

МИНИСТЕРСТВО ОБОРОНЫ РОССИЙСКОЙ ФЕДЕРАЦИИ

Военный авиационный инженерный университет (г. Воронеж)

**МАКАРЕНКО С. И., САПОЖНИКОВ В. И.,
ЗАХАРЕНКО Г. И., ФЕДОСЕЕВ В. Е.**

СИСТЕМЫ СВЯЗИ

Учебное пособие

под ред. С. И. Макаренко

Воронеж
Издание ВАИУ
2011

ББК 32.884: 68.651.4 р

Системы связи: учебное пособие для студентов (курсантов) вузов / С. И. Макаренко, В. И. Сапожников, Г. И. Захаренко, В. Е. Федосеев; под общ. ред. С. И. Макаренко. - Воронеж, издание ВАИУ, 2011. – 285 с.: ил.

В учебном пособии рассматриваются основные положения систем связи. Излагаются вопросы приема/передачи сообщений с различными видами модуляции и манипуляции. Уделено внимание помехоустойчивому кодированию информации. Цель учебного пособия – оказать помощь обучающимся в освоении основных положений систем связи, которые позволят на достаточно высоком уровне изучить конкретные образцы радиостанций и другой аппаратуры связи.

Глава 1-2 написана кандидатом технических наук Захаренко Г.И.

Глава 3 написана кандидатом военных наук Сапожниковым В.И.

Глава 4 написана кандидатом технических наук Федосеевым В.Е.

Глава 5 написана кандидатом военных наук Сапожниковым В.И.

Глава 6 написана кандидатом технических наук Макаренко С. И.

и кандидатом военных наук Сапожниковым В.И.

Главы 7-8 написаны кандидатом технических наук Макаренко С. И.

Общая редакция материалов выполнена С. И. Макаренко.

Для студентов (курсантов), обучающихся по специальности «Техническая эксплуатация транспортного радиооборудования», может быть использовано специалистами в области проектирования радиотехнических систем передачи информации.

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ	3
ПРЕДИСЛОВИЕ.....	11
1. ОБЩИЕ ПРИНЦИПЫ ОРГАНИЗАЦИИ РАДИОСВЯЗИ. ПЕРЕДАЧА НЕПРЕРЫВНЫХ СООБЩЕНИЙ.....	12
1.1 Рода и виды связи.....	12
1.2 Требования к связи.....	15
1.3 Структурная схема системы связи. Классификация систем и каналов авиационной радиосвязи	18
1.4 Модели непрерывных сообщений.....	21
1.4.1 Статистические характеристики и математическая модель непрерывных сообщений	21
1.4.2 Влияние разборчивости речи на безопасность поле- тов.....	25
1.5 Компрессия речевого сигнала. Вокодеры.....	26
1.5.1 Методы компрессии речевого сигнала.....	27
1.5.2 Непосредственная компрессия речевого процесса...	27
1.5.3 Методы компрессии с функциональным преобразо- ванием речи	35
1.6 Виды помех в каналах радиосвязи	42
1.7 Виды сообщений и сигналов	49
1.8 Оптимальный прием непрерывных сообщений и его поме- хоустойчивость	54
1.8.1 Принцип формирования и прием сигналов с относи- тельной фазовой модуляцией.....	58
1.8.2 Оптимальный прием аналоговых сигналов с частот- ной модуляцией	64
1.8.3 Оптимальный прием аналоговых сигналов с ампли- тудной модуляцией	
2. ПЕРЕДАЧА НЕПРЕРЫВНЫХ СООБЩЕНИЙ ПО ЦИФРОВЫМ КАНАЛАМ	73
2.1 Импульсная передача непрерывных сообщений	73
2.1.1 Характеристика цифровой обработки сигналов	73

2.1.2	Процессы преобразования сигналов при цифровой обработке	75
2.1.3	Импульсные радиосигналы с различными видами модуляции. Каналы связи с импульсной модуляцией	82
2.2	Цифровая передача непрерывных сообщений	86
2.2.1	Квантование сообщений по уровням	86
2.2.2	Каналы связи с импульсно-кодовой модуляцией и дельта-модуляцией	87
2.2.3	Методы получения сигналов ИКМ	89
2.2.4	Принципы построения декодеров	90
3.	ПЕРЕДАЧА ДИСКРЕТНЫХ СООБЩЕНИЙ	93
3.1	Уравнения дискретно-непрерывной фильтрации сигналов со случайно изменяющимися параметрами	93
3.1.1	Постановка задачи синтеза оптимального различителя сигналов на основе теории статистических решений.....	93
3.1.2	Основные положения теории статистических решений	94
3.2	Системы передачи с когерентной обработкой сигналов.....	100
3.2.1	Алгоритм оптимального демодулятора	100
3.2.2	Потенциальная помехоустойчивость	105
3.3	Синтез приемников цифровых радиосигналов со случайными параметрами	111
3.3.1	Синтез приемников цифровых амплитудно-манипулированных радиосигналов со случайными параметрами	111
3.3.2	Цифровые радиосигналы с различными видами манипуляции.....	115
4.	КОДИРОВАНИЕ СИГНАЛОВ ДАННЫХ.....	120
4.1	Назначение и классификация кодов.....	120
4.1.1	Модель цифрового канала связи. Характеристика сигналов данных	121
4.1.2	Основные определения и классификация кодов	123
4.1.3	Первичные коды, используемые в системах связи ВВС	129

4.1.4	Характер распределения ошибок в каналах связи.....	129
4.2	Помехоустойчивые коды и их основные характеристики.....	131
4.2.1	Принципы построения избыточных помехоустойчивых кодов.....	131
4.2.2	Классификация и основные характеристики избыточных кодов.....	136
4.3	Помехоустойчивые коды.....	146
4.3.1	Линейные коды.....	146
5.	СИСТЕМЫ СВЯЗИ С РЕТРАНСЛЯЦИЕЙ И МНОГО СТАНЦИОННЫМ ДОСТУПОМ.....	176
5.1	Системы авиационной наземной и воздушной радиосвязи..	176
5.1.1	Системы радиосвязи прямой видимости.....	176
5.1.2	Системы коротковолновой авиационной радиосвязи.....	179
5.1.3	Особенности радиосвязи с ВС на малых высотах. Ретрансляция сигналов.....	182
5.2	Общие сведения о многоканальной связи.....	183
5.2.1	Многоканальные системы связи.....	183
5.2.2	Методы линейного разделения сигналов.....	184
5.2.3	Временное разделение каналов.....	186
5.2.4	Частотное разделение каналов.....	189
5.2.5	Фазовое разделение каналов.....	191
5.2.6	Кодовое разделение каналов.....	193
5.2.7	Методы комбинированного разделения каналов.....	194
5.2.8	Помехи в многоканальных системах связи.....	195
5.3	Авиационно-космические системы радиосвязи.....	198
5.3.1	Спутниковые системы связи с воздушными судами и их особенности.....	198
5.3.2	Принципы построения перспективных систем спутниковой связи с многостанционным доступом.....	203
5.3.3	Использование космических ретрансляторов.....	205
6.	СИСТЕМЫ ПЕРЕДАЧИ ДАННЫХ.....	213
6.1	Общие сведения о системах передачи данных.....	213
6.1.1	Назначение и структура систем передачи данных (СПД).....	213

6.1.2 Структура сообщений в СПД.....	217
6.2 Сети передачи данных.....	219
6.2.1 Назначение и структура сетей передачи данных.....	219
6.2.2 Структура информационных сетей.....	221
6.3 Методы коммутации в сетях передачи данных.....	224
6.3.1 Выделенные каналы в сетях передачи данных.....	224
6.3.2 Метод коммутации каналов в сетях передачи дан- ных.....	225
6.3.3 Метод коммутации сообщений в сетях передачи данных.....	226
6.3.4 Метод коммутации пакетов в сетях передачи дан- ных.....	227
6.4 Методы повышения достоверности передачи информации в сетях передачи данных.....	230
6.4.1 Передача с повторением сообщений	231
6.4.2 Передача сообщений по параллельным каналам	231
6.4.3 Метод использования помехоустойчивого кодиро- вания.....	232
7. ПРИНЦИПЫ ПОСТРОЕНИЯ СОВРЕМЕННЫХ СЕТЕЙ СВЯЗИ....	233
7.1 Международные стандарты построения сетей связи.....	233
7.1.1 Эталонная модель взаимодействия открытых систем - Open System Interconnection (OSI).....	233
7.1.2 Уровни модели ЭМВОС (OSI).....	236
7.2 Протоколы обмена данными в сетях связи.....	241
7.2.1 Протоколы канального уровня.....	241
7.2.2 Протоколы сетевого уровня.....	247
7.2.3 Протоколы транспортного и вышестоящих уровней	254
7.3 Архитектура пакетных радиосетей и принципы их построе- ния.....	256
8. КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ ЗАЩИТЫ ДАННЫХ.....	260
8.1 Основы криптографической защиты данных.....	260
8.1.1 Понятие криптосистемы.....	260
8.1.2 Симметричные и асимметричные криптосистемы...	262
8.1.3 Понятие криптоанализа.....	263
8.2 Симметричные криптосистемы.....	264

8.2.1 Шифры перестановки.....	266
8.2.2 Система шифрования Цезаря.....	269
8.2.3 Аффинная система подстановок Цезаря.....	270
8.2.4 Система Цезаря с ключевым словом.....	271
8.2.5 Шифры сложной замены.....	272
8.2.6 Одноразовая система шифрования.....	273
8.2.7 Шифрование методом гаммирования.....	274
8.2.8 Стандарт шифрования данных DES.....	275
8.3 Концепция криптосистемы с открытым ключом.....	278
8.3.1 Однонаправленные функции.....	279
8.3.2 Криптосистема шифрования данных RSA.....	281
ЗАКЛЮЧЕНИЕ.....	284
ЛИТЕРАТУРА.....	286

СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ

АК	– авиационный комплекс
АКП	– авиационный комплекс перехвата
АИФ	– аналоговый интерполяционный фильтр
АСУ	– автоматизированная система управления
АЦП	– аналого-цифровой преобразователь
АО	– авиационное оборудование
АРМ	– автоматизированное рабочее место
АСК	– автоматизированная система контроля
АСУ	– автоматизированная система управления
АУСП	– авиационное управляемое средство поражения
АФ	– аналоговый фильтр
БАСК	– бортовая автоматизированная система контроля
ВВС	– Военно-воздушные силы
ВзПУ	– воздушный пункт управления
ВМФ	– Военно-Морской Флот
ВПП	– взлетно-посадочная полоса
ВС	– воздушное судно
ВСИ	– войсковое средство измерения
ВТА	– военно-транспортная авиация
ДА	– дальняя авиация
ЗАС	– засекречивающая аппаратура связи
ЗИП	– запасные части, инструмент, принадлежности
ЗРК	– зенитный ракетный комплекс
ИА	– истребительная авиация
ИСЗ	– искусственный спутник Земли
КС	– канал связи
КП	– командный пункт
КРУ	– командная радиолиния управления
МКК	– метод коммутации каналов
МКС	– многоканальные системы связи
НПУ	– наземный пункт управления
ОК	– объективный контроль
ОСНОД	– объединенная система навигации и обмена данными
ПВО	– противовоздушная оборона

ПД ТСР	– противодействие техническим средствам разведки
ППМ	– промежуточный пункт маршрута
ПУ	– пункт управления
ПН	– пункт наведения
ПЭВМ	– персональная электронно-вычислительная машина
РЛС	– радиолокационная станция
РСПИ	– радиосистемы передачи информации
РТК	– радиотехнический комплекс
РЭБ	– радиоэлектронная борьба
РЭО	– радиоэлектронное оборудование
РЭС	– радиоэлектронное средство
САЗО	– систем активного запроса-ответа
САЧ	– служебно-адресная часть
СВРС	– система воздушной радиосвязи
СОД	– система обмена данными
СОК	– средство объективного контроля
СПД	– система передачи данных
СПУ	– самолетное переговорное устройство
СС	– система синхронизации
ССС	– спутниковые системы связи
ТТХ	– тактико-технические характеристики
ФА	– фронтовая авиация
ФАПЧ	– фазовая автоподстройка частоты
ЦОС	– цифровая обработка сигналов
ЦАП	– цифро-аналоговый преобразователь
ЦВУ	– цифровое вычислительное устройство
УВХ	– устройство выборки-хранения
УЗО	– устройство защиты от ошибок
УС	– устройство синхронизации
ЭМО	– электромагнитная обстановка
ЭВМ	– электронно-вычислительная машина
OSPF	– <i>Open Shortest Path First</i> - протокол маршрутизации оценки состояния канала
IP	– <i>Internet Protocol</i> — адресный протокол
BGP	– Протокол функционирования межсетевого шлюза

- TCP – *Transport Control Protocol* — протокол управления передачей данных
- RSA – Криптосистема шифрования данных
- FTP – *File Transfer Protocol* — протокол передачи файлов
- DES – *Data Encryption Standard* — Стандарт шифрования данных
- FDDI – *Fiber Distributed Data Interface* — распределённый волоконный интерфейс данных
- DNS – *Domain Name System* — система доменных имён

ПРЕДИСЛОВИЕ

В соответствии с ГОС ВПО по специальности «Техническая эксплуатация транспортного радиооборудования» предусматривается изучение учебной дисциплины «Системы связи». Данное учебное пособие раскрывает вопросы государственного образовательного стандарта по указанному курсу.

Сложность и многообразие современных систем связи делают обязательным условием усвоения содержащегося в учебном пособии материала прочное знание обучающимися предусмотренных ГОС ВПО и изученных ранее учебных дисциплин: «Статистическая теория радиоэлектронных систем», «Прием и обработка сигналов», «Формирование и передача сигналов», «Антенны и устройства СВЧ», «Радиоизмерения», «Основы электродинамики и распространение радиоволн», «Радиотехнические цепи и сигналы».

В учебном пособии нашел отражение многолетний опыт преподавания учебной дисциплины на соответствующей кафедрах Ставропольского высшего военного авиационного инженерного училища (военного института) имени маршала авиации В. А. Судца, а так же Военного авиационного инженерного университета (г. Воронеж).

При написании учебного пособия соблюдалась его инженерная направленность. Поэтому авторы стремились, с одной стороны, помочь читателю овладеть современным математическим аппаратом, используемым для анализа и синтеза систем связи и отдельных их элементов, а с другой, не перегружать материал математическими выкладками и доказательствами. В этой связи авторы по возможности использовали упрощенные математические модели, для которых получаемые результаты оказываются достаточно понятными, и уделяли соответствующее внимание физическому толкованию получаемых результатов и их прикладной направленности.

Предисловие, гл. 3, 5, 6 и заключение написаны В. И. Сапожниковым, гл. 1, 2 – Г. И. Захаренко, гл. 4 – В. Е. Федосеевым, гл. 6, 7, 8 – С. И. Макаренко.

Авторы выражают глубокую признательность доктору технических наук, доценту П. А. Будко, кандидату технических наук, доценту А. П. Курляндчику за полезные замечания и советы, которые были учтены при подготовке пособия к изданию, а также С. Л. Щербаку за большой и кропотливый труд, проделанный при подготовке пособия к изданию.

ГЛАВА 1. ОБЩИЕ ПРИНЦИПЫ ОРГАНИЗАЦИИ РАДИОСВЯЗИ. ПЕРЕДАЧА НЕПРЕРЫВНЫХ СООБЩЕНИЙ

1.1 Рода и виды связи

Для реализации процесса связи в сложных условиях боевой обстановки и реальной местности применяются различные рода и виды связи.

Род связи – классификационная группировка военной связи, выделенная по среде распространения сигналов и типу линейных средств.

На рисунке 1.1 представлены **рода и виды военной связи**.

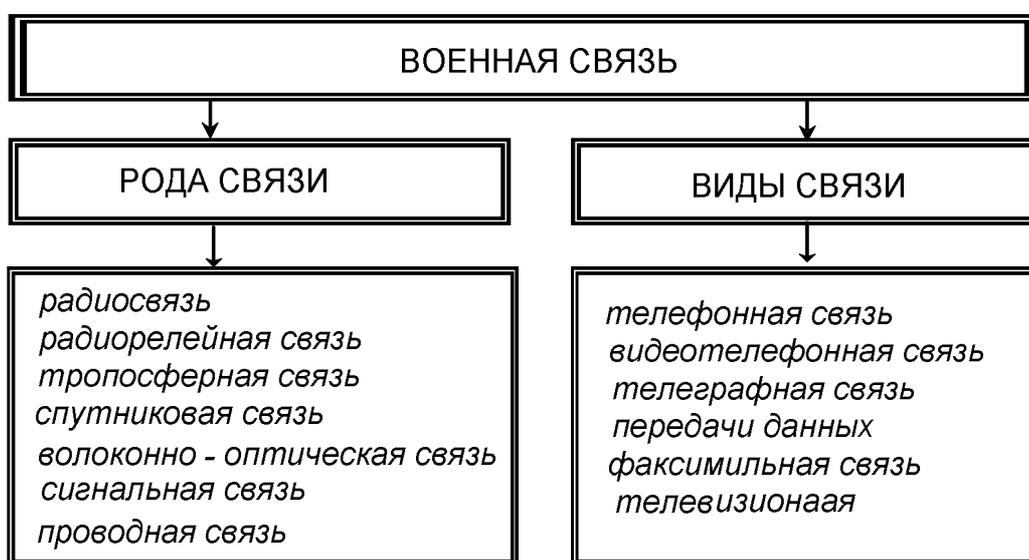


Рисунок 1.1 – Рода и виды связи

Радиосвязь – это род связи, который реализуется с использованием радиосредств, земных и ионосферных радиоволн. Радиосвязь применяется во всех звеньях управления. В тактическом звене управления радиосвязь является важнейшей, а во многих случаях единственной связью, способной обеспечить управление частями и подразделениями в самой сложной обстановке и при нахождении командиров в движении.

Радиорелейная связь – это род связи, который реализуется с использованием радиорелейных средств связи и радиоволн в ультракоротковолновом диапазоне. Радиорелейная связь применяется в звеньях управления от полка и выше.

Тропосферная связь – это род связи, который реализуется с использованием тропосферных средств связи и физического явления дальнего тропосферного распространения ультракоротких волн (ДТР УКВ). По своему на-

значению, боевому применению и качеству тропосферная связь аналогична радиорелейной. Тропосферная связь применяется в звеньях управления от дивизии и выше.

В настоящее время имеет место устойчивая тенденция повышения роли космической и спутниковой связи в военных системах связи. **Под космической связью понимается радиосвязь в интересах корреспондентов наземного, воздушного и морского базирования, имеющая общие участки распространения радиоволн за пределами ионосферы.**

Спутниковая связь – это частный случай космической связи, когда между двумя и более корреспондентами наземного, воздушного или морского базирования связь осуществляется с использованием ретранслятора, размещенного на искусственном спутнике земли.

Современные военные станции спутниковой связи обеспечивают связь на расстояниях от 5000 километров и более. В системе военной связи спутниковая связь находит применение в звене от батальона и выше, а также для связи с разведывательными группами и специальными отрядами (подразделениями).

Проводная связь – это связь, осуществляемая по проводным (кабельным) линиям связи. В системах проводной связи электрический сигнал передается по кабельной линии. Средства проводной связи обеспечивают высокое качество каналов, простоту организации связи, относительно большую скрытность по сравнению с радиосвязью, почти не подвержены воздействию преднамеренных помех. Проводная связь применяется во всех звеньях управления (от взвода (роты) и выше).

Волоконно-оптическая связь – это связь, осуществляемая по волоконно-оптическому кабелю и специальной аппаратуре преобразования электрических сигналов в оптические.

Все рода связи реализуются конкретными средствами связи: радиостанциями, радиорелейными, тропосферными станциями, станциями спутниковой связи, проводными средствами связи, волоконно-оптическими средствами связи. Эти средства образуют каналы связи: радио-, радиорелейные, тропосферные и т.д. Для каналов образующих средств каждого рода военной связи установлены условные обозначения, применяемые при разработке документов по связи.

Одна и та же по содержанию информация может быть представлена сообщениями различного вида: текстом, данными, изображением или речью. Так, например, боевая задача подразделению может быть поставлена в виде текстового документа на телеграфном бланке или на экране дисплея, в виде соответствующих условных знаков на топографической карте или доведена до командира подразделения в речевой форме. В зависимости от способа представления сообщений к виду удобному для восприятия различают виды связи.

Вид военной связи - это классификационная группировка военной связи, выделенная по виду передаваемого сообщения (оконечного оборудования или средства связи). При использовании соответствующей оконечной аппаратуры по каналам радио-, радиорелейных, тропосферных, спутниковых, проводных (кабельных) линий связи обеспечиваются *следующие виды связи*:

- телефонная связь,
- видеотелефонная связь,
- телеграфная связь,
- передача данных,
- факсимильная связь,
- телевизионная связь.

Сигнальная связь - это связь, осуществляемая с помощью заранее определенных зрительных и звуковых сигналов управления. В настоящее время для управления боем используются зрительные средства (световые ракеты, цветные дымы и др.) и звуковые средства (сирены, свистки и др.).

Телеграфную связь, передачу данных и факсимильную связь принято объединять понятием «документальная связь». Дадим предназначение и краткую характеристику каждого вида связи.

Телефонная связь – это вид электросвязи, обеспечивающий передачу (прием) речевой информации, переговоры должностным лицам органов управления. Телефонная связь создает условия, близкие к личному общению, поэтому является наиболее удобной в тактическом звене управления, но сохраняет свое значение и в других звеньях управления. С целью скрытия от противника содержания телефонных переговоров в каналах связи применяется аппаратура засекречивания или устройства технического маскирования речи. В зависимости от применяемой оконечной и специальной аппаратуры телефонная связь может быть открытой, маскированной, засекреченной временной или гарантированной стойкости.

Телеграфная связь – вид электросвязи, обеспечивающий обмен телеграммами (краткими текстовыми сообщениями) и переговоры должностным лицам органов управления с применением средств телеграфной связи. Кроме того она предназначается для передачи документальных сообщений в виде шифрограмм, кодограмм.

Телеграфная связь может быть буквопечатающей или слуховой, засекреченная или открытая (с применением аппаратуры засекречивания или без ее применения). Телеграммы, несущие важную информацию, могут предварительно шифроваться или кодироваться.

Факсимильная связь – это вид электросвязи, обеспечивающий обмен документальной информацией в цветном и черно-белом изображении. Она

предназначается для передачи документов в виде карт, схем, чертежей, рисунков и буквенно-цифровых текстов в черно-белом или цветном изображении. Данная связь представляет большое удобство должностным лицам органов управления, так как на приемном устройстве получается готовый для дальнейшей работы документ с соответствующими подписями и печатями. Факсимильная связь находит применение в оперативном и стратегическом звеньях управления.

***Передача данных** – это вид электросвязи, обеспечивающий обмен формализованными и неформализованными сообщениями между электронно-вычислительными комплексами, автоматизированными рабочими местами должностных лиц пунктов управления.* Она предназначается для обмена информацией в автоматизированных системах управления войсками и оружием (АСУВО). Под **данными** будем понимать информацию, представленную в виде, пригодном для автоматической обработки.

***Видеотелефонная связь** – это вид электросвязи, обеспечивающий переговоры должностных лиц органов управления с одновременной передачей подвижных изображений.* Данный вид связи применяется только в высших звеньях управления.

***Телевизионная связь** – это вид электросвязи, обеспечивающий передачу боевой обстановки и других событий на местности в реальном масштабе времени.* Она используется в высших звеньях управления.

1.2 Требования к связи

С целью обеспечения эффективного функционирования системы управления войсками и оружием к процессу управления предъявляются требования: *устойчивость, непрерывность, оперативность и скрытность.* Эти требования выражают основные свойства процесса управления. Выполнение этих требований в основном определяется качеством информационного процесса в системе управления, т. е. качеством связи. Военная связь должна обеспечить передачу сообщений от источника к получателю в установленный срок без искажений и с сохранением в тайне от противника содержащейся в нем информации. Другими словами к связи *предъявляются требования: по своевременности, по достоверности, по безопасности.* Только при выполнении этих трех основных условий можно полагать, что связь выполнит свои задачи. Если хотя бы одно из этих требований не будет выполнено, сообщение считается не доставленным. При перехвате сообщения противником связь может оказать отрицательное влияние на успех боя.

Предъявление требований к связи, системам, сетям, комплексам и средствам связи является сложным процессом и основывается на теории эффек-

тивности. Объектом исследования теории эффективности являются системы и процессы оценки эффективности, а предметом исследования – закономерности создания, функционирования и развития систем и процессов оценки эффективности. Военные связисты должны уметь оценивать эффективность процесса связи и функционирования системы связи. Оценка эффективности связи и системы связи связана с такими понятиями как: свойство, характеристика, показатель, критерий, требование. Рассмотрим эти понятия.

Свойство – это выражение такой стороны предмета (объекта, процесса), которая обуславливает его различие или общность с другими предметами (объектами, процессами) или обнаруживается в его отношении с ними. Каждый предмет, объект или процесс имеет основные и второстепенные свойства. Основными свойствами связи являются своевременность, достоверность и безопасность.

Характеристика – это описание отличительных качественных свойств чего-либо или кого-либо. Характеристика может быть качественной или количественной. Для качественной характеристики систем и процессов используют понятия "лучше" или "хуже". Количественные характеристики часто называют параметрами. Для оценки эффективности процесса связи чаще пользуются понятием показателя эффективности.

Показатель эффективности (показатель) – это характеристика, функция характеристик или правило, выбранное для оценки некоторого свойства объекта или совокупности его свойств. Показатели разделяют на количественные и качественные. Пример: своевременность связи может оцениваться такими показателями как вероятность своевременной передачи сообщений и среднее время пребывания сообщения в канале связи.

Количественные показатели выражаются в виде функционала от параметров объекта и определяют абсолютную или относительную (доля, часть) числовую меру проявления свойства или совокупности свойств.

Качественные или порядковые показатели оперируют понятиями какой-либо установленной шкалы суждений, отражающей предпочтение (типа: хуже, лучше, больше, меньше и т.п.), либо бальной шкалы или шкалы весовых коэффициентов.

Оценка эффективности процесса связи напрямую связана с ответом на вопрос: «В какой степени достигнута требуемая эффективность?» Для ответа на этот вопрос используют понятие критерия.

Критерий (эффективности) – это признак, правило, мера суждения, на основании которых проводится оценка или классификация чего-либо. Критерий может быть абсолютным или относительным. **Абсолютный критерий** выражает предельную меру достигаемого эффекта для сравнительной оценки возможных альтернативных решений. **Относительный критерий** – это нор-

ма оценки показателя для достижения требуемой эффективности. Практика показывает, что не всегда экономически целесообразно достигать максимальных или оптимальных критериальных значений. В таких случаях устанавливают целесообразную величину требований к показателям.

Требование – это установленный количественный уровень значения характеристики или показателя оцениваемого объекта, либо высказывательная форма, устанавливающая его качественное соответствие поставленным целям или решаемым задачам.

Таким образом, под **эффективностью связи (системы связи)** будем понимать ее свойство соответствовать цели функционирования.

К современным системам радиосвязи предъявляются следующие основные требования:

Оперативность (своевременность) – обмен информацией осуществляется в сроки, обусловленные потребностями управления. Оценивается вероятностью того, что сообщение будет полностью доставлено требуемому адресату в течение заданного времени.

Достоверность связи характеризует способность системы связи обеспечить воспроизведение переданных сообщений в пунктах приема с заданной точностью. Достоверность передачи цифровых сообщений определяется вероятностью правильного приема код/комбинаций первичного кода, соответствующих переданным символам.

Помехоустойчивость связи – способность (свойство) системы выполнять прием/передачу сообщений в условиях действия помех. Для количественной оценки помехоустойчивости используют критерий достоверности передачи сообщений.

Надежность связи – свойство системы связи выполнять передачу сообщений, сохраняя при этом в течение заданного промежутка времени значение основных характеристик (достоверность) в заданных пределах. Количественные характеристики надежности: вероятность отказа в заданном интервале времени, время наработки на отказ.

Скрытность радиосвязи – способность системы противостоять раскрытию противником содержания передаваемой информации и мест расположения пунктов управления и аэродромов. Скрытность осуществляется за счет: минимальной мощности ПРД; направленности антенны; уменьшения времени передачи; увеличения времени между сеансами связи; радиомаскировки и соблюдения правил радиообмена применение специальной аппаратуры и документов связи.

1.3 Структурная схема системы связи. Классификация систем и каналов авиационной радиосвязи

Обобщенная структурная схема простейшей одноканальной системы радиосвязи представлена на рисунке 1.2.



Рисунок 1.2 – Обобщенная структурная схема простейшей одноканальной системы радиосвязи

Отправителем и получателем сообщения могут выступать как человек, так и различного рода технические устройства.

Радиосвязь осуществляется посредством радиоволн, которые распространяясь в воздушном пространстве образуют линию связи. *Линия связи* – физическая среда и совокупность аппаратных средств, используемых для передачи сигналов от ПРД к ПРМ.

Совокупность передающего устройства, линии связи, приемного устройства образуют канал связи.

Канал связи – совокупность средств, обеспечивающих передачу сигнала от отправителя сообщения к получателю сообщения.

Совокупность отправителя сообщения, канала радиосвязи, получателя сообщения образуют **простейшую систему радиосвязи**. Характерным для самолетных радиостанций является объединение приемного и передающего устройств, расположенных в одном конструктивном блоке - **радиостанции**.

В системах управления различного назначения процесс управления непосредственно с передачей информации. Поэтому передача информации занимает важное место в системе управления. Осуществление функций управления вообще невозможно без организации обмена информацией между управляемыми объектами системы.

Для организации авиационной радиосвязи ВС и НПУ оборудованы соответствующими радиотехническими средствами. Важное место среди них занимают бортовые и наземные радиостанции. Современные отечественные радиостанции имеют относительно высокие ТТХ и обеспечивают надежную радиосвязь при управлении авиацией на всех этапах ее применения.

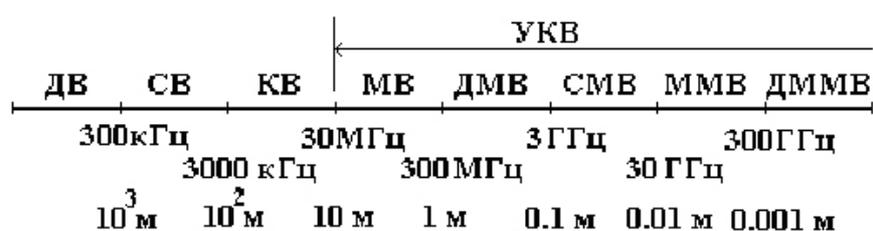
При этом системы радиосвязи классифицируются по следующим признакам.

Авиационная радиосвязь подразделяется на воздушную и наземную.

В зависимости от диапазона используемых радиоволн в соответствии с Регламентом радиосвязи выделяют девять диапазонов. Разделение радиоволн на диапазоны в первую очередь связано с особенностями их распространения и использования. Принятая в настоящее время классификация видов радиоволн приведена в таблице 1.1.

