

***«МАТЕМАТИЧЕСКИЕ МОДЕЛИ
ДЛЯ ИСТОЧНИКОВ
ИНФОРМАЦИИ»***

Простейший тип дискретного источника – это такой, который выдаёт последовательность букв (символов), выбираемых из определенного алфавита.

Например, **двоичный источник** выдает двоичную последовательность вида 100101110..., причём алфавит состоит из двух символов $\{0,1\}$. В более общем случае источник дискретен информации с алфавитом из L символов, $\{x_1, x_2, \dots, x_L\}$ выдает последовательность букв, выбираемых из этого алфавита.

Чтобы конструировать математическую модель для дискретного источника предположим, что каждый символ алфавита $\{x_1, x_2, \dots, x_L\}$ имеет заданную вероятность выбора P_k т.е

$$p_k = P(X = x_k), \quad 1 \leq k \leq L$$

где:

$$\sum_{k=1}^L p_k = 1$$

Рассмотрим две математические модели для дискретных источников.

Предположим, что символы выходной последовательности источника статистически независимы, т. е. выбираемый текущий символ статистически независим от всех предыдущих и последующих. Источник, выход которого удовлетворяет условиям статистической независимости символов в выбранной последовательности, называется источником без памяти. Такой источник называется **дискретным источником без памяти (ДИБП)**.

Если отдельные выходные символы дискретного источника статистически взаимозависимы, как, например, в английском тексте, мы можем сконструировать математическую модель, основанную на статической стационарности. По определению дискретный источник называется **стационарным**, если совместные вероятности двух последовательностей длины n , допустим $\{a_1, a_2, \dots, a_n\}$ и $\{a_{1+m}, a_{2+m}, \dots, a_{n+m}\}$ одинаковые для всех $n \geq 1$ и при всех сдвигах m . Другими словами, совместные вероятности для последовательностей источника произвольной длины инвариантны по отношению к произвольному сдвигу во времени.

Аналоговый источник выдает сигнал $x(t)$, который является реализацией случайного процесса $X(t)$.

Предположим, что $X(t)$ - стационарный случайный процесс с автокорреляционной функцией $\phi_{xx}(\tau)$ и спектральной плотностью мощности $\Phi_{xx}(f)$.

Если $X(t)$ - частотно-ограниченный случайный процесс, т.е. $\Phi_{xx}(f) = 0$ для $|f| \geq W$, можно использовать теорему отсчетов для представления $X(t)$ в виде:

$$X(t) = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2W}\right) \frac{\sin\left[2\pi W\left(t - \frac{n}{2W}\right)\right]}{2\pi W\left(t - \frac{n}{2W}\right)}$$

Где $X\{(n/2W)\}$ - отсчёты процесса $X(t)$ взятые со скоростью Найквиста $f_s=2W$, 1/с.

Используя теорему отсчётов, мы можем преобразовать аналоговый источник в эквивалентный источник с дискретным временем. После этого выход источника характеризуется совместной ФПВ $p(x_1, x_2, \dots, x_m)$ для всех $m \geq 1$, где $X_n\{(n/2W)\}$, $1 \leq n \leq m$ - случайные величины, соответствующие отсчётам $X(t)$.

Заметим, что выходные отсчёты стационарного источника обычно непрерывны, и, следовательно, их нельзя представить в цифровой форме без потери точности представления. Например, мы можем квантовать каждый отсчёт рядом дискретных значений, но процесс квантования вносит потери в точность представления, и, следовательно, исходный сигнал не может быть восстановлен точно по квантованным отсчётам.

«ЛОГАРИФМИЧЕСКАЯ МЕРА ИНФОРМАЦИИ»

Чтобы разработать подходящую меру для информации, рассмотрим две дискретные случайные величины X и Y с возможными значениями соответственно $x_i, i=1,2,\dots,n$ и $j=1,2,\dots,m$

Допустим, мы наблюдаем некоторый выход $Y=y_j$ и мы желаем количественно определить величину информации, которую обеспечивает выборка события $Y=y_j$ относительно события $X=x_i, i=1,2,\dots,n$

Заметим, что если X и Y статистически не зависят друг от друга, выбор $Y=y_j$ не даёт информации о выборе события $X=x_i$.

С другой стороны, если X и Y полностью зависимы, так что выбор $Y=y_j$ однозначно определяет выбор $X=x_i$ информационное содержание этого выбора точно такое же, как при выборе события $X=x_i$.

Подходящая мера информации, которая удовлетворяет указанным условиям, - это логарифм отношения условной вероятности:

$$P(X=x_i | Y=y_j) = P(x_i | y_j)$$

к вероятности:

$$P(X = x_i) \equiv P(x_i)$$

Это значит, что количество информации, полученное при появлении события $Y=y_j$ относительно события $X=x_i$ определяется как:

$$I(x_i ; y_j) = \log \frac{P(x_i | y_j)}{P(x_i)} \quad (1)$$

названа **взаимной информацией** между x_i и y_j

Единица измерения $I(x_i; y_j)$ определяется основанием логарифма, в качестве которой обычно выбирается или 2, или e . Когда основание логарифма равно 2, единицей измерения $I(x_i; y_j)$ является **бит**, а когда основание равно e , единицей измерения $I(x_i; y_j)$ является **нат** (натуральная единица). (Стандартная аббревиатура для \log_e - это **ln** .)

Когда случайные величины X и Y статистически независимы, то $P(x_i|y_j)=P(x_i)$, следовательно, $I(x_i; y_j)=0$.

С другой стороны, когда выбор события $Y=y_j$ полностью определён выбором события $X=x_i$ условная вероятность в числителе (1) равна единице и, следовательно:

$$I(x_i ; y_j) = \log \frac{1}{P(x_i)} = -\log P(x_i) = I(x_i ; x_i) \quad (2)$$

Но (2) как раз определяет информацию $X=x_i$

Исходя из этих соображений, её называют **собственной информацией события $X=x_i$** . Она обозначается :

$$I(x_i) = \log \frac{1}{P(x_i)} = -\log P(x_i) \quad (3)$$

Пример 1

Предположим, что имеется дискретный источник, который выдаёт двоичную цифру 0 или 1 с равной вероятностью каждые τ_x секунд. Количество информации при каждом появлении новой цифры:

$$I(x_i) = -\log_2 P(x_i) = -\log_2 \frac{1}{2} = 1 \text{ (бит)}, x_i = 0,1$$

Теперь предположим, что последовательные цифры на выходе источника статистически независимы, т.е. источник не имеет памяти. Рассмотрим блок символов источника из k двоичных цифр, который существует на интервале kT_x . Имеется $M=2^k$ таких возможных k -битовых блоков, каждый с равной вероятностью $1/M=2^{-k}$. Собственная информация k -битового блока равна:

$$I(x'_i) = -\log_2 2^{-k} = k \text{ бит}$$

Она выдаётся на временном интервале $K\tau x$.

Таким образом, логарифмическая мера количества информации обладает желаемыми свойствами аддитивности, когда определённое число единичных выходов источника рассматривается как один блок.

вернёмся к определению взаимной информации, определяемой (1), и умножим числитель и знаменатель отношения вероятностей на $p(y_j)$:

$$\frac{P(x_i|y_j)}{P(x_i)} = \frac{P(x_i|y_j)P(y_j)}{P(x_i)P(y_j)} = \frac{P(x_i, y_j)}{P(x_i)P(y_j)} = \frac{P(y_j|x_i)}{P(y_j)}$$

Отсюда делаем вывод:

$$I(x_i ; y_j) = I(y_j; x_i)$$

Таким образом, информация, содержащаяся в выборе события $Y=y_j$ относительно события $X=x_i$ идентична информации, содержащейся в выборе события $X=x_i$ относительно события $Y=y_j$.

Пример 2

Предположим, что X и Y - двоичные $\{0,1\}$ случайные величины, представляющие вход и выход канала с двоичным входом и двоичным выходом. Входные символы равновероятны, а условные вероятности выходных символов при заданном входе определяются так:

$$P(Y=0|X=0)=1-p_0$$

$$P(Y=1|X=0)=p_0$$

$$P(Y=1|X=1)=1-p_1$$

$$P(Y=0|X=1)=p_1$$

Определим, сколько информации об $X=0$ и $X=1$ содержится в событии $Y=0$. Из заданных вероятностей получим:

$$P(Y=0) = P(Y=0|X=0)P(X=0) + P(Y=0|X=1)P(X=1) = \frac{1}{2}(1-p_0+p_1)$$

$$P(Y=1) = P(Y=1|X=0)P(X=0) + P(Y=1|X=1)P(X=1) = \frac{1}{2}(1-p_1+p_0)$$

Тогда взаимная информация о символе $X=1$ при условии, что наблюдается $Y=0$, равна:

$$I(x_1; y_1) = I(0; 0) = \log_2 \frac{P(Y=0|X=0)}{P(Y=0)} = \log_2 \frac{2(1-p_0)}{1-p_0+p_1}$$

Аналогично взаимная информация о символе $X=1$ при условии, что наблюдается $Y=0$, равна

$$I(x_2; y_1) = I(1; 0) = \log_2 \frac{2p_1}{1-p_0+p_1}$$

Рассмотрим несколько частных случаев. В первом, когда, $p_0=p_1=0$ канал называют каналом **без шумов** и $I(0;0)=\log_2 2=1$ бит.

Следовательно, когда выход точно определяет вход, нет потери информации. С другой стороны, если $p_0=p_1=1/2$, канал становится непригодным так как $I(0;0)=\log_2 1=0$

Если $p_0=p_1=1/4$, то

$$I(0;0) = \log_2 \frac{3}{2} = 0,587$$

$$I(0;1) = \log_2 \frac{1}{2} = -1$$

Помимо определения взаимной информации и собственной информации полезно определить **условную собственную информацию** как:

$$I(x_i | y_j) = \log \frac{1}{P(x_i | y_j)} = -\log P(x_i | y_j) \quad (5)$$

Тогда, комбинируя (1), (3) и (5), получаем соотношение:

$$I(x_i; y_j) = I(x_i) - I(x_i | y_j) \quad (6)$$

Мы интерпретируем $I(x_i|y_j)$ как собственную информацию о событии $X=x_i$ после наблюдения события $Y=y_j$. Из условия $I(x_i) \geq 0$ и $I(x_i|y_j) \geq 0$ следует, что $I(x_i, y_j) < 0$ когда $I(x_i|y_j) > I(x_i)$, и $I(x_i, y_j) > 0$, когда $I(x_i|y_j) < I(x_i)$. Следовательно, взаимная информация между парой событий может быть или положительной, или отрицательно или равной нулю.

«Средняя взаимная информация и энтропия»

Зная взаимную информацию, связанную с парой событий $(x_i; y_j)$, которые являются возможной реализацией двух случайных величин X и Y , мы можем получить среднее значение взаимной информации простым взвешиванием $I(x_i; y_j)$, с вероятностью появления этой пары и суммированием по всем возможным событиям.

Таким образом получим:

$$I(X;Y) = \sum_{i=1}^n \sum_{j=1}^m P(x_i; y_j) I(x_i; y_j) = \sum_{i=1}^n \sum_{j=1}^m P(x_i; y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} = I(X;Y) \quad (1)$$

как среднюю взаимную информацию между X и Y

Видно, что $I(X; Y) = 0$, когда X и Y статистически независимы и $P(x_i; y_j) = P(x_i)P(y_j)$

Важным свойством средней взаимной информации является то, что $I(X; Y) \geq 0$

Аналогично определим среднюю собственную информацию, обозначенную $H(X)$

$$H(X) = \sum_{i=1}^n P(x_i) I(x_i) = - \sum_{i=1}^n P(x_i) \log P(x_i) \quad (2)$$

Если X представляет собой алфавит возможных символов источника, $H(X)$ представляет среднюю собственную информацию на символ источника, и её называют **энтропией источника**. В частном случае, когда символы источника равновероятны, $P(x_i) = 1/n$ для всех i , и, следовательно

$$H(X) = - \sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} = \log n \quad (3)$$

В общем случае $H(X) \leq \log n$ при любых заданных вероятностях символов источника. Другими словами, энтропия источника максимальна, когда выходные символы равновероятны.

Пример 1

Рассмотрим двоичный источник, который выдаёт последовательности независимых символов, причём выходной кодовый символ «0» с вероятностью q , а символ «1» с вероятностью $1-q$. Энтропия такого источника

$$H(X) = H(q) = -q \log q - (1-q) \log (1-q) \quad (4)$$

Функцию $H(q)$ иллюстрирует рисунок 1.

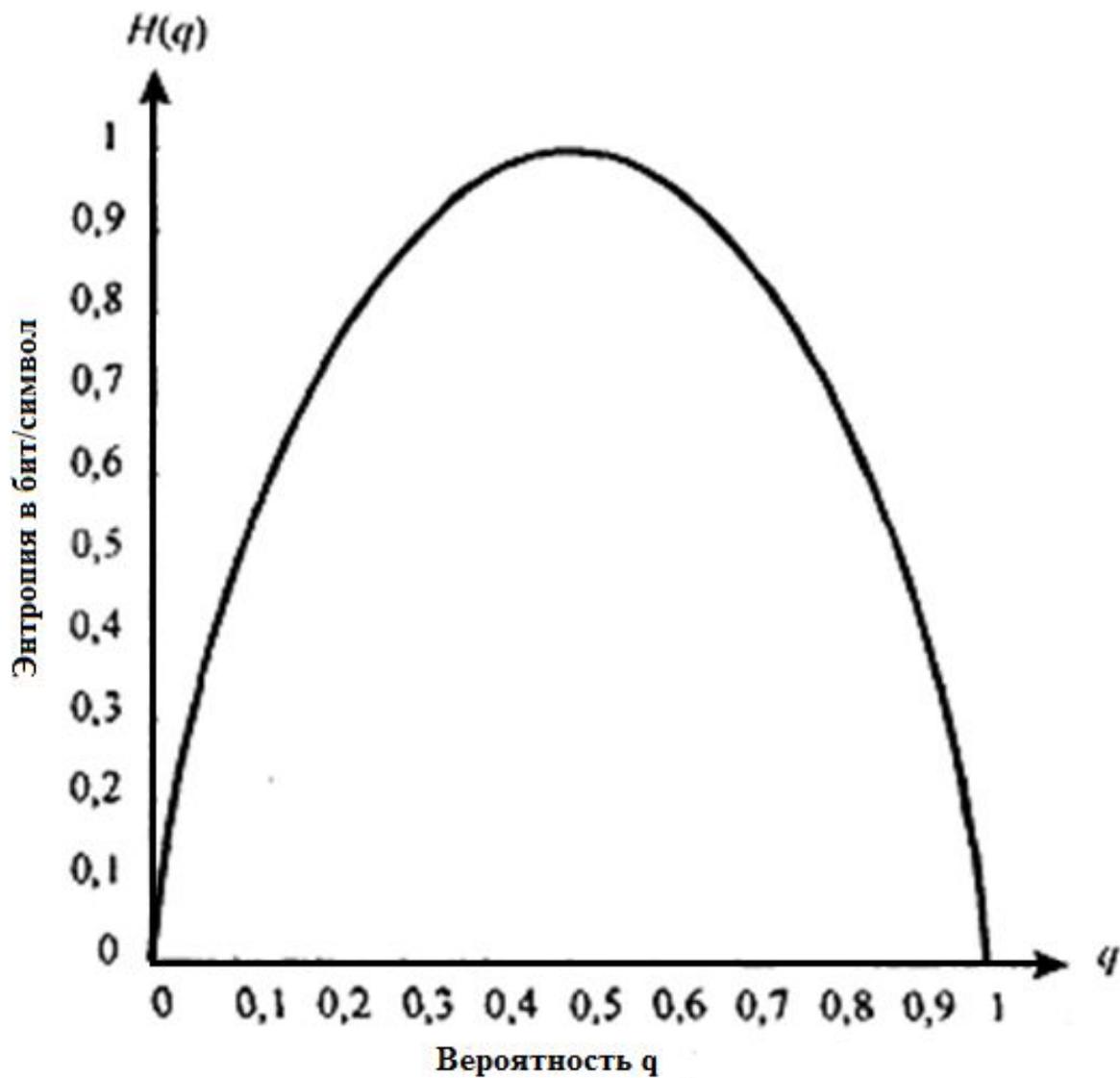


Рисунок 1. Энтропия двоичного источника

Среднее значение условной собственной информации называется условной энтропией и определяется как:

$$H(X|Y) = \sum_{i=1}^n \sum_{j=1}^m P(x_i, y_j) \log \frac{1}{P(x_i|y_j)} \quad (5)$$

Мы интерпретируем $H(X|Y)$ как неопределённость X (дополнительную информацию, содержащуюся в Y) после наблюдения Y . Комбинация (1), (2) и (4) даёт соотношение:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (6)$$

Из условия $I(X, Y) \geq 0$ следует, что $H(X) \geq H(X|Y)$ и $H(Y) \geq H(Y|X)$, причём равенство имеет место тогда, и только тогда, когда X и Y статистически независимы.

Если мы интерпретируем $H(X|Y)$ как среднее значение неопределённости (условной собственной информации) X после наблюдения Y и $H(X)$ как среднее значение априорной неопределённости (собственной информации)

т.е. имевшейся до наблюдения, тогда $I(X;Y)$ определяет взаимную информацию (уменьшение среднего значения неопределённости, имеющейся относительно X после наблюдения Y).

Так как $H(X) \geq H(X|Y)$ то ясно, что при условии наблюдения Y энтропия $H(X)$ не увеличится.

Пример 2

Определим $H(X|Y)$ и $I(X, Y)$ для канала с двоичным входом и выходом, рассмотренного выше в примере 1, для случая, когда $p_0=p_1=p$. Пусть вероятность входных символов равна $P(X=0)=q$ и $P(X=1)=1-q$. Тогда:

$$H(X)=H(q)=-q \lg q - (1-q) \lg(1-q)$$

Где $H(q)$ - функция энтропии. Зависимость $H(X|Y)$ в бит/символ как функция от q и p параметра показана на рис. 3. График средней взаимной информации $I(X, Y)$ в бит/символ дан на рис. 3.1.

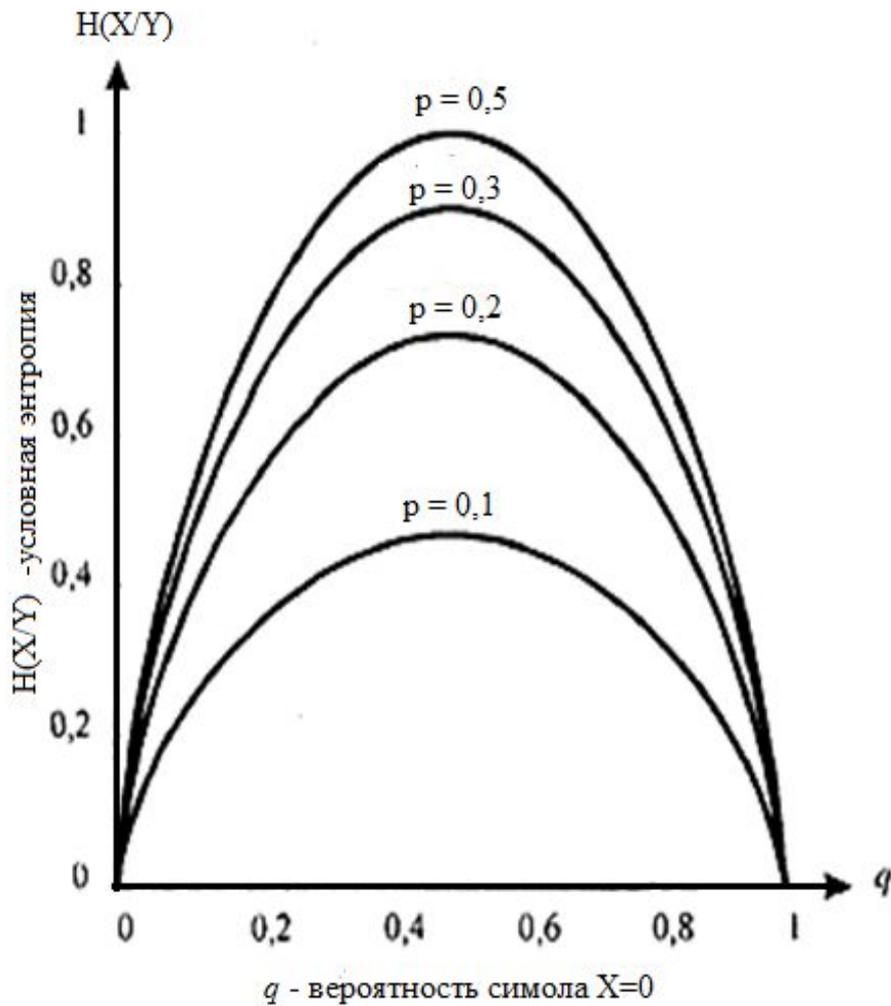


Рис. 3. Условная энтропия для двоичного симметричного канала

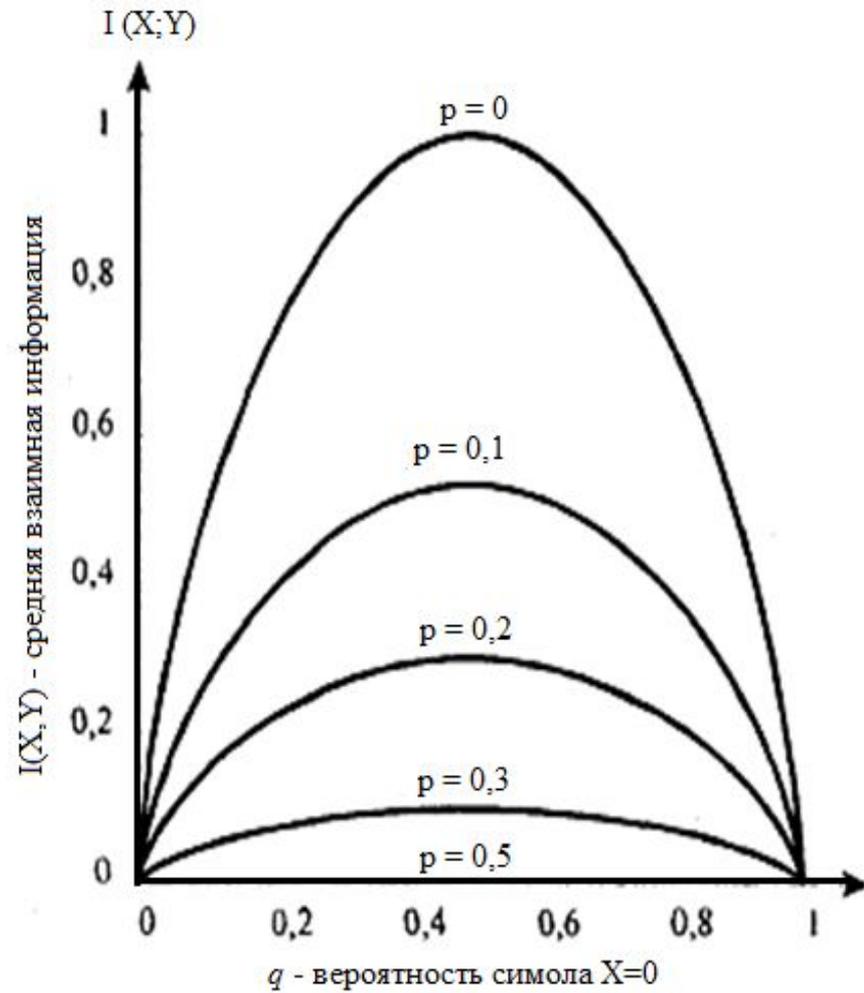


Рис.3.1. Средняя взаимная информация для двоичного симметричного канала

Когда условную энтропию $H(X|Y)$ рассматривают применительно к каналу с входом и выходом Y , то называют ненадёжностью канала на символ и её интерпретируй как величину средней неопределённости, оставшейся в X после наблюдения Y .

Результаты, приведённые выше, легко обобщаются на случай произвольного числа случайных величин. В частности, предположим, что мы имеем блок из k случайных величин $X_1 X_2 X_3 \dots X_k$ с совместной вероятностью $P(X_1 X_2 \dots X_k) \equiv P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$

Тогда энтропия определяется как:

$$H(X_1 X_2 \dots X_k) = - \sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} \dots \sum_{j_k=1}^{n_k} P(x_{j_1} x_{j_2} \dots x_{j_k}) \log P(x_{j_1} x_{j_2} \dots x_{j_k})$$

Поскольку совместную вероятность $P(X_1 X_2 \dots X_k)$ можно выразить в виде:

$$P(x_1 x_2 \dots x_k) \equiv P(x_1) P(x_2 | x_1) P(x_3 | x_1 x_2) \dots P(x_k | x_1 x_2 \dots x_{k-1})$$

то следует

$$\begin{aligned} H(X_1 X_2 X_3 \dots X_k) &= H(X_1) + H(X_2 | X_1) + H(X_3 | X_1 X_2) + \dots + H(X_k | X_1 \dots X_{k-1}) = \\ &= \sum_{i=1}^k H(X_i | X_1 \dots X_{i-1}) \end{aligned} \quad (7)$$

С учётом результата $H(X) \geq H(X|Y)$ где $X = X_m$ и $Y = X_1 X_2 \dots X_{m-1}$ из (7) следует

$$H(X_1 X_2 \dots X_k) \leq \sum_{m=1}^k H(X_m)$$

причём равенство имеет место тогда, и только тогда, когда случайные величины $X_1 X_2 X_3 \dots X_k$ статистически независимы.

*«Кодирование для дискретных
источников без памяти»*

Предположим, что ДИБП выдает буквы или символы каждые τ_x секунд. Каждый символ выбирается из конечного алфавита x_i , $i=1,2,\dots,L$, с вероятностью $P(x_i), i=1,2,\dots,L$. Энтропия ДИБП в битах на символ:

$$H(X) = - \sum_{i=1}^L P(x_i) \log_2 (x_i) \leq \log_2 L \quad (1)$$

причем равенство имеет место, если все символы равновероятны. $H(X)$ определяет среднее число бит на символ источника, а производительность источника в битах определяется как $H(X)/\tau_x$

Кодовые слова фиксированной длины

Сначала рассмотрим схему блокового кодирования, которая сопоставляет уникальный ряд из R двоичных символов с каждым символом источника. Поскольку имеется L возможных символов источника, то числа двоичных символов кодера на один символ источника при уникальном кодировании

$$R = \log_2 L \quad (2)$$

когда L равно целой степени основания 2, и

$$R = [\log_2 L] + 1 \quad (3)$$

когда не равно целой степени основания 2. Здесь $[x]$ означает наибольшее целое, меньшее, чем x .

R будем называть скоростью кодирования. Она определяет число символов кодера на один символ источника. Поскольку,

$$H(X) \leq \log_2 L, \text{ то } R \geq H(X)$$

Эффективность кодирования для ДИБП определяется отношением $H(X) / R$. Видим, что если L равно степени числа 2 и символы источника равновероятны, то $R=H(X)$. Следовательно, код фиксированной длины с R двоичными символами на символ источника в данном случае обеспечивает стопроцентную эффективность. Однако, если L не равно степени 2, но символы источника всё ещё равновероятны, R отличается от $H(X)$ самое большее на один бит на символ.

Если $\log L \gg 1$, эффективность такой схемы кодирования высока. С другой стороны, если L мало, эффективность кода с фиксированной длиной можно увеличить путем кодирования последовательности из J символов источника за время $J\tau_s$. Чтобы выполнить такое кодирование, мы должны L^J выбрать уникальных кодовых слов. Используя кодовую последовательность из N двоичных символов, мы можем образовать 2^N возможных кодовых слов. Число N должно быть выбрано так, чтобы:

$$N \geq J \log_2 L$$

Следовательно, требуется минимальное целое значение для N , равное

$$N = \lceil J \log_2 L \rceil + 1 \quad (4)$$

Теперь среднее число символов кода на символ источника $R=N/J$, и, таким образом, неэффективность кодирования сокращается примерно в J раз по сравнению с посимвольным кодированием, описанным выше.

Взяв J достаточно большим, можно делать эффективность процедуры кодирования, измеренную отношением $JH(X)/N$, как угодно близкой к единице.

Методы кодирования, не приводят к искажениям, так как кодирование символов источника или блоков таких символов в кодовые слова выполняется однозначно (уникально). Такие типы кодов названы бесшумными.

Предположим, что мы пытаемся уменьшить скорость кодирования R путем смягчения условия однозначности процесса кодирования. Предположим, что только доля L^J блоков символов источника кодируется однозначно. Выберем наиболее вероятных J -символьных блоков и будем кодировать каждый из них однозначно, в то время как оставшиеся $L^J - J$ блоков длины J представим одним оставшимся кодовым словом. Эта процедура кодирования вызовет ошибку декодирования каждый раз, когда источник выдаст такой маловероятный блок.

Теорема кодирования источника I

Пусть X - это ансамбль символов ДИБП с конечной энтропией $H(X)$. Блоки из J символов источника кодируются в двоичные кодовые слова длиной N . Для любого $\varepsilon > 0$ вероятность P_e ошибки декодирования можно сделать сколь угодно малой, если

$$R = \frac{N}{J} \geq H(X) + \varepsilon$$

J достаточно велико.

Наоборот, если $R \leq H(X) - \varepsilon$ тогда P_e сколь угодно близка к 1 при достаточно больших J .

Исходя из этой теоремы мы видим, что среднее число бит на символ источника требуемое для кодирования выхода ДИБП с произвольно малой вероятностью ошибки декодирования, ограничено снизу энтропией источника $H(X)$. С другой стороны, если $R < H(X)$, вероятность ошибки декодирования приближается к 100%, если J произвольно увеличивать.

*Кодовые слова переменной
длины*

Если символы источника неравновероятны, более эффективный метод кодирования сводится к использованию кодовых слов переменной длины. Примером такого кодирования является код Морзе. В коде Морзе символам, возникающим более часто, сопоставляются более короткие кодовые слова, а символам, возникающим менее часто, сопоставляются более длинные кодовые слова.

Таблица 1. Коды переменной длины

Символ	$P(a_k)$	Код I	Код II	Код III
a_1	1/2	1	0	0
a_2	1/4	00	10	01
a_3	1/8	01	110	011
a_4	1/8	10	111	111

Для примера предположим, что выходные символы ДИБП $a_1 a_2 a_3 a_4$ соответствующими вероятностями $P(a_1)=1/2$, $P(a_2)=1/4$, $P(a_3)=P(a_4)=1/8$, кодируются так как показано в табл.1. Код I имеет переменную длину и имеет принципиальный недостаток. Чтобы увидеть этот недостаток, предположим, что мы приняли последовательность 001001... Ясно, что 00 декодируется как a_2 . Однако последующие четыре бита декодируются неоднозначно. Они могут декодироваться или как $a_4 a_3$, или как $a_1 a_2 a_1$.

Код II в табл. 1 обеспечивает однозначное и немедленное декодирование. Удобно представлять кодовые слова этого кода графически как узлы на дереве, как показано на рис. 1. Видно, что 0 указывает на окончание кодового слова в первых трех кодовых словах. Эта характеристика вместе с тем обстоятельством, что ни одно кодовое слово не содержит более трех двоичных символов, что делает этот код немедленно декодируемым.

Заметим, что ни одно кодовое слово этого кода не является префиксом (началом) другого кодового слова. В общем, **префиксное условие** кода требует, чтобы для данного кодового слова C_k длины k с элементами (b_1, b_2, \dots, b_k) не существовало других кодовых слов длины $l < k$ с элементами (b_1, b_2, \dots, b_l) для $1 \leq l \leq k-1$

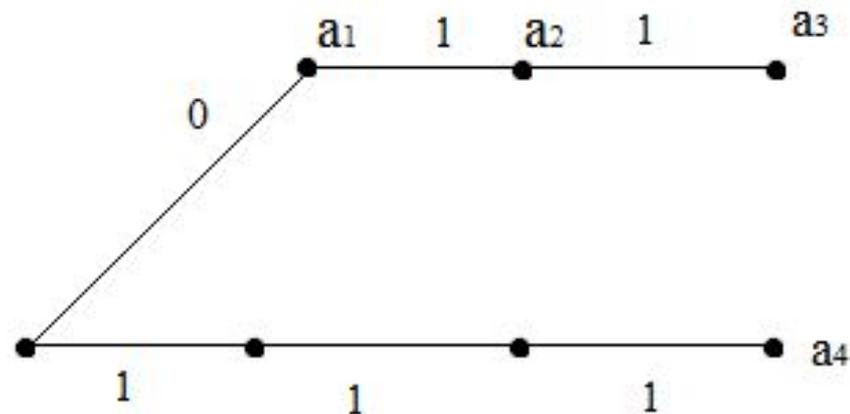
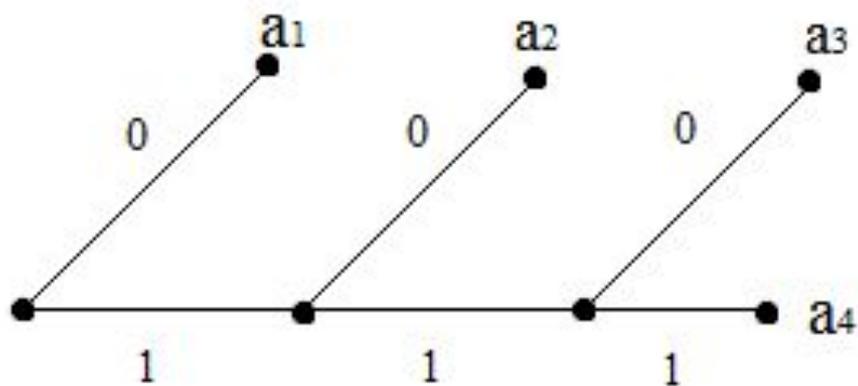


Рис. 1. Кодовое дерево для кода II в табл.1

Рис. 2. Кодовое дерево для кода III в табл.1

Другими словами, нет кодовых слов длины $l < k$, которые совпадают с первыми l двоичными символами другого кодового слова длины $k > l$. Это свойство делает кодовые слова немедленно декодируемыми.

Код III из табл. 1 имеет кодовое дерево, показанное на рис. 2. Видим, что в этом случае имеет место однозначное декодирование, однако требующее задержки. Ясно, что этот код не удовлетворяет префиксному условию.

Наша главная цель – создать систематическую процедуру для конструирования однозначных декодирующих кодов переменной длины, эффективных в том смысле, что среднее число бит на один символ источника, определяемое соотношением, было бы минимальным. Условие существования кода переменной длины, которое удовлетворяет префиксному условию, дается неравенством Крафта.

$$\bar{R} = \sum_{k=1}^L n_k P(a_k) \quad (1)$$

Неравенство Крафта

Необходимым и достаточным условием существования двоичного кода с кодовыми символами длины $n_1 \leq n_2 \leq \dots \leq n_L$, удовлетворяющего условию префиксности, является

$$\sum_{k=1}^L 2^{-n^k} \leq 1 \quad (2)$$

Сначала мы докажем, что (2) является достаточным условием для существования префиксного кода. Чтобы построить такой код, мы начнем с 2^n полного двоичного дерева порядка $n = n_L$, которое имеет конечных узлов, причем от каждого узла порядка $k-1$ «растут» по два узла порядка k , $1 \leq k \leq n$.

Выберем некоторый узел порядка n_1 в качестве первого кодового слова C_1 . Этот выбор устраняет 2^{n-n_1} конечных узлов. От остающихся доступных узлов порядка n_2 мы выбираем один узел для второго кодового слова C_2 . Этот выбор устраняет 2^{n-n_2} конечных узлов. Этот процесс продолжается, пока последнее кодовое слово не определено в конечном узле $n = n_L$. Следовательно, в узле порядка $j < L$ доля числа отсечённых конечных узлов:

$$\sum_{k=1}^J 2^{-n^k} < \sum_{k=1}^L 2^{-n^k} \leq 1 \quad (3)$$

Таким образом, создали кодовое дерево, которое встроено в полное дерево из 2^n узлов, как иллюстрируется на рис. 3. для дерева, имеющего 16 конечных узлов, и источника, состоящего из пяти символов, отображаемых кодовыми словами длиной $n_1=1, n_2=2, n_3=3, n_4=4$.

Чтобы доказать, что (2) является необходимым условием, мы заметим, что в дереве порядка $n = n_L$, число конечных узлов, отсечённых от общего числа 2^n конечных узлов равно:

$$\sum_{k=1}^L 2^{n-n^k} \leq 2^n$$

Следовательно,

$$\sum_{k=1}^L 2^{-n^k} \leq 1$$

Неравенство Крафта можно использовать для доказательства следующей теоремы кодирования источника (без шумов), которое применяется к кодам, удовлетворяющим префиксному условию.

Теорема кодирования источника II.

Пусть X - ансамбль символов двоичного источника без памяти с конечной энтропией $H(X)$ и выходными символами x_k , $1 \leq k \leq L$ с соответствующими вероятностями выбора p_k , $1 \leq k \leq L$. Существует возможность создать код, который удовлетворяет префиксному условию и имеет среднюю длину \bar{R} , которое удовлетворяет неравенству:

$$\bar{R} \geq H(X) \quad (4)$$

Чтобы установить нижнюю границу в (4), обратим внимание на то, что для кодовых слов, которые имеют длину n_k , $1 \leq k \leq L$, разность $H(X) - \bar{R}$ может быть выражена в виде

$$\sum_{k=1}^L p_k (n_k - \bar{R}) \quad (5)$$

Используя неравенство $\ln x \leq x-1$, из (5) находим:

Верхняя граница в (4) может быть установлена при предположении что n_k , $1 \leq k \leq L$ - целые числа, выбираемые из условия $2^{-n_k} \leq p_k \leq 2^{-n_k+1}$. Но если $p_k \geq 2^{-n_k+1}$ просуммированы по $1 \leq k \leq L$, получаем неравенство Крафта, для которого демонстрировали, что там существует код, удовлетворяющий префиксному условию другой стороны, если мы берем логарифм, получаем: $\log p_k < -n_k + 1$

или, что эквивалентно: $n_k < 1 - \log p_k$ (6)

Если умножить обе части неравенства (6) на p_k и просуммировать по $1 \leq k \leq L$, получаем желательную верхнюю границу, данную в (6). Это завершает доказательство (6).

