until 1997, when he became Professor Emeritus. He was a founder and president of the Association for Computing Machinery.

В частном случае, когда требуется исправить однократные ошибки, имеем зависимость $\mathbf{2^r - r - 1 \geq k.}$

Оценить количество проверочных символов и избыточность кода можно из таблицы 4.3, построенной по указанной зависимости (4.5).

Таблица 4.3. Соотношение контрольных и информационных символов для кода, исправляющего однократные ошибки

r	2	3	4	5	6	7	8	9	10
k	1	4	11	26	57	120	247	502	1013
F	0,67	0,43	0,27	0,16	0,10	0,06	0,03	0,02	0,01

Вспомнив вторую теорему Шеннона, можно убедиться, что действительно, увеличивая длину блока, можно бесконечно уменьшать избыточность (**F**), обеспечивая вместе с тем помехоустойчивость кода.

Сравните с аналогичной таблицей 4.4 для помехоустойчивого кода, исправляющего двукратные ошибки.

Таблица 4.4. Соотношение контрольных и информационных символов для кода, исправляющего двукратные ошибки

r	2	3	4	5	6	7	8	9	10	11	12	13
k	1	1	2	3	5	9	15	23	35	53	79	115
F	0,67	0,75	0,67	0,63	0,55	0,44	0,35	0,28	0,22	0,17	0,13	0,10

С увеличением длины блока избыточность уменьшается, хотя и не так активно, как для кода, исправляющего однократные ошибки.

Выполняется ли неравенство Хемминга в задаче о разведчиках?

Ответ: да, но только для двукратных ошибок. Если же требуется исправлять и однократные, и двукратные ошибки, то $\mathbf{r \geq \log_2(1+8+28)=6}$.

4.2. Матричные коды

Матричные коды относятся к группе блочных разделимых кодов. Для матричного кода сумма по модулю **2** двух разрешенных комбинаций также дает разрешенную комбинацию (**свойство замкнутости**). Все разрешенные комбинации матричного (**n,k**)-кода можно получить, располагая **k**-значными исходными комбинациями. При этом:

- в число исходных комбинаций не должна входить тривиальная (нулевая);
- все кодовые комбинации должны быть линейно независимы, т.е. ни одна из них не может быть получена путем суммирования других;
- для обеспечения требуемой корректирующей способности минимальное кодовое расстояние $\mathbf{d_{min}}$ кодовых комбинаций должно удовлетворять условиям Хемминга.

Получение кодовых комбинаций производится с помощью порождающих матриц, состоящих из **k** строк и **n** столбцов:

$$\mathbf{G}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} & b_{11} & b_{12} & \dots & b_{1r} \\ a_{21} & a_{22} & \dots & a_{2k} & b_{21} & b_{22} & \dots & b_{2r} \\ a_{k1} & a_{k2} & \dots & a_{kk} & b_{k1} & b_{k2} & \dots & b_{kr} \end{bmatrix}$$

В классической (канонической) форме кода элементы первых **k** столбцов служат для информационных целей, а оставшихся – для проверочных. Соответственно, порождающую матрицу **G** можно представить в виде двух подматриц – информационной $\mathbf{I_k}$ и проверочной **P**. $\mathbf{G}=[\mathbf{I_k}\|\mathbf{P}]$, где

$$\mathbf{I_k}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix},\ \mathbf{P}=\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1r} \\ b_{21} & b_{22} & \dots & b_{2r} \\ b_{k1} & b_{k2} & \dots & b_{kr} \end{bmatrix}$$

Информационную подматрицу часто берут в виде квадратной единичной матрицы:

$$\mathbf{I_k}=\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

При этом проверочная подматрица **P** должна строиться с соблюдением следующих условий:

- все строки должны быть различны,
- кодовое расстояние между любыми двумя строками подматрицы должно быть не менее $\mathbf{d_{min}-2}$,
- строки должны быть линейно независимыми, то есть ни одна из них не должна являться суммой других.

Пусть, например, порождающая матрица (7,4)-кода имеет вид:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

и на вход вторичного (помехоустойчивого) кодера поступила исходная комбинация **m = [0 0 1 1].** Тогда кодовое слово образуется умножением исходной комбинации на порождающую матрицу.

$$\mathbf{u} = \mathbf{m}\cdot\mathbf{G} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Обнаружение ошибок с помощью матричных кодов

Имея порождающую матрицу кода **G**, строят проверочную матрицу **H**, посредством которой можно обнаруживать и, по возможности, исправлять ошибки. Проверочная матрица должна быть ортогональна любой разрешенной комбинации кода, то есть произведение проверочной матрицы на любую разрешенную кодовую комбинацию должно давать нулевой вектор. Порядок построения проверочной матрицы следующий.

1) В порождающей матрице **G** выделяют информационную квадратную подматрицу $\mathbf{I_k}$ и проверочную подматрицу **P**.
2) Проверочную подматрицу транспонируют, получая подматрицу $\mathbf{P^T}$.

$$\mathbf{P^T} = \begin{bmatrix} \mathbf{b_{11}} & \mathbf{b_{21}} & \mathbf{\ldots} & \mathbf{b_{k1}} \\ \mathbf{b_{12}} & \mathbf{b_{22}} & \mathbf{\ldots} & \mathbf{b_{k2}} \\ \hdashline \mathbf{b_{1r}} & \mathbf{b_{2r}} & \mathbf{\ldots} & \mathbf{b_{kr}} \end{bmatrix}$$

3) Справа к подматрице $\mathbf{P^T}$ приписывают квадратную единичную матрицу $\mathbf{I_r}$ размера **r×r**.

Получается матрица $\mathbf{H} = \left[\mathbf{P^T} \middle\| \mathbf{I_r}\right]$, которую используют для обнаружения ошибок путем проверки ее ортогональности полученному кодовому слову **û**. Вычисляют вектор **s**, называемый **синдромом ошибки**: **s = H·û**. Если синдром ошибки равен нулю, то комбинация передана безошибочно, в противном случае ошибка существует, ее даже можно исправить. Можно показать, что синдром ошибки однозначно определяется вектором ошибки. Действительно: **s = H·û**, но в свою очередь **û = u ⊕ e**, значит **s = H·(u ⊕ e) = H·u ⊕ H·e**. Матрица **H** составляется так, чтобы выполнялось условие ортогональности **H·u = 0**, следовательно, **s = H·e**.

Возьмем порождающую матрицу (7.4)-кода из предыдущего примера и составим проверочную матрицу. Проверочная подматрица **P** имеет вид:

$$\mathbf{P}=\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}\\ \mathbf{1}&\mathbf{1}&\mathbf{0}\\ \mathbf{1}&\mathbf{1}&\mathbf{1}\end{bmatrix}.$$ Транспонируем: $$\mathbf{P}^{\mathbf{T}}=\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}\end{bmatrix}.$$

Составим проверочную матрицу, приписав справа единичную матрицу 3×3:

$$\mathbf{H}=\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}.$$

Проверим полученное в предыдущем примере кодовое слово **u = [0 0 1 1 0 0 1]**.

$$\mathbf{s}=\mathbf{H}\cdot\mathbf{u}=\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}=\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

Синдром равен **0**, следовательно, ошибки нет. Комбинацию можно декодировать. Это просто. Первые четыре разряда принятого кодового слова и представляют собой исходную комбинацию. Последние три разряда можно отбросить.

Допустим, произошла ошибка во втором разряде и была принята ошибочная комбинация **û = [0 1 1 1 0 0 1]**. Вычислим синдром

$$\mathbf{s}=\mathbf{H}\cdot\hat{\mathbf{u}}=\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}=\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\\ \mathbf{1}\end{bmatrix}$$

Синдром ненулевой, следовательно, обнаружена ошибка. Можно ли эту ошибку не только обнаружить, но и исправить? Обратим внимание, что всего возможно **7** различных однократных ошибок (по числу разрядов). Синдром трехзначный, значит, существует **7** ненулевых синдромов. То есть, каждой ошибке можно сопоставить свой синдром. Следовательно, по виду синдрома можно локализовать, а значит, исправить ошибку. Синдром совпадает с тем столбцом проверочной матрицы, номер которого соответствует позиции ошибки (правило верно только для однократных ошибок!).

Коды Хемминга

В отличие от канонического матричного кода, в коде Хемминга информационные и проверочные биты не разнесены в отдельные подматрицы,

а чередуются. Если биты кодовой комбинации пронумеровать, начиная с **1**, слева направо, то контрольными (проверочными) оказываются биты с номерами **1**, **2**, **4**, **8** и т.д., все остальные являются информационными. Цель этих перестановок – сделать так, чтобы синдром ошибки непосредственно указывал на локализацию ошибок, минуя промежуточную таблицу соответствия синдромов и ошибок или исключая необходимость сравнения синдрома ошибки со столбцами проверочной матрицы. Код Хемминга начинают строить с проверочной матрицы **H**, так как ее вид очевиден: столбцы проверочной матрицы представляют собой набор синдромов, соответствующих двоичному представлению номера столбца. Затем строят порождающую матрицу **G**, исходя из того, что матрицы **H** и **G** ортогональны, т.е. скалярное произведение каждой строки матрицы **G** на каждую строку матрицы **H** равно нулю, то есть

$$\mathbf{H}\cdot\mathbf{G}^{\mathbf{T}}=\mathbf{0} \text{ и } \mathbf{G}\cdot\mathbf{H}^{\mathbf{T}}=\mathbf{0}\ .$$

Построим код Хемминга для (7,4)-кода. Чтобы синдром ошибки указывал на место ошибки, проверочная матрица должна иметь вид:

$$\mathbf{H}=\begin{bmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}&\mathbf{1}\\ \mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{1}\end{bmatrix}$$

Обратите внимание, каждый столбец матрицы **H** является двоичным представлением номера этого столбца.

В матрице **H** на **1**-м, **2**-м и **4**-м местах стоят столбцы матрицы, имеющей единицы на второстепенной диагонали, а на остальных – столбцы, соответствующие информационным разрядам кода. Выделим подматрицу, соответствующую информационным разрядам (то есть **3**-й, **5**-й, **6**-й и **7**-й столбцы):

$$\mathbf{H}_{\mathbf{I}}=\begin{bmatrix}\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{1}\end{bmatrix}$$

Повернем матрицу $\mathbf{H}_{\mathbf{I}}$ по часовой стрелке и получим проверочную подматрицу порождающей матрицы:

$$\mathbf{P}=\begin{bmatrix}\mathbf{1}&\mathbf{1}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{1}\\ \mathbf{0}&\mathbf{1}&\mathbf{1}\\ \mathbf{1}&\mathbf{1}&\mathbf{1}\end{bmatrix}$$

Поставим столбцы матрицы P на **1**, **2** и **4**-е места, а остальные столбцы будут столбцами единичной матрицы. Получим порождающую матрицу кода Хемминга:

$$\mathbf{G_X} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Закодируем по Хеммингу входную комбинацию **m = [0 0 1 1]**.

$$\mathbf{u} = \mathbf{m} \cdot \mathbf{G_X} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Проверим разрешенное кодовое слово, вычислив синдром:

$$\mathbf{s} = \mathbf{H} \cdot \mathbf{u} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Синдром нулевой, следовательно, получена разрешенная кодовая комбинация.

Предположим, что произошла ошибка в третьем разряде и принята кодовая комбинация **û=[1 0 1 0 0 1 1].** Вычислим синдром:

$$\mathbf{s} = \mathbf{H} \cdot \hat{\mathbf{u}} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Синдром представляет собой двоичное число **3**, что локализует ошибку в третьем разряде.

Построить блочный систематический код для исправления двукратных ошибок

Решение:

Порождающая матрица **(8, 3)**-кода уже была получена:

$$\mathbf{G} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Представим ее в каноническом виде, переставив столбцы:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Выделим проверочную подматрицу:

$$\mathbf{P}=\begin{bmatrix}\mathbf{1}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{1}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}$$ и транспонируем ее: $$\mathbf{P}^{\mathrm{T}}=\begin{bmatrix}\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}$$

Составим проверочную матрицу (8,3)-кода:

$$\mathbf{H}=\begin{bmatrix}\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}.$$

Код построен.

Как можно использовать полученный код для решения задачи о разведчиках?

Командир должен пронумеровать своих бойцов от **1** до **8**. В соответствии с порождающей матрицей **G** бойцы № **1, 4, 5** должны отправиться по 1-й дороге, номера **2, 6, 7** – по второй дороге, а разведчики с номерами **3** и **8** – по третьей дороге. Проверка результатов разведки может осуществляться следующим образом.