Таблица 1.1 – Классификация видов радиосвязи

Номер диапазона	Наименование диапазона		Границы диапазона		Способ распространения		
	радиоволн	радиочастот	радиоволн	радиочастот	на близкие расстояния	на большие расстояния	
						днем	ночью
4	Мериаметровые сверхдлинные (СДВ)	Очень низкие (ОНЧ)	100...10 км	3...10 кГц	Земные и ионосферные волны	Ионосферные волны	Ионосферные волны
5	Километровые длинные (ДВ)	Низкие (НЧ)	10...1 км	30...300 кГц	То же	То же	То же
6	Гектометровые средние (СВ)	Средние (СЧ)	1 км...100 м	0,3...3 МГц	Земные волны	Земные волны	–"
7	Декаметровые короткие (КВ)	Высокие (ВЧ)	100...10 м	3...30 МГц	То же	Ионосферные волны	–"
8	Ультракороткие метровые (МВ)	Очень высокие (ОВЧ)	10...1 м	30...300 МГц	–"	Тропосферные и ионосферные	Тропосферные и ионосферные
9	Дециметровые (ДЦВ)	Ультравысокие (УВЧ)	1 м...1 дм	300...3000 МГц	–"	Прямые и тропосферные волны	Прямые и тропосферные волны
10	Сантиметровые (СМВ)	Сверхвысокие (СВЧ)	10...1 см	3...30 ГГц	–"	То же	То же
11	Миллиметровые (ММВ)	Крайне высокие	10...1 мм	30...300 ГГц	Прямые волны	Прямые волны	Прямые волны
12	Децимиллиметровые (ДЦМВ)	Гипервысокие (ГВЧ)	1...0,1 мм	(0,3...3)·10 ³ ГГц	То же	То же	–"



В зависимости от используемой линии связи различают системы радиосвязи:

- прямой видимости;
- тропосферные;
- ионосферные;
- космические (спутниковые системы связи);

- системы радиорелейной связи;
- подвижные средства связи (сотовые системы связи).

В зависимости от вида передаваемых сообщений различают системы:

- телефонной;
- телеграфной;
- телевизионной;
- факсимильной связи;
- системы передачи данных (телекодовая связь).

По виду радиосигналов системы радиосвязи делятся на три группы:

- системы передачи аналоговых сигналов – передаваемое сообщение и радиосигнал являются непосредственными функциями времени (аналоговые системы связи с АМ, ЧМ, ФМ...);
- системы передачи цифровых сигналов – передаваемое дискретное сообщение и радиосигнал имеют конечное число состояний (цифровые системы радиосвязи с АМн, ЧМн, ФМн, ОФМн);
- системы передачи импульсных сигналов (импульсные системы радиосвязи).

В зависимости от значения базы сигнала различают:

- широкополосные системы связи;
- узкополосные системы связи.

По количеству одновременно передаваемых сообщений различают:

- одноканальные системы связи;
- многоканальные системы связи;

По направлению обмена сообщениями различают:

- односторонние (одна из радиостанций осуществляет только передачу, а другая или другие – только прием);
- двухсторонние (радиостанции осуществляют передачу и прием).

По порядку обмена сообщениями:

- симплексные системы связи (передачи и прием в радиостанции осуществляется отдельно);
- дуплексные системы связи (передача осуществляется одновременно с приемом);
- полудуплексные системы связи (относятся к симплексным системам, в которых предусматривается автоматический переход от передачи на прием и возможность переспроса).

По способу защиты передаваемой информации:

- открытой радиосвязи;
- закрытой (засекреченной) радиосвязи (сам факт передачи информации не скрывается, а лишь засекречивается информация).

По степени автоматизации обмена информацией различают цифровые сис-

темы радиосвязи:

- неавтоматизированные (оператор);
- автоматизированные (вручную только ввод, а кодирование и т.д. автоматически);
- автоматические (обмен между автоматическими устройствами и ЭВМ без участия оператора).

По виду модуляции и классу излучения – классификация приведена в таблице 1.2

Таблица 1.2 - Краткие обозначения вида модуляций и классов излучений радиопередатчиков

Обозначение класса излучения	Вид модуляции, соответствующий классу излучения
A0	Немодулированная несущая
A1	Амплитудная манипуляция
A2	Амплитудная тональная манипуляция
A3	Амплитудная модуляция
A3A	Однополосная модуляция с ослабленной несущей
A3H	Однополосная модуляция с полной несущей
A3J	Однополосная модуляция с подавленной несущей
F1	Частотная манипуляция
F2	Частотная тональная манипуляция
F3	Частотная модуляция
F6(2F1)	Двойная частотная манипуляция
F9	Относительная фазовая манипуляция
P0	Импульсы с высокочастотным заполнением без модуляции
P3D	Амплитудно-импульсная модуляция
P3E	Широтно-импульсная модуляция
P3F	Фазоимпульсная модуляция
P3G	Импульсно-кодовая модуляция

1.4 Модели непрерывных сообщений

1.4.1 Статистические характеристики и математическая модель непрерывных сообщений

Самым распространенным аналоговым сообщением в авиационных системах радиосвязи является речь.

Она представляет собой непрерывный случайный процесс, который об-

разуется в результате произношения звуков. В речи не существует двух одинаковых звуков. Каждый звук в отдельности представляет собой реализацию случайного процесса с соответствующими вероятностными характеристиками.

Различают (рисунок 1.3) **гласные и согласные** звуки. В свою очередь согласные звуки бывают:

- глхими; (С, Ш, Ф, Х, Ц, Ч ...);
- звонкими; (З, Ж, М, Н, Л ...);
- взрывными (Г, Б, П, К, Т....).

При этом:

- дисперсия гласных звуков равна $D_г \approx 2 \times D_\lambda$;
- дисперсия звонких согласных несколько ниже и равна $D_{зс} \approx 0,1 \times D_\lambda$;
- дисперсия глухих согласных мала и выражается следующим равенством $D_{гс} \approx 0,01 \times D_\lambda$;
- дисперсия взрывных согласных равна $D_{вс} \approx 0$, где D_λ - средняя дисперсия речи.

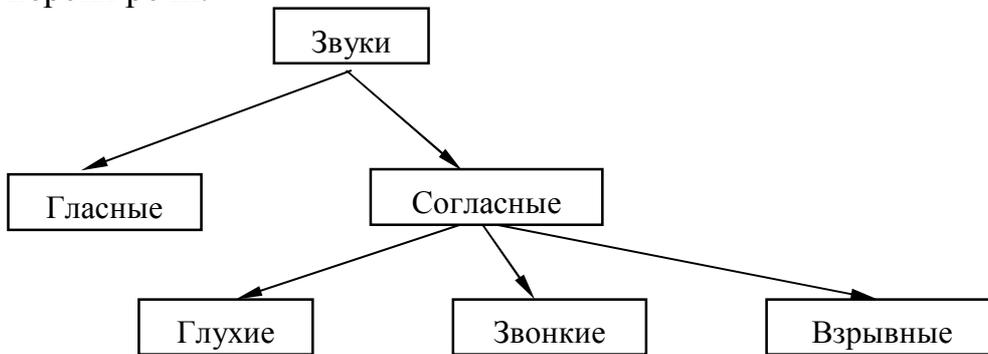


Рисунок 1.3 – Классификация звуков

Дисперсия звуков речи изменяется в пределах: 35...250 мс, а средняя дисперсия равна $T_{ср} = 120$ мс. В симплексных системах воздушной телефонной радиосвязи передаются короткие речевые сообщения длительностью 10...15 с. При этом вероятности передачи в данный момент:

- гласного звука $P_г = 0,5$;
- звонкого, глухого и взрывного согласных звуков соответственно равны $P_{зс} = 0,13$, $P_{гс} = 0,15$, $P_{вс} = 0,22$.

Плотность вероятности отдельных звуков близка к нормальной и поэтому в общем случае плотность вероятности речи, как случайного процесса, может быть представлена формулой полной вероятности в виде:

$$p(\lambda) = P_z \cdot p_z(\lambda) + P_{zc} \cdot p_{zc}(\lambda) + P_{zc} \cdot p_{zc}(\lambda) + P_{вс} \cdot p_{вс}(\lambda) =$$

$$= \frac{1}{2\pi D_\lambda} \left[\frac{0,5}{2} e^{\left(\frac{\lambda^2}{4D_\lambda}\right)} + \frac{0,13}{0,1} e^{\left(\frac{\lambda^2}{0,2D_\lambda}\right)} + \frac{0,15}{0,01} e^{\left(\frac{\lambda^2}{0,02D_\lambda}\right)} + 0,22\delta(\lambda) \right], \quad (1.1)$$

где $p_z(\lambda)$, $p_{zc}(\lambda)$, $p_{zc}(\lambda)$, $p_{вс}(\lambda)$ – указанные выше плотности вероятности речевого процесса при условии, что в данный момент времени передается гласный, звонкий или взрывной согласный звук.

При передаче речи в условиях шума в последнем выражении к дисперсии звуков необходимо добавить дисперсию шума.

Возможны и другие аппроксимации плотности вероятности: экспоненциальная, функция Лапласа, гамма-функции.

График плотности вероятности речевого процесса представлен на рисунке 1.4.

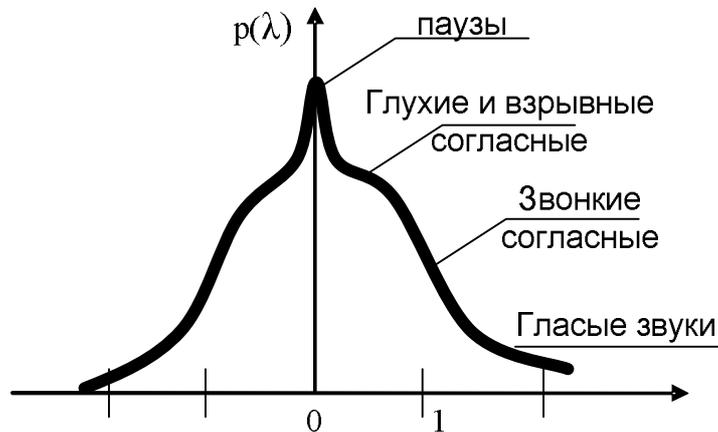


Рисунок 1.4 – График плотности вероятности речевого процесса

На приведенном графике следует отметить три характерных участка:

- верхний соответствует паузам;
- средний соответствует согласным звукам;
- нижний соответствует гласным звукам.

Нормированная корреляционная функция речевого сообщения аппроксимируется выражением:

$$r(\tau) = e^{-2\pi f_1|\tau|} \cdot \cos(2\pi f_2\tau), \quad (1.2)$$

где $f_1 = 500$ Гц, $f_2 = 1000$ Гц.

Соответствующий график представлен на рисунке 1.5.

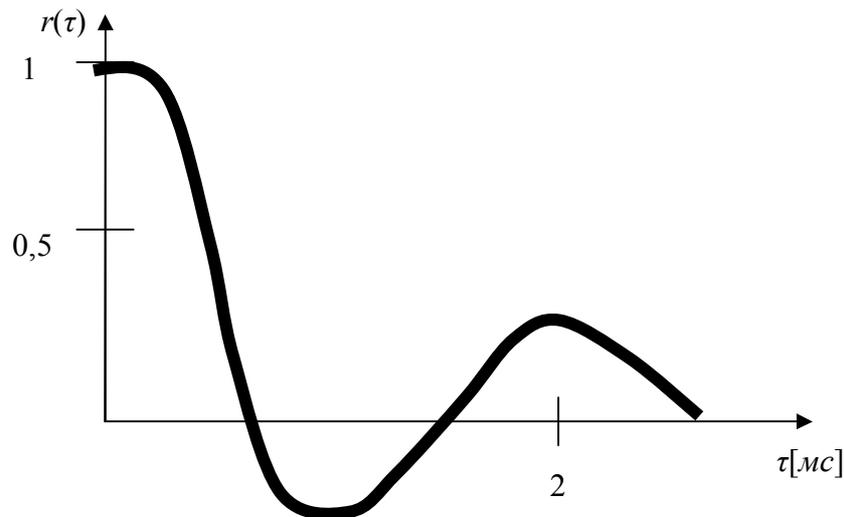


Рисунок 1.5 – Нормированная корреляционная функция речевого сообщения

Спектральную плотность речевого сообщения достаточно хорошо можно аппроксимировать выражением вида:

$$S(f) = 2f_1 \cdot D_\lambda \left[\frac{1}{f_1^2 + 4\pi^2 (f_2 - f)^2} + \frac{1}{f_1^2 + 4\pi^2 (f_2 + f)^2} \right], \quad (1.3)$$

где : $f_1 = 1000$ Гц, $f_2 = 500$ Гц.

Зависимость спектральной плотности от частоты представлена на рисунке 1.6.

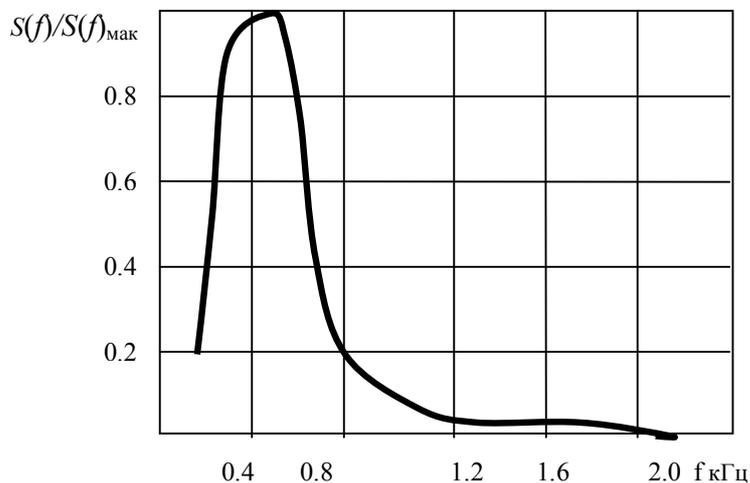


Рисунок 1.6 – Зависимость спектральной плотности от частоты

Из представленной зависимости (1.3) видно, что основная доля энергии речевого сообщения сосредоточена в диапазоне 200...800 Гц. Максимум спек-

тральной плотности наблюдается на частоте $f = 500$ Гц.

В качестве модели речевого сообщения целесообразно рассматривать процесс на выходе интегрирующей цепочки, на вход которой подается белый формирующий шум. Более точная модель речевого процесса может быть представлена формирующим фильтром, состоящим из последовательно соединенных интегрирующей и дифференцирующей цепочек.

При дальнейшем изложении будем пользоваться первой моделью.

1.4.2 Влияние разборчивости речи на безопасность полетов

Восприятие человеком речи осуществляется органами слуха, которые представляют собой спектроанализатор параллельного действия. Эквивалентная полоса частот фильтров слухового аппарата среднестатистического человека составляет от 50 Гц на нижних частотах до 16000 Гц на высоких частотах. Звуковое колебание на частотах меньше 50 Гц непосредственно слуховыми органами человека не воспринимаются.

Качество телефонной связи характеризуется **разборчивостью речи**. Различают разборчивость:

- звуков,
- слов,
- фраз.

Разборчивость определяется экспериментально и расчетным методом. В основе экспериментального определения разборчивости речи положены артикуляционные измерения (ГОСТ-16600). Артикуляционные измерения могут проводиться в лабораторных или в реальных условиях применения средств радиосвязи.

В лабораторных условиях обычно измеряется разборчивость звуков (Z), которая определяется как отношение:

$$Z = \frac{N_{\text{правильно принятых звуков}}}{N_{\text{переданных звуков}}} \cdot 100\% . \quad (1.4)$$

При необходимости по известной разборчивости звуков определяется разборчивость слов (W) путем использования соответствующих графиков пересчета.

Характеристика классов качества ТЛФ связи (согласно ГОСТ-16600) приведены в таблице 1.3.

Таблица 1.3 - Характеристики класса качества речи

Класс качества	Характеристика класса качества	Z, %	W, %
I	Понимание речи без малейшего напряжения внимания	> 91	> 98
II	Понимание речи без затруднений	85-90	94-97
III	Понимание речи с напряжением внимания без переспросов и повторений	78-84	89-93
IV	Понимание речи с напряжением внимания без переспросов и повторений	61-77	70-88
V	Срыв связи	< 60	< 69

1.5 Компрессия речевого сигнала. Вокодеры

1.5.1 Методы компрессии речевого сигнала

Речевые сообщения (сигналы) обладают естественной избыточностью. Она настолько значительна, что объем информации, переносимой речевым сигналом, оказывается по крайней мере на порядок меньше объема речевого сигнала. Поэтому объем сигнала можно сократить без сокращения информации, содержащейся в нем. Сокращение объема сигнала представляет большой интерес для практики, т.к. сигналы меньшего объема можно передавать по более простым и дешевым каналам (каналам, имеющим малый объем V_k).

Уменьшение объема речевого сигнала, т.е. его сжатие (компрессия), достигается при помощи соответствующих преобразований речевого сигнала. Последние делятся на методы непосредственной компрессии и методы функционального преобразования.

Для методов непосредственной компрессии характерным является частичное сохранение микроструктуры речи. При использовании методов непосредственной компрессии преобразованный речевой сигнал, передаваемый по каналу связи, сохраняет явные признаки речи.

Для методов функционального преобразования речевого сигнала характерным является "разрушение" речевого сигнала, замена речевого сигнала его медленно изменяющимися параметрами. Последние, а не сам речевой сигнал, передаются по каналу связи.

Из выражения для объема сигнала $V_c = D_c \cdot T_c \cdot \Delta F_c$ следует, что сжатие (непосредственная компрессия) может быть осуществлено уменьшением:

- динамического диапазона (амплитудная компрессия);
- ширины частотного спектра (частотная компрессия);
- длительности сигнала (временная компрессия).

1.5.2 Непосредственная компрессия речевого процесса

Амплитудная компрессия (сжатие) речевого сигнала предназначена для сжатия динамического сигнала в устройстве передачи, что позволяет улучшить соотношение сигнала к шуму в канале связи.

На приемной стороне производится обратная операция над принятым сигналом - экспандирование (расширение), т.е. восстановление сжатого ранее динамического диапазона до его первоначального значения, вследствие чего уровни сигнала достигают исходных величин.

Совместно эти два устройства - компрессор и экспандер - называются **компандером**.

Амплитудная компрессия осуществляется с помощью нелинейных устройств путем автоматического регулирования уровня речи, либо путем амплитудного ограничения речи.

В системе с автоматическим регулированием усиление системы уменьшается при увеличении уровня речи, в результате чего происходит сближение ее максимального и минимального уровней. Искажение естественной речи приводит к уменьшению разборчивости даже при отсутствии помех. Однако опыт показывает, что это уменьшение незначительно даже при сильном сжатии динамического диапазона. С другой стороны, при заданной пиковой мощности передатчика такие искажения позволяют повысить разборчивость при наличии помех, т.к. максимальный уровень речи на выходе компрессора будет таким же, как и на входе, а более низкие уровни речи окажутся приподнятыми. Таким образом, благодаря «подтягиванию» слабых составляющих речи увеличивается ее разборчивость на фоне помех, т.к. уровень спектра речи на выходе приемника окажется приподнятым относительно уровня шума при сохранении общего уровня ниже порога болевого ощущения. Описанный принцип компрессии динамического диапазона используется в системах автоматической регулировки глубины модуляции (АРГМ). Принцип действия такой системы показан на рисунке 1.7.

При увеличении уровня речи напряжение с выхода выпрямителя, минусом подаваемое на управляющий электрод усилительного элемента 1-ой ступени предварительного усилителя низкой частоты (УНЧ), возрастает, благодаря чему коэффициент усиления уменьшается, а модулирующее напряжение изменяется незначительно.

ДОСТОИНСТВА: незначительное повышение помехоустойчивости.

НЕДОСТАТОК: инерционность схемы (компрессора) вызывает искажения взрывных согласных звуков (Б, П, Д, Т и т.д.).

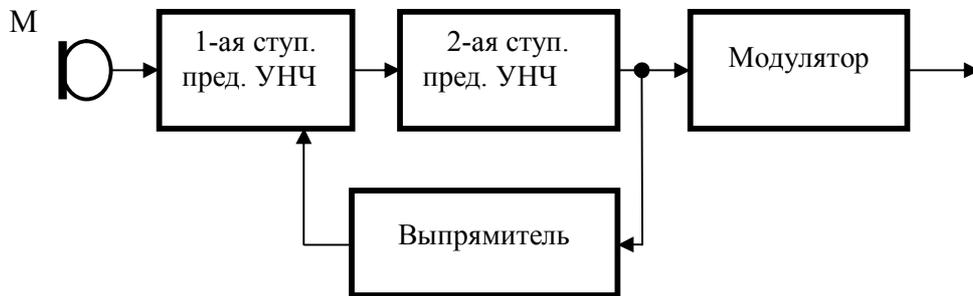


Рисунок 1.7 – Структурная схема системы автоматической регулировки глубины модуляции

Амплитудное ограничение. Более эффективной является компрессия динамического диапазона речи путем амплитудного ограничения. Этот метод компрессии является безинерционным. На рисунке 1.8 приведены характеристики амплитудных ограничений по максимуму (сверху) (а) и по минимуму (снизу) (б).

Различают жесткое и плавное амплитудное ограничения. Жесткому ограничению соответствует линейно-ломаная характеристика ограничителя плавному - характеристика с плавным загибом. Ограничение сверху чаще бывает жестким, ограничение снизу - плавным.

Под степенью ограничения при ограничении сверху понимается величина:

$$\gamma = 20 \lg \frac{U_m}{U_{огр}}, [\text{дБ}], \quad (1.5)$$

где U_m - максимальное значение напряжения на входе ограничителя;
 $U_{огр}$ - напряжение, соответствующее началу ограничения.

При ограничении снизу степень ограничения равна:

$$\gamma = 20 \lg \frac{U_m}{U_m - U_{огр}}, [\text{дБ}]. \quad (1.6)$$

Влияние ограничения снизу и сверху на разборчивость речи различно. Незначительное ограничение снизу вызывает существенное понижение разборчивости речи. Наоборот, даже при достаточно сильных ограничениях сверху разборчивость речи падает незначительно. На рисунке 1.9 показана зависимость разборчивости звуков при $Z = 80\%$ от степени ограничения γ (Дб). Кривая 1 - ограничение сверху, кривая 2 - ограничение снизу.

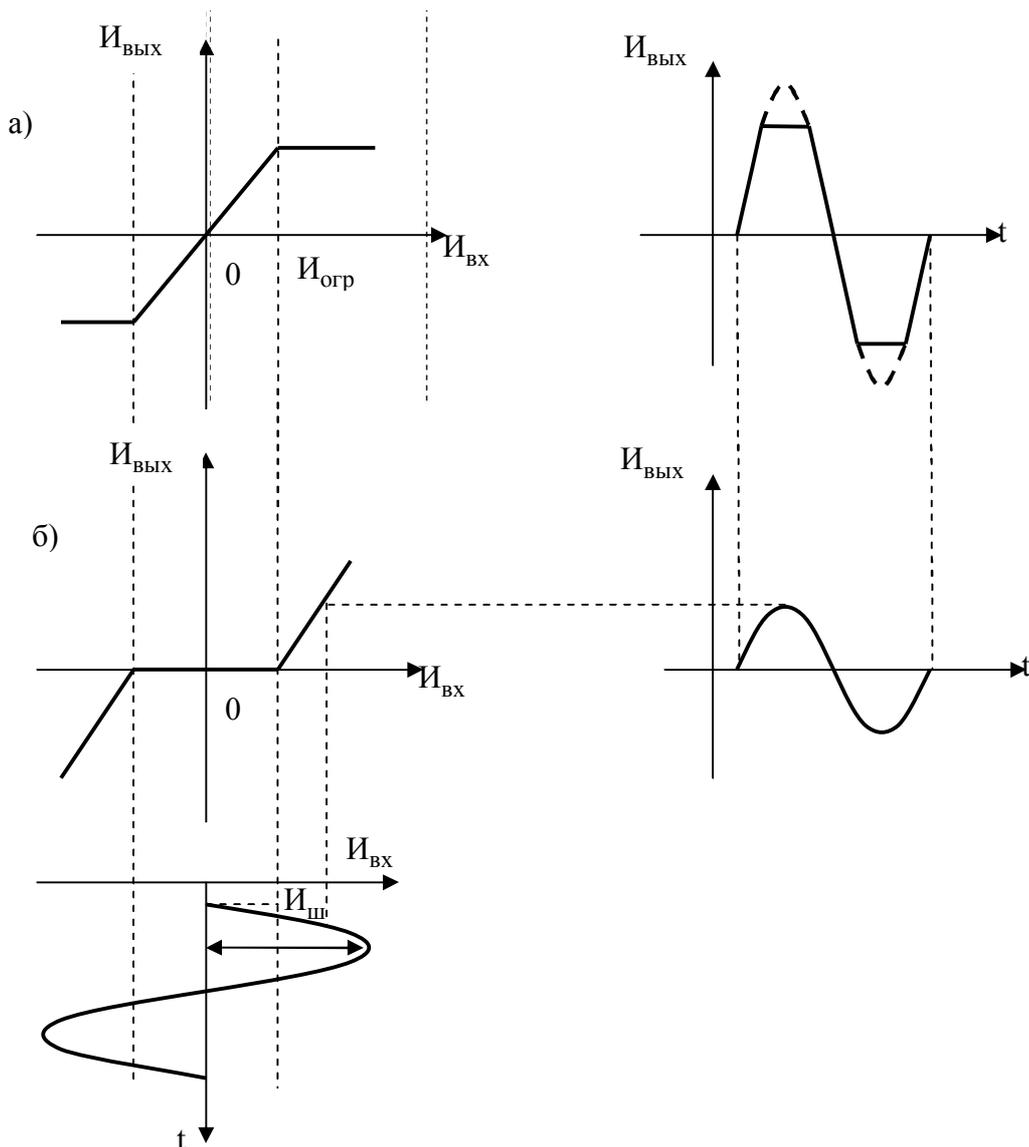


Рисунок 1.8 – Характеристика амплитудных ограничений по максимуму и по минимуму

Установлено, что высокая разборчивость речи сохраняется даже при предельном ограничении, называемом клиппированием. Клиппированная речь имеет вид прямоугольных импульсов различной длительности. Единственным признаком клиппированной речи является последовательность нулевых переходов напряжения, отображающего речевой процесс (рисунок 1.10 б). Клиппирование речи сводится к устранению амплитудной информации, к выделению и передаче частотной информации, содержащейся в функции $\cos(\varphi(t))$, т.е. передаче речи постоянного уровня функции $F(t) = U F(t) \cdot \cos(\varphi(t))$.

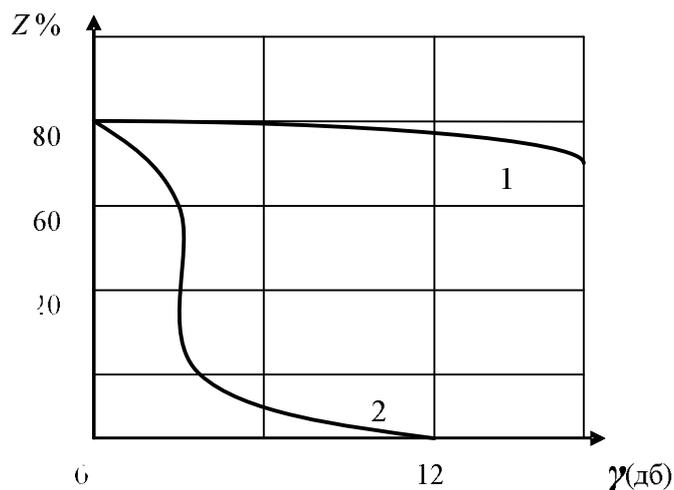


Рисунок 1.9 – Зависимость разборчивости звуков от степени ограничения

Разборчивость сохраняется еще лучше, если речь предварительно дифференцируется, а затем производится ограничение производной речевого процесса (рисунок 1.10 *в* и *г*). Это объясняется тем, что число экстремальных значений в речевом сообщении больше числа нулевых значений примерно в 2 раза для мужского голоса. Следовательно, при сохранении положений экстремальных значений с помощью дифференцирования сохраняется и передается в линию связи более полная информация о структуре сигнала.

Оптимальная степень ограничения в основном зависит от вида модуляции, от мер принятых для борьбы с нелинейными искажениями, а также от условий, в которых ведется передача речевых сообщений. Опыты показывают, что в системе связи с АМ при передаче из тишины оптимальным является ограничение на 18...24 Дб относительно пикового уровня речи. Достаточная разборчивость сохраняется при степени ограничения 40 Дб, но без помех. При наличии помех ограничение на 18...24 Дб обеспечивает более высокую разборчивость по сравнению с разборчивостью неограниченного сигнала до тех пор, пока отношение сигнал/шум остается больше единицы. При более интенсивных помехах разборчивость ограниченного сигнала резко снижается, а неограниченного сигнала - менее резко.

При передаче в условиях сильных акустических помех ($B_{ш} = 110...115$ Дб), степень ограничения приходится уменьшать до 12...16 Дб относительно пикового уровня речи в тишине. Это объясняется тем, что в условиях сильных акустических шумов оператор непроизвольно повышает громкость речи при надетом шлемофоне примерно на 6...8 Дб.

Если уровень ограничения относительно максимального уровня речи в тишине оставить прежним, тогда сильному ограничению, а следовательно, и сильному искажению будут подвергаться и наиболее) слабые согласные звуки.

Искажение согласных звуков приведет к значительному снижению разборчивости.

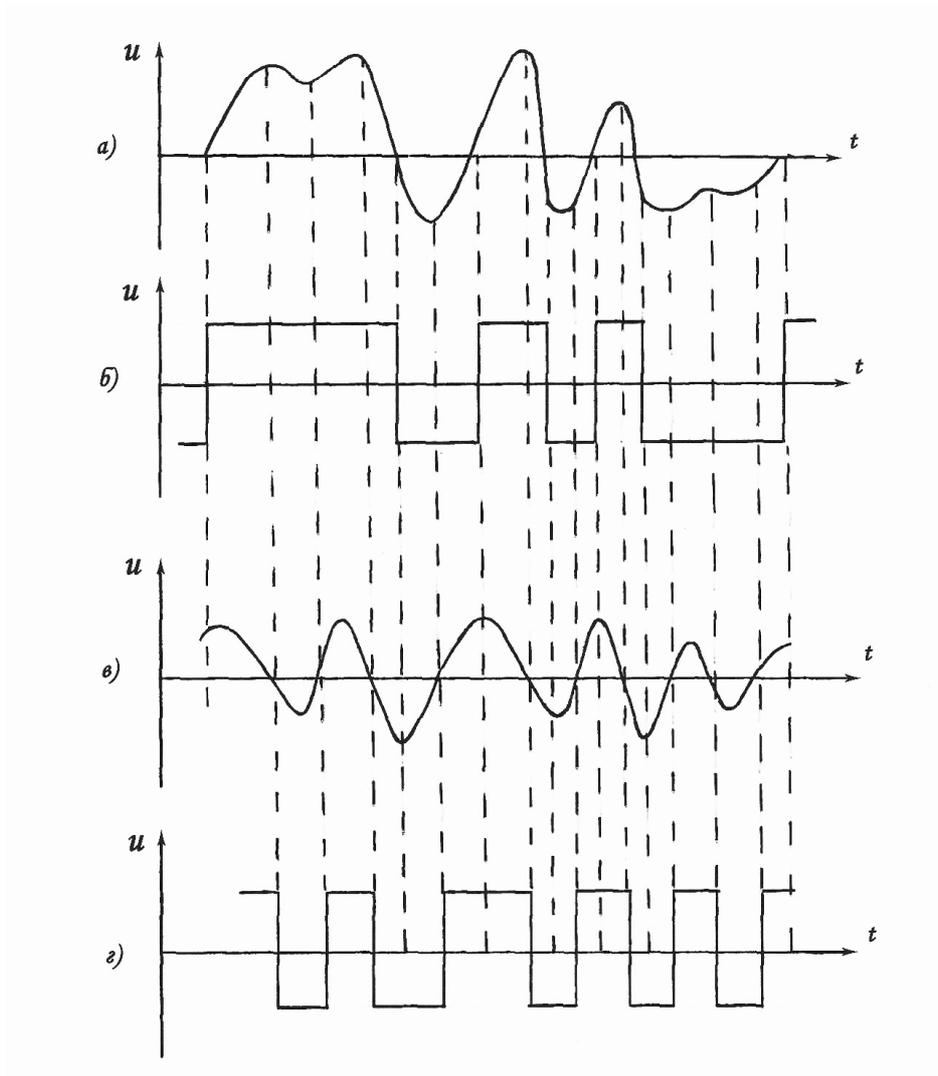


Рисунок 1.10 - Зависимость разборчивости звуков от степени ограничения

Устранение указанных искажений достигается снижением степени ограничения на 6...8 дБ, доводя её до 12...16 дБ, за счет непроизвольного увеличения уровня речи оператором в условиях сильных помех на 6...8 дБ фактическая степень ограничения опять будет равна 18...24 дБ. На рисунке 1.11 показана зависимость разборчивости звуков от степени ограничения в шумах ($B_{ш} = 115...120$ Дб).