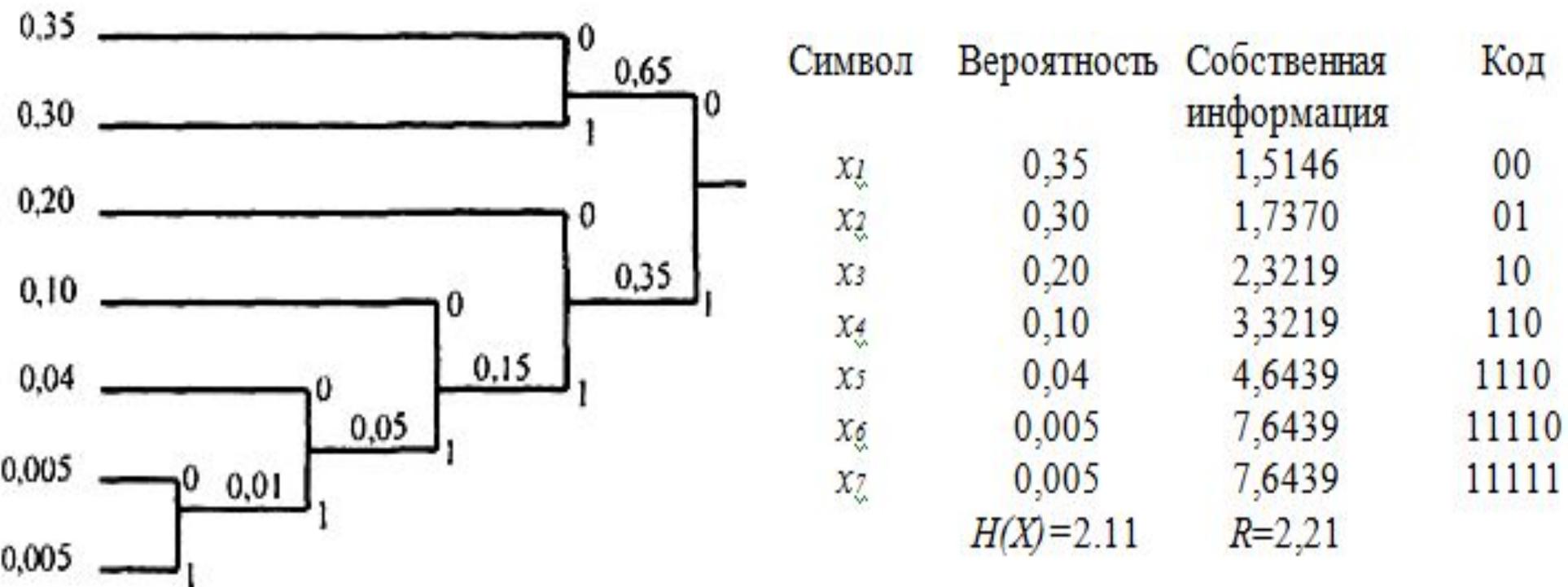
Мы установили, что коды переменной длины, которые удовлетворяют префиксному условию, - это эффективные коды для любого дискретного источника без памяти (ДИБП) с символами, имеющими различную априорную вероятность. Опишем теперь алгоритм для построения таких кодов.

*Алгоритм кодирования
Хаффмена*

Хаффмен (1952) разработал алгоритм кодирования переменной длины, основанный на знании априорных вероятностей символов $P(x_i)$, $i=1,2,\dots,L$. Этот алгоритм оптимален в том смысле, что среднее число двоичных символов, требуемых для представления исходных символов, минимально. Получаемые кодовые слова удовлетворяют префиксному условию, что позволяет уникально и мгновенно декодировать полученную последовательность.

Пример 1.

Рассмотрим ДИБГТ с семью возможными символами, имеющими вероятности выбора, иллюстрируемое ниже.



Среднее число двоичных элементов на символ этого кода 2,21 бит/символ. Энтропия источника – 2,11 бит/символ.

Заметим, что полученный код не единственно возможный. Например, на предпоследнем шаге процедуры кодирования мы имеем равный выбор между x_1 , и x_2 имеющими одинаковые вероятности. В этом пункте мы соединили x_1 и x_2 . В альтернативном коде мы можем соединить x_2 и x_3' . Результирующий код для этого случая иллюстрируется на рис.2

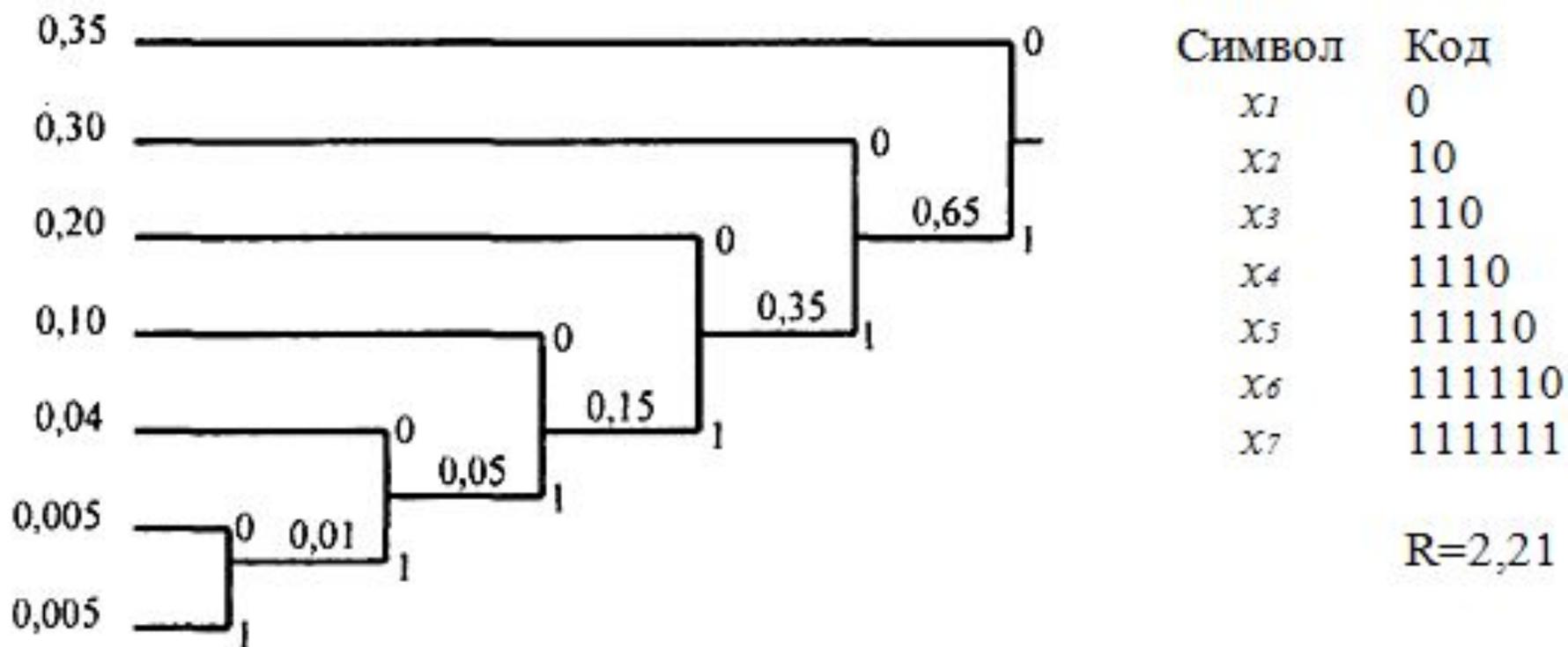


Рис. 2. Альтернативный код для ДИБП в примере 1

Пример 2.

В качестве второго примера определим код Хаффмена для выхода ДИБП, иллюстрируемый на рис. 3. Энтропия этого источника $H(X)=2,63$ бит/символ. Код Хаффмена, показанный на рис. 3, имеет среднюю длину $R=2,70$ бит/символ. Следовательно, его эффективность составляет 0,97.

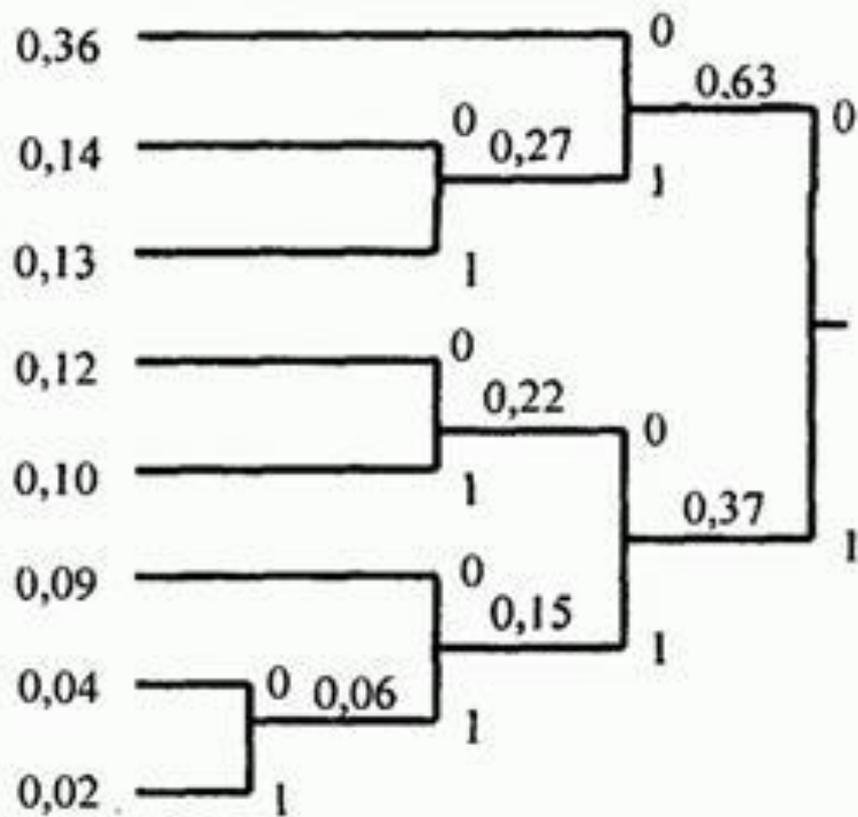
Алгоритм кодирования переменной длины (Хаффмена), генерирует префиксный код, имеющий среднюю длину. Однако вместо посимвольного кодирования более эффективной является процедура, основанная на кодировании блоков из символов одновременно, таком случае границы в $H(X) \leq \bar{R} < H(X)+1$ в теореме кодирования источника Π становятся другими:

$$JH(X) \leq \bar{R}_J < JH(X) + 1 \quad (1)$$

так как энтропия J -символьного блока от ДИБП равна $JH(X)$, и \bar{R}_J - среднее число бит в J -символьном блоке. Если мы разделим (1) на J , то получим:

$$H(X) \leq \frac{\bar{R}_J}{J} < H(X) + \frac{1}{J} \quad (2)$$

Где $\bar{R}_J / J \equiv \bar{R}$ - среднее число битов на исходный символ. Следовательно, \bar{R} можно сделать как угодно близким к $H(X)$, выбирая J достаточно большим.



Символ	Код
x_1	00
x_2	010
x_3	011
x_4	100
x_5	101
x_6	110
x_7	1110
x_8	1111

$H(X) = 2,63$ $\bar{R} = 2,70$

Рис. 3. Код Хаффмена для примера 2

Пример 3

Выход ДИБП состоит из символов x_1 , x_2 и x_3 с вероятностями $0,45$, $0,35$ и $0,20$ соответственно. Энтропия этого источника $H(X)=1,518$ бит/символ. Код Хаффмена для этого источника, данный в табл.1, требует $R_1=1,55$ бит/символ и приводят к эффективности $97,9\%$. Если посредством алгоритма Хаффмена символы закодированы парами, результирующий код выглядит так, как показано в табл.1. Энтропия источника для пар символов $2H(X)=3,036$ бит/пара символов. С другой стороны, код Хаффмена требует $R_2=3,0675$ бит/пара символов. Таким образом, эффективность кодирования увеличилась до $2H(X)/R_2=0,990$ (до $99,0\%$)

Таблица 1. Код Хаффмена

Символ	Вероятность	Собственная информация	Код
x_1	0,45	1,156	1
x_2	0,35	1,520	00
x_3	0,20	2,33	01
		$H(X)=1,518$ бит/символ	
		$R_1=1,55$ бит/символ	
		Эффективность 97,9 %	

Таблица 2. Код Хаффмена для кодирования пар символов

Пара символов	Вероятность	Собственная информация	Код
$x_1 x_1$	0,2025	2,312	10
$x_1 x_2$	0,1575	2,676	001
$x_2 x_1$	0,1575	2,676	010
$x_2 x_2$	0,1225	3,039	011
$x_1 x_3$	0,09	3,486	111
$x_3 x_1$	0,09	3,486	0000
$x_2 x_3$	0,07	3,850	0001
$x_3 x_2$	0,07	3,850	1100
$x_3 x_3$	0,04	4,660	1101

$2H(X)=3,036$ бит/пара символов; $\bar{R}_2 = 3,0675$ бит/пара символов
 $\frac{1}{2} \bar{R}_2 = 1,534$ бит/символ; Эффективность 99,0 %

Дискретные стационарные источники

Рассмотрим дискретные источники, для которых последовательность символов выхода является статистически зависимой. Мы ограничим наше исследование источниками, которые являются статистически стационарными (однородными во времени)

Оценим энтропию некоторой последовательности символов от стационарного источника. Энтропия блока случайных переменных $X_1, X_2 \dots X_k$ равна

$$H(X_1 X_2 \dots X_k) = \sum_{i=1}^k H(X_i | X_1 X_2 \dots X_{i-1}) \quad (1)$$

Где $H(X_i | X_1 X_2 \dots X_{i-1})$ - условная энтропия i -го символа при условии, что источник выдал предыдущие $i-1$ символов. Энтропия на символ для k -символьного блока определяется как:

$$H_k(X) = \frac{1}{k} H(X_1 X_2 \dots X_k) \quad (2)$$

Мы определяем количество информации стационарного источника как энтропию на символ в (2) в пределе при $k \rightarrow \infty$ т.е.

$$H_{\infty}(X) = \lim_{k \rightarrow \infty} H_k(X) = \lim_{k \rightarrow \infty} \frac{1}{k} H(X_1 X_2 \dots X_k) \quad (3)$$

В качестве альтернативы мы можем определять энтропию на символ источника как условную энтропию $H(X_k | X_1 X_2 \dots X_{k-1})$ в пределе при $k \rightarrow \infty$. Этот предел также существует и идентичен пределу в (3). То есть

$$H_{\infty}(X) = \lim_{k \rightarrow \infty} H(X_k | X_1 X_2 \dots X_{k-1}) \quad (4)$$

Этот результат также установлен ниже. Наше изложение использует подход Галлагера (1968). Во-первых, мы покажем, что:

$$H(X_k|X_1X_2\dots X_{k-1}) \leq H(X_{k-1}|X_1X_2\dots X_{k-2}) \quad (5)$$

для $k \geq 2$. С учётом предыдущего результата, согласно которому наложение условий на случайную переменную не может увеличивать её энтропию, мы имеем:

$$H(X_k|X_1X_2\dots X_{k-1}) \leq H(X_k|X_2X_3\dots X_{k-1}) \quad (6)$$

В силу стационарности источника имеем:

$$H(X_k|X_2X_3\dots X_{k-1}) = H(X_{k-1}|X_1X_2\dots X_{k-2}) \quad (7)$$

Во-вторых, мы имеем результат:

$$H_k(X) \geq H(X_k | X_1 X_2 \dots X_{k-1}) \quad (6)$$

В-третьих, по определению $H_k(X)$:

$$H_k(X) = \frac{1}{k} [H(X_1 X_2 \dots X_{k-1}) + H(X_k | X_1 \dots X_{k-1})] = \quad (7)$$

$$= \frac{1}{k} [(k-1)H_{k-1}(X) + H(X_k | X_1 \dots X_{k-1})] \leq \frac{k-1}{k} H_{k-1}(X) + \frac{1}{k} H_k(X)$$

что приводит к:

$$H_k(X) \leq H_{k-1}(X) \quad (8)$$

Следовательно, $H_k(X)$ - не возрастающая последовательность (с ростом k)

Поскольку $H_k(X)$ и условная энтропия $H(X_k|X_1X_2\dots X_{k-1})$ не отрицательны и не возрастающие (с ростом k), оба предела должны существовать. Их предельные выражения могут быть установлены с использованием (1) и (2), чтобы выразить $H_{k+j}(X)$ как:

$$H_{k+j}(X) = \frac{1}{k+j} H(X_1X_2\dots X_{k-1}) + \frac{1}{k+j} [H(X_k|X_1\dots X_{k-1}) + H(X_{k+1}|X_1\dots X_k) + \dots + H(X_{k+j}|X_1\dots X_{k+j-1})] \quad (9)$$

Так как условная энтропия не возрастает, первый член в квадратных скобках является верхней границей для других слагаемых. Следовательно,

$$\frac{1}{k+j} H(X_1X_2\dots X_{k-1}) \leq H_{k+j}(X) \leq \frac{1}{k+j} [H(X_k|X_1\dots X_{k-1}) + H(X_{k+1}|X_1\dots X_k) + \dots + H(X_{k+j}|X_1\dots X_{k+j-1})] \quad (10)$$

Для фиксированного k в пределе для (10) при $j \rightarrow \infty$ получаем:

$$H_\infty(X) \leq H(X_k | X_1 X_2 \dots X_{k-1}) \quad (11)$$

Но (11) справедливо для всех k ; следовательно, это справедливо и для $k \rightarrow \infty$. Поэтому

$$H_\infty(X) \leq \lim_{k \rightarrow \infty} H(X_k | X_1 X_2 \dots X_{k-1})$$

С другой стороны, с учётом (6) мы получаем в пределе для $k \rightarrow \infty$

$$H_\infty(X) \geq \lim_{k \rightarrow \infty} H(X_k | X_1 X_2 \dots X_{k-1}) \quad (13)$$

устанавливает (4).

Предположим, что мы имеем дискретный стационарный источник, который даёт J символов с энтропией на символ $HJ(X)$. Мы можем кодировать последовательность J символов кодом Хаффмена переменной длины, который удовлетворяет префиксному условию при использовании процедуры. Результирующий код имеет среднее число бит для блока с J символами, который удовлетворяет условию:

$$H(X_1 \dots X_J) \leq \overline{R}_J < H(X_1 \dots X_J) + 1 \quad (13)$$

Деля обе части (13) на J , мы получаем границы для среднего числа $\bar{R} = \bar{R}_J / J$ бит на исходный символ как:

$$H_J(X) \leq \bar{R} < H_J(X) + 1/J$$

Увеличивая размер блока J :

$$H_\infty(X) \leq \bar{R} < H_\infty(X) + \xi$$

где ξ стремится к нулю как $1/J$. Таким образом, эффективное кодирование стационарных источников может быть выполнено, если кодировать большие блоки символов в кодовые слова.

Алгоритм Лемпела-Зива

Алгоритм кодирования Хаффмена приводит к оптимальному кодированию источника в том смысле, что кодовые слова удовлетворяют префиксному условию и средняя длина кодового блока минимальна. Конструируя код Хаффмена для ДИБП, мы должны знать вероятности появления всех исходных символов. В случае дискретного источника с памятью мы должны знать совместные вероятности всех блоков длины $n \geq 2$.

В отличие от алгоритма кодирования Хаффмена алгоритм кодирования Лемпела-Зива разработан так, чтобы быть независимым от статистики источника. Следовательно, алгоритм Лемпела-Зива принадлежит классу универсальных алгоритмов кодировать источника.

В алгоритме Лемпела-Зива последовательность с выхода дискретного источника делится на блоки переменной длины, которые называются фразами. Каждая новая фраза; представляет собой последовательность символов источника, отличающуюся от некоторой предыдущей фразы в последнем символе. Фразы перечислены в словаре, который сохраняет расположение существующих фраз. При кодировании новой фразы мы определяем адрес существующей фразы в словаре и добавляем в конец новый символ.

Рассмотрим бинарную последовательность :
10101101001001110101000011001110101100011011.

Деление последовательности, как описано выше, производит следующие фразы:

1,0,10, 11,01,00. 100, 111,010, 1000,011,001, 110, 101,
10001, 1011.

Видим, что каждая фраза в последовательности - соединение одной из предыдущих фраз с новым выходным символом источника. Для кодирования фразы мы конструируем словарь, как показано в табл. 1

Таблица 1. Словарь для алгоритма Лемпела-Зива

	Расположение в словаре	Содержимое словаря	Кодовое слово
1	0001	1	00001
2	0010	0	00000
3	0011	10	00010
4	0100	11	00011
5	0101	01	00101
6	0110	00	00100
7	0111	100	00110
8	1000	111	01001
9	1001	010	01010
10	1010	1000	01110
11	1011	011	01011
12	1100	001	01101
13	1101	110	01000
14	1110	101	00111
15	1111	10001	10101
16		1011	11101

Ячейки словаря пронумерованы последовательно, начиная с 1 и далее, в данном случае до 16, что является числом фраз в последовательности. Различные фразы, соответствующие каждой ячейке, также перечислены, как показано в таблице. Кодовые слова конструируются путём соединения двух частей. Первая часть представляет собой номер ячейки словаря (в двоичной форме) предыдущей фразы, которая соответствует новой фразе, кроме последнего символа. Вторая часть — это новый символ, выданный источником. Он добавляется в конец к первой части, т.е. к номеру ячейки предыдущей фразы. Первоначальный номер ячейки 0000 используется, чтобы кодировать «пустую» фразу

Декодер источника создает идентичную таблицу на приемном конце системы связи и соответственно декодирует полученную последовательность.

Можно заметить, что таблица закодировала 44 исходных бита в 16 кодовых слов по пять битов каждый, что привело к 80 кодированным битам. Следовательно, алгоритм вообще не обеспечил никакого сжатия данных. Однако неэффективность является следствием того, что последовательность, которую мы рассмотрели, очень коротка. По мере увеличения длины последовательности процедура кодирования становится более эффективной и приводит к сжатию последовательности на выходе источника.

Алгоритм Лемпела-Зива широко используется при сжатии компьютерных файлов. «Сжимающие» и «разжимающие» программы (утилиты) в операционной системе UNIX и многочисленные алгоритмы в операционной системе MS DOS являются воплощениями различных версий этого алгоритма.

***«Кодирование для аналоговых
источников - оптимальное
квантование»***

Аналоговый источник выдаёт непрерывный сигнал $x(t)$, который является выборочной функцией случайного процесса $X(t)$. Если $X(t)$ является стационарным случайным процессом с ограниченной полосой, теорема отсчётов позволяет нам представить $X(t)$ последовательностью отсчётов, выбираемых равномерно со скоростью Найквиста.

Применяя теорему отсчётов, выход аналогового источника преобразуется в эквивалентную дискретную во времени последовательность отсчётов. Затем отсчёты квантуются по уровням и кодируются. Один тип простого кодирования - представление каждого дискретного уровня амплитуды последовательностью двоичных символов.

Следовательно, если мы имеем L уровней, нам необходимы $R = \log_2 L$ бит/отсчёт (если L есть степень числа 2) или $R = [\log_2 L] + 1$ (в противном случае).

Если уровни не равновероятны, но вероятности уровней на выходе источника известны, мы можем использовать процедуру кодирования Хаффмена (называемую также энтропийным кодированием), чтобы улучшить эффективность процесса кодирования.

Квантование амплитуд дискретизированного во времени сигнала обеспечивает сжатие данных, но это также приводит к некоторому искажению формы сигнала или потере его точности. Минимизация этих искажений является предметом рассмотрения в данном разделе. Многие результаты, непосредственно применимы к дискретному во времени, непрерывному по амплитуде гауссовскому источнику без памяти. Такой источник служит хорошей моделью для нахождения остаточной ошибки в ряде методов кодирования источника.

Функция скорость-искажение $R(D)$

Под термином «искажение» мы понимаем некоторую меру разности между фактическими выборками источника $\{x_k\}$ и соответствующими квантованными значениями, \tilde{x}_k которую обозначим $d\{x_k, \tilde{x}_k\}$. Например, обычно используемая мера искажения - квадрат ошибки, определенная как:

$$d\{x_k, \tilde{x}_k\} = (x_k - \tilde{x}_k)^2 \quad (1)$$

Используемое для определения ошибки квантования при ИКМ:

$$d\{x_k, \tilde{x}_k\} = |x_k - \tilde{x}_k|^p \quad (2)$$

Где p - принимает значения из ряда положительных целых чисел. Случай $p=2$ имеет предпочтительную математическую трактовку.

Если $d\{x_k, \tilde{x}_k\}$ - мера искажения на отсчёт, искажение между последовательностью отсчётов n и X_n соответствующими n квантованными значениями X_n является средним значением искажения по n отсчётам, т.е.

$$d(X_n, \tilde{X}_n) = \frac{1}{n} \sum_{k=1}^n d(x_k, \tilde{x}_k) \quad (3)$$

На выходе источника имеет место случайный процесс, и, следовательно, n отсчётов X_n в $\{X_n\}$ являются случайными величинами. Поэтому $\{X_n\}$ - случайная величина. Её математическое ожидание определяет искажение D , т.е.

$$D = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2 \quad (4)$$

предположим, что мы имеем источник без памяти с непрерывно-амплитудным выходом X , который имеет ФПВ отсчёта $p(x)$, квантованный амплитудный алфавит \tilde{X} и меру искажения на отсчёт $d\{x_k, \tilde{x}_k\}$ где $x \in X$ и $\tilde{x} \in \tilde{X}$

Тогда минимальная скорость в битах на отсчёт, требуемая для представления выхода X источника без памяти с искажением, меньшим или равным D называется функцией скорость-искажение и определяется как:

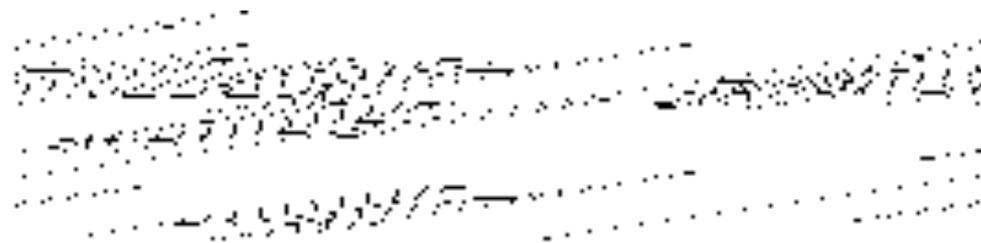
$$R(D) = \min_{p(\tilde{x}|x)} E[d(X, \tilde{X})] \leq D I(X, \tilde{X}) \quad (5)$$

где $I(X, \tilde{X})$ - средняя взаимная информация между X и \tilde{X} . Вообще, скорость $R(D)$ уменьшается при увеличении D или, наоборот, $R(D)$ увеличивается при уменьшении D .

*«Функция скорость-искажение
для гауссовского источника без
памяти»*

Теорема: Функция скорость-искажение для гауссовского источника без памяти (Шеннон, 1959).

Минимальная скорость кодирования, необходимая для представления выхода дискретного во времени, непрерывного по амплитуде гауссовского источника без памяти, при использовании в качестве меры искажения среднеквадратической ошибки на символ (односимвольная мера искажения)

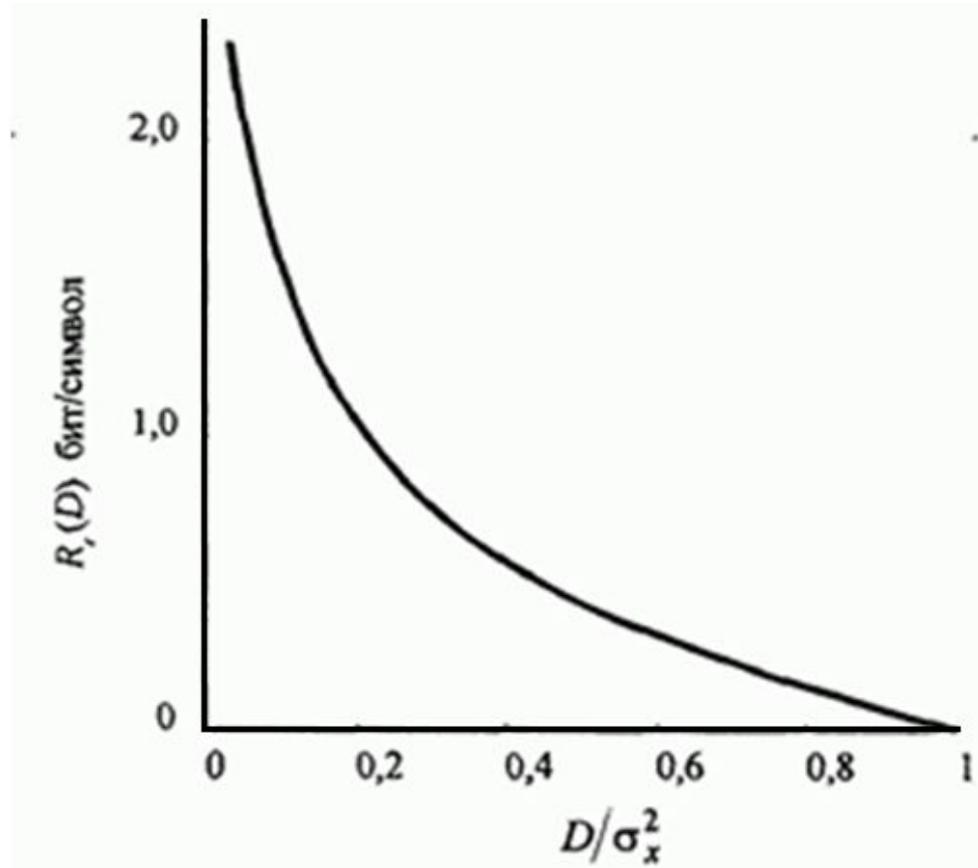


(1)

где σ_x^2 - дисперсия выхода, гауссовского источника.

Заметим, что (1) подразумевает, что, если искажение, $D \geq \sigma_x^2$ никакой информации передавать не нужно. Конкретно при $D = \sigma_x^2$ для реконструкции сигнала достаточно воспроизвести нули. При $D > \sigma_x^2$ для реконструкции сигнала мы можем использовать статистически независимые гауссовские шумовые выборки с дисперсией $D - \sigma_x^2$. График функции $R_g(D)$ представлен на рис.1.

Рис. 1. Функция скорость-искажение для непрерывного по амплитуде гауссовского источника без памяти



Функция скорость-искажение $R(D)$ источника связана со следующей основной теоремой кодирования источника в теории информации.

Теорема: Кодирование источника с заданной мерой искажения (Шеннон, 1959).

Существует схема кодирования, которая отображает выход источника в кодовые слова так, что для любого данного искажения D минимальная скорость $R(D)$ бит на символ (на отсчёт) источника является достаточной для восстановления исходного сигнала со средним искажением, которое является произвольно близким к D .

Это очевидно, потому что функция скорость-искажение $R(D)$ для любого источника представляет нижнюю границу скорости источника, которая является возможной для данного уровня искажения.

Вернёмся к результату в (1) для функции скорость-искажение гауссовского источника без памяти. Если мы поменяем функциональную зависимость между D и R , мы можем выразить D_g через R как

$$D_g(R) = 2^{-2R} \sigma_x^2 \quad (2)$$

Эта функция называется функцией искажение-скорость для дискретного во времени гауссовского источника без памяти

Если искажение в (2) выразить в децибелах, мы получаем

$$10 \log_{10} D_g(R) = -6R + 10 \log_{10} \sigma_x^2 \quad (3)$$

*Верхняя граница для функции
скорость-искажение*

Теорема: Верхняя граница для $R(D)$

Функция скорость-искажение непрерывного по амплитуде источника без памяти с нулевым средним и конечной дисперсией σ_x^2 при использовании среднеквадратичной меры искажений ограничена сверху величиной:

$$R(D) \leq 1/2 \log_2 (\sigma_x^2 / D); \quad (0 \leq D \leq \sigma_x^2) \quad (1)$$

Доказательство этой теоремы дано Бергером (1971). Подразумевается, что гауссовский источник требует максимальную скорость кодирования среди всех других источников при заданном уровне среднеквадратической ошибки. Следовательно, функция скорость-искажение $R(D)$ для произвольного непрерывного источника без памяти с нулевым средним и конечной дисперсией σ_x^2 удовлетворяет условию $R(D) \leq R_g(D)$.

Аналогично функция искажение-скорость того же источника удовлетворяет условию:

$$D(R) \leq D_g(R) = 2^{-2R} \sigma_x^2 \quad (2)$$

Существует также нижняя граница функции скорость-искажение. Её называют нижней границей Шеннона для среднеквадратической ошибки искажения, и она определяется так:

$$R^*(D) = h(X) - 1/2 \log_2 2\pi e D \quad (3)$$

где $h(X)$ - дифференциальная энтропия источника без памяти с непрерывной амплитудой. Функция искажение-скорость, соответствующая (3), равна:

$$D^*(R) = \frac{1}{2\pi e} 2^{-2[R-h(X)]} \quad (4)$$

Следовательно, функция скорость-искажение для произвольного источника без памяти с непрерывной амплитудой ограничена сверху и снизу:

$$R^*(D) \leq R(D) \leq Rg(D) \quad (5)$$

и соответствующая функция искажение-скорость ограничена:

$$D^*(R) \leq D(R) \leq Dg(R) \quad (6)$$

Дифференциальная энтропия гауссовского источника без памяти:

$$h_g(X) = 1/2 \log_2 2\pi e \sigma_x^2 \quad (7)$$

так что нижняя граница $R^*(D)$ в (3) уменьшается до $R_g(D)$. Теперь, если выразить $D^*(R)$ в децибелах и нормировать к $\sigma_x^2=1$ мы получаем из (4):

$$10 \log_{10} D^*(R) = -6R - 6[h_g(X) - h(X)] \quad (8)$$

или, что эквивалентно,

$$10 \log_{10} D^*(R) = -6R - 6 \log_{10} \left(\frac{1 + \sqrt{1 + 2^{2R/3}}}{2} \right) \quad (9)$$

Соотношения в (8) и (9) позволяют сравнивать нижнюю границу искажений с верхней границей, которая определяет искажения для гауссовского источника.

В таблице 1 даны четыре типа ФПВ, которые являются моделями распределения, обычно используемыми для источника сигнала. В таблице даны значения дифференциальной энтропии, различия в скорости (бит на отсчёт) и различия в искажении между верхней и нижней границами. Распределение Лапласа наиболее близко к гауссовскому, а равномерное распределение занимает второе место по близости среди ФПВ, показанных в таблице. Эти результаты дают некоторое представление о различии между верхними и нижними границами искажений и скорости.

Таблица 1. Дифференциальная энтропия и сравнение скорости и искажений четырёх распространённых ФПВ для моделей сигнала

ФПВ	$p(x)$	$h(X)$	$R_g(D) - R^*(D)$ бит/отсчёт	$D_g(R) - D^*(R)$ (дБ)
Гауссовское	$\frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/2\sigma_x^2}$	$\frac{1}{2} \log_2 2\pi e \sigma_x^2$	0	0
Равномерное	$\frac{1}{2\sqrt{3}\sigma_x}, x \leq \sqrt{3}\sigma_x$	$\frac{1}{2} \log_2 12\sigma_x^2$	0,255	1,53
Лапласа	$\frac{1}{\sqrt{2}\sigma_x} e^{-\sqrt{2} x /\sigma_x}$	$\frac{1}{2} \log_2 2e^2 \sigma_x^2$	0,104	0,62
Гамма	$\frac{\sqrt[3]{3}}{\sqrt{8\pi\sigma_x x }} e^{-\sqrt{3} x /2\sigma_x}$	$\frac{1}{2} \log_2 (4\pi e^{0,423} \sigma_x^2 / 3)$	0,709	4,25

Следовательно, эквивалентный дискретный во времени гауссовский источник является источником без памяти. Поэтому функция скорость-искажение для белого гауссовского источника с ограниченной полосой частот в бит/отсчёт равна:

$$R_g(D) = W \log_2 \frac{\sigma_x^2}{D} \quad (0 \leq D \leq \sigma_x^2)$$

Соответствующая функция искажение-скорость:

$$D_g(R) = 2^{-R/W} \sigma_x^2$$

Выражая в децибелах и нормируя к σ_x^2 , получаем:

$$10 \log D_g(R) / \sigma_x^2 = -3R/W$$

Большое количество случаев, в которых гауссовский процесс не является ни белым, ни с ограниченной полосой, было рассмотрено Галлагером (1968) и Гобликом и Холсингером (1967).

