МАТЕМАТИЧЕСКИЕ ОСНОВЫ ПОМЕХОУСТОЙЧИВОГО КОДИРОВАНИЯ

Лекций – 20 час., практических занятий – 16 час., лабораторных – 14 час.

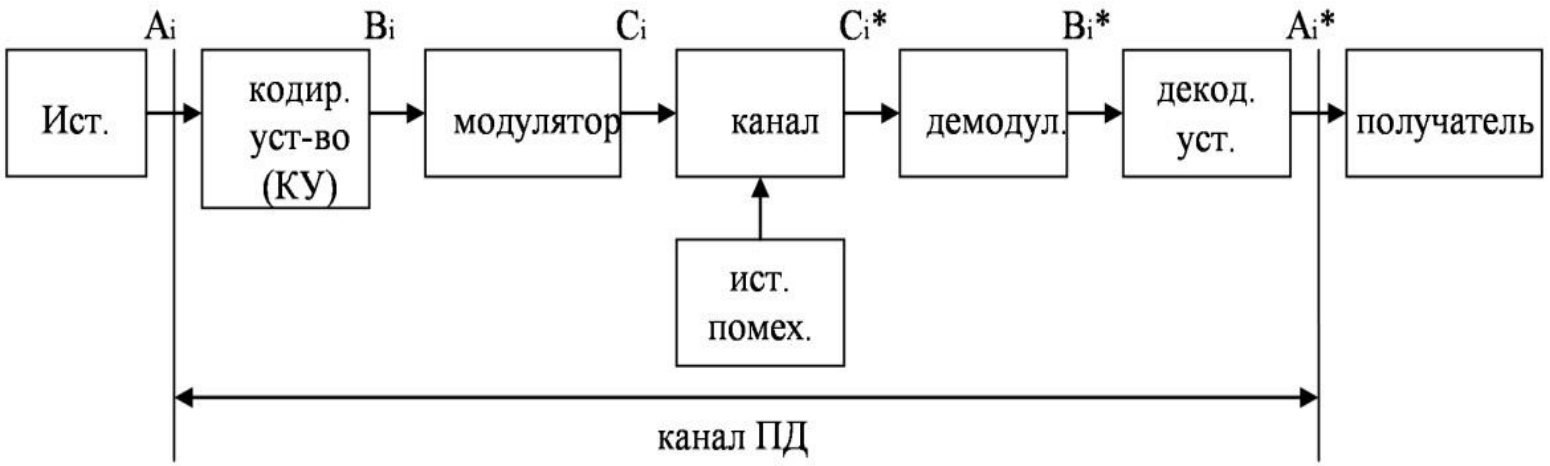
Итоговый контроль - зачет

Занятия ведут:
Профессор Когновицкий Олег Станиславович (лекции);
Доцент Владимиров Сергей Сергеевич (упражнения и лабораторные работы)

Содержание Лекций №1,2:

- **Введение. Обобщенная структурная схема системы передачи сообщений. Дискретный источник и его информационные характеристики. Первичное и вторичное кодирование, их целевое предназначение.**
- **Формулировка первой теоремы Шеннона для канала без помех.**
- **Помехоустойчивое (вторичное) кодирование и его целевые задачи. Ошибки в канале и причины их возникновения.**
- **Модели ошибок.**
- **Вторая теорема Шеннона.**
- **Показатели эффективности применения ПК.**
- **Основные термины и определения помехоустойчивого кодирования. Характеристики помехоустойчивых кодов. Классификация ПК.**

Обобщенная структурная схема системы передачи дискретных сообщений (передачи данных)



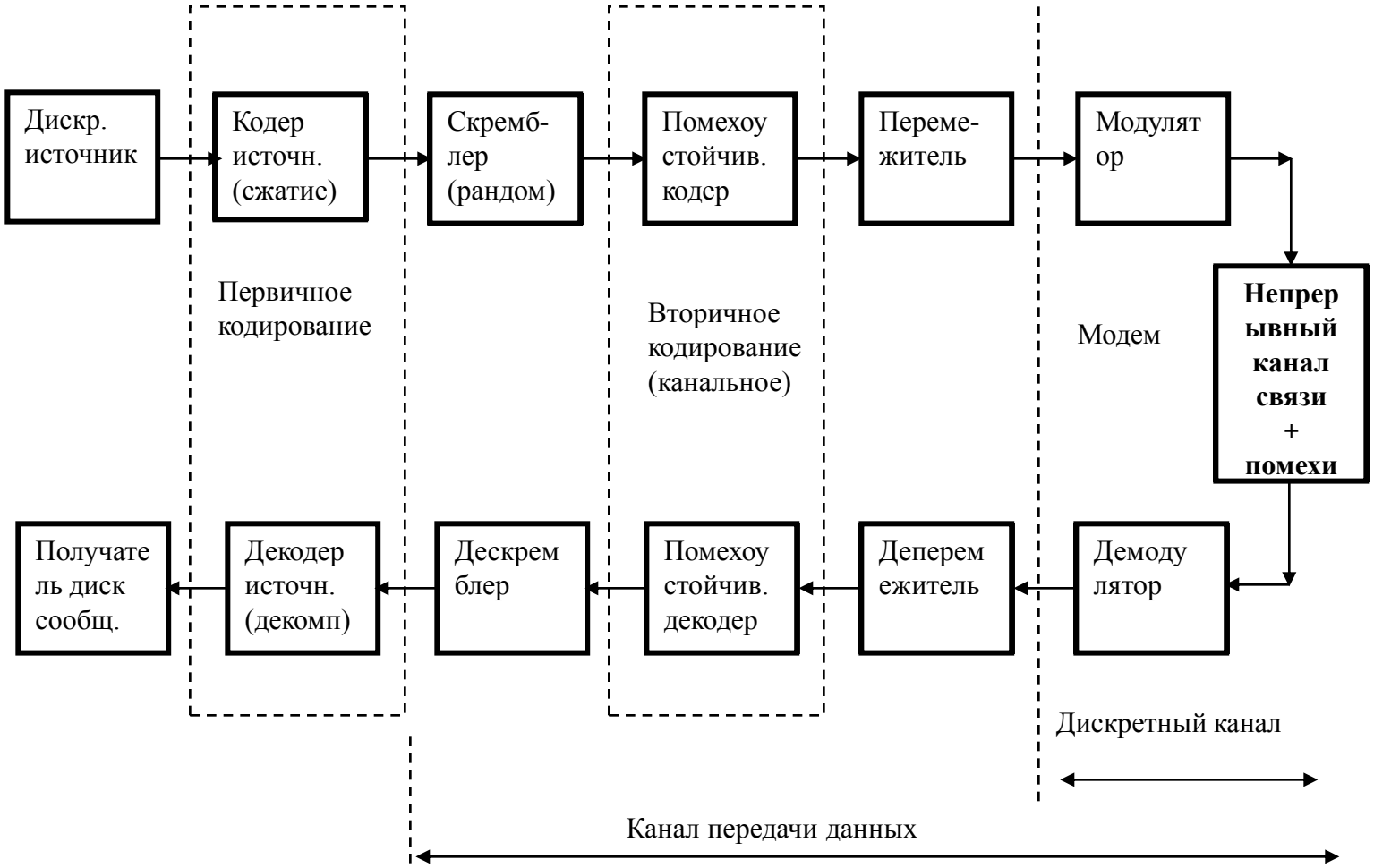
Дискретный источник:

$\{N\}=\{A_1, A_2, A_3,\ldots A_N\}$,
$P_1, P_2, P_3,\ldots P_N$

где: N – множество состояний на выходе источника;

P_i – вероятность появления на выходе источника сообщения A_i.

Обобщенная структурная схема системы передачи дискретных сообщений (передачи данных)



Первичное и вторичное кодирование

Назначение помехоустойчивого кодирования.

Информационные характеристики дискретного источника

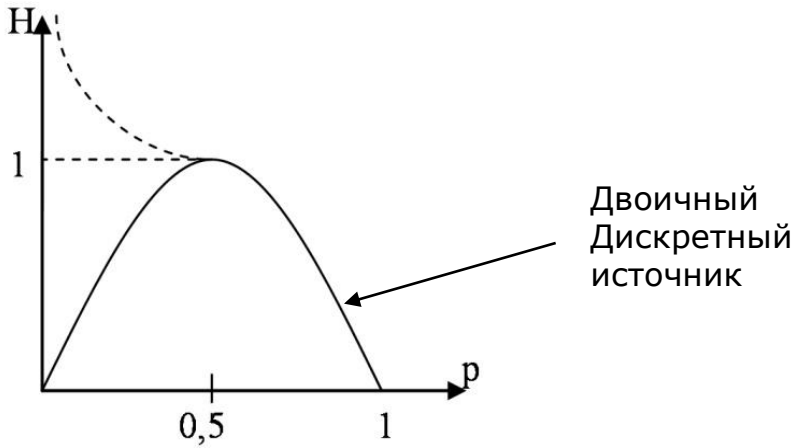
Неопределенность потребителя.
Степень неопределённости. Что влияет на степень неопределенности? Число сообщений N и вероятности появления отдельных сообщений.

Количественная мера неопределенности оценивается информационной емкостью дискретного источника, т.е. количеством информации .
Минимум и максимум неопределенности.
Если $P_i \to 1$, то неопределенность$\to 0$;
если $P_1=P_2=\ldots=P_N=1/N$, то неопределенность будет максимальной.
Формула Шеннона: $\boldsymbol{I_i=\log 1/P_i}$.
Количество информации может быть измерено в любых единицах, но обычно используют **двоичную единицу** – бит. Тогда: $\boldsymbol{I_i=-\log_2 P_i}$.
1 двоичная единица информации – это информация, появляющаяся на выходе источника с двумя равновероятными исходами: $P_1=P_2=0{,}5$.
Энтропия дискретного источника H– это среднее количество информации, содержащейся в одном сообщении, т.е. в двоичных единицах:

$$H = \sum_{i=1}^{N} I_i P_i = -\sum_{i=1}^{N} P_i \log_2 P_i \ \frac{дв.ед}{сообщ.}$$

Для равновероятных сообщений $H = \log_2 N$ *дв. ед./сообщ.*

Энтропия зависит от количества событий и их вероятностей P_i;
Зависимость $H=f(P)$
Двоичный дискретный источник
Рис. 1.2. График зависимости H от P



Можно показать, что если $N>2$, то при увеличении N точка максимума будет смещаться влево и возрастать (на графике это показано штрихпунктирной линией). H_{max} будет при равновероятных событиях.

Энтропия сложного события

Пусть имеются два опыта λ и β.

Опыт λ имеет исходы: $x_1, x_2,\ldots,x_m$

с вероятностями – $p(x_1), p(x_2),\ldots,p(x_m)$;

а опыт β имеет исходы: $y_1, y_2,\ldots,y_k$

с вероятностями - $p(y_1), p(y_2),\ldots,p(y_k)$.

В качестве сложного опыта будем рассматривать совместное появление опытов λ и β.

Тогда исходами сложного опыта будут $(x_1y_1), (x_1y_2),\ldots,(x_my_k)$;

вероятности этих исходов: $p(x_1y_1),\ldots, p(x_iy_j)$.

Энтропия сложного события:

$$H(\lambda,\beta) = -\sum_i \sum_j p(x_i y_j) \log_2 p(x_i y_j) \frac{дв.ед}{сл.событие}.$$

Рассмотрим два случая, когда λ и β независимы и зависимы друг от друга.

1). Опыты λ и β имеют независимые исходы.

Тогда: $p(x_i y_j) = p(x_i) \cdot p(y_j)$.

$$H_1(\lambda,\beta) = -\sum_i \sum_j p(x_i)\, p(y_j) \log_2 [\, p(x_i)\, p(y_j)] \; \frac{дв.ед}{сл.событие}.$$

Раскрыв логарифм произведения, получим:

$$H_1(\lambda,\beta) = \underbrace{-\sum_i p(x_i) \log_2 p(x_i)}_{H(\lambda)} \underbrace{-\sum_j p(y_j) \log_2 p(y_j)}_{H(\beta)} \; \frac{дв.ед}{сл.событие}.$$

Окончательно получаем:

$$H_1(\lambda,\beta) = H(\lambda) + H(\beta).$$

2). Опыты λ и β зависят друг от друга.

Тогда $p(x_i y_j) = p(x_i) \cdot p(y_j / x_i)$

$$H_2(\lambda,\beta) = -\sum_i \sum_j p(x_i y_j) \times \log_2 p(x_i y_j) = -\sum_i \sum_j p(x_i)\, p(y_j / x_i) \times \log_2 [\, p(x_i)\, p(y_j / x_i)] =$$

$$= -\sum_i \sum_j p(x_i)\, p(y_j / x_i) \times \log_2 p(x_i) - \sum_i \sum_j p(x_i)\, p(y_j / x_i) \times \log_2 p(y_j / x_i)] =$$

$$= -\sum_i p(x_i) \times \log_2 p(x_i) \sum_j p(y_j / x_i) - \sum_i p(x_i) \sum_j p(y_j / x_i) \times \log_2 p(y_j / x_i) \, \frac{дв.ед}{сл.событие}.$$

При условии, что произошло событие x_i, имеем:

$$\sum_j p(y_j / x_i) = 1.$$

т.к. суммируются вероятности всех событий y_j при условии появления x_i на выходе источника λ.

Выражение $$-\sum_j p(y_j / x_i) \times \log_2 p(y_j / x_i) = H(\beta / x_i)$$

представляет собой энтропию опыта β, при условии, что в опыте λ был исход x_i . Здесь усреднение результатов по опыту β.

Тогда

$$H_2(\lambda, \beta) = -\sum_i p(x_i) \times \log_2 p(x_i) + \sum_i p(x_i) H(\beta / x_i).$$

Но первое слагаемое есть H(λ), а второе – условная энтропия H(β/λ), где усреднение уже по опыту λ.

Тогда:

$$H_2(\lambda, \beta) = H(\lambda) + H(\beta / \lambda)$$

или аналогично

$$H_2(\lambda, \beta) = H(\beta) + H(\lambda / \beta).$$

Вывод: энтропия сложного события равна сумме энтропий каждого события, если опыты независимы и энтропия равна сумме энтропий одного опыта и условной энтропии другого опыта относительного первого, при двух зависимых опытах (дискретных источниках).

