Теория информации

- ЧТО ТАКОЕ ИНФОРМАЦИЯ?
- КАК ОНА ВОЗНИКАЕТ?
- КАК МЫ УЗНАЕМ ИНФОРМАЦИЮ?
- КАКИМИ ОНА ОБЛАДАЕТ СВОЙСТВАМИ?
- МОГУТ ЛИ ЖИВОТНЫЕ ВОСПРИНИМАТЬ ИНФОРМАЦИЮ?
- МОГУТ ЛИ КОМПЬЮТЕРЫ СОЗДАВАТЬ ИНФОРМАЦИЮ?
- …….??????????????

ЧТО ТАКОЕ ИНФОРМАЦИЯ? КАК ОНА ВОЗНИКАЕТ? КАК МЫ УЗНАЕМ ИНФОРМАЦИЮ? КАКИМИ ОНА ОБЛАДАЕТ СВОЙСТВАМИ? ....

- Человеческая речь?
- Пение птиц?
- Музыка симфонического оркестра?
- Картинная галерея?
- Спортивный матч?
- Учебник по общей теории связи?
- Разговор с соседями?
- Телевизионная передача?

ЧТО ТАКОЕ ИНФОРМАЦИЯ? КАК ОНА ВОЗНИКАЕТ? КАК МЫ УЗНАЕМ ИНФОРМАЦИЮ? КАКИМИ ОНА ОБЛАДАЕТ СВОЙСТВАМИ? ....

- **Вам сообщили, что 2+2=4. А ещё Вам сказали, что Вы выиграли в лотерею 1 млн. рублей. Как Вы оцените полученную информацию? В каком случае ее будет больше?**
- **У Вас было яблоко. Вы его отдали соседу. У соседа есть яблоко, а у Вас нет. У Вас была информация. Вы сообщили ее соседу. Она есть и у Вас и у него. Яблоко осталось одно. А что стало с**

Теория информации

№	Информация	Ценность	Вероят.
1.	С завтрашнего дня стипендия студентов увеличится в 10 раз		
2.	Вы играли в лотерею и Вам сообщают, что Вы выиграли 1 млн. рублей		
3.	У Вашего соседа сегодня день рождения		
4.	В сентябре 30 дней		
5.	Сборная России по футболу на чемпионате мира займет второе место		

Теория информации

Теория информации

- ***Информация*** *- это объективно существующее неотъемлемое* ***свойство*** *объектов, процессов, явлений, отражающее их внутренние и внешние особенности и разнообразие в различных* ***метрических*** *или* ***топологических*** *пространствах.*
- ***Информация*** *отображается в форме (в виде)* ***сообщений****, имеющих ту или иную материальную основу.*

Теория информации

- ***Сообщения*** *преобразуются в* ***сигналы****, которые непосредственно передаются, например, по системе связи.*
- ***Сигналы*** *могут многократно преобразовываться в другие* ***сигналы*** *с целью согласования со средой передачи.*
- ***Информация*** → ***Сообщение*** → ***Сигнал*** →***…***→ ***Сигнал*** → ***Сообщение*** → ***Информация.***

Теория информации

- **Метри́ческим простра́нством** называется множество, в котором между любой парой элементов определено обладающее заранее оговоренными свойствами расстояние, называемое *ме́трикой*. Например, расстояние между точками A, B и C на плоскости: (A,B), (A,C) и (B,C)

A • • B

C •

Теория информации

- **Тополōги́ческое простра́нство** — множество с дополнительной структурой определённого типа (так называемой топологией) или другими словами – это совокупность двух объектов: множества ***X***, состоящего из элементов произвольной природы, называемых точками данного пространства, и из введенной в это множество топологической структуры, или топологии.

Теория информации

- **Множество** - набор, совокупность, собрание каких-либо объектов, называемых его элементами, обладающих общим для всех их характерным свойством.
- **Расстоя́ние**, в широком смысле, степень удаленности или отличий объектов друг от друга.

Теория информации

- **Математи́ческая структу́ра** — название, объединяющее понятия, общей чертой которых является их применимость к множествам, природа которых не определена. Для определения самой структуры задают отношения, в которых находятся элементы этих множеств. Затем постулируют, что данные отношения удовлетворяют неким условиям, которые являются аксиомами рассматриваемой структуры.

Теория информации

- **Отноше́ние** в теории множеств — математическая структура, которая формально определяет свойства различных объектов и их взаимосвязи. Распространёнными примерами отношений в математике являются равенство, делимость, подобие, параллельность и т.д. Наглядно теоретико-множественное отношение можно представить в виде таблицы, каждая строка которой содержит конкретные примеры объектов, связанных данным отношением. Например, телефонный справочник можно рассматривать как отношение, отражающее связь между следующими объектами: **Телефон** ↔ **ФИО абонента** ↔ **Почтовый адрес** ↔ **…** Отношения обычно классифицируются по количеству связываемых объектов и собственным свойствам (симметричность, транзитивность и пр.).

Теория информации

Отношение R симметрично, если

□ □□, □□ □ □□, □□Rb □ bRa

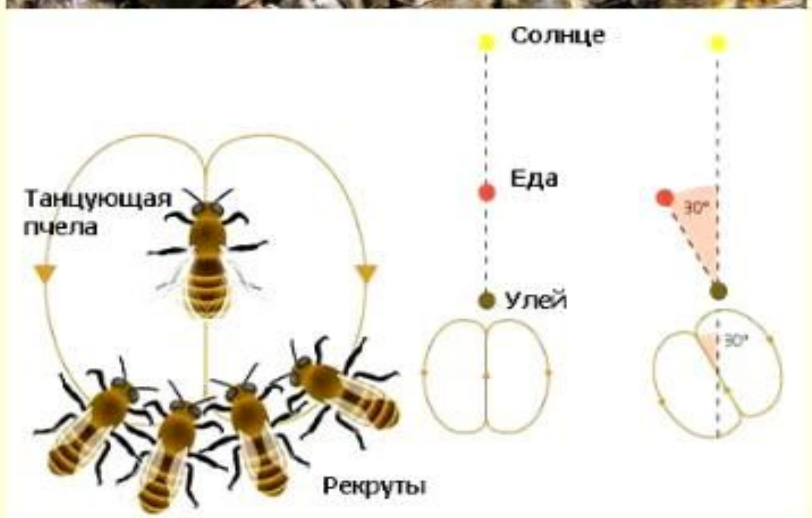
Отношение R транзитивно, если

□ □□, □□, □□ □

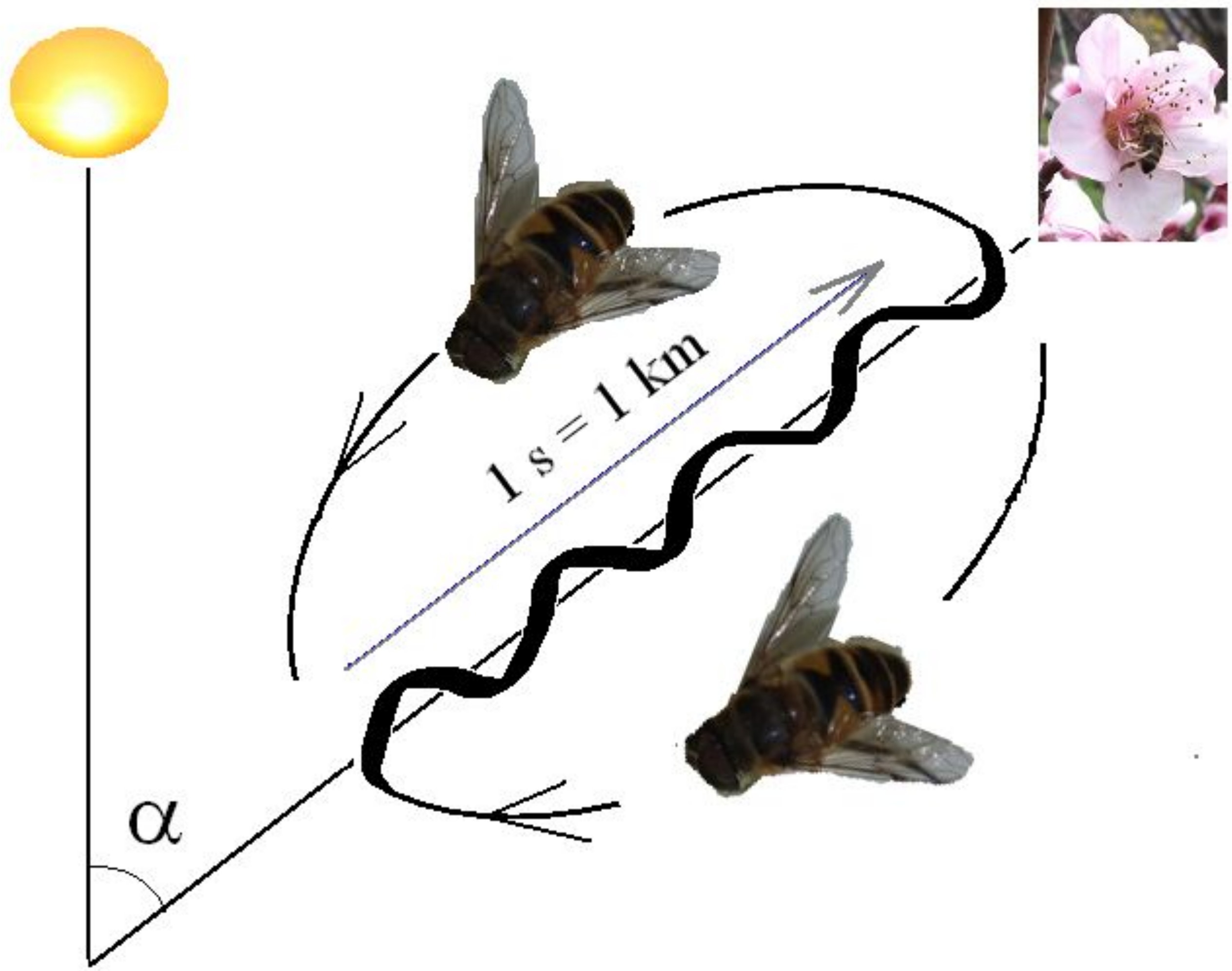
□□, *при* □□Rb *и* bRc будет выполняться □□Rc.