Скалярное квантование

При кодировании источника квантователь может быть оптимизирован, если известна ФПВ уровней сигнала на входе квантователя. Например, предположим, что последовательность $\{x_n\}$ на входе квантователя имеет ФПВ $p(x)$ и $L=2^R$ - желаемое число уровней квантования. Необходимо рассчитать оптимальный скалярный квантователь, который минимизирует некоторую функцию ошибки квантования $q = \tilde{x} - x$, где \tilde{x} - квантованное значение x . Для дальнейшей разработки предположим, что $f(\tilde{x} - x)$ определяет желательную функцию ошибки. Тогда искажение, возникающее за счёт квантования сигнальных уровней, равно:

$$D = \int_{-\infty}^{\infty} f(\tilde{x} - x)p(x)dx \quad (1)$$

В общем, оптимальный квантователь минимизирует D путём оптимального выбора выходных уровней и входного диапазона для каждого выходного уровня. Эту оптимизационную проблему рассматривали Ллойд (1982) и Макс (1960), и полученный оптимальный квантователь назван квантователем Ллойда-Макса.

У равномерного квантователя выходные уровни определяются как $\tilde{x}_k = \frac{1}{2}(2k - 1)\Delta$ входного сигнала в диапазоне $(k - 1)\Delta \leq x \leq k\Delta$, где Δ - размер шага. Если квантователь симметричен (относительно нуля) с конечным числом уровней, среднее искажение (1) может быть выражено в виде:

$$D = 2 \sum_{k=1}^{\frac{L}{2}-1} \int_{(k-1)\Delta}^{k\Delta} f\left(\frac{1}{2}(2k-1)\Delta - x\right) p(x) dx + 2 \int_{\left(\frac{L}{2}-1\right)\Delta}^{\infty} f\left(\frac{1}{2}(2k-1)\Delta - x\right) p(x) dx \quad (2)$$

В этом случае минимизация D выполняется с учётом параметра размера шага Δ . Путём дифференцирования D по Δ получаем:

$$\sum_{k=1}^{\frac{L}{2}-1} (2k-1) \int_{(k-1)\Delta}^{k\Delta} f\left(\frac{1}{2}(2k-1)\Delta - x\right) p(x) dx + (L-1) \int_{\left(\frac{L}{2}-1\right)\Delta}^{\infty} f'\left(\frac{1}{2}(L-1)\Delta - x\right) p(x) dx = 0 \quad (3)$$

где $f'(x)$ означает производную $f(x)$. При выборе критериальной функции ошибки $f(x)$ можно получить численное решение (3) для оптимального размера шага на компьютере для произвольной заданной ФПВ $p(x)$.

Для среднеквадратичного критерия ошибки, когда $f(x)=x^2$, Макс(1960) рассчитал оптимальный размер шага $\Delta_{\text{опт}}$ и минимальное значение среднеквадратической ошибки, когда ФПВ $p(x)$ является гауссовской с нулевым средним и единичной дисперсией. Некоторые из этих результатов даны в табл.1

Таблица 1. Оптимальные размеры шага при равномерном квантовании гауссовских случайных величин

Число выходных уровней	Оптимальный размер шага $\Delta_{\text{опт}}$	Минимум СКО $D_{\text{мин}}$	$10 \lg D_{\text{мин}}$ (дБ)
2	1,596	0,3634	-4,4
4	0,9957	0,1188	-9,25
8	0,5860	0,3744	-14,27
16	0,3352	0,01154	-19,38
32	0,1881	0,00349	-24,57

Видим, что минимальная среднеквадратическая ошибка $D_{\text{мин}}$ уменьшается немного больше, чем на 5 дБ, при каждом удвоении числа уровней L . Следовательно, каждый бит, который используется равномерным квантователем с оптимальным размером числа $\Delta_{\text{опт}}$ для гауссовского входного сигнала уменьшает искажение более чем на 5 дБ.

Результирующее искажение:

$$D = \sum_{k=1}^L \int_{x_{k-1}}^{x_k} f(\tilde{x}_k - x)p(x)dx \quad (4)$$

снова минимизируется путём оптимального выбора $\{\tilde{x}_k\}$ и $\{x_k\}$

Необходимые условия для минимальных искажений можно получить дифференцированием D по $\{\tilde{x}_k\}$ и $\{x_k\}$ в такой оптимизации выражается двумя уравнениями:

$$f(\tilde{x}_k - x_k) = f(\tilde{x}_{k+1} - x_k) \quad (5)$$

$$\int_{x_{k-1}}^{x_k} f'(\tilde{x}_k - x)p(x)dx = 0 \quad (6)$$

Как частный случай рассмотрим минимизацию среднеквадратических значений искажений. В этом случае, $f(x)=x^2$, и, следовательно, из (5) следует:

$$x_k = \frac{1}{2}(\tilde{x}_k + \tilde{x}_{k+1}), \quad k=1, 2, \dots, L-1 \quad (7)$$

Соответствующие уравнения, определяющие $\{\tilde{x}_k\}$

$$\int_{x_{k-1}}^{x_k} (\tilde{x}_k - x)p(x)dx = 0, k=1, 2, \dots, L \quad (8)$$

Таким образом, \tilde{x}_k является центроидом области $p(x)$ между x_{k-1} и x_k .

Эти уравнения могут быть решены численно для произвольных ФПВ $p(x)$. Таблицы 2 и 3 дают результаты оптимизации Макса (1960) для оптимального четырёхуровневого и восьмиуровневого квантователя сигнала, распределённого по Гауссу с нулевым средним и единичной дисперсией.

Таблица 2. Оптимальный 4-уровневый квантователь для гауссовской случайной величины

Уровень k	x_k	\tilde{x}_k
1	-0,9816	-1,510
2	0,0	-0,4528
3	0,9816	0,4528
4	∞	1,510

$D_{\min}=0,1175$

$10\lg D_{\min} = -9,3\text{Дб}$

Таблица 3. Оптимальный 8-уровневый квантизатор для гауссовской случайной величины (Макс, 1960)

Уровень k	x_k	\tilde{x}_k
1	-1,748	-2,152
2	-1,050	-1,344
3	-0,5006	-0,7560
4	0	-0,2451
5	0,5006	0,2451
6	1,050	0,7560
7	1,748	1,344
8	∞	2,152

$D_{\text{мин}}=0,03454$

$10\lg D_{\text{мин}} = - 14,62\text{Дб}$

Таблица 4. Сравнение оптимальных равномерного и неравномерного квантизаторов для гауссовской случайной величины (Макс, 1960; Паез и Глиссон, 1972)

R (бит/отсчёт)	$10 \lg D_{\text{мин}}$	
	Равномерное (дБ)	Неравномерное (дБ)
1	-4,4	-4,4
2	-9,25	-9,30
3	-14,27	-14,62
4	-19,38	-20,22
5	-24,57	-26,02
6	-29,83	-31,89
7	-35,13	-37,81

В таблице 4 сравниваются минимальные среднеквадратические искажения для гауссовской амплитуды сигнала в равномерном и неравномерном квантователях.

Поучительно построить кривые зависимости минимальных искажений от битовой скорости $R = \log_2 L$ бит на отсчёт (на символ) источника для равномерного и неравномерного квантователей.

Эти кривые даны на рис.1. Функциональную зависимость D искажений от битовой скорости R можно выразить как $D(R)$ - функцию искажение-скорость. Мы видим, что функция искажение-скорость для оптимального неравномерного квантователя лежит ниже, чем для равномерного квантователя. Если отсчёты сигнала амплитуды статистически независимы, то на выходе квантователя имеем дискретный источник без памяти, и, следовательно, его энтропия:

$$H(\tilde{X}) = -\sum_{k=1}^L p_k \log_2 p_k \quad (9)$$

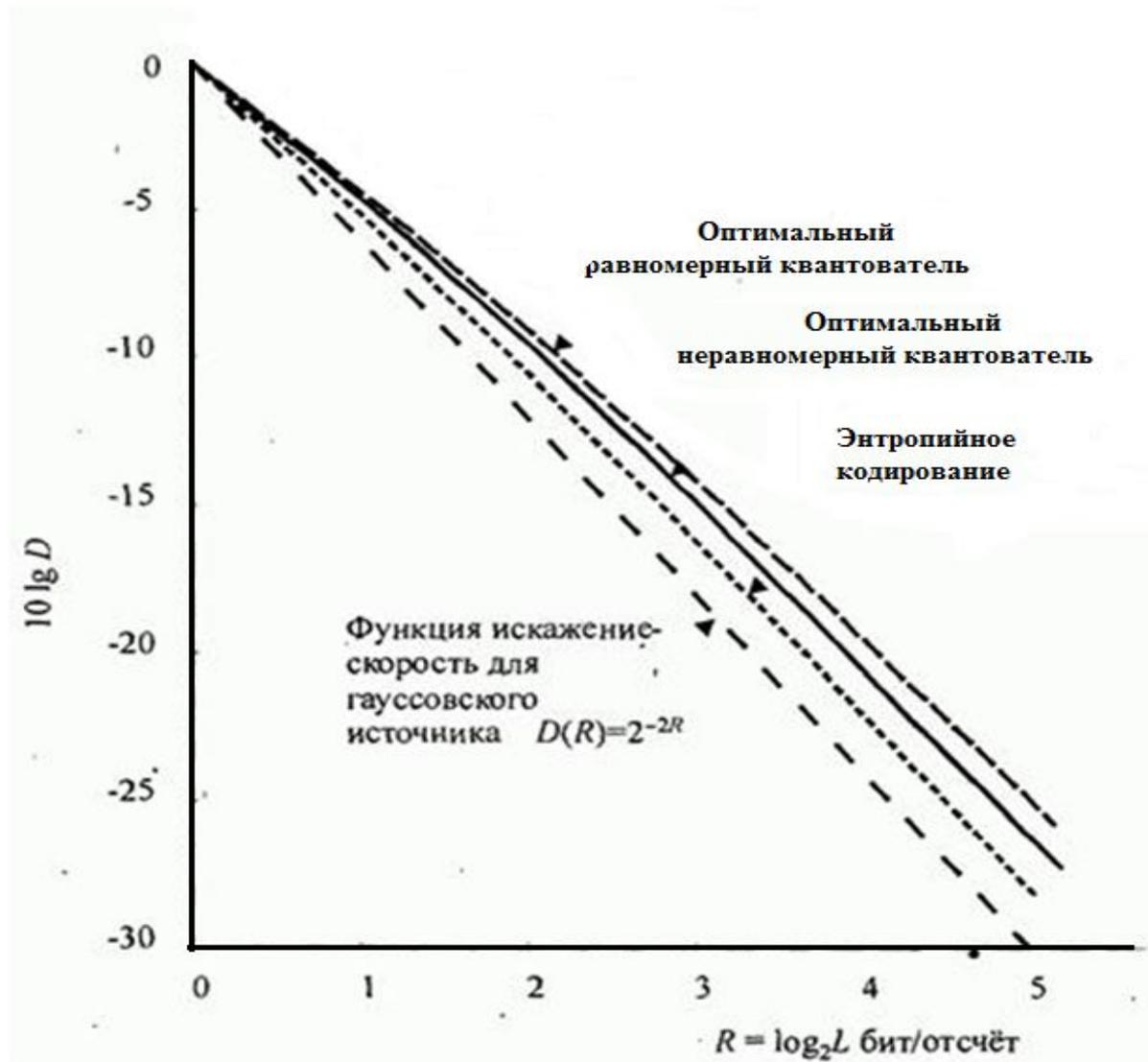


Рис. 1. Кривые зависимости искажение-скорость для гауссовского источника без памяти с дискретным временем

Для примера: оптимальный четырёхуровневый неравномерный квантователь для распределённой по Гауссу амплитуды приводит к вероятностям $p_1=p_4=0,1635$ для двух внешних уровней и $p_2=p_3=0,3365$ для двух внутренних уровней. В этом случае энтропия дискретного источника $H(X)=1,911$ бит/символ. Следовательно, при помощи энтропийного кодирования (кодирование Хаффмена) блоков выходных символов мы можем достичь минимальных искажений (-9,30 дБ) посредством 1,911 бит/символ вместо 2 бит/символ. Макс (1960) определил энтропию для дискретных символов источника после процесса квантования. Зависимость $R(D)$ для этого случая также показана кривой на рис. 1 и обозначена как энтропийное кодирование.

Таблица 5. Энтропия выхода оптимального неравномерного квантователя гауссовской случайной величины (Макс, 1960)

R (бит/отсчёт)	Энтропия (бит/символ)	Искажения $10 \lg D_{\min}$
1	1,0	-4,4
2	1,911	-9,30
3	2,825	-14,62
4	3,765	-20,22
5	4,730	-26,02

Из этого обсуждения мы заключаем, что качество квантователя можно анализировать, когда известна ФПВ непрерывного выхода источника. Оптимальный квантователь с M уровнями обеспечивает минимальное искажение $D(R)$, где $R = \log_2 L$ бит/отсчёт. Такого уровня искажений можно достичь простым представлением каждого квантованного отсчёта битами. Однако возможно более эффективное кодирование.

Дискретные выходы квантователя характеризуются рядом вероятностей $\{p_k\}$, которые можно использовать для расчёта эффективных неравномерных кодов для выхода источника. Эффективность какого-либо метода кодирования можно сравнить с функцией искажение-скорость или, что эквивалентно, с функцией скорость-искажение для дискретного времени и непрерывных амплитуд источника, характеризуемого данной ФПВ.

Если мы сравним характеристики оптимального неравномерного квантователя с функцией искажение-скорость, мы найдём, например, что для искажения в -26 дБ энтропийное кодирование требует скорость на 0,4 бит/отсчёт больше, чем минимальная скорость, а простое блоковое кодирование каждого символа требует скорость на 0,68 бит/отсчёт больше, чем минимальная скорость. Мы также видим, что функция искажение-скорость для оптимального равномерного и неравномерного квантователей гауссовского источника асимптотически приближается к наклону -6 дБ/бит для больших .

Теория построения инфокоммуникационных систем и сетей

Тема: Векторное представление сообщений и сигналов.

Векторное представление сообщений и сигналов

В современной теории передачи информации для описания, анализа и преобразования сообщений и сигналов широко используется геометрическое представление, при котором сигналы рассматриваются – как свойства пространства, преобразование сигналов – как отображение одного пространства в другое.

Введем понятие пространства сигналов. Пусть имеется множество сигналов S , обладающий некоторым общим свойством. Элементы этого множества отличаются друг от друга теми или другими параметрами (амплитудой, длительностью, частотой и т. п.). В общем случае отличия между двумя любыми элементами можно характеризовать некоторым положительным числом, которое трактуется как количественная мера различия сигналов и называется **расстоянием**. Множество сигналов S с подходящим образом определенным расстоянием между элементами называется **пространством** сигналов.

Векторное представление сообщений и сигналов

Для определения расстояния между элементами пространства используют некоторый функционал d , который обычно берется таким, чтобы удовлетворялись требования, являющиеся формализацией свойств, интуитивно связываемых с понятием расстояния:

$$\begin{aligned}d(s_i, s_j) &\geq 0, d(s_i, s_j) = d(s_j, s_i); \\d(s_i, s_k) &\leq d(s_i, s_j) + d(s_j, s_k); \\d(s_i, s_j) &= 0, \text{ только, если } s_i = s_j.\end{aligned}\tag{1}$$

Функционал, удовлетворяющий условиям (1), называется **метрикой**.
Множество S с метрикой d называется **метрическим пространством**.

Векторное представление сообщений и сигналов

Сигналы можно алгебраически суммировать друг с другом. При этом результатом сложения является также сигнал. Сигналы можно усиливать или ослаблять. Все эти свойства находят отражения, если в качестве пространства сигналов взять так называемое **линейное**, или **векторное**, пространство. Оно удовлетворяет следующим условиям:

1. Для любых двух элементов пространства можно определить третий элемент, называемый суммой и входящий в данное пространство, такой, что

$$s_i + s_j = s_j + s_i; s_i + (s_j + s_k) = (s_i + s_j) + s_k$$

2. В пространстве сигналов имеется нулевой элемент 0, такой, что

$$s_i + 0 = s_i \text{ для любых } s_i.$$

3. Для любого элемента s_i , существует противоположный ему элемент $(-s_i)$, принадлежащий данному пространству, такой, что $s_i + (-s_i) = 0$.

Векторное представление сообщений и сигналов

4. Любой элемент пространства можно умножить на любой элемент, принадлежащий скалярному множеству $\{\gamma_i\}$, на котором определены операции сложения и умножения с коммутативными и дистрибутивными свойствами и которое содержит в качестве элементов нуль и единицу, причем $\gamma_i s$ также является элементом пространства сигналов,

$$1 \cdot s = s, (\gamma_i + \gamma_j) s = \gamma_i s + \gamma_j s, \gamma_i (s_k + s_l) = \gamma_i s_k + \gamma_i s_l, \gamma_i (\gamma_j s) = \gamma_i \gamma_j s$$

Элементы линейного пространства обычно называются **векторами**.

Практически все реальные сигналы можно рассматривать как векторы в некотором пространстве. Так, если сигналы представлены последовательностями N действительных чисел, то такие сигналы можно представить N -мерными векторами. В общем случае любой непрерывный сигнал можно рассматривать как бесконечномерный вектор.

Векторное представление сообщений и сигналов

Векторное пространство определяет простые алгебраические взаимосвязи между своими элементами. В частности, любой сигнала как вектор может быть представлен в виде комбинация независимых векторов e_i , $i = 1, 2, \dots, N$, т.е.

$$s_j = \sum_{i=1}^N c_i e_i \quad (2)$$

Представление (2) является единственным, если векторы e_i , $i = 1, 2, \dots, N$, образует **линейно независимую систему**.

Множество всех линейных комбинаций (2) образует N -мерное пространство. Множество линейно независимых векторов e_i , $i = 1, 2, \dots, N$, называется **базисом** этого пространства. Упорядоченную последовательность скаляров c_i , в (2) обычно интерпретируют как **координаты** вектора s_j в базис e_i , $i = 1, 2, \dots, N$. При этом базис интерпретируют как некоторую систему координат, в общем случае косоугольную.

Векторное представление сообщений и сигналов

Любой сигнал можно описать действительной или комплексной функцией, определенной на интервале , кото T_c ий может быть и бесконечным. Множество таких функций образует также линейное пространство. Оно называется **функциональным**. В большинстве случаев функциональное пространство бесконечномерное. Для количественной характеристики сигналов в линейном пространстве вводят **норму**, определяющую длину векторов , обычно обозначают s_j и удовлетворяющую условиям:

$$\begin{aligned} \|s_j\| \geq 0; \|s_j\| = 0, \text{ если } s_j = 0; \|s_i + s_j\| \leq \|s_i\| + \|s_j\|; \\ \|\gamma s_j\| = |\gamma| \|s_j\|, \text{ где } |\gamma| - \text{модуль скаляра } \gamma \end{aligned}$$

Векторное представление сообщений и сигналов

Для N -мерного линейного пространства действительных или комплексных чисел $\{s_j = (s_{j1}, s_{j2}, \dots, s_{jN})\}$, $j = 1, 2, \dots$, норма определяется как

$$\|s_j\| = \left[\sum_{k=1}^N |s_{jk}|^2 \right]^{1/2} \quad (3)$$

а для функционального пространства – как

$$\|s_j\| = \left[\int_0^{T_c} |s_j(t)|^2 dt \right]^{1/2} \quad (4)$$

При таком определении квадрат нормы представляет собой энергию сигнала.

В линейном нормированном пространстве в качестве метрики используется функционал

$$d(s_i, s_j) = \|s_i - s_j\| \quad (5)$$

Векторное представление сообщений и сигналов

Для N -мерного линейного пространства действительных или комплексных чисел с учетом (3) и(5)

$$d(s_i, s_j) = \left[\sum_{k=1}^N |s_{ik} - s_{jk}|^2 \right]^{1/2} \quad (6)$$

а для функционального пространства

$$d(s_i, s_j) = \left[\int_0^{T_c} |s_i(t) - s_j(t)|^2 dt \right]^{1/2} \quad (7)$$

Метрика (7) имеет определенный физический смысл: ее квадрат равен энергии разности двух сигналов, она полностью характеризует различие между сигналами (чем больше $d(s_i, s_j)$, тем больше это различие) и является удобной при расчетах.

Векторное представление сообщений и сигналов

В линейном пространстве можно ввести понятие **скалярного произведения** двух элементов, которое весьма полезно при рассмотрении линейных способов обработки сигналов. Скалярное произведение определяют как

$$(s_i, s_j) = \int_0^{T_c} s_i(t) s_j^*(t) dt \quad (8)$$

для функционального пространства и как

$$(s_i, s_j) = \sum_{k=1}^N s_{ik} s_{jk}^* \quad (9)$$

для N -мерного линейного пространства, где символ $*$ означает комплексно-сопряженную функцию или величину.

Векторное представление сообщений и сигналов

В функциональном анализе доказывается, что в пространстве со скалярным произведением можно ввести норму, удовлетворяющую соотношению

$$\|s_i\| = (s_i, s_i)^{1/2} \quad (10)$$

и метрику

$$d(s_i, s_j) = \|s_i - s_j\| = (s_i - s_j, s_i - s_j)^{1/2} \quad (11)$$

Таким образом, пространство со скалярным произведением можно всегда сделать нормированным и метрическим. Такое пространство при конечном числе N называется **эвклидовым** (обозначается R_N), а при бесконечном N – **гильбертовым** (обозначается L_2).

Введенные понятия пространства, нормы, метрики, базиса позволяют формализовать процессы, связанные с передачей и приемом сигналов.

Векторное представление сообщений и сигналов

Векторное представление применимо как для детерминированных функций, так и для случайных. В последнем случае скалярные произведения (8) и (9), норма (10) и расстояние (11) – случайные величины. Для случайных процессов также справедливо представление в виде (2). При этом коэффициенты c_i являются случайными величинами, а само разложение понимается в смысле среднеквадратической сходимости, т.е.

$$\lim_{N \rightarrow \infty} M \left\{ \left[s_j - \sum_{i=1}^N c_i e_i \right]^2 \right\} = 0 \quad (12)$$

В общем случае коэффициенты разложения c_i коррелированные. Решение многих задач существенно облегчается, если выбрать ортогональный базис, в котором эти коэффициенты оказываются некоррелированными. Разложение случайного процесса по такому базису называется **каноническим**. Для стационарных процессов каноническое разложение всегда возможно.

Теория построения инфокоммуникационных систем и сетей

Тема: Дискретизация непрерывных сообщений с учетом их характеристик и реальных способов восстановления.

Общие сведения

Под дискретизацией понимается процесс представления непрерывного сообщения $x(t)$, заданного на интервале $(0, T_c)$, совокупностью координат c_1, c_2, \dots, c_N . В общем случае процессы представления и восстановления описываются выражениями:

$$(c_1, c_2, \dots, c_N) = A[x(t)] \quad (1)$$

$$\tilde{x}(t) = A'[(c_1, c_2, \dots, c_N)] \quad (2)$$

где A – оператор дискретного представления; A' – оператор восстановления.

Операторы дискретного представления и восстановления могут быть как линейными, так и нелинейными. На практике обычно используют линейные операторы как более простые в реализации.

Общие сведения

При линейных процессах представления и восстановления выражения (1) и (2) можно представить в виде

$$c_i = \int_0^{T_c} x(t)\phi_i(t)dt, \quad i = 1, \dots, N, \quad (3)$$

$$\tilde{x}(t) = \sum_{i=1}^N c_i \varphi_i(t), \quad (4)$$

где $\phi_i(t)$ и $\varphi_i(t)$ – **весовые** и **базисные (координатные)** функции.

В зависимости от системы используемых весовых функций $\phi_i(t), i = 1, \dots, N$, различают **дискретное временное, дискретное обобщенное и дискретное разностное представления.**

Дискретное временное представление

В данном случае используется система весовых функций $\phi_i(t) = \delta(t - t_i)$, $i = 1, 2, \dots, N$, где $\delta(t - t_i)$ – дельта-функция. При этом, как это следует из (3), координаты $c_i = x(t_i)$, $i = 1, \dots, N$, т.е. совпадают с мгновенными значениями (отсчетами) непрерывной функции $x(t)$ в дискретные моменты t_i .

Представление называется регулярным, если шаг дискретизации $T_D = t_i - t_{i-1}$ постоянный. В противном случае оно называется нерегулярным.

При представлении сообщений регулярными отсчетами основным является выбор частоты дискретизации $F_D = 1/T_D$ и базисных функций $\phi_i(t)$. Особенно важно найти минимальную частоту F_D , при которой еще имеется принципиальная возможность восстановления непрерывного сообщения с заданной погрешностью. При решении этих задач следует принимать во внимание свойства исходных сообщений, способы восстановления и требуемую точность восстановления.

Дискретное временное представление

Для модели сообщения с ограниченным спектром решение указанных задач содержится в теореме Котельникова, на основании которой любую непрерывную функцию со спектром, ограниченным полосой частот от нуля до F_{max} , можно однозначно определить последовательностью ее мгновенных значений, взятых через интервалы времени $T_d = \frac{1}{2F_{max}}$.

Восстановление непрерывной функции производится в соответствии с выражением

$$\tilde{x}(t) = \sum_{i=-\infty}^{\infty} x(iT_d) \frac{\sin 2\pi F_{max}(t - iT_d)}{2\pi F_{max}(t - iT_d)}, \quad (5)$$

которое называется **рядом Котельникова**. Базисными функциями в данном случае служат функции отсчетов

$$\varphi_i(t) = \frac{\sin 2\pi F_{max}(t - iT_d)}{2\pi F_{max}(t - iT_d)}, \quad i = \dots, -1, 0, 1, \dots$$

Дискретное временное представление

Они образуют ортогональную на бесконечном интервале $-\infty < t < \infty$ Систему функций. Любую функцию $\varphi_i(t)$ можно получить на выходе идеального фильтра нижних частот, подав на его вход сигнал $\delta(t - iT_d)$.

Идеальным называется фильтр нижних частот, у которого комплексная частотная характеристика имеет вид

$$K(j\omega) = \begin{cases} K_0, & -\omega_{max} \leq \omega \leq \omega_{max} \\ 0, & \omega < -\omega_{max}, \omega > \omega_{max} \end{cases}$$

В соответствии с (5) непрерывное сообщение восстанавливается, если на входе идеального фильтра нижних частот с полосой пропускания $0 \dots F_{max}$ подать последовательность δ -функций $\delta(t - iT_d), i = \dots, -1, 0, 1, \dots$, умноженных на коэффициенты $x(iT_d)$. Однако ни сигнал в виде δ -функции, ни идеальный фильтр нижних частот физически нереализуемы. Поэтому на практике вместо δ -функций используют короткие импульсы, а вместо идеального фильтра нижних частот – фильтр нижних частот, что, естественно, приводит к погрешности восстановления.

Дискретное временное представление

Теорему Котельникова можно распространить и на случайный сигнал. Тогда она формулируется следующим образом: для случайного процесса с односторонней спектральной плотностью мощности, удовлетворяющей условию $G_x(f) = 0$ при $f > F_{max}$, ряд

$$\sum_{i=-\infty}^{\infty} X(iT_D) \frac{\sin 2\pi F_{max}(t - iT_D)}{2\pi F_{max}(t - iT_D)},$$

где $X(iT_D)$ – случайные величины, представляющие собой отсчеты случайного процесса, взятые через интервал времени $T_D = \frac{1}{2} F_{max}$, сходится в среднеквадратическом смысле (6) к процессу $X(t)$.

$$\lim_{N \rightarrow \infty} M \left\{ \left[s_j - \sum_{i=1}^N c_i e_i \right]^2 \right\} = 0 \quad (6)$$

Дискретное временное представление

Теорема Котельникова дает предельные соотношения для идеализированных условий, среди которых следует отметить ограниченность спектра по частоте и бесконечное время наблюдения. Все реальные сигналы конечны во времени и имеют неограниченный по частоте спектр. Использование модели с ограниченным спектром и конечное время наблюдения приводит к погрешности при восстановлении непрерывного сообщения.

Тем не менее теорема Котельникова имеет большое практическое значение. Дело заключается в том, что спектр сигнала так или иначе ограничивается (например, при передаче непрерывного сообщения спектр $G(f)$ целесообразно ограничить частотой F_{max} , при которой $G(f) < N(f)$, где $N(f)$ – спектральная плотность мощности шума на выходе канала). В этих случаях теорема Котельникова позволит сориентироваться в отношении частоты дискретизации. Обычно ее определяют по приближенной формуле

$$F_d \approx 2\lambda F_{max}, \text{ где } \lambda \text{ – некоторый коэффициент, равный } 1,25 \dots 2,5.$$

Дискретное временное представление

Ограничение спектра сообщения частотой F_{max} путем фильтрации приводит к погрешности восстановления, относительный средний квадрат которой

$$\delta_F^2 = \frac{\int_{F_{max}}^{\infty} G(f)df}{\int_0^{\infty} G(f)df}, \quad (7)$$

т. е. равен отношению мощности отброшенной части спектра к средней мощности исходного сообщения.

При отсутствии предварительной фильтрации в процессе восстановления сообщения ошибка дискретизации возрастает. Пусть $S_x(j\omega)$ – спектральная плотность сообщения $x(t)$. Тогда спектральная плотность дискретизированного сигнала $x_D(t)$

$$S_{xD}(j\omega) = \frac{1}{T_D} \sum_{n=-\infty}^{\infty} S_x \left[j \left(\omega - \frac{2\pi n}{T_D} \right) \right], \quad (8)$$

Дискретное временное представление

т. е. она представляет собой с точностью до несущественного множителя $\frac{1}{T_d}$ сумму бесконечного числа «копий» спектра исходного сообщения (рис. 1). Эти копии располагаются на оси частот через равные промежутки $2\pi/T_d$.

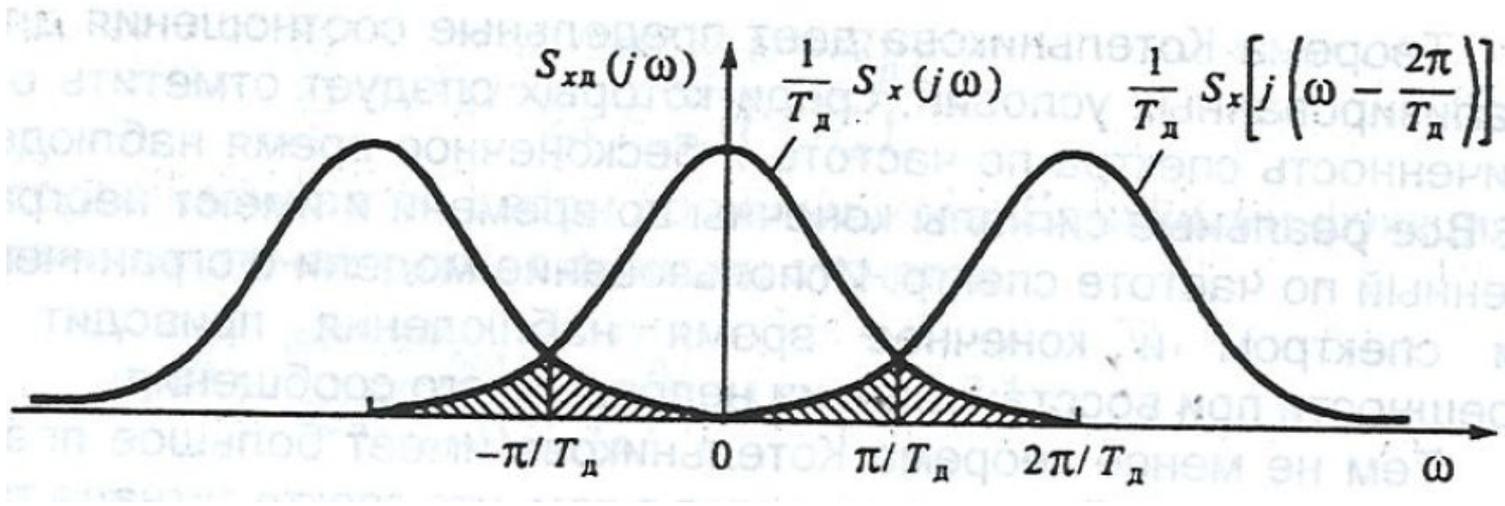


Рис. 1. Спектральная плотность дискретизированного сигнала

Дискретное временное представление

При восстановлении сообщения идеальным фильтром нижних частот с полосой пропускания $-\pi/T_D \leq \omega \leq \pi/T_D$ возникает ошибка, относительный квадрат которой с учетом (8) определяется как

$$\delta_{\text{от.ф}}^2 = \frac{\int_{|\omega| > \frac{\pi}{T_D}} |S_x(j\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |S_x(j\omega)|^2 d\omega} + \frac{\int_{-\frac{\pi}{T_D}}^{\frac{\pi}{T_D}} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \left| S_x \left[j \left(\omega - \frac{2\pi n}{T_D} \right) \right] \right|^2 d\omega}{\int_{-\infty}^{\infty} |S_x(j\omega)|^2 d\omega}. \quad (9)$$

Первое слагаемое (9) характеризует ошибку, обусловленную тем, что составляющие сигнала $x_D(t)$ на частоте $|\omega| > \pi/T_D$ не попадают в полосу пропускания фильтра, и совпадает по значению (7). Второе слагаемое в (9) характеризует ошибку, обусловленную попаданием в полосу фильтра составляющих копий $S_x[j(\omega - 2\pi n/T_D)]$, $n = \pm 1, \pm 2, \dots$. Если ограничить только влиянием копий с $n = \pm 1$, то нетрудно видеть, что второе слагаемое также совпадает по значению с (7).

Дискретное временное представление

При этом

$$\delta_{\text{от.ф}}^2 = 2\delta_F^2, \quad (10)$$

и, следовательно, предварительная фильтрация сообщения с целью ограничения его спектра является целесообразной.

Заметим, что обеспечить условие $G(f) = 0$ при $f > F_{\text{max}}$ физически невозможно. Сообщение на выходе любого реализуемого фильтра будет содержать составляющие на частотах $f > F_{\text{max}}$. Поэтому ошибка (7) является минимальной возможной.

В общем случае восстановление (интерполяция) непрерывного сообщения $x(t)$ по его отсчетам выполняется в соответствии с (4). При этом в качестве базисных функций широко используют алгебраические полиномы. В частности, на практике часто применяются ступенчатая и линейная интерполяции. При ступенчатой интерполяции (рис. 2,а) используется только один отсчет. Функция $\varphi_1(t) = 1$, а $\tilde{x}(t_i), t_i \leq t \leq t_i + T_D$.

Дискретное временное представление

При линейной интерполяции (рис. 2,б) используются два отсчета.

Функции $\varphi_1(t) = 1 - \tau/T_D$, $\varphi_2(t) = \tau/T_D$, а

$$\tilde{x}(t) = x(t_i) + [x(t_{i+1}) - x(t_i)] \tau/T_D, \quad t_i \leq t \leq t_i + T_D = t_{i+1}, \quad \tau = t - t_i.$$

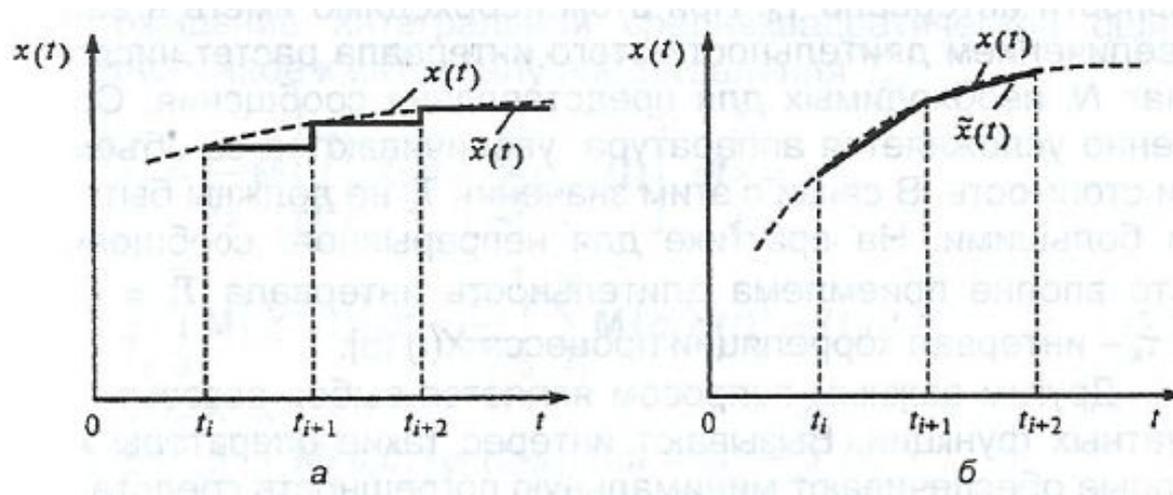


Рис. 2. Диаграммы, иллюстрирующие ступенчатую (а) и линейную (б) интерполяции

Дискретное временное представление

Относительный средний квадрат погрешности интерполяции зависит от нормированной корреляционной функции $r_x(\tau)$ исходного процесса $X(t)$, способа интерполяции и частоты дискретизации. В показано, что для любых стационарных процессов с нулевым математическим ожиданием при ступенчатой интерполяции

$$\delta^2 = 2 \left[1 - \frac{1}{T_D} \int_{-\frac{T_D}{2}}^{\frac{T_D}{2}} r_x(\tau) d\tau \right], \quad (11)$$

при линейной интерполяции

$$\delta^2 = \frac{5}{3} + \frac{1}{3} r_x(T_D) - \frac{4}{T_D} \int_0^{T_D} r_x(\tau) d\tau + \frac{4}{T_D^2} \int_0^{T_D} \tau r_x(\tau) d\tau. \quad (12)$$

Дискретное временное представление

При заданной погрешности интерполяции формы (11) и (12) используется для нахождения частоты дискретизации. Расчеты показывают, что частота F_d существенно превышает частоту дискретизации по Котельникову. Так, для сигнала с прямоугольной спектральной плотностью мощности, ограниченной частотой F_{max} , отношение $F_d/2F_{max}$ равно $\pi/6\delta$ при ступенчатой интерполяции и $\pi/\sqrt[4]{600\delta^2}$ при линейной.