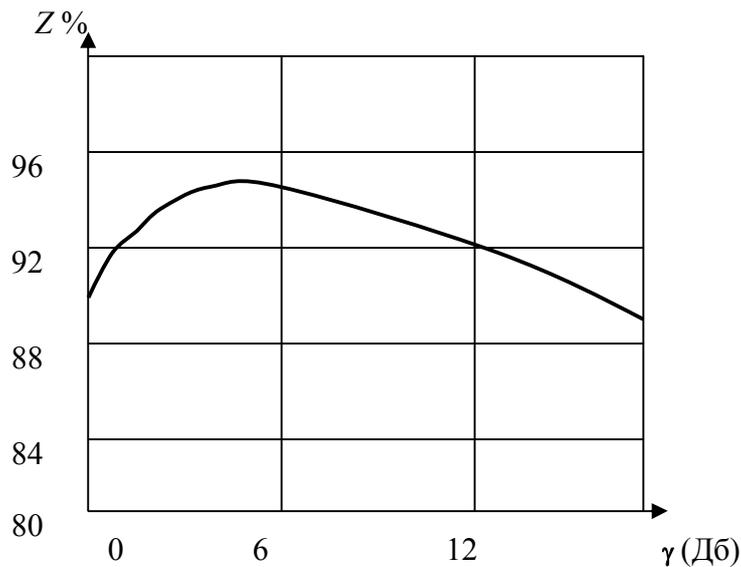


Рисунок 1.11 – Зависимость разборчивости звуков от степени ограничения в шумах

Следует отметить, что пока не накоплен достаточный опыт по применению ограниченной речи в системах связи, но опыты показывают, что предельное ограничение речи при заданной пиковой мощности ПРД и действии флуктуационных помех позволяет получить 5...10 - кратный выигрыш по мощности по сравнению с обычными системами. Ограничение речи может применяться в системах профессиональной связи, работающей в условиях интенсивных помех.

В структуре непосредственной компрессии речевого процесса рассмотрим **метод частотной компрессии**. Спектр речи занимает достаточно широкую полосу частот, примерно 100...9000 Гц. Простейший метод частотной компрессии заключается в ограничении спектра телефонных сигналов, передаваемых по линии связи. Для передачи выбирается участок спектра, являющийся наиболее важным для обеспечения требуемого качества связи. Установлено, что целесообразно срезать низкочастотные составляющие ниже 300 Гц и высокочастотные составляющие свыше 3...5 кГц, которые мало влияют на разборчивость речи. Сужение спектра понижает уровень помех. Большинство радиостанций служебной и коммерческой связей имеют полосу пропускания по НЧ $\Delta F_n = 300...3000$ [Гц] = 2,7 [кГц] или $300...3500$ [Гц] = 3,2 [кГц]. В некоторых устройствах диапазон сужается до $\Delta F_n = 400...2500$ [Гц] = 2,1 [кГц] или расширяется до $200...5000$ [Гц] = 4,8 [кГц].

Возможность компрессии спектра радиотелефонного сигнала зависит от вида модуляции. С точки зрения компрессии спектра целесообразно приме-

нять однополосную модуляцию, т.к. в этом случае ширина спектра радиосигнала сохраняется равной ширине спектра речевого сообщения. Для других видов модуляции радиосигнал имеет более широкий спектр, чем спектр речевого сообщения.

Компрессия спектра телефонного сигнала может быть получена за счет постановки избирательных систем с соответствующей полосой пропускания ΔF_n . Кроме того, за счет увеличения времени передачи. Если речь предварительно записать на пленку, а затем при передаче уменьшить скорость вращения магнитофона по сравнению с нормальной в n раз, то в такое число раз сузится спектр речи. На приемной стороне для восстановления нормальной речи скорость вращения должна быть увеличена в n раз, т.е. восстановлена.

Методы амплитудной и частотной компрессии имеют широкое практическое применение.

Важное значение имеет **метод временной компрессии**. *Временная компрессия телефонного сигнала заключается в сокращении времени передачи и основана на избыточности речи, определяемой её временными характеристиками.* Данная избыточность характеризуется наличием повторяющихся участков во временной функции. Для понимания звуков не обязательно передавать их в течение всего интервала времени, занимаемого или в речевом сообщении. Чтобы понять звук речи, требуется интервал времени 10 мс, тогда как длительность ряда звуков достигает 200...300 мс. Следовательно, имеется возможность сокращения времени передачи путем устранения временной избыточности.

Изложенное подтверждается экспериментально. Так как вырезание производилось при помощи прямоугольных импульсов, то разборчивость прерываемой речи будет зависеть от частоты следования импульсов при соответствующих коэффициентах заполнения d , который равен отношению действительного времени передачи прерываемого речевого сообщения, т.е.

$$d = \frac{1}{t_s} \left(t_s - \sum_{i=1}^n \tau_i \right) = \frac{1}{t_s} \sum_{i=1}^n \tau_s, \quad (1.7)$$

где τ_s, τ_i - соответственно длительность передаваемых участков речи и вырезающих импульсов, t_s - длительность непрерываемой речи.

При этом скорость передачи сообщения определяется соотношением:

$$C = \Delta f_c \log_2 \left(1 + \frac{P}{N} \right), \quad (1.8)$$

где Δf_c – ширина спектра;

P/N – отношение мощности сигнала к мощности помехи;

C – скорость передачи при разговоре¹.

Кривые разборчивости прерываемой речи (W) в зависимости от частоты следования вырезающих импульсов показаны на рисунках 1.11 и 1.12.

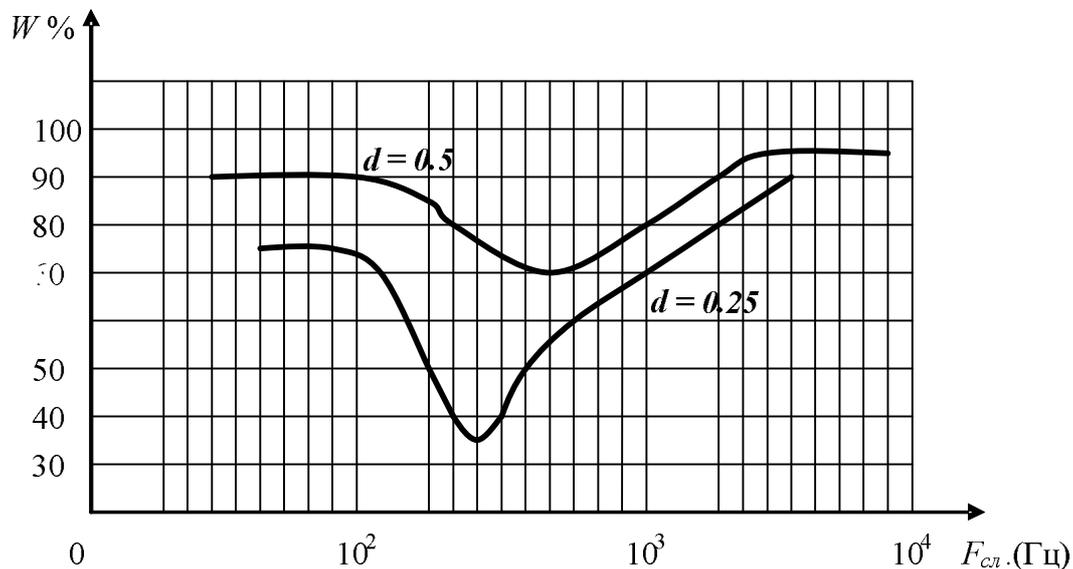


Рисунок 1.12 – Зависимость разборчивости речи от частоты следования вырезающих импульсов

Предположим, что отношение сигнал/шум в канале связи равно по напряжению 3, а по мощности $P/N=9$, что требуется для системы связи с амплитудной модуляцией. Тогда для передачи сигнала при таких условиях в соответствии с формулой Шеннона требуется полоса частот:

$$\Delta f_c = \frac{C}{\log_2 \left(1 + \frac{P}{N} \right)} = \frac{20}{\log_2 10} = \frac{20}{3,32} \approx 6 \text{ [Гц]}, \quad (1.9)$$

Это и есть предел возможного сжатия спектра речи. Оно может быть получено лишь при идеальном кодировании. Применяющиеся на практике системы телефонной радиосвязи с непосредственной компрессией занимают значительно большую полосу частот. Более эффективными являются методы компрессии спектра с функциональным преобразованием речи.

Из рисунка 1.11 видно, что при коэффициенте заполнения $d = 0,5$, т.е. при вырезании половины общей длительности речи, разборчивость сохраняется достаточно высокой. Только при $F_{след\ imp.} = 400...800$ Гц разборчивость существенно уменьшается. Особенно сильное уменьшение разборчивости на-

¹ В среднем равна 20 дв.ед./с.

блюдается при коэффициенте заполнения $d = 0,25$.

Сравнивая рисунки 1.12 и 1.12а можно заметить, что понижение разборчивости менее значительно для использования нерегулярных вырезающих импульсов.

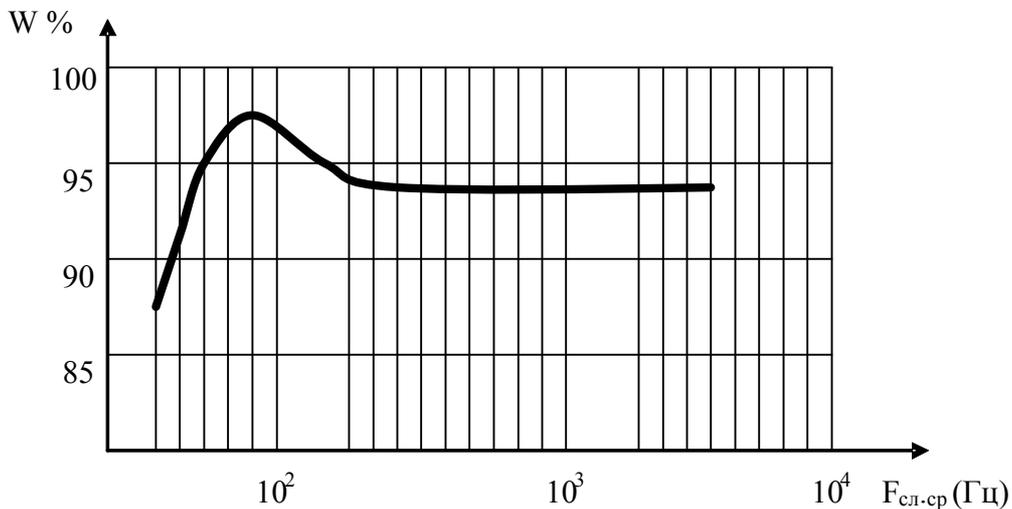


Рисунок 1.12а – Разборчивость речи в зависимости от средней частоты нерегулярных вырезающих импульсов

Преобразование речи может быть использовано для уплотнения канала связи несколькими передачами путем размещения других сообщений в вырезаемых интервалах времени.

1.5.3 Методы компрессии с функциональным преобразованием речи

Рассмотрим сущность методов компрессии с функциональным преобразованием речевого сигнала.

Методы компрессии телефонного сигнала с функциональным преобразованием речи основаны на замене речи её параметрами, сведения о которых передаются в линию связи для восстановления сообщения в приемном устройстве. Выбираются параметры, медленно меняющиеся во времени, поэтому для передачи сведений о таких параметрах требуется более узкая полоса, чем для передачи самой речи, тоже самое и для приема.

Устройства, с помощью которых производится функциональные преобразования речи, называются **вокодерами** (по-английски: войс - голос, кодер - кодировщик). Вокодер состоит из двух основных частей:

- *анализатора*, который производит выделение (т.е. анализ) параметров речи;
- *синтезатора*, который синтезирует, т.е. восстанавливает речь по ее параметрам.

Принципы построения вокодеров могут быть различными, в зависимости от вида используемых параметров речи.

Полосовой вокодер

Принцип действия полосового вокодера основан на использовании спектральных уровней в узких полосах частотного диапазона с учетом средних уровней речи в каналах связи (рисунок 1.13). Структурная схема полосового вокодера представлена на рисунке 1.14.

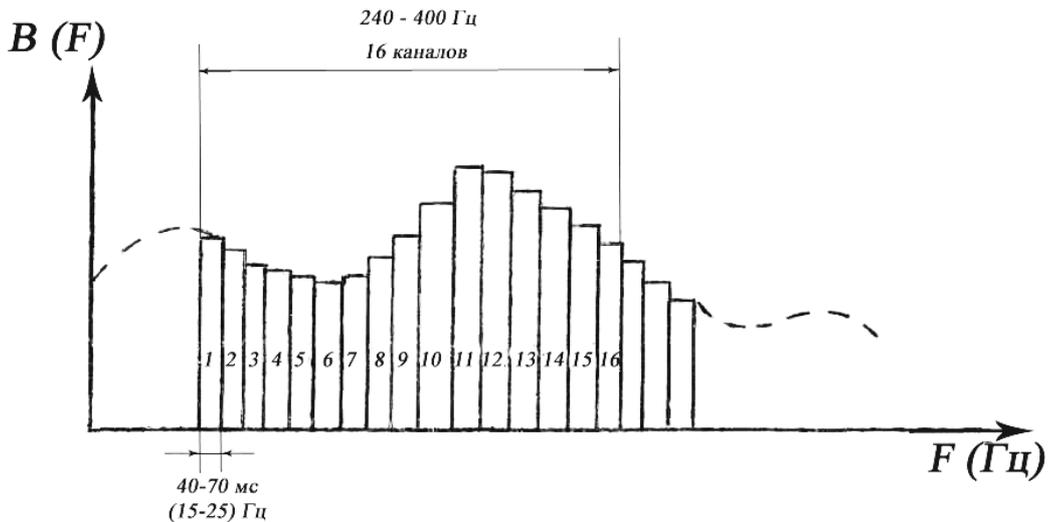


Рисунок 1.13 – Спектральные уровни в узких полосах частотного диапазона

Принцип действия анализатора заключается в том, что речевое сообщение разделяется фильтрами $\Phi_1 \dots \Phi_n$. Амплитудными детекторами $D_1 \dots D_n$ включенными на выходе фильтров, определяется средний уровень речи в соответствующей полосе за промежутки времени порядка 40...70 мс. Зависимости средних уровней от времени, т.е. спектральные огибающие уровней, представляющие напряжения спектральных каналов на выходе анализатора речи, передаются в линию связи на одной несущей частоте одним из методов уплотнения.

В синтезаторе приемного устройства процесс восстановления речи происходит в обратном порядке. С помощью генератора импульсов создается дискретный спектр, аналогичный спектру голосовых связок. Этот спектр с помощью ряда фильтров, аналогичных фильтрам в анализаторе, делится на узкие полосы частот. Напряжения с выходов фильтров подаются на амплитудные модуляторы M_1, M_2, \dots, M_n куда, кроме того, подаются принятые выходные напряжения спектральных каналов. В результате на выходе модуляторов получается изменяющиеся во времени уровни, аналогичные уровням на выходе соответствующих фильтров в анализаторе речи. Поэтому суммарный

процесс, полученный смешиванием всех напряжений на выходе модуляторов, представляет сигнал, примерно совпадающий с исходным, т.е. синтезированную речь.

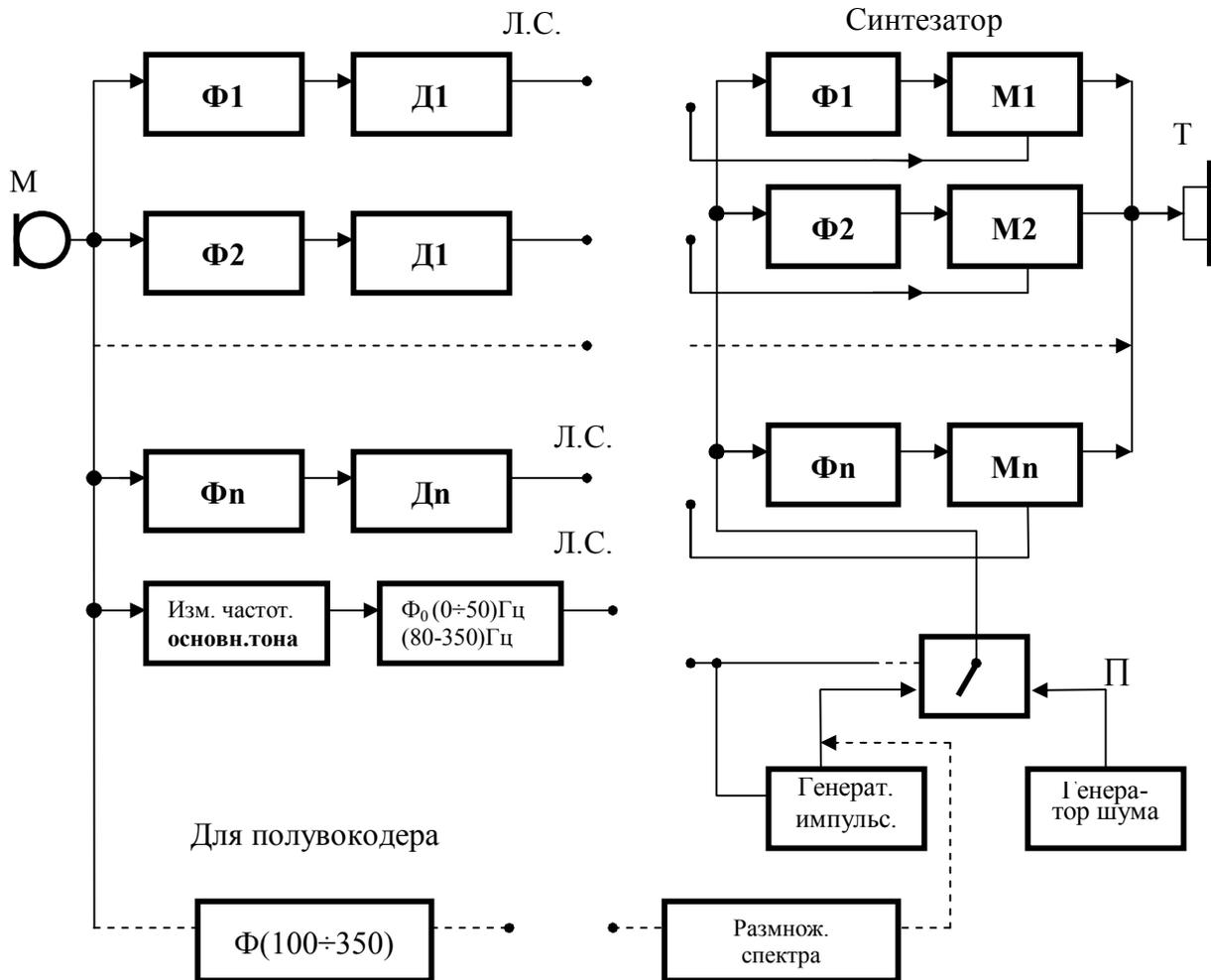


Рисунок 1.14 – Структурная схема полосного вокодера

Полосный вокодер обеспечивает достаточно высокую разборчивость, но не обеспечивает узнаваемости голоса. Для обеспечения узнаваемости голоса используется еще один параметр - частота основного тона, характеризующая высоту голоса. С этой целью в анализаторе измеряется частота основного тона. Полученная информация через фильтр Φ_0 передается в линию связи для управления частотой следования импульсов генератора. В этом случае синтезированная речь будет звучать более естественно.

В образовании глухих согласных звуков основной тон не участвует, поэтому при их произношении информация об основном тоне не передается, а на приемной стороне (синтезаторе) генератор шума создает шум, заменяющий шум глухих звуков. Выбор источника дискретного или непрерывного шума

определяется электронным переключателем Π в синтезаторе в зависимости от наличия или отсутствия сигнала основного тона.

Полосные вокодеры позволяют получить примерно 8...10 - кратное сжатие полосы. Эта возможность достигается за счет того, что выходные напряжения спектральных каналов, описывающие изменения средних уровней в каждой узкой полосе, являются медленно изменяющимися функциями времени и могут быть переданы по каналу связи в узкой полосе. Практически установлено, что полоса 25 Гц или даже 15 Гц для каждого спектрального канала обеспечивает достаточно высокую артикуляцию. Для передачи выходных напряжений шестнадцати (16) каналов требуется полоса 240...400 Гц. Для передачи основного тона требуется полоса около 25...50 Гц. Если еще учесть запас для устранения взаимных помех между каналами, то общая полоса частот должна быть увеличена примерно до 350...600 Гц. Применяется, где не нужна узнаваемость голоса, а основное значение имеет разборчивость речи, например, в устройствах служебной связи.

Гармонический вокодер

Принцип построения гармонического вокодера основан на разложении в ряд Фурье мгновенного спектра речевого сообщения с последующим синтезом речи по найденным коэффициентам разложения.

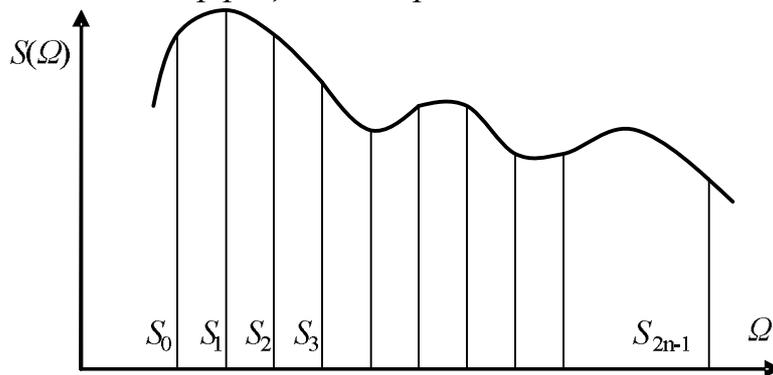


Рисунок 1.15 – Мгновенный спектр речевого сообщения

Мгновенный спектр речевого сообщения $S(\Omega)$ (рисунок 1.15), заданный в интервале частот $[0, \Omega_{\text{макс}}]$ может быть разложен в ряд Фурье, определяемый выражением (1.10).

$$S(\Omega) = \frac{d_0}{2} + \sum_{k=1}^{\infty} \left(d_k \cdot \cos \frac{1PK\Omega}{\Omega_{\text{макс}}} + b_k \cdot \sin \frac{2PK\Omega}{\Omega_{\text{макс}}} \right), \quad (1.10)$$

причем

$$d_k = \frac{2}{\Omega_{\text{макс}}} \int_0^{\Omega_{\text{макс}}} S(\Omega) \cdot \cos \frac{2PK\Omega}{\Omega_{\text{макс}}} \cdot d\Omega, \quad (1.11)$$

$$b_k = \frac{2}{\Omega_{\max}} \int_0^{\Omega_{\max}} S(\Omega) \cdot \sin \frac{2\pi K \Omega}{\Omega_{\max}} \cdot d\Omega. \quad (1.12)$$

Разбивая интервал $[0, \Omega_{\max}]$ на $2 \times n$, узких полос, приближенные значения коэффициентов можно записать в виде:

$$d_k = \frac{1}{n} \sum_{i=0}^{2n-1} S_i \cdot \cos \frac{\pi K_i}{n}; \quad b_k = \frac{1}{n} \sum_{i=0}^{2n-1} S_i \cdot \sin \frac{\pi K_i}{n}; \quad \frac{d_0}{2} = \frac{1}{2n} \sum_{i=0}^{2n-1} S_i, \quad (1.13)$$

где S_i - значения спектра $S(\Omega)$ в точках отсчета.

Для определения коэффициентов Фурье используются спектральные уровни на выходе фильтров полосного вокодера. Для образования сумм (Σ) используются диодно-реостатная матрица со специально подобранными делителями напряжений. Все это составляет анализатор (рисунок 1.17). Полученные в результате анализа напряжения, пропорциональные коэффициентам Фурье, передаются в линию связи одним из методов уплотнения.

Синтез речи заключается в воспроизведении на приемной стороне спектра передаваемого сообщения (рисунок 1.17) В остальном принцип действия гармонического вокодера аналогичен полосному вокодеру (рисунок 1.14). На вход синтезатора подается напряжение от генератора импульсов, создающего дискретный спектр, либо генератор шумов, создающий сплошной спектр. Выбор того или иного генератора производится с помощью различителя глухих и звонких звуков на передающей стороне и переключателя Π на приемной стороне. Различитель глухих и звонких звуков представляет собой два фильтра - фильтр нижних частот и фильтр верхних частот. Граничные частоты обоих фильтров равны 1000 Гц. Напряжения на выходе фильтров выпрямляются детекторами. Интенсивные колебания основной частоты (звонкие звуки) лежат в пределах ФНЧ, а глухие колебания в области ФВЧ. Когда произносится звонкий звук, более сильным будет сигнал в ФНЧ и наоборот. Соответственно этому на выходе образуется постоянное напряжение той или другой полярности. Это напряжение передается в приемное устройство для управления переключателем Π , который подключает ко входу синтезатора либо импульсный генератор, либо генератор шумов.

Полоса гармонического вокодера с учетом полосы основного тона составляет примерно 200...300 Гц вплоть до 5-й гармоники.

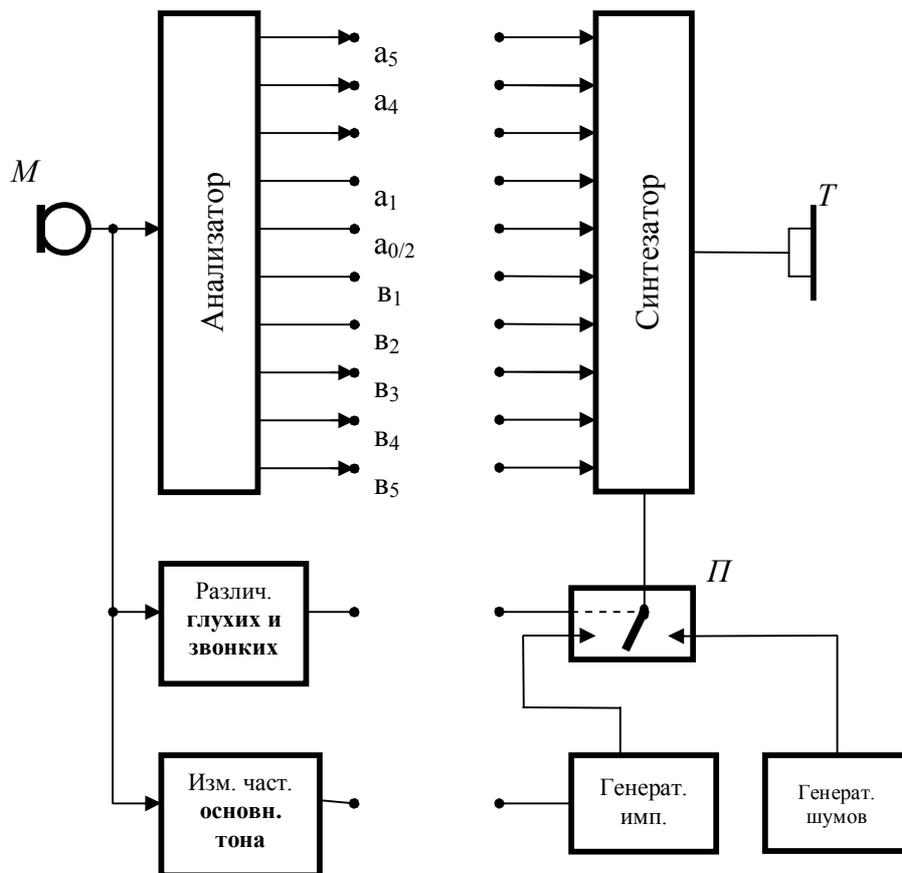


Рисунок 1.17 – Структурная схема гармонического вокодера

Формантный вокодер

В формантном вокодере в качестве параметров речи используются форманты, а именно, частота формант и их уровни. Структурная схема формантного вокодера изображена на рисунке 1.18.

Принцип действия формантного вокодера. В анализаторе формантного вокодера частотный диапазон речи с помощью фильтров $\Phi_0, \Phi_1, \Phi_2, \Phi_3$ делится на четыре части. Фильтр Φ_0 - для выделения диапазона основного тона 0...350 Гц; Φ_1 - для выделения полосы частот 300...1000 Гц - что примерно соответствует местоположению 1-й форманты; Φ_2 - для полосы частот 700...2400 Гц - 2-й форманты и Φ_3 - для полосы частот 2400...3500 Гц - 3-й форманты.

После выделения спектра речи в каждом из диапазонов производится определение частоты и уровня. Частоты $\Phi_0, \Phi_1, \Phi_2, \Phi_3$ определяются с помощью ЧД $Ч_1, Ч_2, Ч_3, Ч_{ш}$, а их уровни с помощью амплитудных $Д_1, Д_2, Д_3$ и $Д_{ш}$.

Информация о найденных параметрах передается в линию связи и на

приемное устройство. В синтезаторе формантного вокодера имеются генератор импульсов, управляемый основным тоном и генератор шума. Создаваемые ими колебания подаются на входы фильтров K_p , K_2 , K_3 и $K_{ш}$ соответственно. В отличие от синтезатора полосного вокодера эти фильтры настраиваемые и их число равно числу формант. Их настройка осуществляется применением резонансной частоты под действием управляющих сигналов, соответствующим частотам F_1 , F_2 , F_3 и $F_{ш}$ в пределах диапазона соответствующей форманты. Напряжения с выхода фильтров подаются на модуляторы M_1 , M_2 , M_3 и $M_{ш}$ вместе с управляющими сигналами, соответствующими уровням I_1 , I_2 , I_3 и $I_{ш}$. В результате суммирования процессов на выходе всех модуляторов будет синтезироваться речь.

