Лекция 3.

1. Условная энтропия. Взаимная информация. Канал связи.
2. Сжатие классических данных. Типичные слова. Теорема Шеннона для незашумленного канала связи.
3. Двоичный симметричный канал связи. Емкость канала.
4. Коды, исправляющие ошибки. Код Хамминга. Теорема Шеннона для зашумленного канала.
   Обратимые логические операции. Универсальные ЛЭ Тоффоли и Фредкина.

(Из прошлой лекции - не успел)

<u>Связь энтропии и информации.</u>

Рассмотрим идеальный газ, состоящий только из одной частицы (Кадомцев, динамика и информация). Это не абсурд. Если одна частица заключена в сосуде объемом V со стенками, находящимися при температуре V, то рано или поздно она придет в равновесие с этими стенками. В каждый момент времени она находится во вполне определенной точке пространства и с вполне определенной скоростью. Будем проводить все процессы настолько медленно, что частица успеет в среднем заполнить весь объем и многократно поменять величину и направление скорости при неупругих столкновениях со стенками сосуда. Таким образом, частица оказывает на стенки среднее давление, имеет температуру T и ее распределение по скоростям является максвелловским с температурой T. Эту систему из одной частицы можно адиабатически сжимать, можно менять ее температуру, давая ей возможность прийти в равновесие со стенками сосуда.

Среднее давление на стенку при $N = 1$, равно $p = T/V$, а средняя плотность $n = 1/V$. Рассмотрим случай изотермического процесса, когда $T = const$. Из первого начала при $T = const$. и $p = T/V$ получаем

$$\Delta Q = TdS = pdV = T\frac{dV}{V}, \text{ поскольку } dE = 0.$$

Отсюда находим, что изменение энтропии не зависит от температуры, так что

$$TdS = T\frac{dV}{V} \to S = \ln\frac{V}{V_0}.$$

Здесь введена постоянная интегрирования: “размер частицы”$\ll V_0 \ll V$ - чтобы не нарушалось приближение идеального газа.

Работа при изотермическом процессе

$$W = \int pdV = T\int\frac{dV}{V} = T\ln\frac{V_2}{V_1} = T(S_2 - S_1)$$

работа определяется разностью энтропий.

Пусть у нас имеются идеальные перегородки, которыми можно поделить сосуд на части без затраты энергии. Разделим наш сосуд на две равные части с объемом $V/2$ каждая. При этом частица будет находиться в одной из половин - но мы не знаем в какой. Допустим, что у нас есть прибор, который позволяет определить в какой из частей находится частица, например, прецизионные весы. Тогда из симметричного распределения вероятностей 50% на 50% нахождения в двух половинках мы получаем 100% вероятности для одной из половин - происходит “коллапс” распределения

вероятностей. Соответственно, новая энтропия $\Delta S = \ln \frac{V}{2V_0}$. окажется меньше исходной энтропии на величину $\Delta S = S_1 - S_2 = \ln \frac{V}{V_0} - \ln \frac{V}{2V_0} = \ln V - \ln V_0 - \ln V + \ln 2 + \ln V_0 = \ln 2.$

За счет уменьшения энтропии можно совершить работу. Для этого достаточно двигать перегородку в сторону пустого объема вплоть до его исчезновения. Работа будет равна $W = T\Delta S = T \ln 2.$ Если бы во внешнем мире ничего не менялось, то повторяя эти циклы, можно построить вечный двигатель второго рода. Но второй закон термодинамики запрещает получение работы только за счет тепла. Значит во внешнем мире должно что-то происходить. Что же это? Обнаружение частицы в одной из половин ***меняет информацию о частице*** **-** из двух возможных половинок указывается только одна, в которой находится частица. Это знание соответствует одному биту информации. Процесс измерения уменьшает энтропию частицу (перевод в неравновесное состояние) и ровно настолько же увеличивает информацию о системе (частице). Если совершать повторные деления пополам полученной ранее половинки, четвертушки, восьмушки и т.д., то энтропия будет последовательно уменьшаться, а информация - увеличиваться! Другими словами