§ 2.1 Первичное кодирование. Первая теорема Шеннона

Процесс передачи сообщений можно представить в следующем виде:

сообщение→комбинация символов→комбинация сигналов.

$$A_i \square \quad B_i \quad \square \quad C_i$$

$$\{N\} \ \{N\} \quad \{N\}$$

Преобразование $A_i \rightarrow B_i$ - **первичное кодирующее** устройство,
обратное преобразование $B_i \rightarrow A_i$ – **первичное декодирующее** устройство.

a – **основание кода,**
n – **длина кодовой комбинации**, например - (101110), то n=6.
Коды **равномерные** и **неравномерные**

Оптимальный код должен иметь **минимальную избыточность**.

Первая теорема Шеннона. Рассмотрим трактовку этой теоремы применительно к дискретному источнику с множеством сообщений, равным N.

Алгоритм определения *n* оптимального первичного кода по первой теореме Шеннона.

а) Имеем множество состояний дискретного **равновероятного источника** N, **основание кода – *а*.**

Для построения оптимального первичного кода производим укрупнение исходных сообщений по *r* и получаем как отдельные события ($A_{i1}\ldots A_{ir}$).

Количество таких укрупненных равновероятных событий будет равно $M=N^r$. Закодируем их равномерным первичным кодом с основанием *а*. При этом получим длину комбинации $n_r=[\log_a M]=\log_a M+\Delta$, где $0<\Delta<1$ – добавка для верхнего округления.

Тогда на одно исходное сообщение будет приходиться

$$n_1 = \frac{n_r}{r} = \left(\frac{\log_a M + \square}{r}\right) = \frac{r\log_a N}{r} + \frac{\square}{r} \quad элементов.$$

Для получения $n_{опт}$ устремим *r* к ∞, тогда $n_1 = n_{опт}=\log_a N = \frac{\log_2 N}{\log_2 a} = \frac{H}{\log_2 a}$.

Вывод: для источника с равновероятными исходами первичный код будет оптимальным тогда, когда число элементов на одно сообщение стремится к $H/(\log_2 a)$.

Если код двоичный, $a = 2$, то следовательно, $n_{опт}=H$.

Пример: События $A_1, A_2, \ldots, A_N$ возникают с равными вероятностями $p_1=p_2=\ldots=p_N=1/N$.
Пусть события равновероятные и $N=10$, $p_i=0{,}1$; $i=1,2\ldots10$.
Энтропия дискретного источника: $H=\log_2 10=3{,}32$ дв.ед./сообщ.
Кодируем сообщение комбинациями одинаковой длины, равной $n=[\log_2 N]$ (верхнее округление). Для нашего примера имеем $\boldsymbol{n_1=[\log_2 10]=4}$.
Имеем следующий вид кодовой комбинации $A_i \rightarrow \{a_1 a_2 a_3 a_4\}$ – 4 дв.ед. информации.
Избыточность первичного кода будет 0,68 дв.ед./сообщ. и, следовательно, такой код не будет оптимальным.
Чтобы добиться более оптимального кода, производим укрупнение событий, т.е. берем сразу по 2 события. Новыми событиями станут пары $\{A_i A_j\}$, получили новый источник с числом укрупненных сообщений $N_2 = N^2 = 10^2$, каждое из которых будет иметь одну и ту же вероятность $p(A_i A_j)=0{,}01$. $n_2=[\log_2 N_2]=[\log_2 100]=7$ дв.ед/на пару $(A_i A_j)$.
$\boldsymbol{n_1=n_2/2=3{,}5}$ дв.ед./сообщ., т.е., мы получили более оптимальный первичный код.

Укрупним по три сообщения:$(A_i A_j A_z)$, $N_3=N^3=10^3$.
$n_3=[\log_2 N_3]=[\log_2 10^3]=10$. Тогда получим $\boldsymbol{n_1=n_3/3=3{,}33}$ дв. элемента/сообщ.

б) **Дискретный источник с различными вероятностями событий:** $p_1 \neq p_2 \neq \ldots \neq p_N$.
Источник содержит N сообщений $A_1, A_2, \ldots, A_N$, вероятности которых $p_1, p_2, \ldots p_N$ будут различны, т.е. $p_i \neq p_j$.
Энтропия дискретного источника будет $H = -\Sigma\, p_i \log_2 p_i$
Найдем $n_{опт}$.
Для этого сообщения на выходе дискретного источника группируют в *S*-цепочки, как отдельные сложные события вида $(A_{i1}A_{i2}\ldots A_{is})$.
Выбираем *S* достаточно большим, чтобы вероятности появления отдельных S-цепочек были статистически одинаковыми.
В длинной цепочке все события можно считать независимы, следовательно A_i появляется в длинной цепочке (p_iS) раз.

$$P(S-цепочки) = \prod_{i=1}^{N} (p_i)^{p_i S}. \qquad M = \frac{1}{P(S-цепочки)}.$$

$m = \log_a M + \Delta$ где $0 < \Delta < 1$.

$$n = \frac{m}{S} = \frac{\log_a M}{S} + \frac{\Delta}{S}.$$

При $S \to \infty$ имеем:

$$n_{опт} = \frac{1}{S}\log_a M = -\frac{1}{S}\log_a \{\prod_{i=1}^{N} (p_i)^{p_i S} = -\frac{1}{S}\sum_{i=1}^{N} (p_i S)\log_a p_i = -\sum_{i=1}^{N} p_i \log_a p_i.$$

$$n_{опт} = \frac{-\sum_{i=1}^{N} p_i \log_2 p_i}{\log_2 a} = \frac{H}{\log_2 a}.$$

Вывод такой же как и у равновероятного источника

§ 2.2 Построение оптимальных кодов

Первая теорема Шеннона дает ответ лишь на вопрос об оптимальной длине кодовой комбинации, но она не дает ответа на вопрос «как построить код?».

Рассмотрим построение конкретных кодов.

Источник с **равновероятными** исходами имеет 10 сообщений

N=10 $\{A_1,\ldots A_{10}\}$ a=2 p_i=0,1. H=3,3 (дв.ед./сообщ.)

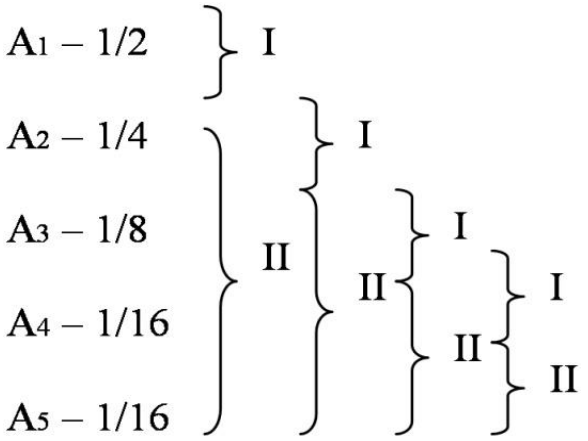
n=[$\log_2 N$]=4. Числа от 0 до 9 кодируем 4-х элементным кодом.

Ранее было доказано по первой теореме Шеннона, что для источника с 10 равновероятными событиями $n_{опт}$=3,3; следовательно, полученный код не оптимальный и надо производить укрупнение событий, кодируя их равномерным кодом.

Пусть теперь имеется простой источник с **неравновероятными** событиями: A_1, A_2, A_3, A_4, A_5: p_i = 1/2, 1/4, 1/8, 1/16, 1/16.

Располагаем в порядке убывания вероятностей и суммарно делим группы пополам.

События, принадлежащим к 1-ой группе, приписывается «1», ко 2-ой – «0».



A1	1
A2	01
A3	001
A4	0001
A5	0000

$$H = -\sum_i p_i \log_2 p_i; \qquad n_{ср} = \sum_i n_i p_i.$$

Легко подсчитать, что H=$n_{ср}$=1,875, значит первичный код оптимальный

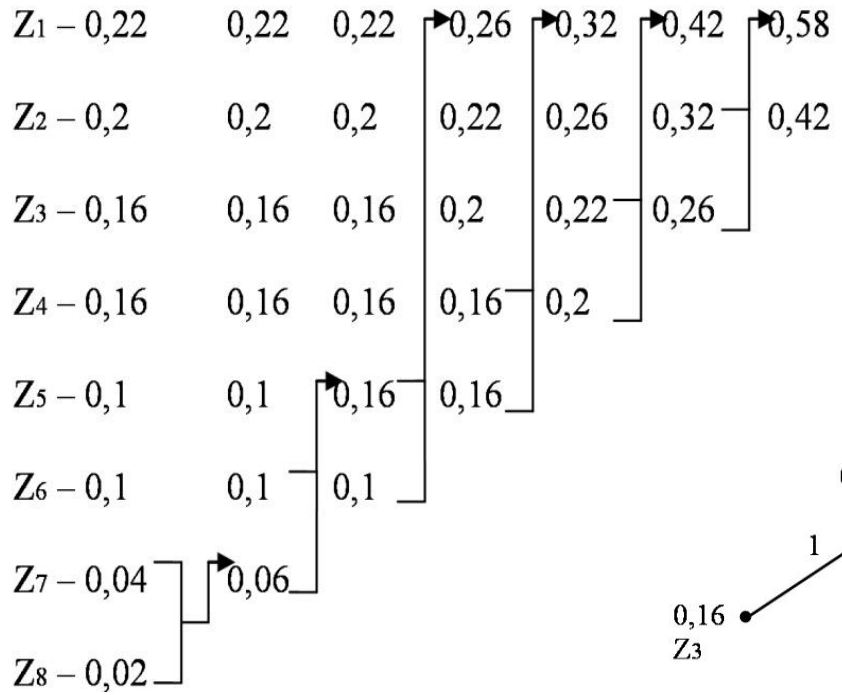
Это пример кода Шеннона-Фено

Ход Хаффмена.

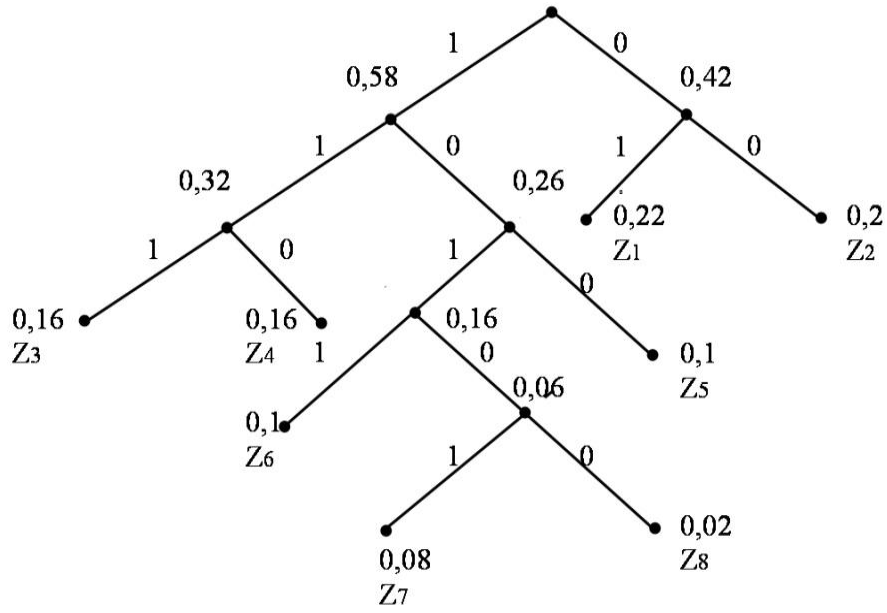
События располагаются в порядке уменьшения их вероятностей, затем группируются так, что 2 события с минимальными вероятностями преобразуются в одну с суммарной вероятностью. Появляется новый ряд вероятностей, расположенных также в порядке убывания вероятностей. Делают так, пока не получат 2 вероятности.

Пусть имеем события $Z_1,\ldots Z_8$ с соответствующими вероятностями:

Z1 – 0,22
Z2 – 0,2
Z3 – 0,16
Z4 – 0,16
Z5 – 0,1
Z6 – 0,1
Z7 – 0,04
Z8 – 0,02



Z1 − 01
Z2 − 00
Z3 − 111
Z4 − 110
Z5 − 100
Z6 − 1011
Z7−10101
Z8 − 10100



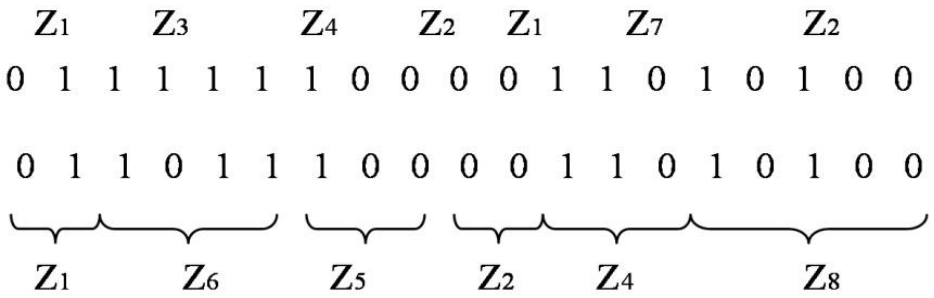
Выводы.

•События с наибольшей вероятностью кодируются меньшим числом элементов.

•$n_{ср}$=2,8; а $n_{опт}$=H=2,76 – следовательно этот код можно считать оптимальным.

•Код обладает перфиксными свойствами.