Теория информации

«Танцы пчел»



Теория информации



Теория информации

- **Курсовая работа**
- **1. Построить код Хаффмана для букв русского алфавита.**
- **2. Выбрав ключи, закодировать фамилию, имя и отчество кодом Цезаря и далее, используя «квадрат» Виженера.**
- **3. Используя код п.1 закодировать свои фамилию, имя и отчество, полученные из п.2.**
- **4. Для двух заданных источников дискретных сообщений рассчитать энтропию объединения. (Индивидуальное задание по номеру в журнале).**

Номер		[illegible]	[illegible]	[illegible]	[illegible]
1	[illegible]	0,5	0,25	0,125	0,125
2	[illegible]	0,5	0,125	0,25	0,125
3	[illegible]	0,5	0,125	0,125	0,25
4	[illegible]	0,25	0,5	0,125	0,125
5	[illegible]	0,125	0,5	0,25	0,125
6	[illegible]	0,125	0,5	0,125	0,25
7	[illegible]	0,25	0,125	0,5	0,125
8	[illegible]	0,125	0,25	0,5	0,125
9	[illegible]	0,125	0,125	0,5	0,25
10	[illegible]	0,25	0,125	0,125	0,5
11	[illegible]	0,125	0,25	0,125	0,5
12	[illegible]	0,125	0,125	0,25	0,5
13	[illegible]	0,25	0,25	0,25	0,25
14	[illegible]	0,625	0,125	0,125	0,125
15	[illegible]	0,125	0,625	0,125	0,125
16	[illegible]	0,125	0,125	0,625	0,125
17	[illegible]	0,125	0,125	0,125	0,625
18	[illegible]	0,375	0,25	0,25	0,125
19	[illegible]	0,375	0,25	0,125	0,25
20	[illegible]	0,375	0,125	0,25	0,25
21	[illegible]	0,25	0,375	0,25	0,125
22	[illegible]	0,25	0,375	0,125	0,25
23	[illegible]	0,125	0,375	0,25	0,25
24	[illegible]	0,25	0,25	0,375	0,125
25	[illegible]	0,25	0,125	0,375	0,25
26	[illegible]	0,125	0,25	0,375	0,25
27	[illegible]	0,25	0,25	0,125	0,375
28	[illegible]	0,25	0,125	0,25	0,375
29	[illegible]	0,125	0,25	0,25	0,375
30	[illegible]	0,375	0,375	0,125	0,125
31	[illegible]	0,375	0,125	0,375	0,125
32	[illegible]	0,375	0,125	0,125	0,375

Теория информации

Матрица условных вероятностей [illegible]

	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	0,25	0,25	0,125	0,125	0,125	0,125
[illegible]	0,25	0,125	0,25	0,125	0,125	0,125
[illegible]	0,125	0,25	0,125	0,25	0,125	0,125
[illegible]	0,125	0,125	0,25	0,125	0,25	0,125

Теория информации

ТЕМЫ РЕФЕРАТОВ:

1. История возникновения и развития информационных устройств.
2. История и современность в первичных способах кодирования сообщений.
3. Аналого-цифровые методы преобразований сигналов.
4. Информатизация современного общества.
5. Информационная безопасность, проблемы и пути решения.
6. Информационные технологии в медицине.
7. Информационные технологии в образовании.

Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики»

Кафедра «Общей теории связи»

КУРСОВАЯ РАБОТА

Выполнил: студент группы (указать номер группы. Фамилию, имя и отчество не указывать, но написать:

Фамилия………………….

Имя………………………..

Отчество…………………..

Фамилия имя и отчество будут заполняться при защите курсовой работы и декодировании выполненных заданий)

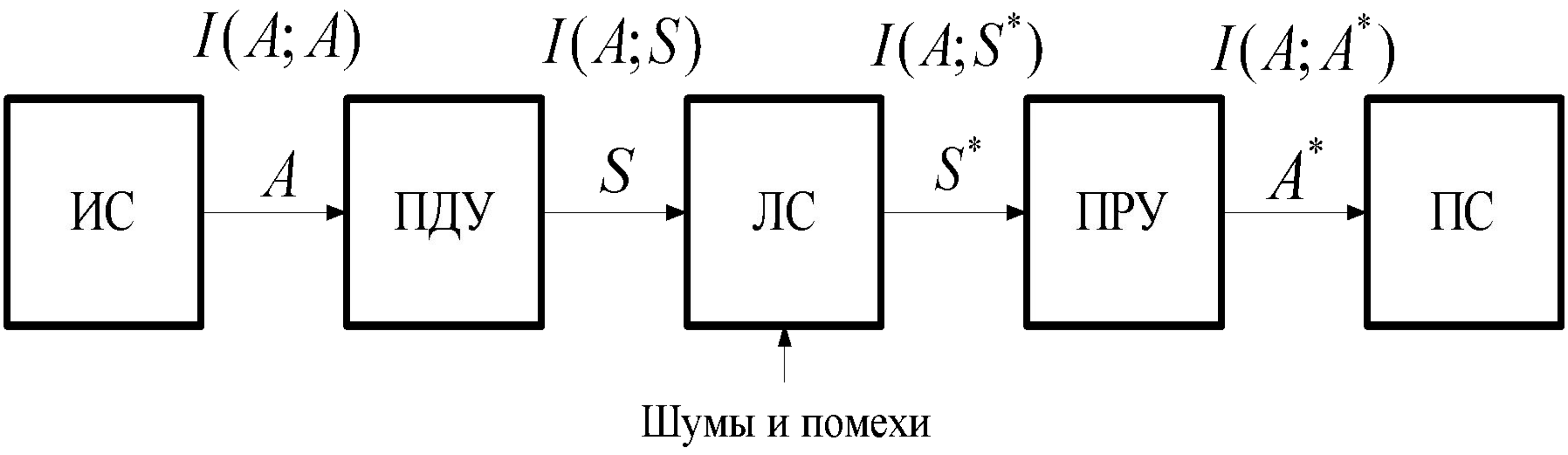
Проверил: (указать должность, фамилию имя и отчество преподавателя)

Москва, (указать год)

- **Фамилия, Имя, Отчество в виде кодовой последовательности кодом Хаффмана ……**
- **Ключ «Цезаря» ……………………………………….**
- **Ключ «Виженера» ……………………………………..**
- **Квадрат «Виженера» …………………………………..**
- **Код Хаффмана (дерево и таблица)……………**
- **Исходные данные и расчет энтропии**

Теория информации

Блок-схема системы передачи информации (по К. Шеннону)




Бара-ба-бры жик-жик…..

Теория информации

- **В процессе передачи по СПИ сообщение подвергается многочисленным преобразованиям, существенно меняющим его электрическое представление и физические характеристики. Например, при передаче речевого сообщения по радиоканалу человек как источник информации с помощью голосового аппарата формирует звуки речи вначале на акустическом уровне. В микрофоне звуковое давление преобразуется в электрический ток. В модуляторе передатчика этот низкочастотный ток (первичный сигнал) преобразуется в высокочастотное напряжение (сигнал), которое затем на выходе антенной системы представляется в виде электромагнитной волны (поля), распространяющейся по радиолинии. Принятая радиоволна (поле) с помощью антенной системы преобразуется в высокочастотное напряжение (принятый сигнал), которое в приемнике преобразуется в низкочастотное напряжение (или ток) – принятое сообщение, и, наконец, с помощью телефона или громкоговорителя это сообщение преобразуется в акустические волны, воспринимаемые слуховой системой человека как получателя информации**

Теория информации

Надо иметь в виду, что в любой системе связи, используемой человеком, конечной целью передачи является не само сообщение, а та информация, которая в нем содержится. Эта информация определяется источником (человеком) и называется полезной информацией. Она не зависит (инвариантна) от физической формы представления сообщения, и, следовательно, сообщение, переданный и принятый сигналы, а также принятое сообщение в идеале должны содержать одно и то же количество информации, вырабатываемое источником. В реальных условиях количество передаваемой источником информации уменьшается из-за действия помех и искажений в различных блоках системы связи.

Теория информации

В теории информации это понятие носит более утилитарный характер. Под *информацией* понимают любые *сведения* о состоянии или поведении некоторого объекта или системы, либо о каких-то событиях, явлениях, предметах, подлежащие передаче от ИС к ПС. При таком определении источником информации уже может быть не только человек, но и компьютер, телеметрический датчик и т.д. При этом можно отметить, что если получатель априори (до передачи) достоверно знает, что будет передано от источника, то количество получаемой им информации нулевое и такая передача (связь) бессмысленна (нецелесообразна), ведь все заранее известно. Поэтому информация, воспринимаемая ПС, будет отличаться от нуля только в случае, если передаваемые сведения являются для него новыми, непредвиденными. Именно эта информация должна оставаться инвариантной при всех преобразованиях сообщений в СПИ; *сообщения и сигналы являются только физическими носителями информации*.

Теория информации

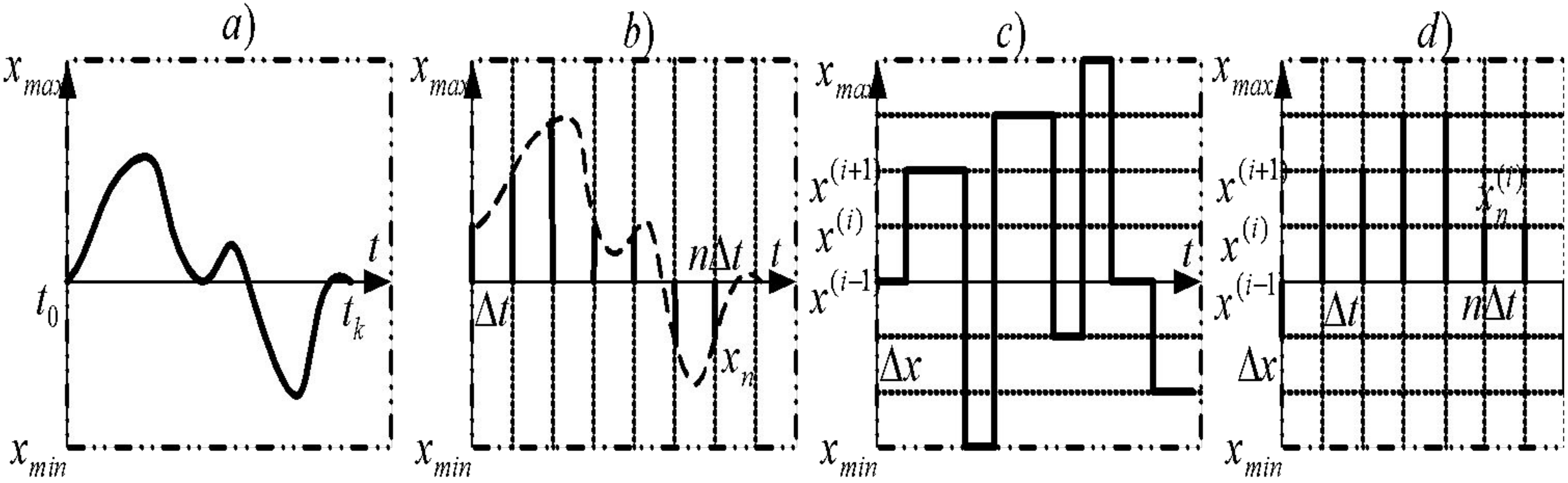
В СПИ сообщение $A \in \mathbf{A}$, выбираемое из множества $\mathbf{A}$, обладает информацией $I(A;A)$, называемой ***собственной информацией***, содержащейся в сообщении A относительно исходного сообщения A. Отклик ПДУ $S \in \mathbf{S}$ переносит информацию $I(A;S)$, называемую ***взаимной информацией***, содержащуюся в сигнале S относительно сообщения A. Аналогично $I(A;S^*)$ - взаимная информация, содержащаяся в сигнале $S^* \in \mathbf{S}^*$, наблюдаемого на приеме относительно сообщения A, а $I(A;A^*)$ - взаимная информация, содержащаяся в восстановленном сообщении $A^* \in \mathbf{A}^*$ относительно передаваемого сообщения A. Поскольку информация при преобразованиях сообщений и сигналов не может увеличиться, то в любой СПИ справедливы следующие неравенства

$$I(A;A) \geq I(A;S) \geq I(A;S^*) \geq I(A;A^*).$$

Равенства достигаются лишь в идеальном случае, когда в СПИ нет ни искажений, ни шумов, ни помех. В этом случае вся информация, вырабатываемая ИС поступает к ПС. В реальном случае в СПИ имеются потери информации. (Следует еще раз подчеркнуть, что в данном случае речь идет об информации, определенной в утилитарном смысле. При более общем определении процесс измерения количества информации приобретает более сложный вид).

Теория информации

Виды сообщений или сигналов



- непрерывный (аналоговый), - дискретно-непрерывный, - непрерывно-дискретный, - дискретный (цифровой).