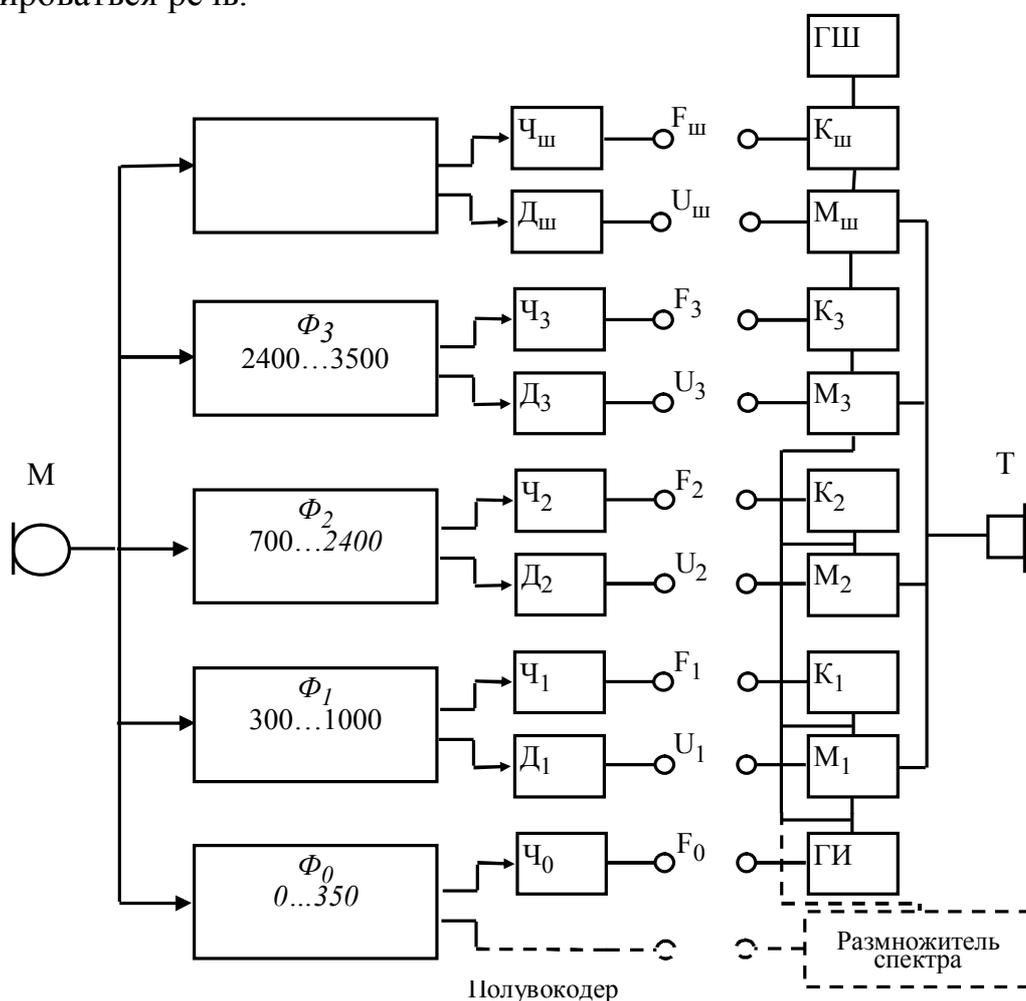


Рисунок 1.18 – Структурная схема формантного вокодера с отдельным управлением уровнями формант звонких звуков

Полувокодер - для определения частоты основного тона, в котором пе-

редается основной тон без преобразования его с использованием для этого всей его полосы частот.

На приемной стороне полоса частот непреобразованной речи с помощью нелинейных элементов "размножается", т.е. спектр расширяется до 3000 ... 4000 Гц и используется вместо генератора импульсов.

Если считать, что для передачи каждого параметра требуется полоса 15...25 Гц, то тогда общая полоса частот будет примерно равна 100...175 Гц, что примерно в 3...5 раз меньше полосы полосного вокодера, что делает их перспективными.

Фонемный вокодер

В фонемном вокодере (рисунок 1.19) в качестве параметров используются фонемы.

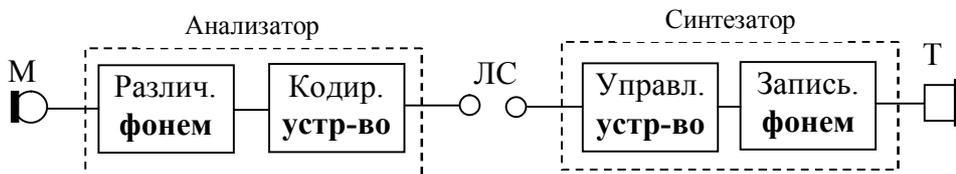


Рисунок 1.19 – Структурная схема фонемного вокодера

Принцип действия фонемного вокодера сводится к следующему. В анализаторе производится определение фонем произносимой речи. В линию связи каждая фонема передается в виде: некоторой кодовой комбинации, подобно кодовым комбинациям букв в системах телеграфии. В синтезаторе имеется запись всех фонем речи. Принимаемые кодовые комбинации включают на телефон соответствующие им фонемы, благодаря чему восстанавливается речь.

Особенностью фонемного вокодера является то, что восстановленный в синтезаторе голос не передает индивидуальных особенностей говорящего и при данном источнике звука будет всегда одним и тем же, поэтому такую систему называют звуковым телеграфом.

Возможность сужения полосы частот до 20...30 Гц делает возможным применение фонемных вокодеров для целей связи, но из-за трудностей реализации они не нашли пока практического применения.

1.6 Виды помех в каналах радиосвязи

Электромагнитные помехи (ЭМП) представляют собой случайные процессы, которые действуют в свободном пространстве или проводящей среде и могут ухудшать качество приема сигналов. В теории ЭМС рассматриваются только непреднамеренные ЭМП, не предназначенные для целена-

правленного нарушения работы радиосредств. Действие ЭМП проявляется в виде искажений сигналов, повреждения элементов радиоаппаратуры (полупроводниковых приборов, ИМС и др.), нарушения функционирования отдельных узлов.

По типу источников различают:

- *естественные помехи* - вызываются различными физическими явлениями в окружающем нас мире;
- *искусственные ЭМП* порождаются электромагнитными процессами в технических устройствах и подразделяются на стационарные, промышленные и контактные.

По виду среды распространения различают:

- *радиационные помехи* - атмосферные, космические, стационарные, большинство промышленных и контактные помехи;
- *кондуктивные помехи* - распространяются от источника к приемнику по физическим цепям питания, заземления и др.

По частотно-временным свойствам ЭМП могут быть:

- *сосредоточенными по спектру (узкополосные),*
- *сосредоточенными во времени (импульсные),*
- *флуктуационными.*

Понятия узкополосности и сосредоточенности помех во времени относятся к величинам Π и $1/\Pi$, где Π — ширина полосы пропускания радиоприемника.

По характеру взаимодействия с полезным сигналом различают:

- *аддитивные помехи* - суммируются с сигналами,
- *мультипликативные помехи* — выступают в роли случайного «коэффициента передачи канала» - коэффициента передачи канала.

При вероятностной классификации ЭМП используются признаки гауссовости и стационарности. *Большинство помех представляет собой нестационарные случайные процессы, что затрудняет анализ их воздействия на РПрУ.* Однако часто при действии помех на коротких по сравнению с сигналом временных интервалах нестационарность их проявляется слабо, и тогда говорят о локально-стационарных помехах.

Относительно самого рецептора помеха может быть:

- *собственной* - создаются элементами самой радиоаппаратуры; источники внешних помех лежат вне ее;
- *внутрисистемной* – помеха формируются источниками, входящими в данную систему связи;
- *внешней и межсистемной* - помехи создаются источниками, внешними для данной системы.

К числу аддитивных ЭМП *естественного происхождения* относят кос-

мические шумы, атмосферные помехи, шумы теплового излучения Земли, собственные шумы радиоаппаратуры.

Космические шумы обусловлены термоядерными процессами в радиозвездах, движением заряженных элементарных частиц в космическом пространстве, поглощением и переизлучением атмосферой и поверхностью Земли излучения Солнца и других космических источников. По характеру спектра их относят к широкополосным флуктуационным помехам.

Атмосферные помехи - следствие грозовых и электростатических разрядов, возникают на частотах до 25 МГц и по уровню превышают космические шумы. Для одиночных близких грозовых разрядов в спектре помех преобладает импульсная компонента, а суперпозиция многих дальних разрядов формирует флуктуационную компоненту. Электростатические помехи создаются токами стека-ния с участков поверхности быстро движущихся объектов, электризуемых потоком водных и пылевых частиц, а также в результате пробоя воздушных промежутков между металлизированными элементами конструкции объекта. Они могут быть заметны на частотах до 1 ГГц. Атмосферные помехи бывают узко- и широкополосными. Для описания узкополосных помех обычно используется **полимодалный** или **модуляционный подход**.

При использовании **полимодалного подхода** помеха представляется в виде суммы гауссовского шума с дисперсией σ_{n1}^2 и импульсной случайной последовательности с частотой повторения импульсов F_u ; сами импульсы имеют постоянную длительность T_u и гауссовское шумовое заполнение с дисперсией σ_{n2}^2 причем полоса тракта $\Pi = 1/T_u$ и $\sigma_{n2}^2 \approx (1,5...2)\sigma_{n1}^2$. Законы распределения мгновенных $X(t)$ и амплитудных $A(t)$ значений суммы обеих компонент негауссовские. Такая модель отражает физический механизм формирования помех и удовлетворительно воспроизводит их вероятностные характеристики в диапазоне частот до 100 кГц.

При **модуляционном подходе** помеха представляется в виде узкополосного гауссовского процесса $y(t)$ с нулевым средним значением, модулированного низкочастотным стационарным случайным процессом $a(t)$, верхняя частота спектра которого существенно ниже центральной частоты спектра $y(t)$, т.е. результирующий процесс имеет вид $x(t) = a(t) y(t)$; при этом законы распределения $W(X)$ и $W(A)$ негауссовские. Задаваясь законами распределения $a(t)$, можно добиться удовлетворительного описания помех в диапазоне до 10 МГц.

Для широкополосных атмосферных помех характерна более четкая выраженность флуктуационной и импульсной компонент. При этом помеха представляется в виде суммы гауссовского процесса с нулевым средним значением и дисперсией σ_n^2 и одного или нескольких случайных потоков импуль-

сов, для которых вероятность появления N импульсов за время T подчиняется закону Пуассона:

$$p = \frac{(F_u T)^N}{N!} \cdot e^{-F_u T}$$

Импульсы помех имеют тенденцию к группированию в пакеты с пуассоновской потоковой статистикой: пуассоновская статистика импульсов внутри пакетов с частотой F_p имп/с, и самих пакетов с частотой F_i , пак/с. Амплитуды импульсов таких помех распределены по логарифмически нормальному закону (ЛНЗ), одна из форм представления которого имеет вид

$$W(A_n) = \frac{1}{\sqrt{2\pi\sigma_n}} \cdot \exp\left[-\frac{(A_n - \mu_n)^2}{2\sigma_n^2}\right].$$

Здесь параметры A_i, σ_i, μ_i , дБмкВ, нормированы к 1 мкВ.

Станционные ЭМП создаются излучениями РПДУ, гетеродинов радиоприемников, генераторами строчной развертки телевизоров и др. Наиболее интенсивные из них — излучения РПДУ, которые нередко являются основным фактором нарушения связи. Такие помехи, как правило, представляют собой узкополосные случайные процессы с ЛНЗ распределения амплитуд; при этом на частотах около 5 МГц в полосе 1 кГц значения $\sigma_n \leq 13$ дБмкВ, $\mu_n \leq 30 \dots 40$ дБмкВ, а на частотах около 10 МГц в той же полосе $\sigma_n \leq 6 \dots 8$ дБмкВ, $\mu_n \leq 15 \dots 20$ дБмкВ.

Для описания станционных ЭМП используются пуассоновская модель и модели, основанные на теории случайных импульсных потоков. Несмотря на некоторую некорректность пуассоновской модели для задач присвоения частот, она часто используется для оценки вероятности попадания N_f помех в заданную полосу частот. Однако из-за нестационарности потока помех вдоль оси частот более точной оказывается аппроксимация распределения интервала между несущими частотами помех законом Эрланга:

$$W(\Delta f) = \beta_n^{N_n-1} \cdot \exp\left(\frac{-\Delta f \beta_n}{(N_n - 1)!}\right),$$

где N_n, β_n - параметры, зависящие от класса излучения помех. Это выражение позволяет строить гистограммы интенсивности помех, полезные при выборе присвоенных частот.

Индустриальные радиопомехи (ИРП) создаются техническими средствами, использующими в той или иной форме энергию электрического тока. Данные помехи могут иметь дискретный и сплошной спектр, быть кондуктивными и радиационными. К первым относятся помехи, создаваемые медицинскими установками ВЧ, ЭВМ, устройствами развертки электронно-лучевых

трубок, ко вторым — помехи от электротранспорта, промышленной и бытовой электроаппаратуры, линий электропередачи (ЛЭП), систем зажигания автомобилей. Индустриальные радиопомехи наблюдаются в широкой полосе частот — от единиц килогерц до единиц гигагерц.

В ЛЭП имеют место длительные (более 0,5...1 с) флуктуации напряжения, превышающие нормированные отклонения $\pm(10...15)\%$, причем провалы наблюдаются значительно чаще, чем перенапряжения. При этом возникают нестационарные потоки одиночных импульсов или пакетов импульсов с амплитудой до 1,5 кВ длительностью от десятков наносекунд до сотен микросекунд, распределенных по экспоненциальному закону или ЛНЗ. На частотах до 10 кГц могут возникать интенсивные ИРП в виде комбинационных составляющих тока промышленной частоты 50 Гц, создаваемые генераторным и трансформаторным оборудованием.

Уровень таких ИРП в течение суток может изменяться на 20...50%. На более высоких частотах (до 30 кГц) помехи приближаются по характеру к флуктуационным; с ростом частоты до 100 кГц и более появляются одиночные импульсы и пакеты импульсов значительной амплитуды. Такие помехи оказывают заметное влияние на цифровые элементы и вычислительные средства радиоаппаратуры при ее близком расположении к источнику.

Разнородность источников ИРП обусловила многообразие их вероятностных законов. Так, для ИРП со сплошным спектром применима модель ЛНЗ распределения амплитуд и пуассоновская или пуассон-пуассоновская потоковая статистика; при этом значения параметров следующие:

$$\begin{aligned} \sigma_n &\approx 7...10 \text{ дБмкВ}, & \mu_n &\approx 19...27 \text{ дБмкВ}, \\ F_u &= 780...9100 \text{ имп/с}, & F_n &= 0,5...100 \text{ пак/с} \end{aligned}$$

Универсальные измерители ИРП позволяют определять действующее $U_{\text{действ}}$, пиковое $U_{\text{пик}}$ и квазипиковое $U_{\text{кп}}$ значения напряжения. Вызвано это тем, что амплитудные и потоковые параметры ИРП по-разному влияют на восприимчивость слуховых, цифровых и других рецепторов. Для флуктуационных и сосредоточенных по спектру ИРП целесообразно измерять $U_{\text{действ}}$. Пиковый вольтметр используют для измерения одиночных помех. Квазипиковое напряжение удовлетворительно характеризует действие ИРП при слуховом приеме сигналов, в то же время оно не адекватно учитывает сбои ИМС, вызываемые широкополосными ИРП.

Для вероятностной характеристики ИРП используется коэффициент импульсности $k_u = U_{\text{действ}} / U_{\text{нм}}$, где $U_{\text{нм}}$ — амплитудное значение помехи. При ЛНЗ распределения амплитуд $k_u \approx 0,7...7,3$; при рэлеевском законе $k_u \approx 0,53$.

Радиационные ИРП оцениваются напряженностью поля на заданном расстоянии R от источника и выражается в дБмкВ/м. Такие помехи наиболее

опасны в ближней зоне излучения на расстояниях $R < \lambda / 2\pi$ и затухают с коэффициентом $\frac{1}{R^{2,2}}$. Помехи от автотранспорта ощутимы на расстояниях до 15 м и достигают наибольшего значения на частотах около 30 МГц; на частотах более 1 ГГц поле помех ослабевает со скоростью около 20 дБ/окт. Уровень ИРП от ЛЭП составляет 40...160 дБмкВ/м и зависит от напряжения сети и насыщенности воздуха частицами воды и пыли.

Кондуктивные ИРП затухают медленно и могут распространяться на значительные расстояния. Для оценки влияния таких помех служит коэффициент переноса помех

$$k_{\text{ИП}} = \frac{U_{\text{ПИ}}}{U_{\text{ПР}}},$$

где $U_{\text{ПИ}}, U_{\text{ПР}}$ — напряжение помех на выходных зажимах источника и входных зажимах рецептора. Уровень таких помех нормируется в полосе частот 0,15...30 МГц.

Связные и телевизионные РПрУ содержат функциональные элементы, создающие ИРП — генераторы строчной развертки, усилители синхроимпульсов и сигналов изображения, гетеродины и др. Создаваемые ими помехи могут быть как радиационными, так и кондуктивными (через сеть питания) и нормируются. Так, допустимая напряженность поля радиационных помех телевизионных приемников в полосе частот 30...1000 МГц составляет 54...66 дБмкВ/м; для вещательных радиоприемников АМ сигналов в диапазоне 0,15...30 МГц напряжение помех на сетевых клеммах 46...74 дБмкВ.

Контактные помехи возникают, как правило, на движущихся объектах при действии электромагнитного поля источника на находящиеся в ближней зоне излучения токопроводящие с переменным сопротивлением механические контакты конструкций объектов. Индуцированные в них токи проводимости приводят к возникновению вторичного поля помех, отличающегося по спектральному составу от первичного поля источника. Действие таких помех наблюдается на частотах до 100 МГц. Спектр контактных помех всегда шире спектра источника, причем его ширина и уровень составляющих возрастают с повышением частоты и мощности первичного сигнала и усилением флуктуации контактного сопротивления. При попадании такого контакта в ближнюю зону излучения нескольких РПДУ возможно образование контактных помех комбинационного типа, которые могут оказывать воздействие на РПрУ при его значительной расстройке Δf относительно несущей частоты радиопередатчика. При этом амплитуда помех удовлетворительно аппроксимируется выражением $U_{\text{Пм}} = (3...15) / \sqrt[3]{\Delta f^4}$, где Δf — расстройка, МГц. Контактные помехи

на интервалах 20...30 мин можно считать стационарными случайными процессами, содержащими гауссовскую флуктуационную и импульсную составляющие с ЛНЗ распределения амплитуд. Флуктуационная компонента преобладает на самолетах, а импульсная - на объектах железнодорожного и автотранспорта.

В реальном канале связи передаваемый сигнал искажается, в результате чего сообщение воспроизводится с ошибкой. Причинами таких искажений являются:

- искажения, вносимые самим каналом,
- помехи, воздействующие на сигнал.

В общем виде влияние помехи на полезный сигнал $S(t)$ выражается оператором:

$$\xi(t) = L[S(t), n(t)], \tag{1.14}$$

Если оператор $L[S(t), n(t)]$ выражается суммой

$$\xi(t) = S(t) + n(t), \tag{1.15}$$

то помеха называется **аддитивной**.

Если оператор $L[S(t), n(t)]$ представляется произведением

$$\xi(t) = k(t) S(t), \tag{1.16}$$

где $k(t)$ – случайный процесс, то помеха называется **мультипликативной**.

В реальных каналах связи имеют место как аддитивные, так и мультипликативные помехи

$$\xi(t) = k(t) S(t) + n(t), \tag{1.17}$$

В целом классификацию помех можно представить в виде рисунка 1.201..

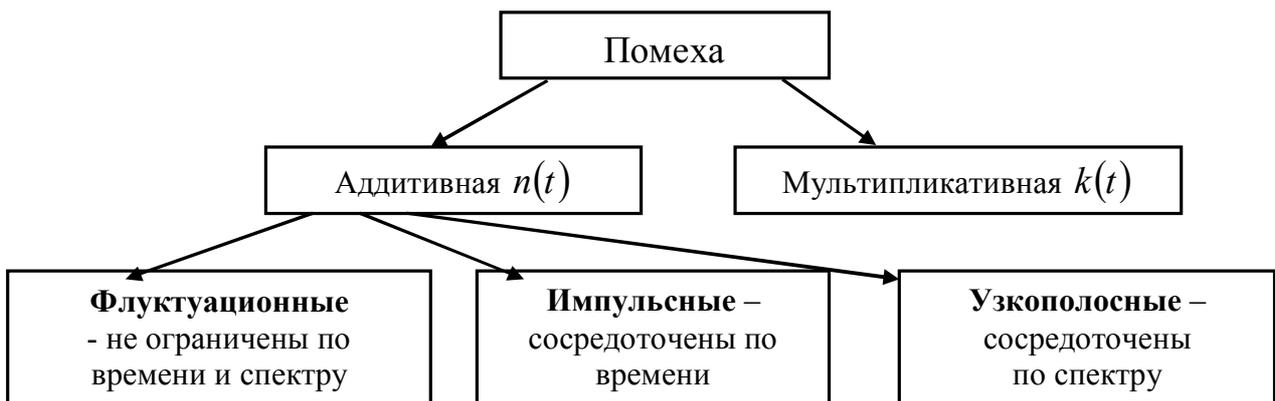


Рисунок 1.20 – Классификация помех

Импульсные помехи являются основным видом помех в проводных каналах связи. Они связаны с автоматической коммутацией и перекрестными на-

водками.

Узкополосные помехи – помехи от соседних радиостанций и различных радиотехнических устройств, а также созданные искусственно прицельные помехи, характеризующиеся тем, что $\Delta f_{\Pi} \approx \Delta f_{\text{ПРМ}}$.

Мультипликативные помехи обусловлены случайными изменениями параметров канала связи. В частности, эти помехи проявляются в изменении уровня сигнала (приход в точку приема нескольких реализаций сигнала, сдвинутых во времени).

1.7 Виды сообщений и сигналов

По своему характеру сообщения могут быть дискретнозначными или дискретными и непрерывнозначными или непрерывными (рисунок 1.21).

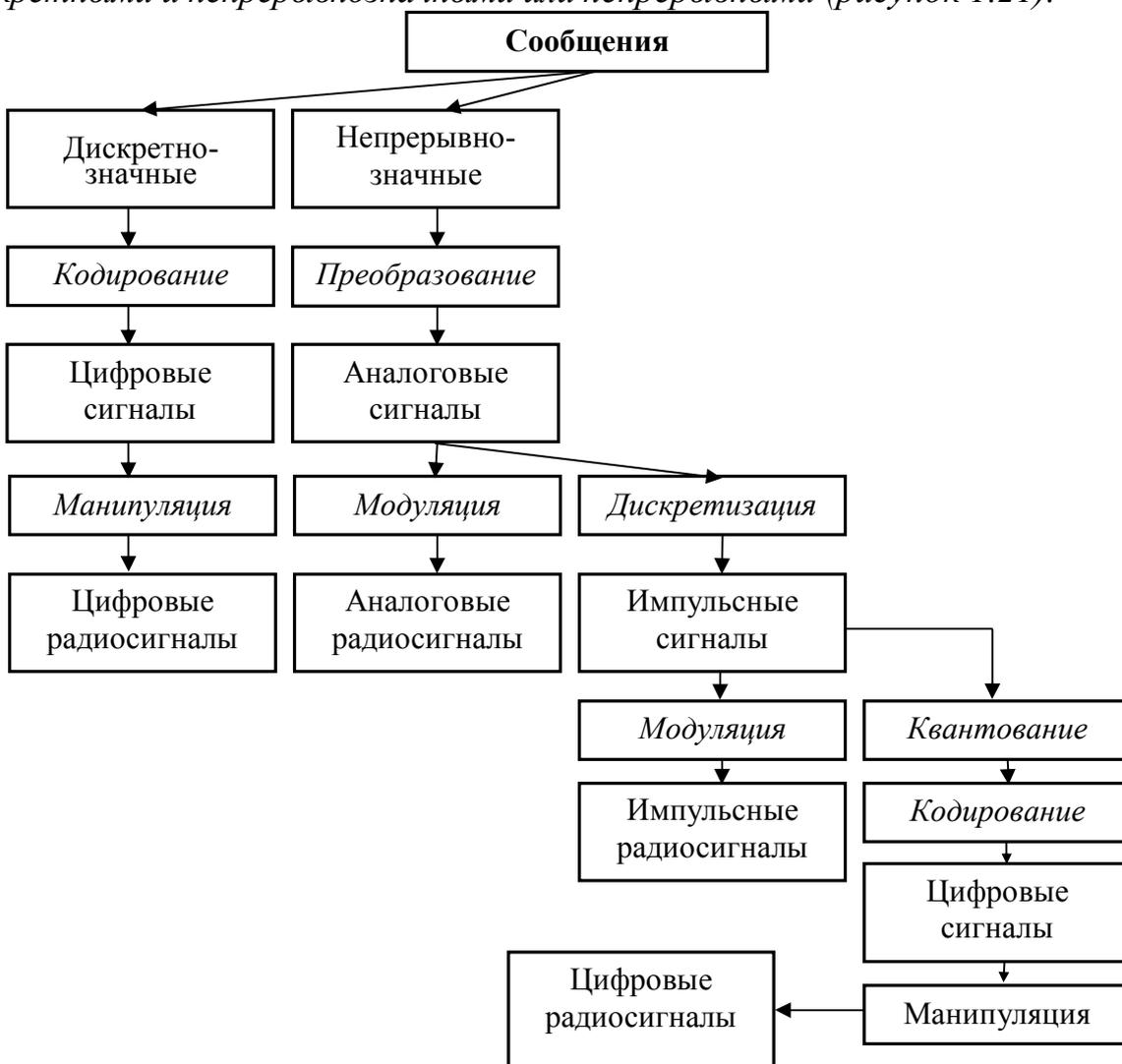


Рисунок 1.21 - Классификация сообщений и сигналов

Дискретнозначными называются сообщения, принимающие конечное или счетное число значений. Множество возможных сообщений с их вероятностными характеристиками образуют **ансамбль сообщений**. Выбор конкретных сообщений из ансамбля осуществляет отправитель сообщений. Типичным примером дискретнозначных сообщений может служить буквенно-цифровой текст, состоящий из букв, цифр и знаков препинания.

Непрерывнозначными называются сообщения, возможные значения которых неотделимы и непрерывно заполняют некоторую область значений. Они характеризуются плотностями вероятности.

К непрерывнозначным сообщениям относятся речь, музыка, подвижное изображение и т. д.

Для передачи по каналу связи любой вид сообщения должен быть преобразован в первичный электрический сигнал. Между сообщением и сигналом должно быть однозначное соответствие, чтобы при обратном преобразовании в пункте приема можно было получить переданное сообщение. Например, звуковое давление при передаче речевых сообщений преобразуется микрофоном в электрическое напряжение. Электрические сигналы, являющиеся аналогами непрерывнозначных сообщений, называются аналоговыми.

Первичные электрические сигналы, соответствующие дискретнозначным сообщениям, называют **цифровыми**. Процесс преобразования дискретнозначных сообщений в цифровые сигналы называется **кодированием**. При кодировании каждому возможному сообщению из ансамбля ставится в однозначное соответствие кодовая комбинация единичных элементов цифрового сигнала.

Систему соответствия между дискретнозначными сообщениями и кодовыми комбинациями единичных элементов принято называть **первичным кодом**. Код обычно задается в виде таблицы, в которой приведены возможные сообщения и соответствующие им кодовые комбинации единичных элементов. В качестве единичных элементов кодовых комбинаций обычно используются электрические импульсы, которые характеризуются вполне определенными значениями амплитуды - представляющего (информационного) параметра цифрового сигнала. В системах передачи цифровых сообщений обычно используются двоичные коды, при которых возможные значения амплитуды единичных импульсов принято условно отождествлять с символами «1» и «0». Это позволяет широко использовать в аппаратуре связи стандартные элементы цифровой техники. Символы единичных элементов кодовых комбинаций «1» и «0» называют **битами**.

Аналоговые сигналы можно преобразовать в импульсные и цифровые сигналы. Преобразование аналогового сигнала в импульсный достигается его дискретизацией по времени. Возможность такого преобразования вытекает из теоремы отсчетов. При дискретизации аналоговый сигнал представляется со-

вокупностью значений отсчетов в дискретные моменты времени. При этом в соответствии со значениями аналогового сигнала в дискретные моменты времени может меняться один из представляющих параметров импульсной последовательности, а именно, амплитуда, длительность или фаза (положение импульсов на оси времени).

Преобразование аналогового сигнала в цифровой достигается его дискретизацией по времени и квантованием по уровню. При квантовании отсчетов аналогового сигнала диапазон возможных значений делится на конечное число областей и каждая из этих областей представляется фиксированным значением представляющего параметра импульсного сигнала. Установленные уровни квантованных отсчетов могут быть преобразованы в кодовые комбинации цифрового сигнала. Указанная процедура преобразования называется импульсно-кодовой модуляцией. Таким образом, при импульсно-кодовой модуляции аналоговый сигнал путем дискретизации, квантования отсчетов этого сигнала и их кодирования преобразуется в цифровой сигнал. В дальнейшем будем предполагать, что сообщение от отправителя поступает в канал связи в виде аналогового, цифрового или импульсного первичного сигнала.

Поскольку передача первичного электрического сигнала на большие расстояния невозможна, то он в передающем устройстве при помощи модуляции или манипуляции преобразуется в радиосигнал. Этот радиосигнал передается через свободное пространство - линию связи к приемному устройству.

Модуляцией называют процесс изменения одного или нескольких параметров радиочастотного колебания в соответствии с представляющим параметром первичного электрического сигнала. Изменяемые при этом параметры также называют **информационными**, а все остальные - **сопутствующими**.

Модуляцию радиочастотного колебания первичным **цифровым** сигналом называют **манипуляцией**.

Модуляцию радиочастотного колебания первичным **импульсным** сигналом (последовательностью импульсов) называют **импульсной модуляцией**.

1.8.1 Математическая модель непрерывного сообщения

При решении задач анализа и синтеза систем передачи информации широко используются математические модели сообщений. В таких моделях сохраняются те свойства процессов, которые являются существенными для решаемой задачи. Модель сообщения должна быть представительной. Это значит, что по результатам анализа качества передачи модели сообщения можно делать выводы о качестве передачи реального сообщения. *В современной теории передачи информации принят вероятностный подход, при котором отдельные сообщения рассматриваются как реализации случайного процесса.*

Математической моделью непрерывного сообщения является непре-

рывный случайный процесс $X(t)$. Наиболее полное описание такого процесса дается n -мерной функцией распределения

$$F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{x(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$$

или n -мерной плотностью распределения вероятности

$$\omega_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{\partial^n F(x_1, \dots, x_n; t_1, \dots, t_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

при $n \rightarrow \infty$.

Указанные многомерные функции определить сложно, а зачастую и невозможно. В то же время для решения многих практических задач, связанных с передачей сообщений, не требуется знания многомерных законов распределения. Поэтому в качестве моделей сообщений обычно используются случайные процессы, задаваемые одномерным и двумерным законами распределения, а во многих случаях - более простыми характеристиками - моментными функциями.

Реальные сообщения, как правило, являются нестационарными. Соответственно их моделями должны служить нестационарные случайные процессы. Чаще всего нестационарные модели допускают квазистационарную трактовку: их можно считать практически стационарными на промежутках времени небольшой длительности. Переход к стационарной модели обусловлен тем, что решение задач с учетом нестационарности сообщений весьма затруднительно и требует сложного математического аппарата.

На практике в качестве стационарных моделей сообщений и помех часто используют *гауссовский случайный процесс*. Гауссовская модель достаточно хорошо описывает речевые и телевизионные сообщения, а также некоторые типы телеметрируемых процессов.

Среди моментных функций наибольшее применение получили:

- *математическое ожидание* случайного процесса

$$m_x(t) = M\{X(t)\} = \int_{-\infty}^{\infty} x \omega(x, t) dx \quad (1.18)$$

- *дисперсия* случайного процесса

$$D_x(t) = M\{[x(t) - m_x(t)]^2\} = \int_{-\infty}^{\infty} [x - m_x(t)]^2 \omega(x, t) dx, \quad (1.19)$$

- *корреляционная функция* случайного процесса

$$R_x(t_1, t_2) = M\{[X(t_1) - m_x(t_1)][X(t_2) - m_x(t_2)]\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - m_x(t_1)][x_2 - m_x(t_2)] \omega_2(x_1, x_2; t_1, t_2) dx_1 dx_2. \quad (1.20)$$

Для стационарных случайных процессов

$$\begin{aligned} m_x(t) &= m_x = \text{const}; \quad D_x(t) = D_x = \text{const}; \\ R_x(t_1, t_2) &= R_x(t_2 - t_1) = R_x(\tau). \end{aligned} \quad (1.21)$$

Иногда модель задается *спектральной плотностью мощности*, которая для стационарного центрированного процесса определяется как

$$G_x(\omega) = \int_{-\infty}^{\infty} R_x(\tau) \exp(-j\omega\tau) d\tau. \quad (1.22)$$

В качестве моделей сообщений, сигналов и помех часто используется эргодический случайный процесс. Для него все характеристики, найденные путем статистического усреднения, совпадают с характеристиками, найденными по его одной реализации путем усреднения по времени. Так, для эргодического процесса

$$\begin{cases} m_x = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t) dt \\ D_x = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [x(t) - m_x]^2 dt \\ R_x(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [x(t) - m_x][x(t - \tau) - m_x] dt \end{cases} \quad (1.23)$$

В ряде случаев достаточными для расчета систем связи характеристиками непрерывных сообщений являются полоса частот F_x , средняя мощность P_x , пик-фактор и динамический диапазон.