Обобщенное дискретное преобразование

В данном случае координаты сообщения $c_i, i = 1, \dots, N$, (3) представляют собой коэффициенты некоторого ряда. При решении рассматриваемой задачи важным вопросом является выбор длительности интервала T_c . При этом необходимо иметь в виду, что с увеличением длительности этого интервала растет число координат N , необходимых для представления сообщения. Соответственно усложняется аппаратура, увеличивается ее объем, масса и стоимость. В связи с этим значение не должно быть слишком большим. На практике для непрерывного сообщения $X(t)$ вполне приемлема длительность интервала корреляции процесса $X(t)$.

$$T_c = (5 \dots 6) \tau_K, \text{ интервал}$$

Другим важным вопросом является выбор весовых и координатных функций. Вызывают интерес такие операторы A и A' , которые обеспечивают минимальную погрешность представления δ^2 при заданном числе координат N или минимальное число координат N при заданной погрешности δ^2 .

Обобщенное дискретное преобразование

Пусть $X(t)$ – случайный процесс с нулевым математическим ожиданием и непрерывной корреляционной функцией $R(t, t')$. Тогда можно показать, что математическое ожидание интегральной среднеквадратической ошибки при представлении процесса $X(t)$ (4)

$$M \left\{ \int_0^{T_c} \left[X(t) - \sum_{i=1}^N c_i \varphi_i(t) \right]^2 dt \right\} \quad (13)$$

при любом фиксированном N будет минимальным, если весовые функции ϕ_i , $i = 1, \dots, N$, совпадают с базисными функциями φ_i , $i = 1, \dots, N$, которые функции удовлетворяют однородному интегральному уравнению Фредгольма второго рода:

$$\sigma_i^2 \varphi_i(t) = \frac{1}{T_c} \int_0^{T_c} R_x(t, t') \varphi_i(t') dt', \quad (14)$$

где $\varphi_i(t)$ – собственная функция; σ_i^2 – собственные значения ядра $R_x(t, t')$ уравнения.

Обобщенное дискретное преобразование

Собственные функции φ_i , $i = 1, \dots, N$, являются ортогональными и определяются уравнениями с точностью до постоянного множителя, который можно выбрать таким, чтобы функции φ_i , $i = 1, \dots, N$, были ортонормированными. При этом координаты в разложении (4) случайного процесса оказываются некоррелированными, а для случайного процесса – статистически независимыми. Кроме того, при $M\{c_i\}$ дисперсии D_{c_i} координаты c_i равны σ_i^2 .

Если базисные функции ортонормированные, то математическое ожидание интегральной среднеквадратической ошибки (13), отнесенное к интервалу представления T_c ,

$$\begin{aligned} \overline{\varepsilon^2} T_c &= \frac{1}{T_c} M \left\{ \int_0^{T_c} \left[X(t) - \sum_{i=1}^N c_i \varphi_i(t) \right]^2 dt \right\} = \\ &= \frac{1}{T_c} \int_0^{T_c} M\{X^2(t)\} dt - \frac{2}{T_c} \int_0^{T_c} \sum_{i=1}^N M\{c_i X(t)\} \varphi_i(t) dt + \\ &+ \frac{1}{T_c} \int_0^{T_c} \sum_{j=1}^N \sum_{i=1}^N M\{c_i c_j\} \varphi_i(t) \varphi_j(t) dt = D_x - \sum_{i=1}^N D_{c_i}. \end{aligned} \quad (15)$$

Обобщенное дискретное преобразование

Выражение (15) позволяет находить число координат N , при котором обеспечивается заданная погрешность дискретного представления.

Разложение случайного процесса с непрерывной корреляционной функцией в ряд (4), в котором базисные функции являются собственными функциями уравнения (14), называется **разложением Карунена-Лоэва**. Хотя это разложение обеспечивает минимальное число координат N при заданной погрешности дискретного представления случайного процесса $\overline{\varepsilon^2}$, однако его применение на практике ограничено. Это обусловлено следующими причинами: корреляционная функция случайного процесса не всегда оказывается известной, процедура отыскания решения уравнения (14) в общем случае неизвестна, техническая реализация устройства разложения сигнала за исключением случая, когда функция $\varphi_i(t)$ гармонические, сложная. Поэтому на практике в качестве базисных часто используют ортогональные функции, при которых погрешность представления близка к минимальной при сравнительно простой аппаратуре. К ним относятся тригонометрические функции, полиномы Чебышева и Лежандра, функции Уолша и др.

Дискретное разностное представление

В данном случае в качестве весовых функций $\phi_i(t)$ используют линейные комбинации дельта-функций:

$$\phi_i(t) = \sum_{k=0}^L (-1)^k C_L^k \delta(t - t_i + kT_D), L = 1, 2, \dots, \quad (16)$$

где C_L^k — число сочетания из L по k . При этом, как следует из (3), координатами являются конечные разности L -го порядка

$$\Delta^L x(t_i) = \sum_{k=0}^L (-1)^k C_L^k x(t_i - kT_D).$$

В частности, при $L = 1$

$$\phi(t) = \delta(t - t_i) - \delta(t - t_{i-1}), \quad c_i = \Delta x(t_i) = x(t_i) - x(t_{i-1}).$$

Адаптивная дискретизация непрерывных сообщений

В данном случае координатами являются мгновенные значения непрерывного сигнала в некоторых точках опроса, неравноотстоящих друг от друга (рис. 3). На интервалах, где функция меняется в больших пределах, отсчеты берутся чаще, а на интервалах медленного изменения – реже. Для представления сообщения стараются использовать как можно меньшее число отсчетов, но достаточное для восстановления сообщения с заданной погрешностью. Отсчеты, позволяющие восстановить непрерывное сообщение на приемной стороне с заданной точностью, называются обычно **существенными**.

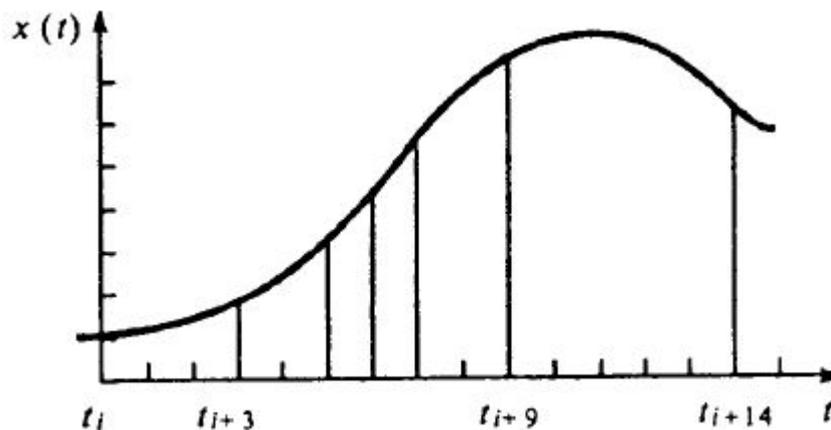


Рис. 3. Пример размещения существенных выборок при линейной интерполяции

Адаптивная дискретизация непрерывных сообщений

Известны различные способы адаптивной дискретизации, отличающиеся алгоритмом формирования существенных отсчетов и видов служебной информации. Простейшим алгоритм формирования существенных отсчетов заключается в следующем. Пусть последний существенный отсчет был в момент t_i . Для формирования следующей выборки сравнивают текущее значение функции $x(t)$ с $x(t_i)$. Ближайший момент t_{i+j} , при котором $|x(t_{i+j}) - x(t_i)| \geq \varepsilon_m$, соответствует очередной существенной выборке.

Адаптивная дискретизация непрерывных сообщений

При адаптивной дискретизации отсчеты передаются в случайные моменты. Поэтому для восстановления непрерывного сообщения по отсчетам приемная сторона должна знать, к каким тактовым моментам относятся принятые отсчеты. В связи с этим на приемную сторону приходится передавать дополнительную служебную информацию. Такой информацией могут быть значения тактовых моментов, соответствующих существенным выборкам. При сравнении различных способов представления это обстоятельство необходимо учитывать.

Адаптивные способы дискретизации широко применяют при отсутствии априорной информации о корреляционной функции или спектральной плотности мощности непрерывных сообщений.

Искажения сигналов в непрерывных каналах

Физическая модель непрерывного канала связи, представляющего наибольший интерес при анализе работы РСПИ, включает в свой состав технические средства, расположенные между выходом модулятора и входом демодулятора (рис. 1).

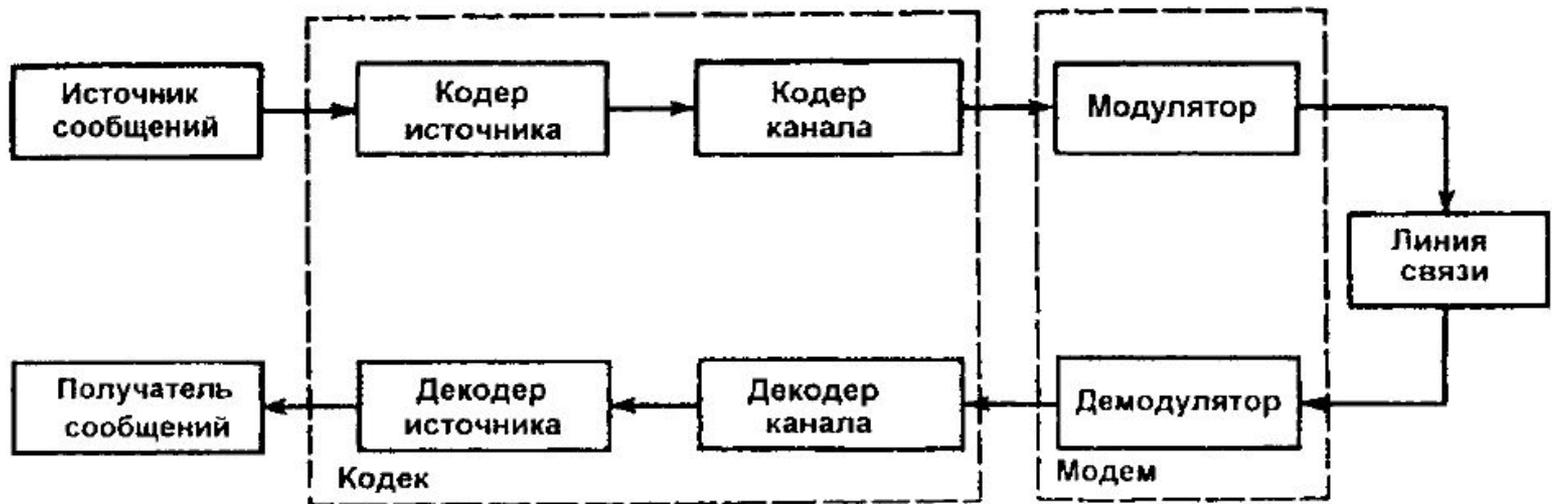


Рис. 1. Структурная схема системы передачи дискретных сообщений

Искажения сигналов в непрерывных каналах

Проходя по непрерывному каналу связи, сигнал претерпевает ряд изменений. Эти изменения сводятся к ослаблению, искажению сигнала и наложению на него помех. В отдельных случаях искажению подвергается смесь сигнала и помех, например, во входных цепях приемника или при ретрансляции в радиорелейных линиях. Для анализа системы важно знать характер искажений и уметь их моделировать. Реальные искажения имеют достаточно сложный характер. Однако для решения большинства задач непрерывный канал можно смоделировать в виде последовательно включенных линейных инерционных и нелинейных безынерционных четырехполюсников, обуславливающих соответственно линейные и нелинейные искажения сигналов (рис. 2). Помехи принципиально могут накладываться на сигнал в любой точке цепи. Несмотря на кажущуюся простоту такой модели канала, нахождение отклика на ее выходе в тех случаях, когда помеха действует на входе нелинейного звена, является сложной математической задачей. Поэтому часто при решении подобных задач обращаются к машинному или физическому моделированию.

Искажения сигналов в непрерывных каналах

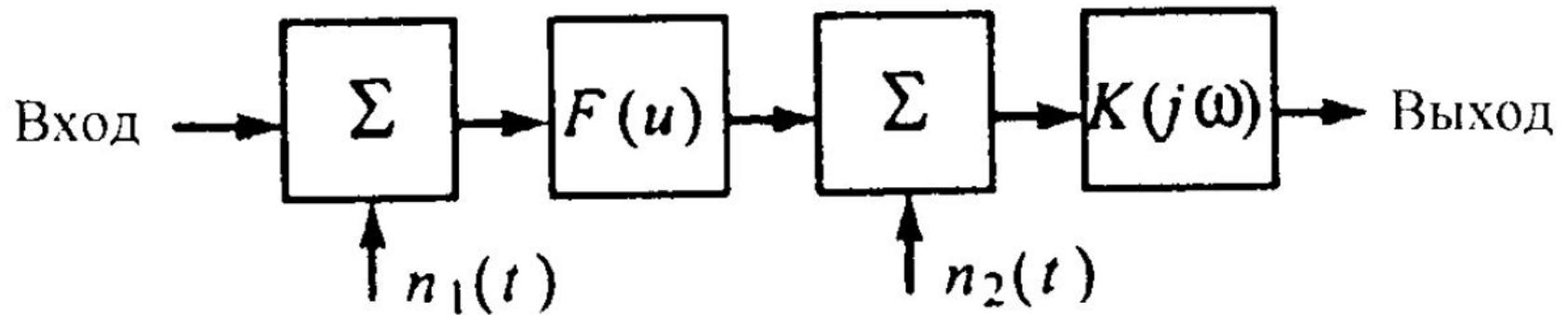


Рис. 2. Модель непрерывного канала связи

Линейные искажения в непрерывных каналах

Линейные искажения проявляются в изменении спектра (корреляционной функции) сигналов и помех. В зависимости от того, каковы эти искажения: регулярны или случайны, различают соответственно каналы с **детерминированными** или **случайными** линейными искажениями. Детерминированные линейные искажения в реальных каналах связаны с наличием частотно-избирательных цепей (фильтров во входных каскадах приемника и в выходных каскадах передатчика, коаксиальных и волноводных трактов, антенн и т.д.).

Случайные линейные искажения определяются средой распространения и связаны в основном с прохождением сигнала от передающей антенны к приемной антенне разными путями (лучами). Этот эффект называется **рассеянием** сигнала. Различают два вида рассеяния сигнала: **дискретное**, когда запаздывание между сигналами в соседних лучах принимает конкретное значение (многолучевой канал), и **дисперсное**, когда запаздывание между соседними лучами бесконечно мало, а число лучей бесконечно велико.

Линейные искажения в непрерывных каналах

Характер рассеяния сигнала определяется диапазоном используемых частот и типом системы. Если раньше типичными каналами с рассеянием сигнала являлись тропосферный и ионосферный, в которых связь за пределами прямой видимости достигалась за счет переотражения сигналов, и были найдены способы борьбы с многолучевостью, то в последние десятилетия в связи с развитием мобильных систем связи, действующие в условиях городской застройки, борьба с многолучевостью приобрела еще большую актуальность. Искажения сигналов, особенно применительно к мобильным системам, носят достаточно сложный характер. Однако даже упрощенные модели позволяют разобраться в характере искажения сигналов и находить способы повышения качества передачи информации по каналам с рассеянием.

Линейные искажения в непрерывных каналах

Пусть в точку приема приходят сигналы, переотраженные от совокупности бесконечно малых по размеру отражателей, размещенных в некотором пространстве, которое случайным образом перемещаются, сохраняя постоянным в среднем объем занимаемого пространства. Тогда на входе приемника будем иметь сумму сигналов с разной амплитудой и временем прихода, которые свою очередь случайно изменяются с некоторой скоростью. Максимальную разницу во времени прихода сигналов называют временем рассеяния сигнала, расширением задержки или памятью канала. Естественно, что эта величина также носит случайный характер, но можно указать ее среднее значение. Рассмотрим характер искажения гармонического сигнала при прохождении по такому каналу. Сигнал на входе приемника представляет сумму синусоид со случайными амплитудами и фазами. Если число переотраженных сигналов велико, то в соответствии с центральной предельной теоремой теории вероятностей суммарные ортогональные сигналы будут иметь нормальные законы распределения амплитуд, а результирующий сигнал будет иметь случайную огибающую и

Линейные искажения в непрерывных каналах

фазу, изменяющиеся соответственно по рэлеевскому и равномерному законам.

Скорость изменения (ширина спектра флуктуации или время корреляции) определяются доплеровским сдвигом по частоте при движении отражателей. Обычно в мобильных системах связи переотраженные сигналы действуют на фоне достаточно мощного прямого сигнала. Тогда результирующий сигнал будет иметь райсовский закон распределения огибающей.

Если число лучей ограничено, например только два, то результирующий сигнал будет представлять биения последних. Случайный характер огибающей и фазы результирующего сигнала будет определяться характером изменения амплитуд и фаз суммируемых сигналов.

Для сигналов с фиксированной шириной спектра F канала с рассеянием представляет собой фильтр со случайно изменяющимися во времени параметрами.

Линейные искажения в непрерывных каналах

В общем случае сигнал на выходе линейного канала с изменяющимися параметрами можно найти, используя интеграл Дюамеля

$$u(t) = \int_0^t h(t, \tau) s(t - \tau) d\tau,$$

где $h(t, \tau)$ – импульсная характеристика канала.

Таким образом, для оценки линейных искажений необходимо знать функцию $h(t, \tau)$ или связанную с ней преобразованием Фурье комплексную частотную характеристику $K(j\omega, t)$.

Решение задач анализа и синтеза устройств обработки сигналов существенно упрощается, если перейти к дискретной модели канала и сигналов. Дискретное представление математических моделей каналов основывается на конечном времени рассеяния сигнала τ_{Π} , определяемой протяженностью импульсной характеристики $h(t, \tau)$, и конечной ширине спектра передаваемого сигнала F_c . Формальным способом введения дискретной модели может быть разложение функции в ряд Котельникова,

Линейные искажения в непрерывных каналах

Фурье и т.п. Если полоса частот сигнала, передаваемого по каналу, ограничена интервалом $F_c = F_B - F_H$, то достаточно рассматривать функцию $K(j2\pi f, t)$ по переменной f только в интервале F_c . При этом импульсную характеристику $h(t, \tau)$ можно представить в виде ряда Котельникова для сигнала с полосовым спектром

$$h(t, \tau) = \sum_{i=-\infty}^{\infty} h_i(t, \tau) = \sum_{i=-\infty}^{\infty} H(i, t) \frac{\sin \left[\pi F_c \left(\tau - \frac{i}{F_c} \right) \right]}{\pi F_c \left(\tau - \frac{i}{F_c} \right)} \times \quad (1)$$
$$\times \cos \left[\pi (F_B + F_H) \left(\tau - \frac{i}{F_c} \right) + \varphi(i, t) \right],$$

где $H(i, t)$ – значение огибающей импульсной характеристики при $\tau = i/F_c$,
 $\varphi(i, t)$ – значение фазы.

Линейные искажения в непрерывных каналах

Физическая модель канала, построенная в соответствии с (1), (рис. 3), содержит линию задержки с L отводами через $1/F_c$, усилитель, комплексный коэффициент которых $h(i, t) = H(i, t) = \exp[j\varphi(i, t)]$ может изменяться, и сумматор.

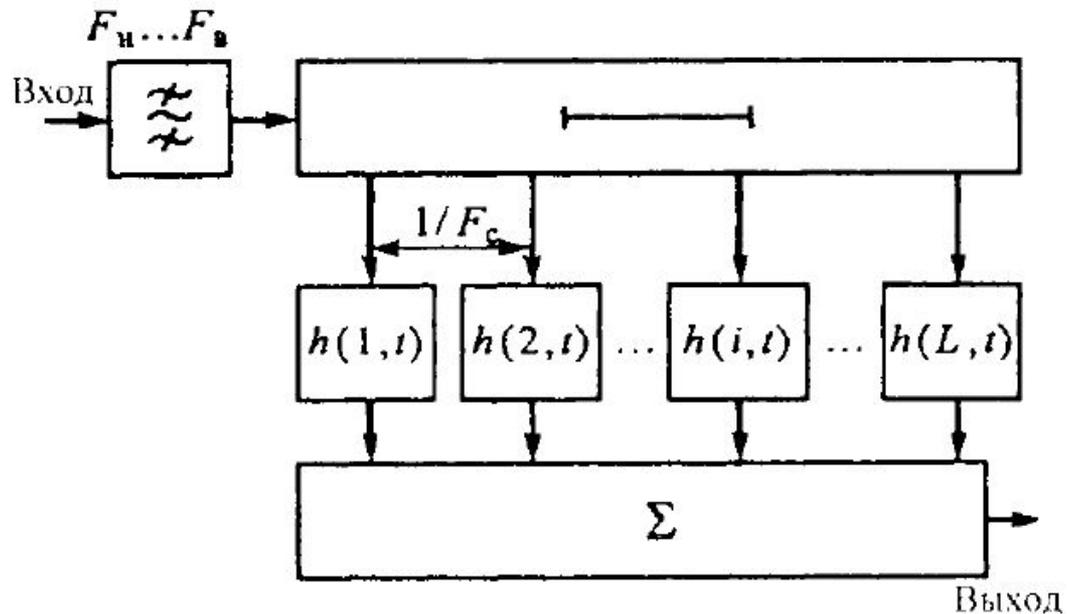


Рис. 3. Модель канала с рассеянием для сигналов с ограниченной шириной спектра

Линейные искажения в непрерывных каналах

В частотной области модель канала можно построить в предположении конечности времени рассеяния сигнала. Тогда функция по $K(j\omega, t)$ может быть $f = \omega/2\pi$ комплексными значениями $K(i, t)$ взятыми через частотный интервал $\Delta f = 1/\tau_{\Pi}$ в герцах. Дискретная модель канала содержит набор полосовых фильтров с примыкающими частотными характеристиками, полоса пропускания каждого из которого $1/\tau_{\Pi}$ и усилителей с управляемыми комплексными коэффициентами передачи $K(i, t)$ (рис.4). Величина Δf иногда называют полосой когерентности. Гармонические сигналы с разносом по частоте, превышающим Δf , будут иметь некоррелированные случайные огибающую и фазу. Этот параметр определяет и характер замираний. Если ширина спектра передаваемого по каналу сигнала меньше Δf , то все спектральные составляющие сигнала изменяются одновременно и такие замирания называются общими. В том случае, когда $F \gg 1/\tau_{\Pi}$, отдельные участки спектра сигнала изменяются независимо и замирания называются селективными.

Линейные искажения в непрерывных каналах

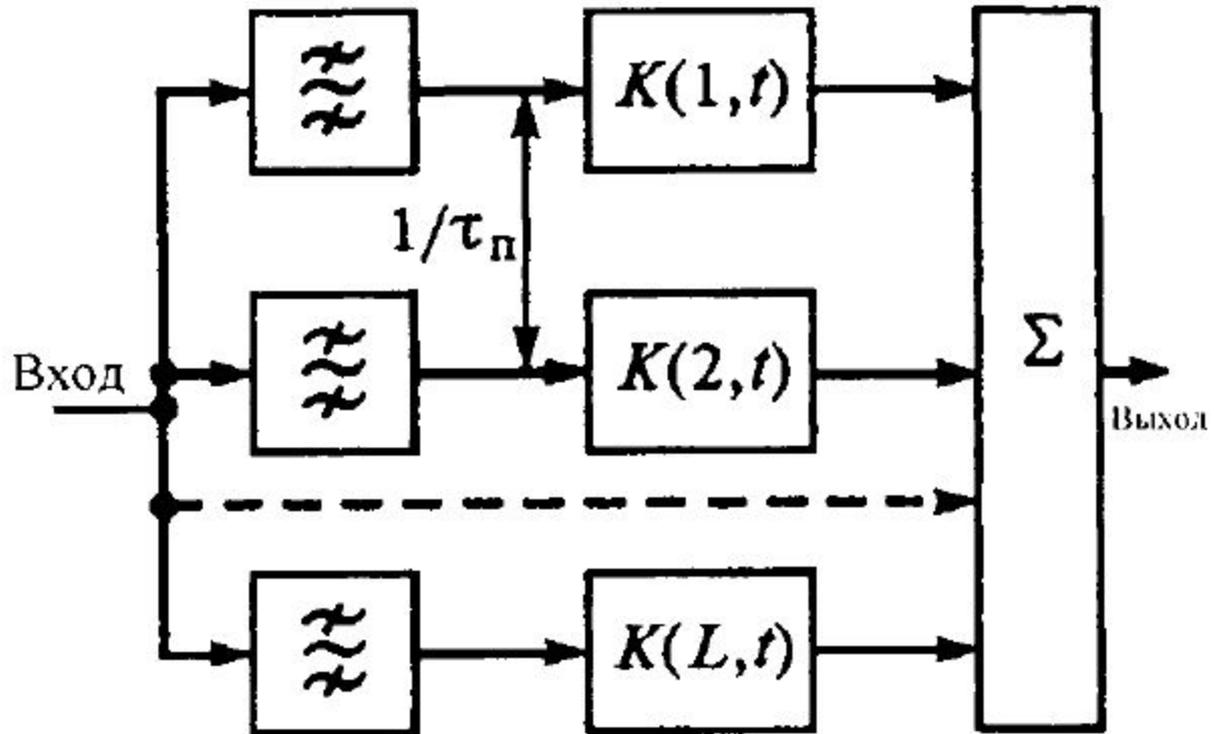


Рис. 4. Модель канала с ограниченным временем рассеяния сигнала

Линейные искажения в непрерывных каналах

Важно знать характер изменения комплексных коэффициентов передачи $h(i, t) = H_D(i, t) + jH_M(i, t) = H(i, t)\exp[j\varphi(i, t)]$ $K(i, t) = K_D(i, t) + jK_M(i, t) = K(i, t)\exp[j\psi(i, t)]$ в каждой ветви. Если рассеивающий объем состоит из большого числа независимых отражателей, то в соответствии с центральной предельной теореме теории вероятности коэффициенты при действительной и мнимой частях будут гауссовскими независимыми случайными величинами с нулевыми математическими ожиданиями и дисперсиями, равными σ_i^2 . Тогда модули $H(i, t), K(i, t)$ и фазы будут $\varphi(i, t), \psi(i, t)$ будут подчиняться соответственно закону Рэлея и равномерному закону. В тех случаях, когда кроме рассеянной составляющей канал имеет и регулярную, модули коэффициентов передачи будут подчиняться обобщенному распределению Рэлея.

Информацию о динамике изменения коэффициента передачи дает корреляционная функция или спектральная плотность мощности флуктуации этого коэффициента.

Линейные искажения в непрерывных каналах

Время корреляции τ_k или ширина спектра флуктуации $F_{\text{фл}}$ характеризуют скорость изменения параметров канала. Например, для КВ канала ширина спектра флуктуации составляет 0,1...1 Гц. В мобильных системах, где диапазон используемых частот много больше, даже в предположении равенства скоростей перемещения отражателей спектра флуктуаций оказывается значительно шире.

В системах передачи дискретной информации рассеяние во времени сигнала приводит также к эффекту межсимвольной интерференции, заключающемуся в наложении следующих друг за другом посылок. Это имеет место, если длительность передаваемых посылок оказывается соизмеримой со временем рассеяния сигнала. Чтобы избежать этого вида искажений в простейшем варианте, приходится снижать скорость передачи в канале.

Теория построения инфокоммуникационных систем и сетей

Тема: Модели дискретно-непрерывных каналов.

Дискретно-непрерывные каналы

Дискретно-непрерывный канал имеет дискретный вход и непрерывный выход. Примером такого канала является канал, образованный совокупностью технических средств между выходом кодера канала и входом демодулятора (рис. 1).

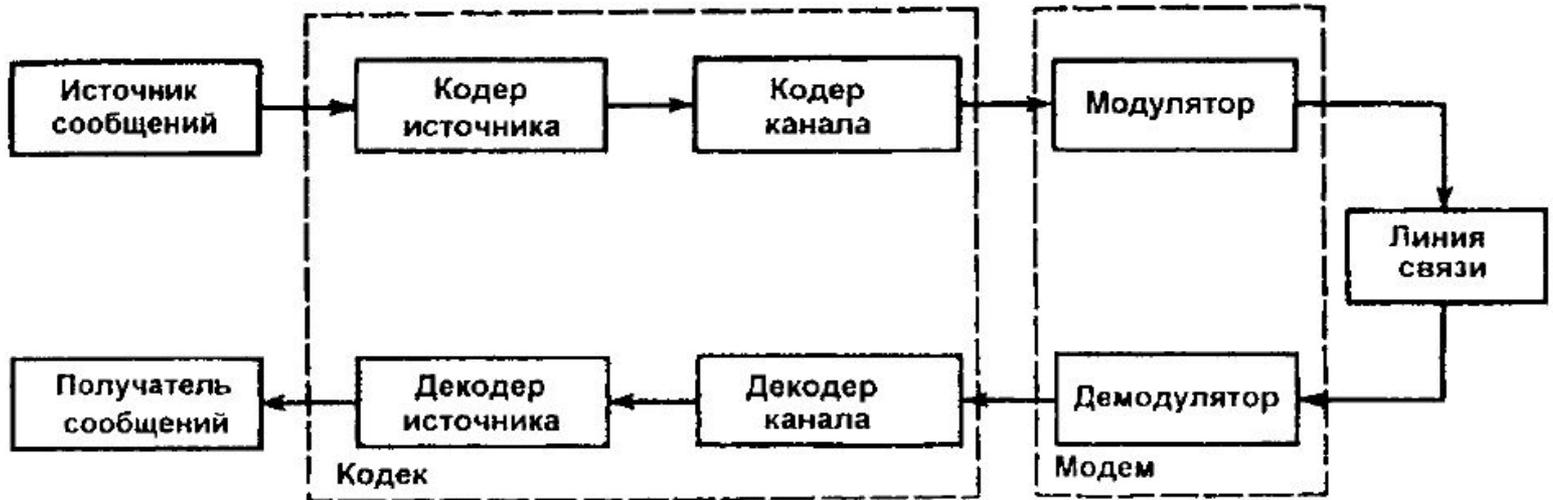


Рис. 1. Структурная схема системы передачи дискретных сообщений

Дискретно-непрерывные каналы

Для его описания необходимо знать алфавит входных символов $\alpha_r, r = 1, \dots, m$, вероятность появления символов алфавита $\rho(\alpha_r), r = 1, \dots, m$, полосу пропускания непрерывного канала F_k , входящего в рассматриваемый канал, и плотности вероятности $w[u|\alpha_r]$ явления сигнала $u(t)$ выходе канала при условии, что передавался символ α_r .

Зная вероятность $\rho(\alpha_r)$ и плотность распределения вероятностей $w[u|\alpha_r]$, можно найти апостериорные вероятности

$$\rho(\alpha_r|u) = \frac{\rho(\alpha_r)w(u|\alpha_r)}{\sum_{r=1}^m \rho(\alpha_r)w(u|\alpha_r)}, \quad r = 1, \dots, m,$$

на основе которых, как правило, и принимается решение о переданном символе.

Дискретно-непрерывные каналы

Ширина спектра сигнала $u(t)$ не может превышать значение F_k . Поэтому в соответствии с теоремой Котельникова его можно представить совокупностью $M = 2F_k T_c$ В, где T_c – длительность сигнала. Соответственно условные плотности вероятности $w(u|\alpha_r), r = 1, \dots, m$, задать как M -мерные плотности вероятности совокупности M отсчетов сигнала

$$u(t)$$

В тех случаях, когда сигнал $u(t)$ является аддитивной смесью полезного сигнала $s_r(t)$ с известными параметрами, несущего информацию о символе α_r , и шума $n(t)$, M -мерная плотность вероятности $w_M(u_1, u_2, \dots, u_M / \alpha_r)$ будет полностью определяться M -мерной плотностью вероятности шума т. е.

$$\begin{aligned} w_M(u_1, u_2, \dots, u_M / \alpha_r) &= w_M[(u_1 - s_1^r), (u_2 - s_2^r), \dots, (u_M - s_M^r)] = \\ &= w_M[n_1, n_2, \dots, n_M], \end{aligned}$$

где u_i, s_i^r и n_i – отсчеты сигналов $u(t), s_r(t)$ и шума $n(t)$ в момент времени t_i .

Дискретно-непрерывные каналы

При независимых отсчетах шума

$$w_M(u_1, u_2, \dots, u_M / \alpha_r) = \prod_{i=1}^M w(n_i).$$

Если плотность вероятности $w(u / \alpha_r)$ для любого сочетания $u(t)$ и α_r не зависит от времени, то канал называется **стационарным**.

Если выполняется условие

$$w_M = (u | X_k^{r_k}, X_{k-1}^{r_{k-1}}, \dots, X_{k-N}^{r_{k-N}}) = w(u | X_k^{r_k}),$$

где $X_{k-N}^{r_{k-N}} \dots X_{k-1}^{r_{k-1}} X_k^{r_k}$ – последовательность передаваемых символов, то такой канал называется **каналом без памяти**.

Реальные каналы являются обычно нестационарными и обладают памятью. Тем не менее модель дискретно-непрерывного стационарного канала с памятью часто применяют благодаря ее простоте.

Теория построения инфокоммуникационных систем и сетей

Тема: Модели непрерывных каналов связи.

Непрерывные каналы

Искажения сигналов и помехи в реальных каналах связи весьма многообразны. Тем не менее математическая модель канала должна по возможности точно описывать основные особенности реального канала и в то же время быть достаточно простой для получения конечных результатов при анализе и синтезе систем передачи. Рассмотрим наиболее простые и часто встречающиеся модели каналов связи.

Идеальный канал без помех вносит детерминированные искажения, связанные с изменением амплитуды и временного положения сигнала. Переданный сигнал может быть полностью восстановлен на приемной стороне в новом временном отсчете. Эта модель используется для описания каналов с закрытым распространением малой протяженности (кабель, провод, волновод, световод и т. д.).

Непрерывные каналы

Канал с гауссовским белым шумом представляет собой идеальный канал, в котором на сигнал накладывается помеха

$$u(t) = \mu s(t - \tau) + n(t).$$

Коэффициент передачи μ и запаздывание τ постоянны и известны в точке приема. Такая модель, например, соответствует радиоканалам, работающим в пределах прямой видимости.

Гауссовский канал с неопределенной фазой сигнала отличается от предыдущего тем, что фаза несущего колебания в точке приема предполагается случайной с плотностью распределения $w(\phi)$ в интервале $-\pi \leq \phi \leq \pi$. Эта неопределенность вызвана двумя причинами: отсутствием устройств оценки и предсказания фазы либо ошибками в оценке фазы при их работе. Важно знать скорость флуктуации фазы. В дискретных системах различают каналы с быстрыми флуктуациями, когда интервал их корреляции меньше длительности посылки, и с медленными, когда это условие не выполняется. При медленных флуктуациях фазы несущего колебания за длительность посылки практически не изменяется.

Непрерывные каналы

Гауссовский канал с неопределенной амплитудой и фазой сигнала вносит в сигнал наряду с флуктуациями фазы и флуктуации амплитуды, которые связаны с изменением во времени по случайному закону коэффициента передачи μ . Как и в предыдущем случае, флуктуации могут быть быстрыми и медленными. Для определения модели канала необходимо задать плотность распределения $w(\mu)$ и корреляционную функцию флуктуации

В гауссовском канале с линейными искажениями форма сигнала изменяется из-за наличия избирательных цепей. В общем случае линейные искажения носят случайный характер. Частотная характеристика канала неравномерна в полосе частот сигнала и изменяется во времени $K(j\omega, t)$. Импульсная характеристика имеет F_c длительность (время памяти канала), превышающую величину $h(t, \tau)$. Такая модель предложена при анализе систем, использующих, например, $1/F_c$ ты с рассеянием сигнала. Сигнал на выходе канала с линейными искажениями

Непрерывные каналы

$$u(t) = \int_0^t h(t, \tau) s(t - \tau) d\tau + n(t).$$

В радиосистемах передачи дискретной информации, когда время памяти канала τ_{Π} соизмеримо с длительностью посылки T_c (а тем более превышает ее), имеет место **межсимвольная интерференция** (МСИ), которая проявляется в наложении друг на друга соседних посылок. Одной из причин возникновения МСИ является увеличение скорости передачи при ограниченной полосе пропускания канала.

В **гауссовском канале с нелинейными искажениями сигнала**, как и в предыдущем случае, аддитивная помеха предполагается в виде гауссовского белого шума, однако смесь сигнала и помехи, проходя по каналу, претерпевает нелинейные искажения так, что на входе приемника $u(t) = F[s(t) + n(t)]$, где $F[\cdot]$ – амплитудная характеристика нелинейного звена канала.

Непрерывные каналы

Возможно дальнейшее усложнение модели с нелинейными искажениями, если предположить наличие в канале еще и линейных искажений, вызванных частотно-избирательными звеньями системы.