$S + I = \text{const}.$

Чем больше известно о физической системе, тем меньше ее энтропия. Если о системе известно все - это значит, что мы перевели ее в сильнонеравновесное состояние, когда ее параметры максимально удалены от равновесных значений. Если в нашей модели частицу удастся поместить в элементарную ячейку объема V_0, то при этом $S = 0$, а информация достигает своего максимального значения $I_{\max} = -\ln p_{\min} = \ln \frac{V}{V_0}$, поскольку вероятность p_{min} найти частицу в данной ячейке равна V_0/V. Если в последующие моменты времени частица начнет заполнять больший объем, то информация будет утрачиваться, а энтропия - расти. Подчеркнем, что за информацию нужно платить (по второму началу) увеличением энтропии S_e внешней системы, причем $\Delta S_e > I$. Действительно, если бы за один бит информации прибор (внешняя система) увеличивал свою энтропию на величину ΔS_e меньшую одного бита, то мы могли бы обратить тепловую машину. А именно, расширяя объем, занятый частицей, мы бы увеличивали ее энтропию на величину $ln2$, получая работу $Tln2$, а суммарная энтропия системы частица плюс прибор уменьшилась бы. Но это невозможно по второму началу.

Итак, информационная энтропия - это мера недостатка (или степень неопределенности) информации о действительном состоянии физической системы.

Информационная энтропия Шеннона:

$$H = \log_2 \Delta\Gamma = -\sum_n p_n \log_2 p_n \text{, где } \sum_n p_n = 1. \qquad (1)$$

(это относится к двухуровневым системам, типа бит: “0” и “1”. Если размерность равна n, то $H = log_n \Delta\Gamma$. Так, для $n = 3$, $H = log_3 \Delta\Gamma$, причем, $\Delta\Gamma = 3$.)

В идеальном случае, когда отсутствуют шумы и помехи, создаваемые внешними источниками в канале связи, конечное распределение вероятностей после измерения сводится к одному определенному значению $p_n = 1$, т.е. $H = 0$, а максимальное значение полученной при измерении информации будет определяться : $I_{max} = H_0$. Таким образом, информационная энтропия Шеннона системы имеет смысл максимальной информации, заключенной в системе; она может быть определена в идеальных условиях измерения

состояния системы в отсутствие шумов и помех, когда энтропия конечного состояния равна нулю:
$H = 0$.
Часто величину (1) называют информационным содержанием
Пусть X - случайная величина, принимающая значения $X = \{x_1, x_2 ... x_n\}$, а $p(x)$ - ее функция распределения. Тогда информационное содержание или информационная энтропия величины X:

$$H = \log_2 \Delta\Gamma = -\sum_x p(x) \log_2 p(x). \qquad (2)$$

Рассмотрим два примера.
<u>Пример 1.</u> Если мы знаем, что $X = 2$, то $p(2) = 1$ и в сумме (2) нет других слагаемых, то $H = 0$, т.е. информационное содержание величины X равно нулю. С другой стороны, если величина X получается при подбрасывании кости с равновероятным распределением $p(x) = 1/6$ для $x \in \{1, 2, 3, 4, 5, 6\}$. Тогда $H = -\log_2\left(1/6\right) \simeq 2.58$. Если X может принимать N различных значений, то ***информационное содержание величины X максимально, когда распределение вероятностей равномерное,*** т.е. $p(x) = 1/N$. Таким образом, для честной кости $H \cong 2.58$, а для нечестной, когда, например, $p(6) = 1/2$, $P(1, 2,...5) = 1/10$ получаем $H = 2.16$. Это утверждение можно строго доказать. Заметим, что оно сочетается с нашим пониманием физического смысла энтропии в том смысле, что информационное содержание (энтропия) максимально, если априорное знание об X минимально. Это свойство используется, например, в криптографии, где необходимо выбирать (неортогональные) базисы равновероятным образом.
Таким образом, максимальная информация, которая может быть в принципе запасена в переменной, принимающей N различных значений составляет $-\log_2\left(1/N\right) = \log_2\left(N\right)$.
Выбор основания “2” у логарифма в теории информации обусловлен требованием $H(X) = 1$, когда X может принимать два значения с одинаковой вероятностью ($N = 2$). Двухуровневые переменные, таким образом, содержат единицу информации- *бит*.