Теория информации

- **Сообщение состоит из слов.**
- **Слово состоит из символов (букв).**
- **Символы принимают значение из некоторого алфавита.**
- **Например, сообщение: «Добрый день» состоит из двух слов. Первое слово состоит из 6 символов (букв), а второе – из 4 символов (букв). Каждый символ (буква) берется из алфавита русского языка,**

Теория информации

- **В общем виде слово состоит из (n) символов (букв), каждая из которых может принимать одно из (m) значений алфавита.**
- **Общее количество различных слов равно:**
  $N=m^n$.
- **Логично предположить, что чем больше N, тем больше информации.**
- **Если один источник порождает N_1 слов, второй N_2 и т.д., а r-ый N_r, то всего будет:**
  $N = N \; N \; \; N$

Теория информации

Впервые логарифмическую меру информации ввел *Хартли:*

[illegible]

Для [illegible] источников будет выполняться условие аддитивности:

[illegible]

При [illegible] информация оценивается в *натах,* при [illegible] в *битах*, при [illegible] информация оценивается в *дитах*. В инженерно-технических приложениях теории информации чаще всего используется единица *бит* и её производные: *байт* ($8=2^3$ бит), *килобит* ($1024=2^{10}$ бит), *килобайт* ($8192=2^{13}$ бит), *мегабит* ($1048576=2^{20}$) и т.д.

Теория информации

Удельное количество информации слова называют его энтропией:

$$[illegible]$$

Реальные источники вырабатывают символы с разной вероятностью, что не учитывается у *Хартли.* Поэтому *К.Шеннон* предложил энтропию слова ([illegible]), вырабатываемого с вероятностью [illegible]([illegible]), определить как

$$[illegible]$$

Эта частная энтропия является величиной случайной. Ее среднее значение – энтропия источника:

$$[illegible]$$

Теория информации

Энтропия максимальна и равна: $H_{max} = \log_b N$

Для доказательства этого воспользуемся методом неопределенных множителей Лагранжа. С учетом условия нормировки, когда $\sum_{i=1}^{N} p_i = 1$ и, полагая $b = e$, найдем экстремум функционала следующего вида:

$$F(p_1, p_2, \dots p_N) = -\sum_{i=1}^{N} p_i \ln p_i + \lambda \sum_{i=1}^{N} p_i$$

Дифференцируя по p_l, и приравнивая производные нулю, получаем

$$\frac{\partial}{\partial p_l} F(p_1, p_2, \dots p_N) = \frac{\partial}{\partial p_l}(-p_l \ln p_l + \lambda p_l) = -\ln p_l - 1 + \lambda = 0$$

Отсюда следует, что $\ln p_l = \lambda - 1$ и не зависит от номера l, что может быть только при равенстве всех p_l, а значит $p_l = 1/N$.

Подставляя эту величину, вычислим:

$$H = -\sum_{i=1}^{N} p_i \log_b p_i = -\sum_{i=1}^{N} \frac{1}{N} \log_b \frac{1}{N} = \log_b N$$

Теория информации

Энтропия есть величина вещественная и неотрицательная, а так же ограниченная.

Первые два утверждения следуют из определения энтропии и очевидного неравенства: [illegible] ≤ [illegible] ≤ [illegible].

Для доказательства третьего рассмотрим слагаемое – [illegible]. При [illegible] = [illegible] результат очевиден. При [illegible] имеем – [illegible] = [illegible] ∞. Раскроем неопределенность по правилу Лопиталя, положив [illegible] = [illegible]:

[illegible] – [illegible] = – [illegible] = – [illegible] ([illegible]) = – [illegible] = [illegible] = [illegible].

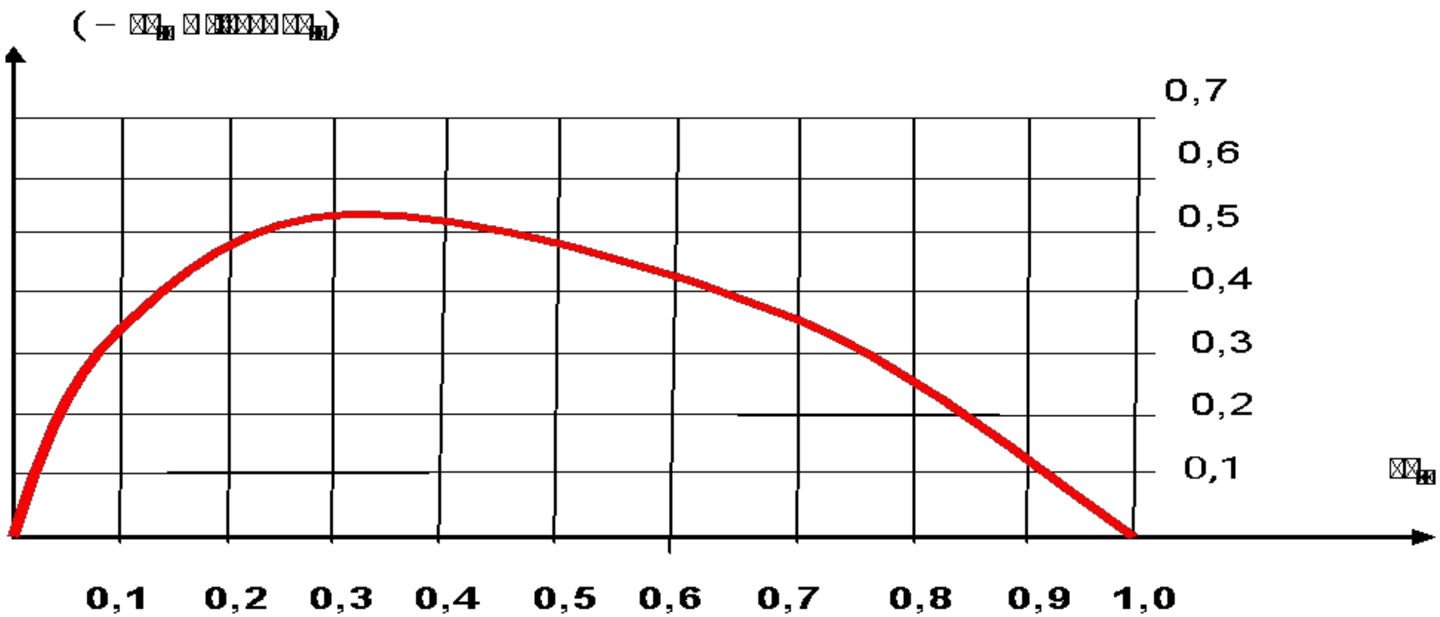
Теория информации

Величина (− [illegible]) принимает максимальное значение при [illegible] = [illegible]. Докажем.

$$[illegible]$$

Следовательно, [illegible] = [illegible], а [illegible] = [illegible]. Построим график зависимости (− [illegible]) от [illegible]. Максимум равен

$$[illegible]$$



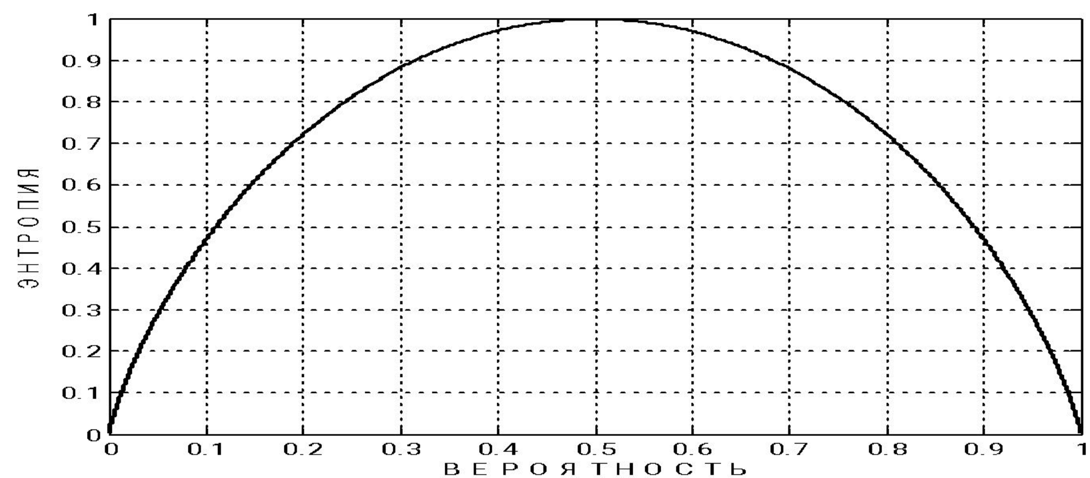
Теория информации

Энтропия двоичного ансамбля.

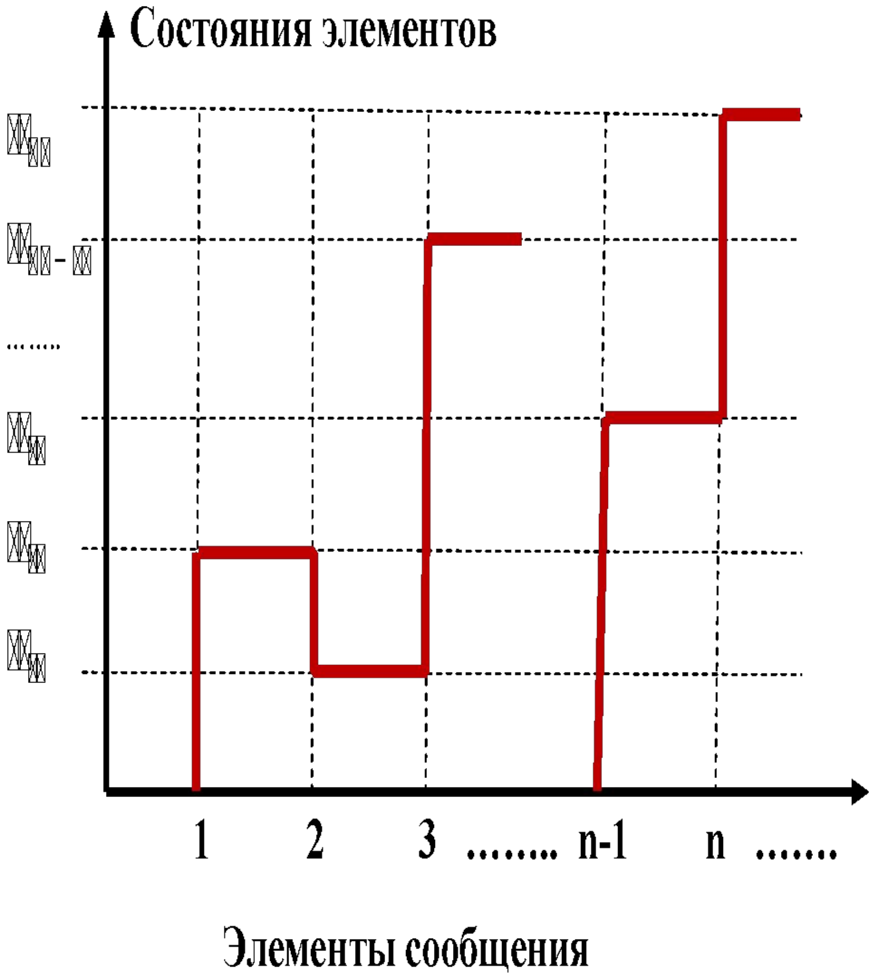
Пусть $m = 2$; $p_1 = p$; $p_2 = (1 - p)$. Тогда