Линейный канал со сложной аддитивной помехой характеризуется тем, что на сигнал могут действовать помехи любого вида: сосредоточенные по спектру, по времени, гауссовские, негауссовские и т. д. Модель помех можно определить, указав способ вычисления многомерной плотности распределения вероятностей. Эта модель наиболее полно отображает реальный шум в каналах связи, однако редко используется из-за сложности. Наиболее просто задать модель сложных аддитивных помех в виде небелого гауссовского шума с изменяющейся во времени и по частоте спектральной плотностью $N(f, t)$, характеризуемой как случайный процесс плотностью распределения $w(N)$ и корреляционными функциями во временной $R_N(\tau)$ и частотной $R_N(\nu)$ областях.

Теория построения инфокоммуникационных систем и сетей

Тема: Нелинейные искажения в каналах связи.

Нелинейные искажения в каналах связи

Нелинейные искажения возникают в результате прохождения сигнала по звеньям с нелинейной амплитудной характеристикой $F(u)$. Так как среда распространения, как правило, линейна, то нелинейные искажения определяются техническими устройствами, входящими в канал связи. Часто они возникают в ретрансляторах радиорелейных линий, в которых для получения максимальной мощности излучения передатчика умышлено переводят в режим работы с ограничением сигнала. Это имеет место, например, в спутниковых ретрансляторах.

Для узкополосных радиосигналов

$$s(t) = A(t)\cos[\omega_0 t + \varphi(t)] = A(t)\cos[\theta(t)]$$

Сигнал на выходе нелинейного звена является периодической функцией θ и может быть представлен в виде ряда Фурье от аргумента θ :

$$S_{\text{ВЫХ}}(t) = g_0(A) + g_1(A) \cos(\theta) + g_2(A) \cos(2\theta) + \dots .$$

Нелинейные искажения в каналах связи

Так как приемное устройство обычно содержит на входе полосовой фильтр, пропускающий только спектральные составляющие в области несущей частоты ω_0 , то составляющая сигнала в полосе пропускания такого фильтра будет равна

$$S_{\omega_0}(t) = g_1(A) \cos[\omega_0 t + \varphi(t)],$$

где $g_1(A) = \frac{1}{\pi} \int_0^{2\pi} F\{A \cos[\theta(t)]\} \cos(\theta) d\theta$ – преобразование Чебышева первого

порядка характеристики $F(u)$, которое определяет огибающую выходного сигнала в основной полосе частот.

Таким образом, нелинейные искажения сигнала сводятся к появлению новых спектральных составляющих на частотах $n\omega_0$ ($n = 0, 2, 3, \dots$) и изменению огибающей $A(t)$. Моменты перехода через нуль сигнала с частотой ω_0 не изменяют своего положения на оси времени.

Нелинейные искажения в каналах связи

Картина искажения сигнала существенно усложняется, когда одновременно с полезным сигналом $s(t)$ действуют другие сигналы или помехи. В этом случае на сигнал воздействует еще и комбинационные составляющие, обусловленные взаимодействием сигнала и помех на нелинейный элемент. Это приводит к потере мощности полезного сигнала и к дополнительным помехам. Подавление полезного сигнала на нелинейности, которое обычно оценивают уменьшением отношения сигнал/шум в децибелах, зависит от формы кривой $F(u)$ и вида помеховых сигналов.

Нелинейные искажения в каналах связи

Особый интерес как нелинейность представляет так называемый предельный ограничитель, для которого $F(u) = \text{sign}(u)$. Пусть на его входе действует два сигнала с разными амплитудами (рис. 1), один из которых полезный, а другой мешающий. На выходе ограничителя будем иметь либо только полезный сигнал, либо только мешающий, в зависимости от соотношения амплитуды. Таким образом, сильный сигнал полностью подавляет слабый сигнал. При других формах сигнала и помехи степень подавления имеет конечное значение. Например, если входной полезный сигнал представляет собой узкополосный радиосигнал, то при любом виде модуляции степень подавления его сильным синусоидальным мешающим сигналом составляет около 6 дБ.

Нелинейные искажения в каналах связи

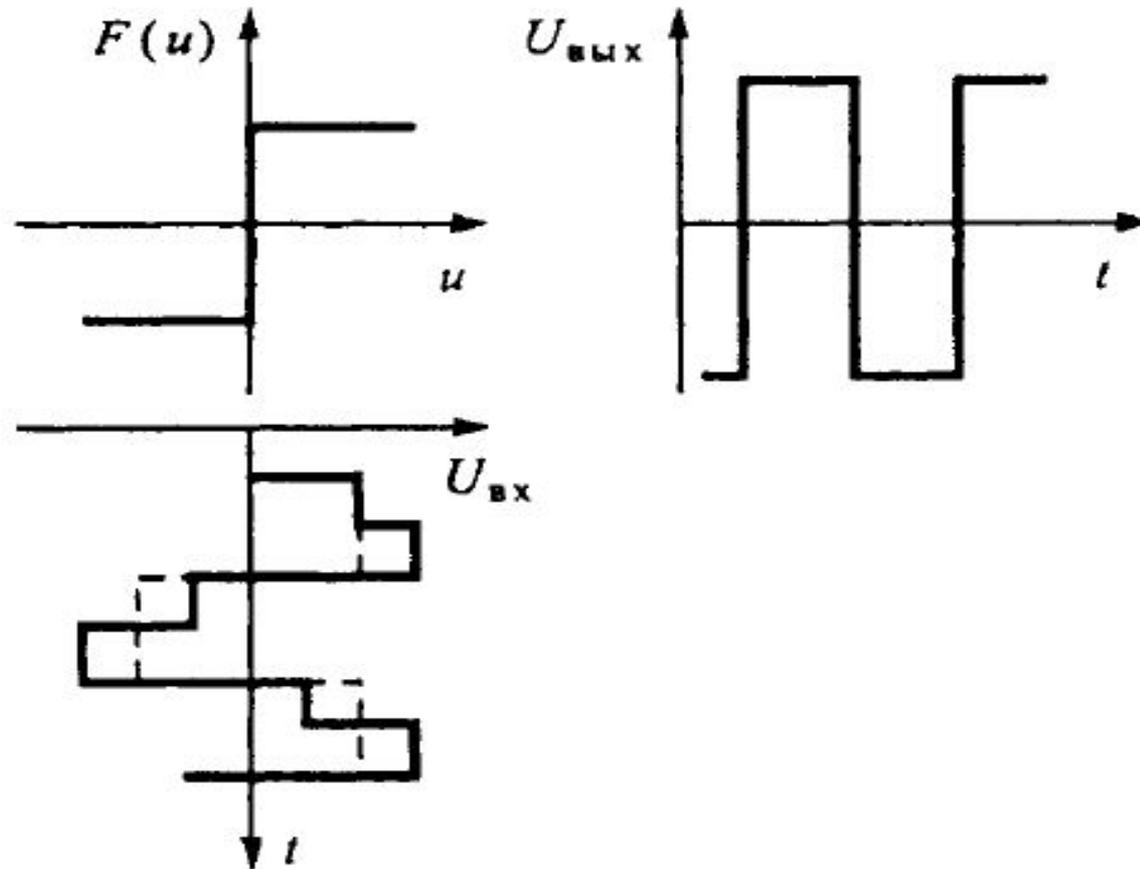


Рис. 1. Диаграмма подавления слабого сигнала сильным на нелинейном элементе (штриховой линией обозначен сильный сигнал)

Нелинейные искажения в каналах связи

Для помехи, представляющей собой сумму гармонического сигнала и гауссовской помехи, коэффициент подавления полезного сигнала $K_{\text{под}}$ можно рассчитать по формуле

$$K_{\text{под}} = \frac{\pi}{4} (1 + \alpha) \left[\exp\left(-\frac{\alpha}{2}\right) I_0\left(\frac{\alpha}{2}\right) \right]^2,$$

где α - отношение мощности синусоидальной составляющей помехи к флуктуационной; I_0 - модифицированная функция Бесселя первого рода нулевого порядка от аргумента $\alpha/2$.

Нелинейные искажения в каналах связи

Из рис. 2 видно, что в предельном ограничитель подавления гармонического сигнала будет наибольшим при воздействии гармонической помехи и наименьшим при воздействии гауссовской помехи.

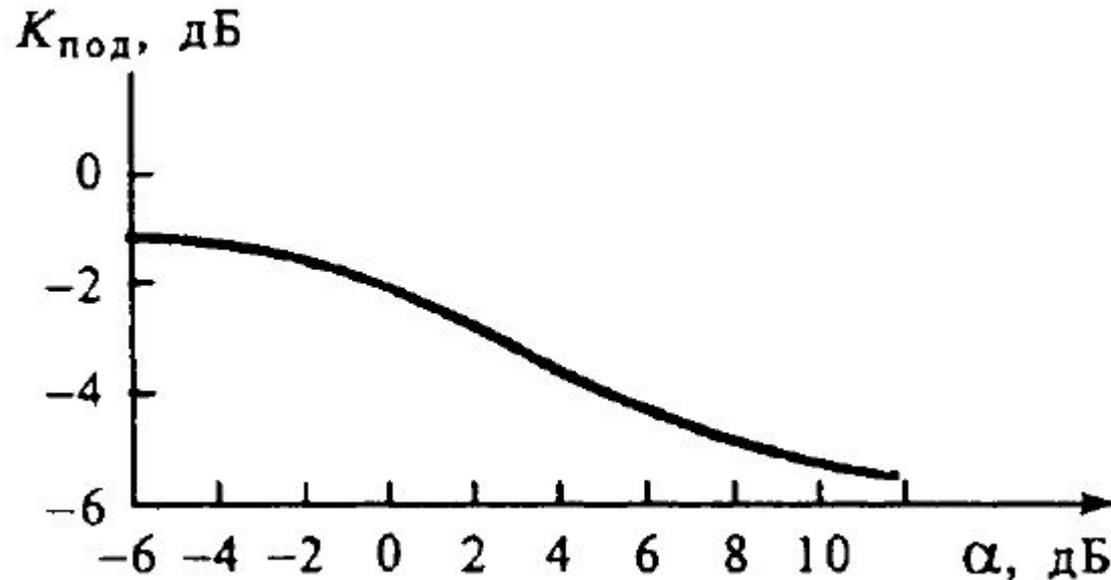


Рис. 2. Зависимость коэффициента подавления узкополосного радиосигнала суммой синусоидальной и гауссовской помех от отношения мощностей этих помех

Теория построения инфокоммуникационных систем и сетей

Тема: Неравномерное квантование.

Неравномерное квантование

Неравномерное квантование, хотя и сложнее в реализации, чем равномерное, довольно часто используется при передаче речевых сигналов. Это объясняется следующими причинами. Одна из них заключается в том, что распределение мгновенных значений речевых сигналов отлично от равномерного; как правило, малые значения гораздо более вероятны, чем большие. Поэтому при равномерном квантовании вероятности попадания сигнала в разные интервалы квантования различны. Соответственно неодинаковым является вклад интервалов квантования в общую погрешность квантования. Очевидно, что погрешность квантования можно уменьшить, если шаг квантования брать меньшим для более вероятных значений сообщения и большим для менее вероятных.

Вторая причина заключается в том, что в телефонных системах различие в средних значениях речевых сигналов может достигать 30 дБ и более. Чтобы сохранить разборчивость речи «тихого» абонента, шаг квантования в области малых значений сигнала должен быть небольшим. В области больших значений сигнала можно допустить более крупный шаг. Таким образом, вновь приходим к неравномерному квантованию.

Неравномерное квантование

Неравномерное квантование можно реализовать различными способами, например квантованием с соответствующей амплитудной характеристикой (непосредственное неравномерное квантование). При этом длины интервалов и уровни квантования обычно выбираются из условия получения минимальной дисперсии погрешности $D_{\text{кв}}$. Такой же эффект можно получить путем сжатия (компрессирования) динамического диапазона сигнала, применения равномерного квантования и последующего расширения (экспандирования) после восстановления отсчетов на приемной стороне (рис. 1). Характеристики компрессора и экспандера должны быть взаимно обратными. Этот метод получил название **квантование с компандированием сигнала**. Характеристику компрессора выбирают из условия обеспечения минимума дисперсии погрешности квантования.

Неравномерное квантование

Неравномерное квантование можно реализовать различными способами, например квантованием с соответствующей амплитудной характеристикой (непосредственное неравномерное квантование). При этом длины интервалов и уровни квантования обычно выбираются из условия получения минимальной дисперсии погрешности $D_{\text{кв}}$. Такой же эффект можно получить путем сжатия (компрессирования) динамического диапазона сигнала, применения равномерного квантования и последующего расширения (экспандирования) после восстановления отсчетов на приемной стороне (рис. 1). Характеристики компрессора и экспандера должны быть взаимно обратными. Этот метод получил название **квантование с компандированием сигнала**. Характеристику компрессора выбирают из условия обеспечения минимума дисперсии погрешности квантования.

Неравномерное квантование

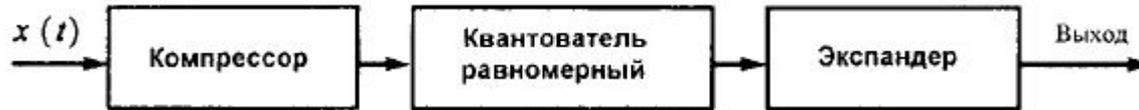


Рис. 1. Структурная схема компандирования

При неравномерном квантовании дисперсия погрешности квантования

$$D_{\text{КВ}} = \sum_{i=1}^{L_{\text{КВ}}} \int_{x^{(i)}}^{x^{(i+1)}} (x - x_{\text{КВ}}^{(i)})^2 w(x) dx, \quad (1)$$

где $x^{(i)}$ и $x^{(i+1)}$ – нижняя и верхняя границы i -го интервала квантования, называемые обычно **порогами квантования**.

Пороги и уровни квантования выбирают из условия минимизации дисперсии (1). Для их нахождения продифференцируем выражение (1) по переменным $x_{\text{КВ}}^{(i)}$ и $x^{(i)}$ равняем производные к нулю. В результате получаем:

Неравномерное квантование

$$\int_{x^{(i)}}^{x^{(i+1)}} (x - x_{\text{KB}}^{(i)}) w(x) dx = 0, \quad (2)$$

$$\left[x^{(i)} - x_{\text{KB}}^{(i-1)} \right]^2 = \left[x^{(i)} - x_{\text{KB}}^{(i)} \right]^2. \quad (3)$$

Из выражений (2) и (3) находим, что оптимальным значение $x_{\text{KB}}^{(i)}$ является абсцисса центра тяжести криволинейной трапеции (рис. 2) по кривой $w(x)$ и основанием $x^{(i)} x^{(i+1)}$, а порог квантования $x^{(i)}$ равен $(x_{\text{KB}}^{(i)} + x_{\text{KB}}^{(i-1)})/2$.

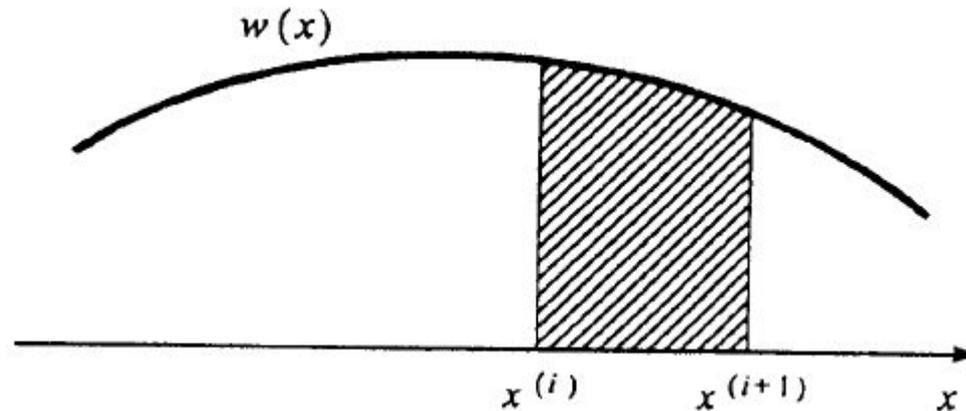


Рис. 2. Диаграмма, иллюстрирующая выбор уровня квантования

Неравномерное квантование

В частности, из (2) и (3) нетрудно видеть, что если распределение $x(t)$ равномерное, то квантование с постоянным шагом оптимальное.

Квантованные отсчеты можно передавать различными способами. На практике для этого чаще всего используют кодовые комбинации, каждая из которых соответствует определенному уровню квантованию. При равномерном коде с основанием m длина кодовых комбинаций не может быть меньше k , где k выбирается из условия $L_{\text{кв}} \leq m^k$.

При выборе основания кода в первую очередь необходимо учитывать простоту, экономичность и удобство реализации цифрового представления непрерывных сообщений. На практике обычно применяют простые (безызбыточные) двоичные коды, среди которых наибольшее применение нашли двоичный натуральный код, симметричный двоично-числовой код и код Грея.

Неравномерное квантование

Двоичный натуральный код – это код, комбинации которого представляют собой двоичные номера уровней квантования. Он прост в реализации и удобен при обработке на ЭВМ.

Симметричный двоично-числовой код используется для предоставления биполярных квантованных отсчетов. При этом высший разряд несет информацию о знаке отсчета, а остальные разряды – об абсолютном значении отсчета в натуральном двоичном коде.

Неравномерное квантование

Код Грея связан с двоичным натуральным кодом следующими соотношениями: $a_0^{\Gamma} = a_0 \oplus a_1$; $a_1^{\Gamma} = a_1 \oplus a_2$; ...; $a_{k-2}^{\Gamma} = a_{k-2} \oplus a_{k-1}$; $a_{k-1}^{\Gamma} = a_{k-1}$, где $a_{k-1} a_{k-2} \dots a_0$ – кодовая комбинация натурального кода, $a_{k-1}^{\Gamma} a_{k-2}^{\Gamma} \dots a_0^{\Gamma}$ – кодовая комбинация кода Грея, символ \oplus означает суммирование по модулю 2. Этот код обладает следующими двумя особенностями, которые способствуют повышению быстродействия кодирующих устройств по сравнению с применением двоичного натурального кода: любые две кодовые комбинации, соответствующие соседним уровням квантования, отличаются друг от друга только в одном разряде; смена значений элементов в каждом разряде при переходе от одной комбинации к другой происходит вдвое реже, чем в двоичной натуральном коде.

Неравномерное квантование

Рассмотренные коды обеспечивают одинаковую погрешность восстановления из-за ошибок в канале связи при условии, что ошибки возникают независимо от передаваемого сигнала и соседние ошибки независимы.

Кроме простых двоичных кодов, при передаче непрерывных сообщений используются помехоустойчивые коды, позволяющие обнаруживать и исправлять ошибки, возникшие из-за действий помех в канале связи.

Теория построения инфокоммуникационных систем и сетей

Тема: Общие сведения о каналах связи.

Общие сведения о каналах связи

Каналы связи можно классифицировать по различным показателям.

В теории передачи сигналов каналы классифицируются по характеру сигналов на входе и выходе. Различают **непрерывные, дискретные и дискретно-непрерывные** каналы. В непрерывных каналах сигналы на входе и выходе непрерывны по уровням, в дискретных – они соответственно дискретны, а в дискретно-непрерывных – сигналы на входе дискретны, а на выходе непрерывны, и наоборот.

Возможно также классификация каналов по назначению РСПИ (телеграфные, телефонные, телевизионные, телеметрические и др.), по виду физической среды распространения (проводные, кабельные, волноводные и др.) и по диапазону используемых ими частот. К радиодиапазону относят частоты в пределах $30 \dots 30 \cdot 10^{12}$ Гц, что соответствует длинам волн от 10^8 м до 0,1 мм. Кроме радиодиапазона, в настоящее время широкое распространение нашел и оптический диапазон волн. В силу дискретного характера электромагнитного излучения в оптическом диапазоне волн такие волн принято называть **квантовыми**.

Общие сведения о каналах связи

Таблица 1. Классификация каналов по диапазонам используемых ими частот

Диапазон частот	Диапазон волн	Название частот	Название волн
30...300 Гц	1000... 10000 км	Сверхнизкие (СНЧ)	–
300...3000 Гц	100... 1000 км	Инфранизкие (ИНЧ)	–
3...30 кГц	10... 100 км	Очень низкие (ОНЧ)	Мириаметровые
30...300 кГц	1... 10 км	Низкие (НЧ)	Километровые
300...3000 кГц	100... 1000 м	Средние (НЧ)	Гектометровые
3...30 МГц	10... 100 м	Высокие (ВЧ)	Декаметровые
30...300 МГц	1... 10 м	Очень высокие (ОВЧ)	Метровые
300...3000 МГц	10... 100 см	Ультравысокие (УВЧ)	Дециметровые
3...30 ГГц	1... 10 см	Сверхвысокие (СВЧ)	Сантиметровые
30...300 ГГц	1... 10 мм	Крайне высокие (КВЧ)	Миллиметровые
300...3000 ГГц	0,1... 1 мм	Гипервысокие (ГВЧ)	Децимиллиметровые

Общие сведения о каналах связи

По способу распространения радиоволн различают каналы с **открытым** и с **закрытым** распространением. В каналах с закрытым распространением электромагнитная энергия распространяется по направляющим линиям (кабельные, проводные, волноводные СВЧ тракта и др.). Для них характерны малый уровень помех и постоянство параметров сигнала, что позволяет передавать информацию с высокой скоростью и достоверностью.

Рассмотрим кратко особенности использования радиоволн различных диапазонов в каналах с открытым распространением. В диапазонах ИНЧ, ОНЧ, НЧ на небольших расстояниях поле в месте приема создается за счет дифракционного огибания волнами выпуклой поверхности Земли. На больших расстояниях радиоволны распространяются в своеобразном сферическом волноводе, внутренняя стенка которой образуется поверхностью Земли, а внешняя – ионосферой. Такой механизм распространения позволяет принимать сигналы в любой точке Земли, причем параметры принятых сигналов отличаются достаточно высокой

Общие сведения о каналах связи

стабильностью. Особенностью этих диапазонов является также способность волн проникать в толщу Земли и воды на глубину в десятки метров. Принципиальным недостатком таких каналов является: ограниченная полоса частот (единицы герц) и очень большие линейные размеры антенных устройств, соизмеримых с длиной волны, составляющей километры. Сверхдлинные волны применяются для навигации и передачи ограниченного объема информации на подводные объекты.

В распространении волн диапазона ВЧ принимает участие ионосфера. Однако если волны длиннее 1 км отражаются от нижнего ее слоя практически зеркально, то декаметровые волны достаточно глубоко проникают в ионосферу, что приводит к эффекту **многолучевости**, когда в точку приема приходит одновременно несколько сигналов с различными временами запаздывания. Многолучевость может носить **дисперсный** или дискретный характер. Дисперсия (рассеяние) сигнала определяется отражением радиоволн от некоторого объема ионосферы, а дискретная многолучевость – отражением от разных слоев ионосферы. Так как глубина

Общие сведения о каналах связи

проникновения в ионосферу зависит от длины волны, то для передачи информации между двумя пунктами можно указать оптимальную рабочую частоту (ОРЧ), на которой связь будет наиболее надежной (максимум мощности принимаемого сигнала, минимум эффекта многолучевости). Значение ОРЧ рассчитывают для определенных трасс и времени связи. Для этого составляют долговременные и кратковременные прогнозы по данным мировой сети станций ионосферного зондирования. Декаметровые волны широко применяются для глобальной связи и радиовещания. С их помощью можно передавать информацию сравнительно большого объема в пределах всего земного шара при ограниченной мощности передатчика и небольших по размеру антеннах. Полоса частот передаваемых сигналов в декаметровом канале не превышает десяти кГц. До появления спутниковых систем связи этот диапазон был единственным пригодным для организации связи между двумя любыми пунктами на Земле без промежуточной ретрансляции. Однако эффект глобального распространения коротких волн имеет и свою отрицательную сторону – в точке приема могут появиться сильные помехи от

Общие сведения о каналах связи

дальних радионавигации.

Гектометровые волны днем распространяются как земные, а ночью – как ионосферные. Дальность распространения земной волны над сушей не превышает 500 км, а над морем – 1000 км. Диапазон СЧ широко используется в радиовещании, связи и радионавигации.

Волны диапазона частот от 30 МГц и выше слабо дифрагируют и поэтому распространяются в пределах прямой видимости. Расстояние прямой видимости (радиогоризонта) по поверхности Земли в километрах примерно равно $5\sqrt{A+B}$, где A и B – высота передающей и приемной антенны в метрах. Если предположить, что $A = 30^a$ $B = 3^b$ радиогоризонт равен 30 км и для всех радиоволн, имеющих путь распространения менее 30 км, потери сигнала не будут связаны с кривизной поверхности Земли. Для всех остальных лучей возникает дополнительное затухание, обусловленное экранированием сигнала земной поверхностью. Однако резкое снижение уровня сигнала на расстояниях, превышающих прямую видимость, имеет и положительную сторону, связанную со снижением уровней помеховых

Общие сведения о каналах связи

сигналов и возможностью использования одинаковых частот для организации связи в зонах, удаленных на расстояния, превышающие радиогоризонт. Некоторого увеличения дальности можно достичь, применив поднятые антенны, а для организации связи на расстояния, превышающие прямую видимость, ретрансляция сигналов. Системы с ретрансляцией сигналов называются **радиорелейными линиями**. Одним из основных достоинств высокочастотных диапазонов является большой частотный ресурс, что позволяет создавать радиосистемы передачи информации с высокой скоростью передачи и радиосети с большим числом одновременно работающих радиостанций. В последние 30 лет диапазон ОВЧ и СВЧ нашел широкое использование для создания сетей мобильной связи.

Общие сведения о каналах связи

Стремление увеличить дальность радиолинии в этих диапазонах без промежуточной ретрансляции нашел свое решение в РСПИ, использующих рассеяние радиоволн на неоднородностях тропосферы, ионосферы и метеорных следах. Однако такие системы по качеству передачи информации не могут конкурировать с радиорелейными линиями того же диапазона, поэтому их имеет смысл применять тогда, когда ретрансляция сигналов по тем или иным причинам невозможна.

Стремление увеличить ширину полосы частот канала, а также повысить пространственную селекцию сигналов за счет использования остронаправленных антенн при их ограниченных размерах привело к освоению диапазона миллиметровых волн. Главной особенностью их с точки зрения распространения является сильное поглощение в дожде и тумане, что ограничивает их применение в наземных системах большой дальности. Однако в космических и спутниковых системах они весьма перспективны.

Общие сведения о каналах связи

Новую эру в освоении высокочастотной области радиодиапазона для средств связи открыл запуск искусственных спутников Земли (ИСЗ). Обычно ИСЗ находятся на высоте от 500 до 40000 км от поверхности Земли и поэтому обеспечивают радиосвязь между земными станциями, удаленными на расстоянии до 10...17 тыс. км. Линия спутниковой связи состоит из двух конечных земных станций и одного или нескольких спутников-ретрансляторов, обращающихся вокруг Земли по заданным орбитам.

Из всего многообразия орбит ИСЗ особый интерес представляют экваториальная круговая орбита, удаленная от поверхности Земли на расстояние около 36000 км (стационарная экваториальная орбита). Когда направление движения ИСЗ по такой орбите совпадает с направлением вращения Земли, спутник будет казаться наземному наблюдателю неподвижным (**стационарным** спутником). При использовании трех стационарных спутников, расположенных в экваториальной плоскости через 120° по дуге, принципиально оказывается возможным организовать глобальную систему связи. Максимальный от горизонта до горизонта обзор

Общие сведения о каналах связи

земной поверхности от одного ИСЗ или, иначе говоря, максимальное расстояние вдоль поверхности Земли между двумя станциями будет практически составлять 15...17 тыс. км. Существенные преимущества стационарной орбиты заключается в снижении требования к системам слежения за спутником, сведении к минимуму доплеровских сдвигов частот сигналов, что упрощает приемное устройство при большом обзоре поверхности Земли. Недостатком стационарной орбиты является плохой охват приполярных зон. Поэтому в России для систем связи широко применяются сильно вытянутые эллиптические орбиты с большой полуосью до 5 земных радиусов с эксцентриситетом 0,8...0,9 и углом наклона примерно 65° . Три спутника, выведенные через равномерные интервалы времени на аналогичные эллиптические орбиты, восходящие узлы которых смещены относительно друг друга на 120° , могут обеспечить круглосуточную непрерывную связь между земными станциями, расположенными в северном полушарии Земли, на глобальные расстояния. Другая возможность в создании глобальных РСПИ заключается в использовании созвездий

Общие сведения о каналах связи

спутников, орбиты которых выбраны так, что над любой точкой на поверхности Земли находится по крайней мере один спутник.

Выбор рабочих частот для линии радиосвязи через ИСЗ определяется следующими факторами: условиями распространения и поглощения радиоволн, уровнем внешних помех, принимаемых антенной, техническими средствами (коэффициент шума приемного устройства, ширина лепестка диаграммы направленности антенны, точность ориентации и т.п.), взаимными помехами между системами связи через ИСЗ и другими службами, работающими в смежных или совмещенных диапазонах частот. Ограничение диапазона частот снизу определяется экранирующим действием ионосферы, а сверху – поглощением в тропосфере. Эти два фактора предопределили диапазон рабочих частот 40 МГц...40 ГГц. В настоящее время наибольшее использование находит диапазон 1...12 ГГц.

Теория построения инфокоммуникационных систем и сетей

Тема: Помехи в каналах связи.

Помехи в каналах связи

Ошибки, возникающие при приеме сообщения, в значительной степени определяются видом и интенсивностью помех, действующих в канале. В зависимости от места нахождения источника помех различают **внутренние** и **внешние** помехи. Внутренние помехи возникают в самой системе. К ним относятся шумы входных каскадов приемника, приемной антенны, линии канализации сигнала и электрические сигналы, попадающие в приемник по внутренним цепям вследствие плохого экранирования или развязки между каскадами. Последний вид помех связан с ошибкой в конструкции и по возможности должен быть устранен. Внутренний шум, обусловленный хаотическим движением носителей зарядов, принципиально неустраним, хотя может быть в значительной степени ослаблен применением качественных узлов и деталей, а также снижением рабочей температуры.

Помехи в каналах связи

Различают **тепловой** и **дробовый** шумы. Тепловой шум обусловлен тепловым движением носителей заряда, приводящим к появлению случайной разности потенциалов. Он представляет собой гауссовский случайный процесс с нулевым средним и спектральной плотностью мощности

$$N_0(f) = \frac{hf}{\exp(hf/kT^0) - 1},$$

где $h = 6,6 \cdot 10^{-34}$ Дж·с – постоянная Планка, $k = 1,38 \cdot 10^{-23}$ Дж/град – постоянная Больцмана, T^0 – абсолютная температура источника шума, f – частота.

В диапазоне частот, в котором работают радиосистемы, выполняется условие $hf \ll kT^0$, и поэтому $N_0(f) = kT^0 = N_0$, Вт/Гц.

Таким образом, тепловой шум можно рассматривать как белый с односторонней спектральной плотностью $N_0 = kT^0$.

В реальных системах полоса частот пропускания ограничена и мощность шума $P_{ш} = N_0 F$.

Помехи в каналах связи

Шумы электровакуумных и полупроводниковых приборов (дробовые шумы) обусловлены дискретной природой носителей заряда. Статистические характеристики дробового шума такие же, как у теплового.

Внешние помехи возникают из-за различных электромагнитных процессов, происходящих в атмосфере, ионосфере, космическом пространстве, а также излучением земной поверхности (**естественные помехи**). Кроме того, они создаются различными радиостанциями (**станционные помехи**), промышленными установками, медицинской аппаратурой, электрическими двигателями и т. п. В зависимости от диапазона частот и условий, в которых работает СПИ, преобладает тот или иной вид помех.

Помехи в каналах связи

Атмосферные помехи возникают в результате различных электрических процессов, происходящих в земной атмосфере. Наиболее мощным источником является электрические грозовые разряды, которые приводят к излучению электромагнитной энергии практически во всем радиочастотном диапазоне. Максимум излучения разряда приходится на полосу частот 5...30 кГц. Интенсивность поля помех, создаваемых электрическими разрядами, в пределах прямой видимости с увеличением частоты уменьшается примерно обратно пропорционально частоте.

Для диапазона частот выше 30 МГц заметными становятся помехи, связанные с источниками, находящимися в пределах нашей Галактики и вне ее (космические шумы). Причиной возникновения этих помех является тепловое излучение межзвездных газов, Солнца, радиозвезд. Большинство известных радиозвезд находятся в пределах нашей Галактики и их излучение во много раз превышает по интенсивности излучение тепловых источников. Интенсивность космических шумов так же, как и внутренних, оценивается шумовой температурой.

Помехи в каналах связи

Земная поверхность, как и всякое нагретое тело, излучает электромагнитные волны. Они могут попадать в антенну по основному или боковым лепесткам диаграммы направленности. Мощность этих шумов в значительной степени определяется положением и формой диаграммы направленности, а также температурой и электрическими характеристиками земной поверхности. По своим статистическим характеристикам они аналогичны тепловому шуму.

Промышленные помехи создаются различными электрооборудованием промышленных предприятий, транспорта, линиями электропередач и другими электроустановками. В большинстве случаев они представляют собой последовательности импульсов с постоянным или переменным периодом следования. Распространение промышленных помех происходит в основном земной волной, однако часто они канализируются линиями связи, электропередач, железнодорожными линиями и т. п. Уровень промышленных помех зависит от места расположения приемника относительно промышленных объектов.

Помехи в каналах связи

Одним из распространенных видов помех являются помехи от посторонних радиостанций. Насыщенность радиосредствами (радиосвязь, радиолокация, радионавигация и т. п.) и, следовательно, загрузка радиодиапазонов таковы, что весьма часто помехи от посторонних радиосредств превышают прочие виды помех. Станционные помехи обусловлены целым рядом причин: нарушение регламента распределения рабочих частот, недостаточной стабильностью генераторов и плохой фильтрацией гармоник сигнала, нелинейными искажениями в канале, ведущими к перекрестным помехам. Снизить уровень станционных помех можно с помощью организационно-технических мероприятий. Это направление в радиоэлектронике последнее время усиленно развивается под названием «Электромагнитная совместимость радиоэлектронных средств».

Помехи в каналах связи

Станционные помехи присутствуют практически во всех диапазонах, и особенно в коротковолновом, где из-за ионосферного распространения радиоволн часто складываются благоприятные условия для прохождения радиоволн от посторонних, очень далеко расположенных передатчиков, работающих на той же частоте. Появление станционных помех в полосе принимаемого сигнала, их уровень и амплитуда является в большинстве случаев случайными процессами. Если число помех, попадающих в полосу сигнала, велико, то в соответствии с центральной предельной теоремой теории вероятностей мгновенные значения результирующего сигнала будут подчиняться гауссовскому закону. В то же время изменение загрузки канала во времени и по частоте приводит к тому, что станционная помеха оказывается нестационарным случайным процессом. Упрощенную физическую модель образования стационарных помех при высокой загрузке канала можно представить в виде последовательно включенных генератора белого шума и фильтра с изменяющейся во времени по случайному закону частотной характеристикой.

Помехи в каналах связи

Спектральную плотность мощности помех $N(f, t)$ (рис. 1) как случайный процесс можно достаточно полно охарактеризовать плотностью вероятности $w_{f,t}(N)$ и корреляционными функциями флуктуации во временной и частотной областях $R_N(\tau)$ и $R_N(\nu)$. Параметрами корреляционных функций являются интервал корреляции во времени τ_k и интервал корреляции по частоте F_{Π} . На практике распространена модель с логнормальным распределением помех в частотной и временной областях.

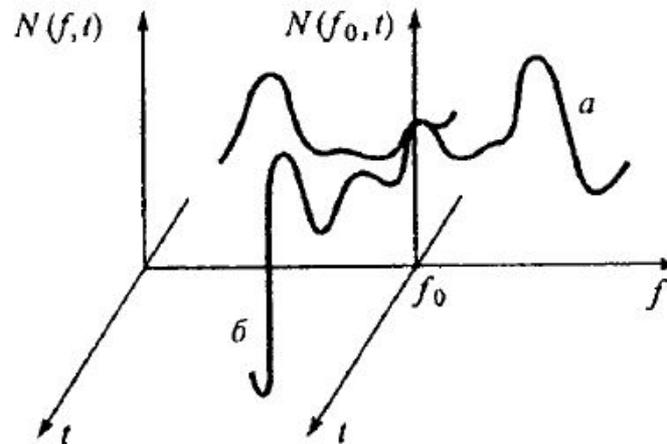


Рис. 1. Изменение спектральной плотности помех по частоте и во времени

Помехи в каналах связи

Если число стационарных помех, попадающих в полосу сигнала, ограничено, то рассмотренная модель не всегда применима. В этом случае поступающую на вход приемника смесь приходится представлять в виде суммы полезного сигнала и ограниченного числа аддитивных помех с известными или неизвестными статистическими характеристиками:

$$u(t) = s(t) + n(t) + \sum_{i=1}^k \gamma_i(t)$$

где $\gamma_i(t) = A_i \cos[\omega_i t + \theta_i(t)]$

Огибающая A_i и фаза $\theta_i(t)$ помехи могут быть как случайными, так и детерминированными процессами.

Теория построения инфокоммуникационных систем и сетей

Тема: Преобразование непрерывных сообщений в цифровую форму. Операция квантования по уровню.

Операция квантования по уровню

Непрерывные сообщения можно передавать по дискретным системам связи. Для этого их преобразуют в цифровую форму (в последовательность символов некоторого алфавита, например двоичного) с помощью операции дискретизации по времени, квантования по уровню и кодирования. В общем случае квантованию могут подвергаться коэффициенты c_i , полученные в результате обобщенного или разностного дискретного преобразования.