<u>Условная энтропия.</u>
Пусть для двухуровневой переменной X, вероятность того, что $X = 1$, равна p, а вероятность того, что $X = 0$ равна $1 - p$. Тогда информационная энтропия есть функция только p:

$$H = -p \log_2 p - (1-p) \log_2 (1-p) \le 1. \qquad (3)$$

Далее в этой лекции под H(p) понимается именно энтропия дихотомных сигналов.
Рассмотрим условную вероятность $p(y|x)$ - вероятность того, что величина Y принимает значение y при условии, что величина X принимает значение x. Условной энтропией называется величина $S(Y|X)$:

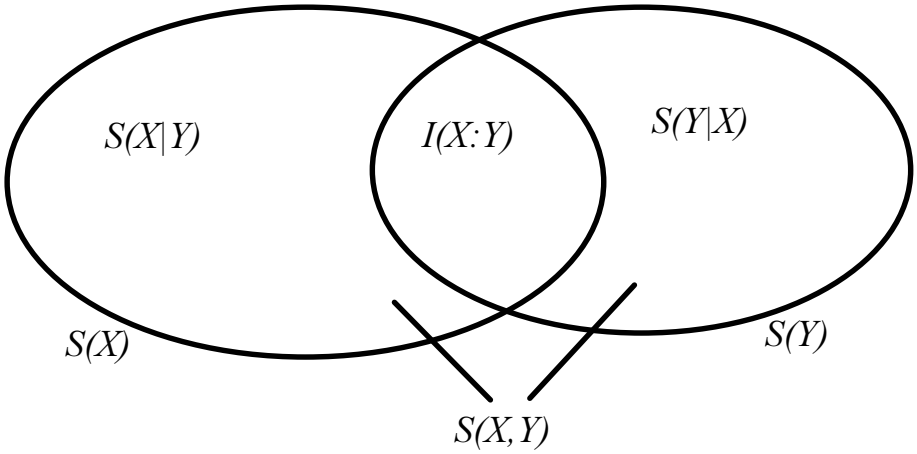
$$S\left(Y \mid X\right) = -\sum_x p(x) \sum_y p(y \mid x) \log p(y \mid x) = -\sum_x \sum_y p(x, y) \log p(y \mid x), \qquad (4)$$

где во втором равенстве использовано понятие совместной вероятности
$p(x, y) = p(y \mid x) p(x)$ - есть вероятность того, что X принимает значение x, а Y принимает значение y.
Из определения (4) следует, что $S(Y|X)$ есть мера того, сколько информации, в среднем, оставалось бы в Y при условии, что мы бы знали X. Заметим, что всегда $S(Y|X) \le S(Y)$ и обычно $S(Y|X) \ne S(X|У)$.

Понятие условной энтропии служит краеугольным камнем для другой величины - ***взаимной информации***, определяемой как

$$I(X:Y)=\sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}=\sum_x\sum_y p(x,y)\log p(x,y)-\sum_x\sum_y p(x,y)\log p(x)p(y)=S(X)-S(X\mid Y). \qquad (5)$$

По определению величина $I(X:Y)$ есть мера того сколько информации содержат X и Y друг о друге. Например, если X и Y - независимые величины, то $p(x,y) = p(x)p(y)$, так что $I(X:Y) = 0$. Соотношения между основными мерами классической информации показаны на рисунке.



Можно показать, что $S(X,Y)$ - информационное содержание величин X и Y (информация, которую мы бы получили, если бы не зная ничего изначально, мы бы узнали значения X и Y) удовлетворяет

$$S(X,Y)=S(X)+S(Y)-I(X:Y).$$

Информация может исчезнуть, но она не может возникнуть из ничего. Этот важный факт отражается в математической формулировке “неравенства получения данных”:

$$\text{если}\quad X\to Y\to Z,\quad \text{то}\quad I(X:Z)\le I(X:Y). \qquad (6)$$

Символы со стрелками означают, что величины X, Y, Z, составляют марковский процесс, в котором Z зависит от Y, и не зависит непосредственно от X:

$$p(x,y,z)=p(x)p(y\mid x)p(z\mid y).$$

Содержание “неравенства получения данных” состоит в том, что “data processor” Y может передать к Z информации не больше, чем он получил от X.

<u>Сжатие данных (data compression)</u>

Как доказать, что определение (1) служит хорошей мерой информации? На первый взгляд кажется непонятно, как разрешить эту задачу. Рассмотрим следующую простую ситуацию (см. рисунок).

