

СТАЖ: 6 (У)

П 952

САМАРСКИЙ  
ГОСУДАРСТВЕННЫЙ  
АЭРОКОСМИЧЕСКИЙ  
УНИВЕРСИТЕТ  
имени академика  
С.П. КОРОЛЕВА

*А.Ю ПРИВАЛОВ*

# ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕЛЕКОММУНИКАЦИЙ

Раздел 1  
Введение в теорию передачи  
дискретных сообщений

САМАРА  
2003

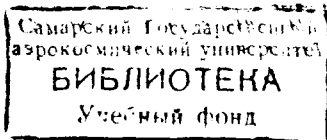
МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ  
УНИВЕРСИТЕТ имени академика С. П. КОРОЛЕВА

*А.Ю. ПРИВАЛОВ*

# ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕЛЕКОММУНИКАЦИЙ

Раздел 1  
Введение в теорию передачи  
дискретных сообщений

*Учебное пособие*



САМАРА 2003

УДК 621.39 : 519.2

Рецензенты: д-р техн. наук, проф. С.А. Прохоров  
д-р физ.-мат. наук, проф. А.И. Жданов

**Привалов А.Ю.** Теоретические основы телекоммуникаций. Разд. 1: Введение в теорию передачи дискретных сообщений: Учебное пособие / Самар. гос. аэрокосм. ун-т. Самара, 2003, 91 с.: ил. ISBN 5-7883-0244-7

Рассмотрены теория оптимального приёма дискретных сообщений, а также последовательная передача дискретных сообщений и теоремы Шеннона о надёжной передаче в каналах с шумами. Предназначено для студентов СГАУ обучающихся по специальности "Прикладная математика и информатика".

Учебное пособие опубликовано при финансовой поддержке Министерства образования Российской Федерации, администрации Самарской области и Американского фонда гражданских исследований и развития (CRDF Project SA-014-02) в рамках российско-американской программы "Фундаментальные исследования и высшее образование" (BRHE).

Печатается по решению редакционно-издательского совета Самарского государственного аэрокосмического университета.

ISBN 5-7883-0244-7 © А.Ю.Привалов, 2003  
© Самарский государственный  
аэрокосмический университет, 2003

# **ВВЕДЕНИЕ В ТЕОРИЮ ПЕРЕДАЧИ ДИСКРЕТНЫХ СООБЩЕНИЙ**

Главная проблема, которую необходимо решить в области сетей передачи данных – это обеспечение надёжной передачи. В данном разделе мы рассмотрим теоретические подходы к этой проблеме, лежащие в основе инженерных решений, относимых моделью ВОС к уровню 1 (физическому).

## **§1. Необходимые сведения из физики и теории обработки сигналов**

В современных сетях телекоммуникаций вообще, и в сетях передачи данных в частности, для переноса сигналов – носителей информации используются почти исключительно колебания электромагнитного поля и связанные с этим явления. Существующие в природе электромагнитные колебания различной частоты обладают чрезвычайно разнообразными свойствами относительно их генерации, распространения, взаимодействия с веществом и т.д. Достаточно вспомнить такие явления, как электрический ток, излучение радиоволн, СВЧ, оптические явления, рентгеновские волны и гамма-излучение. Не все из них используются в системах телекоммуникации. Используется нижняя часть спектра электромагнитных волн вплоть до оптического диапазона. Но и здесь свойства электромагнитных колебаний чрезвычайно разнообразны и также разнообразны технические решения, использующие тот или иной вид электромагнитных колебаний для переноса информации. Примерами являются проводные и кабельные линии связи, радиосвязь как наземная, так и со спутниковой компонентой, оптоволокно, инфракрасные беспроводные сети. В процессе же приёма и обработки электромагнитные сигналы преобразуются, как правило, в сигналы электрические. Поэтому наше особое внимание мы уделим общим свойствам электрических сигналов.

С самой общей точки зрения способность сигнала взаимодействовать с чем-либо и совершать какие-либо изменения в окружающем мире, одним из следствий которых

может являться перенос информации, определяется *энергией сигнала*. Из физики известно, что мгновенная мощность электрического сигнала пропорциональна квадрату напряжения, а полная энергия сигнала  $s(t)$ , представляющего изменение напряжения во времени, пропорциональна интегралу от квадрата сигнала по всему интервалу времени, когда  $s(t)$  отличен от нуля. Коэффициент пропорциональности для нас не важен, и для простоты мы будем считать, что энергия  $E_s$  сигнала  $s(t)$  равна величине этого интеграла, то есть

$$E_s = \int_{-\infty}^{+\infty} s^2(t) dt.$$

Основной вопрос, который нас сейчас интересует – это вопрос о том, что происходит с переносимой информацией сигналом при его распространении через какую-либо среду, которая в этом случае является *каналом связи*. Вопрос этот на самом деле очень сложен, и ответом на него (в том или ином смысле) занимается целый ряд специальных научных дисциплин, среди которых одной из самых важных является *теория обработки сигналов*.

Значение теории обработки сигналов только областью телекоммуникаций не исчерпывается – оно гораздо шире. Именно поэтому мы предполагаем, что читатель уже знаком с ней в объёме, по крайней мере, вводного курса, и с целью небольшого напоминания приведём далее только некоторые нужные для нас результаты.

Из теории обработки сигналов известно, что весьма широкое множество сигналов допускает представление в виде интеграла или ряда Фурье. Такое представление принято называть *спектральным разложением*, а получившуюся при этом функцию (непрерывного или дискретного аргумента) – *спектром* сигнала.

Математически спектр определён на всей оси частот от  $-\infty$  до  $\infty$ , физический же смысл имеют только неотрицательные частоты. Для всех вещественных сигналов их спектр является чётной функцией, поэтому когда говорят о преобразованиях спектра, описывают их только для неотрицательных частот, предполагая что для отрицательной части формальной оси частот над спектром

производятся аналогичные преобразования. Будем придерживаться такого соглашения и мы.

Пара преобразований Фурье (прямое и обратное) – это взаимно однозначные преобразования. В частности, это означает, что изменив спектр сигнала, например, сделав его равным нулю на некотором интервале оси частот, где он до этого нулю не был равен, мы с неизбежностью изменим и сам сигнал. Но отличие изменённого сигнала от исходного может быть достаточно малым, если на интервале, где мы сделали спектр равным нулю, его значения до этого были малыми.

Так как для всех встречающихся на практике сигналов значение функции спектра стремится к нулю при  $f \rightarrow \infty$  и  $f \rightarrow -\infty$ , то можно ввести приближённое понятие *полосы частот сигнала* – т.е. интервала оси частот, где спектр сигнала существенно отличен от нуля, и связанное с ним понятие *ширины спектра* сигнала – т.е. величины этого интервала. Понятие "существенно" здесь имеет тот смысл, что если приравнять к нулю части спектра, выходящие за полосу частот сигнала, то сигнал исказится незначительно.

Следующее понятие, которое нам надо вспомнить, – понятие *линейного четырёхполюсника*. Это объект, имеющий один вход (два полюса), куда поступает входной сигнал, и один выход (два полюса), на котором появляется отклик четырёхполюсника на входной сигнал. Свойство линейности четырёхполюсника состоит в том, что отклик на сумму сигналов на входе равен сумме откликов на каждое слагаемое из этой суммы в отдельности. При этом природа самого четырёхполюсника для возможности его математического описания роли не играет – в частности, в виде линейных четырёхполюсников могут быть представлены многие каналы связи, используемые в сетях телекоммуникаций. Также в виде линейных четырёхполюсников могут быть представлены и многие радиотехнические устройства, которые называются *линейными фильтрами*.

Из теории обработки сигналов известно, что прохождение сигнала  $s(t)$  со спектром  $S(f)$  через линейный четырёхполюсник может быть описано как умножение исходного спектра  $S(f)$  на комплексную функцию  $H(f)$ , называемую *частотной характеристикой* четырёхполюсника.

При этом получается спектр сигнала на выходе четырёхполосника:

$$S'(f) = S(f)H(f).$$

Обратное преобразование Фурье от частотной характеристики  $H(f)$  четырёхполосника называется *импульсной характеристикой* четырёхполосника (будем обозначать её  $h(t)$ ). Заметим, что  $h(t)$  равна отклику фильтра на поданную ему на вход дельта-функцию. Вид сигнала  $s'(t)$  на выходе четырёхполосника можно получить либо взяв обратное преобразование Фурье от  $S'(f)$ , либо найдя свёртку  $s(t)$  с  $h(t)$  (это следствие свойств преобразования Фурье).

Проходя через линейный четырёхполосник, представляющий, например, модель реального канала связи, сигнал в общем случае искажается. Однако можно считать, что практически все реальные каналы связи для некоторого интервала частот приближённо обладают *свойством неискажающей передачи*, т.е.

$$s'(t) \approx A s(t + \tau),$$

где  $A$  – некоторая константа, называемая *коэффициентом ослабления*, а  $\tau$  – *задержкой* сигнала. Такой интервал частот называется *полосой частот канала*.

Иногда полосу частот реального канала связи можно расширить, применяя *корректирующие фильтры*, однако на практике такое расширение не может быть бесконечным – начиная с некоторого момента такой фильтр будет слишком сложным или даже нереализуемым, поэтому будем считать, что у каждого канала связи (рассматриваемого вместе со всеми возможными корректирующими фильтрами) существует определённая полоса частот, заданная и неизменяемая.

Как правило, в сетях связи используются сигналы, полоса частот которых лежит внутри полосы частот используемого канала. Связано это с тем, что при нарушении этого правила сложность (а значит, и стоимость) системы связи, которая бы удовлетворительно работала в таких условиях, резко возрастает. Поэтому, при дальнейшем рассмотрении следующее требование всегда будем считать выполненным: по каналу связи могут передаваться

только сигналы, полоса частот которых не выходит за полосу частот канала.

На оси частот полоса частот канала может быть расположена по-разному. Обозначим нижнюю границу полосы частот канала  $f_b$ , верхнюю –  $f_u$ , а  $f_0 = (f_u + f_b)/2$  – середину полосы частот. Принято говорить, что канал является *узкополосным*, если

$$f_u - f_b \ll f_0.$$

Если же  $f_u - f_b$  того же порядка, что и  $f_0$ , то канал называется *широкополосным*.

По узкополосным каналам могут передаваться сигналы, спектр которых весь сосредоточен вблизи частоты  $f_0$ , и потому они сами похожи на синусоиду этой частоты.

По широкополосным каналам могут передаваться сигналы, в спектре которых присутствуют очень разные частоты, и потому эти сигналы могут иметь гораздо более разнообразную форму.

Теперь мы готовы перейти к тем вопросам, которые изучает *теория передачи дискретных сообщений*.

## **§2. Предмет теории передачи дискретных сообщений**

Упростим рассматриваемую нами сеть связи – исключим множество пользователей и множество каналов связи, присутствующих в сети. Пусть канал связи у нас будет один, а пользователей два – один источник информации, а другой получатель.

Рассмотрим упрощенную модель системы связи на рис. 2.1 и сформулируем исходные общие положения, лежащие в основе дальнейших построений.

Прежде всего, в основе проблемы связи вообще, и надёжной связи в частности, лежит наличие в реальных системах всевозможных случайностей. Например, если бы получатель всегда заранее точно знал, что именно и в какой момент ему должен передать источник сообщений, то вообще отсутствовала бы необходимость в системе связи. Поэтому в теории связи предполагается, что выход источника сообщений является случайным и его нельзя с



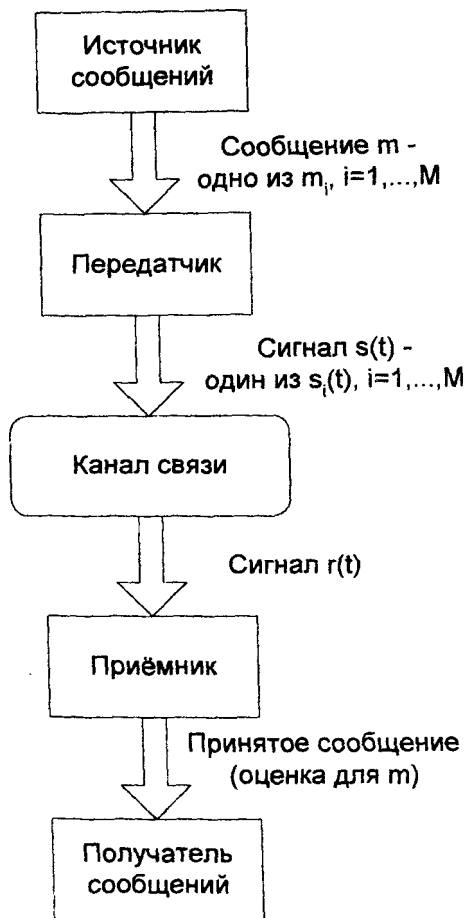


Рис. 2.1. Исходная модель системы связи

определённостью предсказать на приемном конце системы связи.

Менее очевидно то, что проблемы надёжной связи не существовало бы, если бы передаваемый сигнал не искажался случайным образом во время генерации, распространения по каналу связи и приема. Для иллюстрации этого факта представим себе, что наш передатчик может генерировать любое значение напряжения между 0 и 1 В и оно без искажений передается на приемник. Предположим, что передаваемое сообщение представляет собой текст из латинских букв, цифр, знаков препинания и некоторых специальных знаков, всего 128 возможных символов. Каждому из них можно сопоставить семизначное двоичное число, например, согласно известному всем коду ASCII. Тогда все содержимое этого текста, независимо от длины, представляется в виде двоичной последовательности, где первые семь двоичных символов – это первая буква, следующие семь – вторая и т.д. Получившуюся последовательность можно интерпретировать как двоичное число между нулем и единицей, если перед ней поставить ноль и запятую, и передать это число одним значением сигнала.

Такой способ, очевидно, практически непригоден, так как малые искажения сделают либо прием, либо передачу сигнала с такой невероятной точностью невозможными. Нетрудно также прийти из общих соображений к выводу, что чем больше сигнал подвержен неконтролируемым случайным искажениям, тем хуже условия для возможности передавать информацию с помощью этого сигнала. Поэтому одна из центральных проблем теории связи – борьба с влиянием случайных искажений на передаваемый сигнал.

Борьба со случайными искажениями может пониматься по-разному, в зависимости от того, что мы знаем о сигналах или о сообщениях, для передачи и приёма которых создана система связи. Например, если мы можем ожидать любой сигнал (это означает, что мы ничего или почти ничего о нем не знаем), то необходимо воспроизвести его на приемном конце как можно точнее. Эта задача возникает, в частности, при передаче и приеме аналоговых сигналов (в телевидении, радиовещании, аналоговой

телефонии и т.д. ).

Другая задача - когда мы почти всё о сигнале знаем, кроме нескольких его параметров, например, времени прихода и амплитуды. В этом случае нам надо понять, пришел сигнал или нет, и если пришёл, то в какой именно момент и с какой амплитудой. Это задача называется задачей обнаружения сигнала, и возникает, в частности, в радиолокации.

Третья ситуация – когда у нас есть несколько возможных сигналов, и приёмник должен сделать выбор относительно того, какой именно из этих сигналов передавался. Это и есть задача передачи и приема дискретных сообщений, возникающая, в частности, в цифровых сетях телекоммуникаций.

Для возможности содержательного анализа, определим математические модели отдельных частей системы связи, изображённой на рис.2.1.

Начнем с построения модели источника дискретных сообщений. Будем предполагать, что на выходе одного источника может появляться одно из множества  $\{m_i\}, i = 1, \dots, M$  возможных сообщений. С достаточной степенью общности можно считать, что при появлении на выходе источника сообщения  $m_i$ , передатчик передаёт в канал сигнал  $s_i(t)$ , т.е. передаваемое сообщение  $m$  однозначно соответствует сигналу  $s(t)$ :

$$m = m_i \Leftrightarrow s(t) = s_i(t).$$

Далее при передаче по каналу сигнал случайным образом искажается и вместо  $s(t)$  в приёмник приходит сигнал  $r(t)$ . Приемник по принятому сигналу  $r(t)$  должен построить оценку  $\hat{m}$  переданного сообщения. Если при этом  $\hat{m} \neq m$ , то произошла ошибка, если же  $\hat{m} = m$ , то передача прошла без ошибок.

Наша первая обсуждаемая в этой главе задача состоит в нахождении алгоритма работы *оптимального приемника*, то есть алгоритма определения  $\hat{m}$ , минимизирующего вероятность ошибки  $P_E = Pr\{\hat{m} \neq m\}$  (или, что тоже самое, максимизирующего вероятность правильного приема  $P_R = Pr\{\hat{m} = m\} = 1 - P_E$ ).

Понятно, что в общем случае оптимальный приёмник использует для принятия решения принятый сигнал  $r(t)$

и искажающие свойства канала, которые можно описать условным распределением  $Pr\{r(t) | s(t)\}$  (для всех интересующих нас  $s(t)$  и  $r(t)$ ).

К сожалению, в общем случае семейство условных распределений непрерывных сигналов  $Pr\{r(t) | s(t)\}$  чересчур сложно для содержательного анализа, и потому нам необходимо упростить нашу модель.

### §3. Переход от непрерывного канала к векторному

Предположим, что наша система связи такова, что сигнал на выходе передатчика может быть полностью описан  $N$ -мерным вектором  $\vec{s} = (s_1, \dots, s_N)$ , а сигнал, на основании которого приёмник принимает решение о том, какое именно сообщение передавалось, также описывается  $N$ -мерным вектором  $\vec{r} = (r_1, \dots, r_N)$ . Тогда искажающие свойства канала связи можно описать с помощью условного распределения  $Pr\{\vec{r} | \vec{s}\}$  — это семейство условных распределений двух векторов, что для анализа существенно проще.

Сразу же обсудим вопрос о корректности (в принципе) рассмотрения реальных каналов связи с непрерывными сигналами как каналов, передающих векторные сигналы. Прежде всего заметим, что любой сигнал из конечного набора сигналов  $\{s_i(t)\}, i = 1, \dots, M$ , может быть представлен в виде

$$s_i(t) = \sum_{k=1}^N s_{ik} \phi_k(t), N \leq M,$$

для некоторого фиксированного набора функций  $\{\phi_k(t)\}$ .

Набор функций  $\{\phi_i(t)\}, i = 1, \dots, N$ , будем называть *базисом* для системы сигналов  $\{s_i(t)\}, i = 1, \dots, M$ , а сами  $\phi_i(t)$  — базисными функциями. Очевидно, что такое представление всегда возможно, по крайней мере, для  $N = M$  (на практике же часто  $N < M$  и даже  $N \ll M$ ).

Более того, для любого конечного набора сигналов с *конечной энергией*  $\{s_i(t)\}$ , для которых

$$\int_{-\infty}^{+\infty} s_i^2(t) dt < \infty, i = 1, \dots, M,$$

всегда можно найти так называемый *ортонормальный базис*, то есть набор функций  $\{\phi_i(t)\}$ , таких что

$$\int_{-\infty}^{+\infty} \phi_j(t)\phi_k(t)dt = \delta_{jk} = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

Ортонормальные базисы обладают рядом свойств, которые понадобятся нам в дальнейшем, поэтому далее везде, где не оговорено обратное, будем считать, что базис является ортонормальным. Процедура построения такого базиса для заданного набора сигналов называется *процедурой ортогонализации Грамма-Шмидта* и в данном случае выглядит следующим образом:

Пусть задан набор  $\{s_i(t)\}$ ,  $i = 1, \dots, M$ , сигналов с ограниченной энергией.

1) Начнем с  $s_1(t)$ . Если  $s_1(t) \equiv 0$ , перенумеруем сигналы так, чтобы  $s_1(t) \rightarrow s_M(t)$ ,  $s_2(t) \rightarrow s_1(t)$  и т.д., и вернемся к началу шага 1. Если же  $s_1(t) \neq 0$ , то положим

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{E_1}}, \quad \text{где } E_1 = \int_{-\infty}^{+\infty} s_1^2(t)dt.$$

Так как  $E_1$  — энергия сигнала  $s_1(t)$ , то  $\phi_1(t)$  имеет единичную энергию и  $s_1(t) = s_{11}\phi_1(t)$ , где  $s_{11} = \sqrt{E_1}$ , то есть  $\bar{s}_1 = (\sqrt{E_1}, 0, \dots, 0)$ .

2) Пусть далее  $s_2(t) \neq 0$ . Определим

$$\theta_2(t) = s_2(t) - s_{21}\phi_1(t), \quad \text{где } s_{21} = \int_{-\infty}^{+\infty} s_2(t)\phi_1(t)dt.$$

Если  $\theta_2(t) \neq 0$ , то положим

$$\phi_2(t) = \frac{\theta_2(t)}{\sqrt{E_2}}, \quad \text{где } E_2 = \int_{-\infty}^{+\infty} \theta_2^2(t)dt,$$

тогда  $\phi_2(t)$  также имеет единичную энергию и  $s_{22} = \sqrt{E_2}$ . При этом  $\phi_1(t)$  и  $\phi_2(t)$  ортогональны, так как

$$\begin{aligned} \int_{-\infty}^{+\infty} \phi_2(t)\phi_1(t)dt &= \frac{1}{\sqrt{E_2}} \int_{-\infty}^{+\infty} \theta_2(t)\phi_1(t)dt = \\ &= \frac{1}{\sqrt{E_2}} \int_{-\infty}^{+\infty} (s_2(t) - s_{21}\phi_1(t))\phi_1(t)dt = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{E_2}} \left( \int_{-\infty}^{+\infty} s_2(t) \phi_1(t) dt - s_{21} \int_{-\infty}^{+\infty} \phi_1^2(t) dt \right) = \\
&= \frac{1}{\sqrt{E_2}} (s_{21} - s_{21}) = 0.
\end{aligned}$$

При этом  $\vec{s}_2 = (s_{21}, s_{22}, 0, \dots, 0)$ . Если же  $\theta_2(t) \equiv 0$ , то сигнал  $s_2(t)$  не порождает новой базисной функции и выражается через  $\phi_1(t)$  как  $s_2(t) = s_{21}\phi_1(t)$ , а его векторное представление в данном базисе будет  $\vec{s}_2 = (s_{21}, 0, \dots, 0)$ .