•Недостаток – возможное размножение ошибок



Общее представление об ошибках в системах передачи сообщений и об их вероятности

Основные причины возникновения ошибок:

- Последствия фильтрации сигналов в передатчике, канале и приемнике – неидеальная фильтрация приводит к появлению *межсимвольной интерференции* (ISI – intersymbol interference).
- Электромагнитные излучения различных других источников.
- Электрические помехи: *тепловой шум*, который аддитивно накладывается на передаваемый сигнал (аддитивный шум), *импульсные помехи, кратковременные перерывы* (замирания), которые представляют собой мультипликативные помехи.
- Природа шумов – *искусственные и естественные* шумы.

$$p(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1(x-a)^2}{2\sigma^2}} \qquad (1.1)$$

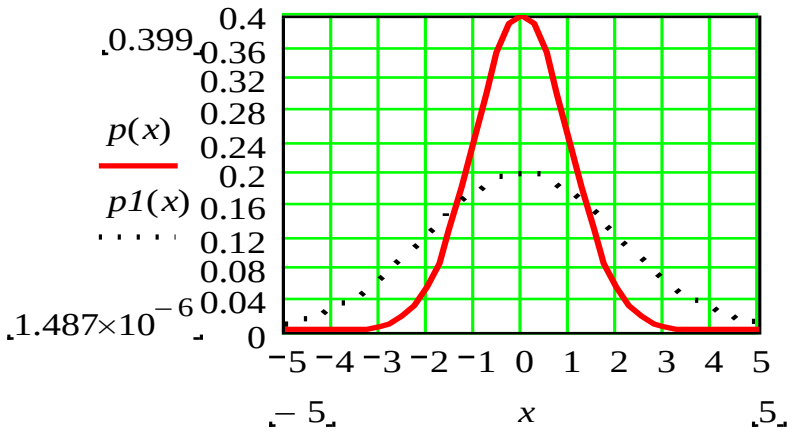


Рис. 1.1. Нормальный закон распределения амплитудных значений гауссова шума при a=0 и σ=1 (сплошная линия) и σ=2 (пунктирная линия).

Основной *спектральной характеристикой* аддитивного шума является спектральная плотность мощности $G_n(f)$, которая будет равномерной для всего частотного диапазона от $-\infty$ до $+\infty$ и записываться в виде

$$G_n(f) = \frac{N_0}{2} \; Вт / Гц. \qquad (1.2)$$

Здесь коэффициент 2 в делителе означает, что функция $G_n(f)$ является двусторонней спектральной плотностью мощности (как в положительной, так и в отрицательной области частот). Из (1.2) следует, что спектральная плотность мощности гауссова шума одинакова на любой частоте (рис. 1.2, *а*), такой шум принято называть *белым*, по аналогии с белым светом, содержащего равные доли всех частот видимого диапазона электромагнитного излучения.

Обратное преобразование Фурье спектральной плотности мощности белого шума определяет его автокорреляционную функцию, равную

$$R_n(\tau) = F^{-1}\{G_n(f)\} = \int_{-\infty}^{\infty} G_n(f) e^{2\pi i f \tau} df = \int_{-\infty}^{\infty} \frac{N_0}{2} e^{2\pi i f \tau} df = \frac{N_0}{2}\delta(\tau) \qquad (1.3)$$

и имеющую вид дельта-функции Дирака, взвешенной множителем $N_0/2$ и находящейся в точке $\tau=0$ (рис. 1.2, *б*).

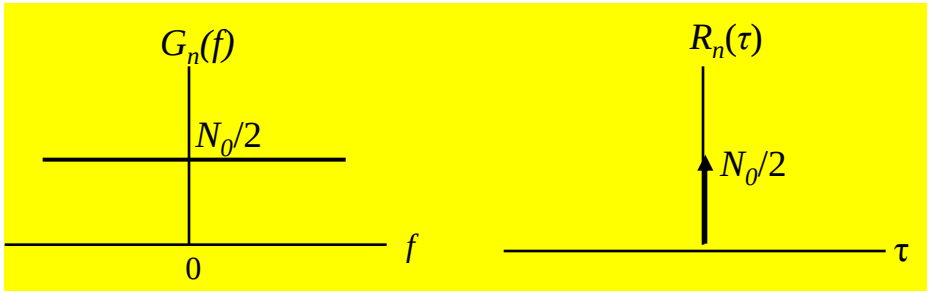


Рис. 1.2. Спектральная плотность мощности (*а*) и автокорреляционная функция (*б*) белого шума.

Равенство автокорреляционной функции $R_n(\tau)$ нулю для всех $\tau \neq 0$ говорит о том, что различные выборки белого шума во времени не коррелируют между собой, т.е. являются независимыми. По этой причине канал с *аддитивным белым гауссовым шумом* (АБГШ) называется *каналом без памяти*.

Вероятность ошибки и отношение сигнал/шум.

Общепринятым критерием качества *аналоговой связи* принято считать отношение сигнал/шум, представляющее собой отношение средней мощности сигнала к средней мощности шума S/N.

В *цифровой связи* в качестве критерия качества связи как правило используется нормированное отношение сигнал/шум E_b/N_0, которое определяется как

$$\frac{E_b}{N_0} = \frac{ST_b}{N/W} = \frac{S/R_b}{N/W} = \frac{S}{N}\left(\frac{W}{R_b}\right). \qquad (1.5)$$

где E_b – это энергия битового сигнала, которую можно определить как мощность сигнала S, умноженную на время передачи битового сигнала T_b;

N_0 – спектральная плотность мощности шума, которую можно выразить как мощность шума N, деленную на ширину полосы частот W;

R_b – битовая скорость передачи равная $R_b = 1/T_b$,

В дальнейшем в нашем курсе речь будет идти о цифровой связи, поэтому в качестве основного критерия связи естественно выбрано отношение E_b/N_0.

Одной из важнейших характеристик цифровой системы связи является зависимость вероятности битовой ошибки P_b от E_b/N_0, имеющая характерный вид, представленный на рис. 1.3

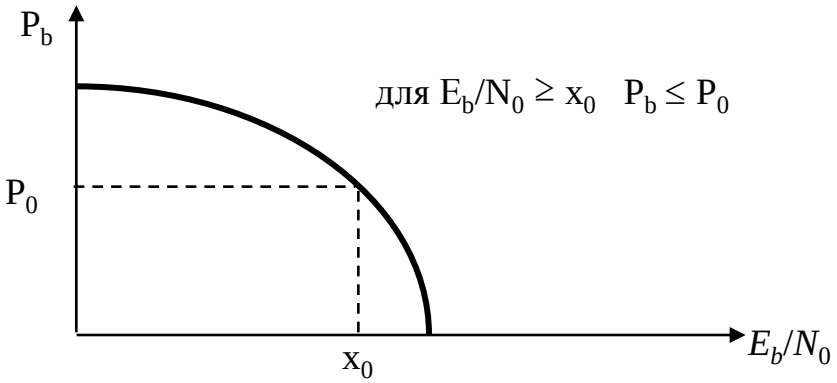


Рис. 1.3. Общий вид зависимости битовой вероятности ошибки от E_b/N_0.

Таким образом, отношение E_b/N_0 может рассматриваться как метрика, позволяющая сравнивать различные цифровые системы связи.

Как отмечено выше, другой причиной возникновения ошибок является *межимпульсная (межсимвольная) интерференция* (intersymbol interference – ISI), которая непосредственно зависит от импульсной характеристики линейной системы, в первую очередь от её полосы пропускания

Рис. 1.4. Импульсная характеристика линейной системы h(t)

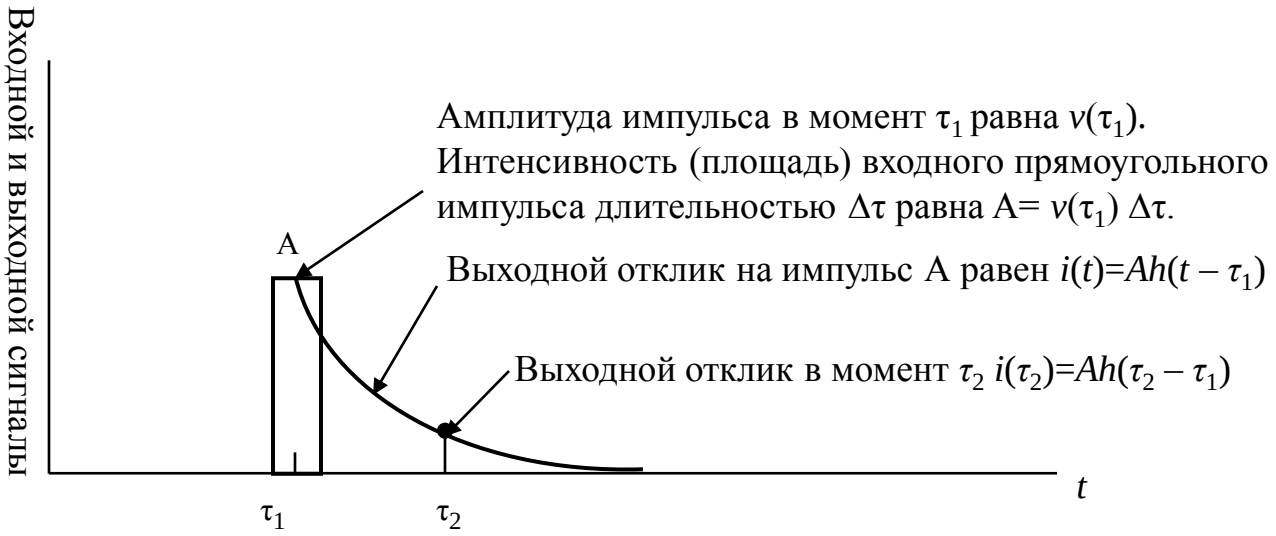


Рис. 1.5. Реакция линейной системы на прямоугольный импульс в момент времени τ_1

На рис.1.6 показаны два примера прохождения прямоугольного импульса $x(t)$ с амплитудой V_m через линейную систему (фильтр) с ограниченной полосой пропускания и выходным сигналом $y(t)$.

а) *б*)

Рис. 1.6. Примеры прохождения прямоугольного импульса длительности T с полосой $W_u = 1/T$ через фильтр с полосой пропускания W_f
а) «хорошая точность воспроизведения», $W_u \approx W_f$;
б) «плохое воспроизведение», $W_u >> W_f$.

На рис. 1.7 показан пример межимпульсной интерференции в тактовый момент времени τ_3. Очевидно, что в момент τ_3 отклик линейной системы на предыдущие импульсы превышает порог и приводит к ошибке.

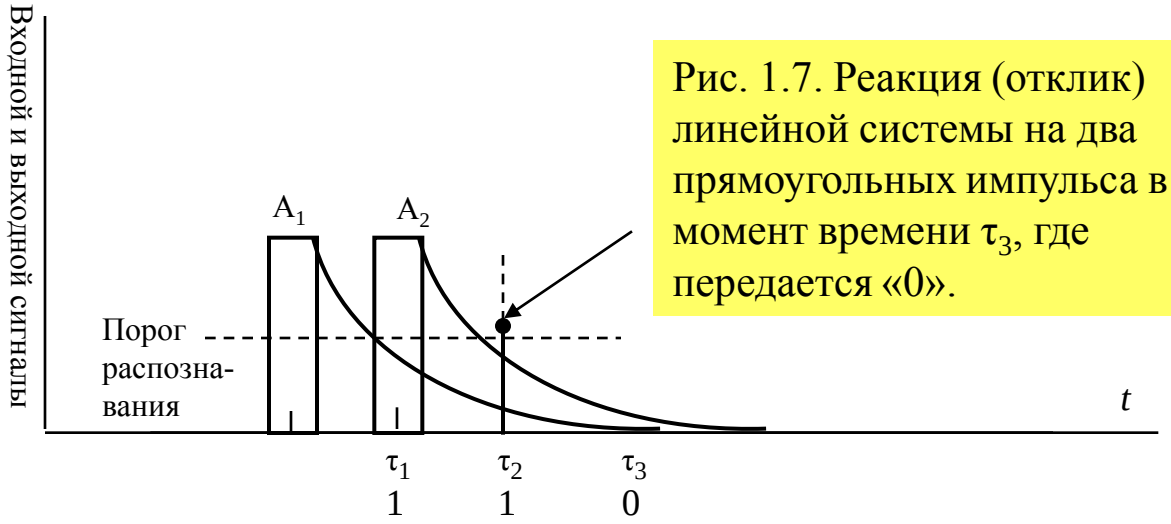


Рис. 1.7. Реакция (отклик) линейной системы на два прямоугольных импульса в момент времени τ_3, где передается «0».

При N импульсах на входе линейной системы отклик на её выходе в последующий момент времени t будет определяться выражением

$$i(t) = A_1 h(t-\tau_1) + A_2 h(t-\tau_2) + \ldots + A_N h(t-\tau_N), \quad (1.6)$$

где $\tau_1, \tau_2, \ldots, \tau_N$ – моменты поступления импульсов на вход.

Для устранения влияний межимпульсной интерференции применяют специальные корректоры, называемые *эквалайзерами* [1] или цифровыми трансверсальными фильтрами.
Другой широко используемый в настоящих системах связи метод предотвращения межимпульсной интерференции (ISI) заключается в выборе специальной формы принимаемого импульса. Над этой проблемой долгое время занимался Найквист. Он показал, что минимальная теоретическая ширина полосы частот, требуемая для передачи со скоростью R=1/T символов/секунду без ISI, должна быть равна ΔF=1/2T. При этом, форма импульса, не вызывающего ISI, должна описываться функцией $\frac{\sin x}{x}$. Такой импульс называют [1] идеальным импульсом Найквиста, форма которого показана на рис.1.8. Как видно из рисунка, передача разных импульсов не создает межимпульсной интерференции в моменты *t* взятия выборок, *t*=±1, ±2, ±3,…
Разумеется, что это будет иметь место в предположении идеальной синхронизации моментов взятия выборок (отсчетов) сигналов.
На практике сегодня такие сигналы применяют при ортогональном многочастотном уплотнении OFDM.