Пик-фактор K_n сообщения - это отношение его максимальной мгновенной мощности к средней:

$$K_{\Pi} = P_{\max} / P_x \quad (1.24)$$

Часто пик-фактор выражается в децибелах:

$$K_{\Pi} = 10 \lg [P_{\max} / P_x]. \quad (1.25)$$

Динамическим диапазоном называется отношение максимальной мгновенной мощности сообщения к минимальной мгновенной мощности, выраженное в децибелах:

$$D = 10 \lg [P_{\max} / P_{\min}]. \quad (1.26)$$

Например, для телефонного речевого сообщения верхняя частота спектра $F_B = 3400$ Гц, нижняя частота спектра $F_H = 300$ Гц, $F_x = F_B - F_H = 3100$ Гц, $K_n = 13 \dots 17$ дБ, $D = 35 \dots 40$ дБ.

При выборе модели необходимо учитывать содержание решаемой задачи, особенности математического аппарата, соответствующего данной модели,

и ряд других факторов. Как правило, рациональной окажется наиболее простая модель, позволившая с требуемой точностью решить поставленную задачу.

1.8 Оптимальный прием непрерывных сообщений и его помехоустойчивость

Гармоническое колебание имеет вид

$$S(t) = A_0 \cos(\omega_0 t + \Theta_0) = A_0 \cos \psi(t). \quad (1.27)$$

Мгновенная угловая частота связана дифференциальным соотношением с фазой

$$\omega(t) = \frac{d\psi(t)}{dt}. \quad (1.28)$$

Исходя из этого полную фазу высокочастотного колебания в момент t определяют как

$$\psi(t) = \int_0^t \omega(\tau) d\tau + \Theta_0, \quad (1.29)$$

где Θ_0 начальная фаза колебания.

Для узкополосного радиосигнала, когда можно выделить некую частоту ω_0 и случайную фазу $\Theta(t)$ за счёт модулирующего сигнала и являющуюся медленной функцией времени. При таком подходе полную фазу сигнала можно представить как

$$\psi(t) = \omega_0 t + \Theta(t) + \Theta_0, \quad (1.30)$$

Таким образом, сигнал с угловой модуляцией записывается в виде:

$$S(t) = A_0 \cos [\omega_0 t + \Theta(t) + \Theta_0], \quad (1.31)$$

Эти соотношения устанавливают связь между изменениями частоты и фазы.

Поясним эти соотношения на примере простейшей гармонической ЧМ, когда мгновенная частота колебания определяется выражением

$$\omega(t) = \omega_0 + \omega_D \cos \Omega t, \quad (1.32)$$

где ω_D представляет собой амплитуду частотного отклонения, которая называется девиацией частоты или просто девиацией.

Составим выражение для мгновенного значения колебания. Для полной фазы имеем

$$\psi(t) = \int_0^t (\omega_0 + \omega_D \cos \Omega t) dt + \Theta_0, \quad (1.33)$$

или

$$\psi(t) = \omega_0 t + \frac{\omega_D}{\Omega} \sin \Omega t + \Theta_0, \quad (1.34)$$

таким образом

$$S(t) = A_0 \cos \left[\omega_0 t + \frac{\omega_D}{\Omega} \sin \Omega t + \Theta_0 \right]. \quad (1.35)$$

Фаза колебания $S(t)$ состоит из линейно-возрастающего слагаемого $\omega_0 t$ и нелинейной составляющей $(\omega_D/\Omega)\sin\Omega t$, т.е. его можно рассматривать как колебание, модулированное по фазе. Закон этой модуляции является интегральным по отношению к закону изменения частоты. Именно модуляция по закону $\omega_D \cos\Omega t$ приводит к модуляции фазы по закону $(\omega_D/\Omega)\sin\Omega t$.

Амплитуда изменения фазы $\Theta_{\max} = \omega_D/\Omega = m_\phi$ часто называют **индексом угловой модуляции**.

Индекс модуляции не зависит от частоты ω_0 , а определяется девиацией ω_D и моделирующей частотой Ω .

Рассмотрим случай, когда стабильное по частоте и фазе колебание подвергается модуляции по закону

$$\Theta(t) = \Theta_{\max} \sin \Omega t, \quad (1.36)$$

т.е. колебание имеет вид:

$$S(t) = A_0 \cos(\omega_0 t + \Theta_{\max} \sin \Omega t + \Theta_0). \quad (1.37)$$

Частота этого колебания будет определяться как:

$$\omega(t) = \frac{d\psi(t)}{dt} = \frac{d}{dt}(\omega_0 t + \Theta_{\max} \sin \Omega t + \Theta_0) = \omega_0 + \Theta_{\max} \Omega \cos \Omega t.$$

Сравнивая последние два выражения приходим к тому, что $\Theta_{\max} \Omega = \omega_D$.

Таким образом, гармоническая модуляция фазы с индексом Θ_{\max} эквивалентна частотной модуляции с девиацией $\omega_D = \Theta_{\max} \Omega$. Из этого примера видно, что при гармонической модуляции нельзя отличить ФМ от ЧМ, так как угол Θ при ФМ изменяется по закону $\Theta = \Theta_{\max} \sin \Omega t$, и при ЧМ -

$$\Theta = \left(\frac{\omega_D}{\Omega} \right) \sin \Omega t.$$

Система связи с частотной модуляцией обладает более высокой помехоустойчивостью по сравнению с системой связи с амплитудной модуляцией. Указанное свойство и определило широкое применение ЧМ для радиосвязи в УКВ диапазоне. Это вытекает из сопоставления спектров АМ и ЧМ сигналов.

При АМ полоса частот сигнала равна:

$$\Delta f_{AM} = 2F_{МАКС}; \quad (F_{МАКС} \text{ для речи}) \approx 4 \text{ [кГц]}. \quad (1.38)$$

Ширина спектра ЧМ радиосигнала зависит от индекса модуляции β , ко-

торый определяется выражением:

$$S_{ЧМ} = S_{АМ} \quad (\text{При } \beta \ll 2). \quad (1.39)$$

Отношение девиации к частоте модуляции (манипуляции) называется **индексом частотной** модуляции (манипуляции).

$$\beta = \frac{f_d}{F_m} = \frac{2f_d}{V_m} = \frac{f_p}{V_m}.$$

При $\beta \ll 1$ спектр частот как при частотной модуляции, так и при амплитудной модуляции состоит из несущей и двух боковых полос, т.е. в этом случае ЧМ радиосигнал занимает полосу

$$\Delta f_{ЧМ} = 2 F_{МАКС}, \quad (1.40)$$

то есть $\Delta f_{ЧМ} \leftrightarrow \Delta f_{АМ}$.

При больших индексах модуляции β спектр частотно-модулированного колебания будет состоять из несущей частоты и бесконечного множества частот.

Теоретически полоса ЧМ сигнала является бесконечной. Практически во внимание следует принимать составляющие спектра с достаточной амплитудой.

При увеличении β приходится учитывать большее количество боковых частот. Обычно полоса ЧМ спектра сигнала учитывает боковые частоты и связана с индексом модуляции β , т. е. частоты $[\omega_0 - \beta\Omega, \omega_0 + \beta\Omega]$.

В этом случае ширина спектра ЧМ сигнала равна:

$$\Delta f_{ЧМ} = 2\beta F_{МАКС} = 2f_d. \quad (1.41)$$

В качестве особенностей следует отметить, что при ЧМ модуляции уменьшение модулирующей частоты приводит к увеличению индекса модуляции:

$$U_{ЧМ}(t) = U_m \cdot \cos(\omega_0 t + \omega_d / \Omega \cdot \cos \Omega t). \quad (1.42)$$

Это приводит к увеличению количества составляющих боковых частот, превосходящих заданную интенсивность. Хотя расстояние между смежными составляющими (равные частоты модуляции) уменьшается. При этом полоса частот остается примерно постоянной.

При ЧМ для всех модулирующих частот легче согласовать полосу пропускания приемника с полосой сигнала. (При ФМ сигнале - это сложнее).

При ФМ уменьшение модулирующей частоты приводит к уменьшению полосы частот ФМ сигнала, т. к. расстояние между смежными составляющими уменьшается, следовательно, количество составляющих боковых частот заданного уровня остаются неизменными.

$$U_{ФМ} = U_m \cdot \cos(\omega_0 t + \beta \cdot \cos \Omega t). \quad (1.43)$$

Ширина спектра при ФМ радиосигнале определяется спектром сообще-

ния и значением индекса фазовой модуляции, то есть:

$$S_{\Phi M}(\omega) = f(\alpha, m\Phi). \quad (1.44)$$

Зависимость ширины спектра ФМ радиосигнала на уровне 0,1 от максимального значения представлена на рисунке 1.22.

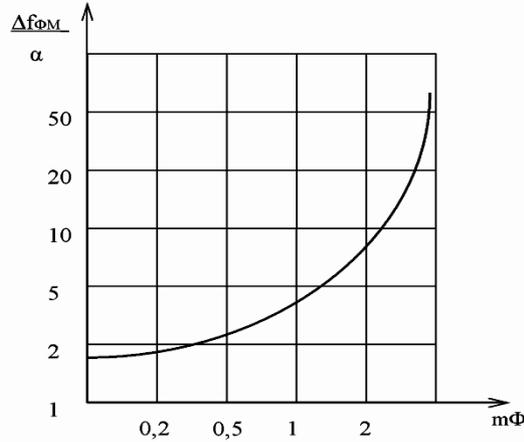


Рисунок 1.22 - Зависимость ширины спектра ФМ радиосигнала на уровне 0,1 от максимального значения

Недостатком ФМ радиосигналов является сильная зависимость помехоустойчивости от стабильности частоты радиоканалов.

Зависимость ширины спектра ЧМ радиосигналов представлена на рисунке 1.23.

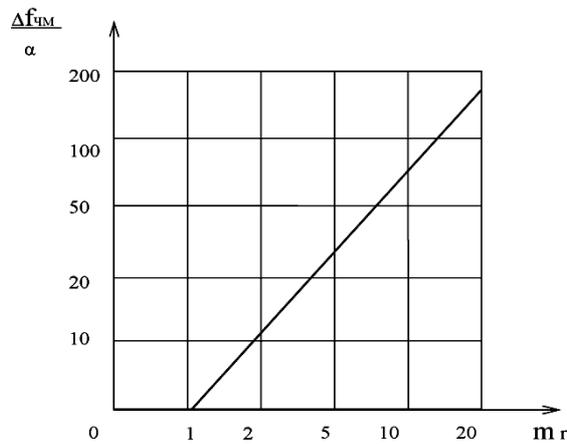


Рисунок 1.23 - Зависимость ширины спектра ЧМ радиосигналов

Из представленных зависимостей следует, что:

- ширина спектра ЧМ радиосигналов определяется индексом частотной модуляции;
- помехоустойчивость увеличивается по мере расширения спектра сигнала;

– при частотной модуляции снижается уровень порогового эффекта.

1.8.1 Принцип формирования и прием сигналов с относительной фазовой модуляцией

Как известно, наименьшая вероятность ошибочного приема при $m = 2$ может быть достигнута при использовании фазоманипулированных на π сигналов: $s_1(t) = -s_2(t) = s_0(t)$, $0 \leq t \leq T_c$ причем обычно $s_0(t) = A_0 \sin \omega_0 t$.

Однако практическая реализация демодуляторов ФМ сигналов встречает определенные трудности, связанные с созданием опорного напряжения с не-изменной начальной фазой.

В существующих СПИ опорный сигнал формируется из принимаемого сигнала. В системах с ФМ задача затрудняется тем, что при равновероятных сигналах в их спектре отсутствует составляющая на частоте несущей и ее невозможно получить путем фильтрации. В этих случаях приходится применять способы формирования опорного напряжения, основанные на снятии манипуляции принятого сигнала. Примерами соответствующих устройств служат схемы Пистолькорса, Сифорова, Костаса и др.

Однако всем известным схемам формирования опорного сигнала в системах с ФМ присущ одинаковый недостаток: из-за воздействия различных неконтролируемых факторов возможны случайные изменения фазы опорного сигнала на π . При этом даже в отсутствие помех передаваемый символ 1 регистрируется как 0, а передаваемый символ 0 - как 1. Возникает явление, называемое «обратной работой», которое будет продолжаться до следующего случайного скачка фазы опорного сигнала.

Поэтому противоположные сигналы невозможно использовать в режиме обычной ФМ в радиоканалах. Однако существует метод, позволяющий ценой небольшого энергетического проигрыша реализовать в радиоканалах преимущества таких сигналов. Этот метод, предложенный Н. Т. Петровичем и получивший название относительной фазовой модуляции (ОФМ), заключается в том, что полезная информация содержится не в абсолютном значении начальной фазы сигнала, а в разности начальных фаз двух соседних сигналов. Для передачи символа 0 начальная фаза передаваемого колебания сохраняется неизменной по отношению к начальной фазе колебания на интервале длительности предшествующего символа. Для передачи символа 1 начальная фаза излучаемого колебания поворачивается на π .

При когерентной обработке сигналов с ОФМ, как и при приеме сигналов с ФМ, прежде всего определяются начальные фазы принятых сигналов, а затем на основе сравнения начальных фаз соседних сигналов принимаются решения о переданных информационных символах. В данном случае при каж-

дом случайном скачке фазы опорного колебания в приемнике будет ошибочно принят только один символ.

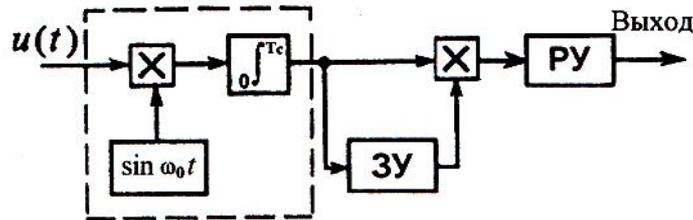


Рисунок 1.24 – Структурная схема оптимального когерентного демодулятора ОФМ сигналов

Алгоритм оптимального приема сигналов с ОФМ реализует устройство (рисунок 1.24), включающее запоминающее устройство ЗУ любого типа, сохраняющее на время T_C информацию о выборочном значении напряжения на выходе интегратора в момент окончания каждого сигнала (аналоговую линию задержки, схему выборки и хранения и т.п.). Как видно из рисунка 1.24, приемное устройство состоит из двух частей, первая из которых (обведена штриховой линией) по существу представляет собой устройство оптимального приема противоположных сигналов. Тогда вероятность ошибочного приема легко определить из следующих соображений. Ошибочная регистрация сигнала в схеме на рисунке 1.24 будет происходить в одном из двух возможных случаев:

1. ошибочно определена начальная фаза предыдущего сигнала, в то время как начальная фаза последующего определена верно;
2. второй сигнал принят ошибочно, а первый - верно.

Вероятность каждого из этих несовместимых событий равна $p_{ош.фм}(1-p_{ош.фм})$, где $p_{ош.фм}$ - вероятность ошибочного приема сигналов с ФМ, определяемая выражением (3.26). Тогда искомая вероятность ошибочного приема сигналов с ОФМ

$$p_{ош.фм} = 2p_{ош.фм}(1-p_{ош.фм}) \approx 2 \left[1 - F \left(\sqrt{2E/N_0} \right) \right] = 2 \left[1 - F(\sqrt{2}h) \right] \quad (1.45)$$

Из (1.45) можно определить энергетический проигрыш метода ОФМ по сравнению с ФМ. Он не превосходит 1 дБ.

Относительную ФМ можно рассматривать как обычную, но при соответствующем дополнительном кодировании передаваемого сообщения. Действительно, если передаваемые символы обозначить $a_1, a_2, \dots, a_k, \dots$, то после такого дополнительного кодирования имеем $b_1, b_2, \dots, b_k, \dots$, причем совпадение символов b_{k-1} и b_k означает передачу исходного символа 0, а в противном случае (символы b_{k-1} и b_k различны) - передачу символа 1. Иначе говоря,

$$b_{k-1} \oplus b_k = a_k \quad (1.46)$$

При случайной начальной фазе каждого передаваемого сигнала применять ОФМ, строго говоря, невозможно. Однако на практике часто встречается ситуация, когда эта начальная фаза претерпевает достаточно медленные изменения, так что разность начальных фаз двух соседних сигналов можно считать не меняющейся случайным образом ни при излучении, ни в процессе распространения на протяжении времени передачи большого числа информационных символов. В то же время для упрощения приемного устройства само абсолютное значение начальной фазы каждой такой пары сигналов при приеме будем полагать случайным. Тогда можно использовать метод ОФМ, но в сочетании с оптимальной некогерентной обработкой пар последовательно передаваемых сигналов. Каждая такая пара сигналов, рассматриваемая как некоторый эквивалентный сигнал $s_{r_3}(t)$ ($r=1,2$), либо имеет скачок на π начальной фазы второго сигнала относительно первого ($r=1$; передается символ 1), либо не имеет такого скачка ($r=2$; передается символ 0).



Рисунок 1.25 – Структурная схема оптимального когерентного демодулятора ОФМ сигналов

В этих условиях задача приема сигнала с ОФМ аналогична задаче оптимального некогерентного приема сигналов $s_{1_3}(t)$ и $s_{2_3}(t)$. Алгоритм такого приема следует из того, что для оптимального демодулятора $V_1 \geq V_2$: регистрируется сигнал $s_{1_3}(t)$ (символ 1), если $V_{1_3} \geq V_{2_3}$, где

$$V_{r_3} = \sqrt{\left[\int_{-T_c}^{T_c} u(t) s_{r_3}(t) dt \right]^2 + \left[\int_{-T_c}^{T_c} u(t) \hat{s}_{r_3}(t) dt \right]^2}$$

Можно показать, что сигналы $s_{1_3}(t)$ и $s_{2_3}(t)$ ортогональны в усиленном смысле. Следовательно, вероятность ошибочного приема $P_{OШ}$ определяется формулой (1.47).

$$P_{OШ} = \left(\frac{1}{2} \right) \exp \left(\frac{-h_s^2}{2} \right) \quad (1.47)$$

где $h_s^2 = \frac{1}{N_0} \int_{-T_c}^{T_c} s_{1_3}^2(t) dt = \frac{1}{N_0} \int_{-T_c}^{T_c} s_{2_3}^2(t) dt = \frac{2E}{N_0} = 2h^2$.

Таким образом

$$P_{out} = \left(\frac{1}{2}\right) \exp(-h^2) \quad (1.48)$$

Как видно из сравнения (1.48) и (1.47), применение ОФМ позволяет без снижения скорости передачи информации уменьшать вдвое энергию излучаемых сигналов (вдвое снизить среднюю мощность передатчика) по сравнению с использованием сигналов, ортогональных в усиленном смысле. Иначе говоря, при оптимальном некогерентном приеме сигналов с ОФМ имеется энергетический выигрыш 3 дБ по сравнению с таким же приемом ортогональных в усиленном смысле сигналов. В то же время, сравнивая (1.48) с (1.45), легко убедиться, что вероятность ошибок при оптимальном некогерентном приеме сигналов с ОФМ несколько выше, чем при оптимальном когерентном приеме тех же сигналов, однако это различие очень мало. Соответствующий энергетический проигрыш некогерентного приема не превышает 1 дБ.

Многократная фазовая модуляция

Многократная фазовая (МФМ) и амплитудно-фазовая (МАФМ) модуляция позволяют эффективно использовать полосу частот канала, но при большом отношении сигнал/шум и отсутствии линейных и нелинейных искажений сигнала. Это ограничивает область применения таких видов модуляции в основном кабельными и радиорелейными каналами связи. В радиоканалах с ненаправленным распространением радиоволн, к которым относятся, например, каналы мобильных систем связи, несмотря на достаточно жесткие требования к полосе частот, используется в основном **четырёхкратная фазовая модуляция (ФМ-4)** и ее разновидности. Это объясняется тем, что при таком же качестве передачи, как при двукратной фазовой модуляции (ФМ-2), ФМ-4 позволяет уменьшить полосу частот примерно в два раза, а ее разновидности дают возможность дополнительно снизить уровень внеполосных излучений.

При четырехфазной модуляции каждая пара двоичных символов источника преобразуется в радиоимпульс с одним из четырех значений фазы несущей (0, $\pi/2$, π , $3\pi/4$). Так как увеличивается длительность посылки, то сужается спектр сигнала. Следует заметить, что при ФМ-4 в отличие от ФМ-2 чаще имеют место скачки по фазе на $\pi/4$, а поэтому изменяется и огибающая спектра, делая его более компактным.

Сигнал при ФМ-4 можно записать в виде $i = 1, 2, \dots, \infty$

$$S(t) = U \cos \left[\omega_0 t + Y_i \left(\frac{\pi}{4} \right) \right], \quad 0 \leq t \leq 2T$$

где T -длительность символа двоичного источника, $Y_i = 0, 1, 2, 3$ - значения пар двоичных символов, $i = 1, 2, \dots, \infty$ - номера пар двоичных символов. Неудобство такого представления четырехфазной модуляции заключается в том, что сиг-

налы в ансамбле оказываются неэквидистантными (коэффициент корреляции имеет два значения 0 и -1). Это несколько усложняет анализ помехоустойчивости. Поэтому проще рассматривать сигнал ФМ-4 как сумму сигналов с двукратной фазовой модуляцией

$$S(t) = \frac{U}{\sqrt{2}} S_i \cos \omega_0 t + \frac{U}{\sqrt{2}} C_i \sin \omega_0 t, 0 \leq t \leq 2T$$

где несущими являются квадратурные составляющие $\cos \omega_0 t$ и $\sin \omega_0 t$ S_i, C_i - соответственно четные и нечетные символы последовательности двоичного источника. Такое представление соответствует общему понятию **квадратурная модуляция**. Если теперь вместо символов 1 и -1 на позициях S_i и C_i использовать непрерывные сигналы, то можно сформировать радиосигнал с любой амплитудой и фазой (как, например, на рисунке 1.26).

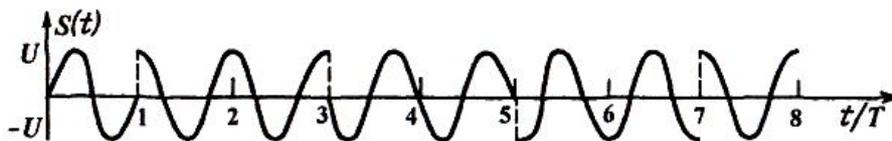


Рисунок 1.26 – Пример сигнала ФМ-4

Поскольку энергия сигналов в квадратурных каналах при четырехфазной модуляции такая же, как при двухфазной, а каналы независимы, то помехоустойчивость ФМ-4 и ФМ-2 при одинаковой скорости передачи должна быть одинаковой.

В общем случае спектр радиосигнала при дискретной фазовой модуляции определяется длительностью радиоимпульса (частотой изменения фазы), величиной изменения фазы и статистическими характеристиками модулирующего случайного дискретного процесса. Для радиосигналов с ФМ-2, когда источник обладает максимальной производительностью и фаза принимает случайные независимые во времени значения 0 и π , огибающая спектра имеет форму $\left(\frac{\sin x}{x}\right)^2$, где $x = (\omega_0 - \omega) \frac{T}{2}$.

Примерно такую же форму огибающей имеет и спектр ФМ-4. Однако энергия сигнала в спектре за пределами полосы $1/T$ достаточно велика. Уменьшить уровень внеполосных излучений можно, изменив форму радиоимпульса, сгладив его. Однако это приведет к появлению амплитудной модуляции и как следствие к требованию линейности амплитудных характеристик приемопередающего тракта. От этого недостатка свободна квадратурная фазовая модуляция со сдвигом и сглаживанием. Модуляция по фазе в квадратурных каналах осуществляется посылками, сдвинутыми относительно друг друга наполовину, т. е. на длительность двоичного символа источника. Сглажива-

ние посылок осуществляется по закону $\cos \frac{\pi t}{2T}$ и $\sin \frac{\pi t}{2T}$. При этом

$$S(t) = \frac{U}{\sqrt{2}} S_i \sin \frac{\pi t}{2T} \cos \omega_0 t + \frac{U}{\sqrt{2}} C_i \cos \frac{\pi t}{2T} \sin \omega_0 t.$$

В результате сложения в квадратурных каналах получаем результирующий сигнал, модуль вектора которого будет иметь постоянную величину, т.е. амплитудная модуляция в сглаженном сигнале отсутствует, а фаза изменяется линейно в пределах длительности посылки T на величину $\pm \pi/2$ (например, как на рисунке 1.27).

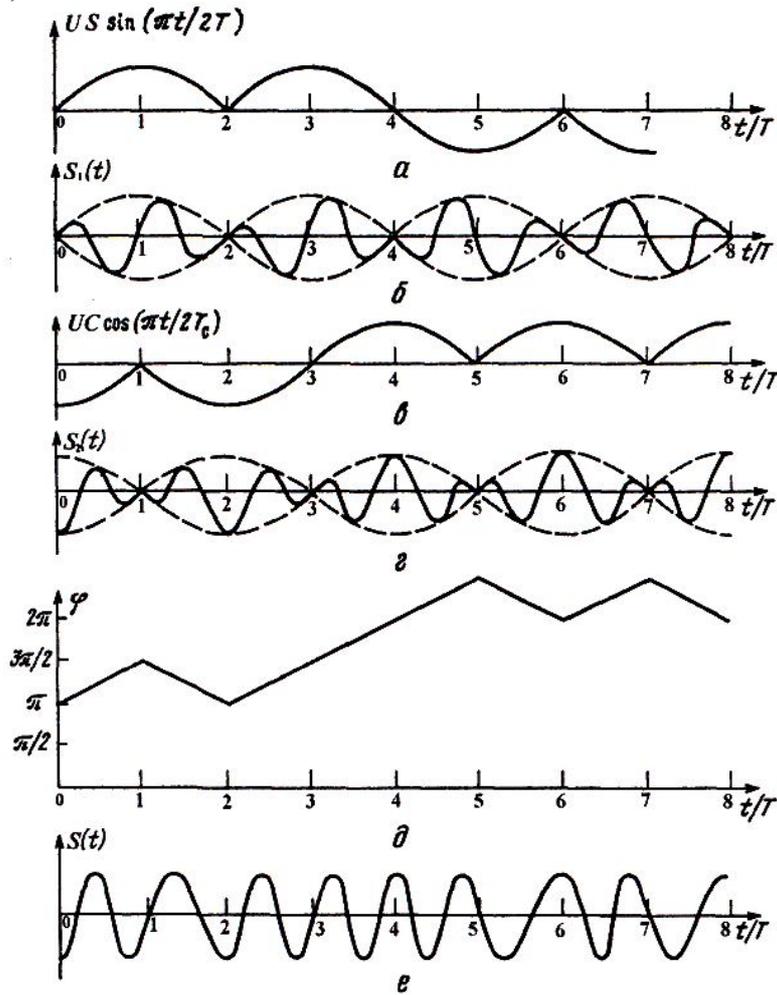


Рисунок 1.27 – Диаграммы формирования частотно-модулированных сигналов с непрерывной фазой

Таким образом, сигнал ФМ-4 со сдвигом и сглаживанием совпадает с сигналом частотно модулированным с непрерывной фазой (ЧМНФ) с индексом модуляции 0,5 или, как его называют в литературе, манипулированным с минимальным сдвигом (ММС):

$$S(t) = U \sin \left[\omega_0 t + \frac{\pi Y_i}{4T} \int_0^{2T} dt \right].$$

Стремление уменьшить уровень внеполосных излучений привело к созданию так называемой квадратурной модуляции с гауссовой фильтрацией, в которой посылки в квадратурных каналах перед сглаживанием пропускаются через низкочастотный фильтр с гауссовой характеристикой. На рисунке 1.28 приведены характеристики мощности внеполосного излучения для всех перечисленных видов квадратурной модуляции.

Достоинства квадратурной модуляции со сглаживанием и гауссовой фильтрацией делает ее перспективной для использования в радиоканалах. Например, в сотовых системах мобильной связи стандарта GSM используется именно такой вид модуляции.

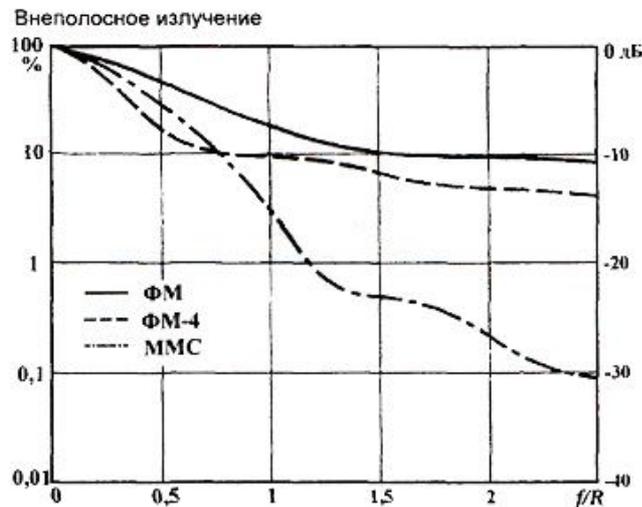


Рисунок 1.28 – Характеристики мощности внеполосного излучения

1.8.2 Оптимальный прием аналоговых сигналов с частотной модуляцией

При частотной модуляции (ЧМ) частота колебания изменяется пропорционально мгновенным значениям модулирующего сообщения $\lambda(t)$.

Поэтому ЧМ радиосигнал может быть представлен в виде:

$$s(t, \lambda) = A_0 \cos (\omega_0 t + \psi (t)),$$

$$\psi (t) = M \int_0^t \lambda (\tau) d \tau, \quad (1.49)$$

где A_0 и ω_0 – априорно известные значения амплитуды и частоты несущей; $\lambda(t)$ -информационное сообщение.

Из (1.49) следует, что полная фаза полезного радиосигнала сигнала при ЧМ складывается из регулярной составляющей $\omega_0 t$ и добавочного члена $\psi(t)$, обусловленного модуляцией, априорные сведения о которой характеризуются двумя уравнениями:

$$\frac{\partial \psi}{\partial t} = M_q \lambda, \quad \frac{\partial \lambda}{\partial t} = -\alpha \lambda + n_\lambda(t). \quad (1.50)$$

Здесь $M_q = \sigma_\omega / \sigma_\lambda$ - крутизна характеристики частотного модулятора, σ_ω - среднеквадратическое значение отклонения частоты радиосигнала от средней частоты ω_0 .

Используя уравнения фильтрации и опустив математические преобразования получим уравнения, определяющие структуру квазиоптимального приемника ЧМ радиосигналов

$$\begin{aligned} \frac{\partial \hat{\psi}}{\partial t} &= -M_q \left(\frac{K}{T_\lambda D + 1} + K_\psi \right) \xi(t) A_0 \sin(\omega_0 t + \hat{\psi}), \\ \hat{\lambda} &= -\frac{K}{T_\lambda D + 1} \xi(t) A_0 \sin(\omega_0 t + \hat{\psi}). \end{aligned} \quad (1.51)$$

где $K = 2R_{\psi\lambda} / \alpha N$, $K_\psi = 2R_{\psi\psi} / M_q N$, $T_\lambda = 1/\alpha$.

Структурная схема квазиоптимального приемника ЧМ радиосигналов приведена на рисунке 1.29.