3) Пусть проведено  $l-1$  шагов процедуры ортогонализации.  $l$ -й шаг состоит в следующем: пусть по сигналам  $s_1(t), \dots, s_{k-1}(t)$  мы определили  $(l-1)$  ортонормальных функций  $\phi_1(t), \dots, \phi_{l-1}(t)$ . Понятно, что  $l \leq k$ , так как каждый новый сигнал порождает не более одной ортонормальной функции.

Пусть  $s_k(t) \neq 0$ , тогда

$$\theta_k(t) = s_k(t) - \sum_{j=1}^{l-1} s_{kj} \phi_j(t), \text{ где } s_{kj} = \int_{-\infty}^{+\infty} s_k(t) \phi_j(t) dt, j = 1, 2, \dots, l-1.$$

Если  $\theta_k(t) \neq 0$ , то положим

$$\phi_l(t) = \frac{\theta_k(t)}{\sqrt{E_k}}, \quad \text{где } E_k = \int_{-\infty}^{+\infty} \theta_k^2(t) dt.$$

Такая  $\phi_l(t)$  имеет единичную энергию, и  $s_{kl} = \sqrt{E_k}$ . Кроме того,

$$\int_{-\infty}^{+\infty} \phi_l(t) \phi_m(t) dt = 0 \quad \forall m = 1, 2, \dots, l-1,$$

так как

$$\begin{aligned}
\int_{-\infty}^{+\infty} \phi_l(t) \phi_m(t) dt &= \frac{1}{\sqrt{E_k}} \int_{-\infty}^{+\infty} \theta_k(t) \phi_m(t) dt = \\
&= \frac{1}{\sqrt{E_k}} \int_{-\infty}^{+\infty} \left( s_k(t) - \sum_{j=1}^{l-1} s_{kj} \phi_j(t) \right) \phi_m(t) dt =
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{E_2}} \left( \int_{-\infty}^{+\infty} s_k(t) \phi_m(t) dt - \sum_{j=1}^{l-1} s_{kj} \int_{-\infty}^{+\infty} \phi_j(t) \phi_m(t) dt \right) = \\
&= \frac{1}{\sqrt{E_k}} (s_{km} - \sum_{j=1}^{l-1} s_{kj} \delta_{jm}) = 0.
\end{aligned}$$

Если  $\theta_k(t) \equiv 0$ , то  $s_k(t)$  полностью раскладывается по  $(l-1)$  имеющимся функциям ортонормального базиса и не порождает новой базисной функции.

Пункт 3 продолжается до тех пор, пока не будут исчерпаны все  $s_i(t)$ . В результате будут найдены  $N \leq M$  ортонормальных функций, причем равенство будет иметь место тогда и только тогда, когда все  $M$  сигналов линейно независимы, то есть ни один из них не может быть представлен линейной комбинацией других.  $N$  называется *размерностью пространства сигналов*, образованного сигналами  $\{s_i(t)\}$ ,  $i = 1, \dots, M$ .

Итак, любой сигнал из произвольного набора из  $M$  сигналов с конечной энергией  $\{s_i(t)\}$ ,  $i = 1, \dots, M$ , может быть представлен в виде вектора  $\vec{s}_i = (s_{i1}, \dots, s_{iN})$  в некотором ортонормальном базисе.

С помощью линейных фильтров несложно построить принципиальную схему передатчика, осуществляющего синтез сигнала, представленного в виде вектора в некотором базисе  $\phi_k(t)$ ,  $i = 1, \dots, N$  (рис. 3.1). Работает схема следующим образом: сообщение  $m_i$ , которому соответствует вектор  $\vec{s}_i = (s_{i1}, \dots, s_{iN})$ , поступает от источника на генератор импульсов, который имеет  $N$  выходов. На каждый из своих выходов этот генератор выдаёт короткий импульс, являющийся практическим приближением к дельта-функции, при этом на выход  $k$  подаётся сигнал, приближённо равный  $s_{ik} \delta(t)$ . Далее этот сигнал поступает на соответствующий линейный фильтр, имеющий импульсную характеристику  $h_k(t) = \phi_k(t)$ . На выходе фильтра получается сигнал  $s_{ik} \phi_k(t)$ , а после суммирования сигналов со всех  $N$  фильтров – сигнал  $s_i(t)$ , который и передаётся в канал.

Для упрощения дальнейших рассуждений примем следующее условие: здесь и далее все базисные функции  $\phi_k(t)$

отличны от нуля только в промежутке времени  $[0, T]$  для некоторого конечного  $T$ . Тогда и все сигналы, которые можно получить в данном базисе, тоже могут быть отличны от нуля только в интервале  $[0, T]$ . На практике такое условие можно считать всегда выполненным, так как в реальных системах связи все сигналы ограничены по времени.

Пусть в нулевой момент времени из канала на вход приёмника, принципиальная схема которого изображена на рис. 3.2, начинает поступать из канала сигнал  $r(t)$ . При этом в тот же самый момент времени на каждый из присутствующих в схеме  $N$  умножителей начинает поступать соответствующая функция  $\phi_k(t)$ . Тогда на выходе соответствующего интегратора в момент времени  $T$  получится выражение

$$r_k = \int_0^T r(t)\phi_k(t)dt = \int_{-\infty}^{+\infty} r(t)\phi_k(t)dt, \quad k = 1, \dots, N. \quad (3.1)$$

Эти значения поступают на решающую схему, которая по ним должна построить оценку  $\hat{m}$  и передать её получателю. Будем считать, что набор этих чисел и образует вектор  $\vec{r}$ . Заметим, что в общем случае

$$r(t) \neq \sum_{k=1}^N r_k \phi_k(t),$$

так как из-за искажений в канале связи, сигнал  $s(t)$ , первоначально представимый в базисе  $\{\phi_k(t)\} k = 1, \dots, N$ , мог превратиться в сигнал, в этом базисе непредставимый. Тем не менее, если кроме  $\vec{r}$  решающая схема приёмника не имеет другой информации о  $r(t)$ , то мы можем считать, что из канала принимается вектор  $\vec{r}$ .

Если теперь перерисовать схему системы связи так, как изображено на рис 3.3, разделив передатчик на *векторный передатчик* и *модулятор*, а приёмник на *демодулятор* и *векторный приёмник*, то считая модулятор, канал связи и демодулятор в совокупности *векторным каналом связи*, получим систему связи с векторным каналом.



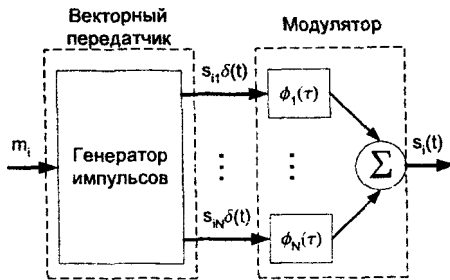


Рис. 3.1. Схема передатчика для модели векторного канала

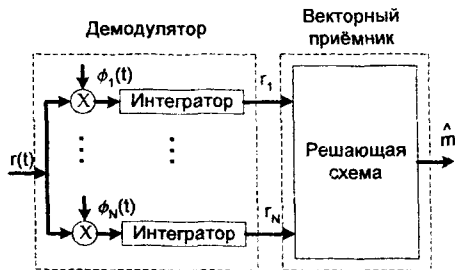


Рис. 3.2. Схема приёмника для модели векторного канала

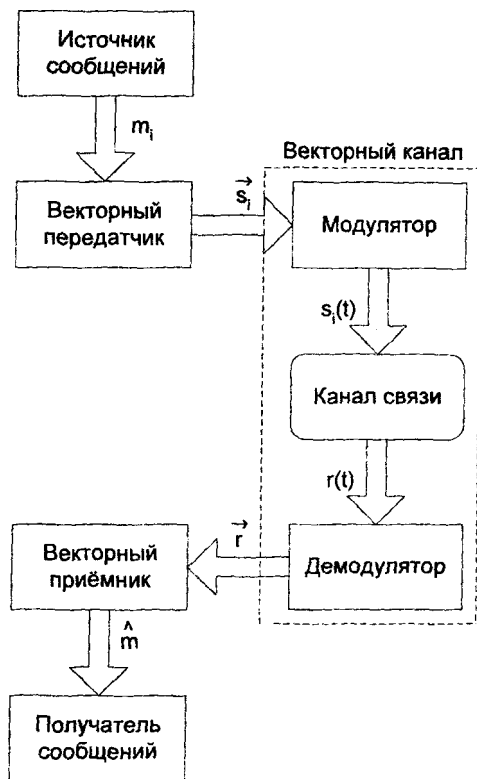


Рис. 3.3. Модель системы связи с векторным каналом

#### §4. Оптимальный приём в векторном канале

Итак, искажающие свойства векторного канала полностью определяются условной вероятностью  $Pr\{\bar{r} | \bar{s}\}$ , которая при условии заданного  $\bar{s} = \bar{s}_i$  есть просто функция от  $N$  переменных, равная величине  $Pr\{\bar{r} | \bar{s} = \bar{s}_i\} = Pr\{\bar{r} | m = m_i\}$ . Отсюда по формуле Байеса мы можем получить величину

$$Pr\{m = m_k | \bar{r}\} = \frac{Pr\{m = m_k\} Pr\{\bar{r} | m = m_k\}}{Pr\{\bar{r}\}}, \quad (4.1)$$

о которой идёт речь в следующей важной теореме:

**Теорема** (Об алгоритме оптимального приема в векторном канале): Оптимальный приемник для векторного канала должен для каждого принятого вектора  $\bar{r}$  положить, что  $\hat{m} = m_k$  всегда, когда

$$Pr\{m = m_k | \bar{r}\} > Pr\{m = m_i | \bar{r}\} \quad \forall i = 1, \dots, M, i \neq k.$$

Если для некоторых  $k$  и  $j$   $Pr\{m = m_k | \bar{r}\} = Pr\{m = m_j | \bar{r}\}$ , то приемник может положить  $\hat{m} = m_k$  или  $\hat{m} = m_j$  произвольным образом, и это не повлияет на вероятность ошибки.

Такой приемник называется *приемником максимальной апостериорной вероятности*.

**Доказательство:** Всякий раз, когда приемник принимает вектор  $\bar{r}$  и делает вывод, что  $\hat{m} = m_k$  (то есть, что передавалось сообщение  $m_k$ ), он делает правильный выбор с вероятностью  $Pr\{m = m_k | \bar{r}\}$ . Полная вероятность правильного приема есть

$$Pr\{m = m_k\} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} Pr\{m = m_k | \bar{r}\} Pr\{\bar{r}\} d\bar{r}.$$

Для приемника максимальной апостериорной вероятности согласно алгоритму его работы

$$Pr\{\hat{m} = m_i\} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \max_{1 \leq i \leq M} (Pr\{m = m_i | \bar{r}\}) Pr\{\bar{r}\} d\bar{r}.$$

Пусть теперь другой приемник при некоторых значениях  $\bar{r}$  принимает другое решение.  $Pr\{\bar{r}\}$  от этого не зависит, и поэтому подынтегральное выражение может только уменьшиться, а это не может увеличить значение интеграла, что и требовалось доказать.

Заметим, что  $Pr\{\bar{r}\}$  не зависит от решающего правила, по которому работает оптимальный приемник, и принимая во внимание взаимную однозначность  $m_k$  и  $s_k(t)$ , откуда имеем  $Pr\{\bar{r} | m = m_k\} = Pr\{\bar{r} | \bar{s} = \bar{s}_k\}$ , а также формулу (4.1), заключаем, что вместо максимума апостериорной вероятности оптимальный приёмник может пользоваться максимумом выражения

$$D_k(\bar{r}) = Pr\{m = m_k\} Pr\{\bar{r} | m = m_k\} = Pr\{m = m_k\} Pr\{\bar{r} | \bar{s} = \bar{s}_k\}.$$

Это выражение принято называть *решающей функцией* оптимального приемника.

Проиллюстрируем работу оптимального приемника с помощью графических представлений. Пусть в базисе системы сигналов имеются две базисные функции:  $\phi_1(t)$  и  $\phi_2(t)$ , и пусть имеются три возможных входных сообщения с вероятностями  $Pr\{m_1\}$ ,  $Pr\{m_2\}$ ,  $Pr\{m_3\}$ , а векторы, соответствующие им в данном базисе, есть  $\bar{s}_1 = (1, 2)$ ,  $\bar{s}_2 = (2, 1)$ ,  $\bar{s}_3 = (1, -2)$ . Изобразим это на координатной плоскости (рис. 4.1). Каждая точка этой плоскости может быть интерпретирована как некоторый возможный вектор  $\bar{r}$ .

Для оптимального приемника алгоритм работы сводится к разделению всей плоскости на некоторые области  $I_i$ ,  $i = 1, 2, 3$ , такие что каждая содержит только те точки  $\bar{r}$ , для которых  $D_i > D_j \forall j \neq i$ , и поэтому для них оптимальный приемник принимает решение о том, что  $m = m_i$ ; тогда и только тогда, когда  $\bar{r} \in I_i$  (точки, для которых  $D_i = D_j$ , образуют границу между  $I_i$  и  $I_j$ ).

Области  $I_i$  называются *оптимальными областями решения*. Оптимальный приемник совершает ошибку, когда из-за искажений в канале  $\bar{r}$  оказывается в другой области, чем был переданный  $\bar{s}_i$ .

Понятно, что вся эта конструкция легко обобщается на случай произвольной размерности базиса и любого числа сигналов.

В соответствии с видом решающей функции  $D_k = Pr\{m = m_k\} Pr\{\bar{r} | \bar{s} = \bar{s}_k\}$  конкретный вид областей ре-

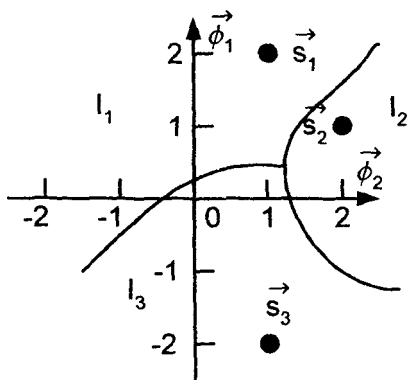


Рис. 4.1. Графическое представление оптимального приёма в векторном канале

нения зависит, во-первых, от априорных вероятностей  $Pr\{m = m_k\}$ , во-вторых, от расположения векторов сигналов  $\{\vec{s}_i\}$  в пространстве, и в-третьих, от свойств канала, описываемых  $Pr\{\vec{r} | \vec{s}\}$ . В общем случае вычисление границ таких областей – весьма сложная задача, однако для некоторых частных случаев она упрощается. Рассмотрим одну из таких ситуаций.

### §5. Канал с аддитивным белым гауссовским шумом

Вернёмся к системе связи, изображённой на рис. 2.1. Предположим, что искажающее действие канала связи на передаваемый сигнал  $s(t)$  сводится к прибавлению к нему некоторого случайного сигнала  $n(t)$ , так что

$$r(t) = s(t) + n(t). \quad (5.1)$$

Сигнал  $n(t)$  в таком случае принято называть *аддитивной помехой* или *аддитивным шумом*. Пусть передатчик, изображённый на рис. 3.1, передаёт в канал сигнал  $s_i(t)$ . После прохождения канала с аддитивным шумом сигнал поступает на вход приёмника, изображённого на рис. 3.2, тогда на вход его решающей схемы, согласно (3.1), поступит

$$r_k = \int_{-\infty}^{+\infty} r(t)\phi_k(t)dt = \int_{-\infty}^{+\infty} (s_i(t) + n(t))\phi_k(t)dt = s_{ik} + n_k.$$

где

$$n_k = \int_{-\infty}^{+\infty} n(t)\phi_k(t)dt,$$

то есть

$$\vec{r} = \vec{s}_i + \vec{n}. \quad (5.2)$$

Наличие аддитивного шума в канале связи привело к тому, что в полученном из него векторном канале появился векторный аддитивный шум – случайный вектор, прибавляющийся к вектору полезного сигнала. Далее будем считать, что  $n(t)$  статистически независим от  $s(t)$ . Тогда и  $\vec{n}$  статистически независим от  $\vec{s}$ , и, в частности,  $Pr\{\vec{n} | \vec{s}\} = Pr\{\vec{n}\}$ . Отсюда и из уравнения (5.2) следует, что

$$Pr\{\vec{r} | \vec{s} = \vec{s}_i\} = Pr\{\vec{n} = \vec{r} - \vec{s}_i | \vec{s} = \vec{s}_i\} = Pr\{\vec{n} = \vec{r} - \vec{s}_i\}.$$

Решающая функция оптимального приёма, следовательно, равна

$$D_i(\bar{r}) = Pr\{m_i\} Pr\{\bar{n} = \bar{r} - \bar{s}_i\}, \quad (5.3)$$

то есть для её вычисления необходимо знать  $Pr\{\bar{n}\}$  – совместное распределение компонент вектора шума (которое, разумеется, зависит от  $n(t)$ ).

Рассмотрим следующий частный случай: пусть  $n(t) = n_w(t)$ , где  $n_w(t)$  – это стационарный в узком смысле гауссовский случайный процесс с нулевым средним и спектральной плотностью  $S_w(f) = \frac{N_0}{2}$ ,  $-\infty < f < \infty$ . Такой шум называется *белым гауссовским шумом*.

Из свойств белого шума как случайного процесса отметим, что его корреляционная функция есть

$$R_w(\tau) = M(n_w(t + \tau)n_w(t)) = \frac{N_0}{2}\delta(\tau), \quad (5.4)$$

где  $\delta(t)$  – дельта-функция. Это следует из того, что  $S_w(f) = \frac{N_0}{2} = \frac{N_0}{2} \int_{-\infty}^{\infty} \delta(t) e^{-i2\pi f t} dt$  и взаимной однозначности преобразования Фурье.

Заметим, что с физической точки зрения белый гауссовский шум невозможен, так как его общая средняя мощность была бы равна  $\overline{n_w^2(t)} = \int_{-\infty}^{\infty} |S_w(f)|^2 df = \infty$ . Полезность введения понятия белого шума заключается в том, что это удобное для анализа приближение к ситуации, когда спектр шума гораздо шире спектра сигнала и на ширине спектра сигнала имеет приблизительно постоянное значение спектральной плотности.

Для аддитивного белого гауссовского шума

$$n_k = \int_{-\infty}^{+\infty} n_w(t)\phi_k(t)dt.$$

Прежде всего заметим, что так как все значения  $n_w(t)$  являются некоррелированными гауссовскими случайными величинами (это следует из вида корреляционной функции), то любая их взвешенная сумма или интеграл от произведения на детерминированную функцию есть тоже величина гауссовская. Поэтому все  $n_k$  – гауссовские случайные величины. Математическое ожидание любой из

них есть:

$$\begin{aligned} M(n_k) &= M\left(\int_{-\infty}^{+\infty} n_w(t)\phi_k(t)dt\right) = \int_{-\infty}^{+\infty} M(n_w(t)\phi_k(t))dt = \\ &= \int_{-\infty}^{+\infty} \phi_k(t)M(n_w(t))dt = 0. \end{aligned}$$

Найдём корреляцию этих величин. Учитывая равенство нулю их математических ожиданий, имеем

$$\begin{aligned} R(n_k, n_j) &= M(n_j n_k) = M\left(\int_{-\infty}^{+\infty} n_w(u)\phi_j(u)du \int_{-\infty}^{+\infty} n_w(v)\phi_k(v)dv\right) = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M(n_w(u)n_w(v))\phi_j(u)\phi_k(v)dvdu = \\ &= \frac{N_0}{2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(u-v)\phi_j(u)\phi_k(v)dvdu = \frac{N_0}{2} \int_{-\infty}^{+\infty} \phi_j(v)\phi_k(v)dv = \\ &= \frac{N_0}{2} \delta_{jk} \approx \begin{cases} N_0/2, & \text{если } i = j, \\ 0, & \text{если } i \neq j, \end{cases} \quad (5.5) \end{aligned}$$

где мы воспользовались видом корреляционной функции белого шума, *фильтрующим свойством* дельта-функции, состоящим в том, что

$$\int_{-\infty}^{+\infty} \delta(u-v)\phi_j(u)du = \phi_j(v)$$

и тем обстоятельством, что  $\{\phi_k(t)\}$  образуют ортонормальный базис.

Получилось, что  $n_k$ ,  $k = 1, \dots, N$ , — гауссовские некоррелированные, а значит и независимые случайные величины с нулевым средним и дисперсией  $\sigma^2 = N_0/2$ . Следовательно, их совместная плотность распределения, то есть плотность распределения вектора шума  $\vec{n} = (n_1, \dots, n_N)$ , есть

$$Pr\{\vec{n}\} = Pr\{\vec{n} = (n_1, \dots, n_N)\} = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^N n_j^2\right) =$$



$$= \frac{1}{(\pi N_0)^{N/2}} \exp\left(-\frac{|\bar{n}|^2}{N_0}\right),$$

где мы использовали тот факт, что в случае ортонормального базиса

$$|\bar{n}|^2 = (\bar{n}, \bar{n}) = \sum_{j=1}^N n_j^2.$$

Итак, наличие в исходном канале связи аддитивного белого гауссовского шума, спектральная плотность которого равна  $N_0/2$ , приводит к появлению в векторном канале аддитивного шума, представляющего собой гауссовский случайный вектор с независимыми одинаково распределёнными компонентами, имеющими нулевое математическое ожидание и дисперсию, равную  $N_0/2$ .