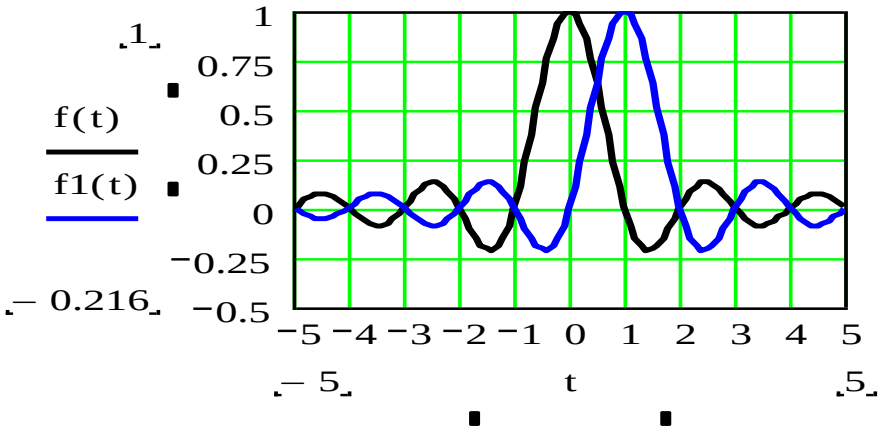


Рис. 1.8. Форма импульсов Найквиста

Вероятность ошибки при передаче двоичных сигналов в канале с гауссовым шумом (АБГШ)

Двоичный сигнал, переданный в течение временного интервала $(0,T)$, будем представлять в следующем виде:

$$s_i(t)=\begin{cases} s_1(t) & 0\le t\le T \quad для\ "1" \\ s_2(t) & 0\le t\le T \quad для\ "0". \end{cases} \qquad (1.7)$$

$$r(t) = s_i(t) + n(t) \qquad i=1,2, \qquad 0\le t\le T. \qquad (1.8)$$

$$z(T) = a_i(T) + n_0(T), \qquad z = a_i + n_0 \qquad i=1,2. \qquad (1.9)$$

где шумовая компонента n_0 является гауссовой с нулевым средним, а компонента a_i является средним значением функции $z(T)$, а именно a_1 или a_2 в зависимости от входного двоичного символа $s_1(t)$ («1») или $s_2(t)$ («0»).

Тогда, с учетом (1.1), плотности условных вероятностей будут определяться выражениями:

$$p(z\mid s_1)=\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{z-a_1}{\sigma_0}\right)^2\right]; \qquad (1.10)$$

$$p(z\mid s_2)=\frac{1}{\sigma_0\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{z-a_2}{\sigma_0}\right)^2\right].$$

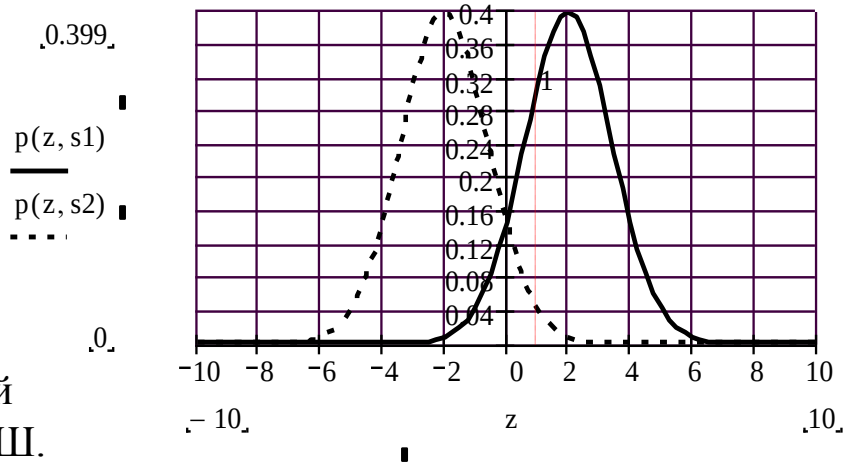


Рис. 1.8. Плотности условных вероятностей передачи двоичных сигналов в канале с АБГШ.

Жесткое и мягкое принятие решений.

На втором этапе после модуляции и дискретизации на приемной стороне принимается решение о переданном сигнале. Решение может быть *жестким* или *мягким*.

Суть **жестких** решений состоит в сравнении значений принятого сигнала $z(T)$ в момент времени T с некоторыми оптимальными порогами. Для двоичных сигналов $s_1(t)=a_1$ и $s_2(t)=a_2$ в канале с АБГШ порог будет один и его оптимальное значение будет равно

$$\gamma = \frac{a_1 + a_2}{2}. \quad (1.11)$$

$$z(T)\begin{cases} > \gamma \to H_1 \quad для \quad сигнала \quad s_1(t) \\ < \gamma \to H_2 \quad для \quad сигнала \quad s_2(t) \end{cases} \quad (1.12)$$

Вероятности ошибок e при передаче сигналов $s_1(t)=a_1$ и $s_2(t)=a_2$ равны

$$P(e|s_1) = P(H_2|s_1) = \int_{-\infty}^{\gamma} p(z|s_1)dz. \quad (1.13) \qquad P(e|s_2) = P(H_1|s_2) = \int_{\gamma}^{\infty} p(z|s_2)dz. \quad (1.14)$$

Суммарная битовая вероятность ошибки при передаче двоичных сигналов в канале с АБГШ

$$P_b = \sum_{i=1}^{2} P(e, s_i) = \sum_{i=1}^{2} P(e|s_i)P(s_i). \quad (1.15)$$

$$P_b = P(e|s_1)P(s_1) + P(e|s_2)P(s_2) = P(H_2|s_1)P(s_1) + P(H_1|s_2)P(s_2). \quad (1.16)$$

Для равновероятных двоичных сигналов суммарная вероятность ошибки равна

$$P_b = \frac{1}{2}P(H_2|s_1) + \frac{1}{2}P(H_1|s_2). \quad (1.17) \qquad P_b = P(H_1|s_2) = P(H_2|s_1). \quad (1.18)$$

$$P_b = \int_{\gamma=\frac{a_1+a_2}{2}}^{\infty} p(z|s_2)dz. \quad (1.19) \qquad P_b = \int_{\gamma=\frac{a_1+a_2}{2}}^{\infty} \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z-a_2}{\sigma_0}\right)^2\right] dz. \quad (1.20)$$

Проведя замену переменной $\frac{z-a_2}{2}=u,$ получим

$$P_b=\int\limits_{u=\frac{a_1-a_2}{2\sigma_0}}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du=Q\left(\frac{a_1-a_2}{2\sigma_0}\right), \qquad (1.21)$$

На рис. 1.9 показан пример **мягких решений** с восьмиуровневой схемой квантования. Уровни квантования условно пронумерованы от нуля до семи в двоичном коде. Если уровень сигнала $z(T)$ попал в зону трех единиц {111}, это свидетельствует о том, что с большой достоверностью переданным символом была «1», а выделение уровня {100} свидетельствует, что была передана «1», но с очень низкой достоверностью. Для более простой и понятной интерпретации мягких решений уровни в данном примере могут быть обозначены [1] как −7,−5,−3,−1,1,3,5, 7. Чем больше абсолютная цифра уровня, тем больше достоверность принятия мягких решений.

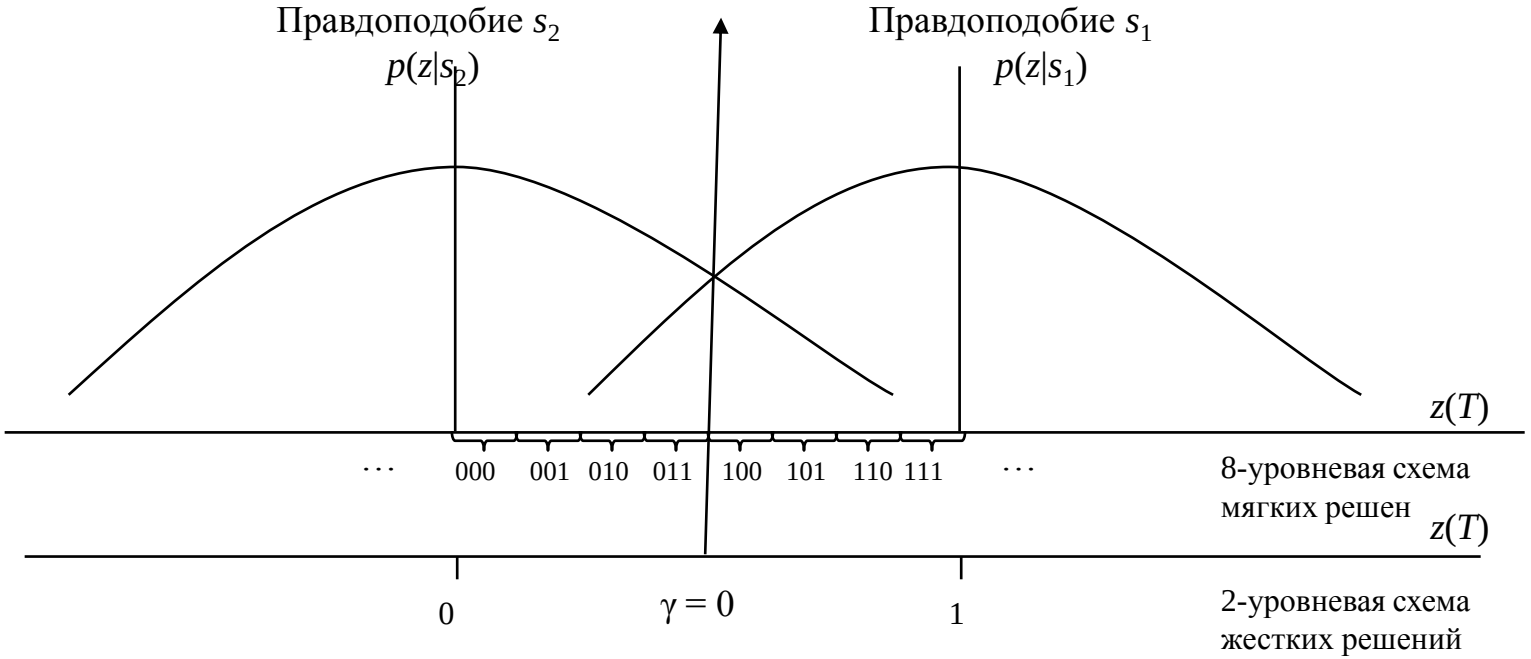


Рис. 1.9. Жесткая и мягкая схемы принятия решений

Знак у цифр характеризует тот или иной двоичный сигнал, например, положительный знак соответствует сигналу s_1 («1»), а отрицательный – сигналу s_2 («0»).
Возможна и другая интерпретация мягких решений для двоичных сигналов в канале с АБГШ. Например, указанные выше цифры в абсолютных значениях более целесообразно соотнести с вероятностями попадания в тот или иной уровень, а именно: −7/8, −5/8, −3/8, −1/8, 1/8, 3/8, 5/8, 7/8. То есть зоны, ближайшие к математическим ожиданиям сигналов s_1 и s_2, имеют наибольшую (близкую к 1) вероятность.
Как утверждается в [1], переход на восьмиуровневую схему мягких решений по сравнению с двухуровневой схемой жестких решений дает выигрыш отношения сигнал/шум в 2 дБ. Это означает, что восьмиуровневая схема мягких решений обеспечивает такую же вероятность битовой ошибки как и двухуровневая схема жестких решений, но при отношении сигнал/шум E_b/N_0 на 2 дБ ниже. Однако за этот выигрыш от применения восьмиуровневой схемы мягких решений приходится платить повышением быстродействия процессора как минимум в 3 раза.

Оптимизация вероятности ошибки - заключается в минимизации вероятности ошибки применительно к используемому каналу связи.

Для канала с АБГШ вероятность ошибки двоичных сигналов P_b определяется выражением (1.21):

$$P_b = \int\limits_{u=\frac{a_1-a_2}{2\sigma_0}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = Q\left(\frac{a_1-a_2}{2\sigma_0}\right),$$

из которого следует, что для её минимизации необходимо обеспечить максимальное значение аргумента в функции $Q(x)$ в выражении (1.21). Очевидно, что этому будет соответствовать максимум величины $\dfrac{(a_1-a_2)^2}{\sigma_0^2}$,

где числитель представляет энергию разностного сигнала на входе согласованного фильтра в момент времени $t=T$, а $(\sigma_0)^2$ − двусторонняя спектральная плотность мощности гауссова шума, равная $N_0/2$. Известно, что если фильтр согласовывает разностный сигнал $[s_1(t)-s_2(t)]$, то соотношение сигнал/шум на его выходе в момент $t=T$ можно записать как

$$\left(\frac{S}{N}\right)_T = \frac{(a_1-a_2)^2}{\sigma_0^2} = \frac{2E_d}{N_0}, \quad (1.22), \quad \text{где} \quad E_d = \int_0^T [s_1(t)-s_2(t)]^2 dt \quad (1.23)$$

энергия разностного сигнала.

Из (1.22) $\dfrac{(a_1-a_2)}{2\sigma_0} = \sqrt{\dfrac{E_d}{2N_0}}$ Тогда $P_b = Q\left(\sqrt{\dfrac{E_d}{2N_0}}\right).$ (1.24)

$$E_d = \int_0^T s_1^2(t)dt + \int_0^T s_2^2(t)dt - 2\int_0^T s_1(t)s_2(t)dt = 2E_b(1-\rho), \quad \text{где} \quad \rho = \frac{1}{E_b}\int_0^T s_1(t)s_2(t)dt.$$

(1.25) (1.26)

коэффициент взаимной корреляции

Подставив (1.25) в (1.24), получим

$$P_b = Q\left(\sqrt{\frac{E_b(1-\rho)}{N_0}}\right). \qquad (1.27)$$

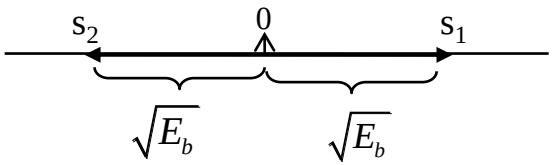
Учитывая, что $-1 \le \rho \le 1$, более удобной для анализа формой записи коэффициента корреляции является запись $\rho = \cos\theta$, где θ – угол между векторами s_1 и s_2.