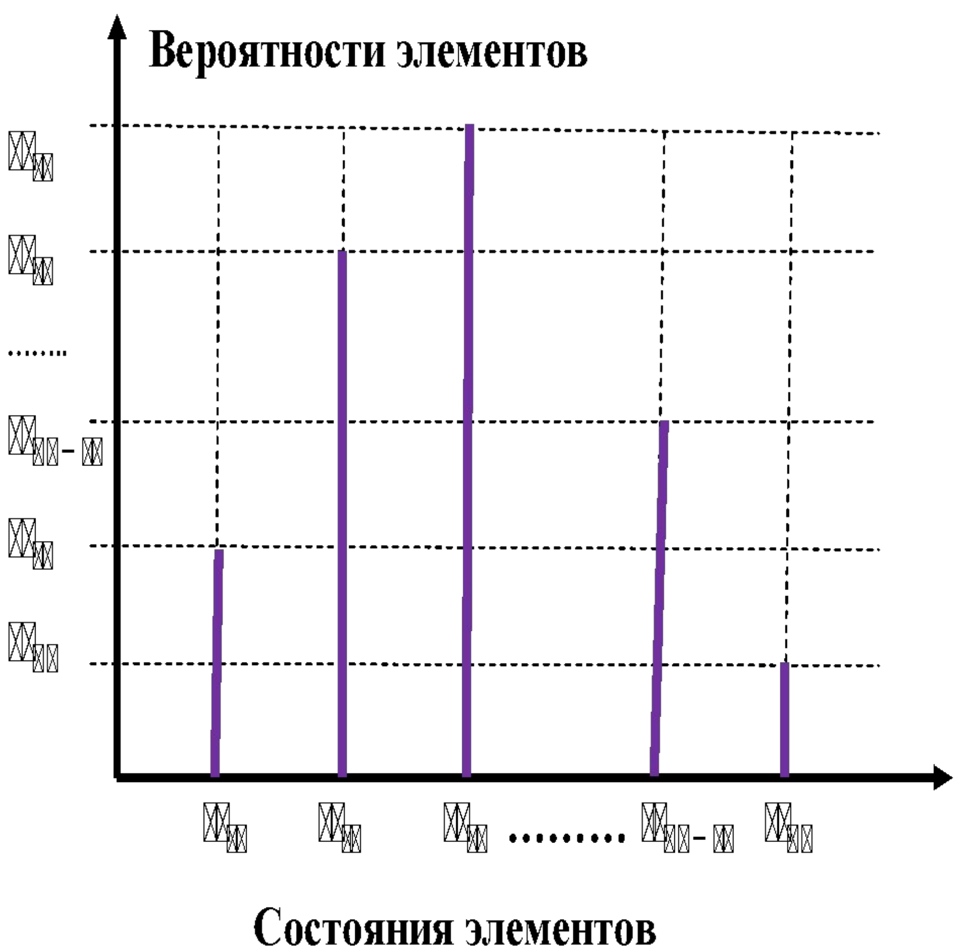
$$H = -p \log_2 p - (1 - p) \log_2 (1 - p).$$

На рисунке показана зависимость энтропии H от вероятности p двоичного ансамбля. Видно, что $H = 0$ при $p = 0$ и $p = 1$; $H_{max} = \log_2 2 = 1$ достигается при $p = 0,5$. Размерность энтропии: [бит/символ].

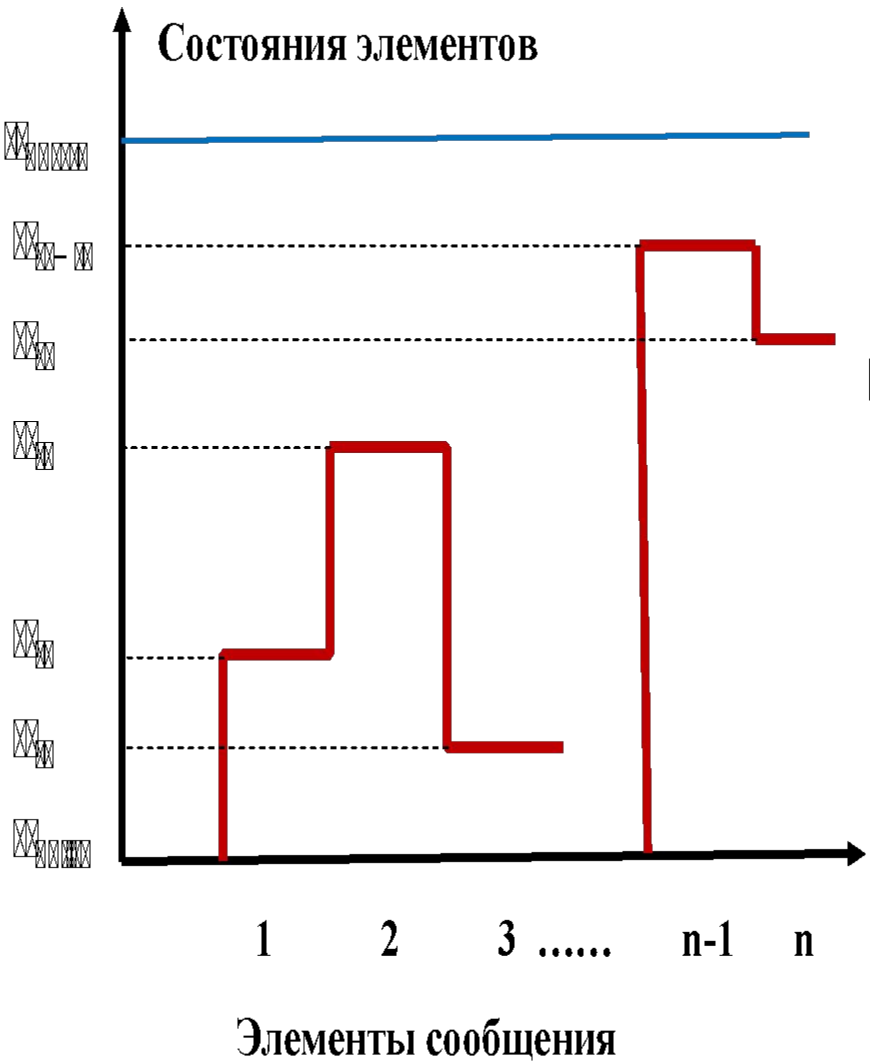


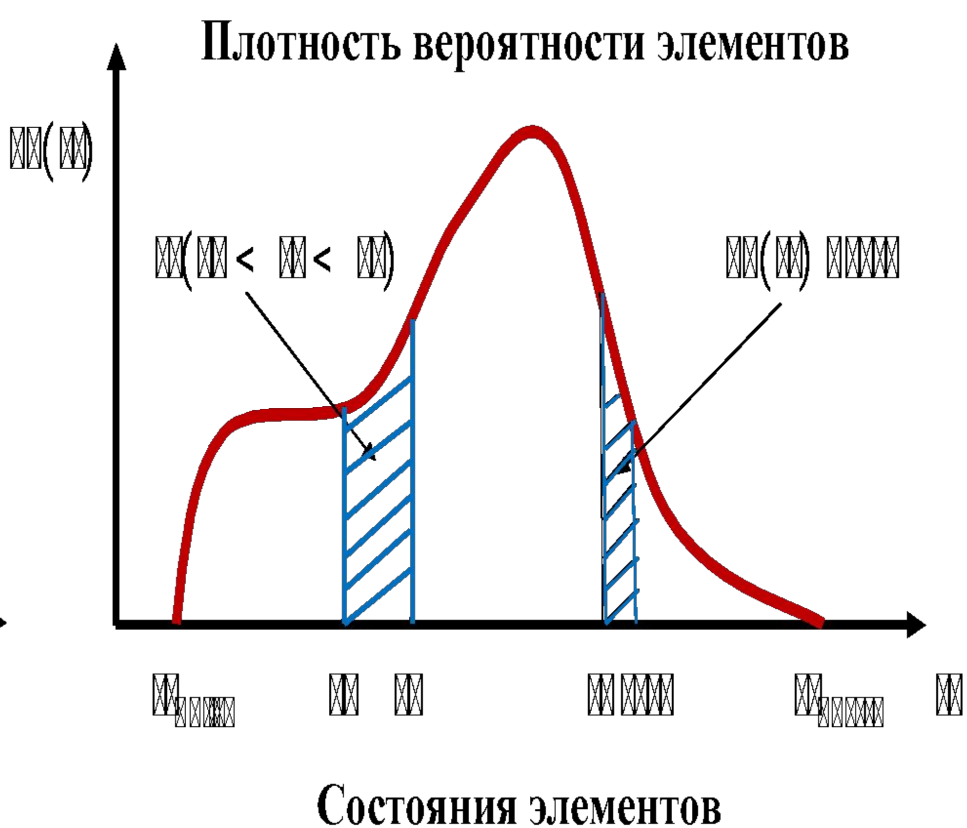
Теория информации





Теория информации





Теория информации

$$H = -\sum_{i=1}^{N} p_i \cdot \log p_i \qquad\qquad p_i = p(x_i)\,\Delta x$$

$$H = -\sum_{i=1}^{N} p(x_i)\,\Delta x \cdot \log\left(p(x_i)\,\Delta x\right) =$$

$$= -\sum_{i=1}^{N} p(x_i)\,\Delta x \cdot \log p(x_i) - \sum_{i=1}^{N} p(x_i)\,\Delta x \cdot \log \Delta x$$

Первое слагаемое при $\Delta x \to 0$ равно: $-\int_{-\infty}^{\infty} p(x) \cdot \log p(x)\, dx$

Т.к. при $\Delta x \to 0$ имеем $\int_{-\infty}^{\infty} p(x)\, dx = 1$, то второе слагаемое равно: $-\log \Delta x$.

Теория информации

Дисперсия состояний элементов сообщения ограничена, а среднее равно 0. Тогда при

$\int_{-\infty}^{\infty}$ [illegible] $=$ [illegible] **и** $\int_{-\infty}^{\infty}$ [illegible] $=$ [illegible] **энтропия**

$$[illegible] = - \int_{-\infty}^{\infty} [illegible]([illegible])\,[illegible][\,[illegible]([illegible])\,[illegible]\Delta[illegible]\,]\,[illegible]$$

максимальна при нормальном центрированном распределении

$$[illegible] = \frac{[illegible]}{[illegible]\,\xi\,\overline{[illegible]}}\,[illegible]^{-\frac{[illegible]}{[illegible]}}$$

и равна [illegible] $=$ [illegible] $\frac{[illegible]}{\Delta [illegible]}\,\xi\,\overline{[illegible]}$

Теория информации

Дисперсия состояний элементов сообщения не ограничена. Тогда при

$$\int_{-\infty}^{\infty} p(x)\,dx = 1 \quad H = \infty \text{ энтропия}$$

$$H = -\int_{-\infty}^{\infty} p(x)\log\left[p(x)\,\Delta x\right]dx$$

максимальна при равномерном распределении на интервале $[a, b]$

$$p(x) = \frac{1}{b-a}$$

и равна $H = \log\frac{b-a}{\Delta x}$

Теория информации

	[illegible]	[illegible]	…	[illegible]	…	[illegible]
[illegible]	[illegible]	[illegible]	…	[illegible]	…	[illegible]
[illegible]	[illegible]	[illegible]	…	[illegible]	…	[illegible]
…	…	…	…	…	…	…
[illegible]	[illegible]	[illegible]	…	[illegible]	…	[illegible]
…	…	…	…	…	…	…
[illegible]	[illegible]	[illegible]	…	[illegible]	…	[illegible]

Теория информации

$p(a_i)$ – **вероятность элемента** a_i **в сообщении A**

$p(b_j)$ – **вероятность элемента** b_j **в сообщении B**

Частная условная энтропия:

$$H(B|a_i) = -\sum_{j=1}^{m} p(b_j|a_i)\log_2 p(b_j|a_i).$$

Общая условная энтропия:

$$H(B|A) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(a_i)\,p(b_j|a_i)\log_2 p(b_j|a_i).$$

Известно, что:

$$p(a_i, b_j) = p(a_i)\,p(b_j|a_i) = p(b_j)\,p(a_i|b_j)$$

Тогда:

$$H(B|A) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(a_i, b_j)\log_2 p(b_j|a_i).$$

Теория информации

Энтропия объединения статистически зависимых сообщений:

[illegible]