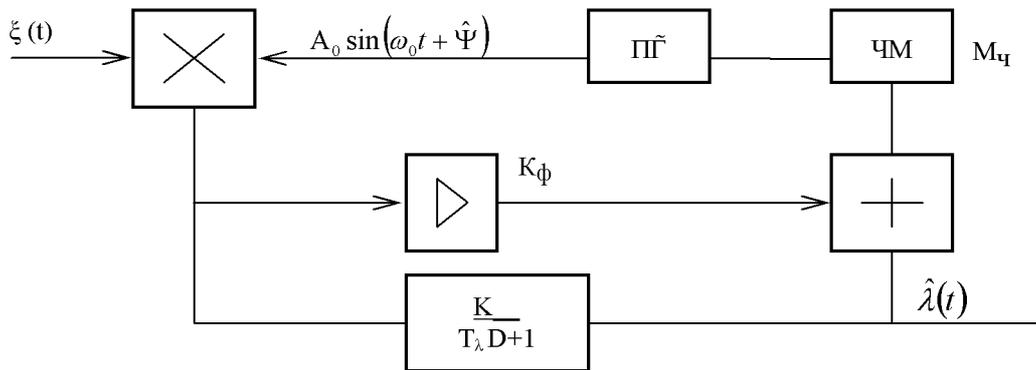


Рисунок 1.29 – Структурная схема квазиоптимального приемника ЧМ радиосигналов

Он представляет собой схему ФАП, в которой управление частотой подстраиваемого генератора осуществляется по двум параллельным каналам.

Помехоустойчивость приема ЧМ- сигнала определяется выражением (1.52)

$$\delta_{\text{чм}}^2 = \frac{R_\lambda}{D_\lambda} = \frac{1}{2q\beta_{\text{хв}}^2} (1 + 2\beta_{\text{хв}}\sqrt{q} - \sqrt{1 + 4\beta_{\text{хв}}\sqrt{q}}) \sqrt{1 + 4\beta_{\text{хв}}\sqrt{q}}, \quad (1.52)$$

где $\beta_{\text{чм}} = \frac{\sigma_\omega}{\alpha}$, $q = \frac{A_0}{2\alpha N}$ - отношение сигнал/шум.

При $q \gg 1$

$$\delta_{\text{чм}}^2 \approx \frac{2}{\sqrt{\beta_{\text{чм}}} \sqrt[4]{q}}, \quad (1.53)$$

Чем больше $\beta_{\text{чм}}$, тем больше помехоустойчивость системы, но при этом расширяется спектр сигнала.

В работе квазиоптимального приемника ЧМ – радиосигналов так же как и при приеме ФМ – радиосигналов наблюдается **демодуляционный порог (пороговый эффект)**.

1.8.3 Оптимальный прием аналоговых сигналов с амплитудной модуляцией

Рассмотрим основные характеристики АМ радиосигнала. Пусть передаваемое сообщение $\lambda(t)$ имеет спектр

$$s(j\omega) = \int_{-\infty}^{\infty} \lambda(t) e^{-j\omega t} dt, \quad (1.54)$$

причем основная доля энергии спектра $s(j\omega)$ сосредоточены в полосе модулируемых частот, которую характеризуют граничными частотами ω_{\min} и ω_{\max} . Удобно рассмотреть характеристики моделируемого сигнала, когда сообщение имеет вид гармонического колебания (рисунок 1.60) с частотой Ω , лежащей в интервале $[\Omega_{\min}; \Omega_{\max}]$, и начальной фазой, равной нулю:

$$a_{\Omega}(t) = A_{\Omega} \cos \Omega t, \quad (1.55)$$

т.е. случай однотоновой (или просто тональной) модуляции при этом

$$s_{\text{AM}}(t) = A_0 [1 + m \cos \Omega t] \cos(\omega_0 t + \varphi_0), \quad (1.56)$$

где A_0 - значение амплитуды в режиме молчания. При неискаженной модуляции ($m \leq 1$) амплитуда высокочастотных колебаний АМ - радиосигнала изменяется в пределах от $A_{0\min} = (1 - m)A_0$ до $A_{0\max} = (1 + m)A_0$, откуда

$$m = \frac{A_{\max} - A_{\min}}{A_{\max} + A_{\min}}, \quad (1.57)$$

Спектр АМ - радиосигнала при тональной модуляции содержит компоненты несущего (основного) колебания с амплитудой A_0 и частотой ω_0 и колебаний двух боковых составляющих с равными амплитудами $A_{\sigma} = \frac{A_0 \omega}{2}$.

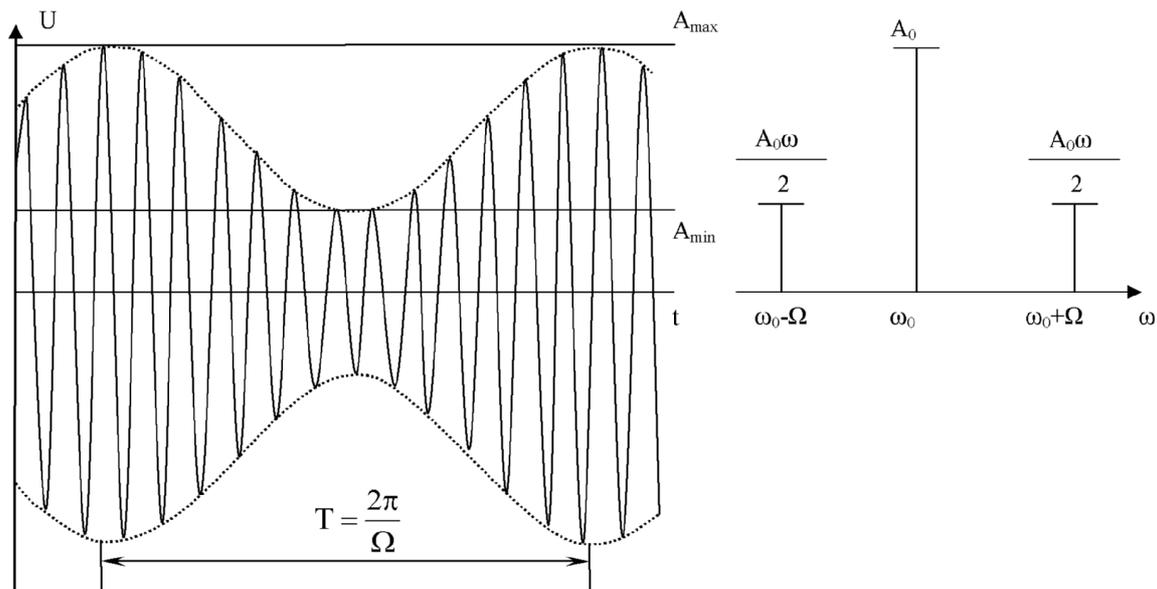


Рисунок 1.60 – АМ – радиосигнал и его спектр

АМ - радиосигнал наглядно можно представить с помощью векторной диаграммы (рисунок 1.31). Для этого сигнал представим в виде суммы гармоник:

$$s(t) = A_0 \cos \omega_0 t + A'_\sigma \cos(\omega_0 + \Omega)t + A''_\sigma \cos(\omega_0 - \Omega)t. \quad (1.58)$$

Каждую составляющую этого сигнала проектируем на вертикальную ось вектора, вращающегося против часовой стрелки.

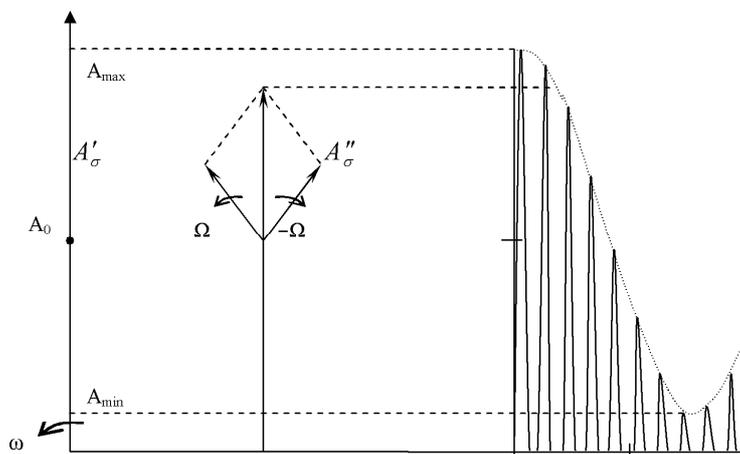


Рисунок 1.31 – Представление АМ-радиосигнала с помощью векторной диаграммы

Спектральная плотность АМ - радиосигнала $s(t)$ с непериодической огибающей $A(t)$ определим с помощью преобразования Фурье, т.е.

$$\begin{aligned} \dot{s}(\omega) = F|s(t)| &= \int_{-\infty}^{\infty} A(t) \cos(\omega_0 t + \psi_0) e^{-j\omega t} dt = \frac{1}{2} \int_{-\infty}^{\infty} A(t) e^{j\psi_0} e^{-j(\omega - \omega_0)t} + \\ &+ \frac{1}{2} \int_{-\infty}^{\infty} A(t) e^{j\psi_0} e^{-j(\omega + \omega_0)t} dt = \frac{1}{2} \dot{s}_A(\omega - \omega_0) e^{j\psi_0} + \frac{1}{2} \dot{s}_A(\omega + \omega_0) e^{-j\psi_0}, \end{aligned} \quad (1.59)$$

где $\dot{s}_A(\omega) = F\{A(t)\}$ - спектральная плотность огибающей сигнала.

При $\psi_0 = 0$

$$\dot{s}(\omega) = \frac{1}{2} \dot{s}_A(\omega - \omega_0) + \frac{1}{2} \dot{s}_A(\omega + \omega_0) \quad (1.60)$$

Выражение (1.60) определяет правило построения спектральных характеристик АМ - радиосигнала $\dot{s}(\omega)$ по спектральным плотностям его огибающей $\dot{s}_A(\omega)$.

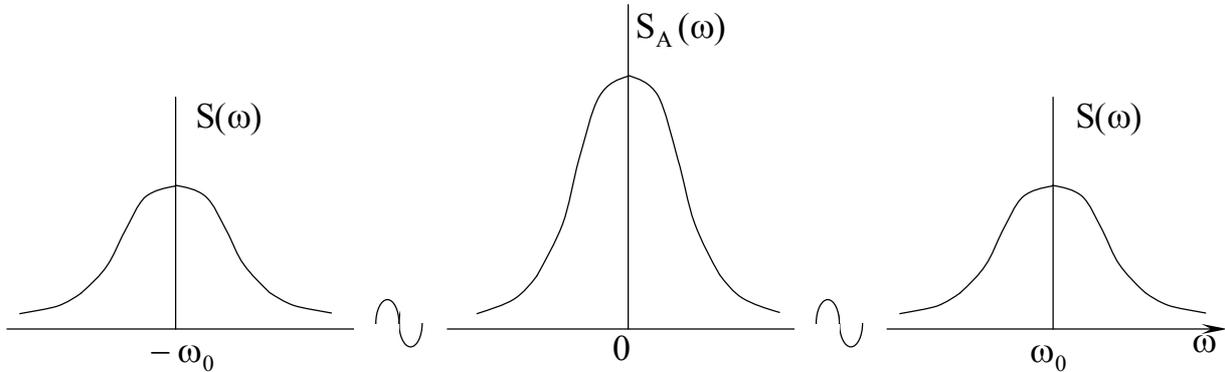


Рисунок 1.32– Спектральная характеристика АМ - сигнала

Для узкополосных сигналов, т.е. при $\frac{\Delta f_p}{f_0} \ll 1$ с высокой степенью точности можно принять:

$$\dot{s}(\omega) \approx \begin{cases} \frac{1}{2} \dot{s}_A(\omega - \omega_0), & \omega > 0, \\ \frac{1}{2} \dot{s}_A(\omega + \omega_0), & \omega < 0. \end{cases} \quad (1.61)$$

Таким образом, модуль спектральной плотности в области $\omega > 0$ определяется модулем спектральной плотности огибающей и может быть получен путем смещения его по оси частот на величину ω_0 и деления его ординат пополам. Аналогично определяется $\dot{s}(\omega)$ и в области $\omega < 0$.

С учетом этого результата активная ширина спектра радиосигнала Δf_p в два раза больше активной ширины спектра Δf_{os} , то есть $\Delta f_p = 2 \Delta f_{os}$.

Это следует из того, что сам радиосигнал $s(t)$ связан со всей спектральной плотностью $\dot{s}(\omega)$ обратным преобразованием Фурье

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{s}(\omega) e^{j\omega t} d\omega \quad (1.62)$$

или через одностороннее преобразование Фурье

$$s(t) = 2 \operatorname{Re} \left\{ \frac{1}{2\pi} \int_0^{\infty} \dot{s}(\omega) e^{j\omega t} d\omega \right\}, \quad (1.63)$$

то есть сигнал полностью определяется комплексной спектральной плотностью $\dot{s}(\omega)$.

В области положительных частот при $\frac{\Delta f_p}{f_0} \ll 1$

$$s(t) \approx \operatorname{Re} \left\{ \frac{1}{2\pi} \int_0^{\infty} \dot{s}_A(\omega - \omega_0) e^{j\omega t} d\omega \right\}. \quad (1.64)$$

Методы формирования и оптимальный прием АМ радиосигналов

АМ радиосигнал может быть записан в следующем виде:

$$s(t, \lambda) = (A_0 + M_a \lambda) \cos(\omega_0 t + \varphi(t)), \quad (1.65)$$

где $M_a = \sigma_a / \sigma_\lambda$ - крутизна характеристики амплитудного модулятора;

$\sigma_a = M_a \sigma_\lambda$ - среднеквадратическое отклонение огибающей от уровня несущей;

$\varphi(t)$ - случайная фаза, рассматриваемая как сопутствующий параметр.

Во многих практических случаях можно считать, что случайная фаза радиосигнала описывается дифференциальным уравнением:

$$\frac{\partial \varphi}{\partial t} = n_\varphi(t), \quad (1.66)$$

где $n_\varphi(t)$ - белый шум с функцией корреляции

$$M[n_\varphi(t) n_\varphi(t + \tau)] = (N_\varphi / 2) \delta(\tau). \quad (1.67)$$

После нормирования вид функциональной зависимости (1.65) АМ радиосигнала получим:

$$\begin{aligned} s(t, \lambda) &= (A_0 + M_a \lambda) \cos(\omega_0 t + \varphi) = A_0 \left(1 + \frac{\sigma_a}{A_0} \frac{\lambda}{\sigma_\lambda} \right) \cos(\omega_0 t + \varphi) = \\ &= A_0 (1 + m_a \lambda') \cos(\omega_0 t + \varphi), \end{aligned} \quad (1.68)$$

где $m_a = \frac{\sigma_a}{A_0}$ - коэффициент амплитудной модуляции;

$\hat{\lambda}' = \frac{\lambda}{\sigma_\lambda}$ - нормированное сообщение.

Уравнения фильтрации в стационарном режиме для случая АМ радиосигнала имеют вид:

$$\begin{cases} \hat{\lambda} = \frac{K}{TD+1} \left(\xi(t) 2 \cos(\omega_0 t + \hat{\phi}) - A_0 \right) \\ \partial \hat{\phi} / \partial t = -K_\phi \xi(t) \left(A_0 + M_a \hat{\lambda} \right) 2 \sin(\omega_0 t + \hat{\phi}) \end{cases} \quad (1.69)$$

где $T = T_\lambda / (1 + K_\lambda)$, $T_\lambda = 1 / \alpha$, $K = K_\lambda / (1 + K_\lambda) M_a$,

$$K_\lambda = \frac{R_{\lambda\lambda} M_a^2}{\alpha N} = \sqrt{1 + 4q \frac{m_a^2}{1 + m_a^2}} - 1, \quad K_\phi = R_{\phi\phi} / N.$$

Уравнения (1.69) определяют структурную схему квазиоптимального приемника, которая приведена на рисунке 1.33.

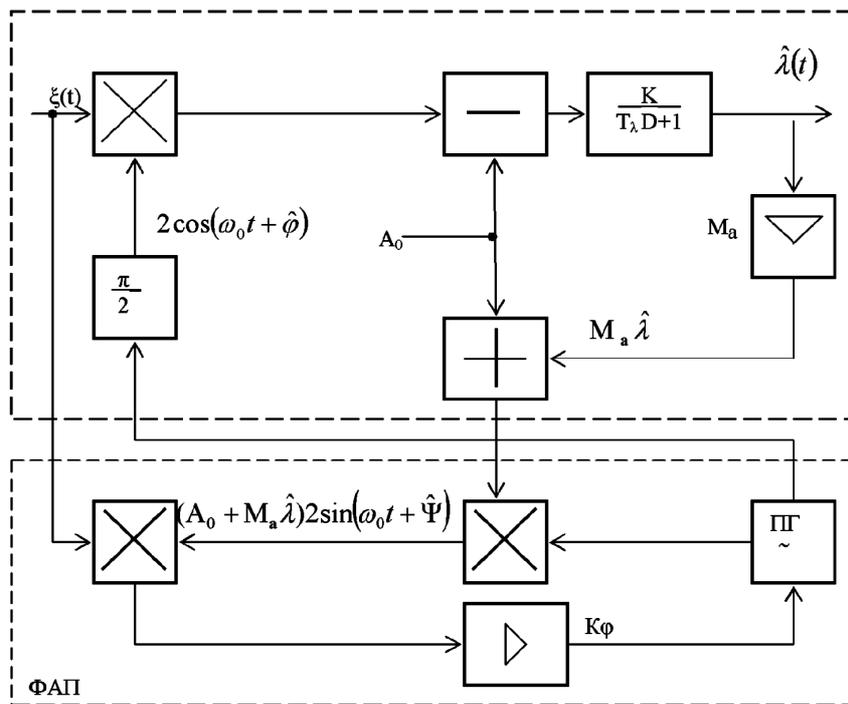


Рисунок 1.33 - Структурная схема квазиоптимального приемника АМ радиосигналов

Особенностью синтезированной схемы оптимального приемника является наличие в нем синхронного детектора, опорное напряжение для которого вырабатывается ФАП и оптимального линейного фильтра.

Основной особенностью синхронного детектора является обеспечение

равенства выходного значения отношения сигнал/шум входному. В существующих радиоприемных устройствах для детектирования АМ радиосигналов обычно используются детекторы огибающей. При малых значениях сигнал/шум в них наблюдается эффект подавления слабого сигнала помехой. Поэтому применение синхронного детектора обеспечивает примерно двукратный выигрыш по мощности по сравнению с детектором огибающей.

Оптимальный линейный фильтр обеспечивает фильтрацию сообщения $\lambda(t)$ от шума с минимально возможной ошибкой. Ошибка фильтрации состоит из двух составляющих. Одна из них – динамическая, обусловлена искажением сообщения $\lambda(t)$ при прохождении его через фильтр, другая – шумовая, обусловлена прохождением шума через фильтр.

Зная операторный коэффициент передачи фильтра $K(p)$ запишем АЧХ линейного фильтра через комплексный коэффициент передачи, который получается заменой p на $j\omega$:

$$K(\omega) = |K(j\omega)| = \frac{K}{\sqrt{1 + T^2 \omega^2}}, \quad (1.70)$$

где $K(j\omega)$ – комплексный коэффициент передачи,

$$T = \frac{T_\lambda}{1 + K_\lambda}, \quad T_\lambda = 1/\alpha, \quad K = \frac{K_\lambda}{(1 + K_\lambda) \cdot M_a}, \quad K_\lambda = \frac{R_{\lambda\lambda} M_a^2}{\alpha N_0} = \sqrt{1 + 4q \frac{m_a^2}{1 + m_a^2}} - 1.$$

АЧХ линейного фильтра определяется отношением сигнал/шум (q). Например, при увеличении q АЧХ расширяется (рисунок 1.34).

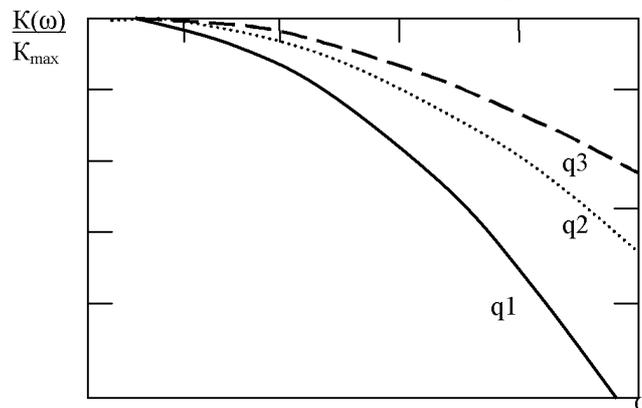


Рисунок 1.34 – Зависимость АЧХ от отношения сигнал/шум

Очевидно, что при этом динамическая составляющая ошибки фильтрации уменьшается, а шумовая – возрастает, но в меньшей степени по сравнению с динамической составляющей.

Проводя анализ помехоустойчивости приема АМ радиосигналов, можно отметить, что относительная ошибка фильтрации в данном случае определяет-

ся выражением вида:

$$\delta_{AM}^2 = \frac{1}{2q \frac{m^2}{1+m^2}} \left(\sqrt{1 + 4q \frac{m^2}{1+m^2}} - 1 \right), \quad (1.71)$$

или при $q \gg 1$

$$\delta_{AM}^2 = \frac{1}{\sqrt{q \frac{m^2}{1+m^2}}}. \quad (1.72)$$

ГЛАВА 2 ПЕРЕДАЧА НЕПРЕРЫВНЫХ СООБЩЕНИЙ ПО ЦИФРОВЫМ КАНАЛАМ

2.1 Импульсная передача непрерывных сообщений

Достаточно широкое распространение в настоящее время получили импульсные системы радиосвязи. Эти системы связи, в частности, позволяют осуществить временное уплотнение каналов. В импульсных системах связи аналоговое сообщение передается своим выборочным значением, определяемые моментами взятия отсчета. Также на практике широкое распространение получили системы, основанные на преобразовании аналогового сообщения в цифровую форму.

2.1.1 Характеристика цифровой обработки сигналов

Цифровая обработка сигналов (ЦОС) в последние годы все шире используется в радиоприемных устройствах. Прогресс в этой области вызван достижениями в микроэлектронике, позволившими создать вычислительные средства, обладающие высоким быстродействием, малыми габаритами, весом и энергопотреблением. Интерес к цифровой обработке сигналов вызван тем, что на ее основе можно создавать устройства с характеристиками, недостижимыми при использовании аналоговых методов обработки сигналов. Кроме того, применение устройств с цифровой обработкой в ряде случаев оказывается более выгодным с технической и экономической точек зрения из-за их универсальности и возможности работать в различных режимах. Сфера применения цифровой обработки непрерывно расширяется. Это радиосвязь, радио-, гидро- и звуколокация, телеметрия, анализ спектров, обнаружение сигналов на фоне помех, адаптивная коррекция каналов связи, адаптивная компенсация помех, анализ и синтез речи, радиовещание, телевидение, цифровые синтезаторы частот, цифровые методы измерений, обработка сигналов в геологоразведке, сейсмологии, медицине и т.д.

Цифровую обработку сигналов следует отличать от цифровых методов передачи сообщений, когда подлежащие передаче аналоговые сигналы преобразуются в цифровую форму уже на передающей стороне. При этом может устраняться избыточность в сообщениях для снижения скорости цифрового потока. Для повышения помехозащищенности применяется избыточное кодирование, перемежение и другие методы.

Цифровая обработка сигналов - это более широкое понятие, чем цифровые методы передачи сообщений. Она включает в себя, кроме описанных выше преобразований, додетекторную обработку (фильтрацию), детектирование

и последетекторную обработку сигналов цифровыми методами. При этом передаваемые по каналам связи сообщения могут быть как цифровыми, так и аналоговыми. Цифровая обработка сигналов может охватывать не только фильтрацию и детектирование на приемной стороне, но и формирование модулированных или манипулированных сигналов на входе канала связи цифровыми методами.

Несмотря на множество уже решенных технических задач, в области применения ЦОС существует ряд проблем, которые сдерживают широкое применение цифровой обработки в радиоприемных устройствах различного назначения. Это ограниченное быстродействие цифровой элементной базы, ограниченные разрядность и быстродействие преобразователей аналоговых сигналов в цифровые, возникающие при ЦОС дополнительные искажения и шумы, ухудшение массогабаритных, энергетических и экономических характеристик устройств ЦОС по сравнению с аналоговыми, недостаточно разработанные теоретические вопросы и методы расчета элементов и устройств ЦОС с заданными качественными показателями. Эти проблемы связаны как с отсутствием требуемой элементной базы, так и со сложностью происходящих в устройствах ЦОС процессов, математическое описание которых во временной и спектральной областях оказывается гораздо более сложным, чем в аналоговых устройствах. Вместе с тем цифровая обработка сигналов, несмотря на указанные недостатки, имеет ряд преимуществ перед аналоговой обработкой:

- значительно более высокую точность обработки сигналов по сложным алгоритмам;
- гибкую оперативную перестройку алгоритмов обработки сигналов, обеспечивающую как создание многорежимных устройств, так и реализацию адаптивных систем;
- высокую технологичность изготовления устройств ЦОС, связанную с отсутствием необходимости настройки при изготовлении и регулировке при эксплуатации;
- высокую степень совпадения и повторяемости характеристик реализованных устройств с расчетными характеристиками;
- возможность построения развивающихся интеллектуальных систем, способных к реконфигурации, поиску и обнаружению неисправностей;
- большие возможности автоматизации проектирования устройств с ЦОС;
- высокостабильные эксплуатационные характеристики устройств с ЦОС.

Эти преимущества позволяют применять цифровую обработку сигналов во многих радиоприемных устройствах.

2.1.2 Процессы преобразования сигналов при цифровой обработке

Процессы цифровой обработки аналоговых сигналов в радиоприемных устройствах состоят из нескольких этапов, которые представлены в виде структурной схемы на рисунке 2.1. Схема содержит следующие последовательно соединенные устройства:

- аналоговый фильтр (АФ),
- устройство выборки-хранения (УВХ),
- аналого-цифровой преобразователь (АЦП),
- цифровое вычислительное устройство (ЦВУ),
- цифро-аналоговый преобразователь (ЦАП),
- аналоговый интерполяционный фильтр (АИФ).

На рисунке 2.2 приведена структурная схема преобразования аналогового сигнала в цифровую форму, которая отражает преобразование аналогового сигнала $x(t)$ последовательность чисел в двоичном коде $x(n)$, следующих с периодом дискретизации T_d на выходе АЦП. Это преобразование включает следующие этапы: дискретизация во времени, квантование по уровню, кодирование.

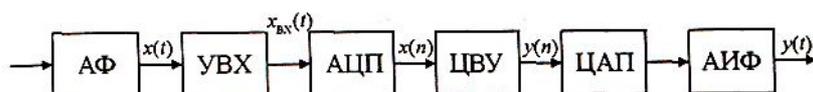


Рисунок 2.1 – Структурная схема процесса цифровой обработки аналоговых сигналов

На рисунке 2.2 $x(t)$ - входной аналоговый сигнал; $x_в(t)$ - сигнал на выходе устройства выборки; $x_{вх}(t)$ - сигнал на выходе устройства выборки-хранения; $x_{к\ вх}$ - квантованный по уровню сигнал на выходе; $W(t)$ - сигнал выборки, представляющий последовательность коротких прямоугольных импульсов.

Задачей УВХ является определение мгновенного значения входного сигнала в момент отсчета и фиксация этого значения на время, необходимое для преобразования его в число. Такое преобразование получило название *дискретизации*.

Дискретизация — это преобразование аналогового сигнала, непрерывного по уровню и во времени, в сигнал, непрерывный по уровню, но дискретный во времени. Под дискретностью во времени понимается тот факт, что полученный в результате этого преобразования дискретизированный сигнал изменяется по уровню в соответствии с аналоговым сигналом лишь в определенные моменты времени. Обычно подразумевается, что эти моменты следу-

ют через постоянный временной интервал T_d , называемый *периодом дискретизации*.

При представлении дискретизированного сигнала решетчатой функцией этот сигнал равен аналоговому в точках отсчета и равен нулю во всех других точках. Ступенчатая функция получается из решетчатой проведением через ее ординаты отрезков прямых на интервале T_d , параллельных оси абсцисс. Площадь взвешенных дельта-функций равна значению аналогового сигнала в точках отсчета. Интегрирование этих функций на интервалах T_d дает ступенчатую функцию.

Высота прямоугольных импульсов равна значению аналогового сигнала в точках отсчета. При длительности импульсов $t_u = T_d$ последовательность прямоугольных импульсов превращается в ступенчатую функцию, а при $t_u \rightarrow 0$ — в решетчатую.

В основе дискретизации лежит теорема отсчетов Котельникова и теория модуляции. На рисунке 2.2 видно, что преобразование аналогового сигнала в дискретизированный происходит в два этапа: выборка и хранение.

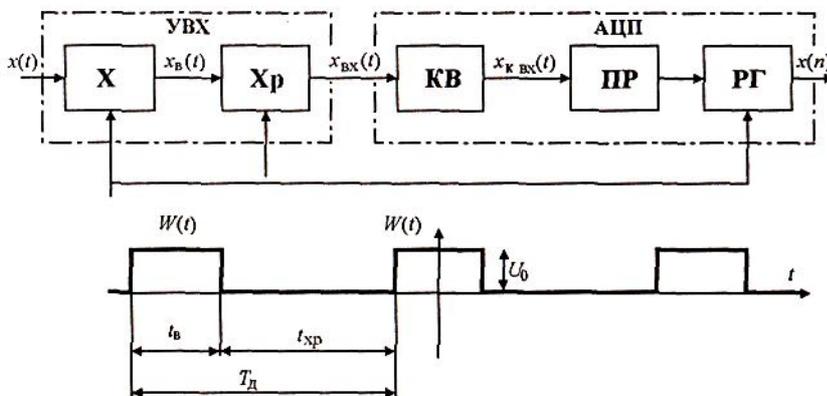


Рисунок 2.2

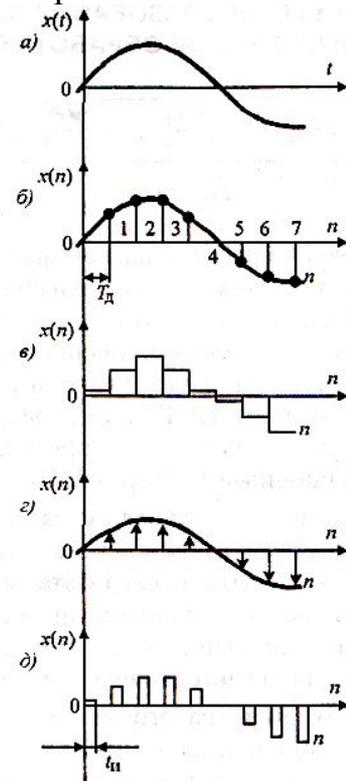


Рисунок 2.3

Представим изображенный на рисунке 2.2 сигнал выборки $W(t)$ рядом Фурье. Для прямоугольных импульсов амплитуды U_0 , длительности t_e пе-

риода следования T_D получим:

$$W(t) = \frac{U_0 t_e}{T_D} \left[1 + 2 \sum_{k=1}^{\infty} \text{sinc}(k \pi t_e F_D \cos(k \omega_D t)) \right],$$

где $\text{sinc}(k \pi t_e F_D) = \frac{\sin(k \pi t_e F_D)}{k \pi t_e F_D}$, $k = 1, 2, 3, \dots$, $F_D = \frac{1}{T_D}$ - частота дискретизации, $\omega_D = 2\pi F_D$

Тогда на выходе множителя в УВХ сигнал

$$x_e(t) = x(t)W(t) = \frac{U_0 t_e}{T_D} \left[x(t) + 2x(t) \sum_{k=1}^{\infty} \text{sinc}(k \pi t_e F_D) \cos k \omega_D t \right]. \quad (2.1)$$

Если входному сигналу $x(t)$ соответствует спектр $S(\omega)$, то каждому произведению $2x(t) \cos k \omega_D t$ соответствует спектр $S(\omega - k \omega_D) + S(\omega + k \omega_D)$

Тогда с учетом того, что спектр суммы равен сумме спектров и $\text{sinc}(0)=1$, спектр сигнала $x_e(t)$ определится из выражения:

$$S_e(\omega) = \frac{U_0 t_e}{T_D} \sum_{k=-\infty}^{\infty} \text{sinc}(k \pi t_e F_D) S(\omega - k \omega_D). \quad (2.2)$$

Из (2.2) следует, что спектр на выходе устройства выборки представляет собой сумму спектров входного сигнала, бесконечно повторяющихся через частотный интервал ω_D вправо и влево от исходного спектра $S(\omega)$ с весовым множителем $\text{sinc}(k U_0 t_e F_D)$. Таким образом, в результате выборки происходит размножение спектра входного сигнала.