Операция квантования по уровню заключается в замене непрерывного множества значений, которое может принимать сообщение $x(t)$, дискретным множеством заранее определенных значений $x_{\text{KB}}^{(i)}, i = 1, \dots, L_{\text{KB}}$, называемых уровнями квантования. Такое преобразование выполняет нелинейное устройство с характеристикой, изображенной на рис. 1, следующим образом. Диапазон возможных значений сообщения разбивается на L_{KB} интервалов. При попадании отсчета сигнала в i -й интервал ему присваивается значение

$$x_{\text{KB}}^{(i)}.$$

Операция квантования по уровню

Различают равномерное и неравномерное квантование. При равномерном квантовании шаг Δx берется постоянным, $x_{\text{кв}}^{(i)}$ уровень соответствует середине i -го интервала квантования. При неравномерном квантовании шаг Δx является переменным.

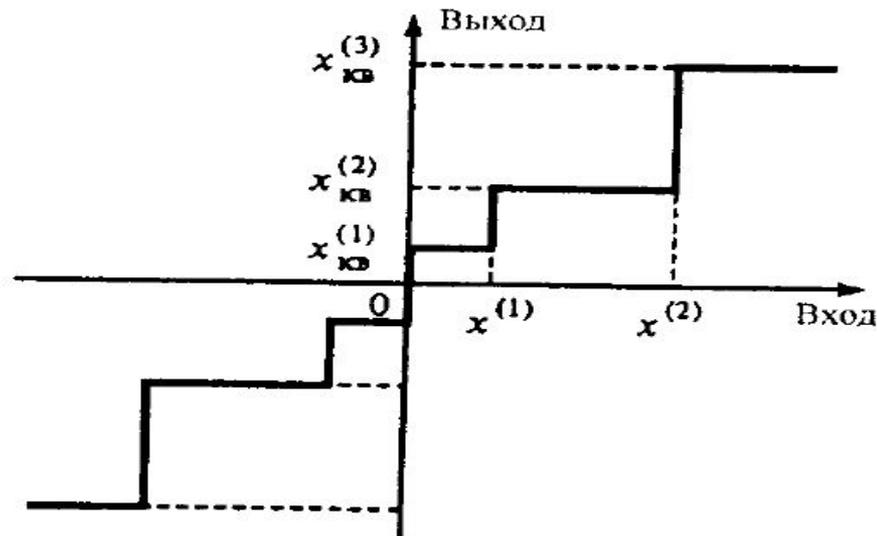


Рис. 1. Характеристика квантователя

Операция квантования по уровню

Замена непрерывного множества возможных значений сообщения дискретным множеством фиксированных значений приводит к погрешности, называемой шумом квантования. При равномерном квантовании дисперсия погрешности квантования определяется как

$$D_{\text{КВ}} = \sum_{i=1}^{L_{\text{КВ}}} \int_{x_{\text{КВ}}^{(i)} - \frac{\Delta x}{2}}^{x_{\text{КВ}}^{(i)} + \frac{\Delta x}{2}} (x - x_{\text{КВ}}^{(i)})^2 w(x) dx, \quad (1)$$

где $w(x)$ – плотность распределения вероятностей мгновенных значений сообщения $x(t)$.

При равномерном распределении значений сообщения из (1) находим, что

$$D_{\text{КВ}} = (\Delta x)^2 / 12. \quad (2)$$

Операция квантования по уровню

Таким образом, для рассматриваемого случая дисперсия погрешности равномерного квантования зависит только от значения шага Δx или при заданном диапазоне изменения значений сообщений – от числа уровней квантования.

Заметим, что при большом числе уровней квантования $D_{\text{КВ}} = (\Delta x)^2 / 12$ при любом законе распределения мгновенных значений сообщения. Действительно, при большем числе уровней квантования плотность вероятности $w(x)$ в пределах любого i -го интервала можно считать постоянной и равной $w(x_{\text{КВ}}^{(i)})$. Тогда

$$D_{\text{КВ}} \approx \sum_{i=1}^{L_{\text{КВ}}} \int_{x_{\text{КВ}}^{(i)} - \frac{\Delta x}{2}}^{x_{\text{КВ}}^{(i)} + \frac{\Delta x}{2}} (x - x_{\text{КВ}}^{(i)})^2 w(x_{\text{КВ}}^{(i)}) dx = \sum_{i=1}^{L_{\text{КВ}}} w(x_{\text{КВ}}^{(i)}) \frac{\Delta x^3}{12}.$$

Учитывая, что $\sum_{i=1}^{L_{\text{КВ}}} w(x_{\text{КВ}}^{(i)}) \Delta x \approx 1$, получаем $D_{\text{КВ}} \approx \Delta x^2 / 12$.

Дискретные каналы

Дискретный канал (ДК) имеет дискретный вход и выход. Пример ДК – совокупность технических средств между выходом кодера канала и выходом демодулятора. Для описания ДК необходимо знать алфавит входных символов $\alpha_r, r=1, \dots, m$, их вероятности появления $p(y_i | \alpha_r)$, скорость передачи символов v , алфавит символов на выходе канала $y_j, j=1, \dots, n$, и значения переходных вероятностей $p(y_j | \alpha_r), l=1, \dots, n, r=1, \dots, m$, появления символа y_j при условии передачи символа α_r .

Первые две характеристики определяются свойствами источника сообщений, скорость v – полосой пропускания непрерывного канала, входящего в состав ДК, объем алфавита выходных символов – алгоритмом работы решающей схемы, переходные вероятности $p(y_j | \alpha_r)$ – характеристиками непрерывного канала.

В ДК объемы алфавитов входных и выходных символов могут не совпадать. Пример – канал со стиранием. Добавочный символ (символ стирания) на выходе канала тогда, когда анализируемый сигнал нельзя с большой вероятностью отождествить ни с одним из передаваемых символов.

Зная вероятности $p(\alpha_r)$ и $p(y_j | \alpha_r), j=1, 2, \dots, n, r=1, \dots, m$, можно вычислить апостериорные вероятности

$$p(\alpha_r | y_j) = \frac{p(\alpha_r)p(y_j | \alpha_r)}{\sum_{r=1}^m p(\alpha_r)p(y_j | \alpha_r)}, \quad r=1, \dots, m, \quad j=1, \dots, n,$$

ДК называется стационарным, если переходные вероятности $p(y_j | \alpha_r)$, $j = 1, 2, \dots, n$; $r = 1, \dots, m$, не зависят от времени. ДК называется без памяти, если переходные вероятности $p(y_j | \alpha_r)$, $j = 1, 2, \dots, n$; $r = 1, \dots, m$, не зависят от того, какие символы передавались и принимались ранее.

Если в стационарном ДК алфавиты на входе и выходе совпадают и

$$p(\alpha_r | y_j) = \begin{cases} p_{ош} & \neq, \\ 1 - (для j) p_{ош} & =, \end{cases}$$

То такой канал называется симметричным.

Наиболее простой является модель стационарного симметричного канала без памяти. В таком канале ошибки возникают независимо друг от друга. Вероятность ошибки $p_{ош}$ при передаче любого символа одинакова и не меняется во времени. Распределение ошибок в нем подчиняется биномиальному закону

$$P_{ош}(r) = C_n^r p_{ош}^r (1 - p_{ош})^{n-r}, \quad (1)$$

Где n – число символов в блоке, r – число ошибочных символов.

Вероятность появления в блоке l и более ошибок

$$P_{ош}(r \geq l) = \sum_{r=l}^n C_n^r p_{ош}^r (1 - p_{ош})^{n-r}$$

Большинство реальных каналов имеют «память», которая проявляется в том, что вероятность ошибки в символе зависит от того, какие символы передавались до него и как они были приняты. Первый факт обусловлен межсимвольными искажениями, являющимся результатом рассеяния сигнала в канале, а второй – изменением отношения сигнал/шум в канале или характера помех.

Одним из распространенных методов описания дискретного канала с памятью, связанной с межсимвольными искажениями, является использование аппарата цепей Маркова (посимвольное описание). В этом случае последовательность состояний двоичного канала рассматривается как N-связная двоичная цепь Маркова, а значения символов на каждой позиции – как состояние цепи, где N – число символов, на которое распространяется память канала.

Основные положения теории статических решений. Оптимальные критерии.

При синтезе оптимальных устройств обработки сигналов одним из исходных положений является выбор критерия качества работы устройства.

Пусть принятый сигнал

$$u(t) = s(t, \lambda) * n(t), \quad (1)$$

Где * - символ, характеризующий способ взаимодействия полезного сигнала $s(t, \lambda)$ и помехи $n(t)$.

Положим, что наблюдения колебания $u(t)$ производится в дискретные моменты времени t_1, \dots, t_M .

С учетом (1), запишем:

$u = s * n$, где $u = (u_1, \dots, u_M)$, $s = (s_1, \dots, s_M)$, $n = (n_1, \dots, n_M)$ - M-мерные векторы соответственно принятого колебания, полезного сигнала и помехи.

Статистические характеристики помехи предполагаются известными и задаются в виде распределения $w(n) = w_M(w_1, \dots, w_M)$. Известными считается так же способ взаимодействия сигнала с помехой.

Статистическое описание вектора u принятого колебания для фиксированного вектора s – условное распределение $w(u | s) = w_M(u_1, \dots, u_M | s)$.

Функцию $w(u | s)$ называют функцией правдоподобия. После получения выборки u_1, \dots, u_M функция $w(u | s)$ зависят только от s и характеризует степень соответствия вектора принятого колебания тому или иному вектору полезного сигнала, т.е. показывает, насколько один вектор s при известной выборке u_1, \dots, u_M более правдоподобен, чем другие.

При известном распределении $w(s) = w_M(s_1, \dots, s_M)$ полезных сигналов, зная $w(u | s)$, можно найти совместное распределение векторов u принимаемых колебаний и векторов s полезных сигналов:

$$w(u, s) = w(s)w(u | s) \quad (2)$$

Задача заключается в том, чтобы на основе полученной выборки u_1, \dots, u_M и априорных данных с помощью взаимодействия сигнала и помехи принять одно вполне определенное решение из набора $\gamma_0, \gamma_1, \dots, \gamma_k$ возможных, образующих пространство решений Γ .

Выбор решения по принятой выборке u_1, \dots, u_M производится в соответствии с алгоритмом работы решающего устройства $\Delta(\gamma | u)$. Функция $\Delta(\gamma | u)$ называется решающей функцией или решающим правилом. Пространство принимаемых сигналов U разбивается на непересекающиеся области U_0, U_1, \dots и устанавливается соответствие между решениями и этими областями. Если каждой области становится в соответствие только вполне определенное решение, то правило $\Delta(\gamma | u)$ называется нерандомизированным (детерминированным).

Если некоторой области U_i ставятся в соответствие решения $\gamma_{j1}, \dots, \gamma_{jk}$ соответственно с

вероятностями p_{i1}, \dots, p_{ik} , причем $\sum_{j=1}^k p_{ij} = 1$, то такие правила называются рандомизированными.

В дальнейшем рассматриваются только нерандомизированные правила, и соответственно функция

$$\Delta(\gamma_i | u_j) = \begin{cases} 1, & \text{если } i = j; \\ 0, & \text{если } i \neq j. \end{cases}$$

При любом решающем правиле при наличии помех всегда возможны ошибочные решения. Для количественной оценки потерь, связанного с принятием решений, вводится так называемая функция потерь $\Pi(s, \gamma)$. Значение $\Pi(s_i, \gamma_j)$ характеризует потери при принятии решения γ_j в то время как правильным является решение γ_i . Функция потерь должны удовлетворять следующим свойствам:

$$\Pi(s_i, \gamma_j) > 0; \Pi(s_i, \gamma_j) > \Pi(s_j, \gamma_i); \Pi(s_i, \gamma_i) \leq 0.$$

Сформулируем математическую задачу выбора решения: на основе априорных данных с пространствах полезных сигналов S и помех N , распределениях вероятностей на этих пространствах $w(s)$ и $w(n)$, способе взаимодействия сигнала и помехи на заданной функции потерь $\Pi(s, \gamma)$ необходимо по полученному сигналу u оптимальным образом найти решение γ о том, какой конкретно из полезных сигналов содержится на входе приемника.

Так как s и γ являются случайными величинами, то $\Pi(s, \gamma)$ так же является случайной величиной.

Поэтому качество решения можно характеризовать математическим ожиданием функции потерь:

$$R = M \{ \Pi(s, \gamma) \} = \sum_j \sum_i w(s_i, \gamma_j) \Pi(s_i, \gamma_j), \quad (3)$$

Где $w(s_i, \gamma_j)$ - совместная вероятность появления на входе приемника сигнала s_i и принятия решения γ_j .

Величина R характеризует средние потери при принятии решения и называется средним риском. Оптимальным решающим правилом будет такое, для которого значение среднего риска будет наименьшим. Правило, при котором минимизируется средний риск, называется байесовским правилом или байесовским критерием.

В случае, когда пространства сигналов S и решений Γ непрерывны, средний риск:

$$R = \int_S \int_{\Gamma} w(s, \gamma) \Pi(s, \gamma) d\gamma ds, \quad (4)$$

Где $w(s, \gamma)$ - совместная плотность вероятности сигнала s и решения γ .

Для нерандомизированных правил принятое колебание и решение связаны детерминированной зависимостью, и поэтому справедливо соотношение

$$w(s, \gamma) ds d\gamma = w(s, u) ds du. \quad (5)$$

С учетом формулы (5) выражение (4) записывается в виде

$$H = \int_S \int_U w(s, u) (\gamma_u) du ds,$$

Где γ_u - решение, соответствующее принятому сигналу $u(t)$.

Заметим, что средний риск не является исчерпывающей характеристикой решающего правила. Более полной характеристикой качества обработки может служить совокупность двух показателей: среднего риска R и дисперсии среднего риска D_R . При этом решающее правило значительно усложняется и на практике ограничиваются критерием минимума среднего риска.

Применение байесовского критерия требует большого объема априорной информации. Если априорное распределение сигналов $w(s)$ неизвестно, то применять критерий Байеса нельзя. В этом случае применяют минимаксный критерий.

Введем понятие условного среднего риска:

$$H_{s_i} = \sum_j w_i(\gamma_j) (\gamma_j | i)$$

Для дискретных пространств S и Γ :

$$E_I \approx \int_{\Gamma} w(\gamma, s) d(\gamma | u) \quad E_U \approx \int_U \psi(\gamma, u, s) d(u | u) .$$

В соответствии с минимаксным критерием наилучшим считается правило, которое обеспечивает наименьшее значение максимального условного риска r_s .

Минимаксный критерий дает наилучшее решение для наихудшего случая. Он гарантирует что минимально возможное значение условного риска не будет больше некоторого значения даже при самом неблагоприятном распределении $w(s)$. Самые неблагоприятные условия бывают редко, поэтому минимаксный критерий обозначают перестраховочным. Если при некотором неминимаксном правиле $\Delta(\gamma | u)$ максимальный условный риск несколько больше r_s , а остальные условные риски значительно меньше тех, которые соответствуют минимаксному правилу $\Delta^*(\gamma | u)$, то целесообразно отдать предпочтение правилу $\Delta(\gamma | u)$.

На практике часты случаи, когда ошибочные решения в одинаковой степени нежелательны. При этом целесообразно функцию потерь задать следующим образом:

$$H(s_i, \gamma_j) = \begin{cases} \text{нпуст} & j \neq i; \\ 0 & \text{при } i = j. \end{cases} \quad (6)$$

Тогда средний риск

$$R = \sum_i \sum_j w(s_i, \gamma_j) \quad (7)$$

Из (7) следует, что величина R с точностью до постоянного множителя совпадает с полной вероятностью ошибочных решений. Поэтому байесовский критерий при задании функции потерь в виде (6) есть смысл назвать критерием минимума полной вероятности ошибки. Его часто называют критерием идеального наблюдателя, а также критерием Котельникова-Зигерта.

К критерию минимума полной вероятности ошибки сводится критерий максимума апостериорной вероятности:

$$w(s|u) = \frac{w(s)w(u|s)}{w(u)},$$

В соответствии с которым решение принимается в пользу сигнала s_j , если

$$w(s_i|u) > w(s_j|u), \quad j = 1, \dots, m; \quad i \neq j. \quad (8)$$

Если решение принимать в соответствии с выражением (8), то гарантируется, что полная вероятность ошибки будет минимальной.

При отсутствии априорных сведений о функции потерь и распределении сигналов $w(s)$ считают, что $w(s) = \text{const}$, т.е. распределение сигналов равномерное. При этом апостериорная вероятность $w(s|u)$ с точностью до постоянного множителя совпадает с функцией правдоподобия $w(u|s)$. Для обеспечения минимального среднего риска решение принимается в пользу сигнала, для которого функция правдоподобия максимальна, т.е. в пользу сигнала s_i , если:

$$w(u|s_i) > w(u|s_j), \quad j = 1, \dots, m; \quad j \neq i. \quad (9)$$

Критерий (9) называют критерием максимального правдоподобия.

На практике, распределения $w(u | s_i)$, $i = 1, \dots, m$, известны не в полной мере, так как они зависят от некоторых неизвестных параметров сигналов (начальных фаз, амплитуд, доплеровских сдвигов частоты). Поэтому само распределение становится не детерминированным а случайным $w(\lambda)$, где λ - вектор случайных неизвестных параметров сигналов. В таком случае оптимальное по Байесу правило сводится к минимизации математического ожидания среднего риска $R(\lambda)$.

$$M\{R(\lambda)\} = \int_{\Lambda} R(\lambda)w(\lambda)d\lambda.$$

При задании функции потерь в виде (6) решающее правила (8) и (9) можно записать:

$$\bar{w}(s_i | u) > \bar{w}(s_j | u), \quad j = 1, \dots, k, \quad j \neq i;$$

$$\bar{w}(u | s_i) > \bar{w}(u | s_j), \quad j = 1, \dots, k, \quad j \neq i.$$

Где $\bar{w}(s_l | u)$, $\bar{w}(u | s_l)$, $l = 1, \dots, k$ - усредненные по всем неизвестным параметрам сигналов s_l апостериорные плотности вероятности и функции правдоподобия.

Алгоритм оптимального демодулятора.

Пусть сигнал на входе приемника представляет смесь одного из возможных полезных сигналов $s_r(t)$, $r = 1, 2, \dots, m$, и шума $n(t)$:

$$u(t) = s_r(t) + n(t), \quad 0 \leq t \leq T_c, \quad (10)$$

Где T_c - длительность полезного сигнала.

Необходимо на основании анализа реализации процесса $u(t)$ решить, какой из полезных сигналов присутствует в принятом сигнале. Будем предполагать, что момент t_0 поступления полезного сигнала известен, а так же известны параметры полезных сигналов $s_r(t)$, в том числе начальные фазы $\varphi_r(t)$.

В дальнейшем будем считать, что $t_0 = 0$, $\varphi_r = 0$. Кроме того, будем полагать, что процесс $n(t)$ - стационарный гауссовский с нулевым математическим ожиданием.

Пусть количество полезных сигналов $m = 2$. С учетом (9) алгоритм работы оптимального различителя сводится к вычислению отношения функций правдоподобия

$$L^M(u) = w(u | s_1) / w(u | s_2)$$

И сравнения его с порогом $L_0 = p_2 / p_1$, где p_1 и p_2 - априорные вероятности появления сигналов $s_1(t)$ и $s_2(t)$ соответственно.

Найдем структуру функции правдоподобия $w(u | s_r)$, $r = 1, 2$. С учетом 10, $w(u | s_r)$ определяется многомерной плотностью $w(n)$ с последующей заменой $n = u - s$.

Многомерную выборку $n = (n_1, n_2, \dots, n_M)$ удобно разложить используя коэффициенты разложения Карунена-Лозва случайного процесса $n(t)$:

$$n(t) = \sum_{k=1}^M n_k \varphi_k(t), \quad 0 \leq t \leq T_c,$$

Где $n_k = \int_0^{T_c} n(t) \varphi_k(t) dt$, а ортонормированные функции $\varphi_k(t)$ удовлетворяют

интегральному уравнению

$$\int_0^{T_c} R(t_2 - t_1) \varphi_k(t_2) dt_2 = \sigma_k^2 \varphi_k(t_1), \quad (11)$$

Где $R(t_2 - t_1)$ - корреляционная функция процесса $n(t)$.

С учетом свойств разложения Каренена-Лоэва и закона распределения процесса $n(t)$, можно утверждать, что координаты n_k , $k = 1, \dots, M$, являются независимыми гауссовскими ожиданиями и дисперсиями σ_k^2 . Соответственно M -мерная плотность вероятности определяется:

$$w(n_1, \dots, n_M) = (\sqrt{2\pi})^{-M} \prod_{k=1}^M \sigma_k^{-1} \exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{n_k^2}{\sigma_k^2}\right).$$

В качестве координат векторов $u = (u_1, \dots, u_M)$ и $s_r = (s_{r1}, \dots, s_{rM})$ возьмем коэффициенты разложения Карунена-Лоэва сигналов $u(t)$ и $s(t)$:

$$u(t) = \sum_{k=1}^M u_k \varphi_k(t),$$

$$s_r(t) = \sum_{k=1}^M s_{rk} \varphi_k(t), \tag{12}$$

$$u_k = \int_0^{T_c} u(t) \varphi_k(t) dt,$$

$$s_{rk} = \int_0^{T_c} s_r(t) \varphi_k(t) dt.$$

Учитывая, что

$$n = u \cdot s_r,$$

$$n_k = u_k \cdot s_{rk},$$

$$w(u | s_r) = w(n) = w(u - s_r),$$

Находим выражение для функции правдоподобия

$$w(u | s_r) = (\sqrt{2\pi})^{-M} \prod_{k=1}^M \sigma_k^{-1} \exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{(u_k - s_{rk})^2}{\sigma_k^2}\right),$$

Соответственно отношение правдоподобия

$$\begin{aligned} L^M(u) &= \frac{w(u | s_1)}{w(u | s_2)} = \\ &= \frac{\exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{(u_k - s_{1k})^2}{\sigma_k^2}\right)}{\exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{(u_k - s_{2k})^2}{\sigma_k^2}\right)} = \frac{\exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{s_{1k}^2}{\sigma_k^2}\right) \exp\left(\frac{M}{\sum_{k=1}^M \frac{u_k s_{1k}}{\sigma_k^2}}\right)}{\exp\left(-\frac{1}{2} \sum_{k=1}^M \frac{s_{2k}^2}{\sigma_k^2}\right) \exp\left(\frac{M}{\sum_{k=1}^M \frac{u_k s_{2k}}{\sigma_k^2}}\right)} \end{aligned} \quad (13)$$

$\Lambda(u) = \lim_{M \rightarrow \infty} \Lambda^M(u)$ называется функционалом отношения правдоподобия.

Логарифмируя полученное выражение, получаем

$$\ln \Lambda(u) = \sum_{k=1}^{\infty} \frac{u_k}{\sigma_k^2} (s_{1k} - s_{2k}) - \frac{1}{2} \sum_{k=1}^{\infty} \frac{s_{1k}^2 - s_{2k}^2}{\sigma_k^2} \quad (14)$$

Пусть $n(t)$ представляет собой белый шум с корреляционной функцией

$$R(\tau) = \frac{N_0}{2} \delta(\tau) \quad (15)$$

Подставляя 15 в 11, видно, что уравнению 11 удовлетворяют любые $\varphi_k(t)$, ортонормированные на интервале $(0, T_c)$, а все коэффициенты σ_k^2 одинаковы и равны $N_0 / 2$. Тогда 14 запишется как

$$\ln \Lambda(u) = \frac{2}{N_0} \sum_{k=1}^{\infty} u_k (s_{1k} - s_{2k}) - \frac{1}{N_0} \sum_{k=1}^{\infty} (s_{1k}^2 - s_{2k}^2) \quad (16)$$

С учетом (12) и ортонормированности функции $\varphi_k(t)$

$$\int_0^{T_c} u(t) s_r(t) dt = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} u_k s_{rl} \int_0^{T_c} \varphi_k(t) \varphi_l(t) dt = \sum_{k=1}^{\infty} u_k s_{rk}$$

Соответственно

$$\begin{aligned} \sum_{k=1}^{\infty} u_k (s_{1k} - s_{2k}) &= \int_0^{T_c} u(t) [s_1(t) - s_2(t)] dt, \\ \sum_{k=1}^{\infty} (s_{1k}^2 - s_{2k}^2) &= \int_0^{T_c} s_1^2(t) dt - \int_0^{T_c} s_2^2(t) dt. \end{aligned} \quad (17)$$

Заменяя в 16 суммы интегралами согласно 17, находим

$$\ln \Lambda(u) = \frac{2}{N_0} \int_0^{T_c} u(t) [s_1(t) - s_2(t)] dt - \frac{1}{N_0} (E_1 - E_2)$$

Где $E_r = \int_0^{T_c} s_r^2(t) dt$ - энергия сигнала $s_r(t)$.

Тогда алгоритм работы оптимального по критерию минимума полной вероятности ошибки различителя для $m = 2$ записывается в виде: принимается решение в пользу сигнала $s_1(t)$, если

$$q = \frac{2}{N_0} \int_0^{T_c} u(t) [s_1(t) - s_2(t)] dt \geq \ln \frac{p_2}{p_1} + \frac{1}{N_0} (E_1 - E_2) = l_0 \quad (18)$$

На практике обычно $p_1 = p_2$ и $E_1 = E_2$. Тогда 18 запишется как

$$q = \frac{2}{N_0} \int_0^{T_c} u(t) [s_1(t) - s_2(t)] dt \geq 0$$

В общем случае ($m > 2$), полагая, что все сигналы равновероятны, алгоритм оптимального демодулятора можно представить в виде следующей системы ($m-1$) неравенств:

регистрируется сигнал $s_l(t)$, если

$$\int_0^{T_c} u(t) s_l(t) dt - \frac{E_1}{2} \geq \int_0^{T_c} u(t) s_r(t) dt - \frac{E_r}{2}, l, r = 1, \dots, m; r \neq l. \quad (19)$$

Демодулятор на рис.1. содержит m идентичных каналов, включающих местный генератор, формирующий опорный сигнал $s_r(t)$, перемножитель и интегратор. На выходе каждого канала установлено устройство, в котором из соответствующего результата интегрирования вычитается постоянное значение, определяемое энергией полезных сигналов.

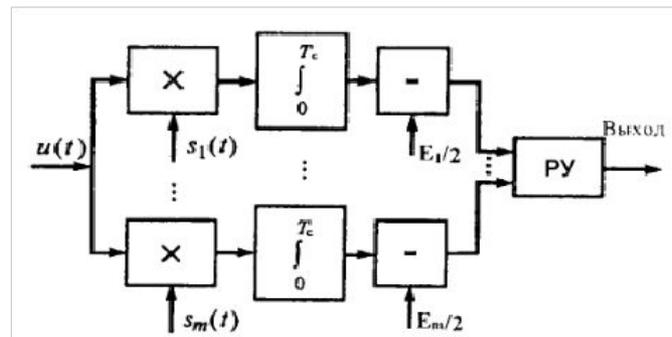


Рис. 1. Структурная схема демодулятора m детерминированных сигналов с использованием корреляторов

Решение о том, какой из полезных сигналов m присутствует на входе, принимается в решающем устройстве (РУ) в момент, совпадающий с моментом окончания полезного сигнала, после чего происходит сброс напряжения на интеграторах до нулевого уровня.

Задача построения демодуляторов облегчается, если выполняется условие $E_r = E$, $r = 1, \dots, m$. Такая система сигналов называется системой с активной паузой.

При этом отпадает необходимость использовать устройства вычитания.

$$\int_0^{T_c} u(t)s_l(t)dt \geq \int_0^{T_c} u(t)s_r(t)dt, \quad l, r = 1, \dots, m; r \neq l.$$

Демодулятор на рис.1. называется корреляционным, поскольку напряжение на выходе интегратора любого канала в момент окончания анализа пропорционально значению функции взаимной корреляции сигнала $u(t)$ и опорного напряжения $s_r(t)$.

Другой метод реализации алгоритма 19 основан на использовании согласованных фильтров. Пусть имеется фильтр, согласованный с сигналом $s_i(t)$. Импульсная характеристика такого фильтра

$$h_i(t) = bs_i(T_c - t) \quad (20)$$

Выходное напряжение фильтра при поступлении на его вход процесса $u(t)$ определяется с помощью интеграла Дюамеля

$$q_i(t) = \int_0^t u(t - \tau)h_i(\tau)d\tau$$

С учетом (20)

$$q_i(t) = b \int_0^t u(t - \tau)s_i(T_c - \tau)d\tau \quad (21)$$

Заменив переменную интегрирования в 21, получим

$$q_i(t) = b \int_{T_c-t}^{T_c} u(y + t - T_c)s_i(y)dy$$

Таким образом, в момент окончания полезного сигнала

$$q_i(t) = b \int_0^{T_c} u(t) s_i(t) dt,$$

т.е. получаем величину, которая должна быть вычислена в соответствии с 19.

Следовательно алгоритм оптимального приема может быть реализован с помощью устройства на основе фильтров $C\Phi_i$, согласованных с сигналами $s_i(t)$.

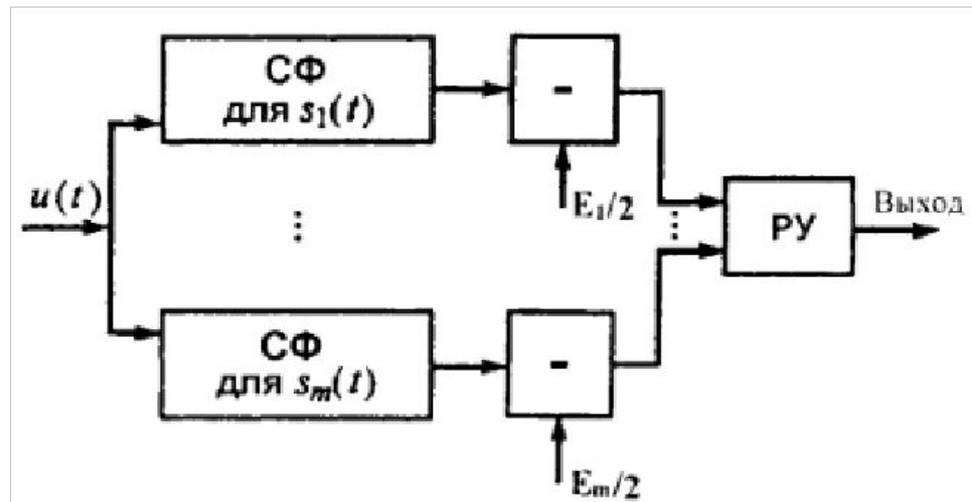


Рис. 2. Структурная схема демодулятора m детерминированных сигналов с использованием согласованных фильтров.

Потенциальная помехоустойчивость. Двоичные системы передачи информации.

Пусть сигнал на входе приемника имеет вид $u(t) = \lambda s_1(t) + (1 - \lambda)s_2(t) + n(t)$, где λ - случайная величина, принимающая значения 1 и 0 с вероятностями p_1 и p_2 соответственно, причем $p_1 + p_2 = 1$; $s_1(t)$ и $s_2(t)$ - полезные сигналы с известными параметрами; $n(t)$ - стационарный гауссовский белый шум с нулевым математическим ожиданием и корреляционной функцией $R_n(\tau) = (N_0 / 2)\delta(\tau)$.

Алгоритм работы демодулятора описывается выражением (18).

Средняя (полная) вероятность ошибки, используемая для количественной оценки помехоустойчивости, для рассматриваемого случая

$$P_{ош} = p_1 P_{ош}(s_2 | s_1) + p_2 P_{ош}(s_1 | s_2), \quad (22)$$

где условные вероятности ошибок

$$P_{ош}(s_2 | s_1) = \int_{q < l_0} w(q | s_1) dq, \quad (23)$$
$$P_{ош}(s_1 | s_2) = \int_{q \geq l_0} w(q | s_2) dq.$$

$w(q | s_1)$ и $w(q | s_2)$ - плотности вероятности случайной величины q при наличии на входе сигналов $s_1(t)$ и $s_2(t)$ соответственно.

Найдем распределение величины q для указанных случаев. Пусть на входе приемника присутствует сигнал $s_1(t)$. Тогда с учетом (18)

$$q = \int_0^{T_c} [s_1(t) + n(t)][s_1(t) - s_2(t)] dt$$

Легко видеть, что величина q является линейным функционалом гауссовского случайного процесса, а следовательно, имеет нормальную плотность вероятности.

Математическое ожидание величины q

$$m_q = M \{q\} = \int_0^{T_c} [s_1(t) + M \{n(t)\}][s_1(t) - s_2(t)] dt = E(1 - r_s)$$

где $r_s = \frac{1}{E} \int_0^{T_s} s_1(t)s_2(t) dt$ условно называют коэффициентом взаимной корреляции между сигналами $s_1(t)$ и $s_2(t)$.

Дисперсия величины q

$$\begin{aligned} D_q &= M \{(q - m_q)^2\} = M \left\{ \left[\int_0^{T_c} n(t)[s_1(t) - s_2(t)] dt \right]^2 \right\} = \\ &= \int_0^{T_c} \int_0^{T_c} M \{n(t_1)n(t_2)\} [s_1(t_1) - s_2(t_1)][s_1(t_2) - s_2(t_2)] dt_1 dt_2 \end{aligned}$$

Учитывая, что $M \{n(t_1)n(t_2)\} = (N_0 / 2)\delta(t_2 - t_1)$, и используя фильтрующее свойство δ -функции, находим

Таким образом,

$$w(q | s_1) = \frac{1}{\sqrt{2\pi} \sqrt{EN_0(1-r_s)}} \exp \left\{ -\frac{[q - E(1-r_s)]^2}{2EN_0(1-r_s)} \right\}.$$

Аналогично можно показать, что если присутствует сигнал $s_2(t)$, то случайная величина

$$q = \int_0^{T_c} [s_2(t) + n(t)][s_1(t) - s_2(t)] dt$$

будет иметь нормальную плотность вероятности с параметрами $m_q = -E(1-r_s)$, $D_q = EN_0(1-r_s)$, т.е.

$$w(q | s_2) = \frac{1}{\sqrt{2\pi} \sqrt{EN_0(1-r_s)}} \exp \left\{ -\frac{[q + E(1-r_s)]^2}{2EN_0(1-r_s)} \right\}$$

Плотности вероятности $w(q | s_1)$ и $w(q | s_2)$ представлены на рис.3.

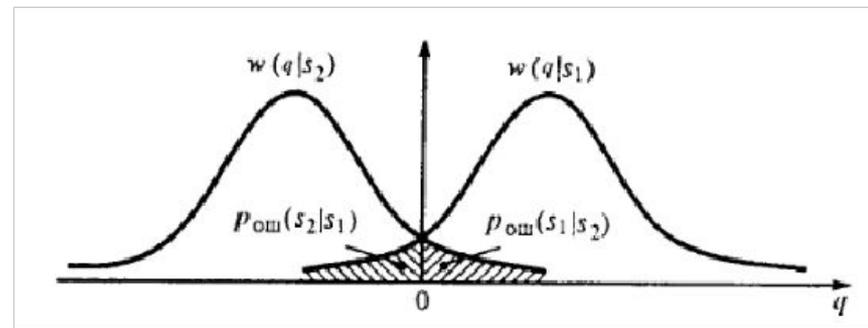


Рис.3. Плотности вероятности сигналов на выходе оптимального демодулятора в системах с активной паузой

Теперь, используя 23, нетрудно определить условные вероятности ошибок:

$$p_{ош}(s_2 | s_1) = \int_{-\infty}^{l_0} \frac{1}{\sqrt{2\pi} \sqrt{EN_0(1-r_s)}} \exp \left\{ -\frac{[q - E(1-r_s)]^2}{2EN_0(1-r_s)} \right\} dq,$$

$$p_{ош}(s_1 | s_2) = \int_{l_0}^{\infty} \frac{1}{\sqrt{2\pi} \sqrt{EN_0(1-r_s)}} \exp \left\{ -\frac{[q + E(1-r_s)]^2}{2EN_0(1-r_s)} \right\} dq.$$

Для $p_1 = p_2 = 1/2$ и $E_1 = E_2 = E$ порог $l_0 = 0$. При этом

$$p_{ош}(s_2 | s_1) = p_{ош}(s_1 | s_2) = 1 - F(\sqrt{E(1-r_s)}/N_0), \quad (24)$$

Где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$ - интеграл вероятности.

Используя (22) и (24), окончательно находим

$$p_{ош} = 1 - F(\sqrt{E(1-r_s)}/N_0) = 1 - F(h\sqrt{E(1-r_s)}) \quad (25)$$

где $h^2 = E/N_0$.

Из (25) видно, что средняя вероятность ошибки зависит не только от энергии сигнала и спектральной плотности мощности шума, но и от коэффициента взаимной корреляции между сигналами, т. е. от используемой системы сигналов. Интеграл вероятности $F(x)$ является монотонно возрастающей функцией аргумента. Поэтому при одном и том же отношении E/N_0 помехоустойчивость системы оказывается тем выше, чем меньше коэффициент взаимной корреляции r_s .

Так как $-1 < r_s < 1$, то наибольшей помехоустойчивостью обладают сигналы с $r_s = -1$. Они имеют одинаковую форму, но противоположные знаки и называются *противоположными*. Для них

$$p_{ош} = 1 - F(\sqrt{2E/N_0}) = 1 - F(\sqrt{2}h). \quad (26)$$

Примером противоположных сигналов являются фазоманипулированные сигналы с манипуляцией фазы на π :

$$s_1(t) = A_0 \cos w_0 t, \quad s_2(t) = A_0 \cos(w_0 t + \pi), \quad 0 \leq t \leq T_c.$$

Меньшей помехоустойчивостью обладают *ортогональные* сигналы ($r_s = 0$). Для них

$$p_{ош} = 1 - F(\sqrt{E/N_0}) = 1 - F(h) \quad (27)$$

Сигналы с $r_s = 1$ являются одинаковыми, т. е. $s_1(t) = s_2(t)$, и их невозможно различить. Для них $p_{ош} = 1/2$.