Можно показать, что:

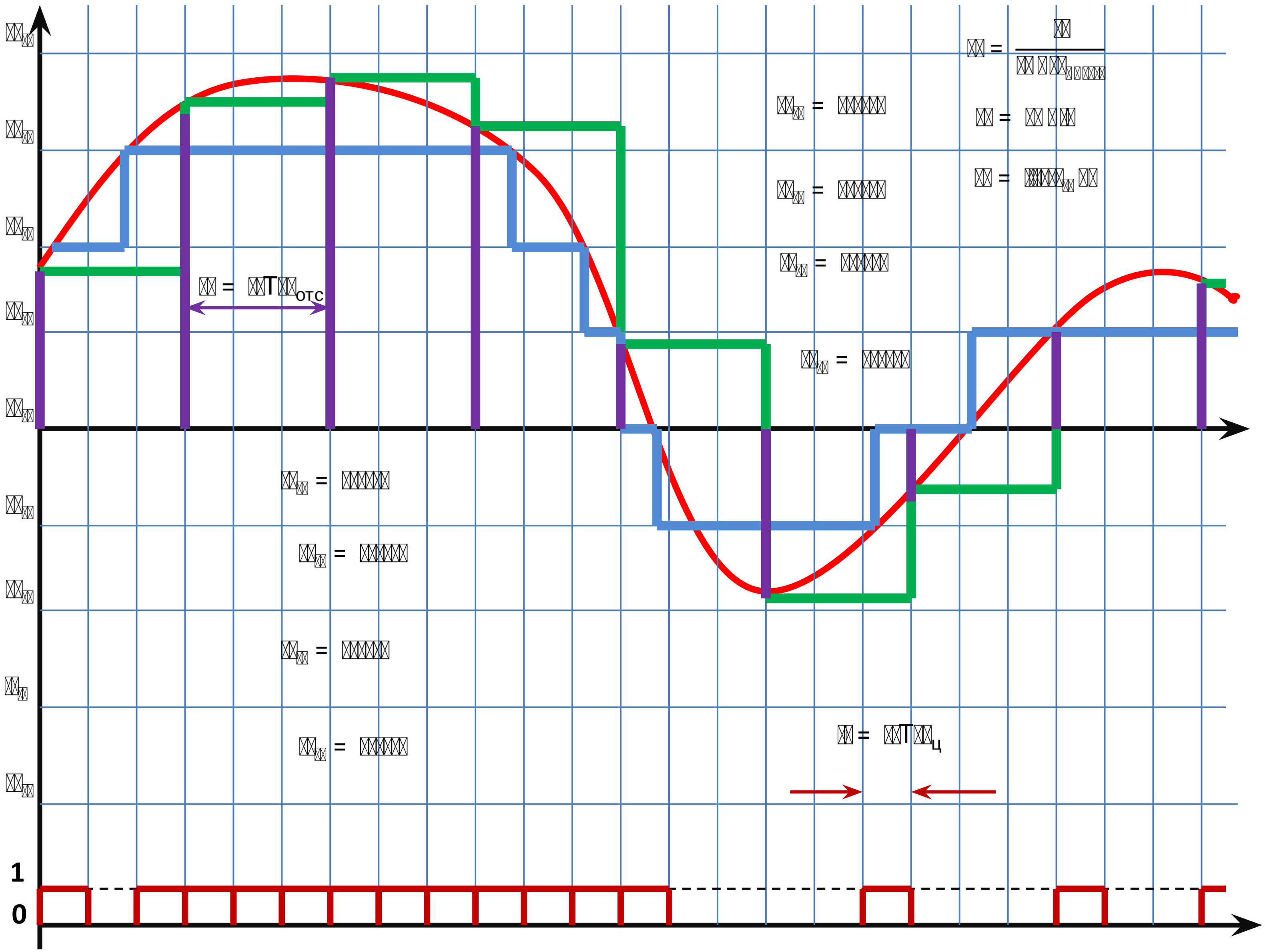
[illegible]

При статистически независимых сообщениях:

[illegible], т.к. [illegible].

При полной статистической зависимости

[illegible] и [illegible].


1
0

Теорема Котельникова

Любая непрерывная функция $x(t)$ с ограниченным (финитным) спектром может быть представлена своими отсчетами $x_k = x(t_k) = x(k\Delta t)$, $k = 0, \pm 1, \pm 2, \ldots,$ взятыми в моменты времени t_k, отстоящими друг от друга на интервал времени Δt (интервал дискретизации), где

$$\Delta t \le \frac{\pi}{\omega_{max}} = \frac{1}{2 \cdot F_{max}}$$

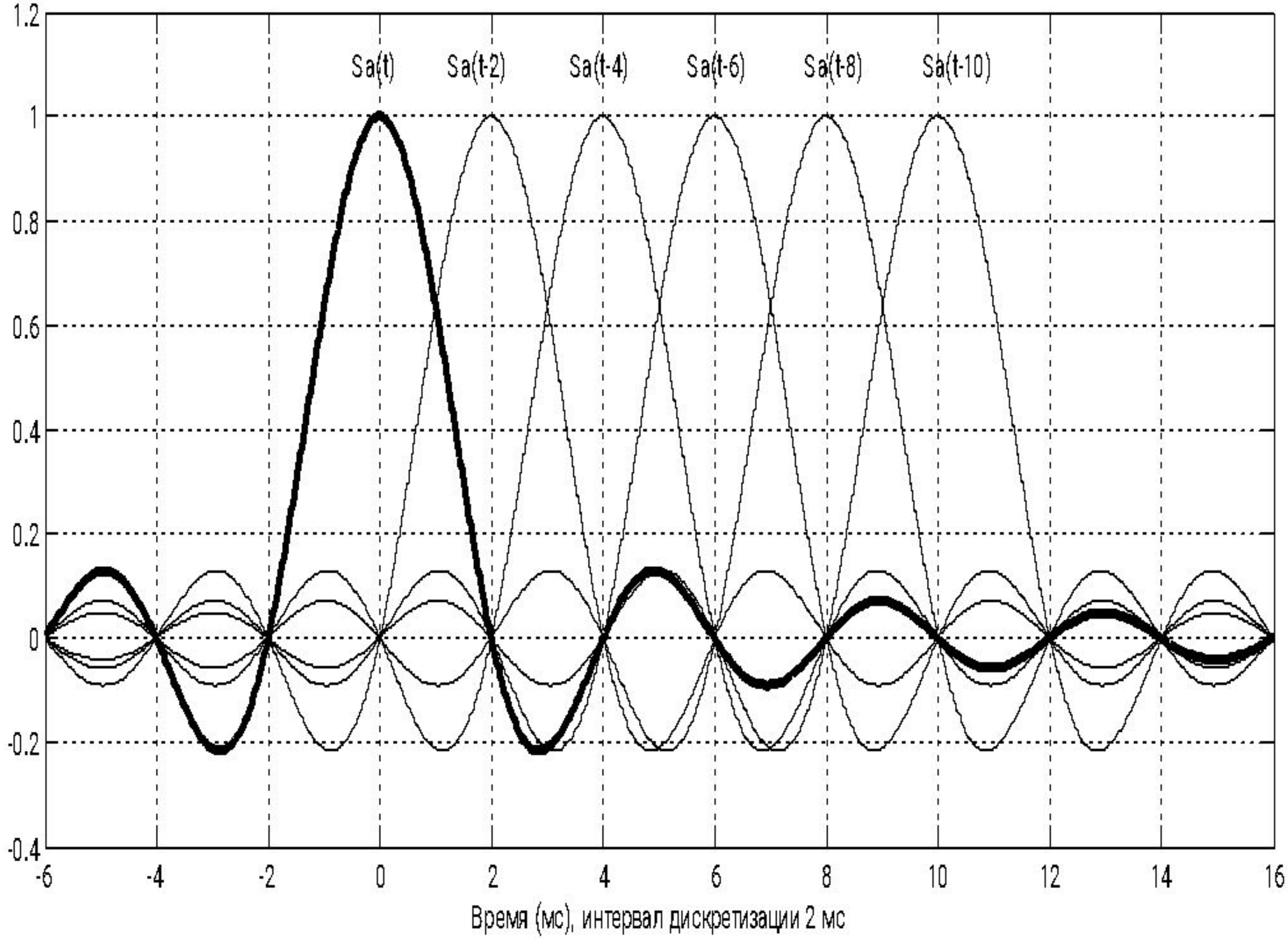
В итоге сигнал представляется в виде ортогонального ряда Котельникова

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t) \cdot \frac{\sin \omega_{max}(t - k\Delta t)}{\omega_{max}(t - k\Delta t)},$$

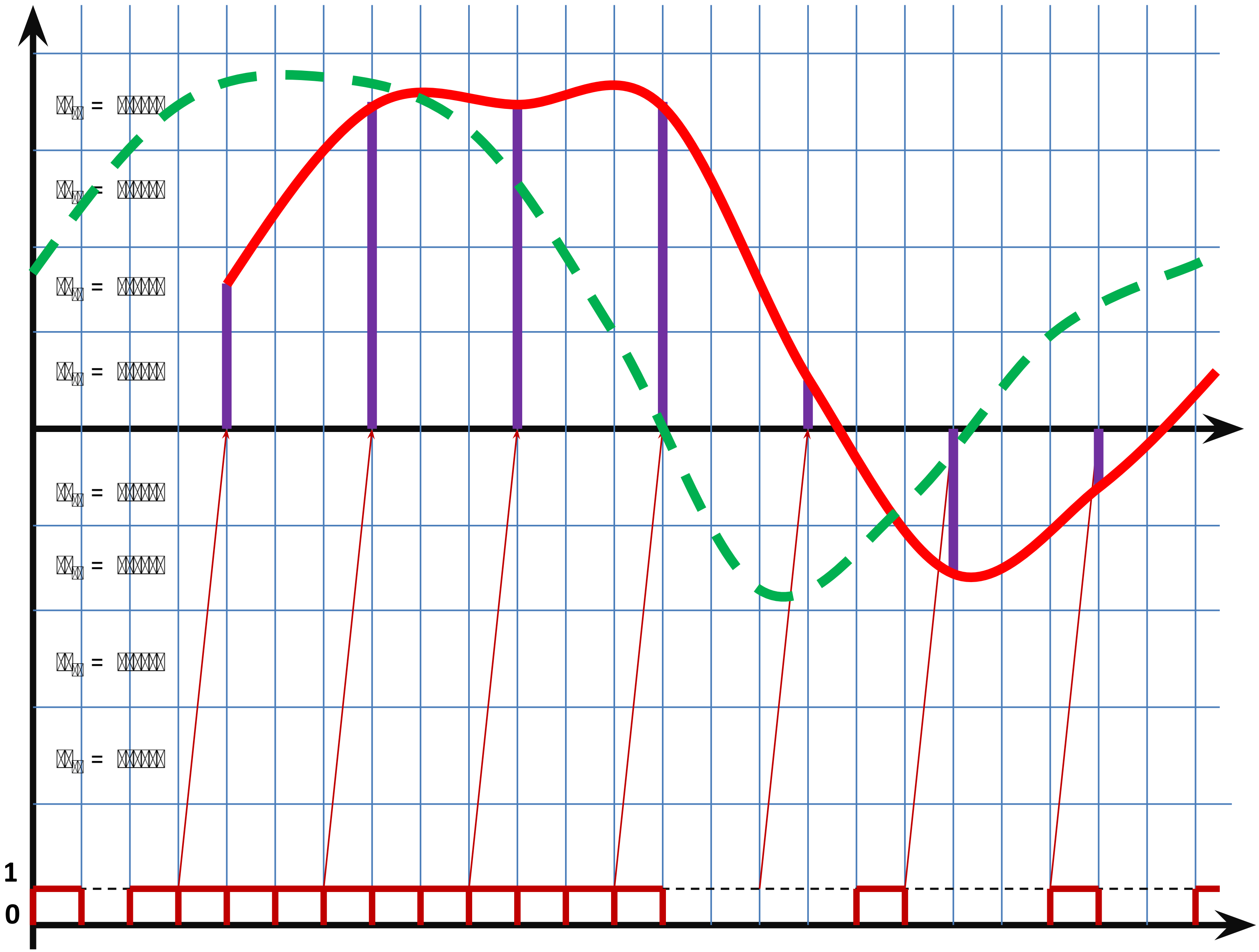
что соответствует дискретной свертке во временной области последовательности отсчетов $\{x_k\}$ с ортогональными функциями отсчетов $\varphi_k(t) = \mathrm{sinc}\,\omega_{max}(t - k\Delta t)$.