Подставив плотность распределения этого векторного шума в выражение (5.3) для решающей функции оптимального приёма, получим, что

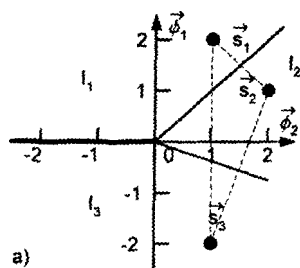
$$D_k(\bar{r}) = Pr\{m_k\} \frac{1}{(\pi N_0)^{N/2}} \exp\left(-\frac{|\bar{r} - \bar{s}_k|^2}{N_0}\right).$$

Нетрудно видеть, что максимизация этого выражения по  $k$  равносильна минимизации по  $k$  выражения

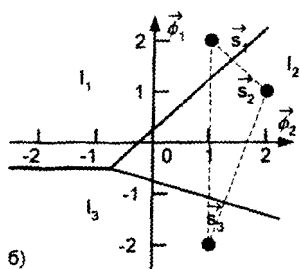
$$|\bar{r} - \bar{s}_k|^2 - N_0 \ln Pr\{m_k\},$$

которому можно дать простое геометрическое толкование. Действительно,  $|\bar{r} - \bar{s}_k|^2$  — это квадрат расстояния между точками  $\bar{r}$  и  $\bar{s}_k$ . Если все  $m_k$  имеют одинаковые вероятности  $Pr\{m_k\}$ , то оптимальное правило приема состоит в том, чтобы определить, какой  $\bar{s}_i$  ближе всего к принятому из канала  $\bar{r}$ , и для найденного таким образом  $i$  положить, что  $\hat{m} = m_i$ .

Очень легко в этом случае строятся границы решающих областей. Для равновероятных сигналов они проходят по серединным перпендикулярам к отрезкам, соединяющим две сигнальные точки. Так, для примера, рассматривавшегося ранее в общем случае на рис. 4.1, в случае векторного аддитивного гауссовского шума с независимыми компонентами, имеющими одинаковую дисперсию, решающие области имеют вид, изображённый на рис. 5.1, а.



a)



б)

Рис. 5.1. Решающие области оптимального приёма в векторном канале с аддитивным гауссовским шумом: а) равновероятные сигналы; б) неравновероятные сигналы.

Если же априорные вероятности сообщений не равны (например,  $Pr\{m_2\} > Pr\{m_1\} > Pr\{m_3\}$ ), то границы областей будут иметь вид, изображённый на рис. 5.1, б.

После того, как области решений  $I_k$  определены, можно записать выражение для условной вероятности правильного приема при условии, что передавалось определенное сообщение:

$$Pr\{\hat{m} = m \mid m = m_k\} = Pr\{\vec{r} \in I_k \mid m = m_k\} = \int_{I_k} Pr\{\vec{r} \mid \vec{s} = \vec{s}_k\} d\vec{r}.$$

Для случая аддитивного гауссовского шума с независимыми компонентами, имеющими одинаковую дисперсию, имеем

$$Pr\{\hat{m} = m \mid m = m_k\} = \int_{I_k} \frac{1}{(\pi N_0)^{N/2}} \exp\left(-\frac{|\vec{r} - \vec{s}_k|^2}{N_0}\right) d\vec{r}. \quad (5.6)$$

Полная вероятность правильного приема

$$Pr\{m = m_k\} = \sum_{k=1}^M Pr\{m_k\} Pr\{\hat{m} = m \mid m = m_k\}. \quad (5.7)$$

В дальнейшем мы рассмотрим случай, когда интегралы по решающим областям могут быть легко вычислены или аппроксимированы. А пока вернёмся к принципиальной схеме оптимального приёмника векторного канала.

### §6. Принципиальные схемы оптимального приёмника для векторного канала с аддитивным гауссовским шумом

Полученный выше алгоритм оптимального приёма позволяет построить несложную решающую схему для оптимального приёмника (рис. 6.1). В этой схеме константы  $c_i = \frac{1}{2}(N_0 \ln Pr\{m_i\} - |\vec{s}_i|^2)$  могут быть вычислены заранее, так как не зависят от  $\vec{r}$ .

Нетрудно видеть, что эта схема производит следующие действия: если в некоторый момент времени на соответствующие входы начинают подаваться  $r(t)$  и все  $\phi_k(t)$ , то схема находит  $r_k = \int_{-\infty}^{+\infty} r(t)\phi_k(t)dt$ , потом для всех  $i =$

$= 1, \dots, M$  находит  $(\bar{r}_i, \bar{s}_i)$  и, наконец, по максимуму выражения  $(\bar{r}_i, \bar{s}_i) + \frac{1}{2}(N_0 \ln Pr\{m_i\} - |\bar{s}_i|^2)$  определяет  $\hat{m}$ .

Этот алгоритм совершенно равносильен алгоритму оптимального приёма, так как

$$|\bar{r} - \bar{s}_i|^2 - N_0 \ln Pr\{m_i\} = |\bar{r}|^2 - 2(\bar{r}, \bar{s}_i) + |\bar{s}_i|^2 - N_0 \ln Pr\{m_i\},$$

где  $|\bar{r}|^2$  одно и то же для всех  $m_i$ . Таким образом, можно минимизировать

$$-2(\bar{r}, \bar{s}_i) + |\bar{s}_i|^2 - N_0 \ln Pr\{m_i\}$$

или максимизировать

$$(\bar{r}, \bar{s}_i) - \frac{1}{2}(|\bar{s}_i|^2 - N_0 \ln Pr\{m_i\}),$$

что наш приемник и делает.

Приемник на рис. 6.1. называется *корреляционным приемником*, так как сочетание перемножителя с интегратором (которое используется при получении величин  $r_k$ ), принято называть *коррелятором*.

На рис. 6.2. приведена схема оптимального приёмника для векторного канала с аддитивным гауссовским шумом, которая для получения из сигнала  $r(t)$  величин  $r_k$  использует не корреляторы, а так называемые *согласованные фильтры*.

По определению, фильтром, согласованным с сигналом  $\phi(t)$ , который отличен от нуля на промежутке  $[0, T]$ , называется фильтр с импульсной характеристикой  $h(t) = \phi(T - t)$ .

В схеме на рис. 6.2  $k$ -й фильтр согласован с сигналом  $\phi_k(t)$ . Если подать на такой фильтр сигнал  $r(t)$ , то согласно формуле вычисления сигнала на выходе фильтра с импульсной характеристикой  $h(t)$  выходной сигнал  $u(t)$  будет

$$u(t) = \int_0^t r(\tau)h(t - \tau)d\tau = \int_0^t r(\tau)\phi_k(T - t + \tau)d\tau,$$

и в момент времени  $t = T$  получится

$$u(T) = \int_0^T r(\tau)\phi(\tau)d\tau.$$

Вспоминая, что все сигналы мы условились считать отличными от нуля только на интервале  $[0, T]$ , можно считать, что интегрирование в предыдущем выражении производится по всей оси. При этих условиях в схеме на рис. 6.2 на выходе  $k$ -го фильтра в момент  $t = T$  получится величина  $r_k$ , которая и подаётся на решающую схему. Такой приёмник называется *приёмником на согласованных фильтрах*.

Для полноты картины разберём ещё один способ построения оптимального приёмника для векторного канала с аддитивным гауссовским шумом (рис. 6.3 и 6.4). В схеме на рис. 6.3 на выходе  $i$ -го коррелятора имеем

$$\begin{aligned} \int_{-\infty}^{+\infty} r(t)s_i(t)dt &= \int_{-\infty}^{+\infty} r(t)\left(\sum_{j=1}^N s_{ij}\phi_j(t)\right)dt = \\ &= \sum_{j=1}^N s_{ij} \int_{-\infty}^{+\infty} r(t)\phi_j(t)dt = \sum_{j=1}^N s_{ij}r_j = (\vec{r}, \vec{s}_i). \end{aligned}$$

Аналогично, для схемы на рис. 6.4 в момент  $t = T$  на выходе  $i$ -го фильтра тоже будем иметь  $(\vec{r}, \vec{s}_i)$  — уже скалярное произведение, которое в схемах на рис. 6.1 и 6.2 получается только после первого блока решающей схемы. Поэтому решающая схема в приёмниках на рис. 6.3 и 6.4 проще, чем на рис. 6.1 и 6.2 — в них отсутствует блок вычисления скалярного произведения. Однако это не означает, что сами приёмники, схемы которых изображены на рис. 6.3 и 6.4 проще, чем приёмники на рис. 6.1 и 6.2. Дело в том, что в схемах на рис. 6.3 и 6.4 корреляторов или фильтров  $M$  штук, а в приёмниках на рис. 6.1 и 6.2 — только  $N$  штук, а  $M \geq N$ , и на практике часто  $M \gg N$ . Так что вопрос о том, какая схема лучше, должен решаться в каждом конкретном случае отдельно.

## §7. Мультивекторные каналы

Модель векторного канала с аддитивным гауссовским шумом оказалась весьма плодотворной. Нам удалось найти и процедуру оптимального приёма для такого канала и составить принципиальные схемы устройств, её реализующих. Но остался невыясненным следующий очень

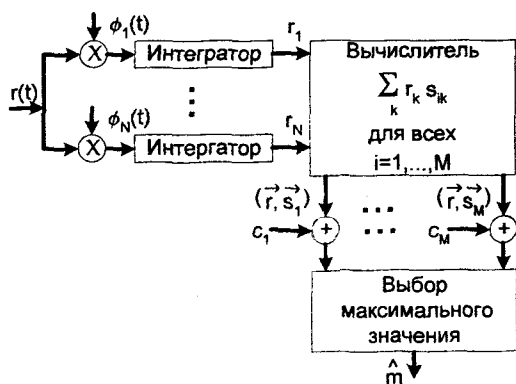


Рис. 6.1. Корреляционный приёмник

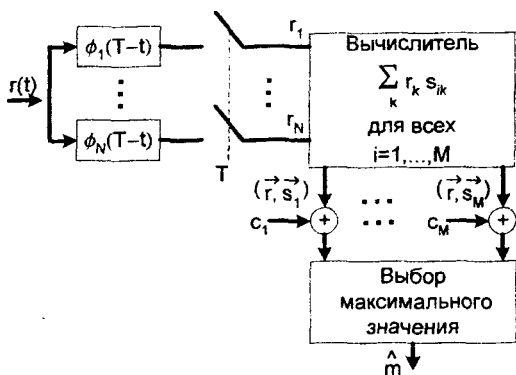


Рис. 6.2. Приёмник на согласованных фильтрах

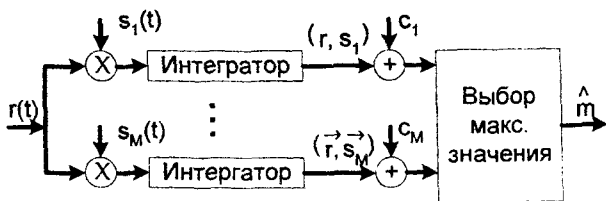


Рис. 6.3. Второй вариант схемы корреляционного приёмника

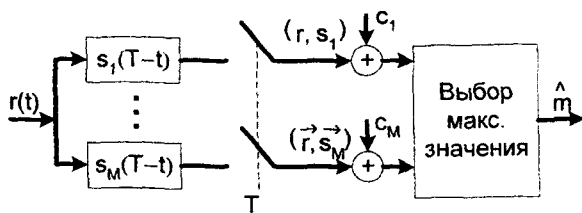


Рис. 6.4. Второй вариант схемы приёмника на согласованных фильтрах

важный вопрос: какое отношение имеет оптимальный приём в векторном канале к оптимальному приёму в исходном канале связи, не являющимся векторным? Ведь из того что приёмник максимальной апостериорной вероятности является оптимальным для векторного канала ещё не следует, что он же будет оптимальным для исходного канала связи. Может быть можно не преобразовывать исходный канал к векторному, а придумать что-то ещё, что даст для исходного канала лучшие результаты? Для того чтобы приступить к ответу на эти вопросы, нам понадобятся некоторые предварительные исследования, которые мы проведём в данном параграфе.

Итак, рассмотрим некоторое обобщение модели векторного канала. Пусть имеется система связи с несколькими одновременно действующими векторными каналами, по каждому из которых передаётся свой вектор, соответствующий передаваемому сообщению. Все эти векторы поступают в приёмник, и приёмник может принимать решение о том, какое именно сообщение передавалось на основании всех принятых векторов. Таковую систему связи будем называть *системой связи с мультивекторным каналом*.

Пусть теперь способ получения передатчиком векторов, представляющих передаваемое сообщение в каждом из каналов, значения не имеет. Эти вектора могут быть наборами коэффициентов разложения непрерывного сигнала по различным базисам, наборами значений сигнала в какие-то моменты времени и т.д. – сейчас для нас это неважно. Вопрос, который нас интересует, состоит в следующем: какую именно информацию в системе связи с мультивекторным каналом необходимо учитывать, чтобы организовать оптимальный приём?

Для простоты будем считать, что векторных каналов в системе два, как на рис. 7.1, и вектора на выходе каждого из каналов обозначим  $\vec{r}^I$  и  $\vec{r}^{II}$ . Полный вход приемника естественно описывать вектором  $\vec{r} = (\vec{r}^I, \vec{r}^{II}) = (r_1^I, r_2^I, \dots, r_K^I, r_1^{II}, \dots, r_L^{II})$ , где  $\vec{r}^I = (r_1^I, \dots, r_K^I)$ ,  $\vec{r}^{II} = (r_1^{II}, \dots, r_L^{II})$ . Апостериорная вероятность сообщения  $m_k$  при условии приема векторов  $\vec{r}^I$  и  $\vec{r}^{II}$  есть, очевидно,

$$Pr\{m = m_k | \vec{r}\} = Pr\{m = m_k | \vec{r}^I, \vec{r}^{II}\}.$$



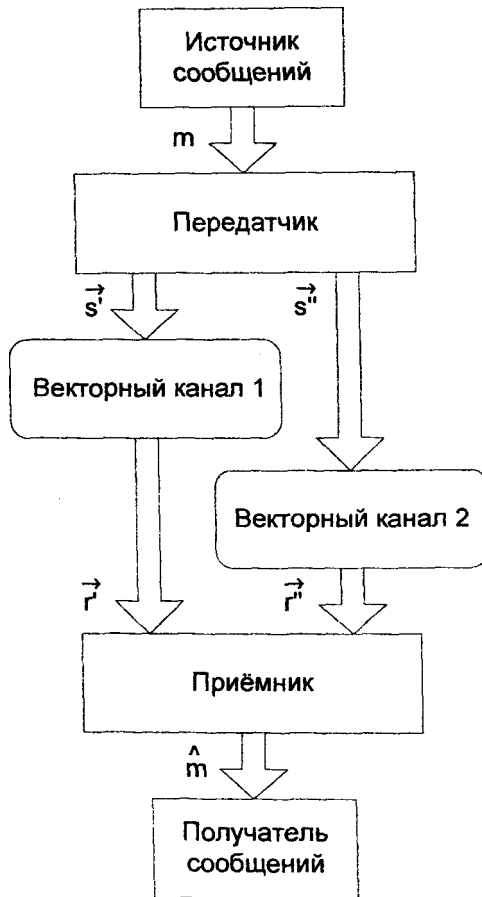


Рис. 7.1. Модель системы связи с мультивекторным каналом

Правило оптимального приема при этом запишется так: полагаем, что  $\hat{m} = m_i$ , если при  $k = i$  достигается максимум выражения

$$D_k = Pr\{m = m_k\} Pr\{\bar{r} | \bar{s} = \bar{s}_k\} = Pr\{m = m_k\} Pr\{r^i, r^j | \bar{s} = \bar{s}_k\}$$

среди всех  $k = 1, \dots, M$ .

Рассмотрим следующий вопрос: всегда ли знание обоих векторов  $r^j$  и  $r^i$  влияет на вероятность ошибки оптимального приема или есть ситуации, в которых оптимальный приемник может пренебречь, например, вектором  $r^i$ , не учитывая его в решающей функции? Условия, при которых это можно сделать, сформулируем в виде теоремы:

**Теорема (о несущественных данных):** В системе связи с мультивекторным каналом оптимальный приемник может пренебречь вектором  $r^i$  тогда и только тогда, когда

$$Pr\{r^i | r^j, \bar{s}\} = Pr\{r^i | r^j\}.$$

**Доказательство:** Итак, решающая функция имеет вид

$$D_k = Pr\{m = m_k\} Pr\{r^j, r^i | \bar{s} = \bar{s}_k\}.$$

По формуле условной вероятности имеем

$$\begin{aligned} Pr\{r^j, r^i | \bar{s} = \bar{s}_k\} &= Pr\{r^i | r^j, \bar{s} = \bar{s}_k\} Pr\{r^j | \bar{s} = \bar{s}_k\} = \\ &= Pr\{r^i, | r^j\} Pr\{r^j | \bar{s} = \bar{s}_k\}. \end{aligned}$$

Тогда решающая функция может быть записана как

$$D_k = Pr\{m = m_k\} Pr\{r^i, | r^j\} Pr\{r^j | \bar{s} = \bar{s}_k\},$$

где множитель  $Pr\{r^i | r^j\}$  не зависит от номера сообщения  $m_k$ . Он один и тот же для всех  $m_k$ , и при отыскании максимума решающей функции по  $k$  его можно не учитывать. Получается, что в решающей функции  $r^i$  не присутствует, что и требовалось доказать.

**Следствие (достаточное условие несущественности данных):** Если для системы связи с мультивекторным каналом справедливо равенство

$$Pr\{r^i | r^j, \bar{s} = \bar{s}_k\} = Pr\{r^i\},$$

то  $\vec{r}^j$  – несущественные для оптимального приёма данные.

Проиллюстрируем смысл и применимость этой теоремы на следующих трех примерах:

1) Мультивекторный канал, изображённый на рис. 7.2, а, в котором  $\vec{n}'$  и  $\vec{n}''$  независимы и не зависят от остальных векторов системы.

Как нетрудно видеть из рисунка, здесь

$$\vec{r}^j = \vec{s} + \vec{n}',$$

$$\vec{r}^j = \vec{n}''.$$

Для этого примера выполняется достаточное условие несущественности данных, а именно, поскольку  $\vec{r}^j$  не зависит ни от  $\vec{s}$ , ни от  $\vec{n}'$ , то

$$Pr\{\vec{r}^j \mid \vec{r}^j, \vec{s}\} = Pr\{\vec{r}^j\} = Pr\{\vec{n}''\},$$

то есть вектором  $\vec{r}^j$  можно пренебречь.

2) Мультивекторный канал, изображённый на рис. 7.2, б, в котором  $\vec{n}'$  и  $\vec{n}''$  независимы и не зависят от других векторов в системе.

Из рисунка следует, что

$$\vec{r}^j = \vec{s} + \vec{n}',$$

$$\vec{r}^j = \vec{r}^j + \vec{n}''.$$

Здесь достаточное условие несущественности данных не выполняется, однако условия теоремы о несущественных данных выполняются. Действительно, так как по условию

$$\vec{n}'' = \vec{r}^j - \vec{r}^j, \tag{7.1}$$

то если задано значение  $\vec{r}^j$ , для любых значений вектора  $\vec{r}^j$  справедливо

$$Pr\{\vec{r}^j \mid \vec{r}^j, \vec{s}\} = Pr\{\vec{n}'' = \vec{r}^j - \vec{r}^j \mid \vec{r}^j, \vec{s}\}.$$

Так как  $\vec{n}''$  независим, в частности, от  $\vec{s}$ , то  $\vec{s}$  из условия можно убрать (заметим, что  $\vec{r}^j$  из условия убрать нельзя,

так как только при условии задания его конкретного значения мы можем записать предыдущее равенство). Таким образом, имеем

$$Pr\{\bar{n}'' = \bar{r}'' - \bar{r}' \mid \bar{r}', \bar{s}\} = Pr\{\bar{n}'' = \bar{r}'' - \bar{r}' \mid \bar{r}'\}.$$

И ещё раз воспользуемся равенством (7.1), но теперь для обратного перехода от  $\bar{n}''$  к  $\bar{r}''$ , которое возможно опять-таки при условии задания конкретного значения  $\bar{r}'$ :

$$Pr\{\bar{n}'' = \bar{r}'' - \bar{r}' \mid \bar{r}'\} = Pr\{\bar{r}'' \mid \bar{r}'\}.$$

Таким образом, имеем

$$Pr\{\bar{r}'' \mid \bar{r}', \bar{s}\} = Pr\{\bar{r}'' \mid \bar{r}'\},$$

то есть для этого случая справедливы условия теоремы о несущественных данных, и оптимальный приёмник может пренебречь вектором  $\bar{r}'$ .

Этот результат понятен и из соображений здравого смысла: в рассматриваемой системе связи  $\bar{r}''$  — это еще более искаженный  $\bar{r}'$  и не может дать дополнительной информации об  $\bar{s}$ , поэтому им можно пренебречь.

3) Мультивекторный канал, изображённый на рис. 7.2, в, в котором  $\bar{n}'$  и  $\bar{n}''$  независимы и не зависят от других векторов в системе.

Из рисунка находим, что

$$\bar{r}' = \bar{s} + \bar{n}',$$

$$\bar{r}'' = \bar{n}' + \bar{n}''.$$

Здесь оптимальный приемник не может игнорировать  $\bar{r}''$ , так как

$$Pr\{\bar{r}'' \mid \bar{r}', \bar{s}\} = Pr\{\bar{n}'' = \bar{r}'' - \bar{r}' + \bar{s} \mid \bar{n}' = \bar{r}' - \bar{s}\},$$

то есть здесь есть явная зависимость от  $\bar{s}$  и потому  $\bar{r}''$  является существенным. Смысл этого понятен из того, что если по  $\bar{r}''$  можно хорошо оценить  $\bar{n}''$ , то значит мы можем найти  $\bar{s}$  точнее (например, если дисперсия  $\bar{n}''$  много меньше дисперсии  $\bar{n}'$ , то с большой вероятностью  $|\bar{n}''| \ll |\bar{n}'|$ , то есть  $\bar{r}''$  близок к  $\bar{n}'$  и является его хорошей оценкой).