Пусть некая персона X (традиционно ее зовут Алиса) хочет передать сообщение приятелю (его зовут Бобом). Ограничим себя только случаем, когда X имеет только два

значения: “нет” и “да”. Мы говорим, что Алиса служит “источником” с “алфавитом” из двух символов. Алиса общается с Бобом посредством пересылки битов (нулей и единиц). Между Алисой и Бобом размещается “***канал связи***” - физическая система, передающая или преобразующая информацию. В качестве канала связи может выступать и логическое устройство.

Если X - дискретная случайная величина, принимающая значения на множестве $\Lambda = \{1, 2 ... |\Lambda|\}$. Рассмотрим случайный источник (Алиса), который порождает последовательность независимых одинаково распределенных случайных величин с распределением p. Последовательность $\omega = (x_1, ...x_n)$ {всего n штук} букв алфавита Λ называется словом длины n. Общее число таких слов $|\Lambda|^n$. Например, Λ = {1, 2, 3}. Пусть n = 2. Сколько всего слов, составленных из n = 2 символов 1, 2, 3? Очевидно их 3^2 = 9 штук. Общее же число слов, составленных из n букв (символов) есть:

$$|\Lambda|^n = \left(2^{\log_2|\Lambda|}\right)^n = 2^{n\log_2|\Lambda|}.$$

Значит, чтобы закодировать все эти слова, используя двоичные последовательности потребуется $n\log_2|\Lambda|$ бит!

Есть лучший способ кодирования - сжатие данных - который использует то обстоятельство, что распределение p - неравномерная величина.

Будем измерять информационное содержание X, подсчитывая *в среднем* число битов, которое должна послать Алиса, для того чтобы Боб распознал X. Очевидно, она должна просто посылать “0” как “нет и “1” как “да”, обеспечивая “скорость битов” в один бит на X. Однако, что будет, если X является существенно случайной переменной, исключая случаи, когда нули идут чаще, чем единицы? Оказывается, что в этом случае Алиса может передавать сообщения Бобу более эффективно, используя следующую процедуру.

Пусть p - вероятность того, что $X = 1$ и $1 - p$ - вероятность того, что $X = 0$. Алиса ждет, пока n значений X будут готовы для того, чтобы их переслать (n - большое число). Среднее число единиц в такой последовательности n значений равно np. Это - наиболее вероятное число единиц в любой последовательности, так, что число единиц будет всегда близко к np. Пусть np - целое число. Найдем вероятность получения любой последовательности, содержащей np единиц. Это будет произведение того, что в последовательности есть np единиц (p^{np}) и того, что остальные элементы - $n(p-1)$ нули $\{(1-p)^{n(1-p)}\}$

$$p^{np}(1-p)^{n-np} = \left\{\textit{Используем соотношение } A = 2^{\log_2 A}, \textit{ где } A = p^{np}(1-p)^{n-np}\right\} =$$

$$= 2^{\log_2\left[p^{np}(1-p)^{n(1-p)}\right]} = 2^{np\log_2 p + n(1-p)\log_2(1-p)} \equiv 2^{-nH(p)}$$

где $H(p)$ - определена в (3) для двоичной кодировки. Такая последовательность называется *типичной последовательностью* или *типичным словом.* Более точно, определим типичное слово как последовательность, которая удовлетворяет условию

$$2^{-n(H(p)+\varepsilon)} \le p(\textit{слово}) \le 2^{-n(H(p)-\varepsilon)},$$

где $p(\textit{слово})$ - вероятность появления последовательности (слова)

Теперь можно показать, что вероятность образования типичной последовательности из n величин Алисы превосходит величину $1-\varepsilon$ для достаточно больших n, вне зависимости от того, насколько мало ε! ***Это означает, что Алисе не нужно передавать n бит Бобу для того, что бы он распознал n исходов***. Ей лишь нужно

сказать Бобу *какова ее типичная последовательность*. Им нужно договориться заранее о том, как выделять (в смысле отмечать) эти типичные последовательности. Например, они могут нумеровать их в порядке увеличения бинарного значения. Алиса просто посылает свою метку, но не саму последовательность.