Первый случай. Допустим, объект находится на 1-й дороге, а шпионами являются бойцы № **1** и **5**. Обнаружив объект, они солгут, и командир получит следующее сообщение: **Û=(0001000)**. Умножив сообщение на проверочную матрицу, командир получит ненулевой синдром и обнаружит ложь:

$$\begin{bmatrix}\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{1}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{1}\end{bmatrix}\cdot[\mathbf{0\ 0\ 0\ 1\ 0\ 0\ 0\ 0}]=\begin{bmatrix}\mathbf{1}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}$$

Второй случай. Допустим, что объект находится на 1-й дороге, а шпионы – это бойцы № **2** и **8**. С учетом искажений во 2-м и 8-м разрядах, командир получит следующее сообщение: **Û=(11011001)**. Проверка вновь осуществляется перемножением проверочной матрицы на сообщение:

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \cdot [1\ 1\ 0\ 1\ 1\ 0\ 0\ 1] = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Ненулевой синдром свидетельствует об ошибке.

Третий случай. Объект находится на 1-й дороге и в докладе разведгрупп нет искажений (шпионы воздержались от лжи). Тогда сообщение будет выглядеть так: **Û=(10011000)**. Умножение проверочной матрицы на эту кодовую комбинацию даст нулевой синдром, свидетельствующий об отсутствии искажений.

Можно ли не только обнаружить объект, но и установить, кто из бойцов является шпионами?

Ответ: Синдром полученного кода – **5**-разрядный, следовательно, имеется $2^5-1=31$ различный ненулевой синдром. В тоже время, возможны $8 \cdot 7/2=28$ различных двукратных ошибок. Следовательно, каждой из двукратных ошибок можно однозначно сопоставить синдром. Таблицу соответствий ошибки и синдрома можно построить следующим образом. Берем вектор ошибки, например, **(11000000)**, умножаем на него проверочную матрицу и получаем синдром **S=(11110)**. Следующий вектор ошибки - **(10100000)**, умножаем на него проверочную матрицу и получаем синдром **S=(11001)**. И так далее.

Однако заметим, что если наряду с двукратными ошибками учитывать однократные (то есть один из шпионов солжет, а второй скажет правду), то всего возможны **28+8=36** различных ошибок. Это превышает количество возможных синдромов и, следовательно, не все такие ошибки могут быть локализованы.

4.3. Циклические коды

Название кодов произошло от их свойства, заключающегося в том, что каждая кодовая комбинация может быть получена путем циклической перестановки символов комбинации, принадлежащей этому же коду. Это значит, что если, например, комбинация $\mathbf{a_0\ a_1\ a_2 \ldots a_{n-1}}$ является разрешенной, то комбинация $\mathbf{a_{n-1}\ a_0\ a_1\ a_2 \ldots a_{n-2}}$ также является разрешенной. Циклические коды обычно имеют полиномиальное представление, означающее, что бинарные элементы кодового слова $\mathbf{a_{n-1}\ a_{n-2} \ldots a_2\ a_1\ a_0}$ интерпретируются как коэффициенты полинома от некоторой фиктивной переменной **x**:

$$\mathbf{a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \ldots + a_2x^2 + a_1x + a_0.}$$

Например, кодовое слово **11010** представляется в виде полинома $\mathbf{x^4+x^3+x.}$

Наибольшая степень **x** в слагаемом с ненулевым коэффициентом называется **степенью полинома**. Таким образом, в полиноме степени **n-1** старший коэффициент $\mathbf{a_{n-1}}$ равен **1**.

Действия над кодовыми комбинациями сводятся к действиям над полиномами. Причем операция сложения производится по модулю **2**. Множество таких полиномов и действий над ними образуют так называемое поле Галуа порядка **2** (обозначается **GF(2)**).

Циклический сдвиг коэффициентов полинома степени **n-1** можно выполнить, умножая полином на **x** с одновременным сложением с двучленом $\mathbf{x^n+1}$. Действительно, пусть $\mathbf{f(x)=1\cdot x^{n-1}+a_{n-2}x^{n-2}+\ldots+a_2x^2+a_1x+a_0}$. Коэффициенты полинома **f(x)** запишем в виде кортежа $\{\mathbf{1, a_{n-2}, \ldots, a_2, a_1, a_0}\}$.

Выполним над полиномом операцию умножения на **x** и сложения с двучленом $\mathbf{x^n+1}$. Тогда $\mathbf{f(x)\cdot x + (x^n+1) = x^n+a_{n-2}x^{n-1}+ \ldots +a_2x^3+a_1x^2+a_0x + x^n+1 =}$

$\mathbf{= a_{n-2}x^{n-1}+ \ldots +a_2x^3+a_1x^2+a_0x +1}$

Кортеж коэффициентов нового полинома выглядит так: $\{\mathbf{a_{n-2}, \ldots, a_2, a_1, a_0, 1}\}$. Таким образом, мы убедились, что указанная операция эквивалентна циклическому сдвигу коэффициентов.

Следовательно, если в качестве исходного взять некоторый полином **g(x)**, то процесс получения разрешенных кодовых комбинаций можно представить в следующем виде:

$\mathbf{u_1(x) = g(x)}$

$\mathbf{u_2(x) = g(x)\cdot x + (x^n+1)}$

$\mathbf{u_3(x) = u_2(x)\cdot x + (x^n+1) = g(x)\cdot x^2 + (x^n+1)(x+1)}$

…

$\mathbf{u_n(x) = g(x)\cdot x^{n-1} + (x^n+1)(x^{n-2}+x^{n-3}+\ldots+1)}$

При таком способе построения полином **g(x)** называется **порождающим** (иногда используют термин **производящий** полином). Если потребовать, чтобы порождающий полином **g(x)** был делителем двучлена $\mathbf{(x^n+1)}$, то все разрешенные комбинации приобретают свойство делимости на **g(x)**. Из этого следует, что можно легко проверить, является ли комбинация разрешенной, для этого достаточно проверить ее делимость на полином **g(x)**.

Основные свойства циклических кодов:

1. В циклическом **(n,k)-**коде каждый кодовый полином должен иметь степень не более **n-1**.

2. Для каждого циклического кода существует собственный единственный полином **g(x)** степени **(n-k)** вида

$\mathbf{g(x)=x^{n-k}+g_{n-k-1}x^{n-k-1}+\ldots+g_2x^2+g_1x+1}$,

называемый порождающим полиномом кода.

3. Каждый разрешенный кодовый полином **u(x)** является кратным **g(x)**, то есть **u(x) = q(x)·g(x)**.

4. Порождающий полином **g(x)** является делителем двучлена **(x^n+1)**.

Кодирование с использованием циклических кодов

Если дана **k**-значная входная комбинация **m(x)**, то **n**-значное кодовое слово **u(x)** циклического кода с порождающим полиномом **g(x)** можно получить двумя способами.

Первый, исходная **k**-значная комбинация, выраженная в виде полинома

$m(x) = m_{k-1}\cdot x^{k-1}+\ldots+m_1\cdot x+m_0$ степени **(k – 1)**, умножается на порождающий полином **g(x)** степени **(n – k)**:

u(x) = m(x)·g(x).

Полученный таким способом код теряет свойство систематичности. То есть невозможно указать, где информационные символы, а где проверочные. Декодирование такого кода представляет определенные трудности.

Второй, используется процедура деления полиномов и вычисления остатков. В соответствии с правилами деления полиномов для каждой пары полиномов **C(x)** и **g(x)** (причем **g(x) ≠ 0**) существует единственная пара полиномов **q(x) —** частное и **ρ(x) —** остаток, такие, что

$$C(x) = q(x)\cdot g(x) \oplus \rho(x),$$

Домножим исходный полином **m(x)** на **x^{n-k}**. Получившийся полином будет иметь степень **(k – 1) + (n – k) = (n-1)**. Представим его в виде

$$m(x)\cdot x^{n-k} = q(x)\cdot g(x) \oplus \rho(x),$$

где **q(x)** – частное от деления **$m(x)\cdot x^{n-k}$** на порождающий полином **g(x)**, а **ρ(x)** – остаток от деления. Поскольку степень **g(x)** равна **(n-k),** то степень **ρ(x)** должна быть **(n-k-1)** или меньше, а сам полином **ρ(x)** будет иметь вид

$$\rho(x)= \rho_{n-k-1}\cdot x^{n-k-1} +\ldots+ \rho_2\cdot x^2+ \rho_1\cdot x +\rho_0.$$

С учетом правил арифметики в GF(2) выражение **$m(x)\cdot x^{n-k}$ = q(x)·g(x) ⊕ ρ(x)** можно переписать следующим образом:

$$\rho(x) \oplus x^{n-k} \cdot m(x) = q(x)\cdot g(x),$$

откуда видно, что полином **$\rho(x)\oplus x^{n-k}\cdot m(x)$** является кратным **g(x)** и имеет степень **n-1** или меньшую. Следовательно, он соответствует свойствам кодовых комбинаций циклических кодов и представляет собой кодовое слово кодируемой информационной последовательности **m(x).**

Раскрыв последнее выражение, получим

$$\rho(x)\oplus m(x)\cdot x^{n-k} = m_{k-1}\, x^{n-1}+\ldots+ m_1\cdot x^{n-k+1} + m_0\, x^{n-k} +\rho_{n-k-1}\, x^{n-k-1} +\ldots+\rho_1\, x +\rho_0 ,$$

что соответствует кодовому слову $\mathbf{u = (m_{k-1} \dots m_1 m_0\ \rho_{n-k-1} \rho_1 \dots \rho_0)}$.

Таким образом, кодовое слово циклического кода состоит из неизменной информационной части (**k** разрядов) и проверочных символов (**n-k** разрядов). Проверочные символы являются коэффициентами полинома **ρ(x),** то есть остатка от деления $\mathbf{m(x)\cdot x^{n-k}}$ на порождающий полином **g(x).**

С использованием кода, задаваемого порождающим полиномом $\mathbf{g(x) = x^3+x+1}$, закодируем произвольную последовательность, например **m=(1001).**

Последовательности **m =(1001)** соответствует полином $\mathbf{m(x)= x^3+ 1}$.

Умножим **m(x)** на $\mathbf{x^{n-k}}$:

$\mathbf{m(x)\cdot x^{n-k} = m(x)\cdot x^3 = (x^3+ 1)\cdot x^3 = x^6 + x^3.}$

Разделим $\mathbf{m(x)\cdot x^{n-k}}$ на порождающий полином **g(x)**:

$$
\begin{array}{l|l}
\_x^6+ \quad x^3 & x^3 + x + 1 = g(x) \\
\underline{\ x^6+ \ x^4+ \ x^3\ \ \ } & \underline{\ x^3+ x\ \ \ \ \ \ \ \ \ \ } \\
\quad \_x^4 & \\
\quad \underline{\ x^4+ \ x^2+ x\ } & \\
\quad\quad x^2+ x & =\rho(x)
\end{array}
$$

Остаток от деления **ρ(x)** будет равен $\mathbf{x^2+ x}$.

Следует иметь ввиду, что в GF(2)-арифметике операция вычитания совпадает с операцией сложения по модулю 2.

Таким образом, кодовый полином, соответствующий информационной последовательности **m = (1001),** будет иметь следующий вид:

$\mathbf{u(x) = 1\cdot x^6+ 0\cdot x^5+ 0\cdot x^4 + 1\cdot x^3 + 1\cdot x^2+ 1\cdot x^1 + 0\cdot x^0}$,

а соответствующее кодовое слово **u = (1001110).**

Циклическое кодирование с помощью сдвиговых схем

Алгоритм кодирования, основанный на делении полиномов, можно реализовать аппаратно, используя так называемую схему деления. Она представляет собой цепочку регистров сдвига, в которой цепи обратной связи замкнуты либо разомкнуты в соответствии с коэффициентами порождающего полинома **g(x)** (см. рис. 4.6).

Кодирование в схеме выполняется следующим образом:

- в течение **k** тактов символы информационной последовательности **m** через переключатель, находящийся в верхнем (правом) положении, один за другим передаются в канал (выходной регистр) и одновременно с этим записываются в регистры проверочных символов, в которых благодаря наличию цепей обратной связи $\mathbf{g_0 \dots g_{n-k-1}}$ формируется остаток от деления

полинома $\mathbf{m(x)} \cdot \mathbf{x^{n-k}}$ на порождающий полином **g(x),** то есть проверочные символы;

- начиная с **(k+1)**-го такта переключатель переводится в нижнее (левое) положение, и из сдвигового регистра выводятся **(n-k)** проверочных символов; цепь обратной связи при этом разомкнута.