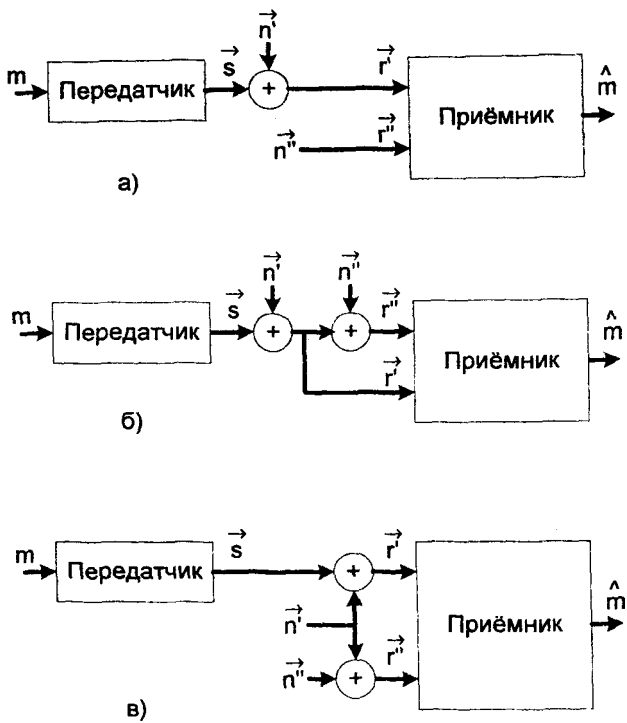


Рис. 7.2. Примеры мультивекторных каналов

## §8. Оптимальный приём в канале с аддитивным белым гауссовским шумом

Теперь мы готовы к доказательству теоремы о том, что при определённых условиях оптимальный приём в векторном канале является оптимальным и для исходного канала связи.

Рассмотрим канал связи с аддитивным белым гауссовским шумом. Как отмечалось ранее, в предположении о том что передавался сигнал  $s_i(t)$ , после прохождения через демодулятор векторного приёмника получается вектор  $\vec{r} = (r_1, \dots, r_N)$ , у которого  $r_k = \int_{-\infty}^{+\infty} r(t)\phi_k(t)dt = s_{ik} + n_k$ , где  $n_k = \int_{-\infty}^{+\infty} n_w(t)\phi_k(t)dt$ ,  $k = 1, \dots, N$ . Заметим, что так как  $n_w(t)$  не зависит от  $s(t)$ , то и  $\vec{n} = (n_1, \dots, n_N)$  не зависит от  $\vec{s} = (s_1, \dots, s_N)$ .

Ещё раз подчеркнём, что  $n_w(t) \neq \sum_{j=1}^N n_j \phi_j(t)$ , и, следовательно, преобразуя  $r(t)$  в векторном приемнике, мы потеряли часть информации о  $r(t)$ . Поэтому наша цель — доказать, что вся потерянная информация несущественна для оптимального приёма.

**Теорема:** Оптимальный приём в канале связи с аддитивным белым гауссовским шумом может быть осуществлён как оптимальный приём в векторном канале.

**Доказательство:** Пусть  $r_1(t) = \sum_{j=1}^N r_j \phi_j(t) = s(t) + n^*(t)$ , где  $s(t) = \sum_{j=1}^N s_j \phi_j(t)$ ,  $n^*(t) = \sum_{j=1}^N n_j \phi_j(t)$ .

Пусть также  $r_2(t) = r(t) - r_1(t) = (s(t) + n_w(t)) - (s(t) + n^*(t)) = n_w(t) - n^*(t)$ . Сигнал  $r_2(t)$  не зависит от передаваемого сигнала  $s(t)$  и содержит всю информацию, которую мы потеряли при векторном приеме, поэтому нам надо показать, что любые данные, получаемые из  $r_2(t)$ , несущественны для оптимального приема.

Пусть  $\vec{r}_2 = (r_2(t_1), r_2(t_2), \dots, r_2(t_m))$  — любая конечная совокупность временных отсчетов сигнала  $r_2(t)$ . Очевидно, что  $\vec{r}_2$  зависит только от  $n_w(t)$ . То есть  $\vec{r}_2$  и  $\vec{n}$  оба не зависят от  $\vec{s}$ .

Пусть  $\vec{r}_1 = \vec{s} + \vec{n}$ . Заметим, что

$$\begin{aligned} Pr\{\vec{r}_2 \mid \vec{r}_1, \vec{s}\} &= Pr\{\vec{r}_2 \mid \vec{n}, \vec{s}\} = \frac{Pr\{\vec{r}_2, \vec{n}, \vec{s}\}}{Pr\{\vec{n}, \vec{s}\}} = \\ &= \frac{Pr\{\vec{r}_2, \vec{n}\} Pr\{\vec{s}\}}{Pr\{\vec{n}\} Pr\{\vec{s}\}} = Pr\{\vec{r}_2 \mid \vec{n}\}. \end{aligned}$$

По теореме о несущественных данных оптимальный приемник может игнорировать  $\bar{r}_2$ , если  $\bar{r}_2$  не будет зависеть и от  $\bar{n}$ . Так как  $\bar{r}_2$  - это произвольный набор значений  $r_2(t)$ , то мы должны доказать независимость случайных процессов  $r_2(t)$  и  $n^*(t)$ . Тогда можно игнорировать весь  $r_2(t)$ .

Для доказательства этого факта прежде всего заметим, что процессы  $r_2(t)$  и  $n^*(t)$  являются результатами линейных операций (сложения, вычитания и интегрирования) над гауссовским процессом  $n_w(t)$ . Поэтому процессы  $r_2(t)$  и  $n^*(t)$  являются совместно гауссовскими процессами, то есть их любое совместное распределение - гауссовское.

По свойству гауссовских процессов они независимы тогда и только тогда, когда некоррелированы, то есть когда их взаимная корреляционная функция равна нулю:

$$M(n^*(s)r_2(t)) - M(n^*(s))M(r_2(t)) = 0 \quad \forall t, s.$$

Так как  $n_w(t)$  - процесс с нулевым средним, то  $r_2(t)$  и  $n^*(t)$  - тоже, поэтому нам достаточно показать, что

$$M(n^*(s)r_2(t)) = 0 \quad \forall t, s.$$

Пользуясь определением  $n^*(t)$ , имеем

$$M(n^*(s)r_2(t)) = M(r_2(t) \sum_{j=1}^N n_j \phi_j(s)) = \sum_{j=1}^N \phi_j(s) M(n_j r_2(t)).$$

Учитывая, что функции базиса  $\{\phi_j(t)\}$  линейно независимы, эта сумма может равняться нулю, только если

$$M(n_j r_2(t)) = 0 \quad \forall j, t.$$

Из определения  $r_2(t)$  имеем:

$$\begin{aligned} M(n_j r_2(t)) &= M(n_j(n_w(t) - n^*(t))) = M(n_j n_w(t) - n_j n^*(t)) = \\ &= M\left(\int_{-\infty}^{+\infty} n_w(s) \phi_j(s) ds n_w(t) - n_j \sum_{k=1}^N n_k \phi_k(t)\right) = \end{aligned}$$

$$= \int_{-\infty}^{+\infty} M(n_w(s)n_w(t))\phi_j(s)ds - \sum_{k=1}^N M(n_j;n_k)\phi_k(t). \quad (8.1)$$

Вычислим первый член этого выражения:

$$\begin{aligned} \int_{-\infty}^{+\infty} M(n_w(s)n_w(t))\phi_j(s)ds &= \int_{-\infty}^{+\infty} R_w(t-s)\phi_j(s)ds = \\ &= \frac{N_0}{2} \int_{-\infty}^{+\infty} \delta(t-s)\phi_j(s)ds = \frac{N_0}{2}\phi_j(t) \end{aligned} \quad (8.2)$$

по свойству дельта - функции (здесь мы воспользовались видом корреляционной функции белого шума).

А второй член ( $M(n_j;n_k)$ ) мы уже вычисляли - см. (5.5):  $M(n_j;n_k) = \frac{N_0}{2}\delta_{jk}$ , то есть

$$\sum_{k=1}^N M(n_j;n_k)\phi_k(t) = \frac{N_0}{2}\phi_j(t). \quad (8.3)$$

Подставив (8.2) и (8.3) в (8.1), получим

$$M(n_j;r_2(t)) = \frac{N_0}{2}\phi_j(t) - \frac{N_0}{2}\phi_j(t) = 0 \quad \forall j, s.$$

Таким образом,  $n^*(t)$  и  $r_2(t)$  независимы, поэтому  $\bar{r}_2$  и  $\bar{n}$  независимы и, следовательно, по достаточному условию несущественности данных  $\bar{r}_2$  можно игнорировать при оптимальном приеме, что и требовалось доказать.

Итак, теперь у нас есть алгоритм оптимального приема для канала связи с аддитивным белым гауссовским шумом. Этот результат очень важен не только с теоретической, но и с практической точки зрения, так как многие реальные каналы связи могут быть с хорошей степенью точности описаны либо математической моделью канала с аддитивным белым гауссовским шумом, либо близкими к ней по свойствам, но более сложными математическими моделями. Это относится к значительной части проводных и оптоволоконных каналов связи, а также некоторым радиоканалам.



## §9. Преобразования пространства сигналов, не изменяющие вероятность ошибки оптимального приёма

Теперь перейдём к решению следующей задачи: вычислению вероятности ошибки оптимального приёма в канале с аддитивным белым гауссовским шумом. С практической точки зрения это весьма важно, так как для практики нам надо знать не только то, что мы используем наилучший алгоритм приёма, но и то, какой результат от его использования мы получим в каждой конкретной ситуации.

Отметим одно важное свойство, которое мы получили при рассмотрении каналов с аддитивным белым гауссовским шумом: мы нигде не использовали ни конкретный вид сигналов  $s_i(t)$ , ни конкретный вид ортонормального базиса  $\{\phi_k(t)\}$ ,  $k = 1, \dots, N$ , в пространстве этих сигналов. Это говорит о том, что для характеристик оптимального приема (таких, например, как вероятность ошибки) важны только взаимное расположение векторов  $\bar{s}_i$ , представленных в каком-либо ортонормальном базисе, и дисперсия шума  $N_0/2$ , а вид самих сигналов не важен.

Для практики это очень важное обстоятельство. Оно позволяет за счёт выбора формы сигналов при сохранении заданной помехоустойчивости одновременно решать и другие задачи, например обеспечить простоту генерации сигналов. Часто поступают так: задаются некоторым базисом пространства сигналов, который легок в реализации, а потом уже в этом базисе выбирают сами сигналы. В качестве примера таких часто используемых на практике ортонормальных базисных сигналов можно привести следующие:

1) Неперекрывающиеся по времени прямоугольные импульсы длительности  $\tau$ , которые математически можно представить так:  $\phi_k(t) = g(t - k\tau)$ ,  $k = 1, \dots, N$ , где

$$g(t) = \begin{cases} \sqrt{\frac{1}{\tau}}, & -\tau \leq t < 0, \\ 0, & \text{иначе.} \end{cases}$$

Понятно, что такой базис может использоваться для представления кусочно-постоянных функций, изменяющих свои значения только в моменты  $k\tau$ .

2) Отрезки синусоиды с целым числом периодов на интервале  $[0, T]$ :

$$\phi_k(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi kt}{T}\right), & 0 \leq t < T, \\ 0, & \text{иначе,} \end{cases}$$

где  $k = 1, \dots, N$ .

3) Система из двух функций:

$$\phi_1(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin(2\pi f_0 t), & 0 \leq t < T, \\ 0, & \text{иначе;} \end{cases}$$

$$\phi_2(t) = \begin{cases} \sqrt{\frac{2}{T}} \cos(2\pi f_0 t), & 0 \leq t < T, \\ 0, & \text{иначе,} \end{cases}$$

где  $f_0 T$  – целое число, больше нуля. Такой базис может использоваться для представления синусоидального сигнала частоты  $f_0$  в интервале  $[0, T]$  с любыми амплитудой и фазой.

При дальнейшем анализе различных систем сигналов мы будем, как правило, иллюстрировать наше рассмотрение с помощью сигналов из пространства самого простого базиса – прямоугольных импульсов. Но не надо забывать, что базис может быть любым – от этого зависит только форма сигналов, но не свойства системы сигналов относительно помехоустойчивости оптимального приёма.

В общем случае вероятность ошибки оптимального приёма в канале с аддитивным белым гауссовским шумом дают формулы (5.6) и (5.7). Заметим, что подынтегральное выражение в (5.6) зависит только от модуля разности  $|\vec{r} - \vec{s}_i|$ , но не от самих векторов  $\vec{r}$  и  $\vec{s}_i$ , а это означает, что значение интеграла не изменяется при преобразованиях системы координат, не изменяющих значение  $|\vec{r} - \vec{s}_i|$ , а именно при сдвиге, повороте и инверсии осей. При вычислении вероятности ошибки мы будем активно пользоваться этими свойствами, так как преобразования системы координат могут значительно упростить нахождение соответствующих интегралов.

Преобразование системы координат – очень полезный приём. Рассмотрим, например, следующую интересную

для практики задачу: по одной заданной системе сигналов найти систему сигналов с таким же взаимным расположением сигнальных точек (и, соответственно, с такой же вероятностью ошибки оптимального приёма), но с наименьшей *средней энергией сигналов*  $\overline{E_m}$ , которая определяется как

$$\overline{E_m} = \sum_{i=1}^M Pr\{m_i\} E_i = \sum_{i=1}^M Pr\{m_i\} |\bar{s}_i|^2.$$

Для данного расположения сигнальных точек среднюю энергию сигнала можно минимизировать, не меняя вероятность ошибки, если вычесть из каждого сигнала постоянный вектор  $\bar{a}$ , такой чтобы  $\sum_{i=1}^M Pr\{m_i\} |\bar{s}_i - \bar{a}|^2$  была минимальной. Способ отыскания такого  $\bar{a}$  становится очевидным, если заметить, что выражение для  $\overline{E_m}$  совпадает с выражением для момента инерции системы  $M$  точек с массами  $Pr\{m_i\}$  и положением  $\bar{s}_i$ . Так как момент инерции, вычисленный относительно центра тяжести, минимален, то искомое  $\bar{a}$  есть центр тяжести:

$$\bar{a} = \sum_{i=1}^M Pr\{m_i\} \bar{s}_i.$$

Докажем это. Пусть мы сдвигаем центр на вектор  $\bar{b}$ , тогда

$$\begin{aligned} \sum_{i=1}^M Pr\{m_i\} |\bar{s}_i - \bar{b}|^2 &= \sum_{i=1}^M Pr\{m_i\} |(\bar{s}_i - \bar{a}) + (\bar{a} - \bar{b})|^2 = \\ &= \sum_{i=1}^M Pr\{m_i\} ((\bar{s}_i - \bar{a})^2 + 2(\bar{s}_i - \bar{a})(\bar{a} - \bar{b}) + (\bar{a} - \bar{b})^2) = \\ &= \sum_{i=1}^M Pr\{m_i\} |\bar{s}_i - \bar{a}|^2 + 2(\bar{a} - \bar{b}) \sum_{i=1}^M Pr\{m_i\} (\bar{s}_i - \bar{a}) + (\bar{a} - \bar{b})^2 = \\ &= \sum_{i=1}^M Pr\{m_i\} |\bar{s}_i - \bar{a}|^2 + (\bar{a} - \bar{b})^2, \end{aligned}$$

которое, очевидно, принимает минимальное значение, когда  $\bar{a} = \bar{b}$  (здесь использовалось определение  $\bar{a}$ , из которого следует, что  $\sum_{i=1}^M Pr\{m_i\}(\bar{s}_i - \bar{a}) = 0$ ).

### §10. Система из двух сигналов

Простейший случай – когда система состоит из двух сигналов. При этом удобно пользоваться системой с наименьшей максимальной энергией, и такой, что  $\bar{s}_1 - \bar{s}_2$  лежит вдоль оси  $\phi_1$  (рис. 10.1, а).

Решающие области оптимального приёма в этом случае определяются выражением:  $\min_{i=1,2} \{|\bar{r} - \bar{s}_i|^2 - N_0 \ln Pr\{m_i\}\}$ . При равных априорных вероятностях  $Pr\{m_1\} = Pr\{m_2\}$  можно рассматривать  $\min_{i=1,2} |\bar{r} - \bar{s}_i|^2$ . В этом случае граница областей решения есть прямая, равноудаленная от  $\bar{s}_1$  и  $\bar{s}_2$ , – ось  $\phi_2$ .

Нетрудно видеть, что в этом случае ошибки приема возникают тогда и только тогда, когда координата  $n_1$  вектора шума  $\bar{n} = (n_1, n_2)$  превышает  $d/2$ , где  $d$  – расстояние между сигналами:

$$Pr\{\hat{m} \neq m \mid m = m_1\} = Pr\{\bar{r} \in I_2 \mid m = m_1\} = Pr\{n_1 > \frac{d}{2}\},$$

где

$$d^2 = |\bar{s}_1 - \bar{s}_2|^2 = \int_{-\infty}^{+\infty} [s_1(t) - s_2(t)]^2 dt.$$

Так как  $n_1$  является гауссовской величиной с нулевым средним и дисперсией  $N_0/2$ , то

$$Pr\{n_1 > \frac{d}{2}\} = \int_{\frac{d}{2}}^{+\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{x^2}{N_0}\right) dx.$$

Введём определение:

$$Q(y) = \int_y^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Функция  $Q(y)$  называется *функцией ошибок*.

Далее используем подстановку  $\gamma = x\sqrt{2/N_0}$  и получим

$$Pr\{\hat{m} \neq m \mid m = m_1\} = \int_{\frac{d/\sqrt{2}}{\sqrt{N_0/2}}}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\gamma^2}{2}\right) d\gamma = Q\left(\frac{d}{\sqrt{2N_0}}\right).$$

Для безусловной вероятности ошибки в этом случае имеем

$$Pr\{\hat{m} \neq m\} = \sum_{i=1}^2 Pr\{m_i\} Pr\{\hat{m} \neq m \mid m = m_1\} = Q\left(\frac{d}{\sqrt{2N_0}}\right).$$

Это есть вероятность ошибки оптимального приема для любой пары равновероятных сигналов, удаленных друг от друга на расстояние  $d$ , независимо от их положения в пространстве сигналов.

Если это сигналы с минимальной энергией и, следовательно, являются *противоположными*, то  $|\bar{s}_1| = |\bar{s}_2| = \sqrt{E_s}$ , так что  $d = 2\sqrt{E_s}$  и

$$Pr\{\hat{m} \neq m\} = Q\left(\sqrt{\frac{2E_s}{N_0}}\right).$$

Для двух *ортогональных* сигналов с амплитудой  $\sqrt{E_s}$  расстояние между ними  $d = \sqrt{2E_s}$  и  $Pr\{\hat{m} \neq m\} = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$ .

Так как  $Q(x)$  – функция убывающая, а  $\frac{2E_s}{N_0} > \frac{E_s}{N_0}$ , то  $Q\left(\frac{2E_s}{N_0}\right) < Q\left(\frac{E_s}{N_0}\right)$ , то есть при той же энергии сигналов система связи с противоположными сигналами обеспечивает лучшую помехоустойчивость, чем система с ортогональными сигналами. Расчеты показывают, что при больших отношениях сигнал/шум разница примерно в 2 раза.

В случае, когда априорные вероятности сообщений  $m_1$  и  $m_2$  не равны (рис. 10.1, б), граница решающих областей выводится из уравнения

$$|\bar{r} - \bar{s}_1|^2 - N_0 \ln Pr\{m_1\} = |\bar{r} - \bar{s}_2|^2 - N_0 \ln Pr\{m_2\},$$

то есть

$$\left(r_1 - \frac{d}{2}\right)^2 - N_0 \ln Pr\{m_1\} = \left(r_1 + \frac{d}{2}\right)^2 - N_0 \ln Pr\{m_2\},$$

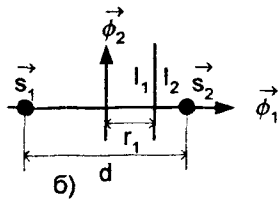
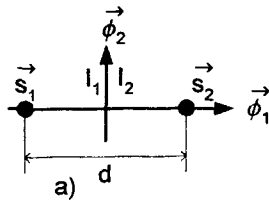


Рис. 10.1. Система из двух сигналов:  
 а) равновероятные сигналы;  
 б) неравновероятные сигналы.

откуда

$$-r_1 d - N_0 \ln Pr\{m_1\} = r_1 d - N_0 \ln Pr\{m_2\}$$

и

$$r_1 = \frac{N_0}{2d} \ln \frac{Pr\{m_2\}}{Pr\{m_1\}}.$$

Тогда вероятность ошибки будет

$$\begin{aligned} Pr\{\hat{m} \neq m\} &= \\ &= Pr\{m_1\} Pr\{\hat{m} \neq m \mid m = m_1\} + Pr\{m_2\} Pr\{\hat{m} \neq m \mid m = m_2\} = \\ &= Pr\{m_1\} Q\left(\frac{d - 2r_1}{\sqrt{2N_0}}\right) + Pr\{m_2\} Q\left(\frac{d + 2r_1}{\sqrt{2N_0}}\right). \end{aligned}$$

Теперь перейдем к разбору случаев  $M > 2$ .