При противоположных или *антиподных* сигналах, для которых $\theta = 180^o$, (Рис. 1.10, *а*) коэффициент взаимной корреляции равен $\rho=-1$. Тогда вероятность битовой ошибки для таких сигналов, как следует из (1.27), будет равна

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \qquad (1.28)$$

Двоичные сигналы с нулевой корреляцией, у которых $\theta = 90^o$ (Рис. 1.10, *б*) и коэффициент взаимной корреляции равен $\rho=0$, называются *ортогональными*. Вероятность битовой ошибки для таких сигналов, соответственно, будет равна

$$P_b = Q\left(\sqrt{\frac{E_b}{N_0}}\right). \qquad (1.29)$$



а)

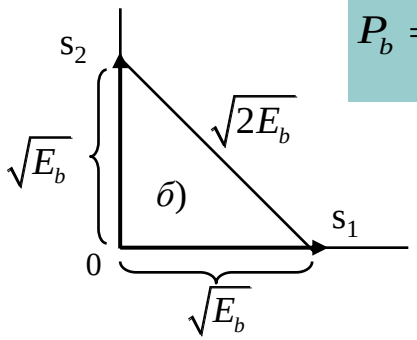


Рис. 1.10. Векторное представление двоичных сигналов s_1 и s_2 : *а*) антиподных; *б*) ортогональных.

M-ичная передача сигналов и вероятность битовой ошибки

При M-ичной передаче сигналов количество возможных сигналов равно $M=2^k$, где показатель k определяет количество бит информации, передаваемых одним M-арным символом. Ошибочный прием одного M-ичного символа приводит к возникновению от одного до k ошибочных бит. Поэтому вероятность битовой ошибки P_b в случае M-ичной передачи зависит от вероятности символьной ошибки $P_E(M)$, которая в свою очередь зависит от режима передачи сигналов и от способов модуляции. Подробно об этом описано в [1]. К примеру, вероятности битовой ошибки и вероятности символьной ошибки для ортогональных M-ичных сигналов связаны соотношением

$$\frac{P_b}{P_E(M)} = \frac{2^{k-1}}{2^k - 1} = \frac{M}{2(M-1)}. \qquad (1.30)$$

Очевидно, что $$\lim_{k\to\infty} \frac{P_b}{P_E} = \frac{1}{2}.$$

Дискретные каналы и их представления.

Вместо непрерывного канала с АБГШ при описании помехоустойчивых кодов часто используют дискретные формы представления каналов. Наибольшее распространение получили двоичные дискретные каналы, в которых информационные двоичные символы «1» и «0» складываются по модулю два с последовательностью ошибок $\{E\}= \{e_1, e_2, e_3, e_4,\ldots\}$

Для двоичного дискретного канала $e_i \in \{0.1\}$.

Если e_i=0, то ошибки нет, если e_i=1− ошибка есть.

<u>Двоичные каналы с независимыми ошибками (каналы без памяти):</u>

Двоичный симметричный канал (ДСК)

Двоичный несимметричный канал (ДНК)

Канал со стиранием.

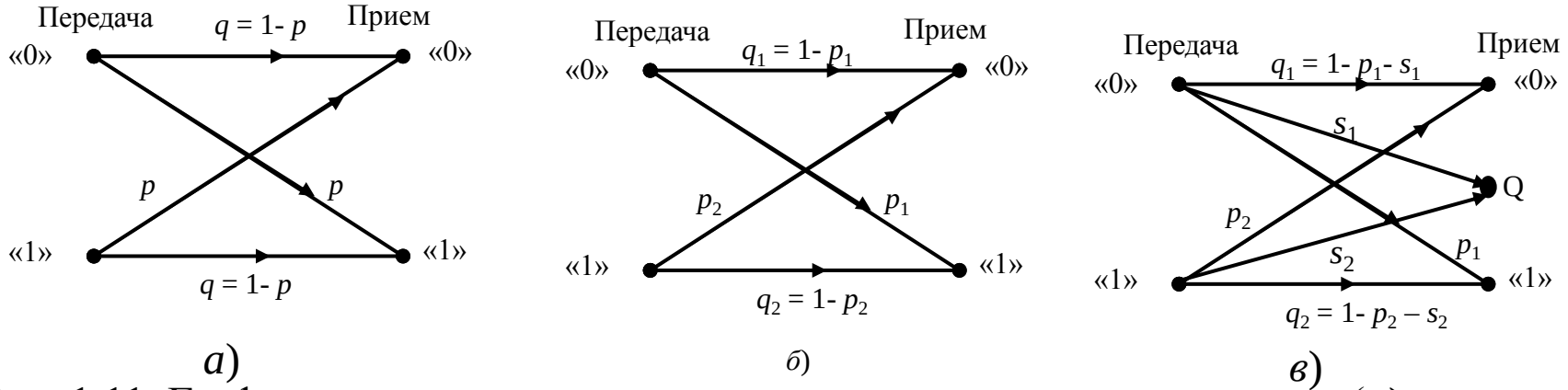


Рис. 1.11. Графическое представление двоичного симметричного канала (*а*), двоичного несимметричного канала (*б*) и канала со стираниями (*в*).

Модель двоичного симметричного канала с независимыми ошибками.

Модель канала ДСК с независимыми ошибками определяется схемой Бернулли (биномиальный закон распределения ошибок).

По формуле бинома Ньютона имеем:

$$(q+p)^n = C_n^0 q^n p^0 + C_n^1 q^{n-1} p + C_n^2 q^{n-2} p^2 + \ldots + C_n^{n-1} q p^{n-1} + C_n^n q^0 p^n. \qquad (1.32)$$

Член ряда $C_n^i q^{n-i} p^i$ представляет собой вероятность появления i ошибок в n-элементной комбинации.

$P(1,n)=P(1,5)= C_n^1 p(1-p)^{n-1} = C_5^1 p(1-p)^4.$ $\quad P(m, n)= C_n^m p^m (1-p)^{n-m}.$ $\quad P(0, n)=(1-p)^n=q^n.$

$$P(\geq 1, n)= \sum_{i=1}^{n} C_n^i p^i (1-p)^{n-i} = 1 - P(0,n) = 1-(1-p)^n.$$

Если $p<<1$ и длительность комбинации такая, что $np<<1$, то можно принять

$$P(\geq 1, n) = 1-(1-p)^n \approx np.$$

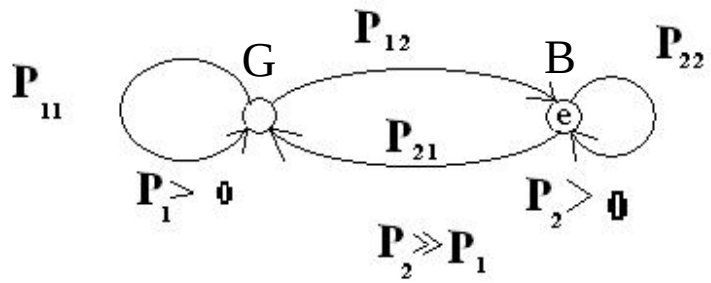
Экспериментально вероятность ошибки p оценивают частостью появления ошибок

$$p_N = \frac{M_{ош}}{N}. \qquad p_N \to p \quad \text{при} \qquad N \to \infty.$$

Модели дискретных каналов с зависимыми ошибками (модели каналов с памятью).

1) Модель Гильберта.

$$P = \begin{vmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{vmatrix}$$



***P*(*B*), *P*(*G*)- вероятности пребывания канала соответственно в плохом и хорошем состоянии.**

Существует обобщенная модель Гильберта, которая предполагает большее число состояний. Увеличение числа состояний позволяет более точно описать канал. Но при этом возникают трудности вычислительного плана. Поэтому часто переходят на модель Гильберта с тремя состояниями: хорошее, плохое, среднее.

2) Модель Беннета-Фройлиха.

В этой модели поток ошибок рассматривается как простейший поток или пуассоновский поток пачек ошибок. Канал может быть в двух состояниях: ошибок нет (идеальный канал) или пачки ошибок

$$P(k,t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \qquad P(0,t) = e^{-\lambda t}. \qquad P(k \geq 1, t) = 1 - e^{-\lambda t}.$$

3) Модель Мертца.

4).Модель Военной академии связи.

Два параметра: *p* – средняя вероятность ошибки двоичного элемента,

α – коэффициент группирования ошибок, (0≤ α≤1)

Предположим, что в результате передачи *N* элементов последовательности возникли *M* ошибок (*M* единиц)

$$p = \frac{M}{N}.$$

$$B_0(n) = \frac{N}{n} \qquad P(\geq 1, n) = \frac{B_{ош}(n)}{B_0(n)}$$

Для канала без памяти (ДСК) (α=0) эта вероятность будет в среднем равна:

$$P(\geq 1, n) = f(n) = \sum_{i=1}^{n} C_n^i p^i q^{n-i} = 1 - q^n = 1 - (1-p)^n \approx 1 - (1 - np) \approx np. \qquad (1.33)$$

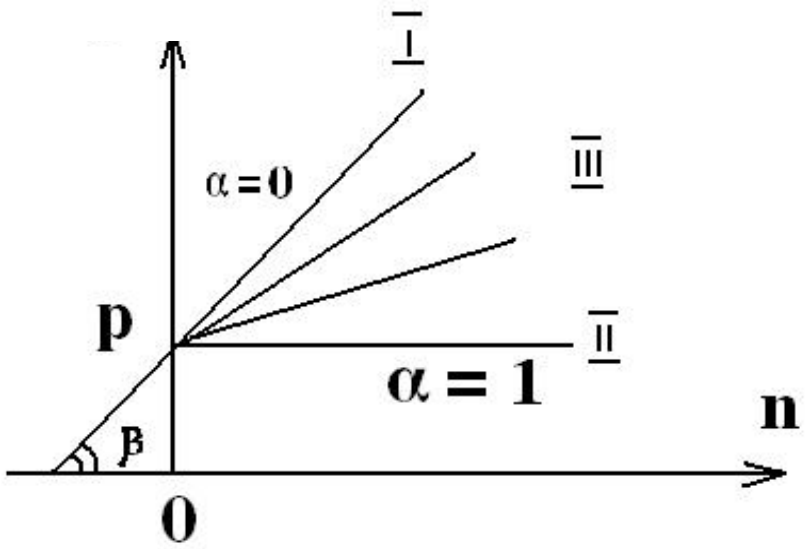
Второй граничный случай для канала спамятью будем иметь, когда все М ошибок гипотетически сгруппированы в единую пачку,т.е. идут непрерывно (α=1).

$$B_{ош}(n) = \frac{M}{n} \qquad P(\geq 1, n) = \frac{B_{ош}(n)}{B_0(n)} = \frac{M}{N} = p \qquad (1.34)$$

α=0: $\log P(\geq 1, n) = \log p + \log n = f(n).$

α=1: $\log P(\geq 1, n) = \log p$

$$\log P(\geq 1, n) = \log p + tg\beta \log n = \log p + (1 - \alpha) \log n \qquad (1.35)$$



$$P(\geq 1, n) = p \cdot n^{1-\alpha}. \qquad (1.36)$$

Вопрос 1.1. *Чем обосновывается нормальный закон распределения АБГШ?*

Вопрос 1.2. *Какой физический и вероятностный смысл имеет дисперсия белого шума* $D(x)= N_0/2$*?*

Вопрос 1.3. Чем объясняется то, что для аналоговой связи в качестве критерия качества выбрано отношение средней мощности сигнала к средней мощности шума S/N, а для цифровой – нормированное отношение энергии битового сигнала к спектральной плотности мощности шума E_b/N_0*?*

Вопрос 1.4. Показать, в каких единицах измеряются величины E_b *и* N_0*?*

Упражнение 1. Используя компьютерные пакеты программ, построить графики вероятности битовой ошибки P_b *по формулам (1.28) и (1.29) в зависимости от отношения сигнал/шум* E_b/N_0*.. Сделать сравнительный анализ графиков.Обосновать выводы.*

Литература.

1. Скляр, Д. Цифровая связь. Теоретические основы и практическое применение / Д. Скляр; пер. с англ. – М. :Издательский дом «Вильямс», 2003. – 1104 с.
2. Морелос-Сарагоса, Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение / Р. Морелос-Сарагоса; пер. с англ. – М. : Техносфера, 2006. – 319 с.

Энергетическая эффективность помехоустойчивого кодирования.

Эффективность помехоустойчивого кода, наряду с такими показателями как скорость R кода и достоверность передачи информации, часто оценивается энергетическим выигрышем от применения такого кодирования [1,2]. В качестве энергетического показателя системы связи выбирают отношение сигнал/шум, которое требуется для достижения заданной вероятности ошибки. Энергетический выигрыш оценивается как разность значений соотношений сигнал\шум, которые обеспечивают заданную вероятность ошибки в системе без помехоустойчивого кодирования и системе с применением помехоустойчивого кода. При этом энергетический выигрыш измеряется, как правило, в дБ.

Ниже приводится пример сравнения вероятностей ошибок при передаче двоичных сигналов с двоичной фазовой модуляцией BPSK информационными комбинациями длиной k=11 элементов каждая. Сравнение проводится для обычной передачи со скоростью R_1=4800 бит/с без помехоустойчивого кодирования и передачей с блочным помехоустойчивым кодированием (n,k)=(15,11) с исправлением однократных битовых ошибок. Рассматривается синхронная система связи с непрерывной передачей, поэтому скорость передачи двоичных сигналов при использовании помехоустойчивого кода должна быть увеличена в n/k раз, т.е. равной R_2=4800 x 15/11=6545 бит/с.

Вероятность битовой ошибки в канале с АБГШ и с модуляцией BPSK с антиподными (противоположными) сигналами и когерентным приемом оценивается выражением

$$p_b = \int_{\sqrt{2E_b/N_0}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = Q\left(\sqrt{\frac{2E_b}{N_0}}\right), \quad p_1 = Q\left(\sqrt{\frac{2E_1}{N_0}}\right), \quad p_2 = Q\left(\sqrt{\frac{2E_2}{N_0}}\right),$$

Сравним теперь вероятности выдачи потребителю после декодирования комбинации, содержащей k=11 информационных элементов, с ошибками.