Определенный интерес представляет так называемое идеальное устройство выборки, у которого время выборки $t_e = 0$, а произведение $U_0 t_e = 1$. Иначе говоря, сигнал выборки у него представляет последовательность дельта-функций, следующих с периодом T_D . Спектр таких импульсов также представляет последовательность дельта-функций через частотный интервал F_D .

Подставляя соотношение $t_e = 0$ и $U_0 t_e = 1$ в (2.2), получим выражение для спектра сигнала на выходе идеального устройства выборки:

$$S_{us}(\omega) = \frac{1}{T_D} \sum_{k=-\infty}^{\infty} S(\omega - k \omega_D). \quad (2.3)$$

Из (2.3) следует, что на выходе идеального устройства выборки спектр представляет сумму спектров входного сигнала, бесконечно повторяющихся через частотный интервал ω_D вправо и влево от исходного спектра $S(\omega)$.

Дискретизации могут подвергаться сигналы двух видов. Первый вид сигнала $x(t)$ — низкочастотный, спектр которого в основном сосредоточен в

интервале от 0 до $0,5\omega_d$, второй - радиочастотный, спектр которого в основном сосредоточен выше частоты $0,5\omega_d$. Второй случай характерен для приемников, в которых цифровая обработка осуществляется на промежуточной частоте.

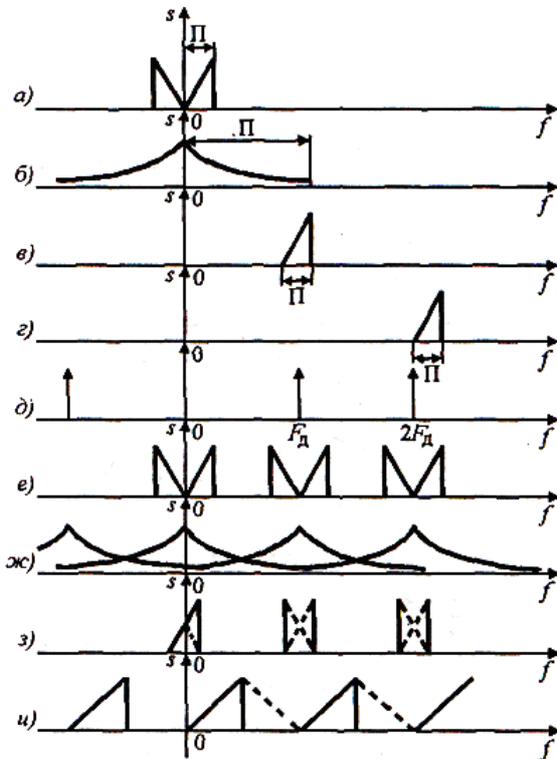


Рисунок 2.4 - Спектры различных входных аналоговых сигналов (а, б, в, г), спектр сигнала выборки в виде последовательности дельта-функций (д) и спектры идеально дискретизированных сигналов (е, ж, з, и)

На рисунке 2.4 изображены спектры различных входных аналоговых сигналов (а, б, в, г), спектр сигнала выборки в виде последовательности дельта-функций (д) и спектры идеально дискретизированных сигналов (е, ж, з, и). Из рисунка видно, что при ширине исходного спектра Π , удовлетворяющего теореме Котельникова ($\Pi < 0,5F_d$), и при несовпадении гармоник y сигнала выборки со спектром входного сигнала участки спектра сигнала на выходе идеального устройства выборки не перекрываются (рисунок 2.4 е, и). В обрат-

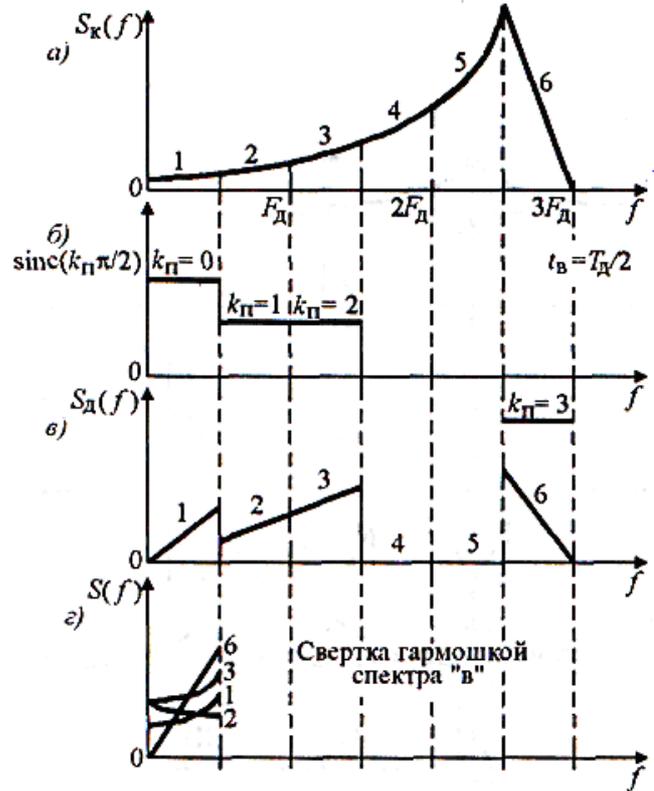


Рисунок 2.5 - Последовательность получения приведенного спектра широкополосного радиочастотного сигнала

ном случае возникают наложения участков спектров друг на друга (рисунок 2.4 ж, з).

При цифровой обработке сигналов важнейшим является участок спектра, находящийся в низкочастотном интервале от 0 до $0,5 \omega_D$. Указанный участок спектра полностью описывает спектр сигнала $x_e(t)$, поэтому ему дано особое название — *приведенный спектр*. Как следует из (2.2), приведенный спектр есть сумма участков исходного спектра $S(\omega)$ шириной $0,5 \omega_D$, умноженных на весовую функцию $\text{sinc}(k_n U_0 t_e F_D)$ и перенесенных в низкочастотный интервал. Здесь k_n - это номер гармоники сигнала выборки, которая осуществляет перенос участка спектра $S(\omega)$ шириной $0,5 \omega_D$ в низкочастотный интервал. Это число определяется из условия $|\omega - k_n \omega_0| \leq 0,5 \omega_0$ откуда получим формулу для определения целого числа k_n :

$$\left(\frac{\omega}{\omega_0} - 0,5 \right) \leq k_n < \left(\frac{\omega}{\omega_0} + 0,5 \right) \quad (2.4)$$

На рисунке 2.5 показана последовательность получения приведенного спектра широкополосного радиочастотного сигнала: исходный спектр (а), весовая функция $\text{sinc}(k_n U_0 t_e F_D)$ при длительности выборки $t_e = 0,5 T_D$ (б), помноженный на весовую функцию исходный спектр (в) и приведенный спектр (г) с указанием номеров участков исходного спектра. Спектр (з) получается в результате свертки «гармошкой» спектра (в), пунктиром показаны линии перегиба рисунка при свертке.

При несовпадении гармоник $k_n \omega_0$ с исходным спектром $S(\omega)$ и выполнении условия $\Pi < 2F_0$ приведенный спектр с точностью до множителя $\text{sinc}(k_n U_0 t_e F_D)$ соответствует исходному спектру без инверсии или с инверсией. В противном случае имеет место наложение участков спектра $S(\omega)$ друг на друга в приведенном спектре (рисунок 2.5, г).

В зависимости от режима работы устройства хранения в УВХ они подразделяются на следящие и интегрирующие. В следящем УВХ устройство хранения безынерционно к сигналу $x_e(t)$ на его входе во время выборки t_e и абсолютно инерционно во время хранения t_{xp} . Иначе говоря, в этом режиме сигнал на выходе устройства хранения (X_p) повторяет входной сигнал $x_e(t)$ во время выборки t_e и остается неизменным в течение времени хранения t_{xp} . Уровень этого неизменного сигнала равен значению сигнала $x_e(t)$ в момент окончания выборки.

В интегрирующем УВХ устройство хранения интегрирует входной сиг-

нал за время выборки t_g и хранит без изменения результат интегрирования в течение времени t_{xp} . Перед очередным импульсом выборки устройство хранения обнуляется. Так как момент начала выборки t произвольный по отношению к сигналу $x(t)$, то сигнал на выходе интегрирующего УВХ в конце выборки

$$x_{ув}(t) = \frac{1}{t_g} \int_{-0.5t_g}^{0.5t_g} x(t-\tau) d\tau$$

При гармоническом сигнале $x(t) = A \cos(\omega t)$, тогда

$$x_{ув}(t) = \text{sinc}(0.5 \omega t_g) x(t) \quad (2.5)$$

Из (2.5) следует, что в интегрирующем УВХ возникают амплитудно-частотные искажения в преобразуемом сигнале. Но у интегрирующего УВХ имеются преимущества перед следящим УВХ: выше частотный диапазон, лучшее использование энергии сигнала, дополнительные фильтрующие свойства, шире динамический диапазон.

С выхода УВХ дискретизированный сигнал $x_{ув}(t)$ поступает на вход АЦП (рисунок 2.2). В АЦП формируются числа в двоичном коде, соответствующие знаку и уровню поступающих с выхода УВХ выборок. Поскольку разрядность чисел на выходе АЦП ограничена, преобразование непрерывных по уровню выборок в числа происходит с округлением или усечением до ближайшего разрешенного разрядной сеткой числа. Таким образом, в АЦП происходит квантование уровня выборок. Если АЦП имеет на выходе p двоичных разрядов, то он может преобразовать в числа $S = 2^p$ уровней входного сигнала.

Процесс преобразования в АЦП состоит из трех этапов: квантование, преобразование в код и занесение кода в регистр. Процесс квантования эквивалентен прохождению сигнала через нелинейный безынерционный четырехполосник. Далее квантованный по уровню сигнал поступает на преобразователь ПР уровней в коды. Сформированные на выходе ПР числа записываются по фронту сигнала выборки в регистр РГ и хранятся в нем в течение времени T_A до следующего фронта сигнала выборки.

В реальных АЦП процессы квантования и преобразования в код обычно взаимосвязаны. Однако для анализа происходящих в АЦП процессов их удобно разделить, как это показано на рисунке 2.2. При этом подчеркнем тот факт, что преобразование квантованного по уровню сигнала в код — операция взаимнооднозначная, где каждому квантованному уровню сигнала ставится в соответствие определенное двоичное число, которое может быть однозначно преобразовано в этот же уровень сигнала. Таким образом, преобразователь уровня в код хотя и изменяет форму представления квантованного по уровню сигнала (вместо уровня — комбинация единиц и нулей), но не изменяет его информационную сторону, отражаемую во временном или спектральном опи-

сании сигнала. Это справедливо при условии, что время хранения t_{xp} выборки из сигнала в УВХ больше времени преобразования t_{ac} уровня сигнала в код, поэтому всегда должно выполняться условие:

$$t_{xp} > t_{ac}.$$

Регистр РГ в АЦП играет роль устройства выборки и хранения на время T_D дискретизированного и квантованного сигнала, представленного в двоичном коде. Регистр выполняется на тактируемых Д-триггерах, запись числа в РГ происходит по фронту синхроимпульса, в качестве которого предназначен сигнал выборки $W(t)$. Запоминание числа в регистре происходит в определенный момент времени и не зависит от длительности сигнала выборки t_B в УВХ. Следовательно, в регистре реализуется идеальная выборка из сигнала в виде числа и идеальное хранение его без изменения в течение времени T_D .

Из-за эффектов усечения в АЦП возникает дополнительный шум. При широкополосном входном воздействии спектр этого шума обелается, т.е. делается равномерным. Этот шум получил название *шума квантования*. Дисперсия этого шума зависит от шага квантования h_y по оси y и определяется из известного выражения

$$D_{шк} = \frac{h_y^2}{12} \quad \text{где} \quad h_y = \frac{2}{2^p - 1} \quad (2.6)$$

На выходе АЦП приведенный спектр шумов квантования

$$S_{шк} = \frac{D_{шк}}{0,5F_D} = \frac{h_y^2}{6F_D} = \frac{1}{3F_D(2^p - 1)^2} \quad (2.7)$$

Из этого выражения видно, что $S_{шк}$ уменьшается как с ростом числа разрядов p АЦП, так и с ростом частоты дискретизации F_D . Таким образом, приведенный спектр на выходе АЦП есть сумма приведенных спектра дискретизированного сигнала на выходе УВХ и спектра шумов квантования. В ЦВУ происходит обработка чисел с выхода АЦП, полученный в результате обработки сигнал подается на вход ЦАП. Цифро-аналоговый преобразователь за время t_{ca} формирует из поступающих на его вход чисел $y(n)$ аналоговый сигнал, уровень и знак которого однозначно соответствует знаку и модулю этих чисел.

Таким образом ЦАП, как и УВХ, осуществляет свертку двух функций, одна из которых дискретизированный и квантованный сигнал, представленный регулярной последовательностью чисел $y(n)$, а другая - импульсная характеристика ЦАП $u(t)$. Следовательно, приведенный спектр на выходе ЦАП есть произведение приведенного спектра на его входе и амплитудно-частотной характеристики ЦАП.

В зависимости от длительности формируемой ЦАП импульсной харак-

теристики $u(t)$ они называются интерполяторами нулевого, первого и более высокого порядков. Цифро-аналоговый преобразователь с импульсной характеристикой в виде прямоугольного импульса единичной амплитуды и длительности $\tau = t_{xp1} \leq T_D$ называется *интерполятором нулевого порядка*. Цифро-аналоговые преобразователи с импульсными характеристиками длительностью $T_D < \tau \leq 2T_D$ называют интерполяторами первого порядка и т.д. Форма импульсной характеристики может быть различной: треугольной, косинусоидальной, вида $\frac{\sin x}{x}$ трапецеидальной и т.д.

2.1.3 Импульсные радиосигналы с различными видами модуляции. Каналы связи с импульсной модуляцией

С помощью дискретизатора непрерывное сообщение $\lambda(t)$ преобразуется в импульсную последовательность, один из параметров которой (амплитуда, частота, положение) меняется по закону передаваемого сообщения.

В этой связи различают системы связи с АИМ, ШИМ, ВИМ. (рисунок 2.6)

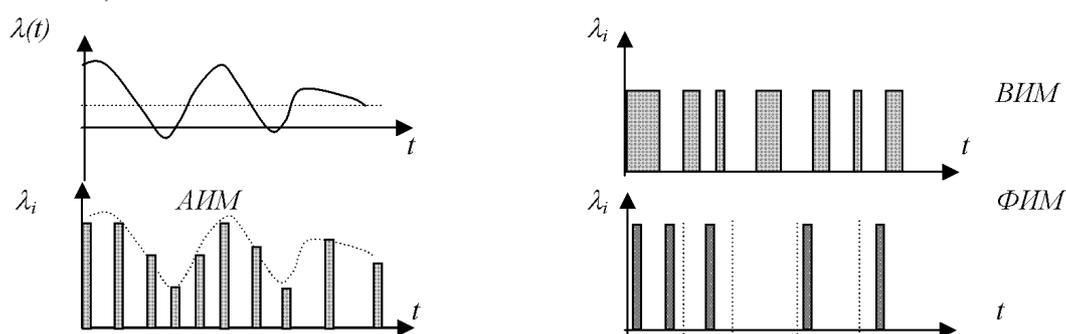


Рисунок 2.6 – Каналы связи с импульсной модуляцией

Структурная схема импульсной системы связи представлена на рисунке 2.7.

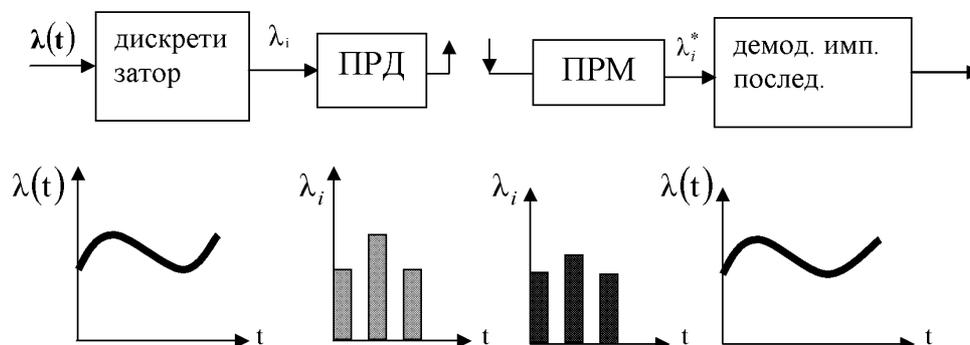


Рисунок 2.7 – Каналы связи с импульсной модуляцией

В зависимости от вида модулируемого параметра ВЧ- колебания различают:

- АИМ/АМ,
- ШИМ/АМ,
- АИМ/ФМ и т.д.

На передающей стороне импульсной системы связи производится комбинированная модуляция радиосигналов, например, амплитудно-импульсная, модуляция видеоимпульсов сообщением и амплитудная модуляция гармонического колебания видеоимпульсами, что сокращенно обозначается как АИМ–АМ. Находят применение и другие варианты комбинированной модуляции. На приемной стороне системы связи может последовательно осуществляться демодуляция гармонического колебания в приемнике, в результате чего получаются (модулированные видеоимпульсы, и демодуляция этих импульсов). В оптимальном варианте эти два преобразования должны осуществляться совместно. Работой приемника и импульсного демодулятора управляет импульсный тактовый генератор, синхронизированный с передающим генератором.

Различают амплитудно-импульсную модуляцию первого и второго рода (АИМ-1 и АИМ-2). Сигнал при АИМ–АМ–1 можно записать в виде:

$$s(t) = \sum_{v=0}^{\infty} \text{rect}(t - v\Delta) A_0 \left(1 + m \frac{\lambda(t)}{\sigma_\lambda} \right) \cos(\omega_0 t + \varphi_0), \quad (2.8)$$

где $\text{rect}(t - v\Delta)$ – прямоугольный импульс с единичной амплитудой и длительностью t_u , расположенный на интервале времени

$$\left(v\Delta - \frac{t_u}{2}, v\Delta + \frac{t_u}{2} \right); m = \frac{\sigma_a}{A_0} \text{ – коэффициент амплитудной модуляции.}$$

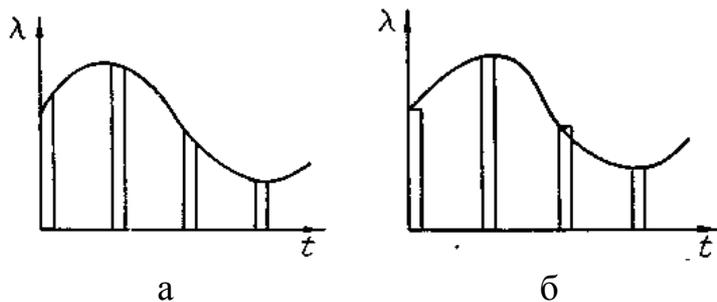


Рисунок 2.8 – Формирование сигналов при АИМ первого (а) и второго (б) рода

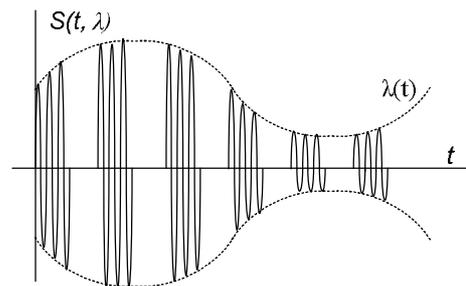


Рисунок 2.9 – Радиосигнал при АИМ-АМ-1

Формирование сигналов АИМ-1 поясняет рисунок 2.8 а. Примерный вид радиосигнала (2.8) показан на рисунке 2.9. Сигнал во время действия импульса изменяется по закону сообщения, т. е. не является прямоугольным. При АИМ-2 (рисунок 2.9 б) используются прямоугольные импульсы и их ампли-

тудные значения равны соответствующим отсчетам сообщения. На практике чаще используется АИМ-1, при которой модуляция импульсов осуществляется с помощью простых ключевых схем.

Могут также применяться каналы связи с АИМ–ЧМ и АИМ–ФМ. В случае АИМ–ЧМ сигнал представляет собой радиоимпульс с постоянной амплитудой, частота колебания которого пропорциональна сообщению, а при АИМ–ФМ – пропорциональна сообщению фаза колебания. Ввиду более простой реализации каналы связи с АИМ–АМ применяются чаще других и поэтому в дальнейшем рассматривается только этот вид амплитудно-импульсной модуляции.

Рассмотрим системы радиосвязи с АИМ/АМ, радиосигналы которых могут быть записаны в виде:

$$S(t, \lambda) = A_0(1 + m\lambda^*(t)) u(t) \cos(\omega t + \varphi(t)), \quad (2.9)$$

где A_0 – среднее значение амплитуды импульсов;

$u(t)$ – функция, описывающая последовательность импульсов с единичной амплитудой (будем считать, что они прямоугольные)

$m = \sigma_a/A_0$ – коэффициент амплитудной модуляции;

$\lambda^*(t) = \lambda(t)/\sigma_\lambda$ – нормированное значение информационного сообщения.

Спектр сигналов АИМ содержит:

- составляющие спектра сообщения,
- составляющие, определяемые параметрами последовательностью импульсных сигналов (рисунок 2.10).

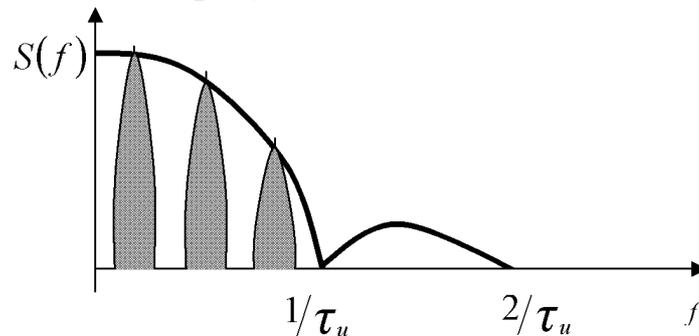


Рисунок 2.10 - Спектр сигналов АИМ

Если частота дискретизации достаточно высока, т.е. выполняется условие $T \ll \tau_\lambda$, то синтез оптимального приемника для АИМ сигнала совпадает со случаем непрерывного сигнала. Единственным отличием является дополнительная операция умножения $\zeta(t)$ на функцию $u(t)$ в приемнике, структурная схема которого при известной начальной фазе имеет вид представленный на рисунке 2.11.

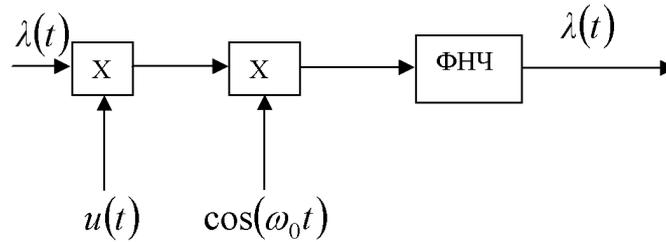


Рисунок 2.11 - Структурная схема оптимального приемника для АИМ сигнала

При случайной начальной фазе источником синхронного напряжения должна быть схема ФАП. Строблирующее напряжение $u(t)$ получается от специальной системы тактовой синхронизации.

Неоптимальная схема приемника АИМ-АМ имеет вид представленный на рисунке 2.12. В приведенной схеме используется детектор огибающей, что существенно упрощает приемное устройство.

АИМ среди различных импульсных систем является наиболее простой, но при ней не полностью используется мощность передатчика.

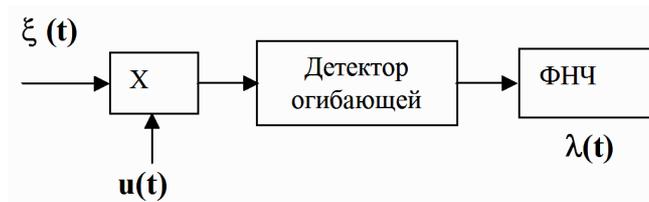


Рисунок 2.12 – Схема неоптимального приемника АИМ-АМ

Схема оптимального приемника АИМ-АМ изображена на рисунке 2.13. Отметим, что помехоустойчивость оптимального приемника АИМ-АМ такая же, как и при приеме АМ – радиосигналов при одинаковой средней мощности.

На входе приемника АИМ-АМ смесь сигнала с шумом умножается на опорное гармоническое колебание, получаемое с помощью схемы ФАП. В результате этого происходит синхронное, детектирование радиосигнала. Далее полученный результат умножается на тактовые импульсы (ТИ), вырабатываемые устройством тактовой синхронизации. Эти импульсы совпадают по времени с импульсами сигнала и поэтому в результате умножения осуществляется стробирование последних. Оценка сообщения получается на выходе линейного фильтра, представляющего собой интегрирующую цепочку с постоянной времени α^{-1} .

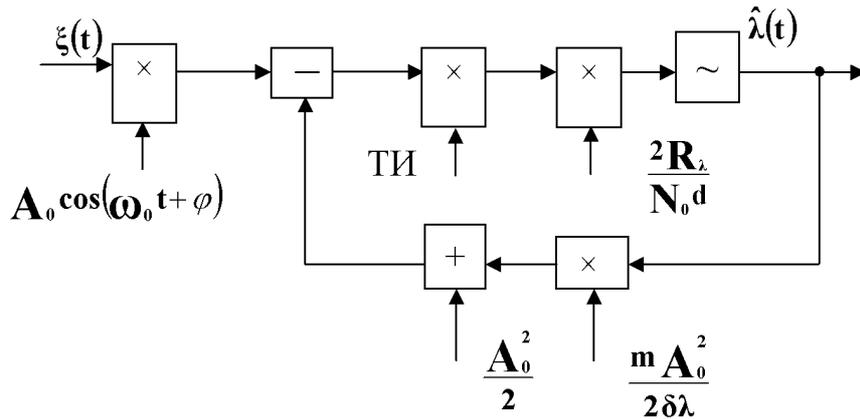


Рисунок 2.13 – Схема оптимального приемника системы связи с АИМ-АМ-1

2.2 Цифровая передача непрерывных сообщений

Достаточно широкое распространение в настоящее время получили импульсные системы радиосвязи. Эти системы связи, в частности, позволяют осуществить временное уплотнение каналов.

В импульсных системах связи аналоговое сообщение передается своим выборочным значением, определяемым моментами взятия отсчета. На практике нашли широкое распространение системы, основанные на преобразовании аналогового сообщения в цифровую форму. Рассмотрим последовательность преобразования аналоговых сигналов в цифровую форму.

2.2.1 Квантование сообщений по уровням

В системах связи с ИКМ осуществляется дискретизация непрерывных сообщений по времени и квантование по уровню. При этом весь диапазон возможных значений сообщения разбивается на конечное число уровней. По каналу связи вместо истинного значения $\lambda(t_i)$ передается ближайшее к нему значение уровня $\lambda^*(t_i)$.

Представление случайного процесса некоторыми дискретными уровнями называется квантованием процесса по уровню. Если по переданным уровням восстановить сообщение, то даже при отсутствии помех возникает ошибка квантования $\lambda(t) - \lambda^*(t)$ (шумы квантования) проиллюстрированная на рисунке 2.14, где каждому уровню ставится в соответствие запись его значения в двоичном коде

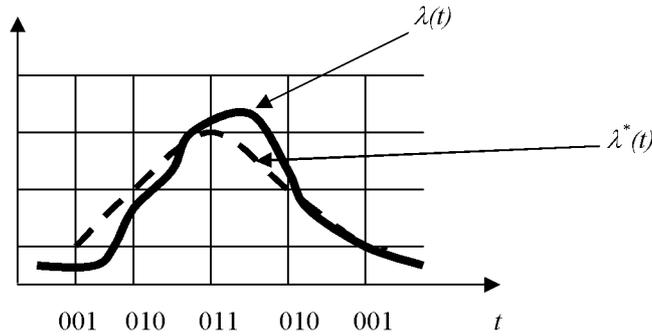


Рисунок 2.14

2.2.2 Каналы связи с импульсно-кодовой модуляцией и дельта-модуляцией

Дельта-модуляцией (ДМ) называется дифференциальная импульсно-кодовая модуляция (ДИКМ) при квантовании сообщения на два уровня. Помимо преимуществ в точности передачи сообщения, о котором говорилось выше, эта система характеризуется простотой технической реализации; квантователь на два уровня представляет собой обычное пороговое устройство. Так как каждый отсчет сообщения преобразуется в один бит, в такой системе не нужна синхронизация по кодовым комбинациям. Укажем, что систему связи с ДМ предложил в 1949 г. советский инженер Л. А. Коробков. За рубежом эта система была описана только в 1952 г.

Кодер и декодер системы связи с ДМ могут быть цифровыми или аналоговыми. Так же могут быть реализованы устройства системы связи с ДМ. Ввиду простоты технической реализации рассмотрим аналоговый вариант системы (рисунок 2.15).

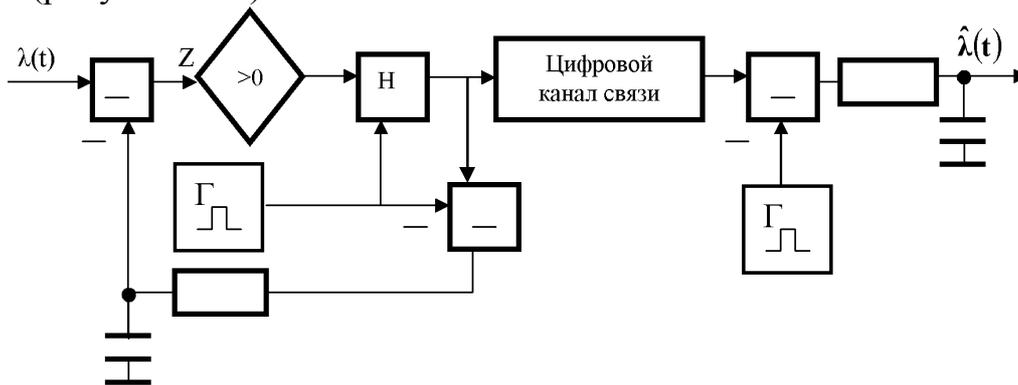


Рисунок 2.15 – Схема системы связи с ДМ

В кодере приведенной структурной схемы из сообщения $\lambda(t)$ вычитается его оценочное значение $\lambda^*(t)$. Далее, в компараторе с нулевым порогом срабатывания осуществляется квантование на два уровня. Полученный процесс совместно с импульсами тактового генератора поступает на схему «И», где осу-

ществляется дискретизация по времени и образуются величины X_v , поступающие в канал связи и цепь обратной связи кодера. В цепи обратной связи кодера из двоичных однополярных импульсов X_v , принимающих значения нуль или $2\sigma_z\sqrt{\frac{\pi}{2}\frac{1}{\alpha t_u}}$ вычитаются периодические импульсы тактового генератора с амплитудами $\sigma_z\sqrt{\frac{\pi}{2}\frac{1}{\alpha t_u}}$. В результате этого образуется последовательность двухполярных импульсов с амплитудами $\frac{\hat{Z}_y}{a \cdot t_u} = \pm \sigma_z\sqrt{\frac{\pi}{2}} \cdot a \cdot t_u$, воздействующая на интегрирующую цепочку. В декодере производятся преобразования сигналов аналогичные описанным.