Чтобы проследить, насколько хорошо работает этот метод можно показать, что все типичные последовательности имеют одинаковые вероятности и, следовательно, их существует $2^{nH(p)}$ штук. Для передачи одной из $2^{nH(p)}$ последовательностей, очевидно, Алиса должна послать *nH(p)* бит. Ясно, что Алиса не может сделать ничего лучше этого (т.е. послать меньшее число бит) т.к. типичные последовательности равновероятны: никакой информации не может быть извлечено при дальнейших манипуляциях. Поэтому информационное содержание или информационная энтропия каждого значения из множества *X* в оригинальной последовательности должно быть *H(p)*, что доказывает справедливость (2)!

Мы не вдаемся в математические детали доказательства. Отметим, лишь, что использован закон больших чисел, который гласит, что для произвольно малых ε и δ выполняется неравенство

$$P\left(\left|m-np\right|<n\varepsilon\right)>1-\delta$$

для достаточно больших *n*, где *m* - число единиц, содержащихся в последовательности длиной n. Для достаточно больших n число единиц *m* будет отличаться от среднего значения *np* на число как угодно малое по сравнению с *n*. Например, в рассмотренном выше случае нули и единицы будут распределены по биномиальному закону

$$P\left(n,m\right)=C\left(n,m\right)p^{m}\left(1-p\right)^{n-m}\simeq\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-(m-np)^{2}\Big/2\sigma^{2}\right\},$$

где распределение Гаусса получается в пределе, когда *n, np* $\to\infty$, стандартное отклонение $\sigma=\sqrt{np(1-p)}$, а $C(n,m)=\dfrac{n!}{m!(n-m)!}$ - число сочетаний из *n* по m.

Все эти соображения, относящиеся к определению информационной энтропии (информационному содержанию) (2) имеют важное практическое значение. Оно состоит в том, что для передачи *n* значений величины *X* нам нужно послать через канал связи только $nH(X)\le n$ бит.

Этот алгоритм получил название *классического сжатия данных* или ***теоремы Шеннона для нешумящего канала связи.***

<u>Пример.</u> Пусть p = 1/4. Тогда по теореме Шеннона лучшее из того, что дает техника сжатия данных это передача каждого сообщения из четырех значений *X* посылкой в среднем *4H(1/4)* ~ 3.245 бит.

Техника сжатия данных нашла огромное применение в телекоммуникации. Например, при сжатии информации для передачи телевизионных изображений и сохранении их в памяти компьютера. С точки зрения инженерного дизайна канала связи сжатие данных может показаться фантастической техникой. Предположим у нас есть телефонная связь с гористой местностью, но скорость связи не очень высока для того, чтобы послать, скажем, видео изображение. Обычное инженерное решение состоит в замене телефонной линии на более быструю, в то время как из теории информации следует, что можно использовать старую линию, но при условии компрессии данных на одном из двух концов (компрессия на одном и декомпрессия на другом конце). Удивительно, что пригодность кабеля может быть улучшена “починкой” информации, а не самого кабеля.

Двоичный симметричный канал связи

До сих пор мы рассматривали случаи идеальной передачи сообщений посредством нешумящих каналов. Теорема Шеннона дает нам меру лучшей компрессии данных в условиях идеальной связи.

Теперь остановимся на случае передачи информации при наличии шума в канале. рассмотрим лишь простейшие случаи.

Предположим, что у нас имеется двоичный канал связи, т.е. когда Алиса посылает Бобу только нули и единицы. Нешумящий канал передает значения по схеме $0 \to 0$ *и* $1 \to 1$. Зашумленный канал иногда выдает нуль вместо единицы и наоборот. Существует большое число разновидностей шума. Например, “инверсия бита” приводит к равновероятному перевороту бита $0 \to 1$ *и* $1 \to 0$. Иногда канал имеет тенденцию к релаксации, т.е. $1 \to 0$, в то время как обратный процесс $0 \to 1$ невозможен. Возможны случаи, когда такие процессы идут случайно от бита к биту или возникают и кончаются внезапно.