Примером ортогональных сигналов являются фазоманипулированные сигналы с манипуляцией фазы на $\pi/2$:

$$s_1(t) = A_0 \cos w_0 t,$$

$$s_2(t) = A_0 \cos(w_0 t + \pi/2), \quad 0 \leq t \leq T_c$$

Ортогональные сигналы можно получить на основе частотной манипуляции.

Действительно, в этом случае

$$s_1(t) = A_0 \cos(\omega_1 t - \varphi_1),$$

$$s_2(t) = A_0 \cos(\omega_2 t - \varphi_2), \quad 0 \leq T \leq T_c$$

При $\varphi_1 = \varphi_2$ коэффициент взаимной корреляции между этими сигналами

$$r_s = \sin[(\omega_2 - \omega_1)T_c] / (\omega_2 - \omega_1)T_c + \\ + \{ \sin[(\omega_2 + \omega_1)T_c - 2\varphi] + \sin 2\varphi \} / [(\omega_1 + \omega_2)T_c]$$

При выполнении условия $(\omega_2 - \omega_1)T_c = 2k\pi$, $k = 1, 2, \dots$, $r_s = 0$ и сигналы оказываются ортогональными.

На практике обычно ортогональность обеспечивается выбором параметров ω_1, ω_2 и T_c так, чтобы $(\omega_2 - \omega_1)T_c \gg 1$. При этом $r_s \approx 0$.

Заметим, что минимальное значение коэффициента взаимной корреляции между частотно-манипулированными сигналами равно $-1/1,5\pi$. Оно достигается, когда $(\omega_2 - \omega_1)T_c = 1,5\pi$. При этом вероятность ошибок

$$P_{ош} = 1 - F(\sqrt{1,21E/N_0}) = 1 - F(\sqrt{1,21}h)$$

Оценим помехоустойчивость системы передачи, использующей амплитудно-манипулированные сигналы

$$s_1(t) = A_0 \cos(\omega_0 t + \varphi), \quad s_2(t) = 0, \quad 0 \leq T \leq T_c.$$

Алгоритм различения сигналов (18) в рассматриваемом случае принимает вид

$$q = \int_0^{T_c} u(t) s_1(t) dt \geq \frac{E}{2} = l_0 \quad (28)$$

Плотности вероятности $w(q|s_1)$ и $w(q|s_2)$ описываются гауссовским законом с параметрами $m_q = E$, $D_q = EN_0/2$ и $m_q = 0$, соответственно (рис.4).

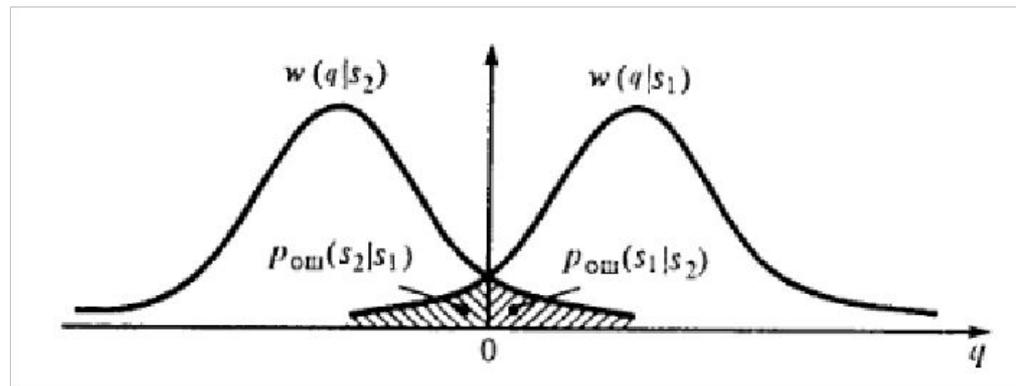


Рис. 4. Плотности вероятности сигналов на выходе оптимального демодулятора в системах с пассивной паузой.

При $p_1 = p_2 = 1/2$ средняя вероятность ошибки

$$P_{ош} = \frac{1}{2} \left[\int_{-\infty}^{l_0} w(q|s_1) dq + \int_{l_0}^{\infty} w(q|s_2) dq \right]$$

Учитывая, что порог $l_0 = E/2$ [см. (28)], находим

$$P_{ош} = 1 - F(0,5\sqrt{2E/N_0}) = 1 - F(h/\sqrt{2}) \quad (29)$$

На рис. 5 представлены зависимости вероятности ошибок от отношения E/N_0 для фазо- (ФМ), частотно- (ЧМ) и амплитудно-манипулированных (АМ) сигналов, рассчитанные соответственно по формулам (26), (27), (29) (сплошные линии).

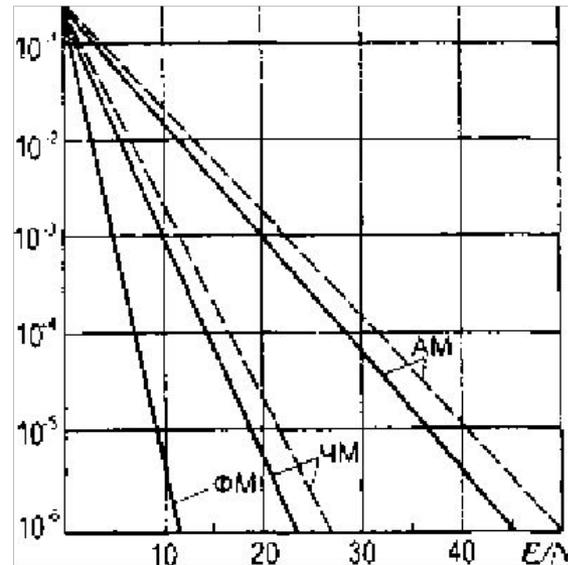


Рис. 5. Зависимость вероятности ошибки от E/N_0 для детерминированных сигналов (сплошные линии) и для сигналов со случайными начальными фазами (штриховые линии)

Таким образом, наибольшей потенциальной помехоустойчивостью обладают фазоманипулированные сигналы. Они обеспечивают выигрыш в энергии сигнала в два раза по сравнению с частотно-манипулированными сигналами и в четыре раза по сравнению с амплитудно-манипулированными сигналами.

Частотно-манипулированные сигналы обеспечивают выигрыш в энергии сигнала по сравнению с амплитудно-манипулированными сигналами в два раза.

Однако следует иметь в виду, что, в отличие от фазовой и частотной манипуляций, при амплитудной манипуляции передается только один сигнал. Поэтому если исходить из среднеэнергетических затрат, то нетрудно видеть, что системы с АМ и ЧМ сигналами обладают одинаковой помехоустойчивостью.

Заметим, что величина $\sqrt{2E(1-r_s)}$ представляет собой расстояние между сигналами

$$d = \left[\int_0^{T_c} [s_1(t) - s_2(t)]^2 dt \right]^{1/2}$$

При этом формулу (25) можно записать в виде

$$p_{ош} = 1 - F(d\sqrt{2N_0}) \quad (30)$$

Из (30) следует, что при действии в канале гауссовского белого шума вероятность ошибки зависит только от расстояния между сигналами и спектральной плотности шума. Этот вывод оказывается справедливым и для $m > 2$.

При высоких требованиях к помехоустойчивости ($p_{ош} < 10^{-3}$) вероятность ошибки удобно определить по приближенной формуле:

$$p_{ош} = \frac{1}{\sqrt{2\pi} \sqrt{h^2_0(1-r_s)}} \exp \left\{ -\frac{h^2(1-r_s)}{2} \right\},$$

Которая получается при асимптотическом представлении интеграла вероятности

$$F(x) \approx 1 - \frac{\exp(-x^2 / 2)}{\sqrt{2\pi} x} \quad (32)$$

Точность вычислений по формуле 31 не хуже 10%, если $\sqrt{h^2(1-r_s)} \geq 3$.

M-ичные системы передачи информации.

Пусть принятый сигнал имеет вид $u(t) = s_r(t) + n(t)$, $0 \leq T \leq c$, где $s_r(t)$, $r = 1, \dots, m$, ($m > 2$) - возможные полезные сигналы на входе приемника; $n(t)$ - помеха типа белого гауссовского шума. Будем считать, что вероятность передачи любого сигнала равна $1/m$. Тогда решение о том, какой из сигналов $s_r(t)$, $r = 1, \dots, m$, был передан, принимается на основе анализа $(m-1)$ неравенств (19). Ошибка при приеме сигнала возникает тогда, когда неравенство (19) не выполняется хотя бы для одного $r \neq l$.

Пусть q_1, \dots, q_m - напряжения на выходах каналов различителя, а $w_m(q_1, \dots, q_m | s_l)$ - m -мерная плотность вероятности совокупности случайных величин q_1, \dots, q_m при условии, что на входе приемника действует сигнал $s_r(t)$. Тогда с учетом алгоритма работы оптимального различителя вероятность правильного приема сигнала

$$p_{np}(s_l) = \int_{-\infty}^{\infty} dq_1 \int_{-\infty}^{q_1} \dots \int_{-\infty}^{q_l} w_m(q_1, \dots, q_m | s) dq_1, \dots, dq_m \quad (33)$$

Соответственно при передаче сигнала S вероятность ошибки

$$p_{ош}(s_i) = 1 - p_{np}(s_i)$$

Она при прочих равных условиях зависит от ансамбля применяемых сигналов $s_r(t)$, $r = 1, \dots, m$. Существует бесконечно большое число систем, отличающихся друг от друга индивидуальными и совместными свойствами сигналов. Представляет интерес система сигналов, обеспечивающая максимальную помехоустойчивость при заданных априорных условиях передачи.

При действии в канале помехи типа белого гауссовского шума помехоустойчивость системы зависит от расстояний между сигналами:

$$d(s_i, s_j) = \left\{ \int_0^{T_c} [s_i(t) - s_j(t)]^2 dt \right\}^{1/2}, \quad i, j = 1, \dots, m \quad (34)$$

причем чем больше минимальное из этих расстояний, тем выше помехоустойчивость системы.

Если сигналы имеют одинаковую энергию E , то (34) можно упростить:

$$d(s_i, s_j) = \left\{ 2E(1 - r_{ij}) \right\}^{1/2} \quad (35)$$

где $r_{ij} = \frac{1}{E} \int_0^{T_c} s_i(t)s_j(t)dt$ - коэффициент взаимной корреляции между сигналами $s_i(t)$ и $s_j(t)$.

Из (35) следует, что для достижения большего расстояния коэффициент взаимной корреляции должен быть как можно меньше. Для обеспечения одинаковой вероятности правильного приема любого сообщения надо потребовать, чтобы все коэффициенты r_{ij} были одинаковыми, т. е. $r_{ij} = r_0$ для всех i и $j, i \neq j$. Значение r_0 удовлетворяет неравенству $r_0 \geq -1/(m-1)$, которое вытекает из следующего соотношения:

$$\int_0^{T_c} \left[\sum_{i=1}^m s_i(t) \right]^2 dt = \int_0^{T_c} \sum_{i=1}^m \sum_{j=1}^m s_i(t) s_j(t) dt = mE + r_0 E(m^2 - m) \geq 0$$

Для оптимальной системы сигналов

$$r_{ij} = -1/(m-1) \tag{36}$$

Сигналы $s_r(t)$, $r = 1, \dots, m$, удовлетворяющие условию

$$\frac{1}{E} \int_0^{T_c} s_i(t) s_j(t) dt = \begin{cases} 1, & j = i, \\ -1/(m-1), & j \neq i. \end{cases}$$

называются *симплексными*, поскольку в $(m-1)$ -мерном пространстве они образуют правильный симплекс с числом вершин m . Симплексные сигналы являются *эквилистантными*, т. е. для всех пар сигналов $s_i(t)$ и $s_j(t)$ расстояние $d(s_i, s_j)$ одинаково.

На практике часто применяют ортогональные сигналы, для которых

$$r_{ij} = \frac{1}{E} \int_0^{T_c} s_i(t) s_j(t) dt = \begin{cases} 1, & j = i, \\ 0, & j = 1, \dots, m, \\ 0, & j \neq i. \end{cases} \quad (37)$$

При больших значениях m ортогональные сигналы по помехоустойчивости близки к симплексным. Это следует из того, что значение r_0 , определяемое формулой (36), при больших m стремится к нулю. Ортогональные сигналы с равной энергией также являются эквидистантными.

Другой системой, близкой при $m \gg 1$ к симплексной, является *биортогональная* система сигналов $s_r(t)$, $r = 1, \dots, m$ (m - четное число), которая характеризуется тем, что для каждого сигнала $s_i(t)$ существует противоположный сигнал $-s_i(t)$, а остальные сигналы ортогональны сигналу $s_i(t)$. Определить помехоустойчивость m -ичных систем в общем случае трудно. Однако для равновероятных симплексных, ортогональных и биортогональных сигналов выражение (33) существенно упрощается и сводится к однократному интегралу, который можно оценить с помощью численных методов.

Системы передачи с ортогональными сигналами.

Пусть сигнал на входе приемника имеет вид $u(t) = s_r(t) + n(t), 0 \leq T \leq T_c$.

Тогда напряжение на выходе 1-го канала

$$q_i = \int_0^{T_c} u(t) s_i(t) dt$$

будет представлять собой гауссовскую случайную величину с математическим ожиданием $M\{q_i\} = E$ и дисперсией $D_{q_i} = EN_0 / 2$, а напряжения на выходах остальных каналов будут представлять гауссовские случайные величины с нулевыми математическими ожиданиями и дисперсиями, равными $EN_0 / 2$. Нетрудно показать, что в рассматриваемом случае величины q_1, \dots, q_m являются некоррелированными

$$M\{q_i, q_j\} = 0, \quad i, j = 1, \dots, m, \quad i \neq j,$$

а следовательно, с учетом их распределения и статистически независимыми.

При этом m -мерная плотность вероятности

$$w_m(q_1, \dots, q_m) = w(q_1)w(q_2) \dots w(q_m)$$

Где

$$w(q_l) = \frac{1}{\sqrt{2\pi} \sqrt{EN_0/2}} \exp \left\{ -\frac{[q-E]^2}{EN_0} \right\},$$
$$w(q_j) = \frac{1}{\sqrt{2\pi} \sqrt{EN_0/2}} \exp \left\{ -\frac{q_j^2}{EN_0} \right\}. \quad (38, 39)$$

Подставляя (37) - (39) в (33), после преобразований получаем:

$$p_{np}(s_j) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left(-\frac{x^2}{2} \right) F^{m-1} \left(x + \sqrt{\frac{2E}{N_0}} \right) dx =$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left(x - \sqrt{\frac{2E}{N_0}} \right)^2 \right] F^{m-1}(x) dx$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \left(-\frac{t^2}{2} \right) dt$ - интеграл вероятности.

Нетрудно видеть, что вероятность правильного приема оказывается одинаковой для всех сигналов $s\{t\}$, $\Gamma=1, \dots, m$.

Поэтому полная вероятность ошибки

$$P_{np}(s_j) = 1 - \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left(x - \sqrt{\frac{2E}{N_0}} \right)^2 \right] F^{m-1}(x) dx \quad (40)$$

Из (40) следует, что при прочих равных условиях с увеличением числа сигналов m вероятность ошибки возрастает. Физически это объясняется увеличением вероятности превышения шумом на выходе какого-либо канала (в момент принятия решения) напряжения на выходе канала, принимающего полезный сигнал. Однако это не означает, что потенциальная помехоустойчивость m -ичных систем меньше, чем двоичных. При сравнении систем необходимо иметь в виду, что каждый равновероятный m -ичный сигнал несет в $\log_2 m$ раз большее количество информации, чем двоичный сигнал, или при той же скорости передачи информации имеет в $\log_2 m$ большую длительность.

На рис.6 построены зависимости вероятности ошибки при когерентном приеме m -ичных ортогональных сигналов от $h_m^2 = E_B / N_0$, где $E_B = E / \log_2 m$ - энергия, затрачиваемая на 1 бит информации.

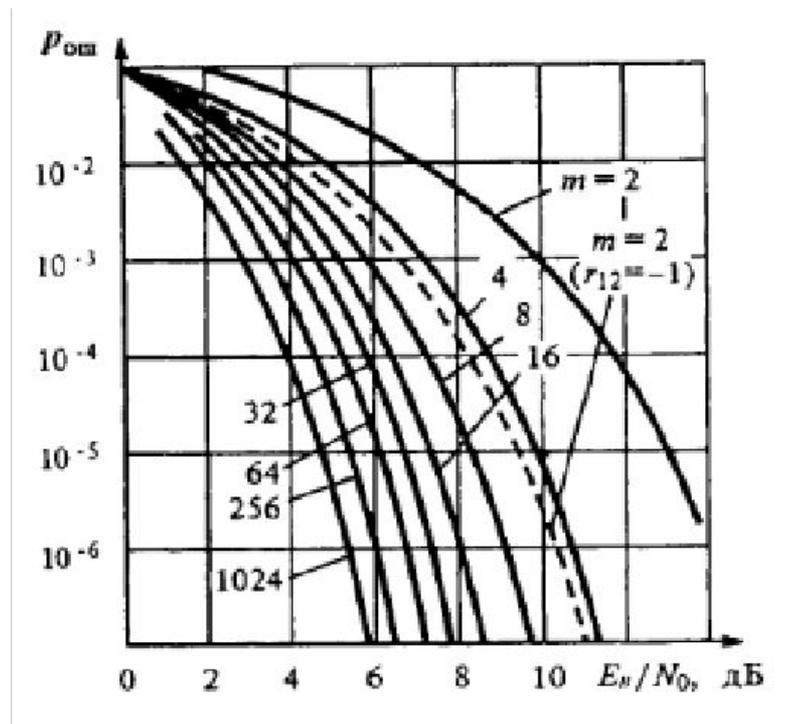


Рис. 6. Зависимость вероятности ошибки от отношения E_B / N_0 при оптимальном приеме m детерминированных ортогональных сигналов.

Системы ортогональных сигналов с $m > 2$ позволяют обеспечить при одинаковой скорости передачи информации существенный выигрыш в энергетике по сравнению с двоичными сигналами. Так, при $m = 32$ и $p_{ош} = 10^{-5}$ он составляет почти два раза.

Платой за энергетический выигрыш является увеличение ширины полосы частот, занимаемой системой, и усложнение приемника, который для сигналов с одинаковыми энергиями содержит m корреляторов (по числу сигналов) и решающее устройство.

Системы передачи с симплексными сигналами.

Вероятность ошибки $p_{ош}$ в системе с симплексными сигналами $s_r(t)$, $r = 1, \dots, m$, связана простым соотношением с вероятностью ошибки для ортогональных сигналов. Действительно, пусть $s_r(t)$, $r = 1, \dots, m$, - симплексные сигналы. Образует новый ансамбль сигналов длительностью $T_c(1 + |r_0|)$:

$$s_r(t) = \begin{cases} s_r(t), & 0 \leq t \leq T_c \\ \sqrt{E/T_c}, & T_c \leq t \leq T_c(1 + |r_0|) \end{cases} \quad (41)$$

где $r_0 = -1/(m-1)$.

Сигналы (41) являются ортогональными

$$\int_0^{T_c(1+|r_0|)} s_i'(t)s_j'(t) dt =$$

$$= \int_0^{T_c} s_i(t)s_j(t) dt + \int_0^{T_c(1+|r_0|)} \frac{E}{T_c} dt = Er_0 + E|r_0| = 0, i \neq j.$$

Энергия каждого сигнала $s_r'(t)$ равна $E(1+|r_0|)$.

Учитывая, что передача сигналов на интервале времени $T_c \leq t \leq T_c(1+|r_0|)$ одинаково влияет на выходные напряжения всех m каналов (на этом интервале времени все сигналы $s_r'(t)$ одинаковы), можно утверждать, что вероятность ошибки для исходного ансамбля сигналов $s_r(t)$, $r = 1, \dots, m$, равна вероятности ошибки для ортогонального ансамбля сигналов с энергией $E(1+|r_0|)$. Таким образом, вероятность ошибки для симплексных сигналов определяется как

$$p_{ош} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left(x - \sqrt{\frac{2E}{N_0} \frac{m}{m-1}} \right)^2 \right] F^{m-1}(x) dx$$

Зависимость вероятности ошибки от отношения $h_m^2 = E_B / N_0$ для симплексных сигналов можно проследить по рис. 6, если по оси абсцисс вместо E_B / N_0

отложить величину $\frac{E_B}{N_0} \frac{m}{m-1}$.

Из (40) следует, что при прочих равных условиях с увеличением числа сигналов m вероятность ошибки возрастает. Физически это объясняется увеличением

Помехоустойчивость симплексных сигналов выше, чем ортогональных. Однако это различие уменьшается с увеличением m и при $m \gg 1$ помехоустойчивость обеих ансамблей оказывается практически одинаковой.

Системы передачи с биортогональными сигналами.

Оптимальный различитель биортогональных сигналов состоит из набора $m/2$ корреляторов, устройства нахождения максимального по абсолютной величине напряжения на их выходах и устройства определения знака этого напряжения.

При передаче любого сигнала $s_r(t)$ ошибка отсутствует, если выполняются неравенства:

$$q_i \gg 0, q_i \gg \max(|q_1|, |q_2|, \dots, |q_{i-1}|, |q_{i+1}|, \dots, |q_{m/2}|).$$

Вероятность правильного приема любого сигнала

$$P_{np}(s_i) = \int_0^\infty dq_1 \int_{-q_1}^{q_1} \dots \int_{-q_1}^{q_1} w_{m/2}(q_1, \dots, q_{m/2} | s_i) dq_1 \dots dq_{m/2}.$$

Случайные величины $q_1, \dots, q_{m/2}$ являются статистически независимыми и распределены так же, как и для ортогональных сигналов. Это позволяет вывести следующее выражение для вероятности ошибки:

$$P_{ош} = 1 - \frac{1}{\sqrt{2\pi}} \int_0^\infty [2F(x) - 1]^{\frac{m}{2}-1} \exp\left[-\frac{1}{2}\left(x - \sqrt{\frac{2E}{N_0}}\right)^2\right] dx.$$

Помехоустойчивость биортогональных сигналов выше, чем ортогональных. Однако при $m \gg 2$ эта разница становится пренебрежимо малой.

Как уже указывалось, вычислить вероятность ошибки в глинной системе в общем случае трудно. Поэтому на практике часто пользуются верхней границей для вероятности ошибки

$$P_{ош}(s_i) \leq \sum_{\substack{j=1 \\ j \neq i}}^m P_{ош}(s_j | s_i) \quad (42)$$

где $p_{ош}(s_j | s_l)$ - вероятность ошибки при передаче сигнала $s_l(t)$ в двоичной системе, использующей сигналы $s_l(t)$ и $s_j(t)$. Оценка (42) справедлива для любой системы сигналов и любого канала.

Более простой, но менее точной является верхняя граница, определяемая как

$$p_{ош} \leq (m-1) \max p_{ош}(s_j | s_l),$$

где $\max p_{ош}(s_j | s_l)$ - максимальная по всем парам l, j вероятность ошибки в двоичной системе, использующей сигналы $s_l(t)$ и $s_j(t)$.

Выбор и формирование сигналов. Ортогональные сигналы.

В общем случае ортогональные сигналы можно сформировать следующим образом. Пусть $\varphi_j(t)$, $j = 1, \dots, N$ - некоторая полная ортонормированная система функций.

Любой сигнал $s_i(t)$, $i = 1, \dots, m$ с полосой частот F_c можно представить в виде

$$s_i(t) = \sum_{j=1}^N b_{ij} \varphi_j(t),$$

где $N = 2F_c T_c$ - число отсчетов на интервале T_c по теореме Котельникова,

$$b_{ij} = \int_0^{T_c} s_i(t) \varphi_j(t) dt, \quad j = 1, \dots, N - \text{коэффициенты разложения.}$$

Геометрически сигнал $s_i(t)$ можно представить вектором в N -мерном пространстве с координатами $(b_{i1}, b_{i2}, \dots, b_{iN})$. Сигналы $s_i(t)$, $i=1, \dots, m$, будут ортогональны, если для любого i -го сигнала удовлетворяется соотношение

$$b_{ij} = \begin{cases} \sqrt{E}, & j = i, \\ 0 & j \neq i. \end{cases}$$

Пусть в качестве примера базисные функции

$$\varphi_j(t) = \begin{cases} \sqrt{2/T_c} \sin w_j t, & 0 \leq t \leq T_c, \\ 0 & \text{при иных } t \end{cases}$$

где частоты w_j , $j=1, \dots, m$, выбираются из условия обеспечения ортогональности функций $\varphi_j(t)$, $j=1, \dots, m$.

Тогда сигналы

$$s_i(t) = \sqrt{\frac{2E}{T_c}} \sin w_i(t), \quad 0 \leq t \leq T_c, \quad i=1, \dots, m,$$

образуют ортогональную систему. Они получаются путем многопозиционной частотной манипуляции.

Существует бесконечно большое число ортогональных систем функций, на основе которых могут быть сформированы ортогональные сигналы. На практике часто для этого используют ортогональные коды. При этом сами сигналы получают фазовой манипуляцией несущего колебания по закону кодовых комбинаций.

В общем случае построение ортогональных кодов связано с матрицами Адамара, под которыми понимаются квадратные ортогональные матрицы с элементами ± 1 . Поэтому строки (или столбцы) матрицы Адамара можно использовать для формирования комбинаций ортогонального кода (символ -1 заменяется символом 0).

Укажем два основных положения, касающихся вопросов существования и построения матриц Адамара. Матрицы Адамара имеют порядок либо $N = 2$, либо $N = 4k$, $k = 1, 2, \dots$.

Матрица $A_{N_1 \times N_2}$ порядка $N_1 \times N_2$ полученная из матрицы Адамара A_{N_2} подстановкой матрицы Адамара A_{N_1} вместо элементов $+1$ и $-A_{N_1}$ вместо элементов -1 , есть тоже матрица Адамара.

Таким способом можно легко строить матрицы Адамара более высокого порядка.

Пусть в качестве примера даны матрицы Адамара:

$$\mathbf{A}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ и } \mathbf{A}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Используя указанный способ, нетрудно получить матрицу Адамара порядка $N=8$:

$$\mathbf{A}_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Если первая строка и первый столбец матрицы Адамара состоят из единиц, то говорят, что матрица записана в нормальной форме.

Ортогональные коды можно построить на основе системы функций Уолша, которые достаточно просто генерируются.

Биортогональные сигналы.

Для построения системы из m биортогональных сигналов берется система из $m/2$ ортогональных сигналов и к каждому из них добавляется противоположный. Простейшим биортогональным является ансамбль из четырех сигналов с одинаковой энергией.

Если в качестве базисных функций $\varphi_j(t)$ использовать

$$\varphi_1(t) = \sqrt{2/T_C} \cos w_0 t, \quad \varphi_2(t) = \sqrt{2/T_C} \sin w_0 t$$

то при $m=4$ биортогональные сигналы отличаются только фазой и совпадают с сигналами, полученными фазовой манипуляцией.

Симплексные сигналы. В общем случае они получаются из ортогональных сигналов следующим образом. Пусть $s_i(t)$, $i=1, \dots, m$ - ортогональные сигналы. Добавив к каждому сигналу $s_i(t)$, $i=1, \dots, m$, один и тот же сигнал $s(t)$, получим новую систему сигналов $s'_i(t) = s_i(t) + s(t)$. Заметим, что обе системы сигналов обеспечивают одинаковую помехоустойчивость.

Суммарная энергия новых сигналов

$$E_{\Sigma} = \sum_{j=1}^m \int_0^{T_c} [s_j(t) + s(t)]^2 dt \quad (44)$$

Для симплексных сигналов E_{Σ} должна быть минимальной. Минимизируя выражение (44) по $s(t)$, можно показать, что симплексные сигналы, получаемые на основе ортогональных $s_i(t)$, $i = 1, \dots, m$, имеют вид

$$s'_i(t) = s_i(t) - \frac{1}{m} \sum_{i=1}^m s_i(t) \quad (45)$$

Так как с учетом (45)

$$\sum_{i=1}^m s'_i(t) = 0$$

то каждый из сигналов $s'_i(t)$ можно представить в виде линейной комбинации остальных. Отсюда следует, что симплексные сигналы $s_i(t)$, $i = 1, \dots, m$, можно представить в виде векторов в $(m-1)$ -мерном евклидовом пространстве.

Задаваясь одной базисной функцией $\varphi(t)$, можно сформировать два симплексных сигнала, которые совпадут с противоположными сигналами. Задаваясь двумя базисными функциями, можно сформировать три симплексных сигнала. Концы их векторов лежат в вершинах равностороннего треугольника.

Симплексные сигналы могут быть получены на основе симплексных кодов. Рассмотрим равномерный код с основанием 2 и длиной m . Пусть его кодовые комбинации представляют последовательности из m символов, принимающих значения -1 и 1. Введем понятие коэффициента взаимной корреляции между любой парой комбинаций b_i и b_j

$$r_{ij} = \frac{1}{m} \sum_{k=0}^{m-1} b_{ik} b_{jk}$$

Потребуем, чтобы коэффициенты взаимной корреляции $r_{ij} = r_0$, $i, j = 1, \dots, m$;

$i \neq j$. Можно показать, что

$$r_0 \geq \begin{cases} -1/(\text{если } m \text{ четное число} & , \\ -1/ \text{если } m \text{ нечетное число} & , \end{cases}$$

Код, все пары кодовых комбинаций которого имеют коэффициенты взаимной корреляции

$$r_0 = \begin{cases} -1/(\sqrt{m}) & \text{если } m \text{ четное число} \\ -1/\sqrt{m} & \text{если } m \text{ нечетное число} \end{cases},$$

называется симплексным.

Симплексные коды можно построить на основе матриц Адамара. Нетрудно показать, что если существует матрица Адамара порядка $N=4k$, то можно построить симплексные коды для $m=4k$, $4k-1$, $2k$ и $2k-1$, где k - целое положительное число. Действительно, пусть A_N - матрица Адамара порядка $N=4k$, записанная в нормальной форме. Тогда, зачеркнув первый столбец, получаем матрицу, строки которой образуют симплексный код для $m=4k$. Если, кроме того, вычеркнуть одну из строк, то получим симплексный код для $m=4k-1$.

Коды для $m=2k$ и $m=2k-1$ получаются следующим образом. Возьмем любой j -й столбец ($j \neq 1$) и вычеркнем в матрице A_N строки, на которых элементы j -го столбца равны 1 (или -1).

Вычеркнем также первый и j -й столбцы. Тогда оказывается, что оставшиеся строки образуют симплексный код для $m = 2k$. Если вычеркнуть еще одну строку, то получим симплексный код для $m = 2k - 1$.

Особый интерес представляют симплексные коды, комбинации которых являются циклическими перестановками одной из них. Выясним, каким условиям должны удовлетворять такие кодовые последовательности.

Введем периодическую корреляционную функцию (ПКФ) кодовой последовательности b_0, b_1, \dots, b_{m-1} , символы которой принимают значения ± 1 :

$$R^{\Pi}(j) = \sum_{i=0}^{m-1} b_i b_{i+j}, \quad j = 0, 1, \dots, m-1.$$

Последовательности, ПКФ которых удовлетворяет условию $R^{\Pi}(j) = \alpha$, $j = 1, \dots, m-1$, называются последовательностями с *двухуровневой ПКФ*. Можно показать, что $|\alpha| \geq 1$. Для построения циклических симплексных кодов пригодны только последовательности с двухуровневой ПКФ, для которых $\alpha = -1$.

К числу таких последовательностей относятся линейные рекуррентные последовательности максимальной длины (M-последовательности), последовательности Лежандра, последовательности Холла и последовательности Якоби. Ниже рассматриваются только M-последовательности как наиболее часто применяемые на практике.

Линейной рекуррентной последовательностью (ЛРП) называется последовательность символов $\{b_i\} = b_0, b_1, \dots$, удовлетворяющая рекуррентному правилу

$$b_i = c_0 + c_1 b_{i-1} + c_2 b_{i-2} + \dots + c_n b_{i-n} \quad (46)$$

где как значения последовательности $\{b_i\}$, так и значения коэффициентов c_0 и c_1 принадлежат некоторому алфавиту $(0, 1, \dots, L-1)$, а операции сложения и умножения производятся по модулю L , причем L предполагается простым числом.

Соотношение (46) называется правилом кодирования, число n - памятью, а число L - основанием последовательности. В дальнейшем будут рассматриваться только двоичные последовательности ($L=2$).

Без потери общности коэффициент c_0 можно положить равным 0. Тогда рекуррентное правило запишется в виде

$$b_i = c_1 b_{i-1} \oplus c_2 b_{i-2} \oplus \dots \oplus c_n b_{i-n} \quad (47)$$

Из (47) следует, что для построения ЛРП необходимо в каждый тактовый момент времени запоминать n последних символов $b_{i-1}, b_{i-2}, \dots, b_{i-n}$ последовательности $\{b_i\}$ и суммировать их по модулю два с весами c_1, c_2, \dots, c_n . Эти операции осуществляет регистр сдвига с обратной связью (рис. 7).

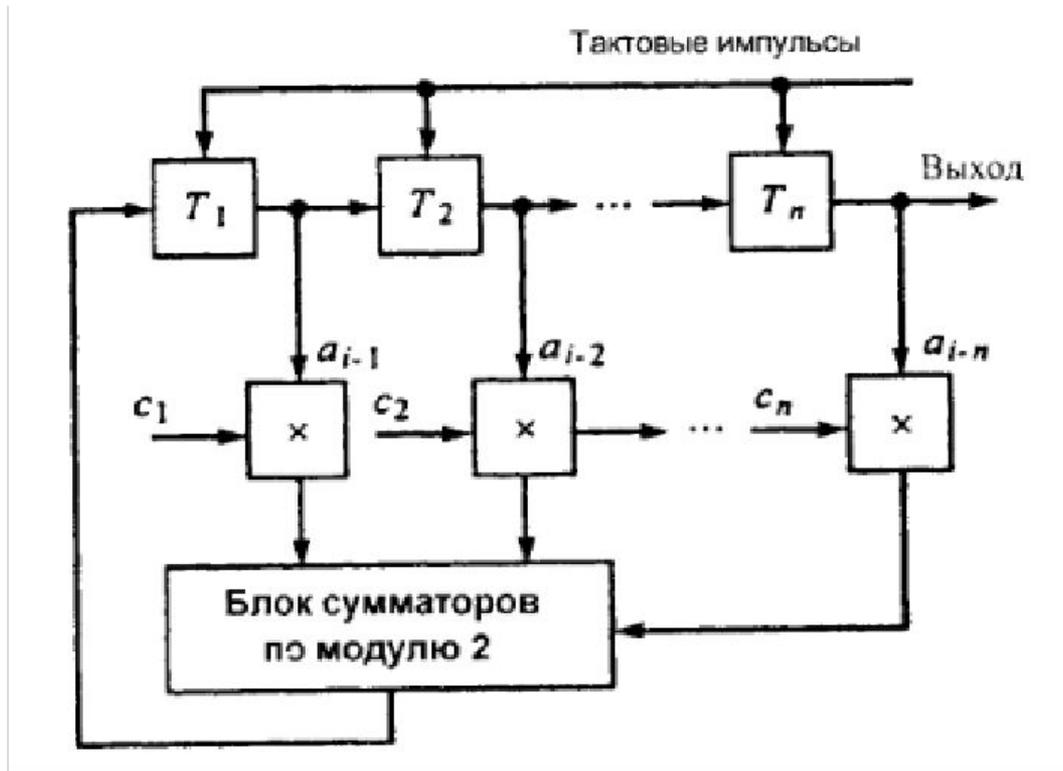


Рис. 7. Структурная схема генератора М-последовательности.

Учитывая, что число различных n -значных двоичных комбинаций равно 2^n и что комбинация из n нулей не содержится в ЛРП, период ЛРП $m < 2^n - 1$. Линейные рекуррентные последовательности с максимальным периодом $2^n - 1$ называются М- последовательностями.

Необходимым и достаточным условием существования М-последовательности является примитивность многочлена

$$f(x) = c_n x^n \oplus c_{n-1} x^{n-1} \oplus \dots \oplus c_1 x \oplus 1$$

где коэффициенты c_1, c_2, \dots, c_n те же, что и в (47). Многочлен $f(x)$ называется примитивным, если он не раскладывается на множители и делит двучлен $x^{2^n-1} \oplus 1$ без остатка и не делит никакой двучлен $x^s \oplus 1$, где $s < 2^n - 1$.

Каждому примитивному многочлену соответствует вполне определенная М-последовательность. Поэтому число различных М-последовательностей памяти n равно числу примитивных многочленов степени n , которое определяется из соотношения $Q = \varphi(2^n - 1) / n$ (табл. 1), где $\varphi(m)$ - функция Эйлера в теории чисел, равная количеству целых чисел, включая единицу, меньших числа m и взаимно простых с ним. В табл. 2 в качестве примера приведены примитивные многочлены по одному для $n = 3, 4, \dots, 10$.

Таблица 1

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Q	1	1	2	2	6	6	18	16	48	60	176	144	630	756

Таблица 2

n	Многочлен	n	Многочлен
3	$x^3 \oplus x^2 \oplus 1$	7	$x^7 \oplus x^6 \oplus 1$
4	$x^4 \oplus x^3 \oplus 1$	8	$x^8 \oplus x^4 \oplus x^3 \oplus x^2 \oplus 1$
5	$x^5 \oplus x^3 \oplus 1$	9	$x^9 \oplus x^5 \oplus 1$
6	$x^6 \oplus x^5 \oplus 1$	10	$x^{10} \oplus x^7 \oplus 1$

Амплитудно-фазоманипулированные сигналы.