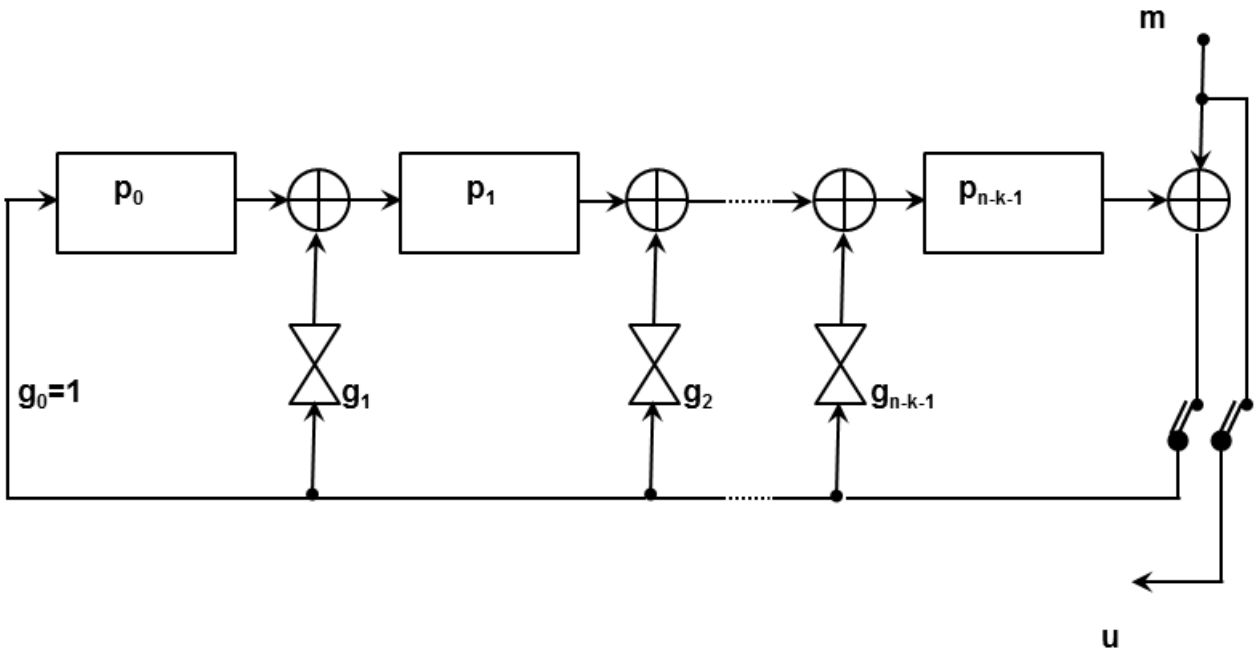


Рис. 4.6. Сдвиговая схема кодирования циклических кодов для порождающих полиномов вида $\mathbf{g(x)=1+g_1x+g_2x^2+\ldots+g_{n-k-1}x^{n-k-1}+x^{n-k}}$

Для циклического **(7,4)**-кода, используемого в качестве примера и имеющего порождающий полином $\mathbf{g(x) = x^3+ x +1}$, схема кодирования показана на рис. 4.7.

В этой схеме, в отличие от обобщенной схемы кодера, отсутствуют вентильные элементы в цепях, где значения коэффициентов обратной связи $\mathbf{g_i}$ равны нулю, там же, где коэффициенты передачи $\mathbf{g_i}$ равны единице, цепь просто замкнута.

Например, закодируем с помощью сдвиговой схемы, изображенной на рис. 4.7, произвольную последовательность **m.** Пусть **m = (1001).** Тогда выходной код будет следующий: **(1001110)**.

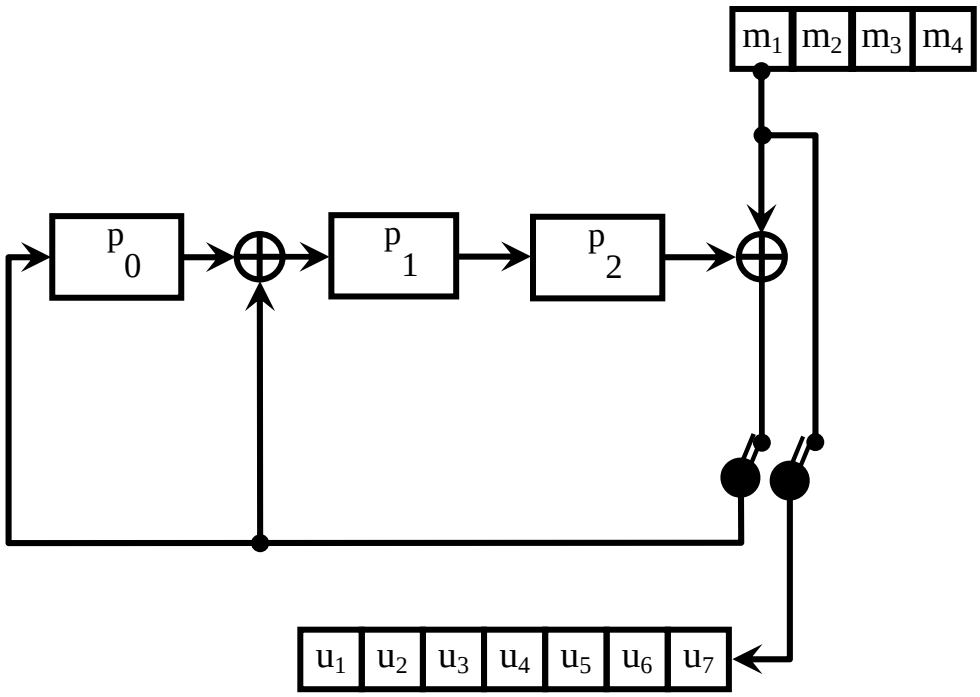


Рис. 4.7. Сдвиговая схема кодирования циклических кодов для порождающего полинома $g(x) = x^3 + x + 1$

Вычисление синдрома и исправление ошибок в циклических кодах

Вычисление синдрома для циклических кодов является довольно простой процедурой, совершенно аналогичной процедуре кодирования. Рассмотрим ее.

Пусть **u(x)** и **û(x)** - полиномы, соответствующие переданному кодовому слову и принятой последовательности. Они могут совпадать, если передача сообщения была осуществлена безошибочно, или не совпадать в противном случае.

Разделив **û(x)** на порождающий полином **g(x),** получим

$$\hat{u}(x) = q(x) \cdot g(x) \oplus s(x),$$

где - **q(x)** – частное от деления, **s(x)** – остаток от деления.

Если **û(x)** является допустимым кодовым полиномом, то он делится на **g(x)** без остатка, то есть $s(x) = 0$.

Следовательно, результат $s(x) \neq 0$ является условием наличия ошибки в принятой последовательности, а полином **s(x)** имеет смысл синдрома ошибки в последовательности **û(x)**.

Синдром **s(x)** имеет в общем случае вид

$$s(x) = s_{n-k-1} \cdot x^{n-k-1} + s_{n-k-2} \cdot x^{n-k-2} + \ldots + s_1 \cdot x + s_0 .$$

Поскольку синдром вычисляется как остаток от деления некоего полинома на порождающий полином, то очевидно, что схема вычисления

синдрома подобна схемам кодирования, рассмотренным выше, с той лишь разницей, что в качестве входной последовательности в нее подается вектор **û** (см. рис.4.8).

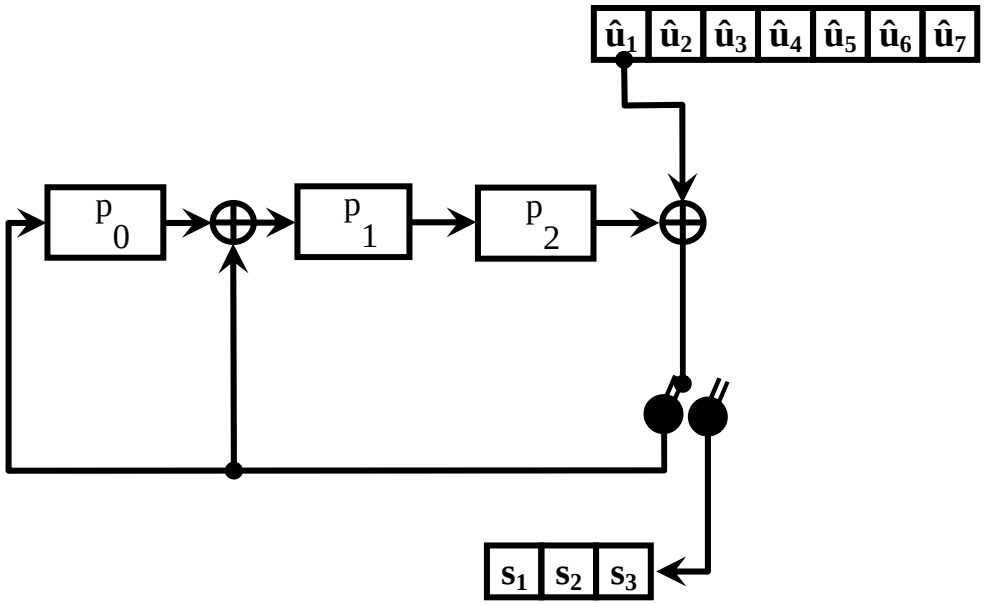


Рис. 4.8. Схема получения синдрома для порождающего полинома $g(x)=x^3+x+1$

При наличии в принятой последовательности **û** хотя бы одной ошибки вектор синдрома **s** будет иметь, по крайней мере, один ненулевой элемент, при этом факт наличия ошибки легко обнаружить, объединив по элементу **ИЛИ** выходы всех ячеек регистра синдрома.

Покажем, что синдромный многочлен **s(x)** однозначно связан с многочленом ошибки **e(x),** а значит, с его помощью можно не только обнаруживать, но и локализовать ошибку в принятой последовательности.

Пусть **e(x)** — полином вектора ошибки.

Тогда полином принятой последовательности

$$\hat{u}(x) = u(x) \oplus e(x).$$

Учтем, что **u(x)** – допустимый кодовый полином, нацело делящийся на порождающий полином **g(x)** и запишем два выражения:

$$\hat{u}(x) = q_1(x)\cdot g(x) \oplus s(x),\ u(x) = q_2(x)\cdot g(x).$$

По определению вектора (полинома) ошибки:

$$e(x) = u(x) \oplus \hat{u}(x) = q_1(x)\cdot g(x) \oplus s(x) \oplus q_2(x)\cdot g(x) =$$

$$= [q_1(x) \oplus q_2(x)]\cdot g(x) \oplus s(x) = q_3(x)\cdot g(x) \oplus s(x).$$

То есть

синдромный полином **s(x)** есть остаток от деления полинома ошибки **e(x)** на порождающий полином **g(x)**.

Отсюда следует, что по синдрому **s(x)** можно однозначно определить вектор ошибки **e(x),** а, следовательно, и исправить эту ошибку.

Получение порождающих полиномов для циклических кодов

Синдромный полином есть остаток от деления полинома ошибки **e(x)** на порождающий полином **g(x)**. Для исправления ошибок необходимо выбрать такой порождающий полином, у которого количество различных остатков не меньше числа возможных ошибок. Только в этом случае все ошибки можно локализовать. Очевидно, что корректирующая способность кода будет тем выше, чем больше различных остатков может быть образовано при делении кодового полинома на порождающий полином. Наибольшее количество остатков обеспечивают т.н. **неприводимые (**синоним **– примитивные) полиномы**. Неприводимый полином в алгебре полей Галуа – это аналог простых чисел.

Неприводимым называется полином, который делится без остатка только на **1** и на себя.

Примером могут служить многочлены: **x+1, x^2+x+1, x^3+x+1** и другие. Если неприводимый многочлен имеет степень **p**, то он дает **2^p-1** остатков. Приводимые полиномы той же степени дают меньшее количество разных остатков.

В этом можно убедиться. Разделим вектор ошибки **100000000**… на неприводимый полином **x^3+x+1**, которому соответствует комбинация **1011**.

```
10000000000 |1011
1011          1011100…     Остатки:
_0110                           011
 0000
 _1100                          110
  1011
  _1110                         111
   1011
   _1010                        101
    1011
    _0010                       001
     0000
     _0100                      010
      0000
       1000                     100
```

Дальше остатки повторяются. Получили следующие остатки: **011, 110, 111, 101, 001, 010, 100**. Всего **7** остатков. Неприводимый полином степени **3** и должен давать $2^3-1=7$ остатков.

Теперь для сравнения разделим вектор ошибки на приводимый полином x^3+x^2+x+1, которому соответствует комбинация **1111**.

```
10000000000 |1111
1111          1100…
_1110
 1111
 _0010
  0000
  _0100
   0000
    1000
```

Дальше остатки повторяются. Получили следующие остатки: **111**, **001**, **010**, **100**. Всего **4** остатка.

Иногда при выборе подходящего неприводимого полинома бывает полезно следующее свойство двучлена x^n+1. Если $n = 2^p-1$, то двучлен x^n+1 можно разложить на неприводимые полиномы степени не больше **p.** Например

$$x^{15}+1=x^{2^4-1}+1=(x+1)(x^2+x+1)(x^4+x^3+1)(x^4+x+1)(x^4+x^3+x^2+x+1).$$

Можно сформулировать следующие правила составления порождающих полиномов:

- порождающий полином должен быть неприводимым;
- порождающий полином должен быть делителем двучлена x^n+1;
- степень полинома **p** должна быть настолько большой, чтобы количество остатков (2^p-1) было не меньше количества ошибок, которые требуется локализовать.