### §11. Прямоугольные системы сигналов

*Прямоугольными системами сигналов* будем называть системы сигналов, у которых все сигналы расположены в вершинах прямоугольников (параллелограммов) (например, рис. 11.1).

Если при этом сигналы равновероятны, то все области решений прямоугольные. Поворотом системы координат можно сделать все границы параллельными осям и тогда многомерное интегрирование по областям решений сведется к перемножению одномерных интегралов. Например, для системы, изображенной на рис. 11.1,

$$\begin{aligned} Pr\{\hat{m} = m \mid m = m_1\} &= \int_{\vec{r} \in I_1} Pr\{\vec{r} \mid m = m_1\} d\vec{r} = \\ &= \int_{-\frac{d}{2}}^{\frac{d}{2}} Pr\{n_1\} dn_1 \int_{-\frac{d}{2}}^{+\infty} Pr\{n_2\} dn_2 = (1 - Q(\frac{d}{\sqrt{2N_0}}))^2. \end{aligned}$$

Из соображений симметрии

$$\begin{aligned} Pr\{\hat{m} = m \mid m = m_1\} &= Pr\{\hat{m} = m \mid m = m_3\} = \\ &= Pr\{\hat{m} = m \mid m = m_4\} = Pr\{\hat{m} = m \mid m = m_6\}. \end{aligned}$$

Из тех же соображений

$$\begin{aligned} Pr\{\hat{m} = m \mid m = m_2\} &= Pr\{\hat{m} = m \mid m = m_5\} = \\ &= \int_{-\frac{d}{2}}^{\frac{d}{2}} Pr\{n_1\} dn_1 \int_{-\frac{d}{2}}^{+\infty} Pr\{n_2\} dn_2 = (1 - 2Q(\frac{d}{\sqrt{2N_0}}))(1 - Q(\frac{d}{\sqrt{2N_0}})). \end{aligned}$$

Обозначив  $p = Q(\frac{d}{\sqrt{2N_0}})$ , имеем

$$Pr\{\hat{m} = m\} = \frac{4}{6}(1-p)^2 + \frac{2}{6}(1-2p)(1-p).$$

Разберем один важный частный случай прямоугольной системы сигналов: когда векторы сигналов являются вершинами гиперкуба. Будем считать, что центр гиперкуба совпадает с началом координат (для  $N = 3$  такая система изображена на рис. 11.2). При этом сигналов может быть  $M = 2^N$  (если заняты все вершины гиперкуба).

Пусть все эти сигналы равновероятны. Многомерный интеграл по какой-либо области решения есть произведение  $N$  одномерных интегралов:

$$\begin{aligned} Pr\{\hat{m} = m \mid m = m_i\} &= \int \dots \int_{I_i} Pr\{\bar{r} \mid m = m_i\} d\bar{r} = \\ &= [\int_{-\infty}^{d/2} Pr\{n_1\} dn_1]^N = (1 - Q(\frac{d}{\sqrt{2N_0}}))^N. \end{aligned}$$

(из соображений симметрии это верно для всех сигналов). Из равновероятности сигналов имеем

$$Pr\{\hat{m} = m\} = (1 - Q(\frac{d}{\sqrt{2N_0}}))^N = (1-p)^N. \quad (11.1)$$

Выразим этот результат через энергию сигнала:

$$|\bar{s}_i|^2 = N \frac{d^2}{4} = E_s \Rightarrow d = 2\sqrt{\frac{E_s}{N}} \Rightarrow p = Q\left(\sqrt{\frac{2E_s}{NN_0}}\right).$$

Можно получить такой же результат и из следующих рассуждений: в данной системе по каждой из  $N$  координат мы имеем два возможных сигнала - на расстояниях  $d/2$  и  $-d/2$  от начала координат. Так что для них  $Pr\{\hat{m} = m\} = 1 - p$ .



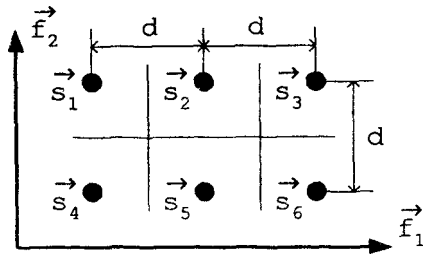


Рис. 11.1. Пример прямоугольной системы сигналов

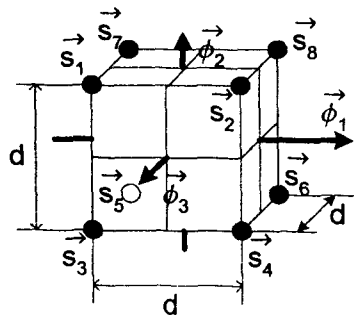


Рис. 11.2. Система сигналов в вершинах гиперкуба, случай трёх измерений

Чтобы не сделать ошибки, мы должны правильно принять все  $N$  координат, то есть  $Pr\{\hat{m} = m\} = (1-p)^N$ , так как шум и, соответственно, прием по каждой координате независимы.

Для примера приведём на рис. 11.3 вид системы сигналов в вершинах гиперкуба в базисе прямоугольных импульсов для  $M = 4$  ( $N = 2$ ). Нетрудно видеть, что в этом случае увеличение размерности пространства сигналов – это увеличение числа прямоугольных импульсов в составе сигнала. Принято говорить, что прямоугольный импульс в данном случае является *элементом* сигнала. Сигнал образуется последовательностью своих элементов, каждый из которых кодирует один бит источника.

## §12. Система ортогональных сигналов и родственные ей системы сигналов

Другим классом сигналов, для которых можно получить относительно простые формулы для минимально достижимой вероятности ошибки, является система  $M$  равновероятных ортогональных векторов равной энергии. Ей родственны системы симплексных и биортогональных сигналов.

Итак, для *ортогональных сигналов*  $M = N$  и

$$\int_{-\infty}^{+\infty} s_i(t)s_k(t)dt = (\vec{s}_i, \vec{s}_k) = E_s \delta_{ik}, \quad i, k = 1, \dots, M.$$

Пусть  $\vec{s}_i = \sqrt{E_s} \phi_i(t)$ , где  $\phi_i(t)$  – единичный вектор вдоль  $i$ -й координатной оси. Заметим, что

$$|\vec{r} - \vec{s}_i|^2 = |\vec{r}|^2 + |\vec{s}_i|^2 - 2(\vec{r}, \vec{s}_i) \sqrt{E_s} = |\vec{r}|^2 + E_s - 2r_i \sqrt{E_s},$$

где  $r_i$  –  $i$ -я составляющая вектора  $\vec{r}$ .

Значит, неравенство

$$|\vec{r} - \vec{s}_k|^2 < |\vec{r} - \vec{s}_i|^2$$

справедливо для всех  $i \neq k$  тогда и только тогда, когда

$$-2r_k \sqrt{E_s} < -2r_i \sqrt{E_s},$$

то есть  $r_i < r_k$  при всех  $i \neq k$ . Отсюда можно записать, что

$$Pr\{\hat{m} = m \mid m = m_1, r_1 = \alpha\} = Pr\{n_2 < \alpha, n_3 < \alpha, \dots, n_M < \alpha\} =$$

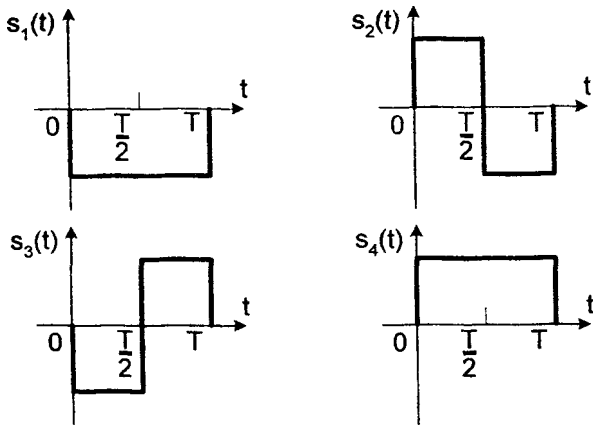


Рис. 11.3. Система из четырёх сигналов  
в вершинах гиперкуба в базе  
прямоугольных импульсов

$$= (Pr\{n_2 < \alpha\})^{M-1},$$

так как если  $m = m_1$ , то  $\vec{s} = (\sqrt{E_s}, 0, \dots, 0)$  и  $\vec{r} = (\sqrt{E_s}, 0, \dots, 0)$ , а  $n_i$  — все независимы и имеют одно и то же распределение. Принимая во внимание, что

$$Pr\{r_1 = \alpha\} = Pr\{n_1 = \alpha - \sqrt{E_s}\}$$

и интегрируя по  $n_1$ , имеем

$$\begin{aligned} & Pr\{\hat{m} = m \mid m = m_1\} = \\ & = \int_{-\infty}^{+\infty} Pr\{n_1 = \alpha - \sqrt{E_s}\} \int_{-\infty}^{+\alpha} (Pr\{n_2 = \beta\} d\beta)^{M-1} d\alpha, \end{aligned}$$

где под интегралами стоит гауссовское распределение  $Pr\{n_i = x\} = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{x^2}{N_0}\right)$ . Вследствие симметрии

$$Pr\{\hat{m} = m \mid m = m_1\} = Pr\{\hat{m} = m \mid m = m_i\} = Pr\{\hat{m} = m\},$$

так что мы получили безусловную вероятность правильного приема.

Пример ортогональной системы сигналов в базе прямоугольных импульсов приведён на рис. 12.1.

*Симплексные сигналы:* Применим идею о минимизации энергии системы сигналов к ортогональным сигналам. В соответствии с ранее выведенными равенствами для нужного сдвига  $\vec{a}$  начала координат, имеем

$$\vec{a} = \frac{1}{M} \sum_{i=1}^M \vec{s}_i = \frac{\sqrt{E_s}}{M} \sum_{i=1}^M \vec{\phi}_i.$$

Результирующая система сигналов  $\{\vec{s}_i^{\vec{a}}\} = \{\vec{s}_i - \vec{a}\}$  называется симплексной.

Так как  $\sum_{i=1}^M \vec{s}_i^{\vec{a}} = \sum_{i=1}^M \vec{s}_i - M\vec{a} = 0$ , то  $\{\vec{s}_i^{\vec{a}}\}$  линейно зависимы и потому лежат в пространстве  $N - 1$  измерения.

Энергия симплексных сигналов:

$$(\vec{s}_i^{\vec{a}}, \vec{s}_k^{\vec{a}}) = (\vec{s}_i - \vec{a}, \vec{s}_k - \vec{a}) = (\vec{s}_i - \vec{a}, \vec{s}_k - \vec{a}) - \vec{a}(\vec{s}_i + \vec{s}_k) + |\vec{a}|^2 =$$

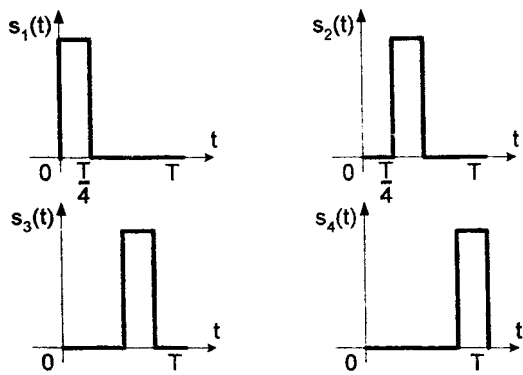


Рис. 12.1. Система из четырёх ортогональных сигналов в базисе прямоугольных импульсов

$$= E_s \delta_{ik} - 2 \frac{E_s}{M} + \frac{E_s}{M} = \begin{cases} E_s (1 - \frac{1}{M}), & \text{при } i = k, \\ -\frac{E_s}{M}, & \text{при } i \neq k. \end{cases} \quad (12.1)$$

Отсюда видно, что энергия симплексных сигналов в  $(1 - 1/M)$  раз меньше, чем ортогональных при той же вероятности ошибки (которая при сдвиге координат не меняется), для  $M = 2$  энергия в 2 раза меньше, при  $M > 2$  разница уменьшается. Равенство (12.1) можно использовать как определение симплексной системы сигналов, и любую симплексную систему сигналов можно превратить в ортогональную сдвигом на  $\sqrt{E_s/M} \vec{\phi}$ , где  $\vec{\phi}$  — единичный вектор, ортогональный ко всем  $\{\vec{s}_i\}$ .

Пример симплексной системы сигналов, полученной из ортогональной системы, изображённой на рис. 12.1, приведён на рис. 12.2.

*Биортогональные сигналы:* Система биортогональных сигналов получается из системы ортогональных сигналов присоединением к ней сигналов, противоположных по знаку исходным. Так что для нее  $M = 2N$ .

Обозначим дополнительные сигналы как  $\{-\vec{s}_i\}$ ,  $i = 1, \dots, N$ , и предположим, что энергия сигналов равна  $E_s$ . Очевидно, что если передавался  $\vec{s}_1$ , то принятый  $\vec{r}$  будет ближе к  $\vec{s}_1$ , чем к  $-\vec{s}_1$ , если  $r_1 > 0$ . Кроме того,  $\vec{r}$  ближе к  $\vec{s}_1$ , чем к любому из  $\vec{s}_i$ , образующих исходную ортогональную систему, тогда и только тогда, когда  $r_1 > r_i, \forall i \neq 1$ . И наконец,  $\vec{r}$  ближе к  $\vec{s}_1$ , чем к любому из добавленных к ортогональной системе противоположных векторов  $-\vec{s}_i$  тогда и только тогда, когда  $r_1 > -r_i, \forall i \neq 1$ . Учитывая, что при передаче  $\vec{s}_1$  на выходе канала связи получится  $r_1 = \sqrt{E_s} + n_1$ , имеем

$$\begin{aligned} & Pr\{\hat{m} = m \mid m = m_1, r_1 = \alpha\} = \\ & = Pr\{-\alpha < n_2 < \alpha, -\alpha < n_3 < \alpha, \dots, -\alpha < n_N < \alpha\} = \\ & = (Pr\{-\alpha < n_2 < \alpha\})^{N-1} = \int_{-\alpha}^{+\alpha} (Pr\{n_2 = \beta\} d\beta)^{N-1}. \end{aligned}$$

Тогда

$$\begin{aligned} & Pr\{\hat{m} = m \mid m = m_1\} = \\ & = \int_0^{+\infty} Pr\{n_1 = \alpha - \sqrt{E_s}\} \int_{-\alpha}^{+\alpha} (Pr\{n_2 = \beta\} d\beta)^{N-1} d\alpha. \end{aligned}$$

Это выражение верно и для безусловной вероятности вследствие равенства априорных вероятностей.

Пользуясь симметрией нормального распределения, получаем

$$\begin{aligned} Pr\{\hat{m} = m\} &= \\ &= \int_0^{+\infty} Pr\{n_1 = \alpha - \sqrt{E_s}\} \left(1 - 2 \int_\alpha^{+\infty} (Pr\{n_2 = \beta\} d\beta)^{M/2-1} d\alpha\right). \end{aligned}$$

Когда  $M$  и  $E_s/N_0$  велики, различие в помехоустойчивости между ортогональными и биортогональными системами сигналов мало, но  $N$  для биортогональных сигналов в два раза меньше.

На рис. 12.3 изображены сигналы, которые надо добавить к системе ортогональных сигналов на рис. 12.1, чтобы получилась система биортогональных сигналов с  $M = 8$  ( $N = 4$ ).

### §13. Аддитивная граница для вероятности ошибки

Практический расчёт вероятности ошибки оптимального приёма по формулам, полученным для ортогональной и родственным ей системам сигналов, не так уж прост. Далее мы представим метод приблизительной оценки вероятности ошибки в случае равновероятных сигналов. Он основан на том, что ошибка происходит тогда, когда при передаче  $\vec{s}_k$  на приеме вектор  $\vec{r}$  ближе к  $\vec{s}_i$ ,  $i \neq k$ , чем к  $\vec{s}_k$ . Если обозначить за событие  $A_i$  событие, когда при передаче  $\vec{s}_k$  принятый  $\vec{r}$  ближе к  $\vec{s}_i$ , чем к  $\vec{s}_k$ , то можно записать, что

$$Pr\{\hat{m} \neq m \mid m = m_k\} = Pr\{A_1 \cup A_2 \dots A_{k-1} \cup A_{k+1} \dots A_M\} \leq \sum_{\substack{i=1 \\ i \neq k}}^M Pr\{A_i\}.$$

Нетрудно видеть, что  $Pr\{A_i\}$  – это вероятность ошибки для системы из двух равновероятных сигналов  $\vec{s}_k$  и  $\vec{s}_i$  (остальные сигналы никак на нее не влияют; заметим также, что  $Pr\{A_i\} \neq Pr\{\hat{m} = m_i \mid m = m_k\}$ ). Для каналов с белым гауссовским шумом имеем  $Pr\{A_i\} = Q\left(\frac{|\vec{s}_i - \vec{s}_k|}{\sqrt{2N_0}}\right)$ .

Для полностью симметричных систем сигналов большинство слагаемых в сумме одинаковы и  $Pr\{\hat{m} \neq m\} = Pr\{\hat{m} \neq m \mid m = m_k\}$ , так что выражение еще упрощается.

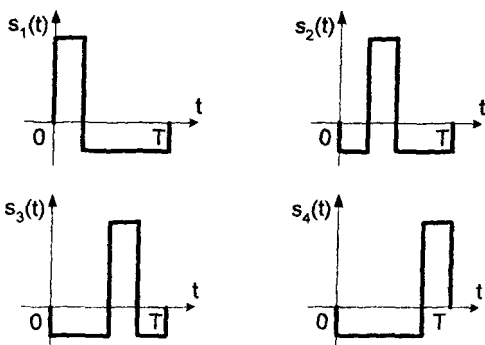


Рис. 12.2. Система из четырёх симплексных сигналов, полученная из системы ортогональных сигналов, изображённой на рис. 12.1.

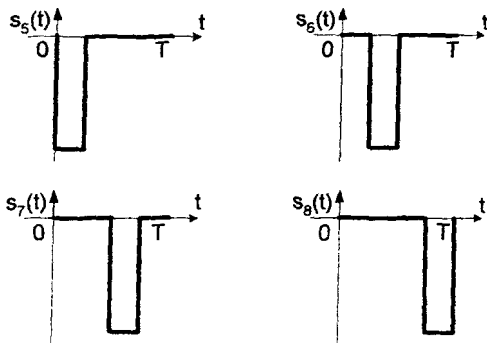


Рис. 12.3. Сигналы, которые надо добавить к системе ортогональных сигналов на рис. 12.1, чтобы получить биортогональную систему.



Так, для ортогональной системы имеем

$$Pr\{\hat{m} \neq m\} \leq (M-1)Q\left(\sqrt{\frac{E_s}{N_0}}\right),$$

для биортогональной

$$Pr\{\hat{m} \neq m\} \leq (M-2)Q\left(\sqrt{\frac{E_s}{N_0}}\right) + Q\left(\sqrt{\frac{2E_s}{N_0}}\right).$$

Во многих случаях аддитивная граница есть полезное приближение к точному значению. Расчеты показывают, что при фиксированном  $M$  точность границы растет с ростом  $E_s/N_0$ .

#### §14. Полностью симметричные системы сигналов и минимаксный приём

Как вы видели, во многих случаях, таких как бинарные сигналы ( $M=2$ ), ортогональные, симплексные, биортогональные и сигналы в вершинах гиперкуба, вычисление вероятности ошибки упрощается из-за полной симметрии геометрической конфигурации множества  $\{\bar{s}_i\}$ . Под полной симметрией мы понимаем свойство, состоящее в том, что любая перенумерация сигнальных точек может быть осуществлена поворотом, сдвигом или инверсией осей координат.

При наличии полной симметрии условие равновероятности  $m_i$ , т.е.

$$Pr\{m_i\} = \frac{1}{M} \quad \forall i,$$

приводит к конгруэнтности областей решения  $\{I_i\}$  и, таким образом, к тому, что условная вероятность правильного приема не зависит от передаваемого сигнала:

$$Pr\{\hat{m} = m \mid m = m_i\} = P^* \quad \forall i = 1, \dots, M,$$

Если такой приемник с конгруэнтными областями решения используется для случая, когда  $\{Pr\{m_i\}\}$  не равны,

то результирующая безусловная вероятность правильного приема

$$Pr\{\hat{m} = m\} = \sum_{i=1}^M Pr\{m_i\} Pr\{\hat{m} = m \mid m = m_i\} = P^*,$$

то есть такая же, как и в случае равновероятных сообщений.

Это свойство называется *инвариантностью данного приемника относительно статистики источника*. Если статистика источника известна, то помехоустойчивость может быть еще увеличена за счет оптимального выбора областей. Но если она не известна, а система сигналов полностью симметрична, то построив инвариантный относительно статистики источника приемник, мы будем знать вероятность ошибки независимо от этой статистики.

В результате можно сделать вывод о том, что приемник, оптимальный для равновероятных сигналов, является *минимаксным* для полностью симметричных систем сигналов. При фиксированном передатчике и канале для приемника вероятность ошибки зависит от статистики источника и при некотором выборе этой статистики достигает максимальной величины. Это максимальное значение  $Pr\{\hat{m} \neq m\}$  является удобным критерием качества работы приемника при отсутствии априорных сведений о вероятностях  $Pr\{m_i\}$  – это гарантированный минимум помехоустойчивости. При таком критерии приемник, для которого максимальная величина  $Pr\{\hat{m} \neq m\}$  оказывается наименьшей среди всех возможных приемников, называется *минимаксным приемником*.

Доказать, что приемник, оптимальный для равновероятных сигналов, – *минимаксный* для полностью симметричной системы  $\{\bar{x}_i\}$ , не трудно. Во-первых, вероятность ошибки такого приемника одинакова для всех статистик источника, то есть она и есть наихудшая его вероятность ошибки. Во-вторых, поскольку он оптимален для равновероятных сообщений, то любой другой приемник даст на равновероятных сообщениях большую вероятность ошибки и, значит, имеет большую наихудшую вероятность ошибки, что и требовалось доказать.