Очень важную разновидность шума представляет собой процесс, при котором воздействие на биты происходит независимо и ошибки возникают по схеме $0 \to 1$ *и* $1 \to 0$. Эти ошибки наиболее близки к реальным шумовым процессам. Если две ошибки $0 \to 1$ *и* $1 \to 0$ равновероятны, то канал связи называется ***двоичным симметричным каналом***. Такой канал характеризуется единственным параметром *p*, который есть просто вероятность ошибки на один посланный бит. Предположим, что Алиса посылает в канал сообщение *X*, а Боб получает зашумленное сообщение *Y*. Задача Боба состоит в оптимальном извлечении *X* из *Y*. Если *X* состоит из единственного бита, то Боб может использовать условные вероятности:

$$p(x=0 \mid y=0) = p(x=1 \mid y=1) = 1-p,$$

$$p(x=0 \mid y=1) = p(x=1 \mid y=0) = p,$$

из которых можно извлечь $S(X|Y) = H(p)$, используя (3, 4). Тогда из определения (5) взаимной информации получаем:

$$I(X:Y) = S(X) - H(p). \qquad (7)$$

Очевидно, что наличие шума в канале ограничивает информацию об Алисиной величине *X*, содержащейся в принятом Бобом сигнале *Y*. Кроме того из неравенства сжатия данных (6) Боб не может увеличить информацию об *X* манипулируя *Y*. Однако из (7) следует, что качество связи между Алисой и Бобом может улучшаться при росте $S(X)$. Оказывается, что информация зависит как от свойств источника, так и от свойств канала. Было бы полезно ввести некую меру, характеризующую только канал связи для того чтобы знать насколько хорошо канал передает информацию. Такая величина называется ***емкостью канала связи***. Она определяется как максимальная взаимная информация $I(X:Y)$ между входом и выходом, причем максимализация происходит по всем возможным источникам:

Емкость канала $C \equiv \max\limits_{\{p(x)\}} I(X:Y)$ (8)

Емкость канала измеряется в битах на символ и для двоичных каналов должна принадлежать интервалу $0 \le C \le 1$. Все это очень здорово, но определенная в (8) емкость не позволяет нам эффективно сравнивать каналы, поскольку процедура максимизации по источникам не вполне тривиальна. Определение емкости канала $C(p)$

является одной из основных проблем теории информации, но к счастью данный случай достаточно прост. Из (7) и (8) можно вывести результат:

$$C(p) = 1 - H(p), \qquad (9)$$

полученный при *S(X) = 1* (т.е. когда *P(x = 0) = P(x = 1) = ½),* т.е. максимум достигается при равномерном распределении.

<u>Коды, исправляющие ошибки.</u>

До сих пор мы интересовались тем как информация проходит по зашумленному каналу связи и каковы потери в нем. Алиса не может передать больше информации, чем C(p) на передаваемый символ. Предположим, что Боб обезвреживает мину, а Алиса, находясь на некоем расстоянии, кричит ему какой провод обрезать. Она не скажет "синий провод" и надеется, что он услышит ее правильно. Она будет повторять свои сообщения много раз, и Боб будет ждать пока не будет уверен, что получил правильное сообщение. Такой способ сообщений свободный от ошибок может быть достигнут даже через зашумленный канал. В этом примере можно получить выгоду от уменьшенного числа ошибок жертвуя количеством передаваемой информации. Рассмотрим более продвинутые стратегии.

Набор {0, 1} рассматривается как группа (GF(2)), в которой операции сложения. Вычитания. Умножения и деления выполняются по модулю 2 (т.е. 1 + 1 = 0). *n*- битовое слово есть вектор, содержащий *n* компонент, например, 011 это вектор (0,1,1). Набор таких векторов образует векторное пространство относительно сложения, т.к., например, 011 + 101 значит (0,1,1) + (1,0,1) = (0 + 1,1 + 0,1 + 1) = (1,1,0) = 110 по стандартным правилам сложения векторов. Это аналогично выполнению операции XOR.

Эффект шума, действующего на слово *u* можно записать в виде $u \to u' = u + e$, где вектор ошибки показывает какой бит в слове *u* перевернулся под действием шума. Например, *u* = 1001101 → *u′* = 1101110 можно переписать в виде *u′* = *u* + 0100011. Код, исправляющий ошибку, есть такой набор слов, что

$$u + e \neq v + f \quad \forall u,\ v \in C\left(u \neq v\right), \forall e, f \in E, \qquad (10)$$

где *E* - набор ошибок, исправляемых кодом *C*, включая случай отсутствие ошибки *e* = 0. Чтобы использовать такой код Алиса и Боб договариваются какому кодовому слову u отвечает какое сообщение и Алиса изредка посылает кодовые слова в канал. Поскольку канал зашумлен, Боб получает не *u* а *u* + *e*. Однако, Боб может однозначно извлечь *u* из *u* + *e* используя (10).