Этим требованиям удовлетворяют полиномы, приведенные ниже в табл. 4.5.

ГОСТ 28082-89 «Методы обнаружения ошибок при последовательной передаче данных» устанавливает в качестве основного порождающий полином **16**-й степени $x^{16}+x^{12}+x^5+1$ **(1 0001 0000 0010 0001)**. Для более высокой степени помехозащищенности передаваемой информации ГОСТом рекомендовано использовать полином **32**-й степени $x^{32}+x^{26}+x^{23}+x^{22}+x^{16}+x^{12}+x^{11}+x^{10}+x^8+x^7+x^5+x^4+x^2+x+1$ **(1 0000 0100 1100 0001 0001 1101 1011 0111).**

Таблица 4.5. Таблица примитивных полиномов

Степень полинома	1	2	3	4	5	6	7	8
Двоичное представление	1	111	1011	10011	100101	1000011	10001001	100011101
			1101	11001	111101	1010111	10001111	101110111
					110111	1100111	10011101	111110011
					101111	1001001	11110111	101101001
					111011	1101101	10111111	110111101
							11010101	111100111
							10000011	100101011
								101100011

4.4. Непрерывные коды

Методы кодирования и декодирования, рассмотренные ранее, относились к блочным кодам. При использовании таких кодов информационная последовательность разбивается на отдельные блоки, размером, как правило, не меньшим, чем требуется для кодирования одного символа первичного алфавита. Эти блоки кодируются независимо друг от друга. Таким образом, закодированная последовательность становится последовательностью независимых слов одинаковой длины.

При использовании непрерывных кодов поток данных разбивается на гораздо меньшие блоки длиной $\mathbf{k_0}$ разрядов (в частном случае $\mathbf{k_0 = 1}$), которые называются **кадрами информационных разрядов**.

Кадры информационных разрядов кодируются **кадрами проверочных разрядов** длиной $\mathbf{r_0}$ (в частном случае $\mathbf{r_0 = 1}$). При этом кодирование кадра информационных разрядов в кадр кодового слова производится с учетом предшествующих **L** кадров. **L** называется **размером памяти** кода. Процедура кодирования, таким образом, связывает между собой последовательные кадры кодовых слов. Передаваемая последовательность становится одним полубесконечным кодовым словом.

Сверточный (цепной) алгоритм непрерывного кодирования

Цепной алгоритм кодирования, известный также как код Финка – Хагельбаргера, является одним из наиболее простых примеров непрерывных кодов. В нем каждый проверочный символ формируется путем сложения двух информационных символов, отстоящих друг от друга на **L** позиций.

Введем следующие характеристики кода:

$k = (L+1) \cdot k_0$ - информационная длина слова;

$n = (L+1) \cdot n_0$ - кодовая длина слова, где $\mathbf{n_0 = k_0 + r_0}$.

Кодовая длина слова – это длина кодовой последовательности, на которой сохраняется влияние одного кадра информационных символов.

R = k/n – скорость кода, которая характеризует степень избыточности кода, вводимой для обеспечения исправляющих свойств кода.

Аналогично блочным, сверточные коды обозначаются как **(n,k)**-коды.

В частном, наиболее распространенном, случае, $\mathbf{k_0 = 1}$, $\mathbf{r_0 = 1}$, $\mathbf{n_0 = k_0 + r_0 = 2}$, $\mathbf{R = ½}$.

Хоть избыточность такого кода достаточно велика, но зато он позволяет исправлять **пачки ошибок**, т.е. непрерывные последовательности ошибочных разрядов. Длина исправляемой пачки зависит от размера памяти кода.

Обозначим поток информационных символов через $\mathbf{a_0\, a_1\, a_2 \ldots a_L\, a_{L+1}} \ldots$ Проверочные символы $\mathbf{b_i}$ получаются по следующему формальному правилу: $\mathbf{b_i = a_i \oplus a_{L+i}}$. Таким образом, формируется следующая последовательность проверочных символов сверточного кода: $\mathbf{b_0 = a_0 \oplus a_L}$; $\mathbf{b_1 = a_1 \oplus a_{L+1}}$; … $\mathbf{b_L = a_L \oplus a_{2L}}$; $\mathbf{b_{L+1} = a_{L+1} \oplus a_{2L+1}}$; … В общем потоке символов цепного кода информационные разряды, начиная с $\mathbf{a_L}$, чередуются с проверочными разрядами, а первые **L** информационных символов удваиваются:

$\mathbf{a_0\, a_0\, a_1\, a_1 \ldots a_L\, b_0\, a_{L+1}\, b_1\, a_{L+2}\, b_2 \ldots}$

Таким образом, каждый символ входной последовательности $\mathbf{a_k}$ участвует в формировании двух проверочных символов: $\mathbf{b_{k-L}}$ и $\mathbf{b_k}$. Например, для размера памяти кода **L=3** и $\mathbf{n_0=2}$ процесс формирования выходной последовательности показан на рис. 4.9.

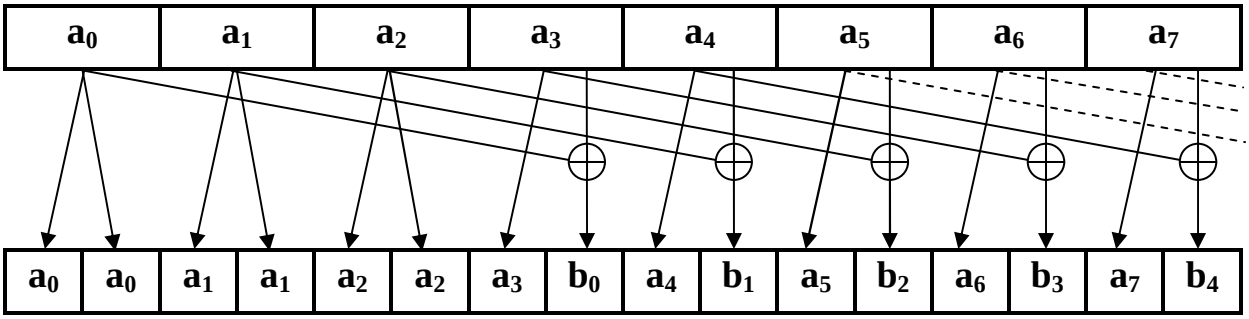


Рис. 4.9. Процесс формирования цепного (8,4)-кода

На приеме информационные и проверочные символы разделяются и регистрируются независимо друг от друга. Из принятой последовательности информационных символов формируются контрольные символы $\mathbf{c_i}$ по тем же правилам, что и проверочные: $\mathbf{c_i = a_i \oplus a_{L+i}}$. Затем каждый контрольный символ $\mathbf{c_i}$ сравнивается с соответствующим проверочным символом $\mathbf{b_i}$. Если произошла ошибка в информационном символе, например, $\mathbf{a_k}$, то это вызовет искажение сразу двух контрольных символов: $\mathbf{c_{k-L}}$ и $\mathbf{c_k}$, что и обнаружится в результате их сравнения с проверочными символами $\mathbf{b_{k-L}}$ и $\mathbf{b_k}$. Отсюда по общему индексу k легко определить и исправить ошибку. Ошибка в принятом проверочном

символе, например, $\mathbf{b_k}$, приводит к несовпадению контрольной и проверочной последовательностей лишь в одном месте. Исправление такой ошибки не требуется. Видно, что проверку надо производить с задержкой на **2L**.

Непрерывное кодирование с помощью импульсной переходной характеристики

Импульсная переходная характеристика (ИПХ) – это реакция кодера на воздействие в виде **δ**-функции. Дискретным аналогом **δ**-функции является полубесконечная последовательность **(100000…)**.

Например, для **(8,4)**-кода память кода **L = 3**. Значит, последовательность **δ = (10000…)** в соответствии с правилами цепного кодирования будет кодироваться так: **u = (11 00 00 01 00 00 00…).** Начиная с 9-го разряда, эта последовательность вырождается в нулевую. Соответственно, первые 8 разрядов и будут представлять собой импульсную переходную характеристику кода. Обозначается так: $\mathbf{H_{(8,4)}}$ **= (11 00 00 01)**. Для кодирования надо просуммировать с соответствующим сдвигом реакцию кодера на каждый входной разряд.

Например, пусть входная последовательность **m = (110100…).** Каждая единица входной последовательности вызывает реакцию в виде ИПХ. Просуммируем эти три реакции:

11 00 00 01

⊕ 11 00 00 01

⊕ 11 00 00 01

u= (11 11 00 10 01 00 01 00 …)

При декодировании полученная последовательность разделяется на информационную (нечетную) и проверочную (четную). Информационная последовательность с помощью ИПХ преобразуется в контрольную последовательность. Если контрольная и проверочная последовательности совпадают, значит, ошибки нет.

Для нашего примера из полученной **û=(11 11 00 10 01 00 01 00 …)** выделяется информационная **â=(1101000 …)** и проверочная $\hat{\mathbf{b}}$ **=(11001010…)**, вычисляется контрольная **ĉ=(11 11 00 10 01 00 01 00 …).** Сравниваем **ĉ** с полученной **û** и убеждаемся, что ошибок нет.

Если в процессе передачи произошла однократная ошибка в пределах **L+1** кадров, ее можно локализовать, вычислив синдромную последовательность, которая представляет последовательность поэлементных сумм проверочных и контрольных символов. Для каждой ИПХ имеется свой синдромный кадр длиной **L+1**. Позиция, в которой синдромный кадр совпадает с синдромной последовательностью, является позицией ошибки.

Допустим, принята последовательность, содержащая ошибки (выделены):

û=(11 01 00 10 01 10 01 00 00…).

Выделяем из нее информационную последовательность (нечетные разряды): **â=(10010100 0…).** Строим контрольную последовательность **ĉ=(11 00 00 10 00 11 01 00 01…)**

Вычисляем синдромную последовательность, сравнивая проверочные (четные) символы принятой последовательности **û** и контрольной последовательности **ĉ**.

S=(010011001…). Синдромный кадр для данного кода равен $\mathbf{S_0=(1001)}$ – это проверочные (четные) символы ИПХ. Сдвигая синдромный кадр вдоль синдромной последовательности, обнаруживаем, что они совпадают во второй и шестой позиции. Следовательно, ошибки локализованы **2**-м и **6**-м разрядом.

4.5. Неалгебраические методы обеспечения помехоустойчивости

Рассмотренные ранее коды предполагают использование алгебраических процедур кодирования-декодирования, то есть разного рода алгебраических уравнений. Эти способы не учитывают вероятностных характеристик передаваемых кодовых последовательностей и помех. Существует ряд методов противодействия помехам, использующих статистические характеристики сообщений, а также учитывающие последствия неверного декодирования.

Пусть по каналу связи передается некоторое сообщение **u** из алфавита **A,** а принимается кодовая последовательность **û**, содержащая искажения вследствие помех, то есть по отдельным символам приемник мог принять неправильные решения (вместо нулей – единицы и наоборот). Эта ошибка случайна, ее вероятность зависит от характеристик канала связи, характеристик кода, метода кодирования и декодирования. Желательно, чтобы вероятность ошибочного декодирования была как можно меньшей.

Введем обозначения:

$\mathbf{u_j}$ − **j**-е слово используемого алфавита **A**; **j=1÷M**;

$\mathbf{u_{ji}}$ − **i**-й символ этого кодового слова;

û - принятое сообщение, возможно, содержащее искажения.

Известны априорные вероятности появления кодовых слов – $\mathbf{P(u_j)}$. Какое кодовое слово содержится в принятом сообщении, мы не знаем. Решение, принятое декодером, будем обозначать **u***. В качестве решения декодера должно быть принято одно из слов алфавита **A**.

Оптимальный декодер должен учитывать всю имеющуюся информацию об используемом коде, канале связи и помехах, действующих в этом канале, и обеспечивать максимальную вероятность правильных ответов о том, какие кодовые слова были переданы по каналу связи. Такой критерий оптимальности называется критерием максимума апостериорной вероятности.

Декодер максимума апостериорной вероятности должен выбирать в качестве решения кодовое слово $\mathbf{u^*} = \mathbf{u_j}$, которое максимизирует условную вероятность $\mathbf{P(u_j \mid \hat{u})}$ — вероятность того, что была передана последовательность $\mathbf{u_j}$, если принята данная реализация сигнала $\hat{\mathbf{u}}$:

$$\mathbf{u^* = \arg\max\{P(u_j \mid \hat{u}); u_j \in A\}}$$

Поскольку $\mathbf{P(u_j \mid \hat{u}) \cdot P(\hat{u}) = P(\hat{u} \mid u_j) \cdot P(u_j)}$, то по формуле Байеса $\mathbf{P(u_j \mid \hat{u}) = P(\hat{u} \mid u_j) \cdot P(u_j)/P(\hat{u})}$. Здесь $\mathbf{P(\hat{u})}$ – это полная безусловная вероятность появления сообщения $\hat{\mathbf{u}}$: $\mathbf{P(\hat{u}) = \sum P(\hat{u} \mid u_j) \cdot P(u_j)}$.