Давайте обсудим, как работает такой минимаксный приёмник для системы сигналов в вершинах гиперкуба. Он предполагает, что все вершины используемого  $N$ -мерного гиперкуба равновероятны. Именно такой случай мы рассматривали в §11, и нашли, что вероятность правильного приёма всего  $N$ -мерного сигнала есть произведение вероятностей правильного приёма по каждому измерению (см. формулу 11.1), так что приём по каждому измерению независим от приёма по другим измерениям.

Часть сигнала, которая соответствует одному измерению, мы назвали элементом сигнала, так что отмеченную выше особенность можно сформулировать так: минимаксный приёмник для данной ситуации может быть построен как *приёмник с поэлементным приёмом* (также говорят *с покомпонентным приёмом*). Например, если рассматриваемая система сигналов в вершинах гиперкуба использует базис из прямоугольных импульсов, то минимаксный приёмник может принимать решение о том, какой именно импульс передавался (с амплитудой  $+\sqrt{E_N}$  или  $-\sqrt{E_N}$ ) независимо от других импульсов, составляющих сигнал.

В отличие от минимаксного покомпонентного приёма, оптимальный приём в общем случае может осуществляться только как *приём в целом*. Для иллюстрации различий между двумя способами приёма рассмотрим следующий простой пример: пусть система сигналов является двумерным гиперкубом, но сигналы не равновероятны,  $Pr\{\bar{s}_1\} = Pr\{\bar{s}_3\} = 1/2$ ,  $Pr\{\bar{s}_2\} = Pr\{\bar{s}_4\} = 0$ , пусть передавался сигнал  $\bar{s}_1$ , и пусть из канала принят сигнал  $\bar{r}$  (рис. 14.1), тогда оптимальный приёмник, осуществляющий приём в целом, не сделает ошибки и решит, что передавался сигнал  $\bar{s}_1$ , а минимаксный приёмник, осуществляющий покомпонентный приём, решит, что передавался сигнал  $\bar{s}_2$  (хоть у этого сигнала и нулевая вероятность). Так что с точки зрения теоретической помехоустойчивости приём в целом, конечно, лучше покомпонентного.

Но у покомпонентных приёмников есть другое очень важное для практики преимущество – они гораздо проще в реализации, и потому в реальных системах телекоммуникаций общего пользования, где простота, а значит и стоимость, очень важны, они используются гораздо чаще настоящих оптимальных приёмников.

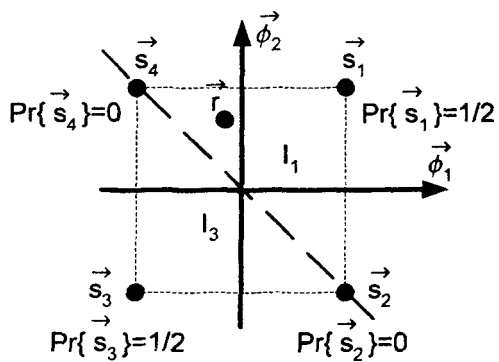


Рис. 14.1. Сравнение оптимального и минимаксного приёмников

## §15. Передача последовательностей сообщений

До сих пор в нашем анализе речь шла о передаче одного сообщения от источника к получателю, и единственное условие относительно поведения системы связи во времени, которое мы учитывали, – это условие о том, что передача одного сообщения осуществляется на промежутке времени  $[0, T]$  для некоторого  $T$ . Теперь от передачи одного единственного сообщения мы переходим к передаче последовательности сообщений, когда, передав одно сообщение, система связи сразу же переходит к передаче следующего. Понятно, что на практике системы связи так и работают.

Конечно, с теоретической точки зрения, если система связи за каждые  $T$  единиц времени передаёт одно сообщение, выбираемое из  $M$  возможных, то последовательность из  $K$  таких сообщений можно представить как передачу одного из  $M^K$  возможных сообщений, осуществляемую за  $KT$  единиц времени. Мы будем иногда использовать такой подход, но более интересна с практической точки зрения формулировка этой проблемы, при которой явно учитывался бы последовательный характер передачи сообщений. Но для этого нам необходимо ввести несколько новых понятий и пересмотреть некоторые предположения о нашей модели системы связи.

Прежде всего, определим понятие *скорости дискретного источника сообщений*, или как её ещё называют, *информационной скорости*. Пусть имеется источник сообщений, который производит последовательность дискретных двоичных символов (т.е.  $M=2$ ). Предположим, что каждый из этих символов может быть либо 0, либо 1 с равной вероятностью и независимо от всех других событий. Пусть этот источник производит за единицу времени  $R$  таких символов. Тогда один символ производится за  $1/R$ , а за время, кратное величине  $1/R$ , источник генерирует  $RT$  двоичных символов, причем каждая из  $2^{RT}$  возможных последовательностей генерируется с равной вероятностью. То есть передатчик, связанный с этим источником, должен быть способен передать одно из  $2^{RT}$  равновероятных сообщений в течение каждого из последовательных интер-

валов длины  $T$ .

В ситуациях, аналогичных рассмотренной выше, будем называть  $R$  скоростью источника и измерять ее числом двоичных символов в единицу времени (бит/с – от английского сокращения bit – Binary digiT – двоичная цифра).

Для источника, который за  $T$  единиц времени генерирует одно из  $M$  равновероятных сообщений, определим скорость источника как

$$R = \frac{1}{T} \log_2 M \text{ (бит/с)}.$$

Для примера рассмотрим источник, который каждые  $\frac{1}{R'}$  единиц времени генерирует один символ из алфавита в  $L$  символов. Если символы равновероятны и выбор каждого производится независимо, то за промежуток  $T$  источник создаст последовательность из  $R'T$  сообщений (каждое выбирается из  $L$  возможных) и эту последовательность можно рассматривать как одно из  $M = L^{R'T}$  равновероятных сообщений. Поэтому скорость источника равна  $R = \frac{1}{T} \log_2 M = R' \log_2 L$  (бит/с).

Разумность определения скорости источника через число бит в единицу времени можно понять, заметив, что совокупность  $M$  сообщений можно пронумеровать и записав их номера в двоичной форме, получить эквивалентное представление в двоичном виде. Если  $M = 2^K$ , то в полученной таким образом двоичной последовательности каждый двоичный символ независим от остальных и с равной вероятностью принимает значения 0 и 1.

Для краткости и простоты мы ограничим наше рассмотрение только вопросами передачи двоичных последовательностей такого типа. Но все, что мы получим, может быть обобщено на случай произвольного алфавита источника.

То что наше понятие информационной скорости имеет весьма полезный смысл, можно увидеть из следующего примера: пусть есть два независимых источника. Один из них за промежуток  $T$  генерирует одно из  $M_1$  равновероятных сообщений, а другой – одно из  $M_2$  равновероятных сообщений. Если каждый из источников связан с отдельным передатчиком, то первый передатчик должен передавать

со скоростью

$$R_1 = \frac{1}{T} \log_2 M_1,$$

а второй – со скоростью

$$R_2 = \frac{1}{T} \log_2 M_2.$$

Если они оба подключены к одному передатчику, то за время  $T$  этот передатчик должен передавать одно из  $M = M_1 M_2$  равновероятных сообщений, и его скорость

$$R = \frac{1}{T} \log_2 M_1 M_2 = \frac{1}{T} \log_2 M_1 + \frac{1}{T} \log_2 M_2 = R_1 + R_2.$$

Таким образом, система, способная передавать информацию со скоростью  $R$ , равноценна двум системам, передающим со скоростью  $R/2$  (заметим, что система, передающая за единицу времени одно из  $M$  равновероятных сообщений, не эквивалентна двум системам, каждая из которых может передавать за единицу времени одно из  $M/2$  сообщений).

Понятно, что нам хотелось бы, чтобы каждая система связи работала с максимальной возможной для неё информационной скоростью при сохранении, разумеется, необходимой надёжности и других своих характеристик. Какие же факторы могут ограничивать это и к чему подобные ограничения могут привести? Именно эти вопросы и будут рассмотрены далее в этом разделе.

Следующее понятие, которое необходимо ввести – *мощность передатчика*. Энергию для процесса, который может длиться произвольно долго, ограничивать не имело смысла, но если на передачу одного сообщения мы намерены тратить как можно меньше времени, то на практике придётся учесть ограничение на мощность передатчика, которую мы будем обозначать  $P_s$ . Для сигнала  $s(t)$  длительности  $T$  средняя мощность определяется как

$$P_s = \frac{1}{T} \int_0^T s^2(t) dt = \frac{E_s}{T}.$$

Это важный момент, так как в канале с аддитивным белым гауссовским шумом, как мы видели, помехоустойчивость

оптимального приёма зависит от отношения сигнал/шум, то есть от  $E_s/N_0$ . Если средняя мощность ограничена, то за время  $T$  это отношение не может быть больше, чем  $P_s T/N_0$ .

Введём ещё понятие *средней энергии на бит*. Если скорость источника равна  $R$ , то можно считать, что каждые  $1/R$  единицы времени генерируется один двоичный символ. Если при этом средняя мощность ограничена величиной  $P_s$ , то средняя энергия, приходящаяся на один бит, которую мы обозначим за  $E_b$ , есть

$$E_b = \frac{P_s}{R}.$$

Средняя энергия на бит, требуемая различными системами связи для достижения заданной помехоустойчивости, является некоторой мерой их относительной эффективности.

Кроме средней мощности, на практике надо учитывать и ограничение на *пиковую мощность*, или *максимальную мощность* передатчика.

Следующий фактор, который мы должны принять во внимание, — это ограниченная полоса частот любого реального канала связи. Вспомним из теории обработки сигналов, что в случае сокращения времени передачи сигнала ширина его спектра возрастает. Действительно, пусть сигнал  $x(t)$  имеет длительность  $T$  (т.е. отличен от нуля на интервале длины  $T$ ), и пусть его спектр  $X(\omega)$  имеет ширину  $W$  (т.е. спектр существенно отличен от нуля на интервале длины  $W$ ). Сигнал, имеющий ту же форму, что и  $x(t)$ , но занимающий интервал в  $n$  раз меньше, имеет вид  $x(nt)$ . Его спектр есть

$$\int_{-\infty}^{+\infty} x(nt)e^{-i\omega t} dt = \frac{1}{n} \int_{-\infty}^{+\infty} x(u)e^{-i\frac{\omega}{n}u} du = \frac{1}{n} X\left(\frac{\omega}{n}\right)$$

— это растянутый в  $n$  раз (и умноженный на  $1/n$ ) спектр сигнала  $x(t)$ . Таким образом, сжатие сигнала приводит к растяжению его спектра. Для выполнения условий искажающей передачи по каналу связи (см. §1) необходимо, чтобы этот спектр и после растяжения не выходил за пределы полосы частот канала связи, и это ставит предел



степени сжатия сигнала для увеличения информационной скорости.

Менее очевидно, что рассматривая систему связи, работающую вблизи предельно достижимых информационных скоростей, необходимо также учитывать неточности в воспроизведении сигналов передатчиком и приёмником. Ранее мы пренебрегали этим и считали, что и приёмник, и передатчик генерируют все необходимые сигналы абсолютно точно. Так как мы могли считать, что рассматриваем сигналы произвольно большой длительности, то это предположение было допустимо. Теперь же нам необходимо учитывать, что реальная аппаратура воспроизводит форму сигналов, включая их расположение на оси времени не идеально, а с некоторыми погрешностями. В частности, если сигнал начался не точно в момент времени  $t$ , а в некоторый момент  $t + \Delta t$ , где  $\Delta t$  меньше, чем некоторый ненулевой предел, зависящий от конструкции аппаратуры, то эти два сигнала на практике будут неотличимы друг от друга.

Кроме того, разница в форме сигналов тоже должна быть достаточно большой, чтобы они были различимы. Будем считать, что для того чтобы сигналы  $\phi(t)$  и  $\psi(t)$  были различимы, должно выполняться следующее условие:

$$\int_{-\infty}^{+\infty} [\phi(t) - \psi(t)]^2 dt \geq \delta > 0$$

для некоторого заданного  $\delta$ , определяемого возможностями аппаратуры. Самое важное для нас следствие из этого условия состоит в том, что в этом случае пространство функций с ограниченной длительностью и ограниченным спектром будет иметь конечное число измерений в том смысле, что в базисе такого пространства будет конечное число базисных функций, с помощью которых все остальные представляются с задаваемой возможностью аппаратуры погрешностью.

Более точно этот результат сформулирован в следующей теореме, которую (в разных формах) первыми доказали Ландау, Поллак и Шеннон:

**Теорема:** (*О максимальной размерности пространства сигналов*)

Пусть  $\{f(t)\}$  - множество функций, удовлетворяющих следующим условиям:

- 1)  $f(t)$  тождественно равна 0 при  $t \notin [0, T]$ .
- 2) Энергия  $f(t)$  конечна:

$$\int_{-\infty}^{+\infty} |F(f)|^2 df = \int_{-\infty}^{+\infty} f^2(t) dt = E_f < \infty.$$

- 3) Спектральная энергия  $f(t)$  вне полосы частот  $[-W, W]$  не превышает  $\eta_w^2$ :

$$\int_{-W}^{+W} |F(f)|^2 df \geq (1 - \eta_w^2) E_f.$$

Тогда существует совокупность  $L$  ортонормальных функций  $\psi_i(t)$ ,  $i = 1, 2, \dots, L$ , для которых:

- а) Выполняются условия 1, 2 и 3.
- б) Для любой функции из множества  $\{f(t)\}$  справедливо, что

$$\int_{-\infty}^{+\infty} [f(t) - \sum_{i=1}^L f_i \psi_i(t)]^2 dt < \epsilon^2,$$

где  $f_i = \int_{-\infty}^{+\infty} f(t) \psi_i(t) dt$ ,  $\epsilon^2 = 12\eta_w^2$  и при этом  $L$  - максимальное целое число, для которого

$$L \leq 2TW + 1.$$

( Без доказательства. )

Замечание: в данной теореме предполагается, что спектр функции выходит немного за полосу частот (на  $\eta_w^2$ ). Это потому, что функция, отличная от нуля на конечном интервале, не может иметь спектр, отличный от нуля только на конечном интервале, - ее спектр будет отличен от нуля при всех  $-\infty < f < +\infty$ . Таким образом, как мы знаем из теории обработки сигналов, сигнал ограниченной длительности невозможно без искажений передать по каналу с ограниченной полосой частот без искажений. Однако с практической точки зрения, если выбрать  $\eta_w$  малым, то искажения будут малыми и ими можно будет пренебречь.

Главный результат этой теоремы состоит в том, что в канале с полосой частот  $W$  может существовать не более

чем  $(2TW + 1)$  различных друг от друга ортонормальных сигналов длительности  $T$ . Можно сказать, что число возможных измерений в пространстве сигналов длительности  $T$  и шириной спектра  $W$  не может расти с ростом  $T$  (или  $W$ ) быстрее, чем линейно.

Теперь давайте покажем, что это число может расти линейно с ростом  $T$ . То есть покажем, что

$$N = DT,$$

где  $N$  – число ортогональных сигналов длительности не больше  $T$ , а  $D$  – коэффициент, не зависящий от  $T$  и линейно зависящий от  $W$ . Пусть  $x(t)$  – некоторый сигнал, тождественно равный нулю вне интервала длительности  $\tau$  и занимающий полосу частот  $W$  (например, прямоугольный импульс). Тогда за время  $T$  можно передать  $T/\tau$  таких сигналов, чтобы они не перекрывались (и, значит, были бы ортогональными). То есть  $D = 1/\tau$  (ортогональных сигналов в единицу времени).

Как мы уже отмечали, если  $X(f)$  – спектр  $x(t)$ , то  $\frac{1}{n}X\left(\frac{w}{n}\right)$  – спектр  $x(nt)$ , который имеет длительность  $\tau/n$  и ширину полосы  $nW$ . Отсюда видно, что коэффициент  $D$  прямо пропорционален  $W$ .

Можно привести и другие примеры систем ортогональных сигналов длительности  $T$ , число которых линейно растет с ростом  $T$ . Например, отрезки синусоид с целым числом периодов на интервале  $[0, T]$  и частотами, не превосходящими  $W$ . (Частоты кратны  $1/T$ ).

В реальных каналах такое поведение сохраняется, но  $D$  имеет меньшую величину, чем в теореме. Обычно число практически реализуемых ортогональных сигналов в полосе  $W$  и длительностью  $T$  лежит между  $TW$  и  $3/2TW$ , в зависимости от определения  $W$  и сложности реализации.

Итак, мы сформулировали ряд ограничений, в которых и будем далее рассматривать передачу последовательных сообщений.

## **§16. Передача отдельных символов и передача блоков ортогональными сигналами**

Чтобы увидеть, что различные системы связи, характеризующиеся одной и той же величиной  $E_b$ , при передаче

последовательных сообщений могут вести себя совершенно по-разному, рассмотрим передачу последовательности  $K = RT$  равновероятных двоичных символов двумя способами.

Первый (весьма естественный) способ – передать информацию последовательностью  $K$  неперекрывающихся сигналов в системе связи с бинарными противоположными сигналами (например, прямоугольными импульсами), где каждый сигнал соответствует передаче одного бита. Энергия каждого импульса  $E_b$ , общая расходуемая энергия  $E_s = KE_b$ . Мы будем интересоваться вероятностью  $P_R$  принять все биты без ошибки (вероятность ошибки  $P_E = 1 - P_R$ ).

Как и ранее, в канале действует аддитивный гауссовский шум.

Сигнал, несущий информацию об этой последовательности, можно записать как

$$s(t) = \sum_{j=1}^K s_j x(t - j\tau_0),$$

где  $\tau$  – длительность единичного сигнала  $x(t)$ ,

$$s_j = \begin{cases} +1, & \text{если } j\text{-й бит } 1, \\ -1, & \text{если } j\text{-й бит } 0, \end{cases}$$

а  $x(t)$  представляет собой сигнал энергии  $E_b$ . Положив

$$\phi_j(t) = \frac{x(t - j\tau_0)}{\sqrt{E_b}}, \quad j = 1, 2, \dots, K,$$

мы поставим в соответствие совокупности  $M = 2^K$  возможных сигналов  $2^K$  вершин  $K$ -мерного гиперкуба. Ранее мы рассматривали такую систему сигналов и нашли, что вероятность ошибки

$$P_E = 1 - (1 - p)^K = 1 - (1 - p)^{RT},$$

где  $p = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) = Q\left(\sqrt{\frac{2P_s}{RN_0}}\right)$  для канала с белым гауссовским шумом со спектральной мощностью  $N_0/2$ . Видно, что с ростом  $T$ , то есть с ростом количества бит  $K$  переданной

последовательности, вероятность ошибки стремится к 1. При фиксированных  $K$  и  $N_0$  вероятность ошибки можно уменьшить, только увеличивая  $E_b$ , что достигается либо увеличением средней мощности, либо уменьшением скорости передачи  $R$ .