В случае передачи простым кодом без помехоустойчивого кодирования вероятность ошибочного приема комбинации будет равна

$$P_1(k{=}11){=}1-(1-p_1)^k.$$

Если же использовать помехоустойчивый блочный код (n,k)=(15,11) с исправлением однократных битовых ошибок, то вероятность получения ошибочного сообщения будет равна

$$P_2(n,k)=P_2(15,11)=\sum_{j=2}^{n=15} C_n^j (p_2)^j (1-p_2)^{n-j}.$$

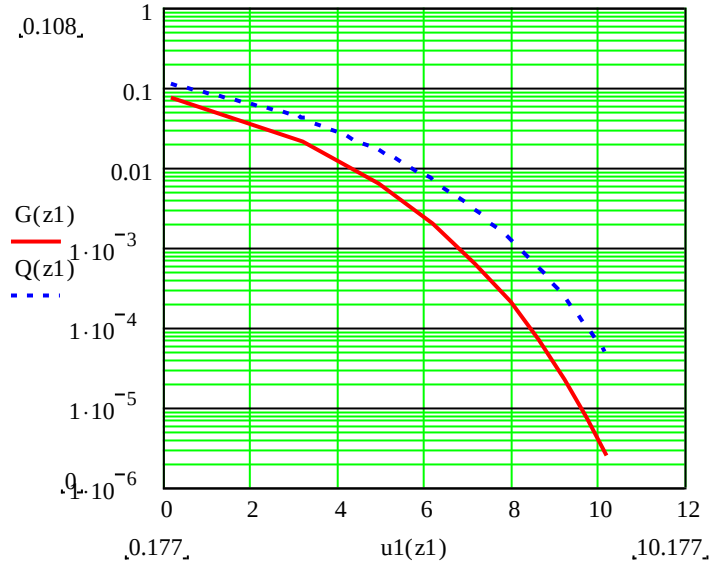


Рис. 2.3. Зависимости вероятностей битовой ошибки при передаче простым кодом длиной k=11 (сплошная кривая) и избыточным кодом (15,11) с исправлением одиночных ошибок (пунктирная кривая)

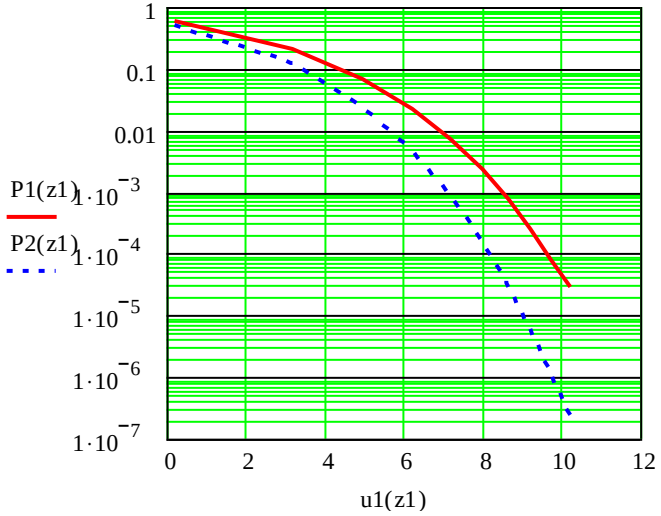


Рис. 2.4. Зависимости вероятностей ошибочного получения комбинации при передаче простым кодом длиной k=11 (сплошная кривая) и избыточным кодом (15,11) с исправлением одиночных ошибок (пунктирная кривая)

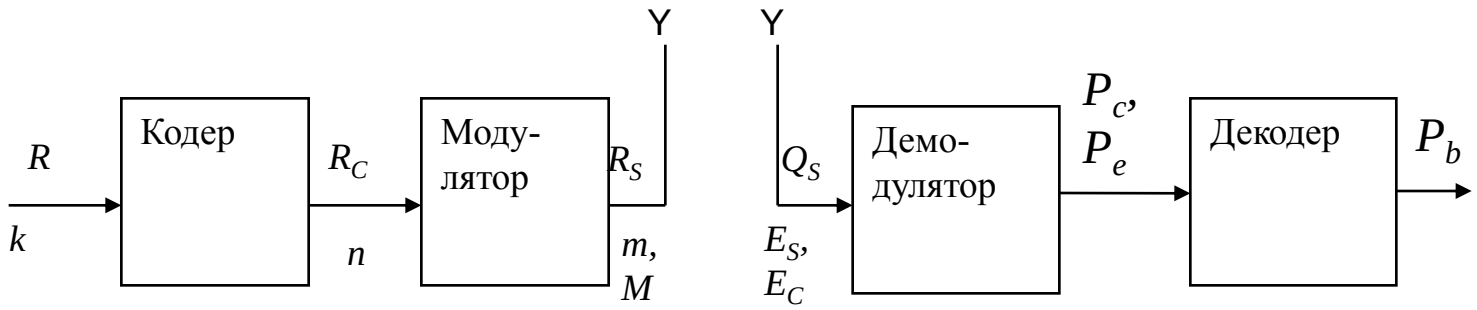
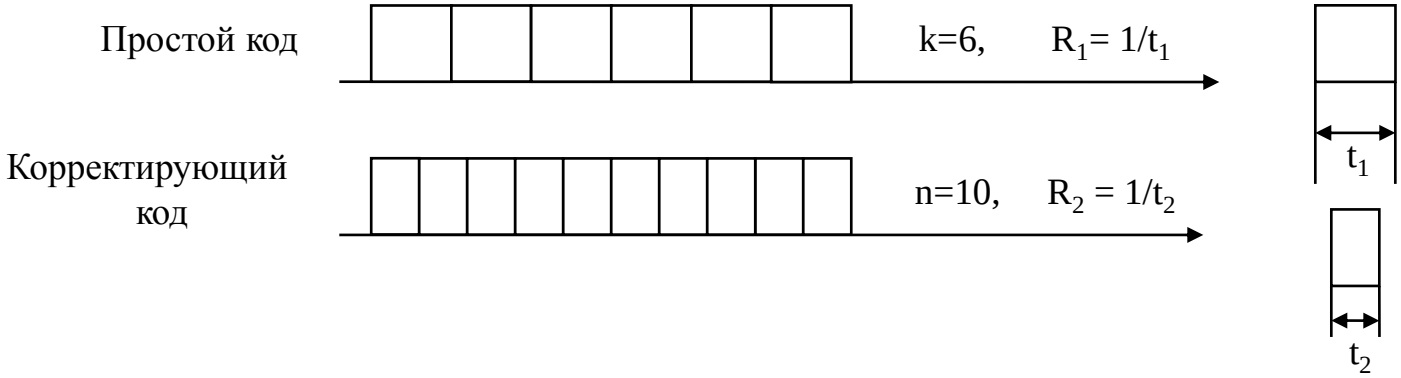


Рис. 2.5. Обобщённая схема системы передачи информации



$R_2 > R_1$, $\Delta F_2 > \Delta F_1$, $t_2 = k/n \cdot t_1$, $R_2 = n/k \cdot R_1$

Рис. 2.6

R – скорость передачи информации от источника, бит/с;
k – количество информационных бит в комбинации на входе кодера;
n – количество элементов в кодовой комбинации избыточного помехоустойчивого кода (n > k) на выходе кодера;
$R_C = (n/k) \cdot R$ – скорость передачи канальных бит на выходе кодера, бит/с;
$R_S = R_C / m = R_C / \log_2 M$ – скорость передачи сигналов на выходе модулятора (количество сигналов в сек, Бод);
M – размерность сигнала (количество возможных различных сигналов на выходе модулятора);
m – количество бит информации в одном сигнале на выходе модулятора;
E_S – энергия сигнала на входе приемника;
E_C – энергия сигнала на входе приемника, приходящаяся на один канальный бит;
E_b – энергия сигнала на входе приемника, приходящаяся на один бит информации;
$P_S = E_S R_S = E_C R_C = E_b \cdot R$ – мощность сигнала на входе приемника;
p_c – вероятность ошибочного приема канального бита на выходе демодулятора;
p_e – вероятность ошибочного приема сигнала на выходе демодулятора;
p_b – вероятность ошибочного приема информационного бита на выходе декодера.

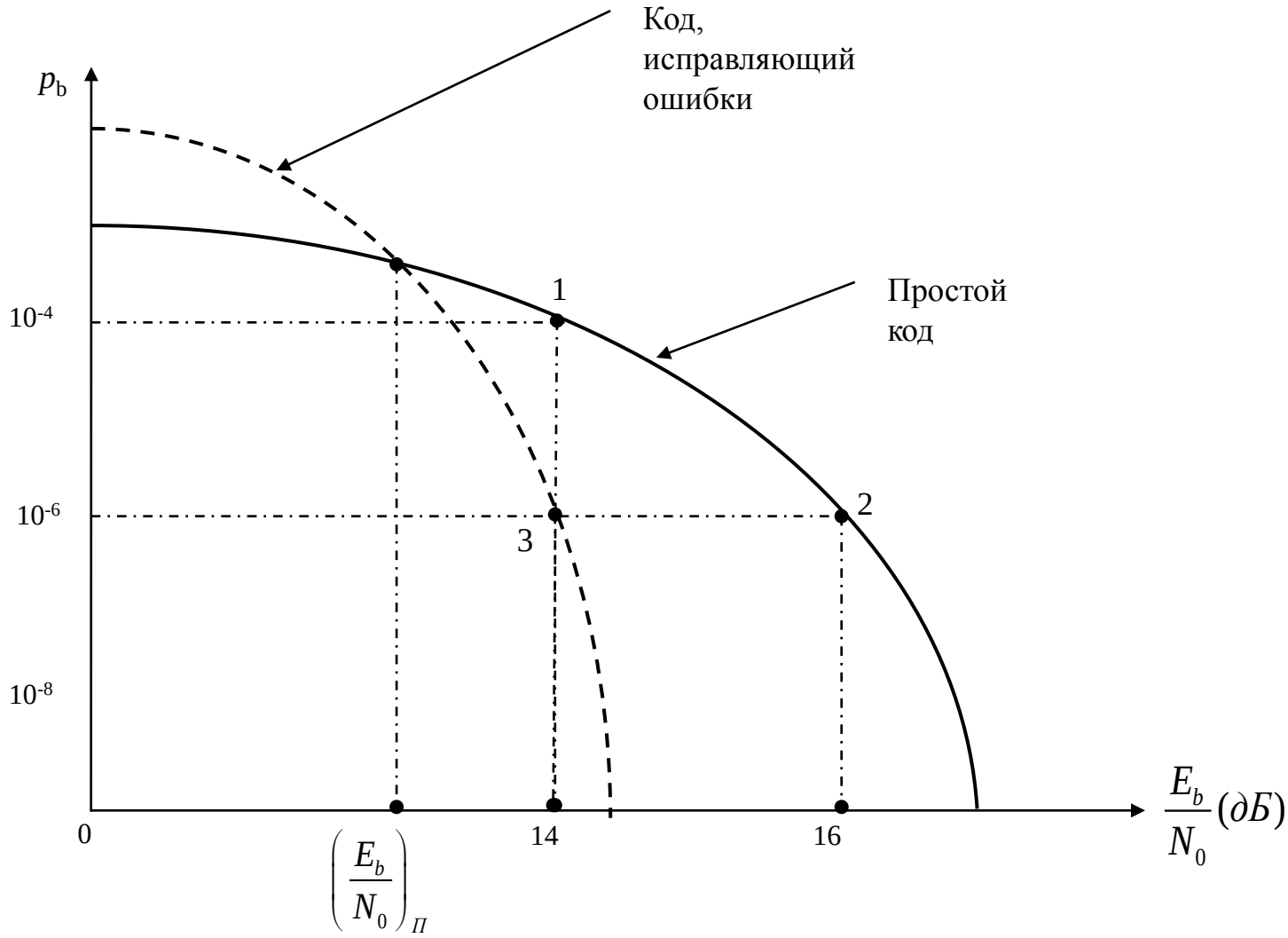


Рис.2.7. Вероятности битовой ошибки при передаче сообщений простым кодом и кодом, корректирующим ошибки

<u>Случай 1</u>. Применение корректирующего ошибки кода с целью повышения достоверности передачи данных при сохранении энергетических затрат на бит передаваемой информации.

Исходное состояние системы при использовании простого кода находится (рис.1) в рабочей точке 1:

$$P_b = 10^{-4}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 14\,дБ; \quad \Delta F_1 = \frac{1}{t_1}.$$

После применения корректирующего кода система переходит в точку 3 на кривой корректирующего кода:

$$P_b = 10^{-6}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 14\,дБ; \quad t_2 < t_1; \quad \Delta F_2 = \frac{1}{t_2} > \Delta F_1 = \frac{1}{t_1}.$$

<u>Случай 2</u>. Целью применения кода, корректирующего ошибки, может быть снижение энергетических затрат на передачу одного бита информации при сохранении требуемой достоверности передачи.

Исходное состояние системы при использовании простого кода находится (рис.1) в рабочей точке 2:

$$P_b = 10^{-6}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 16\,дБ; \quad \Delta F_1 = \frac{1}{t_1}.$$

После применения корректирующего кода система переходит в точку 3 на кривой корректирующего кода:

$$P_b = 10^{-6}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 14\,дБ; \quad t_2 < t_1; \quad \Delta F_2 = \frac{1}{t_2} > \Delta F_1 = \frac{1}{t_1}.$$

Эффективность помехоустойчивого кода составила

$$\left(\frac{E_0}{N_0}\right)_{ПК} - \left(\frac{E_0}{N_0}\right)_{КК} = 2\,дБ.$$

<u>Случай 3.</u> Целью применения помехоустойчивого кода, исправляющего ошибки, является увеличение скорости передачи информации при сохранении требуемой достоверности без увеличения мощности передатчика.

Этот случай наиболее вероятен в различных радиотехнических системах связи, в том числе спутниковых, где накладываются жесткие ограничения на мощность сигналов на выходе передатчика.

Исходное состояние системы при использовании простого кода и однократной модуляции находится (рис.1) в рабочей точке 2:

$$P_b = 10^{-6}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 16\,дБ; \quad R_1 = \frac{1}{t_1}.$$

После перехода на многократную модуляцию помехоустойчивость сигнала уменьшается и система при использовании простого кода переходит в рабочую точку 1:

$$P_b = 10^{-4}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 14\,дБ; \quad R_2 > R_1.$$

После применения корректирующего кода система переходит в точку 3 на кривой корректирующего кода:

$$P_b = 10^{-6}, \quad \left(\frac{E_0}{N_0}\right)_{ПК} = 14\,дБ; \quad R_{кан} > R_2 > R_1 = \frac{1}{t_1}.$$

Задание на самостоятельную работу.