Пример. Источник генерирует **3** кодовых слова: $\mathbf{u_1=(0\ 1\ 0)}$, $\mathbf{u_2=(0\ 0\ 1)}$, $\mathbf{u_3=(1\ 1\ 1)}$ с вероятностями $\mathbf{P_1=0.4}$, $\mathbf{P_2=0.4}$, $\mathbf{P_3=0.2}$. Принята комбинация $\hat{\mathbf{u}}\mathbf{=(1\ 1\ 0)}$. Зная, что вероятность искажения одного бита $\mathbf{p=0.1}$, определить оптимальное по критерию максимума апостериорной вероятности переданное слово.

Оптимальное решение – то, которое максимизирует величину $\mathbf{P(u_j \mid \hat{u}; j=1,2,3)}$. По формуле Байеса $\mathbf{P(u_i \mid \hat{u}) = P(\hat{u} \mid u_j) \cdot P(u_j) / \sum P(\hat{u} \mid u_j) \cdot P(u_j)}$.

$\mathbf{P(\hat{u} \mid u_1)=0.1 \cdot 0.9 \cdot 0.9=0.081}$ (вероятность того, что исказится только 1-й бит)

$\mathbf{P(\hat{u} \mid u_2)=0.1 \cdot 0.1 \cdot 0.1=0.001}$ (вероятность того, что исказится 1-й, 2-й и 3-й биты)

$\mathbf{P(\hat{u} \mid u_3)=0.9 \cdot 0.9 \cdot 0.1=0.081}$ (вероятность того, что исказится только третий бит)

$\mathbf{P(\hat{u}) = \sum P(\tilde{u} \mid u_j) \cdot P(u_j) = 0.081 \cdot 0.4 + 0.001 \cdot 0.4 + 0.081 \cdot 0.2 = 0.049}$

$\mathbf{P(u_1 \mid \hat{u}) = 0.081 \cdot 0.4 / 0.049 = 0.66}$

$\mathbf{P(u_2 \mid \hat{u}) = 0.001 \cdot 0.4 / 0.049 = 0.01}$

$\mathbf{P(u_3 \mid \hat{u}) = 0.081 \cdot 0.2 / 0.049 = 0.33.}$

Максимум апостериорной вероятности (**0.66**) достигается для первого слова - $\mathbf{u_1}$. Следовательно, это решение и будет принято. $\mathbf{u^* = u_1}$

Если считать, что все кодовые слова равновероятны - $\mathbf{P(u_j) = const}$, а также, что безусловная плотность $\mathbf{P(\hat{u})}$ не зависит от $\mathbf{u_j}$, то максимуму $\mathbf{P(u_j \mid \hat{u})}$ соответствует максимум $\mathbf{P(\hat{u} \mid u_j)}$, так называемой функции правдоподобия — условной вероятности того, что сигнал примет значение $\hat{\mathbf{u}}$, если передавалось кодовое слово $\mathbf{u_j}$. **Декодер максимального правдоподобия** выбирает решение, максимизирующее функцию правдоподобия:

$$\mathbf{u^* = \arg\max\{P(\hat{u} \mid u_j); u_j \in A\}}$$

Более развитые методы декодирования учитывают еще и последствия от ошибок декодирования.

Представьте, что кодовое слово $\mathbf{u_3=(1\ 1\ 1)}$ является сигналом боевой тревоги, тогда последствия от ошибочного декодирования могут быть очень велики, несмотря на то, что вероятность такой ошибки мала.

Для оценки последствий ошибочного решения вводится так называемая функция потерь $\mathbf{L(u,u_j)}$.

Функция потерь $\mathbf{L(u,u_j)}$ – это мера негативных последствий при декодировании, являющихся результатом того, что вместо истинного переданного сообщения $\mathbf{u_j}$ принимается решение о приеме сообщения **u**.

В соответствии с критерием Байеса надо декодировать так, чтобы минимизировать средние потери от ошибочного декодирования. Средние потери от принятия решения **u** при получении сообщения $\hat{\mathbf{u}}$ вычисляются по формуле: $\mathbf{W(u|\hat{u}) = \sum L(u,u_j)\cdot P(u_j|\hat{u})}$. **Декодер Байеса** выбирает решение, минимизирующее средние потери: $\mathbf{u}^*=\arg\min\{W(u|\hat{u});u\in A\}$

Пример. Кодовые слова из предыдущего примера имеют следующий смысл: $\mathbf{u_1}$ – проверка, $\mathbf{u_2}$ – учебная тревога, $\mathbf{u_3}$ – боевая тревога. Потери от ошибок декодирования (в условных единицах) сведены в матрицу потерь $\mathbf{L(u,u_j)}$:

$\mathbf{L(u,u_j)}=$

$\mathbf{u_j}$ \ $\mathbf{u}$	$\mathbf{u_1}$	$\mathbf{u_2}$	$\mathbf{u_3}$
$\mathbf{u_1}$	**0**	**5**	**10**
$\mathbf{u_2}$	**8**	**0**	**10**
$\mathbf{u_3}$	**40**	**20**	**0**

$\mathbf{W(u_1|\hat{u}) = \sum L(u_1,u_j)\cdot P(u_j|\hat{u})=0\cdot 0.66+8\cdot 0.01+40\cdot 0.33=13.28}$

$\mathbf{W(u_2|\hat{u}) = \sum L(u_2,u_j)\cdot P(u_j|\hat{u})=5\cdot 0.66+0\cdot 0.01+20\cdot 0.33=9.9}$

$\mathbf{W(u_3|\hat{u}) = \sum L(u_3,u_j)\cdot P(u_j|\hat{u})=10\cdot 0.66+10\cdot 0.01+0\cdot 0.33=6.7}$

Минимальные средние потери при решении $\mathbf{u^*=u_3}$

Декодер Байеса имеет достаточно универсальный характер, так его вид зависит от функции потерь. Задавая функцию потерь тем или иным способом, можно получать различные критерии декодирования. Например, если функция потерь представляет собой кодовое расстояние между истинным и принимаемым сообщением ($\mathbf{L(u,u_j)=\sum u_k \oplus u_{jk}}$), то критерий Байеса превращается в критерий максимального правдоподобия.

Вышерассмотренные примеры относятся к так называемому **жесткому декодированию**, когда в декодере сначала принимается решение относительно значения символов принятой последовательности, а уже затем – относительно значения кодового слова. При жестком декодировании по принятому сигналу сначала определяются символы принятой последовательности $\hat{\mathbf{u}}$, а потом эта последовательность поочередно сравнивается со всеми кодовыми словами данного кода. Решение принимается в пользу кодового слова, оптимизирующего принятый критерий декодирования.

В **мягких декодерах** решения относительно **u** выносятся непосредственно на основе принятого сигнала с учетом статистических

характеристик дискретизации аналогового сигнала. Поскольку в процессе мягкого декодирования информация о сигнале учитывается в большей мере (решение принимается по всему сигналу сразу, а не по частям, для каждого символа в отдельности, и только потом - для всей принятой последовательности), то качество мягкого декодирования должно быть, по идее, выше. Однако реализация жесткого декодера является гораздо более простой – действия выполняются над нулями и единицами. Поэтому такие декодеры используются чаще, хотя и несколько проигрывают мягким декодерам в вероятности правильного декодирования.

Еще один подход к декодированию – использование совокупности критериев и формирование так называемых компромиссных решений на основе теории компромиссов В.Парето.

Задачи и вопросы к главе 4

1. Определить кодовое расстояние между комбинациями **A** и **B** и вес каждой комбинации.

1.1. **A** = **0011011**
B = **1000101**

Решение. Вес комбинации равен количеству единиц в слове: **ω(A) = 4**, **ω(B) = 3**. Для вычисления кодового расстояния сложим комбинации по модулю 2 и найдем вес суммы. **A⊕B** = **1011110**, **ω(A⊕B) = 6**, **d** = **6**.

1.2. **A** = **1011011**
B = **1101111**;

1.3. **A** = **0100010101110**
B = **1010011011110**;

1.4. **A** = **01011**
B = **10101.**

2. Алфавит кода состоит из трех разрешенных комбинаций **A**, **B** и **C**. Найти минимальное кодовое расстояние и оценить корректирующую способность кода.

2.1. **A** = **1011010**
B = **1101111**
C = 1011001

Решение. $\mathbf{d_{min} = min\{4, 4, 2\} = 2}$. Следовательно, в силу (4.2), любая однократная ошибка может быть обнаружена. Ошибки более высокой кратности могут не быть обнаружены. Гарантий исправления даже однократных ошибок в этом коде нет, так как не выполняется условие (4.3). Действительно, ошибки в 6-м или 7-м разрядах комбинаций **A** и **C** не исправляется. А вот ошибку в первом или 4-м разрядах можно обнаружить и исправить.

2.2. **A** = **1011010**
B = **1001101**
C = 1011001;

2.3. **A** = **0011010**
B = **0101101**
C = 1011001;

3. Код Хемминга **(7,4)** имеет порождающую матрицу:

$$\mathbf{G}=\begin{bmatrix}1&0&0&0&0&1&1\\0&1&0&0&1&0&1\\0&0&1&0&1&1&0\\0&0&0&1&1&1&1\end{bmatrix}$$

3.1. зашифруйте число $\mathbf{13_{10}}$;

3.2. зашифруйте число $\mathbf{0101_2}$**;**

3.3. исправьте ошибку в кодовом слове **0110001** и найдите передаваемое число

3.4. исправьте ошибку в кодовом слове **0111010** и найдите передаваемое число

4. Выбрать порождающий полином циклического кода, исправляющего однократные ошибки и позволяющего передать **2000** различных сообщений.

Решение. Определим необходимое количество информационных разрядов из соотношения $\mathbf{k \geq \log 2000}$. $\mathbf{k=11}$.
По правилу Хемминга ($\mathbf{r \geq \log (n+1)}$, **для t=1**) определим, что для исправления однократных ошибок требуется **4** проверочных разряда, следовательно, надо строить **(15,11)**-код.
В **(15,11)**-коде возможно возникновение **15** одиночных ошибок, следовательно, нужен порождающий полином **4**-й степени, так как при этом полиноме получаются $\mathbf{2^4-1=15}$ различных остатков от деления.
Из таблицы примитивных полиномов выбираем порождающий полином $\mathbf{g(x)=(x^4+x+1)}$, так как он проще других, обеспечивающих ту же корректирующую способность кода.

5. Циклический код порождается многочленом $\mathbf{g(x) = x^3 + x + 1}$;

5.1. закодируйте число $\mathbf{7_8}$;

5.2. закодируйте число $\mathbf{10_{10}}$;

5.3. найдите и исправьте ошибку в принятой кодовой комбинации **0111000**;

5.4. найдите и исправьте ошибку в принятой кодовой комбинации **0111001**.

6. Требуется передавать пятиразрядные двоичные слова с помощью кода, исправляющего трёхкратные ошибки. Чему равна общая длина кодовых слов?

7. Определить количество информационных разрядов кода длиной в пятнадцать символов, если код исправляет две ошибки.

Решение. По формуле (4.5) $\mathbf{r \geq \log_2 \sum_{i=0}^{t} C_n^i}$. Здесь **t**=2, **n**=15. Можно записать $\mathbf{r \geq \log_2(1+n+n(n-1)/2)}$. $\mathbf{r \geq \log_2 121 = 7}$. Количество информационных разрядов **k = n – r = 8**.

8. Построить порождающий многочлен для создания циклического кода, обнаруживающего все трёхкратные ошибки при передаче **1000** сообщений.

9. Представьте порождающий многочлен **11001** в полиномиальном виде.

10. Укажите синдром однократной ошибки в **6**-м разряде для **(15,11)**-кода Хемминга.

11. Канонический блочный код имеет порождающую матрицу

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Определите синдром однократной ошибки в **3**-м разряде принятой кодовой комбинации.

12. Помехоустойчивый декодер принял двоичную комбинацию **Ŭ** и вычислил вектор ошибки **e**. Определите переданное десятичное число.

12.1. **Ŭ = 01101101, e = 10001000;**

12.2. **Ŭ = 11000101, e = 00011000;**

12.3. **Ŭ = 00111101, e = 00001011;**

12.4. **Ŭ = 11101011, e = 10100001;**

13. Блочный код имеет порождающую матрицу **G**.

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Определите синдром двукратной ошибки во втором и шестом разрядах принятого кодового слова.

14. Цепной (8,4)-код имеет импульсную переходную характеристику **(11000001)**. На вход помехоустойчивого кодера подается бинарная последовательность **(1010110100…)**. Закодируйте эту последовательность.