Пример работы кода, исправляющего ошибки, приведен в таблице. Это т.н. код Хамминга. Обозначение [*n, k, d*] означает, что кодовые слова имеют длину *n* бит, всего этих кодовых слов 2^k штук и все они отличаются друг от друга по крайней мере в *d* позициях. В силу некоей специфики условие (10) удовлетворяется для любой ошибки, которая воздействует более чем на один бит. Другими словами, набор *E* исправляемых ошибок есть {0000000, 1000000, 0100000, 0010000, 0001000, 0000100, 0000010, 0000001}. Заметим, что *E* может содержать по крайней мере 2^{n-k} членов. Отношение *k/n* называется *нормой кода*, т.к. каждый блок из n передаваемых бит содержит *k* бит информации, т.е. *k/n* бит на бит.

Сообщение	Хаффман	Хамминг
0000	10	0000000
0001	000	1010101
0010	001	0110011

0011	11000	1100110
0100	010	0001111
0101	11001	1011010
0110	11010	0111100
0111	1111000	1101001
1000	011	1111111
1001	11011	0101010
1010	11100	1001100
1011	111111	0011001
1100	11101	1110000
1101	111110	0100101
1110	111101	1000011
1111	1111001	0010110

Левая колонка - 16 возможных 4-битовых сообщений. Две другие колонки-закодированные версии каждого сообщения. Код Хаффмана - сжатие данных. Наиболее часто встречающиеся сообщения имеют более короткую длину. Считается, что в каждом бите сообщения нули встречаются в три раза вероятнее, чем единице. Код Хамминга - исправляющий ошибки. Каждое кодовое слово отличается от всех других по крайней мере тремя позициями. Поэтому любая единичная ошибка может быть исправлена. Код Хамминга линеен: все слова даются линейными комбинациями 1010101, 0110011, 0001111, 1111111. Они удовлетворяют проверке четности 1010101, 0110011, 0001111.

Параметр d называется “минимальным расстоянием” кода. Он важен, когда кодируется сигнал в присутствии шума, действующего на биты независимо , как в двоичном симметричном канале. Код с минимальным расстоянием d может исправить все ошибки, действующие менее чем на $d/2$ бит передаваемого кодового слова и при шуме, действующем независимо на биты, это является наиболее вероятным набором ошибок. На самом деле вероятность, что n-битовое слово получит m ошибок дается биномиальным распределением, поэтому если код может исправить более чем среднее количество ошибок np, исправление будет вероятнее всего удачным.

Центральный результат классической теории информации состоит в том, что существует мощный исправляющий ошибки код:

> **<u>теорема Шеннона</u>**. Если норма $k/n < C(p)$ и n - достаточно велико, то существует двоичный код, позволяющий осуществлять связь с произвольно малой вероятностью ошибки.

Здесь вероятность ошибки - это вероятность того, что происходит неисправленная ошибка, которая заставляет Боба неверно воспринять полученное слово. Теорема Шеннона звучит очень многообещающе, поскольку из нее следует, что не нужно разрабатывать низкошумящие каналы, что является дорогостоящей и трудновыполнимой задачей. Вместо этого мы компенсируем шум техникой коррекцией ошибок при кодировании и раскодировании, т.е. работая непосредственно в рамках теории информации.

В заключение - об универсальных (классических) логических элементах.

Логический элемент осуществляет логически обратимую операцию, если сигнал на входе может быть однозначно определен по сигналу на выходе. Фредкин и Тоффоли показали, что существует два (по крайней мере) универсальных ЛЭ, с помощью

которых можно организовать произвольные логические операции в компьютере. Это логически полные ЛЭ Controlled-Controlled-NOT (Тоффоли) и Controlled SWAP (Фредкин).

ТОФФОЛИ

a	**b**	**c**	**a**	**b**	**c**
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	1
1	1	1	1	1	0

Операция “НЕ” по выходу “c”, когда на входах a, b - логические единицы.

ФРЕДКИН

a	**b**	**c**	**a**	**b**	**c**
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	1	0
1	1	0	1	0	1
1	1	1	1	1	0

входы b, c - обмениваются значениями, когда на входе a -логическая единица.