Решить задачу по выбору помехоустойчивого кода, корректирующего ошибки, с целью обеспечения достоверности приема с вероятностью битовой ошибки не хуже $P_b = 10^{-9}$ при сохранении скорости передачи информации R, энергетических затрат и допустимой полосы частот.

Исходные данные:

В системе используются сигналы с ФМн размерностью M=8;

Скорость передачи информации от источника R = 9600 бит/с;

Допустимая полоса частот канала F = 4000 Гц;

Энергетические затраты в системе

$$\left(\frac{E_b}{N_0} \right) = 13 дБ \, (или \; 20,89);$$

Последовательность решения поставленной задачи

Шаг первый. Если не задан вид модуляции в исходных данных, то необходимо выбрать вид модуляции таким образом, чтобы полоса частот сигнала была меньше от допустимой на величину, достаточную для введения избыточных элементов в исходную комбинацию. Окончательный выбор вида модуляции должен производиться после учета выполнения всех требований к помехоустойчивому коду.

Шаг второй - определение возможных вариантов кода при условии, что требуемая полоса частот после введения избыточных элементов не превысит допустимую.

Третий шаг – оценка вероятности P_E ошибочного приема сигнала на входе демодулятора (выходе канала) в зависимости от заданного в исходных данных отношения сигнал/шум в канале АБГШ.

Четвертый шаг – оценка вероятности P_C ошибочного приема канального бита.

Пятый шаг – оценивание итоговой вероятности битовой ошибки P_b на выходе декодера с учетом корректирующих свойств выбранного помехоустойчивого кода в системе реального времени.

Литература.

1. Скляр, Д. Цифровая связь. Теоретические основы и практическое применение / Д. Скляр; пер. с англ. – М. :Издательский дом «Вильямс», 2003. – 1104 с.

2. Морелос-Сарагоса, Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение / Р. Морелос-Сарагоса; пер. с англ. – М. : Техносфера, 2006. – 319

3. Когновицкий О.С.,Охорзин В. М. Теория помехоустойчивого кодирования. Часть 1. Циклические коды.: Учебное пособие/ СПбГУТ. – СПб., 2013. – 94 с.

Передача сообщений по каналам с помехами. помехоустойчивое кодирование

Вторая теорема Шеннона о возможности передачи информации с произвольно малой частотой ошибок в дискретном канале.

Теорема звучит следующим образом:

«Если скорость производства информации источником сообщений меньше пропускной способности канала, то существует такая система кодирования, что сообщения источника могут быть переданы по каналу с произвольно малой частотой ошибок (или со сколь угодно малой ненадежностью)».

Докажем эту теорему на частном примере двоичного симметричного канала.

Допустим существует источник на выходе которого может появиться одно из N сообщений A_i подлежащих передаче. Обозначим эти сообщения через (A_1, A_2, A_3…, A_N).

Каждое сообщение появляется со своей ему присущей вероятностью P_i, т.е. (P_1, P_2, P_3…P_N)

Энтропия источника H, т.е. количество двоичных единиц информации приходящихся в среднем на одно сообщение, равна:

$$H = -\sum_{i=1}^{N} P_i \log_2 P_i$$

Для передачи сообщений по каналу связи сообщения кодируются двоичным кодом. Предположим теперь, что кодируется не каждое отдельное сообщение A_i, а S сообщений вместе (S-цепочки). **Если брать S достаточно большим**, то каждое сообщение A_i в S-цепочке появится в среднем (с большой вероятностью) $P_i \cdot S$ раз. Таким образом сообщение A_1 появится в какой-либо S-цепочке P_1S раз, сообщение A_2 – P_2S раз и т.д.

Если сообщения являются взаимно независимыми, то можно найти вероятность P какой либо реализации блока из S сообщений как

$$P = \prod_{i=1}^{N} (P_i)^{P_i S}$$

Таким образом S-цепочки можно считать **равновероятными** и, следовательно, их количество можно приближенно определить как: $Мвер = \frac{1}{P}$

Учитывая, что все S-цепочки равновероятны, их можно закодировать в соответствии с первой теоремой Шеннона равномерным k-элементным двоичным кодом, где

$$k = \log_2 Мвер = -S\sum_{i=1}^{N} P_i \log_2 P_i = SH.$$

Для нахождения n определим наиболее вероятное количество ошибок r в одной кодовой комбинации длины n при условии, что S, а следовательно и n, являются достаточно большими. Тогда можно считать $r=Pэ \cdot n$, где $Pэ$ -вероятность ошибки на один кодовый элемент.

Число сочетаний C_n^r дает максимально возможное количество кодовых комбинаций длины n, в которых искажены равно r элементов и которые соответствуют одной S-цепочке.

Таким образом, если на приеме каждому блоку из S исходных сообщений (разрешенной n-элементной комбинации) поставить в соответствие C_n^r кодовых комбинаций, составленных определенным образом, то большая часть ошибок с большой вероятностью будет обнаруживаться и исправляться.

Тогда для однозначно правильного декодирования всех *S*-цепочек понадобится всего Mвер$^* C_n^r$ кодовых *n*-элементных комбинаций. Общее количество комбинаций *n*-элементного двоичного кода равно 2^n. Следовательно, чтобы закодировать все Mвер сообщений (*S*-цепочек) различными *n*-элементными кодовыми комбинациями нужно, чтобы выполнялось условие $2^n \geq M\text{вер}^* C_n^r$

Логарифмируя это неравенство, находим

$$n \geq \log_2 M_{вер} + \log_2 C_n^r = SH + \log_2 \frac{n!}{(p_э \cdot n)!(\vartheta \cdot n)!}, \qquad \text{где} \qquad \vartheta = 1 - P_э$$

$$\log_2 n! \to n \log_2 n$$

$$n \geq SH + n\log_2 n - nP_э \log_2 (nP_э) - n\vartheta \log_2 (n\vartheta) =$$
$$= SH + n\log_2 n - nP_э \log_2 P_э - nP_э \log_2 n - n\vartheta \log_2 \vartheta - n\vartheta \log_2 n =$$
$$= SH - n(P_э \log_2 P_э + \vartheta \log_2 \vartheta).$$

Поделив обе части неравенства на *n*, после перестановки получим: $\frac{SH}{n} \leq 1 + P_э \log_2 P_э + \vartheta \log_2 \vartheta.$

$$R = \frac{SH}{n} = \frac{k}{n}.$$

Учитывая, что правая часть неравенства представляет собой пропускную способность двоичного симметричного канала C, получаем неравенство R≤C, что и требовалось доказать

Классификация помехоустойчивых кодов (ПК):

1) *по корректирующим возможностям*:
-обнаруживающие ошибки;
-исправляющие ошибки;
-исправляющие стирания;
-исправляющие ошибки и стирания;
-комбинированные ПК.

2) *по структурным особенностям*:
-блочные (блоковые);
-непрерывные;
-каскадные

Основание кода.
Информационные и избыточные элементы.
Избыточность. Избыточность по элементам и избыточность по основанию кода.
Систематические и несистематические помехоустойчивые коды.
Весовой спектр кода.

Блочные (n,k)-коды:

k-элементные информационные комбинации формируют множество из $N=2^k$ информационных сообщений; в процессе кодирования каждой k-элементной информационной комбинации взаимно однозначно сопоставляется одна из $M=2^n$ n-элементных комбинаций.

Количество *разрешенных* кодовых n-элементных комбинаций – $N=2^k$,
а количество запрещенных n-элементных комбинаций – $(M-N)=2^n-2^k$.

Избыточность блочного (n,k)-кода:

по комбинациям:

- абсолютная – $(M-N)=2^n-2^k$;
- относительная – $(M-N)/M=(2^n-2^k)/2^n$;

по элементам кодовой комбинации:

- абсолютная – $(n-k)$;
- относительная – $\eta=(n-k)/n$.

Кодовая скорость (степень кодирования) – $R=k/n=1-\eta$.

Достоверность передаваемой информации.

Линейный блочный код:

Пусть V_i и V_j - два кодовых слова в двоичном блочном (n,k)-коде. Код называется линейным тогда и только тогда, когда их поэлементная сумма по модулю 2 ($V_i \oplus V_j$) также является кодовым словом этого же кода.

В линейном блочном (n,k)-коде всегда существует так называемый *базис*, представляющий собой подмножество k линейно независимых комбинаций $V_1, V_2, .. V_k$, используемых для генерации других кодовых слов блочного (n,k)-кода следующим образом:

$$U = m_1 V_1 + m_2 V_2 + \ldots + m_k V_k, \qquad (2.1)$$

где m_i =0 или 1 (для двоичных кодов), i =1,…, k.

Базисные n-элементные вектора $V_i = [v_{i1}, v_{i2}, \ldots, v_{in}]$ определяют матрицу *генерации* G, называемую также *образующей* или *порождающей* матрицей:

$$G = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{bmatrix}. \qquad (2.2)$$

Тогда кодовое слово (2.1) получается как матричное произведение

$$U = m\,G, \qquad (2.3)$$

где $\boldsymbol{m}$ – k-элементная информационная вектор-строка $\boldsymbol{m} = [m_1, m_2, \ldots, m_k]$.

Порождающая матрица систематического линейного блочного (*n*,*k*)-кода

Систематический линейный блочный (*n*,*k*)-код – это такой код, в *n*-элементной комбинации которого четко определены *k* позиций, занимаемых информационными элементами, а остальные (*n*–*k*) позиций занимают проверочные (контрольные) биты. В большинстве систематических блочных линейных (*n*,*k*)-кодов порождающая матрица G имеет вид:

$$G = \left[P \vdots E_k \right] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1(n-k)} & 1 & 0 & \ldots & 0 \\ p_{21} & p_{22} & \cdots & p_{2(n-k)} & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{k(n-k)} & 0 & 0 & \ldots & 1 \end{bmatrix}, \qquad (2.4)$$

где E_k – единичная матрица размерностью $k \times k$, а P – матрица проверочных элементов размерностью $k \times (n\text{-}k)$.

Подставив в (2.3) значения матрицы G из выражения (2.4) и вектор-строки k информационных элементов $\boldsymbol{m}$, получим n-элементное кодовое слово систематического линейного блочного (*n*,*k*)-кода в следующем виде:

$$U = p_1, p_2, \ldots, p_{n-k}, m_1, m_2, \ldots, m_k, \qquad (2.5)$$

где

$$\begin{aligned} p_1 &= m_1 p_{11} + m_2 p_{21} + \ldots + m_k p_{k1}; \\ p_2 &= m_1 p_{12} + m_2 p_{22} + \ldots + m_k p_{k2}; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_{n-k} &= m_1 p_{1(n-k)} + m_2 p_{2(n-k)} + \ldots + m_k p_{k(n-k)}. \end{aligned} \qquad (2.6)$$

Проверочная матрица линейного систематического блочного (*n*,*k*)-кода

Для прождающей матрицы *G* любого блочного (*n*,*k*)-кода существует *проверочная* матрица *H* размерности (*n*-*k*)×*n*, с помощью которой производится декодирование полученной *n*-элементной комбинации. Свойством проверочной матрицы *H* является свойство *ортогональности*, которое записывается как

$$GH^T=Ø, \qquad (2.7)$$

где H^T – транспонированная матрица *H*, а *Ø* – нулевая матрица размерности *k* x (*n*-*k*). Учитывая свойство линейности блочного (*n*,*k*)-кода, из (2.7) следует, что произведение любого генерируемого матрицей *G* кодового вектора *U* на H^T дает нулевую вектор-строку длиной (*n*-*k*), т.е.

$$UH^T=0 \qquad (2.8)$$

Именно это позволяет контролировать принадлежность кодового слова *U* множеству разрешенных комбинаций, порождаемых матрицей G.

Для систематического линейного блочного (*n*,*k*)-кода проверочная матрица *H* может иметь следующий вид:

$$H=\left[\, E_{n-k} \;\vdots\; P^T \,\right], \qquad (2.9)$$

где E_{n-k} – единичная матрица размерности (*n*-*k*)×(*n*-*k*), а P^T – транспонированная матрица *P* проверочных элементов из порождающей матрицы *G* в (2.4). Тогда матрица *HT* будет следующей:

$$H^T = \begin{bmatrix} E_{n-k} \\ \cdots \\ P \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \\ p_{11} & p_{12} & \cdots & p_{1(n-k)} \\ p_{21} & p_{22} & \cdots & p_{2(n-k)} \\ \vdots & \vdots & \cdots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{k(n-k)} \end{bmatrix}. \qquad (2.10)$$

Легко проверить, что умножив вектор-строку кодового слова U в виде (2.5) на проверочную матрицу (2.9), получим для двоичного кодирования

$$UH^T = p_1 + p_1, p_2 + p_2, ..., p_{n-k} + p_{n-k} = \varnothing.$$

Обнаружение и исправление ошибок блочным (*n*,*k*)-кодом на основе синдромов

Пусть $r = [r_1, r_2, \ldots, r_n]$ – принятый вектор после передачи кодового слова $U=[u_1, u_2, \ldots, u_n]$, на которое в канале наложился вектор ошибок $e= [e_1,e_2,\ldots,e_n]$, т.е. r является поэлементной суммой по модулю 2 (для двоичных кодов) векторов U и e

$$r = U + e. \qquad (2.11)$$

Проверка принадлежности принятого вектора r множеству разрешенных комбинаций блочного (n,k)-кода производится при декодировании с помощью *синдрома*, который определяется следующим образом:

$$S=rH^T. \qquad (2.12)$$

Если в принятом векторе ошибки отсутствуют, то $r = U$ и синдром S=0. И напротив, если синдром $S\neq0$, то это свидетельствует о том, что в принятом векторе имеются ошибки. При этом во многих системах связи процесс декодирования данной комбинации r заканчивается. В этом случае блочный (n,k)-код работает в *режиме обнаружения ошибок*.