Системы сигналов $m > 2$ с можно также построить путём манипуляции какого-либо одного параметра переносчика. Число возможных дискретных значений манипулированного параметра должно быть равно m . При изменении частоты получают многочастотные сигналы, а при изменении фазы - многофазные.

Можно одновременно изменять несколько параметров переносчика, например амплитуду и фазу, частоту и фазу и т. п. В последнее время большой интерес проявляется к сигналам, полученным манипуляцией амплитуды и фазы.

Амплитудно-фазоманипулированные (АФМ) сигналы описываются как

$$s_i(t) = \sqrt{2/T_c} a_i \cos w_0 t - \sqrt{2/T_c} b_i \sin w_0 t, \quad (48)$$

$$0 \leq t \leq T_c, \quad i = 1, \dots, m$$

или как

$$s_i(t) = \sqrt{2/T_c} U_i \cos(w_0 t + \varphi_i), \quad i = 1, \dots, m \quad (49)$$

где $U_i = \sqrt{a_i^2 + b_i^2}$, $\varphi_i = \arctg(b_i / a_i)$.

Используя функции $\sqrt{2/T_c} \cos w_0 t$, $-\sqrt{2/T_c} \sin w_0 t$ в качестве базисных, сигнал $s_i(t)$ в соответствии с (48) и (49) можно рассматривать либо как двумерный вектор с координатами a_i и b_i в декартовой системе, либо как вектор с амплитудой U_i и фазой φ_i в полярной системе координат.

В принципе для каждого числа m можно построить бесконечно большое число ансамблей АФМ сигналов. Поэтому важной является задача нахождения оптимальных ансамблей.

Пользуясь геометрической трактовкой, каждому сигналу $s_i(t)$ можно поставить в соответствие некоторую область пространства сигналов, которую обычно называют собственной областью или областью правильного приема. Для дискретных сообщений вероятность правильного приема есть вероятность попадания конца вектора принятого сигнала в собственную область S_i передаваемого сигнала $s_i(t)$.

При равновероятной передаче сообщений оптимизация ансамбля заключается в таком размещении сигнальных точек, при котором собственные области S_i , $i = 1, \dots, m$ примерно одинаковы и имеют максимальный объем (площадь). Эта задача сводится в общем случае к плотнейшей укладке сфер одинакового радиуса в m -мерном пространстве сигналов или, как в рассматриваемом случае, к плотнейшей укладке окружностей на плоскости. При этом центры сфер или окружностей соответствуют сигнальным точкам.

Большинство известных ансамблей АФМ сигналов найдены эвристическим методом. На рис. 8,а показаны ансамбли сигналов на основе так называемой треугольной сети для $m = 3, 7, 19$.

Сигнальные точки лежат в вершинах правильных треугольников, а собственные области сигналов (за исключением периферийных) имеют вид правильных шестиугольников. На рис. 8,б представлены ансамбли сигналов на основе квадратной сети, а на рис. 8,в - различные варианты круговых расположений сигнальных точек. В последнем случае ансамбли обозначаются как (i_1, i_2, \dots, i_k) , где i_j - величина, равная числу сигнальных точек на j -й окружности. Радиусы окружностей r_j или отношения радиусов ($j = 1, \dots, k$) должны быть заданы.

Класс АФМ сигналов включает в себя сигналы с m -ичной фазовой манипуляцией, которые определяются как

$$s_i(t) = A_0 \cos(\omega_0 t + 2\pi i / m), \quad i = 1, \dots, m.$$

Они образуют круговую сеть с равномерным распределением точек по окружности.

Методы формирования АФМ сигналов зависят от вида ансамбля, требований к точности и быстродействию модуляторов. При задании в декартовой системе координат (рис. 9) передаваемые двоичные символы поступают на цифроаналоговые преобразователи (ЦАП) блоками длиной $k = \log_2 m$. Вырабатываемые ЦАП сигналы координатами a_i и b_i модулируют квадратурные составляющие несущего колебания в балансных модуляторах (БМ).

В схеме формирования АФМ сигналов, заданных в полярной системе координат (рис. 10), ЦАП вырабатывает из двоичных символов сигналы, используемые для модуляции несущей последовательно в фазовом (ФМ) и амплитудном (АМ) модуляторах.

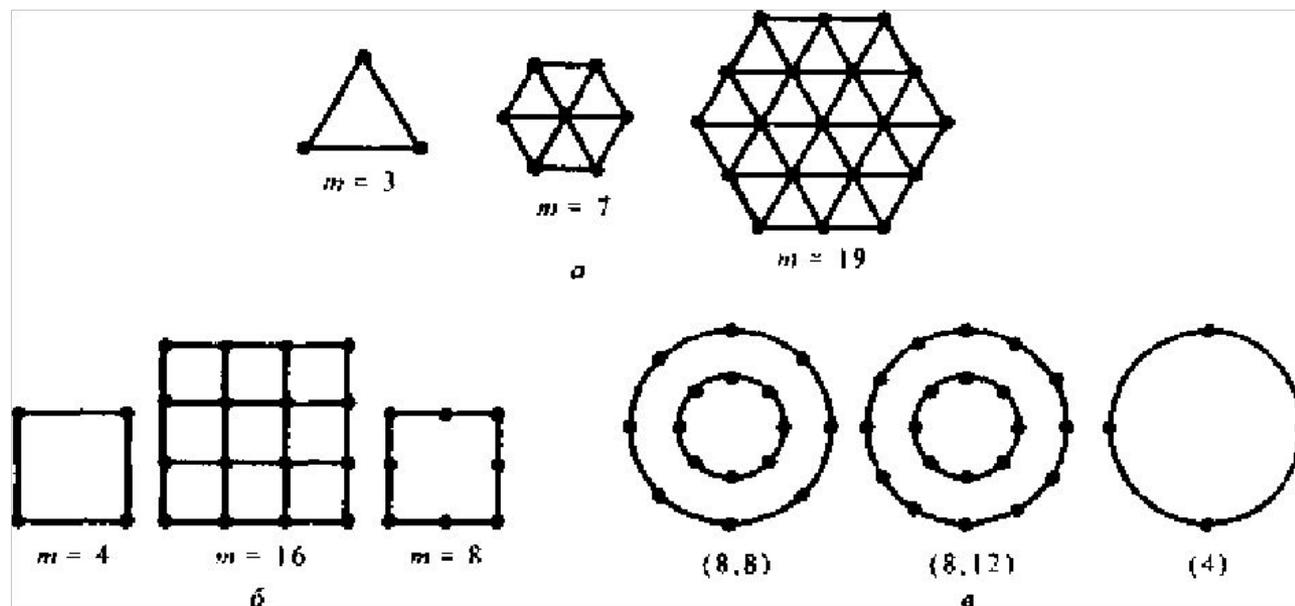


Рис. 8. Ансамбли АФМ сигналов.

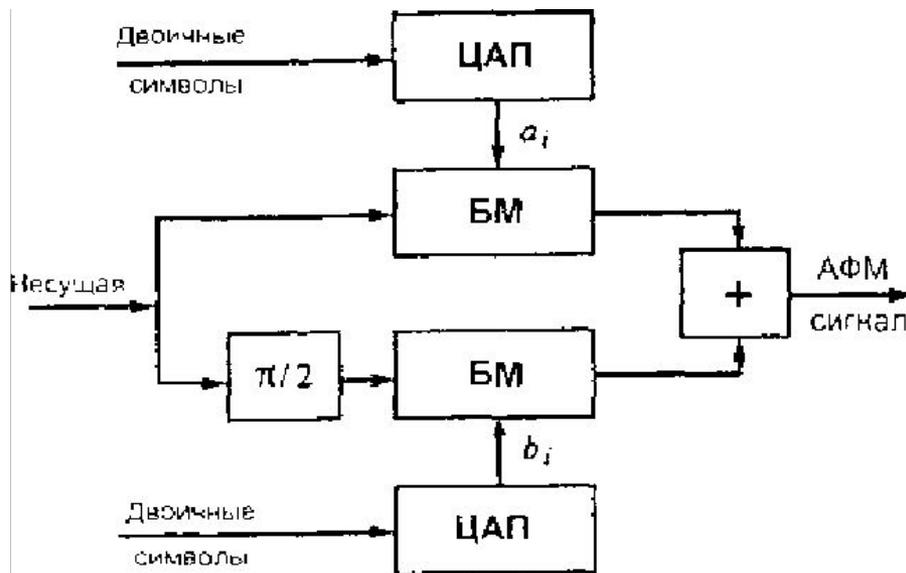


Рис. 9. Структурная схема формирователя АФМ сигналов, заданных в декартовой системе координат.

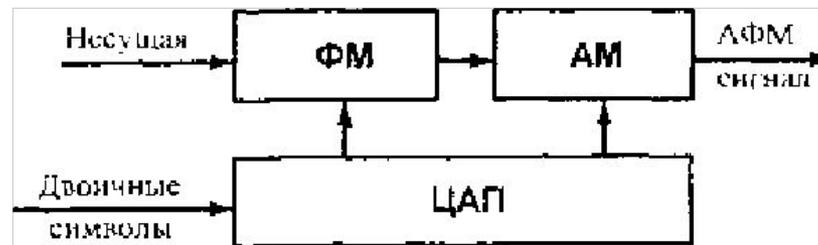


Рис. 10. Структурная схема формирователя АФМ сигналов, заданных в полярной системе координат.

Существуют и другие методы формирования АФМ сигналов. Вычисление средней вероятности ошибки

$$P_{ош} = \sum_{i=1}^m P_{ош}(s_i) p(s_i)$$

при использовании АФМ сигналов в общем случае является весьма громоздким, что обусловлено необходимостью перебора всех собственных областей S_i , $i = 1, \dots, m$, которые, как правило, имеют различную форму и сложную конфигурацию.

Решение задачи упрощается при больших отношениях сигнал/шум. При этом можно воспользоваться верхней границей для вероятности ошибки (42) или (43).

При работе системы в условиях действия гауссовского белого шума с односторонней спектральной плотностью N_0 вероятность ошибки, выраженная через расстояние $d(s_i, s_j)$, находится как

$$P_{ош}(s_j | s_i) = 1 - F \left[d(s_i, s_j) / \sqrt{2N_0} \right]$$

Тогда

$$P_{ош}(s_i) \leq \sum_{\substack{j=1, \\ j \neq i}}^m \left\{ 1 - F \left[\frac{d(s_i, s_j)}{\sqrt{2N_0}} \right] \right\}$$

Используя приближение интеграла вероятности в виде (32), можно записать

$$P_{ош}(s_l) \approx \sum_{\substack{j=1, \\ j \neq l}}^m \frac{\exp[-d^2(s_l, s_j) / 4N_0]}{\sqrt{2\pi d(s_l, s_j) / \sqrt{2N_0}}}$$

Соответственно средняя вероятность ошибки

$$P_{ош}(s_l) \approx \sum_{l=1}^m \sum_{\substack{j=1, \\ j \neq l}}^m \frac{\sqrt{2N_0} \exp[-d^2(s_l, s_j) / 4N_0]}{\sqrt{2\pi d(s_l, s_j)}} p(s_l)$$

что дает удовлетворительную точность при $P_{ош} < 0,01$.

При $m \geq 8$ системы с АФМ сигналами обладают более высокой помехоустойчивостью, чем m -ичные системы с фазовой манипуляцией. Так, при $P_{ош} = 10^{-5}$ и $m = 8$ проигрыш в средней энергии системы с фазовой манипуляцией по сравнению с системой, использующей оптимальный ансамбль сигналов, составляет 1,7 дБ, при $m=16$ -4,3 дБ, при $m=32$ -7,1 дБ, при $m=64$ -10,1 дБ, при $m=128$ -13,1 дБ. Анализ этих результатов позволяет также сделать вывод, что многие из известных ансамблей АФМ сигналов, построенных на основе треугольной и квадратной сетей, и ансамблей с круговым расположением сигнальных точек практически обеспечивают одинаковую помехоустойчивость. По крайней мере, могут быть построены различные типы систем АФМ сигналов, проигрыш которых в средней энергии по сравнению с оптимальными системами не будет превышать 0,5 дБ. Это позволяет выбирать сигналы из соображений простоты построения модулятора и демодулятора.

При выборе ансамбля сигналов необходимо иметь в виду следующее. Все многопозиционные сигналы можно разделить на два класса. К одному из них принадлежат сигналы, для которых характерно, что с увеличением объема ансамбля m растет энергетическая эффективность, но при этом расширяется полоса частот, занимаемая сигналами (снижается частотная эффективность). К этому классу относятся ортогональные, биортогональные и симплексные сигналы. При $m \gg 1$ они обеспечивают практически одинаковую помехоустойчивость и являются наилучшими. В то же время их полосы частот по сравнению с двоичными противоположными сигналами шире соответственно в $m/\log_2 m$, $m/2\log_2 m$ и $(m - 1) / \log_2 m$ раз при той же скорости передачи информации.

К другому классу принадлежат сигналы, для которых характерно, что с увеличением объема ансамбля m расстояние между сигналами уменьшается (снижается энергетическая эффективность), а полоса частот, занимаемая сигналами, не увеличивается (повышается частотная эффективность). К этому классу относятся АФМ сигналы. Очевидно, что применение АФМ сигналов требует высокой линейности и стабильности параметров приемопередающего тракта.

Системы передачи с некогерентной обработкой сигналов.

Алгоритм оптимального демодулятора.

В отличие от когерентного приема в данном случае начальная фаза высокочастотного заполнения сигнала полагается неизвестной и случайной. Подобная ситуация действительно часто имеет место из-за условий распространения радиосигнала, а также использования некоторых методов формирования сигналов. С другой стороны, для упрощения приемного устройства часто оказывается целесообразным не измерять начальную фазу даже тогда, когда она не является случайной и сохраняет свое значение на протяжении большого числа передаваемых сигналов. Таким образом, сигнал на входе приемника

$$u(t) = s_r(t, \varphi_r) + n(t),$$

Где $s_r(t, \varphi_r) = \operatorname{Re} \xi_r(t) e^{-j\varphi}$; φ_r - случайная начальная фаза; $\xi_r(t)$ - аналитический сигнал, соответствующий вещественному $s_r(t)$. Представим $s_r(t, \varphi_r)$ в виде

$$s_r(t, \varphi_r) = \\ = \operatorname{Re} \xi_r(t) \cos \varphi_r + \operatorname{Im} \xi_r(t) \sin \varphi_r = s_r(t) \cos \varphi_r + \hat{s}_r(t) \sin \varphi_r,$$

где $\hat{s}_r(t)$ - преобразование по Гильберту $s_r(t)$.

В рассматриваемых условиях входящий в функционал отношения правдоподобия интеграл

$$\int_0^{T_c} u(t) s_r(t, \varphi_r) dt = V_r \cos(\varphi_r + \phi_r),$$

Где

$$V_r = \sqrt{\left[\int_0^{T_c} u(t) s_r(t) dt \right]^2 + \left[\int_0^{T_c} u(t) \tilde{s}_r(t) dt \right]^2},$$

$$\operatorname{tg} \phi_r = \frac{- \int_0^{T_c} u(t) \tilde{s}_r(t) dt}{\int_0^{T_c} u(t) s_r(t) dt}$$

Таким образом, функционал отношения правдоподобия $\Lambda(u)$ приобретает форму

$$\Lambda(u | \varphi_1, \varphi_2) = \frac{\exp\left(-\frac{E_1}{N_0}\right) \exp\left[\frac{2}{N_0} V_1 \cos(\varphi_1 + \phi_1)\right]}{\exp\left(-\frac{E_2}{N_0}\right) \exp\left[\frac{2}{N_0} V_2 \cos(\varphi_2 + \phi_2)\right]}$$

Положим распределение величины φ равномерным, т. е. $w(\varphi) = 1/2\pi, 0 \leq \varphi \leq 2\pi$. Тогда, усредняя числитель и знаменатель функционала отношения правдоподобия по всем значениям фаз φ_1 и φ_2 , соответственно имеем

$$\bar{\Lambda}(u | \varphi_1, \varphi_2) = \frac{\exp\left(-\frac{E_1}{N_0}\right) \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{2}{N_0} V_1 \cos(\varphi_1 + \phi_1)\right] d\varphi_1}{\exp\left(-\frac{E_2}{N_0}\right) \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{2}{N_0} V_2 \cos(\varphi_2 + \phi_2)\right] d\varphi_2}$$

Положим распределение величины φ равномерным, т. е. $w(\varphi) = 1/2\pi, 0 \leq \varphi \leq 2\pi$. Тогда, усредняя числитель и знаменатель функционала отношения правдоподобия по всем значениям фаз φ_1 и φ_2 , соответственно имеем

$$\bar{\Lambda}(u | \varphi_1, \varphi_2) = \frac{\exp\left(-\frac{E_1}{N_0}\right) \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{2}{N_0} V_1 \cos(\varphi_1 + \phi_1)\right] d\varphi_1}{\exp\left(-\frac{E_2}{N_0}\right) \frac{1}{2\pi} \int_0^{2\pi} \exp\left[\frac{2}{N_0} V_2 \cos(\varphi_2 + \phi_2)\right] d\varphi_2}$$

Учитывая, что

$$\frac{1}{2\pi} \int_0^{2\pi} \exp[\alpha \cos(\varphi + \phi)] d\varphi = l_0(\alpha)$$

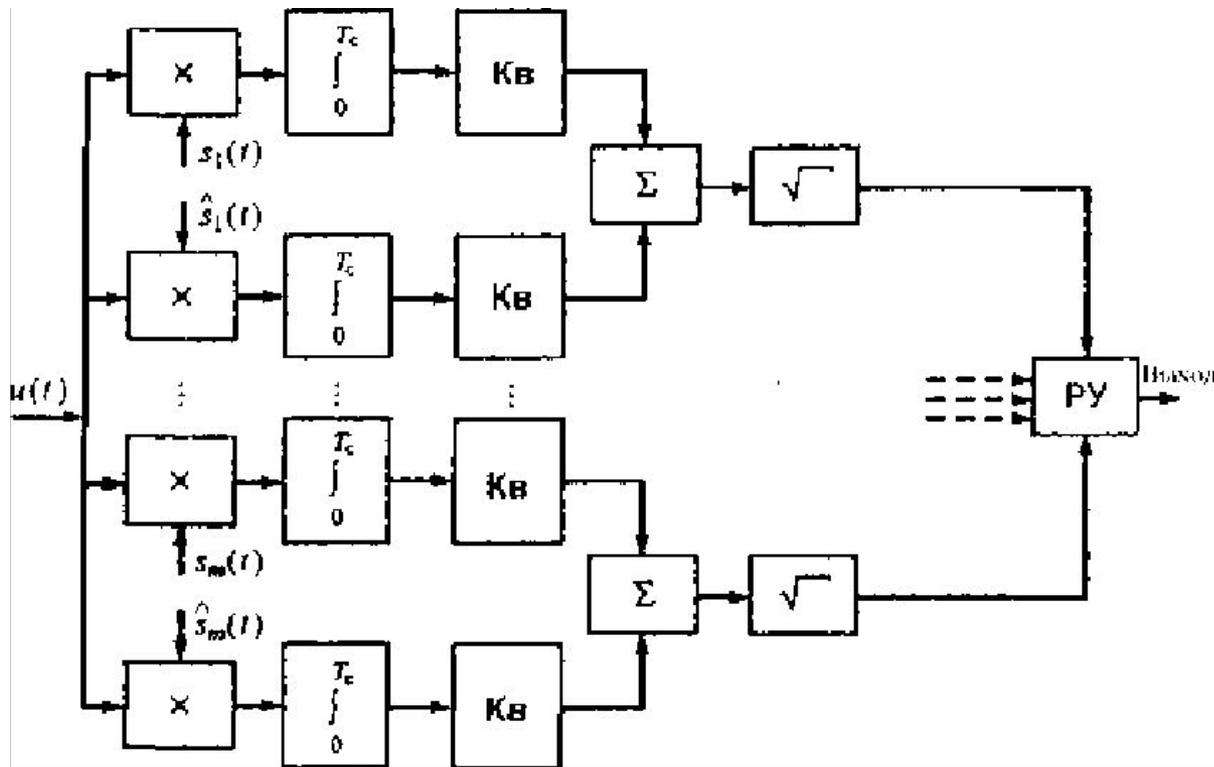
где $l_0(\alpha)$ - модифицированная функция Бесселя первого рода нулевого порядка,

$$\text{получаем } \bar{\Lambda}(u) = \exp\left(-\frac{E_1}{N_0}\right) l_0\left(\frac{2V_1}{N_0}\right) / \exp\left(-\frac{E_2}{N_0}\right) l_0\left(\frac{2V_2}{N_0}\right).$$

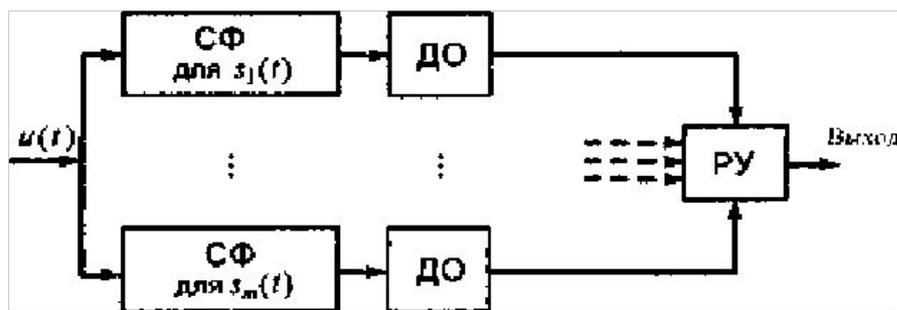
Тогда алгоритм оптимального демодулятора записывается в виде: регистрируется сигнал $s_1(t)$, если

$$\ln l_0\left(\frac{2V_1}{N_0}\right) - \ln l_0\left(\frac{2V_2}{N_0}\right) \geq \frac{E_1 - E_2}{N_0} \quad (50)$$

При $E_1 = E_2$ условие (50) принимает вид $V_1 \geq V_2$.



(а)



(б)

Рис. 11. Структурные схемы некогерентных демодуляторов m сигналов с использованием корреляторов (а) и согласованных фильтров (б).

В общем случае $m > 2$, полагая, что сигналы равновероятны, алгоритм оптимального некогерентного приема аналогично (19) можно представить в форме: регистрируется $s_l(t)$, если

$$\ln l_0 \left(\frac{2V_l}{N_0} \right) - \frac{E_l}{N_0} \geq \ln l_0 \left(\frac{2V_r}{N_0} \right) - \frac{E_r}{N_0},$$

Где $l, r = 1, \dots, m; l \neq r$.

Как и при когерентном приеме, алгоритм (52) существенно упрощается, когда $E_r = E$ для всех $r = 1, \dots, m$. При этом неравенства (52) приобретают вид

$$V_l \geq V_r, \quad l, r = 1, \dots, m; l \neq r.$$

Реализация алгоритма (53) возможна на основе корреляторов (рис. 5.11,а) и согласованных фильтров (рис. 11,б). На рисунках обозначены: Кв - квадратичное устройство и ДО - детектор огибающей.

Потенциальная помехоустойчивость. Двоичные системы передачи информации.

Пусть сигнал на входе приемника имеет вид

$$u(t) = \lambda s_1(t, \varphi_1) + (1 - \lambda) s_2(t, \varphi_2) + n(t)$$

где λ - случайная величина, принимающая значение 1 и 0 с вероятностями p_1 и p_2 соответственно, φ_1 и φ_2 - начальные фазы, представляющие собой независимые случайные величины, распределенные равномерно в интервале $[-\pi, \pi]$, $n(t)$ - помеха типа белого гауссовского шума со спектральной плотностью $N_0 / 2$.

В качестве полезных сигналов $s_i(t)$, $i=1,2$, в данном случае нельзя использовать противоположные сигналы, отличающиеся сдвигом фаз на π , так как при неизвестной начальной фазе такие сигналы неразличимы. Поэтому оценим помехоустойчивость для ортогональной системы сигналов.

Ортогональные системы с активной паузой. Рассмотрим передачу сигналов $s_1(t)$ и $s_2(t)$, ортогональных в усиленном смысле. Для них

$$\int_0^{T_c} s_1(t)s_2(t) dt = 0,$$
$$\int_0^{T_c} s_1(t)\bar{s}_2(t) dt = 0. \tag{54}$$

Примером таких сигналов являются частотно-манипулированные сигналы

$$s_1(t) = A_0 \cos(\omega_1 t + \varphi), \quad s_2(t) = A_0 \cos(\omega_2 t + \varphi),$$

где φ - произвольная начальная фаза, а частоты ω_1 и ω_2 удовлетворяют соотношениям $\omega_1 = 2\pi k_1 / T_c$, $\omega_2 = 2\pi k_2 / T_c$, k_1, k_2 - натуральные числа. Характерной особенностью ортогональных в усиленном смысле сигналов является следующее: если на вход согласованного фильтра, настроенного на сигнал $s_2(t)$ подать сигнал $s_1(t)$, то значение огибающей выходного напряжения в момент $t=T_c$, равно 0.

Исследование показывает, что ортогональные в усиленном смысле сигналы с активной паузой обеспечивают в канале с неопределенной фазой и аддитивной гауссовской помехой минимальную вероятность ошибки, т. е. они являются оптимальными для указанных условий.

Положим, что $p_1 = p_2$ и $E_1 = E_2 = E$. Пусть для определенности передается сигнал $s_1(t)$. Тогда с учетом (51) ошибка возникает, если выполняется неравенство $V_2 > V_1$.

В данном случае значение огибающей

$$V_2 = \left\{ \left[\int_0^{T_C} s_1(t)s_2(t) dt + \int_0^{T_C} n(t)s_2(t) dt \right]^2 + \left[\int_0^{T_C} s_1(t)\tilde{s}_2(t) dt + \int_0^{T_C} n(t)\tilde{s}_2(t) dt \right]^2 \right\}^{1/2}$$

Учитывая (54), получаем

$$V_2 = \left\{ \left[\int_0^{T_C} n(t)s_2(t) dt \right]^2 + \left[\int_0^{T_C} n(t)\tilde{s}_2(t) dt \right]^2 \right\}^{1/2} = (\zeta_2^2 + \tilde{\zeta}_2^2)^{1/2}$$

Случайные величины ζ_2^2 и ζ_2^2 имеют нормальное распределение с нулевым математическим ожиданием и дисперсией $\sigma^2 = EN_0 / 2$. Кроме того, математическое ожидание $M\{\zeta_2, \zeta_2\} = 0$ и, следовательно, величины ζ_2 и ζ_2 статистически независимы. Поэтому огибающая V_2 будет распределена по закону Рэлея (рис. 12):

$$w(V_2 | s_1) = \frac{2V_2}{EN_0} \exp\left(-\frac{V_2^2}{EN_0}\right) \quad (55)$$

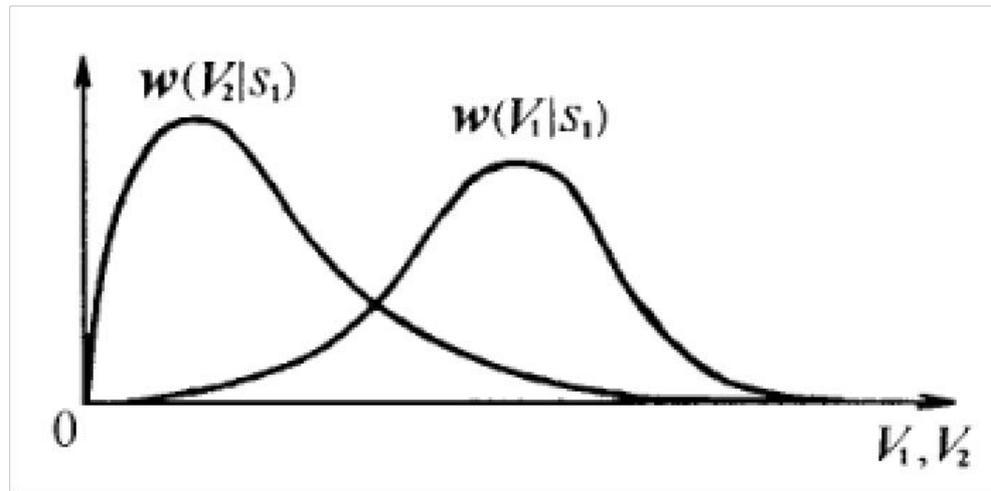


Рис. 12. Плотности сигналов вероятности огибающих на выходах согласованных фильтров (при ортогональных входных сигналах)

Значения огибающей V_1 определяется как

$$V_1 = \left\{ \left[\int_0^{T_c} s_1(t)s_1(t) dt + \int_0^{T_c} n(t)s_1(t) dt \right]^2 + \left[\int_0^{T_c} s_1(t)\tilde{s}_1(t) dt + \int_0^{T_c} n(t)\tilde{s}_1(t) dt \right]^2 \right\}^{1/2} = \left[(E + \zeta_1)^2 + \zeta_1^2 \right]^{1/2}$$

Случайные величины ζ_1 и $\tilde{\zeta}_1$, как и ζ_2 и $\tilde{\zeta}_2$, имеют нормальное распределение с нулевым математическим ожиданием и дисперсией $\sigma^2 = EN_0/2$ и являются статически независимыми. Следовательно, огибающая V_1 будет распределена по закону Райса (см. рис. 12):

$$w(V_1 | s_1) = \frac{2V_1}{EN_0} \exp\left(-\frac{V_1^2 + E^2}{EN_0}\right) I_0\left(\frac{2V_1 E}{EN_0}\right) \quad (56)$$

Нетрудно показать, что в рассматриваемом случае математические ожидания $M\{\zeta_1, \tilde{\zeta}_1\}$ и $M\{\zeta_2, \tilde{\zeta}_2\}$ равны нулю и, следовательно, случайные величины $\zeta_1, \tilde{\zeta}_1, \zeta_2$ и $\tilde{\zeta}_2$ некоррелированные, а с учетом их гауссовского распределения статически независимые. Как следствие, случайные величины V_1 и V_2 также независимы.

Поэтому вероятность ошибки при передаче сигнала $s_1(t)$ можно определить как

$$\begin{aligned}
 P_{ош}(s_2 | s_1) &= \int_0^{\infty} dV_1 \int_{V_1}^{\infty} w_2(V_1, V_2) dV_2 = \\
 &= \int_0^{\infty} w_1(V_1) \int_{V_1}^{\infty} w_1(V_2) dV_2 dV_1
 \end{aligned} \tag{57}$$

Подставляя (55) и (56) в (57), получаем

$$\begin{aligned}
 P_{ош}(s_2 | s_1) &= \int_0^{\infty} \frac{2V_1}{EN_0} \exp\left[-\frac{V_1^2 + E^2}{EN_0}\right] \times \\
 &\times I_0\left(\frac{2V_1}{N_0}\right) \int_{V_1}^{\infty} \frac{2V_2}{EN_0} \exp\left(-\frac{V_2^2}{EN_0}\right) dV_2 dV_1
 \end{aligned} \tag{58}$$

Учитывая, что внутренний интеграл равен $\exp(-V_1^2 / EN_0)$ выражение (58) можно записать в виде

$$P_{ош}(s_2 | s_1) = \int_0^{\infty} \frac{2V_1}{EN_0} \exp\left[-\frac{2V_1^2 + E^2}{EN_0}\right] \times I_0\left(\frac{2V_1}{N_0}\right) dV_1$$

Введем новую переменную $V = \sqrt{2}V_1$ и вынесем за знак интеграла множитель $\exp(-E / 2N_0)$. Тогда

$$P_{ош}(s_2 | s_1) = \frac{1}{2} \exp\left(-\frac{E}{2N_0}\right) \times \int_0^{\infty} \frac{2V}{EN_0} \exp\left[-\frac{V^2 + E^2 / 2}{EN_0}\right] I_0\left(\frac{\sqrt{2}V}{N_0}\right) dV \quad (59)$$

Подынтегральное выражение в (59) представляет собой закон Райса, а следовательно, интеграл равен 1. Таким образом,

$$P_{ош}(s_2 | s_1) = \frac{1}{2} \exp\left(-\frac{E}{2N_0}\right)$$

Учитывая симметричность канала, вероятность ошибки при передаче сигнала $s_2(t)$

$$P_{ош}(s_2 | s_1) = P_{ош}(s_1 | s_2) = \frac{1}{2} \exp\left(-\frac{E}{2N_0}\right)$$

Соответственно средняя вероятность ошибки

$$P_{ош} = \frac{1}{2} \exp\left(-\frac{E}{2N_0}\right) = \frac{1}{2} \exp\left(-\frac{h^2}{2}\right) \quad (60)$$

На рис. 5 (штриховая линия, ЧМ) показана зависимость $p_{ош} = f(E / N_0)$, вычисленная по формуле (60). Анализ показывает, что некогерентный прием ортогональных сигналов дает небольшой энергетический проигрыш по сравнению с когерентным приемом. При малых вероятностях ошибки ($p_{ош} = 10^{-4}$) он не превышает 1 дБ.

Системы с пассивной паузой.

В данном случае $s_1(t) = A_0 \cos(\omega_0 t + \varphi)$, $s_2(t) = 0$, где начальная фаза φ представляет собой случайную величину, распределенную равномерно в интервале $[-\pi, \pi]$. По-прежнему будем полагать, что $p_1 = p_2 = 0,5$.

Решение принимается на основе сравнения значения огибающей напряжения на выходе оптимального приемника (например, согласованного фильтра, настроенного на сигнал $s_1(t)$) с некоторым порогом U_{Π} . При превышении порога принимается решение в пользу сигнала $s_1(t)$, в противном случае - в пользу сигнала $s_2(t)$.

Средняя вероятность ошибки

$$\begin{aligned}
 p_{ош} &= \frac{1}{2} [p_{ош}(s_2 | s_1) + p_{ош}(s_1 | s_2)] = \\
 &= \frac{1}{2} \left[\int_0^{U_{\Pi}} w_1(V_1 | s_1) dV_1 + \int_{U_{\Pi}}^{\infty} w_1(V_2 | s_2) dV_2 \right]
 \end{aligned} \tag{61}$$

где V_1 и V_2 - значения огибающих напряжений на выходе оптимального приемника в момент $t=T_c$ при передаче сигналов $s_1(t)$ и $s_2(t)$ соответственно.

Величина V_1 распределена по закону Райса (56), а величина V_2 - по закону Рэлея (55).

Подставляя распределения огибающих V_1 и V_2 в (61), получаем

$$P_{ош} = \frac{1}{2} \left\{ \int_0^{U_{\Pi}} \frac{2V_1}{EN_0} \exp \left[-\frac{V_1^2 + E^2}{EN_0} \right] \times \right. \\ \left. \times I_0 \left(\frac{2V_1}{N_0} \right) dV_1 + \exp \left[-\frac{U_{\Pi}^2}{EN_0} \right] \right\} \quad (62)$$

Оптимальное значение порога $U_{\Pi \text{ опт}}$ находится из условия минимизации вероятности ошибки (62).

Взяв производную $\frac{dP_{ош}}{dU_{\Pi}}$ и приравняв ее к нулю, имеем

$$\frac{2U_{\Pi}}{EN_0} \exp \left[-\frac{U_{\Pi}^2 + E^2}{EN_0} \right] \times I_0 \left(\frac{2U_{\Pi}}{N_0} \right) - \exp \left[-\frac{U_{\Pi}^2}{EN_0} \right] \frac{2U_{\Pi}}{EN_0} = 0$$

или после упрощений

$$I_0(2U_{\Pi} / N_0) = \exp(E / N_0) \quad (63)$$

Прологарифмируем (63): $\ln I_0(2U_{\Pi} / N_0) = E / N_0$. Учитывая, что

$$\ln I_0(x) = \begin{cases} x, & x \gg 1, \\ x^2 / 4, & x \ll 1. \end{cases}$$

Находим:

$$U_{\text{порт}} = \begin{cases} \sqrt{EN_0} & \text{при больших отношениях сигнал / шум.} \\ \sqrt{E/N_0} & \text{при малых отношениях сигнал / шум.} \end{cases}$$

Таким образом, при больших отношениях сигнал/шум

$$P_{\text{ош}} = \frac{1}{2} \left\{ \int_0^{E/2} \frac{2V_1}{EN_0} \exp\left[-\frac{V_1^2 + E^2}{EN_0}\right] \times \right. \\ \left. \times I_0\left(\frac{2V_1}{N_0}\right) dV_1 + \frac{1}{2} \exp\left[-\frac{E}{4N_0}\right] \right\} \quad (64)$$

При $E / N_0 \geq 10$ первым слагаемым в (64) можно пренебречь. Тогда

$$p_{ош} \approx (1 / 2) \exp(-E / 4N_0). \quad (65)$$

На рис. 5.5 (штриховая линия, АМ) показана зависимость $p_{ош} = f(E / N_0)$, рассчитанная по (65). Сравнение с соответствующей кривой для когерентного случая позволяет сделать вывод, что при вероятности ошибки $10^{-3} \dots 10^{-6}$ некогерентный прием проигрывает в энергетике на 1... 0,5 дБ.

При неоптимальном пороге вероятность ошибки может оказаться значительно больше, чем (65). Поэтому при изменении уровня входящего сигнала порог приходится подстраивать, что является существенным недостатком систем с пассивной паузой.