15. Дайте классификацию помехоустойчивых кодов.

16. В чем состоят основные принципы помехоустойчивого кодирования?

17. Какие характеристики определяют корректирующую способность кода?

18. Что такое синдром ошибки? Чем определяется значность синдрома?

Глава 5
Вычислительный практикум по теории информационных процессов и систем

Вычислительный практикум представляет собой четыре работы, объединенные единым заданием. Структура практикума основана на общей схеме передачи информации по каналу связи (см. рис. 3.1).

Первая работа посвящена расчету характеристик источника информации, в частности определению энтропии и избыточности естественных языков, а также первичному кодированию алфавита источника.

Во **второй работе** требуется смоделировать процесс передачи информации по каналу с помехами. При этом сначала выполняются априорные расчеты, базирующиеся на заданных параметрах канала связи и модели ошибок, а затем имитируется экспериментальные измерения характеристик канала связи.

В **третьей работе** от студентов требуется построить программный кодер и декодер, обеспечивающий помехоустойчивость при передаче сообщений. Программный модуль должен работать автономно, кодируя/декодируя бинарные слова, введенные с клавиатуры.

Четвертая работа объединяет предыдущие. Блок помехоустойчивого кодирования/декодирования встраивается в модель передачи информации по каналу с помехами, и проводятся исследования помехоустойчивой системы передачи информации. Делаются выводы относительно эффективности использования помехоустойчивых алгоритмов.

5.1. Работа №1
Исследование избыточности источника информации

Теоретические сведения

Коэффициент избыточности показывает, какая часть реального сообщения является излишней и могла бы не передаваться, если бы сообщение было организовано оптимально. Коэффициент избыточности выражается формулами:

$$\varphi = \frac{\mathbf{n_p} - \mathbf{n_0}}{\mathbf{n_p}} = 1 - \frac{\mathbf{n_0}}{\mathbf{n_p}}$$

$$\varphi = \frac{\mathbf{H_0} - \mathbf{H_p}}{\mathbf{H_0}} = 1 - \frac{\mathbf{H_p}}{\mathbf{H_0}}$$

где $\mathbf{n_p}$ и $\mathbf{H_p}$ - длина и энтропия реального сообщения, $\mathbf{n_O}$ и $\mathbf{H_O}$ – длина и энтропия оптимального сообщения.

Сообщение будет оптимальным тогда, когда все символы алфавита равновероятны. Энтропия такого сообщения максимальна и равна $\mathbf{\log_2 m}$, где **m** – мощность алфавита. Энтропия реального сообщения при условии независимости символов сообщения вычисляется по формуле:

$$\mathbf{H = -\sum_{i=1}^{m} p_i \cdot \log_2 p_i},$$

где $\mathbf{p_i}$ – вероятность появления **i**-го символа.

Энтропия сообщения при условии попарной зависимости символов вычисляется по формуле:

$$\mathbf{H = -\frac{1}{2} \sum_i \sum_j p_{ij} \log p_{ij}}$$

где $\mathbf{p_{ij}}$ - вероятность появления пары символов.

Справочно: для русского языка значение реальной энтропии равно 4.36 бит на символ, для английского языка – 4.04, для французского – 3.96, для немецкого – 4.10, для испанского – 3.98. С учетом попарной зависимости символов энтропия русского языка – 3.52, английского языка – 3.32 бит на символ.

Задание к работе

Задание А.

А.1. Получить (найти в Интернете) текстовые файлы, содержащие текст на одном из естественных языков или языке эсперанто, в соответствии со своим вариантом. Длина текста – не менее 10 тыс. символов.

А.2. Рассчитать оптимальную энтропию для заданного языка.

А.3. Составить программу для экспериментального определения реальной энтропии сообщения при условии независимости символов. Выполнить расчет для полученного файла.

А.4. Рассчитать избыточность сообщения заданного естественного языка при условии независимости символов алфавита.

А.5. Используя текстовый файл из п. А.1, сформировать усеченный алфавит языка, т.е. оставить в текстовом файле только символы из трети исходного алфавита. Пробелы и знаки препинания не учитывать, прописные и строчные символы не различать. Мощность усеченного алфавита должна быть не более 15.

А.6. Построить таблицу частот символов первичного усеченного алфавита

А.7. Закодировать символы усеченного алфавита равномерным бинарным кодом.

Задание Б.

Б.1. Составить программу для экспериментального определения реальной энтропии сообщения при условии попарной зависимости символов. Выполнить расчет для полученного файла.

Б.2. Рассчитать избыточность сообщения заданного естественного языка с учетом попарной зависимости символов алфавита.

Б.3. Вместе с товарищами сравнить избыточность различных языков.

Б.4. Выполнить пункт Б.1 для усечённого алфавита.

Варианты заданий

№ в.	Название языка	№ в.	Название языка
1.	Английский (English)	13.	Итальянский (Italiano)
2.	Испанский (Español)	14.	Турецкий (Türkçe)
3.	Португальский (Português)	15.	Финский (Suomi)
4.	Японский (中文)	16.	Французский (Français)
5.	Чешский (Česky)	17.	Голландский (Nederlands)
6.	Датский (Dansk)	18.	Польский (Polski)
7.	Немецкий (Deutsch)	19.	Курдский (Kurdî/كوردي)
8.	Греческий (Ελληνικά)	20.	Сербский (Српски/Srpski)
9.	Эсперанто (Esperanto)	21.	Хорватский (Hrvatski)
10.	Румынский (Română)	22.	Шведский (Svenska)
11.	Словацкий (Slovenčina)	23.	Белорусский (Беларуский)
12.	Украинский (Українська)	24.	Норвежский (Bokmål)

Контрольные вопросы

1. Что называется энтропией?

2. Каковы причины появления избыточности в сообщении?

3. Как можно уменьшить избыточность?

4. Как экспериментально определить вероятность появления символа в сообщении?

5. Каковы положительные и отрицательные стороны наличия избыточности сообщения?

6. Сравните величины энтропии, полученные в п.п. А.3 и А.4 задания. Теоретически обоснуйте полученный результат.

7. Сравните полученные значения энтропии естественного языка со справочными. Поясните причину отличия.

5.2. Работа №2
Моделирование передачи сообщения по каналу с помехами

Теоретические сведения

Наличие помех в канале связи приводит к тому, что часть информации при перемещении по каналу теряется, искажается. Информация, принятая приемником, не полностью снимает неопределенность относительно переданной источником, хотя и уменьшает ее. Если на вход канала связи поступил сигнал **u**, а с выхода канала принят сигнал **v**, то говорят о **взаимной (**или **полезной) информации I(u,v)**.

Взаимной информацией между сообщениями **u** и **v I(u,v)** называется величина, определяемая соотношениями:

$$\mathbf{I(u,v) = H(v) - H(v|u)}$$

или

$$\mathbf{I(u,v) = H(u) - H(u|v),}$$

где $\mathbf{H(u)}$ – энтропия передатчика сообщения, $\mathbf{H(u\,|\,v)}$ – апостериорная энтропия, которая учитывает утечку информации при передаче из-за разрушения ее помехами (**ненадежность канала**), $\mathbf{H(v)}$ - энтропия приемника (выхода) канала, $\mathbf{H(v\,|\,u)}$ характеризует постороннюю информацию, вносимую помехами (априорная **энтропия шума**).

Первая из формул взаимной информации представляет собой априорную оценку, а вторая – апостериорную оценку количества информации, передаваемого по каналу с помехами.

Энтропия приемника определяется по формуле:

$$\mathbf{H(v) = -\sum_{i=1}^{Nv} P(v_i)\log P(v_i)}$$

Энтропия шума определяется по формуле:

$$\mathbf{H(v\,|\,u) = -\sum_{i=1}^{Nu} P(u_i)\sum_{j=1}^{Nv} P(v_j\,|\,u_i)\log P(v_j\,|\,u_i)}$$

Для определения информационных характеристик канала с помехами требуется знать статистику входных сообщений (вероятности $\mathbf{P(u_i)}$), а также переходные вероятности $\mathbf{P(v_j|u_i)}$, определяющие характер помех. Переходные вероятности обычно задаются матрицей размерностью **Nu×Nv**, называемой **канальной матрицей**.

Вероятности выходных сообщений $\mathbf{P(v_j)}$ можно определить, зная вероятности входных сообщений и канальную матрицу:

$$\mathbf{P(v_j) = \sum_{i=1}^{Nu} P(u_i) P(v_j \mid u_i)}$$

Если канал передачи информации имеет бинарный характер (то есть передает два символа, например **0** и **1**), то ошибки при передаче могут быть двоякого рода: ошибочная **инверсия** и потеря символа (**стирание**). Ошибочная инверсия описывается условными вероятностями того, что принят **0** при условии, что послана **1** (обозначается **p(1|0)**) и того, что принята **1** при условии, что был послан **0** (обозначается **p(0|1)**). Потеря символа описывается условными вероятностями стирания **0** и стирания **1** (обозначается **p(?|0)** и **p(?|1)** соответственно). Если условные вероятности ошибочной инверсии равны (**p(1|0)** = **p(0|1)**) и условные вероятности потери символа равны (**p(?|0)** = **p(?|1)**), то такой канал связи называется бинарным **симметричным**.

Взаимная информация **I(u,v)** показывает количество информации, содержащееся в среднем в одном принятом сообщении. Если по каналу передается ансамбль из **n** сообщений, то количество информации, поданной в канал, определяется как **n·H(u)**, а количество принятой информации равно **n·I(u,v)**. Разница между этими величинами составляет потери информации. При этом скорость передачи информации будет равна $\mathbf{J = \frac{1}{\tau}(H(v) - H(v \mid u))}$, где τ – среднее время передачи одного сообщения.

Рекомендации по составлению канальной матрицы

Для составления канальной матрицы необходимо,

- во-первых, закодировать входные символы первичного алфавита в виде двоичных последовательностей равномерной длины;
- во-вторых, для каждого символа первичного алфавита определить переходные вероятности его передачи по каналу, то есть вероятность правильной передачи и вероятности его искажения (перехода в другой символ).

При этом переходные вероятности передачи символа определяются, исходя из заданных вероятностей ошибок бинарного канала.

Например, пусть символ **A** закодирован последовательностью **001**, а символ **B** – последовательностью **010**. Тогда вероятность правильной передачи символа **A** определяется так: **p(A|A) = p(0|0)·p(0|0)·p(1|1)**, а вероятность ошибочного перехода символа **A** в символ **B** определяется так: **p(B|A) = p(0|0)·p(1|0)·p(0|1)**. Вероятности правильной передачи бинарных символов, вычисляются из того соображения, что вместе вероятностями ошибок они представляют полную группу событий: **p(0|0) + p(1|0) + p(?|0) = 1** и **p(1|1) + p(0|1) + p(?|1) = 1.**

Следует иметь ввиду, что в результате ошибок передачи по каналу связи в приемнике могут появиться символы, не совпадающие ни с одним из символов первичного алфавита. Это может произойти либо из-за стирания бинарного символа, либо из-за того, что инверсия бинарных символов приведет к возникновению двоичной последовательности, отсутствующей в кодировке первичного алфавита. Значит, алфавит приемника будет иметь мощность **Nv = Nu +1** и содержать дополнительный символ, отсутствующий в алфавите источника. Соответственно, канальная матрица будет иметь размерность **Nu×(Nu+1)**.

Рекомендации по экспериментальному исследованию информационных характеристик

Экспериментальное исследование информационных характеристик канала с помехами основано на определении апостериорных вероятностей.

Энтропия источника рассчитывается на основе апостериорных вероятностей входных сообщений, которые определяются их частотой: $\mathbf{P(u_i) \approx \frac{n_i}{n}}$,

где $\mathbf{n_i}$ – количество переданных по каналу сообщений **i**-го вида.

Апостериорная энтропия канала (ненадежность канала) вычисляется, исходя из апостериорных вероятностей вида $\mathbf{P(u_i \mid v_j)}$, представляющие собой вероятности того, что на вход канала было подано сообщение $\mathbf{u_i}$ при условии, что на выходе принято сообщение $\mathbf{v_j}$. Апостериорные вероятности определяется частотой:

$$\mathbf{P(u_i \mid v_j) \approx \frac{m^*_{ij}}{m_j}},$$

где $\mathbf{m_j}$ – количество полученных сообщений **j**-го вида, $\mathbf{m^*_{ij}}$ - количество переданных сообщений **i**-го вида, воспринятых приемником как сообщение **j**-го вида. В качестве вероятностей выходных сообщений $\mathbf{P(v_j)}$ также используются частоты, полученные экспериментально: $\mathbf{P(v_j) \approx \frac{m_j}{n}}$. Таким образом, ненадежность канала вычисляется по известной формуле условной энтропии $\mathbf{H(U \mid V) = \sum_j P(v_j) \sum_i P(u_i \mid v_j) \log P(u_i \mid v_j)}$.