Теперь рассмотрим другую схему передачи, где каждые  $T$  единиц времени передается один из  $2^K$  ортогональных импульсов, причем длительность одного импульса есть  $\tau_1 = \frac{T}{2^K}$ , а его положение внутри интервала  $T$  определяется передаваемой последовательностью, как двоичным  $K$ -разрядным числом:

$$s_i(t) = \sqrt{E_s} \phi(t - i\tau_1), \quad i = 0, 1, \dots, 2^K - 1$$

(здесь энергия сигнала  $E_s = E_b K = TP_s$ ). Для этой системы мы воспользуемся верхней границей вероятности ошибки передачи одного из  $M$  равновероятных ортогональных сигналов равной энергии:

$$P_E \leq (M - 1)Q \left( \sqrt{\frac{E_s}{N_0}} \right).$$

Нам потребуется оценка для величины

$$Q(\alpha) = \int_{\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx.$$

Докажем следующую оценку для функции ошибок:

$$Q(\alpha) \leq \frac{1}{2} \exp(-\alpha^2/2).$$

Для этого рассмотрим двумерный гауссовский вектор с независимыми компонентами, нулевым средним и единичными дисперсиями компонент  $(X, Y)$ :

$$f(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

Пусть  $D_1$  – область, изображенная на рис.16.1, а. Вероятность для  $(X, Y)$  попасть в область  $D_1$  есть, как нетрудно видеть:

$$Pr\{(X, Y) \in D_1\} = \int_{\alpha}^{+\infty} \int_{\alpha}^{+\infty} f(x, y) dx dy =$$

$$\begin{aligned}
&= \int_{\alpha}^{+\infty} \int_{\alpha}^{+\infty} \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) dx dy = \\
&\int_{\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \int_{\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = [Q(\alpha)]^2.
\end{aligned}$$

Теперь рассмотрим область  $D_2$  на рис. 16.2, б. Очевидно, что

$$Pr\{(X, Y) \in D_2\} \geq Pr\{(X, Y) \in D_1\}.$$

Из соображений симметрии нетрудно видеть, что

$$Pr\{(X, Y) \in D_2\} = \frac{1}{4} Pr\{(X, Y) \in D_3\},$$

где  $D_3$  – внешность круга на рис. 16.2, б. Для отыскания вероятности попадания в  $D_3$  перейдем к полярным координатам:

$$\begin{aligned}
Pr\{(X, Y) \in D_3\} &= \int_{D_3} \int f_{X,Y}(x, y) dx dy = \\
&= \int_{\sqrt{2}\alpha}^{+\infty} \int_0^{2\pi} \frac{1}{2\pi} \exp\left(-\frac{\rho^2}{2}\right) \rho d\rho d\phi = \\
&= \int_{\sqrt{2}\alpha}^{+\infty} \exp\left(-\frac{\rho^2}{2}\right) \rho d\rho = -\exp\left(-\frac{\rho^2}{2}\right) \Big|_{\sqrt{2}\alpha}^{+\infty} = \exp(-\alpha^2),
\end{aligned}$$

откуда следует, что

$$Pr\{(X, Y) \in D_2\} = \frac{1}{4} \exp(-\alpha^2)$$

и

$$Q(\alpha) \leq \sqrt{\frac{1}{4} \exp(-\alpha^2)} = \frac{1}{2} \exp(-\alpha^2/2).$$

Вернёмся к оценке вероятности ошибки передачи для передачи блоков ортогональными сигналами:

$$P_E \leq (M-1)Q\left(\sqrt{\frac{E_s}{N_0}}\right) < M \exp\left(-\frac{E_s}{2N_0}\right).$$

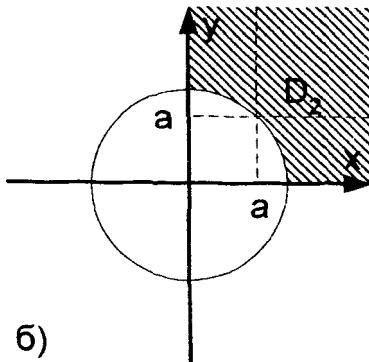
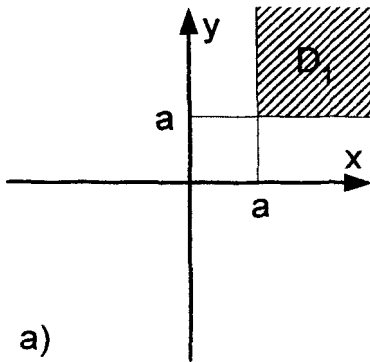


Рис. 16.1. Построение оценки для функции ошибок

Подставив сюда  $M = 2^K = 2^{RT}$  и  $E_s = KE_b = TP_s$ , получим

$$P_E < 2^{RT} \exp\left(-\frac{TP_s}{2N_0}\right) = \exp\left(-T\left(\frac{P_s}{2N_0} - R \ln 2\right)\right),$$

то есть с ростом  $T$  вероятность ошибки будет стремиться к нулю, если

$$R < \frac{P_s}{2N_0 \ln 2} \approx 0,72 \frac{P_s}{N_0}.$$

Если выразить всё через  $E_b$  и  $K$ , то

$$P_E < 2^K \exp -K\left(\frac{E_b}{2N_0}\right) = \exp -K\left(\frac{E_b}{2N_0} - \ln 2\right),$$

которая экспоненциально стремится к нулю при

$$\frac{E_b}{N_0} > 2 \ln 2 \approx 1,39.$$

Как видно, две рассмотренные системы ведут себя относительно вероятности ошибки совершенно по-разному. В первом случае вероятность ошибки с ростом  $K$  стремится к 1, как бы велико ни было отношение сигнал/шум ( $E_s/N_0$ ). У второй системы, увеличивая  $K$ , мы можем сделать вероятность ошибки сколь угодно малой, если только  $E_b/N_0$  больше чем 1,39. Иными словами, отношение сигнал/шум  $E_b/N_0$  устанавливает некоторую границу для максимальной скорости передачи – если скорость меньше этого максимума, то вероятность  $P_E$  может быть сделана сколь угодно малой выбором достаточно больших  $T$ .

Геометрическая интерпретация векторов сигналов позволяет лучше понять причины этого различия: при передаче отдельных символов расстояние между сигналами не меняется, а число соседей и размерность пространства растут линейно с ростом  $K$ .

При передаче блоков ортогональными сигналами расстояние между сигналами растет  $\sim \sqrt{K}$  за счет новых измерений и перенормировки амплитуды. И хотя число соседей тоже растет, именно увеличение расстояния между сигналами при больших  $E_b/N_0$  определяет поведение ошибки. А вот при малых  $E_b/N_0$  поведение определяется ростом числа соседей.



К сожалению, схема передачи блоков ортогональными сигналами на практике не может считаться решением проблемы надежной связи по гауссовскому каналу. При  $R$  близких к  $0,72P_s/N_0$  для надежной передачи необходимы очень большие значения  $T$ . В свою очередь, при больших  $T$  число ортогональных сигналов, равное  $2^{RT}$ , — очень велико, а длительность импульса, равная  $(T/2^{RT})$ , — очень мала. Такой импульс невозможно будет передать по реальным каналам с ограниченной полосой частот. Например, пусть скорость передачи в канале  $R = 100$  бит/с и отношение  $P_s/N_0$  таково, что для достижения нужной помехоустойчивости необходимо  $T = 1$  с (то есть  $K = 100$ ), так что  $M = 2^{100}$ . Если приблизительно принять, что для прямоугольного импульса его ширина спектра  $f$  и длительность  $\tau$ , связаны соотношением  $f\tau \approx 1$ , то в такой системе  $\tau = T/2^{RT} \approx 10^{-30}$  с, и  $f \approx 10^{30}$  Гц, что в несколько раз шире, чем рентгеновский диапазон.

Кроме того, с уменьшением  $\tau$  пиковая мощность, которую должен излучать передатчик, растет и тоже может принимать чрезмерно большие значения. Все это делает рассмотренную схему практически нереализуемой.

Чего же можно добиться в реальных системах связи, когда мы вынуждены учитывать все введенные выше условия? Один из главных результатов теории связи состоит в том, что надежная передача может быть организована и при таких условиях. Теперь мы готовы перейти к доказательству этого факта для канала с аддитивным белым гауссовским шумом. Мы повторим доказательство, принадлежащее Шеннону. Отметим следующее необычное обстоятельство: мы докажем существование системы надёжной передачи последовательных сообщений по каналу с ограниченной полосой частот и аддитивным белым гауссовским шумом без ее построения.

### §17. Надежная передача по каналу с аддитивным белым гауссовским шумом

Рассмотрим систему ортогональных сигналов  $\phi_i(t)$  такую, что число сигналов линейно растёт с ростом  $T$ :  $N = DT$ .

В качестве системы сигналов, используемых для пе-

редачи сообщений, возьмем сигналы вида

$$s_i(t) = \sum_{j=1}^N s_{ij} \phi_j(t), \quad i = 1, \dots, M,$$

где  $s_{ij} = \pm \sqrt{E_N} \forall i, j$ . Это значит, что все сигналы  $s_i(t)$  находятся в вершинах  $N$ -мерного гиперкуба. Энергия такого сигнала, как мы нашли ранее, есть

$$E_s = P_s T = \sum_{j=1}^N s_{ij}^2 = N E_N,$$

где  $E_N = P_s / D = E_s / N$  – энергия сигнала на одно измерение пространства сигналов. Такую систему сигналов мы будем использовать для передачи информации от источника двоичной информации со скоростью  $R$ , причем будем считать, что

$$D > R.$$

Тогда число двоичных последовательностей, порождаемых за время  $T$ , есть  $M = 2^{RT}$ , а число сигналов, являющихся вершинами  $DT$ -мерного гиперкуба, –  $2^{DT}$ . Вершин, которые надо использовать для передачи, меньше чем их общее число. При этом *доля вершин*, которые мы должны использовать, равна

$$\frac{2^{RT}}{2^{DT}} = 2^{-(D-R)T},$$

она стремится к нулю с ростом  $T$  (т.к.  $D > R$ ). Поэтому можно надеяться, что вероятность ошибки не будет стремиться к 1 с ростом  $T$ , как это было для  $D = R$  из-за слишком близкого "соседства" сигналов. Вопрос в том, как выбирать используемые вершины. Но мы обойдем этот вопрос следующим образом: рассмотрим не одну конкретную систему связи данного типа, а сразу *все* возможные системы связи данного типа, каждая из которых состоит из передатчика, канала и оптимального приёмника. Каждая из них использует *свою* систему сигналов  $\{\bar{s}_i\}$ , а в остальном эти системы совершенно идентичны.

Всего для  $2^N = 2^{DT}$  вершин гиперкуба и  $M = 2^{RT}$  сигналов  $\{s_i\}$  имеется  $(2^N)^M = 2^{NM}$  вариантов выбора  $\{s_i\}$  (включая, например, и такой, где  $M$  раз мы выбираем один и тот же вектор для всех сообщений – явно плохой выбор). Предположим, что каждый из этих  $2^{NM}$  вариантов используется в одной и только одной системе связи из нашего набора, и приёмник соответствующей системы оптимален для этого набора.

Следуя установившейся терминологии, будем называть вектор, соответствующий каждому из выбранных сигналов, *кодовым словом*, а совокупность всех таких векторов – *кодом*.

Ясно, что каждая система в нашем наборе обладает определённой вероятностью ошибки. Обозначим вероятность ошибки для  $l$ -й системы  $P_l$ ,  $l = 1, 2, \dots, 2^{NM}$ . Некоторые из этих систем, например, с кодами, в которых все  $M$  векторов  $\{\bar{s}_i\}$  сопоставлены одной и той же вершине гиперкуба, обладают очень большой вероятностью ошибки. Для других систем она меньше, для некоторых из них она может быть очень мала. Мы докажем этот факт, вычислив верхнюю оценку для *средней* (по всему набору систем связи) вероятности ошибки:

$$\bar{P} = M(P_E) = \frac{1}{2^{NM}} \sum_{l=1}^{2^{NM}} P_l.$$

Ясно, что хотя бы одно из  $P_l$  меньше, чем  $\bar{P}_E$ .

Итак, для каждого конкретного набора  $\{\bar{s}_i\}$

$$Pr\{\{\bar{s}_i\}\} = \frac{1}{2^{NM}},$$

$$Pr\{\hat{m} \neq m \mid \{\bar{s}_i\}\} = P_l.$$

Тогда,

$$\bar{P}_E = M(P_E) = \sum Pr\{\hat{m} \neq m \mid \{\bar{s}_i\}\} Pr\{\{\bar{s}_i\}\}.$$

Средняя условная вероятность ошибки при условии, что передаётся сообщение  $m_k$ , есть

$$\overline{Pr\{\hat{m} \neq m \mid m = m_k\}} = \sum Pr\{\hat{m} \neq m \mid m = m_k, \{\bar{s}_i\}\} Pr\{\{\bar{s}_i\}\}.$$

Применяя к каждому конкретному коду аддитивную границу для вероятности ошибки, полученную в §13, имеем

$$Pr\{\hat{m} \neq m \mid m = m_k, \{\bar{s}_i\}\} \leq \sum_{i=1}^M P_2[\bar{s}_i, \bar{s}_k],$$

где  $P_2[\bar{s}_i, \bar{s}_k]$  - вероятность ошибки для системы из двух сигналов  $\bar{s}_i$  и  $\bar{s}_k$ .

Для средней условной вероятности ошибки получаем оценку

$$\overline{Pr\{\hat{m} \neq m \mid m = m_k\}} \leq \sum Pr\{\{\bar{s}_i\}\} \sum_{i=1}^M P_2[\bar{s}_i, \bar{s}_k].$$

Сменим порядок суммирования:

$$\overline{Pr\{\hat{m} \neq m \mid m = m_k\}} \leq \sum_{i=1}^M (\sum Pr\{\{\bar{s}_i\}\} P_2[\bar{s}_i, \bar{s}_k]),$$

где черта означает усреднение по ансамблю кодов.

Для канала с аддитивным белым гаусовским шумом  $P_2[\bar{s}_i, \bar{s}_k]$  зависит, как мы знаем, только от евклидова расстояния между  $\bar{s}_i$  и  $\bar{s}_k$ :

$$P_2[\bar{s}_i, \bar{s}_k] = Q\left(\frac{|\bar{s}_i - \bar{s}_k|}{\sqrt{2N_0}}\right).$$

Если у векторов  $\bar{s}_i$  и  $\bar{s}_k$  различны  $h$  координат, то квадрат расстояния между ними есть

$$|\bar{s}_i - \bar{s}_k|^2 = \sum_{j=1}^N (s_{ij} - s_{kj})^2 = (2\sqrt{E_N})^2 h = 4hE_N.$$

По нашему предположению,  $\bar{s}_i$  с равной вероятностью может быть любой вершиной гиперкуба независимо от  $\bar{s}_k$ . Поэтому вероятность того, что  $s_{ij}$  совпадает с  $s_{kj}$  есть  $\frac{1}{2}$  для всех  $j = 1, \dots, N$ . Следовательно, вероятность того, что у векторов  $\bar{s}_i$  и  $\bar{s}_k$  различны  $h$  координат - это вероятность  $h$  успехов в  $N$  испытаниях Бернули с  $p = q = \frac{1}{2}$ :

$$P_h = \binom{N}{h} \left(\frac{1}{2}\right)^N.$$

Тогда

$$\begin{aligned} \overline{P_2[\bar{s}_i, \bar{s}_k]} &= M[Q\left(\frac{|\bar{s}_i - \bar{s}_k|}{\sqrt{N_0}}\right)] = \\ &= \sum_{h=0}^N Pr\{h\} Q\left(\frac{\sqrt{4hE_N}}{\sqrt{2N_0}}\right) = \sum_{h=0}^N 2^{-N} \binom{N}{h} Q\left(\sqrt{\frac{2hE_N}{N_0}}\right). \end{aligned}$$

Это выражение не зависит от индексов  $i$  и  $k$ , поэтому если ввести обозначение  $\overline{P_{2,E}} = \overline{P_2[s_i, s_k]}$ , получим

$$\overline{Pr\{\hat{m} \neq m \mid m = m_k\}} \leq \sum_{i=1}^M \overline{P_2[\bar{s}_i, \bar{s}_k]} = (M-1)\overline{P_{2,E}} < M\overline{P_{2,E}},$$

и

$$\begin{aligned} \overline{Pr\{\hat{m} \neq m\}} &= \sum_{k=1}^M \overline{Pr\{\hat{m} \neq m \mid m = m_k\}} Pr\{m_k\} < \\ &< M\overline{P_{2,E}} \sum_{k=1}^M Pr\{m_k\} = M\overline{P_{2,E}}. \end{aligned}$$

Используем оценку для  $Q(\alpha) < \frac{1}{2}e^{-\frac{\alpha^2}{2}} < e^{-\frac{\alpha^2}{2}}$ , имеем

$$\begin{aligned} \overline{Pr\{\hat{m} \neq m\}} &< M\overline{P_{2,E}} < M \sum_{h=0}^N 2^{-N} \binom{N}{h} e^{-\frac{hE_N}{N_0}} = \\ &= M \sum_{h=0}^N 2^{-N} \binom{N}{h} \left[e^{-\frac{E_N}{N_0}}\right]^h. \end{aligned}$$

Используем разложение бинома:

$$(1 + \alpha)^N = \sum_{h=0}^N \binom{N}{h} \alpha^h,$$

откуда

$$\overline{Pr\{\hat{m} \neq m\}} < M2^{-N} (1 + e^{-\frac{E_N}{N_0}})^N.$$

Введём обозначение:

$$R_0 = \log_2 \left( \frac{2}{1 + e^{-\frac{E_N}{N_0}}} \right) = 1 - \log_2 \left( 1 + e^{-\frac{E_N}{N_0}} \right).$$

Тогда полученная выше оценка примет вид

$$\overline{Pr\{\hat{m} \neq m\}} < M 2^{-NR_0}.$$

Если определить величину

$$R_N = \frac{R}{D} \text{ (бит/измерение),}$$

равную числу бит, приходящихся на одну размерность гиперкуба или, в более общем случае, на одно измерение пространства сигналов, то  $M = 2^{RT} = 2^{NR_N}$  и оценку можно записать как

$$\overline{Pr\{\hat{m} \neq m\}} < 2^{-N[R_0 - R_N]}.$$

Согласно всем этим соотношениям, независимо от формы их записи, средняя вероятность ошибки, а значит и вероятность ошибки для по крайней мере одного кода в ансамбле, может быть сделана как угодно малой за счёт выбора больших  $N$ , если только  $R_N$  меньше  $R_0$ .

Число измерений пространства используемых сигналов  $N$  часто называют *длиной кодового блока*.

Как мы видели,  $R_0$  зависит только от  $\frac{E_N}{N_0}$ , т.е. от отношения сигнал/шум. При  $E_N \rightarrow \infty$   $R_0 \rightarrow 1$  — это максимальное значение  $R_0$ . Поэтому при любом соотношении сигнал/шум  $R_N \leq 1$  для возможности организации надёжной передачи. При этом  $R_N$  может быть равным единице, только если  $\frac{E_N}{N_0} = \infty$ . Это хорошо согласуется с результатом о том, что при  $R = D$ , когда  $R_N = 1$ , т.е. число возможных сигналов равно числу вершин гиперкуба, вероятность ошибки стремится к 1, если в канале есть хоть какой-нибудь шум.

### §18. Передача многозначными последовательностями

В предыдущем параграфе рассматривалась система сигналов в вершинах гиперкуба. Сейчас мы коротко рассмотрим более общий случай. Будем считать, что по каждому измерению пространства сигналов компонента  $s_i$  векторов сигналов может принимать более, чем два значения. Пусть на отрезке  $[-\sqrt{E_N}, \sqrt{E_N}]$  компонента  $s_i$  может

принимать  $A$  значений, расставленных на этом промежутке через равные расстояния. На рис. 18.1 для примера представлены 16 допустимых сигналов для  $N = 2$ ,  $A = 4$ . В общем случае возможное число сигналов есть  $A^N = 2^{N \log_2 A}$ . Такой способ гарантирует, что

$$|\hat{s}_i|^2 \leq N E_N.$$

Для данной системы могут быть проведены рассуждения, аналогичные двоичному случаю, с той только разницей, что здесь случайный выбор кодовых слов производится среди общего числа возможных сигналов. При этом можно рассмотреть ситуацию, когда разные значения амплитуды выбираются с разной вероятностью. Однако для каждого из этих случаев можно доказать, что

$$Pr\{\hat{m} \neq m\} < 2^{-N(R_0 - R_N)}.$$

Величина  $R_0$  будет, разумеется, иметь разный вид для разных случаев. Для случая, когда все возможные значения  $s$ , выбираются с равными вероятностями  $p = \frac{1}{A}$ , графики зависимости  $R_0$  от  $\frac{E_N}{N_0}$  изображены на рис. 18.2.

Видно, где выгоднее использовать различные  $A$ . При малых отношениях сигнал/шум более высокую скорость  $R_0$  обеспечивает  $A = 2$ , а по мере того как отношение сигнал/шум растёт, становится выгоднее использовать всё большие и большие  $A$ .

Эти результаты хорошо согласуются со здравым смыслом: когда шум велик, то сигналы надо максимально разнести друг от друга и потому сделать  $A = 2$ , а когда шум мал, можно уже не разносить их далеко и, таким образом, уместить в каждом измерении больше бит информации.

## §19. Пропускная способность канала

Мы показали, что может быть построена совокупность сигналов для канала с аддитивным белым гауссовским шумом, такая что при  $R_N < R_0$  вероятность ошибки передачи может быть сделана сколь угодно малой.

Но из того что мы доказали, не следует, вообще говоря, что для скоростей  $R_N > R_0$  не может быть предложен какой-либо другой способ построения совокупности сигналов, обеспечивающий произвольно малую вероятность

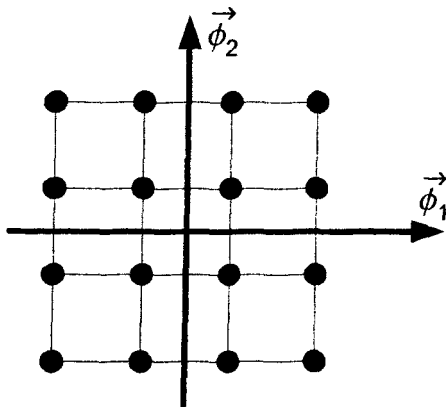


Рис. 18.1. Допустимые сигналы  
для  $A=4$ ,  $N=2$ .



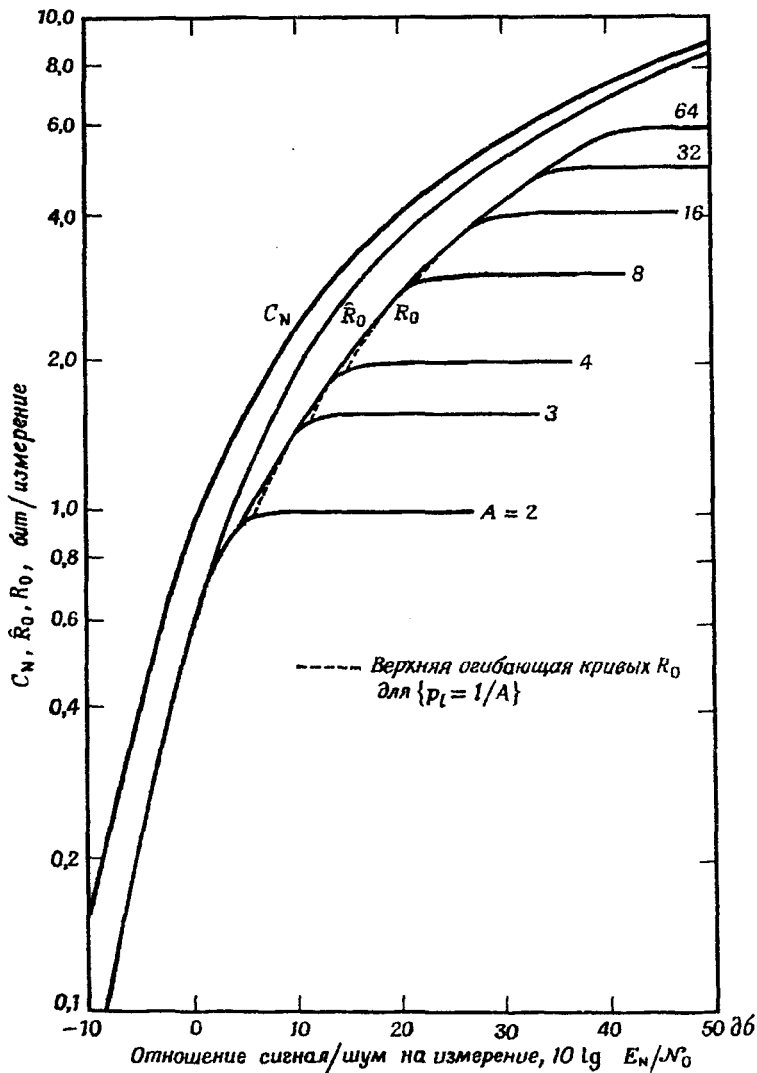


Рис. 18.2. Пропускная способность канала связи

ошибки. Окончательный ответ на этот вопрос даёт доказанная Шенноном теорема о пропускной способности канала связи.