Это можно записать:

$$x(t) = \sum_{k=-\infty}^{\infty} x_k \cdot \mathrm{sinc}\,\omega_{max}(t - k\Delta t)$$



Sa(t)
Sa(t-2)
Sa(t-4)
Sa(t-6)
Sa(t-8)
Sa(t-10)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-6
-4
-2
0
2
4
6
8
10
12
14
16
Время (мс), интервал дискретизации 2 мс


1
0

В реальных условиях обработки аналоговых сигналов их полностью неискаженное восстановление по дискретным отсчетам не возможно по следующим причинам:

1. Сигналы с ограниченным спектром можно наблюдать только на бесконечном временном интервале. В реальных условиях этот интервал ограничен, что приводит к расширению спектра сигнала до бесконечности. Это следует из свойства преобразования Фурье для сигналов, ограниченных во времени.

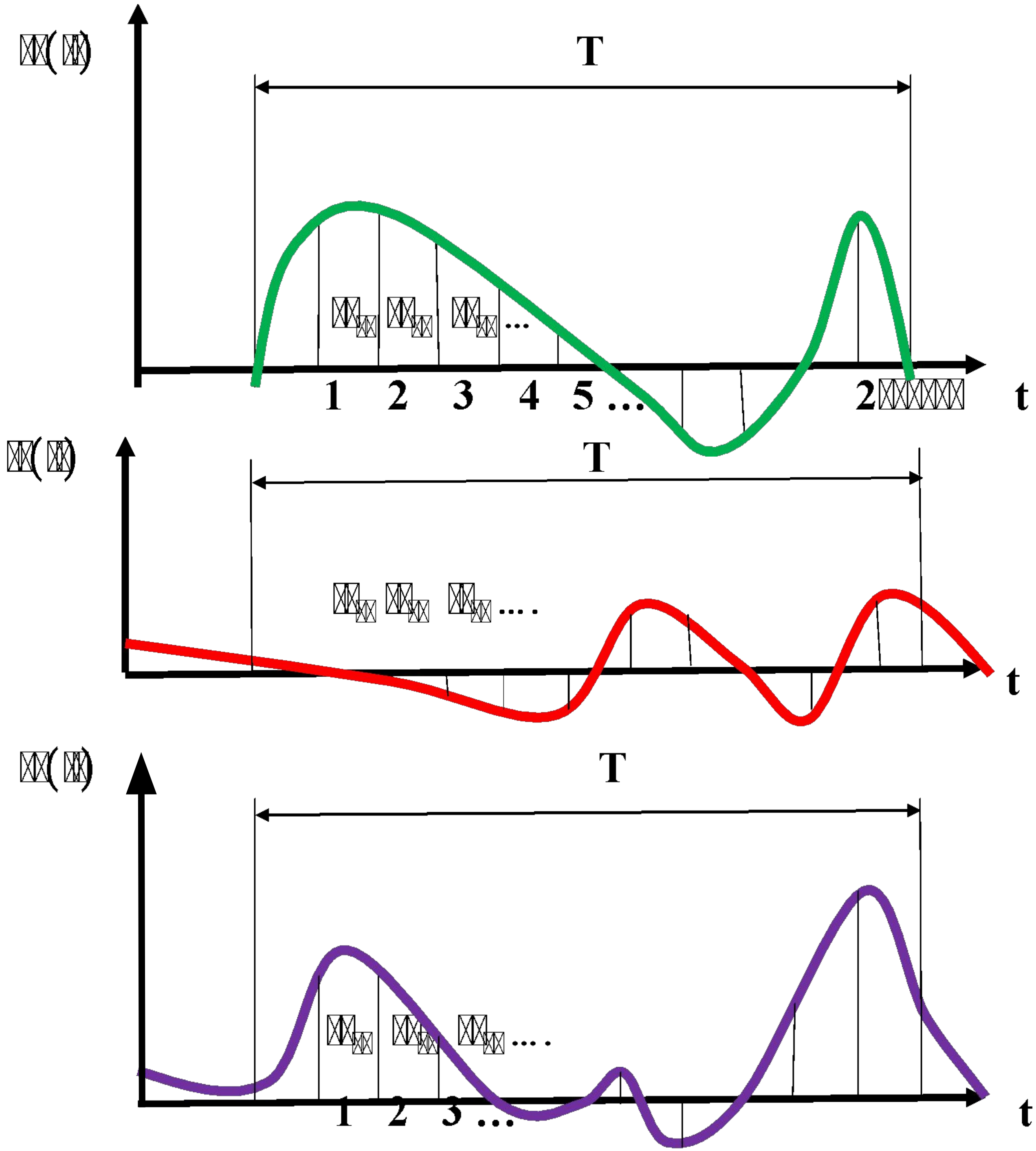
2. На интервале наблюдения аналогового сигнала [illegible] при конечном интервале дискретизации Δ[illegible] наблюдается конечное число отсчетов. Поэтому в реальном случае ряды [illegible]([illegible]) содержат конечное число составляющих.

3. Реальное восстанавливающее устройство отлично от идеального фильтра нижних частот и имеет характеристики, отличные от идеальных.

4. Имеются погрешности в синхронизации, обеспечивающих стабильность частоты отсчетов.

5. Имеются погрешности из-за неточности квантования и проч.

6. Действие помех в канале связи, приводящие к ошибкам в цифровом сигнале.


T
1 2 3 4 5 …
t
T
t
T
1 2 3 …
t

Теория информации

Передается сигнал [illegible]([illegible]) длительностью Т.

В канале действует аддитивная помеха [illegible]([illegible]).

Сигнал [illegible]([illegible]) и канал связи имеют спектр и полосу пропускания [illegible].

На приеме сигнал [illegible] = [illegible] + [illegible]([illegible]).

В соответствии с теоремой В.А. Котельникова рассмотрим отсчеты через интервал [illegible] = [illegible] / [illegible]Δ[illegible], заменив непрерывные функции [illegible] = [illegible] [illegible]Δ[illegible] [illegible] отсчетами.

Количество информации в сигнале [illegible] относительно сигнала [illegible] равно:

[illegible], [illegible] = [illegible] − [illegible]([illegible] | [illegible])

Теория информации

Помеха $n(t)$ и сигнал $x(t)$ статистически независимы. Значит $H(y|x) = H(n)$. Тогда

$$I(x, y) = H(y) - H(n)$$

Скорость передачи информации:

$$R = \lim_{T \to \infty} \frac{I(x, y)}{T} = \lim_{T \to \infty} \frac{H(y) - H(n)}{T}$$

При ограниченной дисперсии максимальная энтропия при нормальном законе распределения. Тогда, полагая статистическую независимость сигналов и помехи, а также нормальным закон распределения отсчетов сигналов и помехи, имеем:

Теория информации

Теория информации

Для оговоренных условий [illegible] = [illegible] + [illegible].

Тогда [illegible] = Δ[illegible]

Заменяя отношение дисперсий на отношение средних мощностей, имеем:

[illegible] = Δ[illegible] + [illegible]

[illegible] = [illegible] = Δ[illegible] + [illegible]

Основы теории кодирования дискретных сообщений

Кодирование – это отображение элементов одного множества элементами другого. При этом правила такого отображения должны быть известны и для обратного преобразования.

В общей теории связи различают следующие виды кодирования: *примитивное*, *эффективное* (статистическое) и *корректирующее* (помехоустойчивое).

Примитивное кодирование (его еще называют *первичным кодированием*) связано с представлением сообщений в виде чисел в той или иной системе счисления. В современных цифровых СПИ часто применяют представление чисел в *двоичной* системе счисления. В рекуррентной форме:

$$Q_k = \pm \sum_{i=-m}^{n-1} a_i k^i,$$

где Q_k - представляемое число, k - основание системы счисления, a_i - разрядный коэффициент, изменяющийся от 0 до $k - 1$. Величина i является номером разряда, n - число целых разрядов, m - число дробных разрядов.

Основы теории кодирования дискретных сообщений

Кодирование – это отображение элементов одного множества элементами другого. При этом правила такого отображения должны быть известны и для обратного преобразования.

В общей теории связи различают следующие виды кодирования: *примитивное*, *эффективное* (статистическое) и *корректирующее* (помехоустойчивое).

Примитивное кодирование (его еще называют *первичным кодированием*) связано с представлением сообщений в виде чисел в той или иной системе счисления. В современных цифровых СПИ часто применяют представление чисел в *двоичной* системе счисления. В рекуррентной форме:

$$A_q = \pm \sum_{i=-m}^{n-1} a_i q^i,$$

где A_q - представляемое число, q - основание системы счисления, a_i - разрядный коэффициент, изменяющийся от 0 до $q - 1$. Величина i является номером разряда, n - число целых разрядов, m - число дробных разрядов.

В десятичной системе счисления, когда $q = 10$ данная формула может быть записана следующим образом:

$$A_{10} = \pm \sum_{i=-m}^{n-1} a_i 10^i,$$

где a = 0, 1, 2, 3…, 9.

В двоичной системе счисления, когда $q = 2$, формула примет вид:

$$A_2 = \pm \sum_{i=-m}^{n-1} a_i 2^i$$

где а = 0, 1.

Достаточно широко при компьютерной обработке информации применяются восьмеричная и шестнадцатеричная системы счисления, которые используются, например, для обозначения адресов расположения данных в памяти компьютера и т. д. Для восьмеричной и шестнадцатеричной системы счисления $q = 8$ и $q = 16$ соответственно, Тогда имеем:

$$A_8 = \pm \sum_{i=-m}^{n-1} a_i 8^i$$

где a = 0, 1, 2, 3…, 7.

$$A_{16} = \pm \sum_{i=-m}^{n-1} a_i 16^i$$

где a = 0, 1, 2, 3…, 9, A, B, C, D, E, F.

Пример:

Двоичное число $K_2 = 11001{,}101 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0, +1 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3}$.

В десятичной форме это же число равно $K_{10} = 25{,}125 = 2 \cdot 10^1 + 5 \cdot 10^0, +1 \cdot 10^{-1} + 2 \cdot 10^{-2} + 510^{-3}$.

В восьмеричной - $K_8 = 31{,}5 = 3 \cdot 8^1 + 1 \cdot 8^0, +5 \cdot 8^{-1}$ и

в шестнадцатеричной - $K_{16} = 19{,}10 = 1 \cdot 16^1 + 9 \cdot 16^0, +10 \cdot 16^{-1}$.