Соответственно, апостериорная оценка полезной информации находится по формуле **I(U,V) = H(U) – H(U|V)**.

Рекомендации по моделированию случайных событий

Если случайное событие **A** имеет вероятность **p(A)**, то порядок его моделирования следующий:

- с помощью генератора случайных чисел (функция random в языке Паскаль или функция rand() в языке C) получаем величину ***x***, находящуюся в интервале [0, 1];
- если **p(A)** < ***x*** принимается решение, что событие **A** произошло, в противном случае – событие **A** не произошло.

Если имеется полная группа событий $\mathbf{A_1, A_2, \ldots, A_k}$ с вероятностями $\mathbf{p(A_1), p(A_2), \ldots, p(A_k)}$ соответственно, причем $\sum \mathbf{p=1}$, то порядок их моделирования следующий:

- разбиваем отрезок [0, 1] на **k** интервалов, каждый из которых имеет размер $\mathbf{p(A_i)}$, где i=1÷k;
- с помощью генератора случайных чисел (функция random() в языке Паскаль или функция rand() в языке C) получаем величину ***x***, находящуюся в интервале [0, 1];
- если величину ***x*** попадает в **i**-тый интервал, то принимается решение, что произошло событие $\mathbf{A_i}$.

Например, требуется смоделировать появление трех событий **A, B** и **C**, имеющих вероятности **p(A)=0.1, p(B)=0.6** и **p(C)=0.3.** Разбиваем отрезок [0, 1] на три интервала: [0, 0.1), [0.1, 0.7) и [0.7, 1]. С помощью функции ***random*** получаем величину ***x***. Если $\boldsymbol{x} \in [0, 0.1)$ принимается решение о том, что произошло событие **A**, если $\boldsymbol{x} \in [0.1, 0.7)$ – событие **B**, если $\boldsymbol{x} \in [0.7, 1]$ – событие **C.**

Задание к работе

Задание А.

Исходя из распределения вероятностей символов первичного алфавита, полученных в задании лабораторной работы №1, и заданных характеристик бинарного канала связи (см. варианты заданий), рассчитать априорные характеристики передачи информации по каналу с помехами.

А.1. Закодировать первичный алфавит, полученный в п.А.5 лабораторной работы №1, двоичным равномерным кодом минимальной длины.

А.2. Построить канальную матрицу передачи сообщений на входном языке, определенном в задании лабораторной работы №1.

А.3. Рассчитать априорные характеристики:

А.3.1. Энтропию источника информации.

А.3.2. Энтропию приемника информации.

А.3.3. Энтропию шума.

А.3.4. Ненадежность канала

А.3.5. Количество полезной информации.

А.3.6. Скорость передачи информации.

А.4. В программе “SPI.exe” смоделировать прохождение сообщений по каналу связи с помехами (без помехоустойчивого кодирования), построить экспериментальную канальную матрицу и рассчитать апостериорные характеристики:

А.4.1. Энтропию источника информации.

А.4.2. Энтропию приемника информации.

А.4.3. Энтропию шума.

А.4.4. Ненадежность канала

А.4.5. Количество полезной информации.

А.4.6. Скорость передачи информации.

А.5. Сравнить и пояснить результаты, полученные в п.п. А.3 и А.4.

Задание Б.

Вместо выполнения п. А.4 написать программу, моделирующую прохождение сообщений по каналу с помехами.

Б.1. Смоделировать генерацию **К** сообщений источника информации в соответствии с заданным распределением вероятности символов первичного алфавита. Результатом должна стать апостериорные вероятности (частоты) входных сообщений.

Б.2. Смоделировать искажения при прохождении входных сообщений по каналу с помехами в соответствии с заданными вероятностными характеристиками. Результатом должна стать апостериорная канальная матрица и апостериорные вероятности (частоты) выходных сообщений.

Б.3. На основании моделирования вычислить экспериментальные характеристики: количество переданной информации (энтропию источника), количество принятой информации (энтропию приемника), потери информации (утечку), излишнюю информацию (шумы), количество полезной информации и скорость передачи информации.

Б.4. Сравнить экспериментальные результаты с теоретическими и пояснить причину отличий.

Варианты заданий

Вариант №	Вероятности искажений	Время передачи одного разряда τ_0, мс.	Кол-во передаваемых символов К	Вариант №	Вероятности искажений	Время передачи одного разряда τ_0, мс.	Кол-во передаваемых символов К
1	P(0 → 0)=0.6 P(0 → 1)=0.4 P(1 → 1)=0.7 P(1 → 0)=0.3	0.1	500	13	P(0 → 0)=0.66 P(0 → 1)=0.34 P(1 → 1)=0.65 P(1 → 0)=0.35	0.3	300
2	P(0 → 0)=0.55 P(0 → 1)=0.45 P(1 → 1)=0.6 P(1 → 0)=0.3	0.2	200	14	P(0 → 0)=0.6 P(0 → 1)=0.4 P(1 → 1)=0.7 P(1 → 0)=0.3	0.4	400
3	P(0 → 0)=0.65 P(0 → 1)=0.35 P(1 → 1)=0.65 P(1 → 0)=0.35	0.3	300	15	P(0 → 0)=0.67 P(0 → 1)=0.33 P(1 → 1)=0.8 P(1 → 0)=0.2	0.5	500
4	P(0 → 0)=0.7 P(0 → 1)=0.3 P(1 → 1)=0.7 P(1 → 0)=0.3	0.4	400	16	P(0 → 0)=0.57 P(0 → 1)=0.43 P(1 → 1)=0.61 P(1 → 0)=0.39	0.6	600
5	P(0 → 0)=0.75 P(0 → 1)=0.25 P(1 → 1)=0.7 P(1 → 0)=0.3	0.5	500	17	P(0 → 0)=0.58 P(0 → 1)=0.42 P(1 → 1)=0.69 P(1 → 0)=0.31	0.7	500
6	P(0 → 0)=0.75 P(0 → 1)=0.25 P(1 → 1)=0.65 P(1 → 0)=0.35	0.6	600	18	P(0 → 0)=0.68 P(0 → 1)=0.32 P(1 → 1)=0.6 P(1 → 0)=0.40	0.8	900
7	P(0 → 0)=0.58 P(0 → 1)=0.42 P(1 → 1)=0.67 P(1 → 0)=0.33	0.7	700	19	P(0 → 0)=0.56 P(0 → 1)=0.44 P(1 → 1)=0.71 P(1 → 0)=0.29	0.5	600
8	P(0 → 0)=0.6 P(0 → 1)=0.4 P(1 → 1)=0.72 P(1 → 0)=0.28	0.5	500	20	P(0 → 0)=0.58 P(0 → 1)=0.42 P(1 → 1)=0.75 P(1 → 0)=0.25	0.6	500
9	P(0 → 0)=0.5 P(0 → 1)=0.4 P(1 → 1)=0.6 P(1 → 0)=0.3	0,3	300	21	P(0 → 0)=0.78 P(0 → 1)=0.22 P(1 → 1)=0.77 P(1 → 0)=0.23	0.7	700
10	P(0 → 0)=0.58 P(0 → 1)=0.42 P(1 → 1)=0.9 P(1 → 0)=0.1	0.2	400	22	P(0 → 0)=0.6 P(0 → 1)=0.4 P(1 → 1)=0.7 P(1 → 0)=0.3	0.9	800
11	P(0 → 0)=0.58 P(0 → 1)=0.42 P(1 → 1)=0.65 P(1 → 0)=0.35	0.1	500	23	P(0 → 0)=0.54 P(0 → 1)=0.46 P(1 → 1)=0.77 P(1 → 0)=0.23	0.3	400
12	P(0 → 0)=0.59 P(0 → 1)=0.41 P(1 → 1)=0.55 P(1 → 0)=0.45	0.2	200	24	P(0 → 0)=0.55 P(0 → 1)=0.45 P(1 → 1)=0.8 P(1 → 0)=0.2	0.4	500

Контрольные вопросы

1. Что такое информационный канал с помехами?

2. Всегда ли количество переданной информации совпадает с количеством принятой информации?

3. За счет чего возникают помехи в канале?

4. Чем отличается априорная вероятность от апостериорной?

5. Что такое скорость передачи информации, пропускная способность информационного канала?

6. Как влияют помехи на пропускную способность информационного канала?

7. Какие типовые модели информационных каналов с помехами Вы знаете?

8. Как соотносятся производительность источника информации и пропускная способность информационного канала?

9. Нарисуйте структурную схему канала передачи информации и поясните назначение ее элементов.

10. Что такое канальная матрица и каковы ее свойства?

5.3. Работа №3
Помехоустойчивое кодирование сообщений

Теоретические сведения

Помехоустойчивыми (корректирующими) называются коды, позволяющие обнаружить и при необходимости исправить ошибки в принятом сообщении. Наибольшее распространение получили блочные систематические коды, в которых существует четкое разделение разрядов на информационные и проверочные. Такие коды обозначаются как **(n,k)-**коды, где **n** – количество разрядов в блоке, **k** – количество информационных разрядов в блоке.

Одними из блочных систематических кодов являются матричный код и циклический код.

Матричный код

Получение кодовых комбинаций производится с помощью порождающих матриц, состоящих из **k** строк и **n** столбцов:

$$\mathbf{G}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} & b_{11} & b_{12} & \dots & b_{1r} \\ a_{21} & a_{22} & \dots & a_{2k} & b_{21} & b_{22} & \dots & b_{2r} \\ \hdots & \hdots & \hdots & \hdots & \hdots & \hdots & \hdots & \hdots \\ a_{k1} & a_{k2} & \dots & a_{kk} & b_{k1} & b_{k2} & \dots & b_{kr} \end{bmatrix}$$

В классической (канонической) форме кода элементы первых k столбцов служат для информационных целей, а оставшихся – для проверочных. Соответственно, порождающую матрицу **G** можно представить в виде двух подматриц – информационной $\mathbf{I_k}$ и проверочной **P**. $\mathbf{G}=\left[\mathbf{I_k}\|\mathbf{P}\right]$, где

$$\mathbf{I_k}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \hdots & \hdots & \hdots & \hdots \\ a_{k1} & a_{k2} & \dots & a_{kk} \end{bmatrix},\ \mathbf{P}=\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1r} \\ b_{21} & b_{22} & \dots & b_{2r} \\ \hdots & \hdots & \hdots & \hdots \\ b_{k1} & b_{k2} & \dots & b_{kr} \end{bmatrix}$$

Информационную подматрицу обычно берут в виде квадратной единичной матрицы размерностью k×k, а проверочная подматрица **P** строится с соблюдением следующих условий:

1) вес (количество единиц) каждой строки подматрицы должен быть не менее $\mathbf{d_{min}-1}$, где $\mathbf{d_{min}}$ – минимальное кодовое расстояние кода,
2) все строки должны быть различны.

Кодовое слово помехоустойчивого кода образуется умножением исходной комбинации на порождающую матрицу: $\mathbf{u = m \cdot G}$.

В отличие от канонического представления, в коде Хемминга информационные и проверочные биты не разнесены в отдельные подматрицы, а чередуются. Если биты кодовой комбинации пронумеровать, начиная с 1,

слева направо, то контрольными (проверочными) оказываются биты с номерами 1, 2, 4, 8 и т.д., а все остальные являются информационными. Таким образом, для получения порождающей матрицы для кода Хемминга нужно в порождающей матрице канонической формы переставить столбцы, поставив столбцы проверочной подматрицы на 1, 2, 4, 8 и т.д. места.

Циклический код

Циклические коды используют полиномиальное представление двоичных последовательностей. Для составления **(n,k)**-кода применяется так называемый порождающий полином степени **n – k**. Кодовое слово циклического кода состоит из неизменной информационной части **m** и **(n-k)** проверочных символов. Проверочные символы являются коэффициентами полинома **ρ(x)**, то есть остатка от деления $\mathbf{m(x)\cdot x^{n-k}}$ на порождающий полином **g(x)**. То есть, при поступлении на вход кодера комбинации **m** кодовое слово **u** получается так:

$$\mathbf{u(x)=\rho(x)\oplus m(x)\cdot x^{n-k} = m_{k-1}\ x^{n-1}+\ldots+ m_1\cdot x^{n-k+1} + m_0\ x^{n-k} +\rho_{n-k-1}\ x^{n-k-1} +\ldots+\rho_1\ x +\rho_0}$$

что соответствует бинарной комбинации $\mathbf{u = (m_{k-1} \ldots m_1\ m_0\ \rho_{n-k-1}\ \rho_1 \ldots \rho_0)}$.

Задание к работе

Задание А.

А.1. Для заданного в лабораторной работе №2 варианта системы передачи информации определить:

А.1.1. Количество информационных разрядов в кодовом слове,

А.1.2. Минимальное кодовое расстояние для однократных и двукратных ошибок,

А.1.3. Количество контрольных разрядов для кодов, исправляющих однократные и двукратные ошибки.