Для её формулировки заметим, что коэффициент пропорциональности  $D$  между длительностью сигналов  $T$  и возможным числом измерений пространства сигналов  $N$  из формулы  $N = DT$  (размерность  $D$  – число измерений в единицу времени) определён неточно, так как неточно определена ситуация, когда спектр сигнала находится в полосе частот канала. Поэтому теорему о пропускной способности сформулируем, используя параметры  $E_N$  и  $R_N = \frac{R}{D}$  (энергию сигнала на одно измерение и скорость поступления данных в битах на измерение пространства сигналов).

**Теорема:** Для канала с аддитивным белым гауссовским шумом существует постоянная

$$C_N = \frac{1}{2} \log_2 \left( 1 + \frac{2E_N}{N_0} \right),$$

называемая *пропускной способностью канала*, которая обладает следующими свойствами:

1) Если  $R_N > C_N$  и число равновероятных сообщений  $M = 2^{NR_N}$  стремится к бесконечности, то для любой совокупности  $M$  передаваемых сигналов вероятность ошибки стремится к 1.

2) Если  $R_N < C_N$  и  $M \rightarrow \infty$ , то существуют такие совокупности  $M$  передаваемых сигналов и соответствующие им оптимальные приёмники, для которых вероятность ошибки будет стремиться к нулю. (Без доказательства.)

Если сравнить  $C_N$  с  $R_0$  или с огибающей для семейства графиков  $R_0$  при разных  $A$  (рис. 18.2), оптимизированных по вероятностям выбора различных амплитуд  $s_i$  (обозначим эту величину как  $\hat{R}_0$ ), то мы увидим, что  $\frac{1}{2}C_N \leq \hat{R}_0 \leq C_N$ . Это означает, что теоретически возможны более эффективные методы надёжной передачи, чем рассмотренная нами блоковая передача.

Приведём ещё выражение  $C_N$  через ширину полосы  $W$ , определённую каким-либо способом. Для этого заметим, что из неравенства  $R_N < C_N$  следует, что  $R < DC_N$  и если

допустимое число измерений в секунду  $D$  оценить как  $D \geq 2W$ , то  $E_N = \frac{P_s}{D} = \frac{P_s}{2W}$ . Отсюда имеем другую известную форму теоремы Шеннона о пропускной способности:

$$R < 2WC_N = W \log_2 \left( 1 + \frac{P_s}{WN_0} \right) = C.$$

Здесь  $C$  – пропускная способность канала в бит/единица времени (бит/с), а  $P_s$  – максимально возможная средняя мощность передаваемого сигнала:

$$P_s = \frac{1}{T} \int_0^T s^2(t) dt = \frac{1}{T} \int_{-\infty}^{\infty} s^2(t) dt = \frac{E_s}{T}.$$

Понятие пропускной способности канала является основой современной теории связи. До работ Шеннона считалось, что при наличии шума в канале и фиксированной скорости источника надёжность передачи не может превзойти некоторый предел. Теорема о пропускной способности утверждает, что шумы, мощность сигнала и полоса частот ограничивают лишь *скорость* надёжной передачи, но возможность надёжной передачи существует всегда.

Модификации этой теоремы верны для очень общих математических моделей каналов. Ещё более важно то, что в реальных каналах также наблюдаются явления, согласующиеся с тем, что надёжная связь возможна, но её скорость не может превзойти некоторый предел. Добавим, что достигнутый на практике предел уже не так далёк от выведенного нами  $\hat{R}_0$  (максимальной скорости блочной передачи).

## §20. Сигналы, используемые на практике

В этом параграфе мы рассмотрим приложения тех теоретических выводов, которые мы получили ранее. Рассмотрим несколько использующихся на практике систем сигналов, и обсудим их достоинства и недостатки.

Независимость помехоустойчивости оптимального приёма от формы сигналов для ситуаций, близких к каналу с аддитивным белым гауссовским шумом, даёт возможность выбором формы сигнала учитывать и другие

практические требования к системам сигналов. К таким требованиям относятся:

- 1) Простота генерации сигналов.
- 2) Согласование спектра сигналов с полосой частот канала.
- 3) Простота реализации оптимального или близкого к оптимальному приёма.

Первый пункт достаточно очевиден - чем проще сигнал генерируется, тем проще, дешевле и надёжней будет устройство передатчика.

Что касается пункта второго, то для узкополосных сигналов он сводится к тому, чтобы спектр сигнала был сосредоточен вблизи центральной частоты канала, и поэтому соответствующие сигналы строятся на основе синусоиды этой частоты. Для широкополосных каналов этот пункт сводится к требованию, чтобы спектр сигнала не выходил как за верхнюю, так и за нижнюю границы полосы частот канала.

Если форма сигнала выбрана, то верхняя граница полосы частот широкополосного канала накладывает ограничение на степень сжатия этой формы, так как при сжатии сигнала расширяется его спектр - мы уже разбирали этот вопрос. Что касается нижней границы полосы частот широкополосного канала, то условие согласования с ним спектра используемой системы сигналов принято формулировать как *условие отсутствия в суммарном сигнале постоянной составляющей*. Дело в том что каналы связи на большие расстояния, например, длинные кабельные или проводные линии, очень плохо пропускают самые низкие частоты, в том числе и "нулевую частоту" - постоянную составляющую. И чем длиннее линия связи, тем хуже она пропускает низкие частоты, тем важнее требование отсутствия постоянной составляющей.

Поясним пункт третий. Как мы видели, построение оптимального приёмника требует использования таких устройств, как перемножители и интеграторы или согласованные фильтры. Это достаточно сложные радиотехнические устройства. Поэтому на практике вместо оптимального приёмника применяют более простые схемы, использующие особенности конкретных систем сигналов и несущественно уступающие по надёжности приёма опти-

мальным. Так, например, если в качестве системы сигналов используется система сигналов в вершинах гиперкуба на основе прямоугольных импульсов (положительной и отрицательной полярности), то близкий к оптимальному приём можно обеспечить, определяя уровень сигнала в середине каждого импульса и сравнивая эту величину с нулём. Такие упрощённые методы приёма сигналов различной формы существуют и для других форм сигналов, но для всех них необходимо соблюдение одного очень важного требования, которое является необходимым и для оптимального приёма – наличие *синхронизации* между приёмником и передатчиком.

Действительно, когда мы разбирали принципы построения оптимального приёмника, то в самом начале обсуждения неявно сделали предположение о том, что приёмник точно знает тот момент, когда на его вход начинает поступать сигнал из канала связи (см. §6). На самом деле, организация синхронизации между приёмником и передатчиком – сложная техническая проблема, один из лучших выходов из которой – использовать такие сигналы, которые кроме переноса информации, помогали бы поддерживать синхронизацию. Такое свойство принято называть *самосинхронизацией* сигнала.

Рассмотрим несколько практических примеров и разберём их достоинства и недостатки с точки зрения введённых нами критериев. Сначала рассмотрим коды для широкополосных каналов связи.

1) Код NRZ (без возвращения к нулю), пример кодирования этим кодом изображён на рис. 20.1, а. Это не что иное, как уже хорошо нам знакомые прямоугольные импульсы.

Достоинства: очень прост в генерации; хорошо использует полосу частот ( $f \sim \frac{1}{T}$ , где  $T$  – время передачи одного бита источника); потенциально помехоустойчив, так как использует противоположные сигналы.

Недостатки: плохо поддерживает синхронизацию – если идёт последовательность одних нулей или одних единиц, то у сигнала нет никаких признаков, помогающих различить, где кончается один импульс и начинается другой; кроме того, если в передаваемой информации число единиц не равно числу нулей или если импульсы одно-

полярные, то в суммарном сигнале будет присутствовать постоянная составляющая.

Используется в тех системах, где главное – простота генерации и приёма, и одновременно важна скорость передачи, а синхронизация может быть обеспечена извне, например, на внутренних шинах ЭВМ. Для телекоммуникаций используется только на небольшие расстояния (связь ЭВМ с периферийными устройствами).

2) АМI-код (Alternate Mark Inversion) – рис. 20.1, б.

Достоинства: простота генерации и хорошее использование полосы частот, как у NRZ; гораздо лучшая, чем у NRZ, способность к самосинхронизации, так как при передаче единицы в сигнале всегда присутствует перепад, по которому приёмник может восстанавливать синхронизацию; отсутствие постоянной составляющей независимо от передаваемой информации.

Недостатки: теоретически обладает худшей потенциальной помехоустойчивостью, чем NRZ, так как вместо противоположных использует систему с тремя уровнями сигналов; при длинной последовательности нулей возможна потеря синхронизации.

В целом, очень хороший код, широко используется на практике в магистральных кабельных системах связи. При этом возможная потеря синхронизации при передаче длинной последовательности нулей исключается путём вставки после определённого числа идущих подряд нулей специального сигнала, восстанавливающего синхронизацию (точнее, заменяют последовательность из нескольких нулей специальным сигналом, в котором присутствует импульс, нарушающий правила кодирования для АМI кода, т.е. имеющий ту же полярность, что и предыдущий).

3) Код Bipolar RZ (биполярный с возвращением к нулю) – рис. 20.1, в.

Достоинства: прост в генерации; обладает очень хорошей способностью к самосинхронизации, так как имеет перепад сигнала в середине интервала передачи каждого бита; потенциальная помехоустойчивость высокая, так как использует противоположные сигналы.

Недостатки: хуже, чем NRZ и АМI, использует полосу частот, так как при той же  $\tau$  (времени передачи одного бита) требует в два раза большую полосу ( $f \sim \frac{2}{\tau}$ ); постоян-

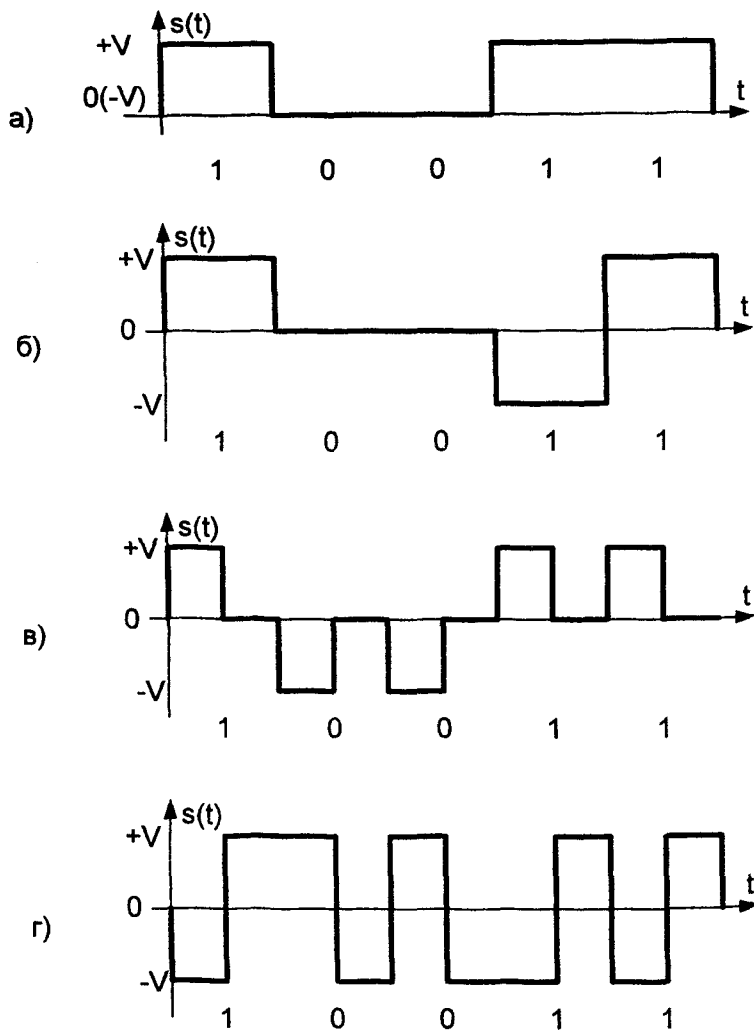


Рис. 20.1. Примеры кодов для широкополосных каналов: а) NRZ; б) AMI; в) Bipolar RZ; г) "Манчестер-2"

ная составляющая суммарного сигнала отлична от нуля, если единиц и нулей в передаваемой последовательности не поровну.

Используется в некоторых среднескоростных локальных сетях, где ширина полосы частот используемого канала не является ограничивающим фактором.

4) Код "Манчестер - 2" – рис. 20.1, г.

Достоинства: обладает очень высокой способностью к самосинхронизации, так как имеет перепад сигнала в середине интервала передачи каждого бита; обладает большой потенциальной помехоустойчивостью, так как использует противоположные сигналы; суммарный сигнал не имеет постоянной составляющей.

Недостатки: по сравнению с NRZ и AMI требует в два раза большую полосу частот при том же времени передачи одного бита ( $f \sim \frac{2}{T}$ ).

Широко используется в среднескоростных локальных сетях ЭВМ, когда ширина полосы частот используемого канала не является ограничивающим фактором – это стандартный код для сети "Ethernet" (до 10 Mbit/s).

Что касается высокоскоростных локальных сетей, таких как "Fast Ethernet" (100 Mbit/s) и "Gigabit Ethernet" (1 Gbit/s), то для них "Манчестер - 2" уже не подходит – слишком широкий у него становится на таких скоростях спектр для передачи сигнала по стандартизованному для этих сетей кабелю.

Для "Fast Ethernet" используется так называемое NRZI кодирование, которое в некотором смысле объединяет в себе свойства NRZ и AMI: используются двухуровневые сигналы, при этом если на интервале передачи одного бита передавалась единица, то в начале этого интервала уровень сигнала меняется на противоположный, а если ноль – то уровень остаётся таким же, как и был. Таким образом, при передаче последовательности единиц перепады в сигнале есть в начале каждого интервала передачи бита, и получающиеся импульсы имеют длительность  $\tau \sim 1/f$ , то есть полоса частот канала используется эффективно. Проблемы с синхронизацией могут возникнуть только при передаче последовательности нулей. Для борьбы с этим применяется так называемое 4B/5B кодирование, когда каждые 4 бита источника, перед тем как



поступить в передатчик, преобразуются в 5 бит согласно следующей таблице:

Биты источника	Биты кода	Биты источника	Биты кода
0000	11110	1000	10010
0001	01001	1001	10011
0010	10100	1010	10110
0011	10101	1011	10111
0100	01010	1100	11010
0101	01011	1101	11011
0110	01110	1110	11100
0111	01111	1111	11101

Нетрудно видеть, что какова бы ни была последовательность бит источника, после кодирования больше трёх нулей подряд не будет, при этом необходимая ширина полосы частот канала возрастает только в 1.25 раза.

Для "Gigabit Ethernet" стандарты кодирования зависят от используемой среды передачи. Для одного из видов кабеля, например, стандартизован сигнальный код NRZI с кодированием 8B/10B для улучшения синхронизирующих свойств сигнала.

Всё это были сигналы, предназначенные для широкополосных каналов. Теперь рассмотрим системы сигналов, предназначенных для работы по узкополосным каналам, у сигналов для которых полоса частот должна быть сосредоточена вокруг некоторой частоты, называемой *несущей*.

Так как здесь очень важно требование того, чтобы спектр сигнала был узкий и весь сосредоточен вокруг частоты  $f$ , то базовыми функциями для системы таких сигналов являются отрезки синусоиды и косинусоиды соответствующей частоты на длине одного периода этой частоты. Будем изображать используемые для передачи информации сигналы для одного элемента суммарного сигнала точками на координатной плоскости. Главные методы передачи информации такими сигналами – это *фазовая* и *амплитудно-фазовая манипуляции*.

1) Двухпозиционная фазовая манипуляция – рис. 20.2, а.

Самая простая в реализации, самая помехоустойчи-

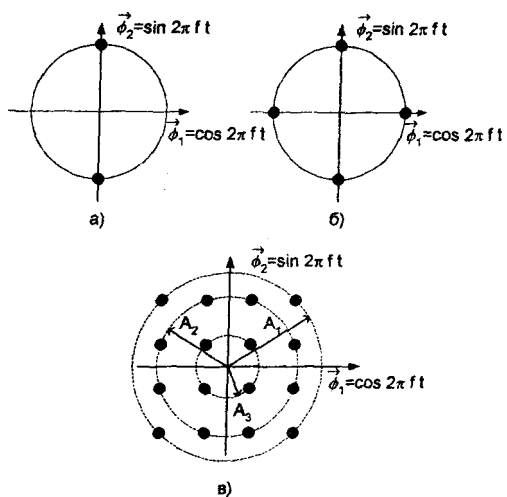


Рис. 20.2. Примеры кодов для узкополосных каналов:  
 а) двухуровневая фазовая манипуляция;  
 б) четырёхуровневая фазовая манипуляция;  
 в) 16-ричная амплитудно-фазовая манипуляция

вая, так как использует противоположные сигналы. Но обеспечивает передачу не более одного бита за период несущей частоты ( $R_0 \rightarrow 1$  при  $\frac{E_s}{N_0} \rightarrow \infty$ ).

2) Четырёхпозиционная фазовая манипуляция – рис. 20.2, б.

Сложнее, требует более качественных (то есть с большим отношением сигнал/шум) каналов, так как менее помехоустойчива. Зато на качественных каналах обеспечивает  $R_0 \rightarrow 2$  при больших отношениях сигнал/шум (случай  $A = 4$  в нашей терминологии), то есть обеспечивает в два раза большую скорость передачи, чем двухпозиционная фазовая манипуляция.

Аналогично строится и восьмипозиционная фазовая манипуляция, но более чем 8-позиционная фазовая манипуляция на практике не используется, так как из-за уменьшающегося расстояния между сигнальными точками помехоустойчивость таких систем сигналов сильно падает.

3) 16-ричная амплитудно-фазовая манипуляция – рис. 20.2, в.

Как видно из рисунка, здесь у сигнала три возможных амплитуды и 12 возможных значений фазы.

Обеспечивает очень высокую скорость передачи ( $R_0 \rightarrow 4$  при  $\frac{E_s}{N_0} \rightarrow \infty$ ), но только для высококачественных каналов связи с большим отношением сигнал/шум.

Существует и множество других подобных систем сигналов. Фактически, каждая из них используется в каком-либо стандарте на передачу данных по телефонным линиям общего назначения – в основном такие сигналы применяются именно там.

### Список литературы

1. Возенкрафт Д., Джекобс И. Теоретические основы техники связи. -М.: Мир, 1969.
2. Харкевич А.А. Борьба с помехами. -М.: Наука, 1965.
3. Бертсекас Д., Галлагер Р. Сети передачи данных. -М.: Мир, 1989.

## ОГЛАВЛЕНИЕ

Введение .....	3
§1. Необходимые сведения из физики и теории обработки сигналов .....	3
§2. Предмет теории передачи дискретных сообщений	7
§3. Переход от непрерывного канала к векторному ..	11
§4. Оптимальный приём в векторном канале .....	18
§5. Канал с аддитивным белым гауссовским шумом ..	21
§6. Принципиальные схемы оптимального приёмника для векторного канала с аддитивным гауссовским шумом .....	26
§7. Мультивекторные каналы .....	28
§8. Оптимальный приём в канале с аддитивным белым гауссовским шумом .....	37
§9. Преобразования пространства сигналов, не изменяющие вероятность ошибки оптимального приёма .....	40
§10. Система из двух сигналов .....	43
§11. Прямоугольные системы сигналов .....	46
§12. Система ортогональных сигналов и родственные ей системы сигналов .....	49
§13. Аддитивная граница для вероятности ошибки ....	54
§14. Полностью симметричные системы сигналов и минимаксный приём .....	56
§15. Передача последовательностей сообщений .....	60
§16. Передача отдельных символов и передача блоков ортогональными сигналами .....	66
§17. Надежная передача по каналу с аддитивным белым гауссовским шумом .....	72
§18. Передача многозначными последовательностями	77
§19. Пропускная способность канала .....	78
§20. Сигналы, используемые на практике .....	82
Список литературы .....	90

**Учебное издание**

**Привалов Александр Юрьевич**

**ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕЛЕКОММУНИКАЦИЙ**

**Раздел 1: Введение в теорию передачи**

**дискретных сообщений**

**Учебное пособие**

Редактор Л.Я. Чегодаева

Корректор Л.Я. Чегодаева

Компьютерная вёрстка А.Ю. Привалов

Подписано в печать 12.05.2003 Формат 60x84 1/16.

Бумага офсетная. Печать офсетная.

Усл.печ.л. 5,34. Усл. кр.-отт 5,46. Уч.-изд.л. 5,75

Тираж 1000 экз. Заказ Арт. С-21/2003

Самарский государственный аэрокосмический  
университет имени академика С.П.Королёва  
443086, Самара, Московское шоссе, 34.

---

АНО "Издательство СНЦ РАН"

443086, Самара, Студенческий переулок, 3а.