Пусть [illegible], [illegible] = 0, [illegible] − 1 – ансамбль сообщений, наблюдаемый на выходе источника ДС. Требуется подобрать такой *эффективный* код, который максимально устраняет избыточность ДС или приводит к минимально возможному объему цифрового представления ДС по сравнению с примитивным кодированием (*максимальному сжатию данных*). *Теоретической основой эффективного кодирования* является **первая теорема Шеннона** (**основная теорема кодирования для канала без шума**), согласно которой при любой статистике источника ДС существует код по основанию [illegible], позволяющий при отсутствии ограничений на задержку получить среднее число $\bar{n} = \mathrm{M}\{n_l\}$ кодовых символов на элемент сообщения, сколь угодно близкое к минимально возможному значению

$$\text{[illegible]}_{\text{[illegible]}} = \frac{\text{[illegible]}(\text{[illegible]})}{\log_{\text{[illegible]}} \text{[illegible]}}$$

или иначе, каким бы ни был источник ДС с энтропией $H(A) < \infty$, всегда существует такой способ кодирования сообщений, для которого выполняется условие:

$$\bar{n} = n_{\min} + \varepsilon,$$

где $\varepsilon > 0$ - сколь угодно малая величина, взятая "для перестраховки", потому что имеем дело со случайными последовательностями.

Метод укрупнения алфавита. Пусть 10-ичный источник создает последовательности цифр: ...3, 1, 5, 7, 9, 8, 0, 2, ... и пусть очередная цифра выбирается независимо от предыдущей с одинаковой вероятностью 1/10 (поскольку цифр всего 10). Если применять примитивное двоичное кодирование, то каждая десятичная цифра заменяется $\lceil \log_2 10 \rceil = 4$ символьной двоичной комбинацией. Для передачи N десятичных цифр придется израсходовать $4N$ двоичных символов. Однако энтропия источника ДС в данном случае $H(A) = \log_2 10 \approx 3.32$ и в соответствии с первой теоремой Шеннона должен существовать код с $\bar{n} \approx 3.32$. Найдем его, укрупняя алфавит. Для этого разобьем десятичную последовательность на пары цифр: ...31, 57, 38, 02, 10, 47,11, 25, и каждую пару будем примитивно кодировать как единый символ нового, большего по мощности, алфавита. Поскольку таких разных пар будет 100, то понадобятся $\lceil \log_2 100 \rceil = 7$ -ми символьные двоичные кодовые комбинации. При этом на одну десятичную цифру будет расходоваться уже 7/2=3.5 двоичных символов.

Этот процесс можно продолжить и попытаться кодировать тройки десятичных цифр: ...315, 798, 021, 047, 112, 5.... и т.д. Постепенно мы приблизимся к теоретически достижимому пределу. Однако при этом растет задержка и ухудшается помехоустойчивость.

Методы статистического кодирования. При статистическом двоичном кодировании элементов ДС используются кодовые комбинации разной длины. Причем более вероятным элементам ДС, приписываются более короткие кодовые комбинации. Наибольшее снижение избыточности достигается кодированием по методу *Фано-Шеннона* или *Хаффмана*. Рассмотрим пример построения неравномерного кода по методу Хаффмана, имеющего более общий алгоритм.

Поясним сказанное на примере двоичного кодирования (ансамбля $[A, P(A)]$, характеризуемого 8-ью сообщениями $\{a_0, a_1, \dots a_7\}$ и соответственно вероятностями их появления $P(a_0) = 1/2$, $P(a_1) = P(a_2) = 1/8$, $P(a_3) = P(a_4) = P(a_5) = 1/16$, $P(a_6) = P(a_7) = 1/32$.

- на 1-ом этапе объединяются два наименее вероятных события в одно составное (с суммарной вероятностью объединяемых событий), располагаемого также в порядке убывания вероятностей нового ансамбля с числом элементов на единицу меньшим;
- на 2-ом и последующих этапах процедура 1-го этапа повторяются до тех пор, пока суммарная вероятность составного события не достигнет единицы;
- на каждом этапе строится сигнальный граф, содержащий две ветви по числу объединяемых событий и узел, характеризующий составное событие;
- кодирование осуществляется, начиная с конечного узла, имеющего вероятность 1, и заканчивается в начале графа, с присваиванием кодового символа 1, если в каждом узле идти к верхней ветви и 0, если идти к нижней ветви. Подсчет единиц и нулей (в каждом промежуточном узле) при переходе от конца графа к L его начальным узлам определяют искомые кодовые комбинации кода Хаффмана.

		A								

	p(A)	A								
	1/2									
	1/8									
	1/8									
	1/16									
	1/16									
	1/16									
	1/32									
	1/32									

	p(A)	A	Шаг 1							
	1/2									
	1/8									
	1/8									
	1/16									
	1/16									
	1/16									
	1/32		1/16							
	1/32									

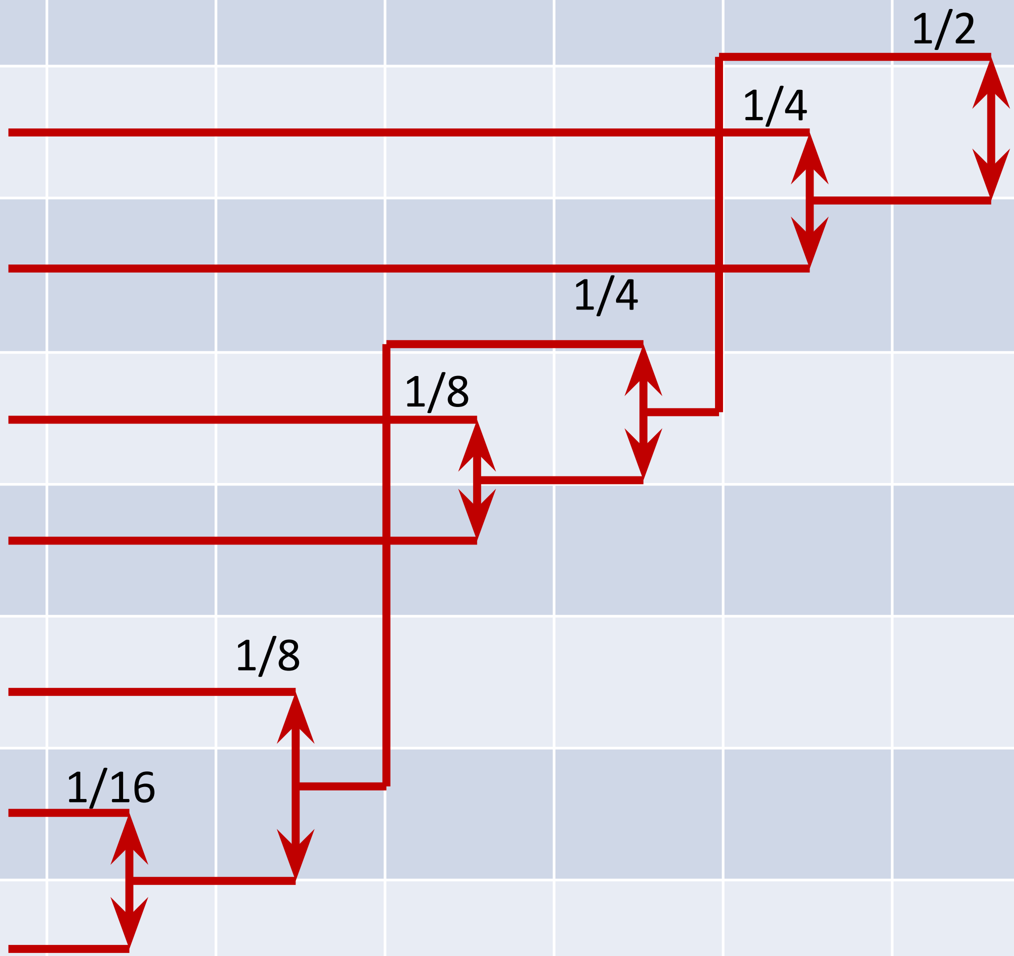
	p(A)	A	Шаг 1	Шаг 2						
	1/2									
	1/8									
	1/8									
	1/16									
	1/16									
	1/16			1/8						
	1/32		1/16							
	1/32									

	p(A)	A	Шаг 1	Шаг 2	Шаг 3					
	1/2									
	1/8									
	1/8									
	1/16				1/8					
	1/16									
	1/16			1/8						
	1/32		1/16							
	1/32									

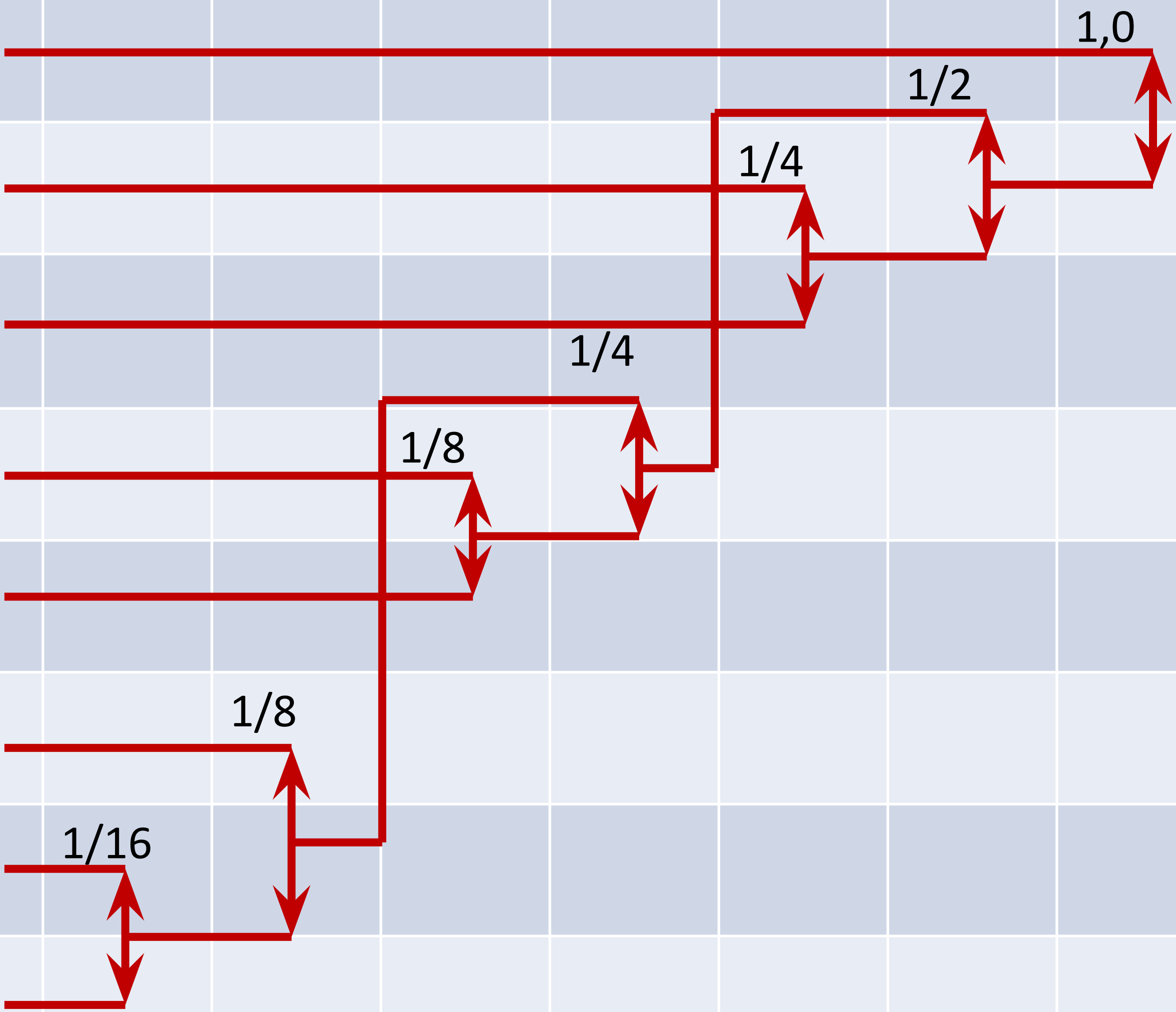
	p(A)	A	Шаг 1	Шаг 2	Шаг 3	Шаг 4				
	1/2									
	1/8									
	1/8					1/4				
	1/16				1/8					
	1/16									
	1/16			1/8						
	1/32		1/16							
	1/32									

	p(A)	A	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Шаг 5			
	1/2									
	1/8						1/4			
	1/8					1/4				
	1/16				1/8					
	1/16									
	1/16			1/8						
	1/32		1/16							
	1/32									