А.2. Составить порождающие матрицы для кода Хемминга, исправляющего однократные ошибки, и канонического матричного кода, исправляющего двукратные ошибки.

А.3. Составить проверочные матрицы для матричных кодов, исправляющих однократные и двукратные ошибки.

А.4. Для матричного кода, исправляющего двукратные ошибки, составить таблицу соответствия синдрома и локализации ошибок (в таблицу включить только ошибки в информационных разрядах).

А.5. Составить программу кодирования, декодирования и локализации однократных ошибок по алгоритму Хемминга.

А.6. Составить порождающий полином **g(x)** для исправления однократных и двукратных ошибок.

Задание Б.

Б.1. Составить программу кодирования, декодирования и локализации однократных ошибок с использованием циклического кода.

Контрольные вопросы

1. Дайте классификацию помехоустойчивых кодов.

2. Что такое корректирующая способность кода?

3. Что такое кодовое расстояние и как оно влияет на корректирующую способность кода?

4. Что такое избыточность кода и от чего она зависит? Как уменьшить избыточность кода?

5. Как определяется длина кодового слова?

6. Как вычисляется синдром ошибки?

7. Опишите правила сложения по модулю два.

8. Опишите арифметические правила для полиномов.

9. Опишите правила кодирования с использованием порождающей матрицы.

10. Опишите правила декодирования с использованием проверочной матрицы.

11. Что такое вектор ошибки?

12. Опишите правила кодирования с помощью порождающих полиномов

13. Как определить соответствие вектора ошибки и синдрома ошибки для циклических кодов?

14. Как определить соответствие вектора ошибки и синдрома ошибки для матричных кодов?

5.4. Работа №4
Исследование характеристик помехоустойчивых кодов

Теоретические сведения

Помехоустойчивость кодирования обеспечивается за счет введения избыточности. Это значит, что из **n** символов кодовой комбинации для передачи информации используется **k < n** символов.

Основные характеристики помехоустойчивых кодов:

- коэффициент избыточности кода,
- доля обнаруживаемых ошибок,
- доля исправляемых ошибок.

Коэффициент избыточности кода – это отношение количества проверочных битов к длине кода: $\mathbf{F = \frac{n-k}{n}}$.

Доля обнаруживаемых ошибок определяется отношением запрещенных принятых комбинаций, которые могут быть обнаружены, ко всему количеству возможных вариантов передачи. Всего возможных кодовых слов – $\mathbf{2^n}$, из них безошибочными могут быть $\mathbf{2^k}$. Соответственно, все множество кодовых комбинаций разбивается на две группы: разрешенные комбинации и запрещенные комбинации. Разрешенных комбинаций $\mathbf{S_p = 2^k}$, запрещенных $\mathbf{S_f = S - S_p = 2^n - 2^k}$. Если на приемной стороне установлено, что принятая комбинация относится к разрешенным, то считается, что сообщение прошло без искажений, а если принята запрещенная комбинация, то делается вывод, что произошла ошибка. Однако, если ошибка такова, что посланная комбинация, претерпев искажения, тем не менее попала в множество разрешенных комбинаций, такая ошибка обнаружена не будет. Каждая из разрешенных комбинаций при передаче может трансформироваться в любую из **S** возможных комбинаций, т.е. всего имеется **S·Sp** возможных вариантов передачи. Из них **Sp** вариантов безошибочной передачи, **Sp·(Sp – 1)** вариантов ошибочной трансформации в другие разрешенные комбинации и **Sp·(S – Sp)** вариантов трансформации в запрещенные комбинации.

Только передача в запрещенные варианты может быть обнаружена. Следовательно, доля обнаруживаемых ошибок определяется так:

$$\mathbf{D_o = \frac{S_p(S - S_p)}{S \cdot S_p} = 1 - \frac{S_p}{S} = 1 - \frac{2^k}{2^n} = 1 - 2^{k-n}}.$$

Доля исправляемых ошибок определяется отношением запрещенных принятых комбинаций, которые могут быть исправлены, к ошибочным комбинациям, которые могут быть обнаружены. Обнаруженную ошибку можно

исправить, если для каждой запрещенной комбинации существует единственная исходная комбинация. Таким образом, ошибка исправляется в **S – Sp** случаях, равных количеству запрещенных комбинаций. Доля исправляемых ошибочных комбинаций от числа обнаруживаемых составляет:

$$\mathbf{D_и = \frac{S - S_p}{S_p(S - S_p)} = \frac{1}{S_p} = 2^{-k}}.$$

Эти две величины априорно характеризуют корректирующую способность кода.

Если известны вероятностные характеристики искажений, то можно определить **априорную вероятность исправления ошибки**. При этом предполагается, что помехоустойчивый код создается с расчетом на определенную кратность ошибок и не исправляет ошибки более высокой кратности. Например, **(7, 4)**-код теоретически должен исправлять однократные ошибки и не рассчитан на ошибки высокой кратности. Таким образом, вероятность исправления ошибки для определенного кода равна вероятности появления ошибок кратности $\leq$ **t**, включая вероятность безошибочной передачи, где **t** – кратность ошибки, на которую рассчитан помехоустойчивый код.

Вероятность появления ошибки кратности **t** рассчитывается по формуле:

$$\mathbf{P_n = C_n^t \cdot p^t \cdot (1 - p)^{n-t}},$$

где **p** – вероятность однократной ошибки.

Для того, чтобы определить апостериорные характеристики, надо провести натурный или модельный эксперимент. В ходе эксперимента требуется собрать следующую статистику: **N** – общее количество переданных кодовых комбинаций, **Nш** – количество комбинаций, которые в процессе передачи получали искажения, **No** – количество комбинаций, в которых ошибки были обнаружены, **Nи** – количество комбинаций, в которых ошибки были исправлены. Комбинациями, получившими искажения, являются те, код которых на входе канала связи не совпала с кодом на выходе канала связи. Обнаруженными ошибками являются те, для которых помехоустойчивый кодер выдал ненулевой синдром. Исправленными ошибками являются такие, для которых код, выданный источником сообщений, совпал с кодом, принятым приемником сообщений, при условии, что кодовая комбинация получала искажения при прохождении канала связи. Апостериорные характеристики определяются так: доля обнаруженных ошибок - $\mathbf{D_o^* = N_o / N_ш}$, вероятность исправления ошибок – $\mathbf{P_n^* = N_и / N_o}$.

Задание к работе

Задание А.

А.1. Для кодов, полученных в лабораторной работе №3, определить априорные характеристики: избыточность, долю обнаруживаемых ошибок, долю исправляемых ошибок, вероятность исправления ошибок.

А.2. Рассчитать полезную информацию и скорость передачи информации с учётом помехоустойчивого кодирования при возникновении однократной ошибки.

А.3. В программе “SPI.exe” смоделировать прохождение сообщений по каналу связи с помехами (с использованием помехоустойчивого кодирования).

А.3.1. По результатам экспериментов п. 3 рассчитать апостериорную полезную информацию и скорость передачи информации с учетом помехоустойчивого кодирования.

А.4. Сравнить и пояснить результаты, полученные в п. А.2 и А.3.

Задание Б.

Б.1. Встроить программные модули кодирования, декодирования и локализации однократных ошибок, разработанные в лабораторной работе №3, в программу моделирования передачи сообщений, разработанную в задании Б лабораторной работы №2.

Б.2. Выполнив моделирование передачи сообщений, собрать необходимую статистическую информацию и рассчитать апостериорные характеристики кода Хемминга.

Б.3. Рассчитать полезную информацию и скорость передачи информации с учетом помехоустойчивого кодирования.

Б.4. Выполнить пп.Б.1-Б.3 для циклического кода.

Б.5. Сравнить априорные и апостериорные результаты, пояснить различия.

Контрольные вопросы

1. Как влияют значность, избыточность кода и кодовое расстояние на его корректирующую способность?

2. Что такое кратность ошибки?

3. Всегда ли обнаруженная ошибка может быть исправлена?

4. Как определить необходимое кодовое расстояние, обеспечивающее заданную корректирующую способность?

5. Что такое граница Хемминга?

Именной указатель

Предметный указатель

Библиографический список

1. Арбатский Е.В. Теория информационных процессов и систем [Электронный ресурс] / Персональная страница Арбатского Е.В. – Режим доступа: http://www.iriit.irk.ru/web-edu/~eugine/old/tipis/index.php

2. Белов В.М. Теория информации. Курс лекций: Учебное пособие для вузов / В.М. Белов, С.Н. Новиков, О.И. Солонская. – М.: Изд-во «Горячая линия-Телеком», 2012. – 144 с.

3. Вдовин В.М. Теория систем и системный анализ: Учебник / В.М. Вдовин, Л.Е. Суркова, В.А. Валентинов. – М.: Издательско-торговая корпорация «Дашков и Ко», 2010. – 640 с.

4. Вернер М. Основы кодирования: Учебник для вузов / М. Вернер: пер. с нем. – М.: Издательский центр «Техносфера», 2004. - 286 с.

5. Волкова В.Н. Искусство формализации / В.Н. Волкова; С.-Петербургский гос. политехнический ун-т.– СПб.: Изд-во СПбГПУ, 2004. – 199 с.

6. Волкова В.Н. Основы теории систем и системного анализа / В.Н. Волкова, А.А. Денисов; С.-Петербургский гос. политехнический ун-т. – СПб: Изд-во СПбГПУ, 2004. – 510 с.

7. Душин В.К. Теоретические основы информационных процессов и систем: Учебник / В.К. Душин. – М.: Издательско-торговая корпорация «Дашков и Ко», 2003. – 348 с.

8. Золотарев В.В. Помехоустойчивое кодирование. Методы и алгоритмы : справ./ В.В. Золотарев, Г.В. Овечкин – М.: Горячая линия – Телеком, 2004.

9. Информационные системы: Учебное пособие для вузов / Под ред. В.Н. Волковой, Б.И. Кузина. - СПб.: Изд-во СПбГТУ, 1998. - 213 с.

10. Красов А.В. Теория информационных процессов и систем [Электронный ресурс] / А.В. Красов; С.-Петербургский государственный электротехнический университет «ЛЭТИ», 2004. – Режим доступа: http://loge.narod.ru/tipis/

11. Ларичев О.И. Теория и методы принятия решений, а также Хроника событий в Волшебных странах / О.И. Ларичев. - М.: Логос, 2003. – 392 с.

12. Литвинская О.С. Основы теории передачи информации: учебное пособие / О.С. Литвинская, Н.И.Чернышев. – М.: КНОРУС, 2010. – 168 с.

13. Морелос–Сарагоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение / Р. Морелос-Сарагоса: пер. с англ. – М.: Издательский центр «Техносфера», 2006. – 320 с.

14. Острейковский В.А. Теория систем: Учебник для вузов / В.А. Острейковский. – М.: Высшая школа, 1997. – 240 с.

15. Петров В.Н. Информационные системы: учебник / В.Н. Петров. - СПб.: Питер, 2002. - 688 с.

16. Подчукаев В.А. Теория информационных процессов и систем / В.А. Подчукаев. – Видное: Гардарики, 2007. - 207 с.

17. Советов Б.Я. Моделирование систем: Учебник для вузов / Б.Я. Советов, С.А. Яковлев. – М.: Высшая школа, 1998. – 295 с.

18. Стариченко Б.Е. Теоретические основы информатики [Электронный ресурс] / Б.Е. Стариченко, 2003. – Режим доступа: http://www.de.uspu.ru/ Informatics/ Metodes/DPP/ F/08/1/index.htm

19. Стариченко Б.Е. Теоретические основы информатики: учеб. пособие для студентов пед. вузов / Б.Е. Стариченко. – М. : Горячая линия –Телеком, 2003 – 256 с.

20. Тартаковский Г.П. Теория информационных систем / Г.П. Тартаковский. – М.: Физматкнига, 2005. – 304 с.

21. Теория информации и кодирование / Б.Б.Самсонов, Е.М. Плохов, А.И. Филоненков и др.- Ростов н/Д.: Феникс, 2002.- 287 с.

22. Теория информационных процессов и систем: Учебник для студентов высших учебных заведений. Под ред. Б.Я.Советова / Б.Я.Советов, В.В. Цехановский, В.А. Дубенецкий. – М.: Академия, 2010. – 432 с.

23. Теория систем и системный анализ [Электронный ресурс] / Тематический сайт. – Режим доступа: http://www.tsisa.ru

24. Яглом А.М. Вероятность и информация / А.М. Яглом, И.М. Яглом. - М.: Наука, 1973. – 512 с.

25. MIT Open Course Ware. Information and Entropy [Электронный ресурс] : открытая библиотека учебных курсов Массачусетского технологического института. – MIT, 2008. – Режим доступа: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-050j-information-and-entropy-spring-2008/