В других системах декодер работает в *режиме прямого исправления ошибок.* Это становится возможным благодаря важной особенности линейных блочных (n,k)-кодов – взаимно однозначному соответствию между синдромом и исправимой комбинацией ошибок. Действительно, используя (2.11) и (2.12) можно записать

$$S=rH^T= (U+e)H^T= U\,H^T +e\,H^T. \qquad (2.13)$$

Учитывая (2.8), имеем

$$S=rH^T= eH^T. \qquad (2.14)$$

Таким образом, найдя синдром S≠0, можно определить соответствующий ему вектор ошибок и произвести их исправление.

Вес и расстояние Хемминга между кодовыми векторами.

Число ненулевых элементов в кодовом слове U называют его *весом* и обозначают $w(U)$.

Обобщённой характеристикой кода, увязывающей избыточность (скорость) и корректирующие способности кода, является *расстояние Хемминга* между векторами U и V, которое обозначается как $d(U,V)$ и определяется количеством одноименных позиций с отличающимися друг от друга кодовыми элементами. Для двоичных кодов расстояние Хемминга между векторами U и V равно весу их поэлементной суммы по модулю 2, т.е.

$d(U,V)= w(U \oplus V)$.

Минимальное расстояние и его влияние на корректирующие способности линейного кода.

Параметром блокового помехоустойчивого кода является наименьшее значение Хеммингового расстояния d всех сравниваемых пар кодовых комбинаций. Этот параметр называется минимальным кодовым расстоянием Хемминга и обозначается $d_{\min}$. Поэтому часто равномерный блоковый (n,k)-код, имеющий параметр $d_{\min}$, записывают как $(n,k,d_{\min})$ или просто (n,k,d).

Блочный линейный код может быть *групповым*, который среди своих кодовых комбинаций содержит также комбинацию, состоящую из одних нулевых элементов. Для такого группового кода характерно, что минимальный вес $w_{\min}$ ненулевой кодовой комбинации равен минимальному расстоянию Хемминга $d_{\min}$.

Для большинства блочных помехоустойчивых кодов декодер принимает решение по принципу *максимального правдоподобия*, суть которого состоит в том, что декодер декодирует принятую *n*-элементную комбинацию в ближайшую к ней разрешенную кодовую комбинацию по расстоянию Хемминга. Другими словами, декодер определяет расстояние Хемминга между принятым вектором *r* и всеми возможными разрешенными кодовыми векторами U_j и выбирает наиболее правдоподобное кодовое слово *Ui*, для которого единственного выполняется условие,что

$$d(r,U_i)<d(r,U_j) \text{ для всех } j\neq i. \qquad (2.15)$$

Если мы хотим построить помехоустойчивый код, исправляющий все сочетания из *t* или менее ошибочных элементов в любой принятой комбинации, то необходимо обеспечить значение минимального кодового расстояния, удовлетворяющее равенству

$$t = \left\lfloor \frac{d_{\min} - 1}{2} \right\rfloor, \qquad (2.16)$$

где $\lfloor \cdot \rfloor$ обозначает целую часть дроби.

Таким образом, для исправления всех сочетаний ошибок из *t* или менее ошибочных символов необходимо и достаточно, чтобы каждая разрешенная комбинация отличалась от любой другой разрешенной комбинации кода не менее, чем в (2*t*+1) позициях.

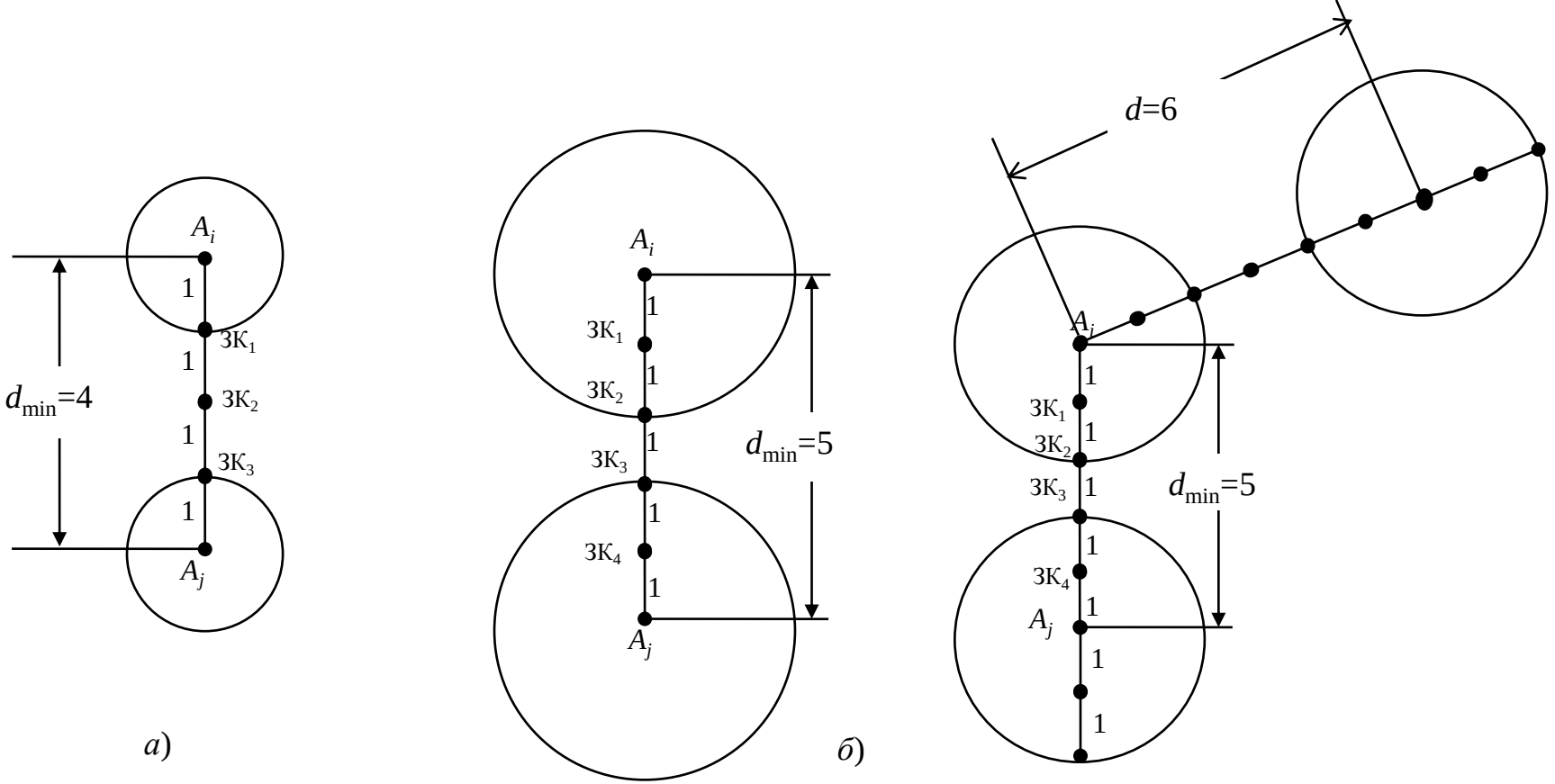


Рис.2.2. Связь минимального кодового расстояния d_{min} и корректирующих свойств помехоустойчивого кода для d_{min}=4 (*а*) и d_{min}=5 (*б*) («1» на рисунке – ошибка)

Из рассмотренных выше примеров следует вывод, что по критерию наименьшего расстояния Хемминга блоковый код гарантированно исправляет все ошибки до половины минимального расстояния Хемминга (так называемого конструктивного расстояния).

Если помехоустойчивый код работает только в *режиме обнаружения* ошибок, то, анализируя рис. 2.2, можно заключить, что в этом режиме помехоустойчивый код гарантировнно обнаруживает все ошибки до кратности σ включительно, равной

$$\sigma \le (d_{\min} - 1). \qquad (2.17)$$

Кроме того, код будет обнаруживать и определенную долю ошибок большей кратности, которые переводят разрешенную кодовую комбинацию в запрещенную. Поэтому блочный (n,k)-код в режиме обнаружения ошибок может обнаружить $2^n - 2^k$ ошибочных комбинаций. Кроме того, из-за ошибок переданная кодовая комбинация U_i может перейти в любую другую разрешенную комбинацию, количество которых равно $2^k - 1$. В этом случае синдром ошибочной комбинвции будет равен 0 и, следовательно, ошибки не будут обнаружены кодом. Вероятность *необнаруженных* ошибок зависит от весового спектра кода. Весовой спектр представляет собой распределение так называемых весовых коэффициентов $A_0, A_1, A_2, \ldots, A_n$, где A_j – количество кодовых слов с весом j. Таким образом, если двоичный блочный (n,k)-код используется только для обнаружения ошибок, то в двоичном симметричном канале вероятность поступления в декодер кодового слова с необнаруженными ошибками будет равна

$$P_{но} = \sum_{j=1}^{n} A_j p^j (1-p)^{n-j}, \qquad (2.18)$$

где p – вероятность битовой ошибки в канале.

Кроме рассмотренных выше двух режимов блочный (n,k)-код может работать в режиме гарантированного исправления ошибок до кратности t включительно и дополнительно обнаруживать все ошибки до кратности σ включительно. При этом граничное максимальное значение σ зависит от четности минимального кодового расстояния $d_{\min}$. Пусть $d_{\min}$ нечетное. Тогда при заданной кратности исправляемых ошибок t от 0 до $\left[\frac{d_{\min}-1}{2}-1\right]$ блочный (n,k)-код может ещё дополнительно обнаруживать также все ошибки кратностью от (t+1) до [$d_{\min}$– (t+1)] включительно. Если же кратность исправляемых ошибок задана от 1 до $t=\left[\frac{d_{\min}-1}{2}\right]$, включительно, то в этом случае код не может еще дополнительно обнаруживать ошибки большей кратности в рамках конструктивного расстояния $d_{\min}$. Пример для $d_{\min}$ =7 приведен в таблице 2.1.

Таблица 2.1

Кратности исправляемых ошибок, t $0\le t\le \left[\frac{d_{\min}-1}{2}\right]$,	Кратности дополнительно обнаруживаемых ошибок, σ	
	от (t+1)	до [$d_{\min}$ – (t+1)] включительно
0 (режим обнаружения)	1	6
1	2	5
≤ 2	3	4
≤ 3	-	-

Для четного $d_{\min}$ максимальная кратность исправляемых ошибок равна $\left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$, где $\lfloor\ \rfloor$ – целая часть дроби. Аналогично, при заданной кратности исправляемых ошибок t блочный (n,k)-код может ещё дополнительно обнаруживать также все ошибки кратностью от $(t+1)$ до $[d_{\min}-(t+1)]$ включительно. Пример для $d_{\min}=6$ приведен в таблице 2.2.

Таблица 2.2

Кратности исправляемых ошибок, t $0 \le t \le \left\lfloor \frac{d_{\min}-1}{2} \right\rfloor$,	Кратности дополнительно обнаруживаемых ошибок, σ	
	от $(t+1)$	до $[d_{\min}-(t+1)]$ включительно
0 (режим обнаружения)	1	5
1	2	4
≤ 2	3	3

Теперь покажем на примере двоичных кодов (основание кода 2), как связана избыточность блочного (*n*,*k*)-кода с кратностью исправляемых ошибок. В случае исправления всех ошибок до кратности *t* включительно подмножество комбинаций длины *n*, соответствующее некоторой разрешенной комбинации *Ai*, будет содержать саму разрешенную комбинцию *Ai* (безошибочный прием) и $(C_n^1 + C_n^2 + ... + C_n^t)$ запрещенных комбинаций длины *n*, где C_n^i число сочетаний из *n* по *i* . Отсюда следует, что число *N* разрешенных комбинаций блочного (*n*,*k*)-кода будет ограничено значением

$$N \le \frac{2^n}{\sum_{i=0}^{t} C_n^i}, \qquad (2.19)$$

где C_n^i – количество запрещенных комбинаций длины *n*, отличающихся от *Ai* в *i* позициях. Наиболее часто встречаются блочные (*n*,*k*)-коды, у которых число разрешенных комбинаций равно $N = 2^k$. В этом случае выражение (2.19) принимает вид

$$2^k \le \frac{2^n}{\sum_{i=0}^{t} C_n^i} \quad \text{или} \quad 2^{n-k} \ge \sum_{i=0}^{t} C_n^i. \qquad (2.20) \qquad n-k \ge \log_2 \sum_{i=0}^{t} C_n^i. \qquad (2.21)$$

Из этого следует, что абсолютная избыточность двоичного помехоустойчивого кода, исправляющего однократные ошибки (*t*=1), определяется неравенством

$$n-k \ge \log_2(1+n).$$

Эта граница для недвоичных кодов (основание кода *p*≠2): $n-k \ge \log_p \left[\sum_{i=0}^{t} C_n^i (p-1)^i \right]$.

Литература.

Когновицкий О.С.,Охорзин В. М. Теория помехоустойчивого кодирования. Часть 1. Циклические коды.: Учебное пособие/ СПбГУТ. – СПб., 2013. – 94 с.

Вопрос 1.1. *Чем обосновывается нормальный закон распределения АБГШ?*

Вопрос 1.2. *Какой физический и вероятностный смысл имеет дисперсия белого шума* $D(x)= N_0/2$*?*

Вопрос 1.3. Чем объясняется то, что для аналоговой связи в качестве критерия качества выбрано отношение средней мощности сигнала к средней мощности шума S/N, а для цифровой – нормированное отношение энергии битового сигнала к спектральной плотности мощности шума E_b/N_0*?*

Вопрос 1.4. Показать, в каких единицах измеряются величины E_b *и* N_0*?*

Упражнение 1. Используя компьютерные пакеты программ, построить графики вероятности битовой ошибки P_b *по формулам (1.28) и (1.29) в зависимости от отношения сигнал/шум* E_b/N_0*.. Сделать сравнительный анализ графиков.Обосновать выводы.*

Литература.

1. Скляр, Д. Цифровая связь. Теоретические основы и практическое применение / Д. Скляр; пер. с англ. – М. :Издательский дом «Вильямс», 2003. – 1104 с.
2. Морелос-Сарагоса, Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение / Р. Морелос-Сарагоса; пер. с англ. – М. : Техносфера, 2006. – 319 с.