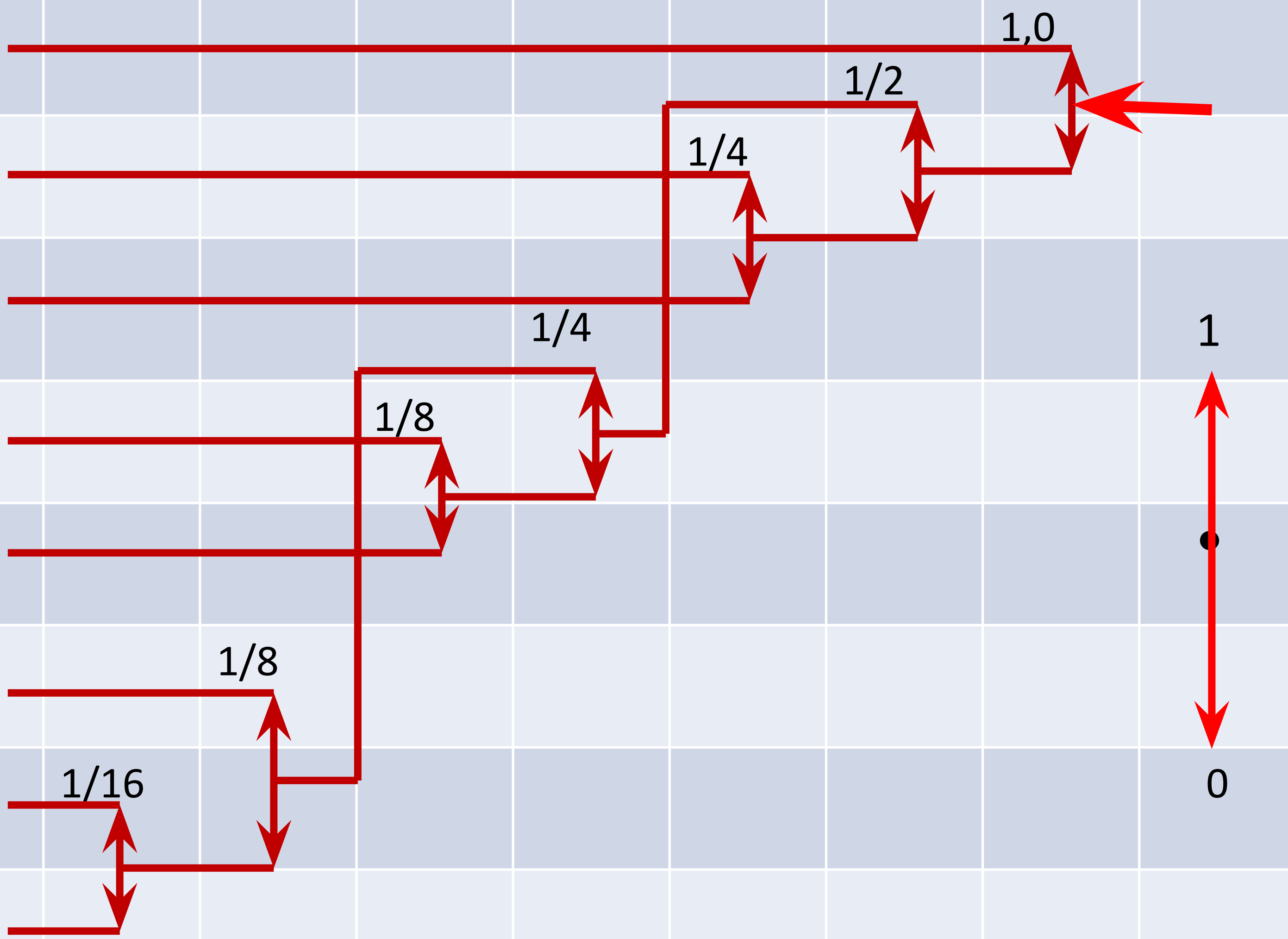
	p(A)	A	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Шаг 5	Шаг 6		
	1/2									
	1/8									
	1/8									
	1/16									
	1/16									
	1/16									
	1/32									
	1/32									



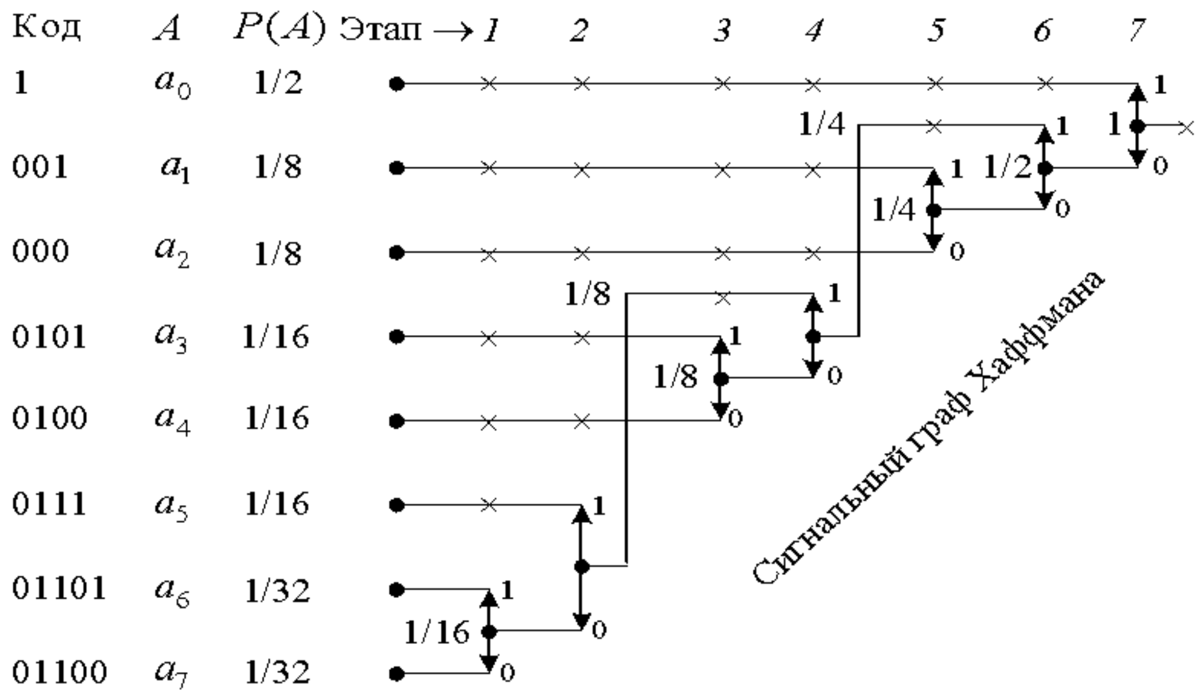
	p(A)	A	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Шаг 5	Шаг 6	Шаг 7	
	1/2									
	1/8									
	1/8									
	1/16									
	1/16									
	1/16									
	1/32									
	1/32									



	p(A)	A	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Шаг 5	Шаг 6	Шаг 7	Старт
	1/2									
	1/8									
	1/8									
	1/16									
	1/16									
	1/16									
	1/32									
	1/32									



Код	p(A)	A	Шаг 1	Шаг 2	Шаг 3	Шаг 4	Шаг 5	Шаг 6	Шаг 7	Старт
1	1/2							1/2	1,0	
001	1/8						1/4			
000	1/8					1/4				1
0101	1/16				1/8					
0100	1/16									
0111	1/16			1/8						
01101	1/32		1/16							0
01100	1/32									



Сигнальный граф эффективного кодирования по Хаффману

Неравномерный код обладает высокой скоростью кодирования, но имеет худшую помехоустойчивость по сравнению с равномерным кодом, а так же более сложную реализацию. Следует так же отметить неравномерность задержки, что так же усложняет восстановление исходного сообщения.

Сообщ.	$\boxtimes_0$	$\boxtimes_1$	$\boxtimes_2$	$\boxtimes_3$	$\boxtimes_4$	$\boxtimes_5$	$\boxtimes_6$	$\boxtimes_7$
Хаффм	1	001	000	0101	0100	0111	01101	01100
Натур.	000	001	010	011	100	101	110	111

Декодируем последовательность:

…010011110010010101010001110110110110011010000010010100…

По **Хаффману**:

$$\ldots\ \boxtimes_4;\ \boxtimes_0;\ \boxtimes_0;\ \boxtimes_0;\ \boxtimes_0;\ \boxtimes_1;\ \boxtimes_1;\ \boxtimes_3;\ \boxtimes_3;\ \boxtimes_2;\ \ldots$$

Натуральный (взвешенный) код:

- необходимо обеспечить обнаружение начало кодовых комбинаций и только потом декодирование:

$$\ldots\ \boxtimes_2;\ \boxtimes_3;\ \boxtimes_6;\ \boxtimes_2;\ \boxtimes_2;\ \boxtimes_5;\ \boxtimes_2;\ \boxtimes_4;\ \boxtimes_3;\ \boxtimes_5;\ \ldots$$

Неравенство **Крафта-Макмиллана** устанавливает, что при заданных кодируемом и кодирующем алфавитах, состоящих соответственно из n и d символов, а так же заданных желаемых длин кодовых слов: $\ell_1, \ell_2, \ldots, \ell_n$, необходимым и достаточным условием существования разделимого и префиксного кодов, обладающих заданным набором длин кодовых слов является выполнение неравенства:

$$\sum_{i=1}^{n} d^{-\ell_i} \leqslant 1.$$

Генетический код

1-е положение		2-е положение									3-е положение
		U		C		A		G			
		UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U	
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C	
		UUA	Leu	UCA	Ser	UAA	ochre	UGA	opal	A	
	U	UUG	Leu	UCG	Ser	UAG	amber	UGG	Try	G	
		CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U	
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C	
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A	
	C	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G	
		AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U	
		AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C	
		AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A	
	A	AUG*	Met	ACG	Thr	AAG	Lys	AGG	Arg	G	
		GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U	
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C	
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A	
	G	GUG*	Val	GCG	Ala	GAG	Glu	GGG	Gly	G	

Триплетные комбинации азотистых оснований мРНК: тимин, цитозин, аденин, гуанин (U, C, A, G) определяют следующие аминокислоты: **Phe** – фениланин, **Leu** – лейцин, **Ile** – изолейцин, **Met** – метионин, **Val** – валин, **Ser** – серин, **Pro** – пролин, **Thr** – треонин, **Ala** – аланин, **Tyr** – тирозин, **His** – гистидин, **Gln** – глутамин, **Asn** – аспарагин, **Lys** – лизин, **Asp** – аспарагиновая кислота, **Glu** – глутаминовая кислота, **Cys** – цистеин, **Try** – триптофан, **Arg** – аргинин, **Gly** – глицин.

Звездочкой обозначены стартовые кодоны, а триплеты **ochre, amber, opal** действуют как стоп кодоны.

(по F. Crick)

Примеры сравнения

16 - 1 0 0 0 0 - Человек

15 - 0 1 1 1 1 - Обезьяна

14 - 0 1 1 1 0 - Кто-то или что-то

……………………………………

2 - 0 0 0 1 0 - Кто-то или что-то

1 - 0 0 0 0 1 - Червяк

0 - 0 0 0 0 0 - Кто-то или что-то

$$N = (d^n)!$$

При d=2 и n=5 имеем N=2,631308369…e+35.

При d=4 и n=3 имеем N=1,268869321… e+89.