Рассмотренная система связи с ДМ синтезирована для гауссовского марковского сообщения, формируемого из белого шума с помощью интегрирующей цепочки. Такие же интегрирующие цепочки используются в полученных схемах для формирования оценки сообщения. На практике могут встретиться другие сообщения, однако во всех случаях оценивающий фильтр должен быть подобен формирующему. Например, спектральная плотность речи максимальна на частоте 500 Гц, и ширина спектра на уровне 0,5 составляет около 300 Гц. Формирующий фильтр речи представляет собой колебательный контур с указанными выше параметрами. В связи с этим в системе телефонной связи с ДМ оценивающими фильтрами вместо интегрирующих цепочек должны быть колебательные контуры.

Речевой процесс является нестационарным: дисперсии глухих и звонких звуков различаются примерно в 200 раз. В связи с этим для передачи речи целесообразно использовать системы с ДМ, адаптивные к изменениям громкости звуков. В таких системах амплитуда импульсов, действующих на фильтры, должна меняться пропорционально среднеквадратическому отклонению речевого процесса.

Разборчивость речи в системе телефонной связи с ДМ зависит от скорости передачи данных в цифровом канале связи. При скорости передачи $V = 9600$ бит/с она составляет около 90% и с увеличением скорости улучшается.

ДМ можно использовать для передачи изображений в телевидении и системах факсимильной связи. Экспериментальные исследования показали, что в такой системе искажения изображения не заметны, если $V > (10—15)f_g$, где f_g — высшая частота спектра видеосигнала.

Достоинствами ДМ являются простота аппаратуры, а также более низкие (по сравнению с ИКМ) требования к скорости передачи данных в канале

связи. Недостатком ДМ является трудность сочленения системы связи с ЦВМ, обусловленная тем, что ЦВМ оперирует с цифрами, непосредственно отображающими значения сообщения, а при ДМ связь сообщения с цифровыми сигналами более сложная.

Таким образом, получение ИКМ включает в себя три операции:

- дискретизация сообщения по времени;
- квантование отсчетов по уровням;
- кодирование уровней квантования.

Система с ИКМ представлена на рисунке 2.16.

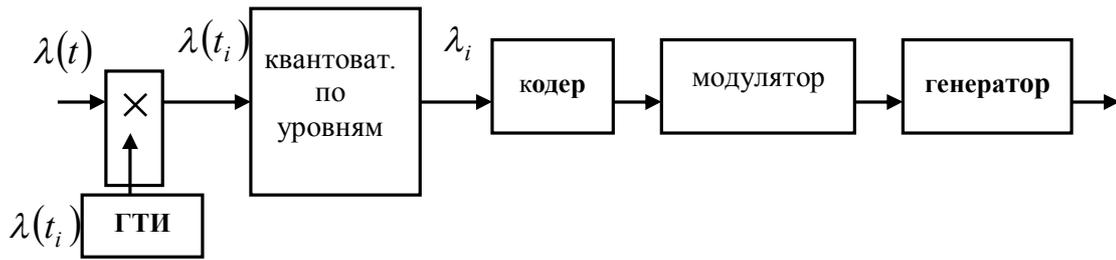


Рисунок 2.16 – Структурная схема системы с ИКМ

2.2.3 Методы получения сигналов ИКМ

Рассмотрим два вида кодеров:

- с преобразованием кодируемой величины во временной интервал с последующим его измерением при помощи счетчиков импульсов;
- поразрядные (взвешивающие) кодеры.

В схеме первого типа кодируемые ИКМ сигналы поступают на вход 1 (рисунок 2.17) и преобразуются в сигналы ШИМ, т.е. длительность импульсов на выходе преобразователя пропорциональна $\lambda(t_i)$.

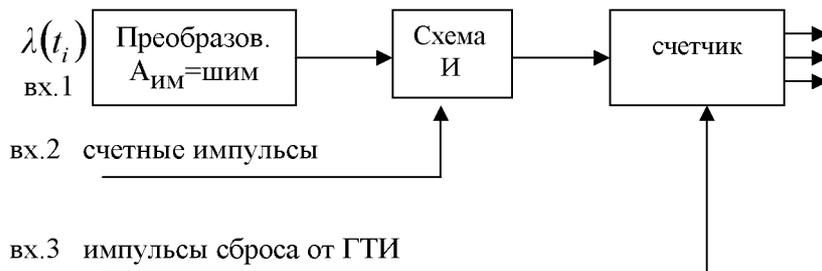


Рисунок 2.17 - Структурная схема кодера с преобразованием кодируемой величины

Далее следует «схема И» на второй вход которой поступают короткие импульсы с частотой следования во много раз превышающей частоту отсчетов. На выходе схемы И_получаются пачки импульсов. Число импульсов в пачке пропорционально значению $\lambda(t_i)$. Далее следует триггерный счетчик импуль-

сов, показания которого считываются в двоичном коде.

Требуемая скорость счета кодера определяется соотношением

$$f_{сч} = \frac{2^M - 1}{T - \tau_3}, \quad (2.10)$$

где τ_3 – защитный временной интервал,
 M – число разрядов двоичного кода,
 T – период отчетов.

Во взвешивающем кодере (рисунок 2.18) происходит последовательное сравнение $\lambda(t_i)$ с порогом $h_1 = L_{max}/2$, потом сигнала $(\lambda(t_i) - h_1)$ с порогом $h_2 = L_{max}/4$ и т.д.

Число уровней квантования есть целая степень числа 2, т.е. $N = 2^M$.

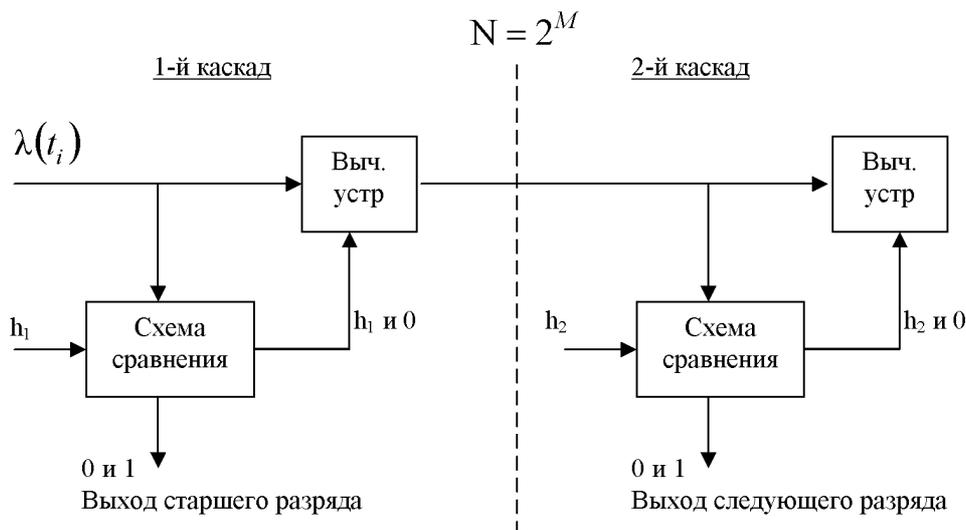


Рисунок 2.18 - Структурная схема поразрядного кодера

Если $\lambda(t_i)$ больше h_1 , то на первом выходе схемы сравнения появляется сигнал равный h_1 , а на втором выходе – кодовая 1. Следующие каскады работают аналогично.

2.2.4 Принципы построения декодеров

При декодировании наибольшее распространение получили декодеры параллельного действия (рисунок 2.19).

При дельта-модуляции по каналу связи передается знак приращения исходного сообщения за тактовый период. Оценка сообщения формируется как сумма полученных приращений.

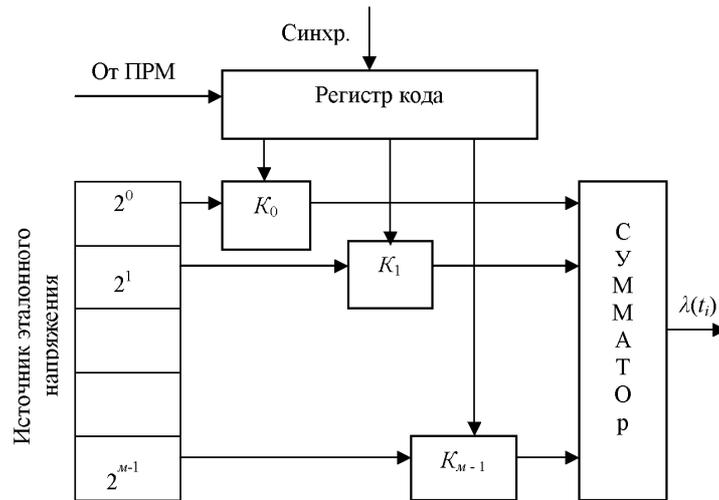


Рисунок 2.19 - Структурная схема декодера параллельного действия

Упрощенная структурная схема системы связи с дельта-модуляцией представлена на рисунке 2.20.

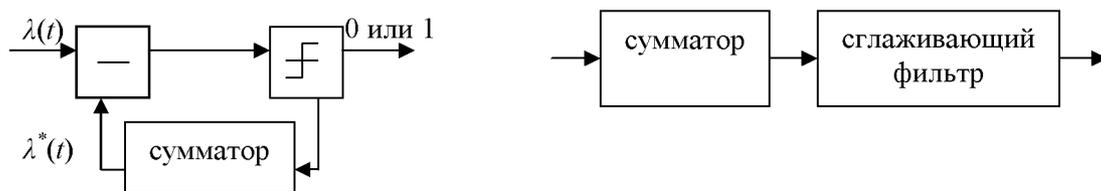


Рисунок 2.20 - Упрощенная структурная схема системы связи с дельта-модуляцией

Положительному приращению сигнала т.е. положительной разности $\lambda(t) - \lambda^*(t)$ соответствует посылка по каналу связи положительного импульса, а отрицательной – отрицательного импульса (или нуля). Процесс формирования сигналов ДМ представлено на рисунке 2.21.

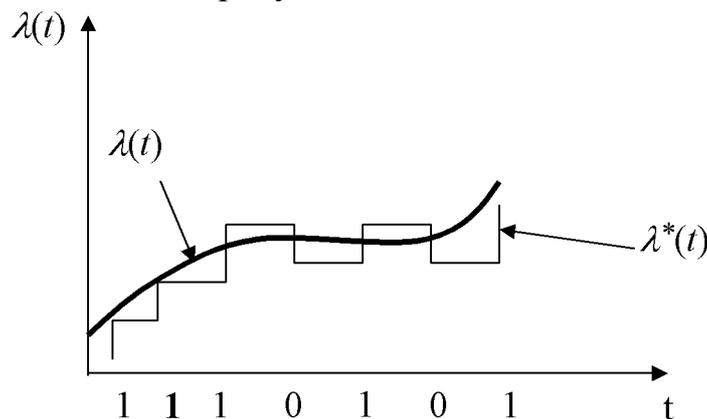


Рисунок 2.21 - Процесс формирования сигналов ДМ

ДМ сопровождается искажениями двух типов:

- Происходит запаздывание процесса $\lambda^*(t)$ по отношению к исходному $\lambda(t)$. Эти искажения называются шумами перегрузки.
- На интервалах, где $\lambda(t)$ почти не меняется, проявляются шумы квантования $\varepsilon(t) = \lambda(t) - \lambda^*(t)$.

Для уменьшения шумов перегрузки полезно увеличить шаг квантования h , однако при этом возрастают шумы квантования.

Системы связи с ДМ обладают всеми положительными качествами КИМ и для ряда случаев, когда не требуется высокая точность, например при передаче речи, является более предпочтительней из-за своей простоты.

Таким образом, преимущества системы связи с ИКМ следующие:

1. возможность регенерации сигналов даже при самых сильных искажениях, т.е. в конечном счете, высокая помехоустойчивость;
2. широкая возможность использования цифровой техники;
3. возможность обработки информации на ЦВМ.

Системы ИКМ перспективны, в частности в лазерных системах связи. Системы связи с ДМ обладают всеми положительными качествами КИМ и для ряда случаев, когда не требуется высокая точность, например при передаче речи, является более предпочтительней из-за своей простоты.

ГЛАВА 3 ПЕРЕДАЧА ДИСКРЕТНЫХ СООБЩЕНИЙ

3.1 Уравнения дискретно-непрерывной фильтрации сигналов со случайно изменяющимися параметрами

Одним из важных требований, предъявляемых к системам радиосвязи, является *помехоустойчивость (ПУ)* – способность системы связи выполнять прием/передачу сообщений в условиях действия помех.

Повышение ПУ может быть достигнуто за счет оптимизации устройств приема и обработки сообщений. Методика оптимального приема сообщений основана на **теории фильтрации**. В упрощенном виде задача фильтрации может быть сформулирована следующим образом. На передающей стороне формируется радиосигнал $S(t, \lambda(t))$, где $\lambda(t)$ – информационное сообщение в непрерывном времени. Считается, что вероятностное описание сигнала $S(t, \lambda(t))$ известно, так как известен способ формирования сигнала на передающей стороне.

3.1.1 Постановка задачи синтеза оптимального различителя сигналов на основе теории статистических решений

В любой радиотехнической системе сигнал при передаче искажается и сообщение на выходе приемника воспроизводится с некоторой ошибкой. Искажения сигнала обусловлены как случайными изменениями характеристик каналов, так и помехами, действующими в нем. *Частотные и временные характеристики канала определяют линейные искажения*. Кроме того, канал может вносить *нелинейные искажения из-за нелинейности переходных и передаточных функций тех или иных звеньев приемника*. Если линейные и нелинейные искажения сигнала обусловлены известными характеристиками канала, то они могут быть устранены соответствующей коррекцией. Искажения от помех носят случайный характер и поэтому не могут быть полностью устранены. В правильно спроектированной системе искажения сообщений обусловлены лишь воздействием помех.

Пусть приемник в отсутствие помех точно воспроизводит сообщение, содержащееся в полезном сигнале. При наличии помех сообщение будет воспроизводиться с некоторой погрешностью. Приемник, обеспечивающий минимальные искажения сообщения, называется оптимальным или идеальным. Критерии оптимальности и количественные характеристики искажений зависят от назначения приемника. *Минимальный уровень искажений при выбранном критерии характеризует потенциальную помехоустойчивость*. При

заданных условиях приема помехоустойчивость реального приемника не может превзойти потенциальную.

Вопросами, связанными с синтезом оптимальных устройств обработки сигналов, занимается **теория оптимальных методов приема**. При некоторых ограничениях на характеристики сигналов и помех она позволяет найти оптимальный алгоритм работы радиоприемного устройства и, соответственно, его структуру, оценить количественно показатели синтезируемого устройства, определить наилучшие виды передаваемых сигналов, а также выяснить степень технического совершенства реальных приемников и возможные пути повышения их помехоустойчивости.

Основная задача теории оптимального приема сигналов формулируется так: предполагая заранее (априори) известными некоторые характеристики передаваемого сигнала, канала и помех, а также их функциональное взаимодействие, необходимо синтезировать оптимальное приемное устройство, которое бы наилучшим образом (в смысле выбранного критерия) воспроизводило сообщение, содержащееся в сигнале на входе приемника, или принимало решение с наименьшими ошибками.

При синтезе оптимальных приемных устройств важными являются: выбор математически продуктивного критерия оптимальности в соответствии с физическим смыслом и целевым содержанием решаемой задачи и четкая математическая формулировка задачи, включающая все априорные сведения и позволяющая решать ее в соответствии с выбранным критерием.

В результате синтеза должны быть найдены оптимальный алгоритм работы приемного устройства и, соответственно, структурная или функциональная схема приемника, определены параметры схемы, вычислены количественные показатели качества работы приемника. Кроме того, при этом важным является выяснение степени чувствительности полученных алгоритмов и количественных характеристик к отклонениям от принятых априорных данных и решение вопроса о возможности технической реализации синтезируемых алгоритмов.

Решение основных задач теории оптимальных методов радиоприема базируется на хорошо разработанных методах математической статистики.

3.1.2 Основные положения теории статистических решений

Оптимальные критерии

Как уже упоминалось выше, при синтезе оптимальных устройств обработки сигналов одним из исходных положений является выбор критерия качества работы устройства. Учитывая, что теория статистических решений охватывает все многообразие статистических оптимальных критериев, целесооб-

разно при решении задач синтеза оптимальных радиоприемных устройств воспользоваться результатами этой теории. Ниже излагаются основные положения теории статистических решений. Пусть принятый сигнал:

$$u(t) = s(t, \lambda) * n(t) \quad (3.1)$$

где * - символ, характеризующий способ взаимодействия полезного сигнала $s(t, \lambda)$ и помехи $n(t)$. Без нарушения общности положим, что наблюдение колебания $u(t)$ производится в дискретные моменты времени t_1, \dots, t_M . Тогда с учетом формулы 3.1 можно записать $u = s * n$ где $u = (u_1, \dots, u_M)$, $s = (s_1, \dots, s_M)$, $n = (n_1, \dots, n_M)$ - M- мерные векторы (M-мерные выборки) соответственно принятого колебания, полезного сигнала и помехи.

Совокупность всех возможных векторов и образует пространство U выборок принимаемого колебания $u(t)$. Аналогично векторы s и n образуют пространство S выборок полезного сигнала $s(t)$ и пространство N выборок помехи $n(t)$.

Статистические характеристики помехи предполагаются известными и задаются в виде распределения $w(n) = w_M(n_1, \dots, n_M)$. Известным считается также способ взаимодействия сигнала с помехой. При этом можно найти статистическое описание вектора и принятого колебания для фиксированного вектора s - условное распределение $w(u | s) = w_M(u_1, \dots, u_M | s)$.

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

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

$$w(u, s) = w(s)w(u | s) \quad (3.2)$$

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

Выбор решения по принятой выборке u_1, \dots, u_M производится в соответствии с *алгоритмом работы решающего устройства* $\Delta(\gamma | u)$. Функция $\Delta(\gamma | u)$ называется **решающей функцией** или **решающим правилом**. Она представляет собой вероятности (или плотность вероятности) принятия реше-

ния γ на основе принятых данных u . Механизм принятия решения можно представить следующим образом. Пространство принимаемых сигналов U разбивается на непересекающиеся области U_0, U_1, \dots и устанавливается соответствие между решениями и этими областями. Если каждой области становится в соответствие только одно вполне определенное решение, то правило $\Delta(\gamma | u)$ называется **нерандомизированным (детерминированным)**.

Существуют правила, при которых для принятой выборки u_1, \dots, u_M допускается несколько решений с соответствующими вероятностями, например, некоторой области u_i ставятся в соответствие решения $\gamma_{i1}, \dots, \gamma_{ik}$ соответственно с вероятностями p_{i1}, \dots, p_{ik} , причем $\sum_{j=1}^k p_{ij} = 1$. Такие правила называются **рандомизированными**.

В дальнейшем рассматриваются только нерандомизированные правила. При этом будем считать, что если выборка принятого колебания попадает в область u_i , то принимается решение γ_i .

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

Очевидно, что при любом решающем правиле при наличии помех всегда возможны ошибочные решения. Для количественной оценки ущерба (потерь), связанного с принятием решений, вводится так называемая **функция потерь (штрафа, стоимости)** $\Pi(s, \gamma)$. Ее конкретное значение $\Pi(s_i, \gamma_j)$ характеризует потери при принятии решения γ_j в то время как правильным является решение γ_j . Функция потерь должна удовлетворять следующим свойствам:

$$\begin{cases} \Pi(s_i, \gamma_j) > 0; \\ \Pi(s_i, \gamma_j) > \Pi(s_j, \gamma_j); \\ \Pi(s_j, \gamma_j) \leq 0 \end{cases}$$

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

Так как появление того или иного полезного сигнала s на входе приемника и принятие решения γ являются случайными событиями, то значение функции потерь $\Pi(s, \gamma)$ является случайной величиной. Поэтому качество решения можно характеризовать математическим ожиданием функции потерь:

$$R = M \{ \Pi(s, \gamma) \} = \sum_j \sum_i w(s_i, \gamma_j) \Pi(s_i, \gamma_j) \quad (3.3)$$

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

Величина R характеризует средние потери при принятии решения и называется средним риском. Чем меньше средний риск, тем лучше решение. *Наилучшим (оптимальным) решающим правилом будет такое, для которого значение среднего риска будет наименьшим. Правило, при котором минимизируется средний риск, называется байесовским правилом или байесовским критерием. Часто его также называют критерием минимума среднего риска.*

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

$$R = \iint_{S \Gamma} \Pi(s, \gamma) w(s, \gamma) d\gamma ds \quad (3.4)$$

где $w(s, \gamma)$ - совместная плотность вероятности сигнала S и решения γ . Для нерандомизированных правил принятое колебание и решение связаны детерминированной зависимостью, а поэтому справедливо соотношение

$$w(s, \gamma) ds d\gamma = w(s, u) ds du \quad (3.5)$$

С учетом формулы (3.5) выражение (3.4) записывается в виде

$$R = \iint_{S U} \Pi(s, \gamma_u) w(s, u) du ds ,$$

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

Заметим, что средний риск не является исчерпывающей характеристикой решающего правила. Более полной характеристикой качества обработки может служить, например, совокупность двух показателей - **среднего риска** R и **дисперсии среднего риска** D_R (в общем случае - совокупность моментов функции потерь $\Pi(s, \gamma)$). Однако синтез решающего правила по совокупности показателей значительно сложнее и на практике ограничивается критерием минимума среднего риска.

Применение байесовского критерия требует большого объема априорной информации: необходимо знать функцию потерь и совместное распределение $w(s, u)$ или, что то же самое, распределения $w(s)$ и $w(u|s)$. Если априорное распределение сигналов $w(s)$ неизвестно, то применять критерий Байеса нельзя, так как при этом задача оптимизации, в смысле минимума среднего риска, оказывается неопределенной.

В этом случае применяют **минимаксный критерий**. Введем понятие условного среднего риска:

$$r_{s_i} = \sum_j \Pi(s_i, \gamma_j) w(\gamma_j | s_i)$$

для дискретных пространств S и Γ и

$$r_s = \int_{\Gamma} \Pi(s, \gamma) w(\gamma | s) d\gamma = \int_U \Pi(s, \gamma_u) w(u | s) du$$

для непрерывных пространств S и Γ , где $w(\gamma_j | s_i)$ - вероятность принятия решения γ_j при условии, что на входе приемника присутствует полезный сигнал s_i , а $w(s)$ и $w(u | s)$ - соответственно условные плотности вероятности случайных величин γ и u .

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

Минимаксный критерий дает наилучшее решение для наихудшего случая. Он гарантирует, что минимально возможное значение условного риска не будет больше некоторого значения даже при самом неблагоприятном распределении $w(s)$. Однако минимаксное правило, предохраняя от наихудшего случая, вероятность появления которого может быть малой, оказывается слишком осторожным для наиболее часто встречающихся ситуаций. Поэтому, если при некотором неминимаксном правиле $\Delta(\gamma' | u)$ максимальный условный риск несколько больше r_s , а остальные условные риски значительно меньше тех, которые соответствуют минимаксному правилу $\Delta^*(\gamma | u)$, то целесообразно отдать предпочтение правилу $\Delta(\gamma | u)$.

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

$$\Pi(s_i, \gamma_j) = \begin{cases} \text{const} & \text{при } j \neq i; \\ 0 & \text{при } j = i. \end{cases} \quad (3.6)$$

При функции потерь, определяемой выражением 3.6, средний риск

$$R = \sum_i \sum_j w(s_i, \gamma_j) \quad (3.7)$$

Из выражения (3.7) следует, что величина R с точностью до постоянного множителя совпадает с полной вероятностью ошибочных решений. Поэтому байесовский критерий при задании функции потерь в виде выражения (3.6) есть смысл назвать **критерием минимума полной вероятности ошибки**. Его часто называют **критерием идеального наблюдателя**, а также **критерием Котельникова - Зигерта**.

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

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

в соответствии с которым решение принимается в пользу сигнала s_i , если

$$w(s_i|u) > w(s_j|u), \quad j=1, \dots, m, \quad j \neq i. \quad (3.8)$$

Действительно, все, что можно узнать о полезном сигнале s_i на основе принятого колебания, заключено в апостериорной вероятности $w(s_i|u)$, которая представляет собой не что иное, как вероятность правильного приема сигналов. Поэтому, если решение принимать в соответствии с выражением (3.8), то гарантируется, что полная вероятность ошибки будет минимальной.

При отсутствии априорных сведений не только о функции потерь, но и о распределении сигналов $w(s)$ считают, что, $w(s) = const$, т.е. принимают распределение сигналов как равномерное. При этом апостериорная вероятность $w(s|u)$ с точностью до постоянного множителя совпадает с функцией правдоподобия $w(s|u)$. Для обеспечения минимального среднего риска используют, так называемый, **критерий максимального правдоподобия** - решение принимается в пользу сигнала, для которого функция правдоподобия максимальна, т.е. в пользу сигнала s_i , если

$$w(u|s_i) > w(u|s_j), \quad j=1, \dots, m, \quad j \neq i. \quad (3.9)$$

До сих пор предполагалось, что условные плотности вероятности $w(u|s_i)$ $i=1, \dots, m$, точно известны. Однако на практике эти распределения часто зависят от некоторых неизвестных параметров полезных сигналов (начальных фаз, амплитуд, доплеровских сдвигов частоты). Если многомерное распределение $w(\lambda)$ случайного вектора, координатами которого являются случайные неизвестные параметры полезных сигналов, известно, то можно построить оптимальное по Байесу правило. Оно сводится к минимизации математического ожидания среднего риска $R(\lambda)$, зависящего от вектора λ .

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

При задании функции потерь в виде (3.6) решающие правила (3.8) и (3.9) можно записать в виде

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

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

3.2 Системы передачи с когерентной обработкой сигналов

3.2.1 Алгоритм оптимального демодулятора

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

$$u(t) = s_r(t) + n(t), \quad 0 \leq t \leq T_c \quad (3.10)$$

где T_c - длительность полезного сигнала.

Необходимо на основании анализа реализации процесса $u(t)$ решить, какой из полезных сигналов присутствует в принятом сигнале. При этом будем предполагать, что момент t_0 поступления полезного сигнала на вход приемника известен, а также известны все параметры полезных сигналов $s_r(t)$, $r=1,2,\dots,m$, в том числе и начальные фазы $\varphi_r(t)$, $r=1,2,\dots,m$, частотного заполнения полезных сигналов. Обработка сигналов, когда используется информация о начальной фазе φ , называется когерентной. В дальнейшем будем считать, что $t_0 = 0$ и $\varphi_r = 0$. Кроме того, будем полагать, что процесс $n(t)$ - стационарный гауссовский с нулевым математическим ожиданием.

Рассмотрим сначала случай, когда число полезных сигналов $m = 2$. С учетом (3.9) алгоритм работы оптимального различителя сводится к вычислению отношения функций правдоподобия

$$L^M(u) = \frac{w(u | s_1)}{w(u | s_2)}$$

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

Найдем структуру функции правдоподобия $w(u | s_r)$, $r=1,2$. С учетом (3.10) многомерная плотность $w(u | s_r)$ определяется многомерной плотностью $w(n)$ с последующей заменой n в соответствии с выражением $n = u - s_r$.

Многомерную выборку $n = (n_1, n_2, \dots, n_M)$ можно задать различными способами. Наиболее удобным является способ, когда в качестве координат n_k , $k=1, \dots, M$, вектора n используют коэффициенты разложения Карунена - Лозева случайного процесса $n(t)$:

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

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

$$\int_0^{T_c} R(t_2 - t_1) \varphi_k(t_2) dt_2 = \sigma_k^2 \varphi_k(t_1), \quad (3.11)$$

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

С учетом свойств разложения Карунена - Лоэва и закона распределение процесса $n(t)$ можно утверждать, что координаты $n_k, k = 1, \dots, M$, являются независимыми гауссовскими случайными величинами с нулевыми математическими ожиданиями и дисперсиями σ_k^2 . Соответственно M -мерная плотность вероятности определяется как

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

Заметим, что при увеличении размера выборки M коэффициенты n_k остаются статистически независимыми гауссовскими величинами и вид распределения сохраняется.

Аналогично в качестве координат векторов $u = (u_1, u_2, \dots, u_M)$ и $s_\Gamma = (s_{\Gamma 1}, s_{\Gamma 2}, \dots, s_{\Gamma M})$ возьмем коэффициенты разложения Карунена - Лоэва сигналов $u(t)$ и $s(t)$:

$$\left\{ \begin{array}{l} u(t) = \sum_{k=1}^M u_k \varphi_k(t), \\ s_\Gamma(t) = \sum_{k=1}^M s_{\Gamma k} \varphi_k(t); \\ u_k = \int_0^{T_c} u(t) \varphi_k(t) dt, \\ s_{\Gamma k} = \int_0^{T_c} s_\Gamma(t) \varphi_k(t) dt. \end{array} \right. \quad (3.12)$$

Учитывая, что

$$n = u \cdot s_\Gamma,$$

$$n_k = u_k \cdot s_{\Gamma k},$$

$$w(u | s_\Gamma) = w(n) = w(u - s_\Gamma),$$

находим выражение для функции правдоподобия

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

Соответственно отношение правдоподобия

$$L^M(u) = \frac{w(u | s_1)}{w(u | s_2)} = \frac{\exp\left[-\frac{1}{2} \sum_{k=1}^M \frac{(u_k - s_{1k})^2}{\sigma_k^2}\right]}{\exp\left[-\frac{1}{2} \sum_{k=1}^M \frac{(u_k - s_{2k})^2}{\sigma_k^2}\right]} = \frac{\exp\left[-\frac{1}{2} \sum_{k=1}^M \frac{s_{1k}^2}{\sigma_k^2}\right] \exp\left[\sum_{k=1}^M \frac{s_{1k}^2}{\sigma_k^2}\right]}{\exp\left[-\frac{1}{2} \sum_{k=1}^M \frac{s_{2k}^2}{\sigma_k^2}\right] \exp\left[\sum_{k=1}^M \frac{s_{2k}^2}{\sigma_k^2}\right]} \quad (3.13)$$

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

Правило принятия решения (3.9) можно для рассматриваемого случая записать в виде

$$\begin{cases} f[\Lambda(u)] > f[\Lambda_0] \\ f[\Lambda(u)] < f[\Lambda_0] \end{cases}^{s_1, s_2},$$

что во многих случаях приводит к более удобной форме записи, где $f(\Lambda)$ - монотонная функция, определенная для всех значений аргумента отношения правдоподобия $\Lambda(u)$. Так как $\Lambda(u) > 0$ и $\Lambda_0 > 0$, то в качестве $f(\Lambda)$ целесообразно выбрать логарифмическую функцию. Тогда, переходя в (3.13) к пределу при $M \rightarrow \infty$ и логарифмируя полученное выражение, получаем

$$\ln \Lambda(u) = \sum_{k=1}^{\infty} \frac{u_k}{\sigma_k^2} (s_{1k} - s_{2k}) - \frac{1}{2} \sum_{k=1}^{\infty} \left(\frac{s_{1k}^2 - s_{2k}^2}{\sigma_k^2} \right). \quad (3.14)$$

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

$$R(\tau) = \frac{N_0}{2} \delta(\tau). \quad (3.15)$$

Тогда, подставляя (3.15) в (3.11), легко видеть, что интегральному уравнению (3.11) удовлетворяют любые функции $\varphi_k(t)$ ортонормированные на интервале $(0, T_c)$ а все коэффициенты σ_k^2 одинаковы и равны $\frac{N_0}{2}$. При этом выражение (3.14) приобретает вид

$$\ln \Lambda(u) = \frac{2}{N_0} \sum_{k=1}^{\infty} u_k (s_{1k} - s_{2k}) - \frac{1}{N_0} \sum_{k=1}^{\infty} (s_{1k}^2 - s_{2k}^2). \quad (3.16)$$

С учетом (3.12) и ортонормированности функции $\varphi_k(t)$

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

Соответственно

$$\sum_{k=1}^{\infty} u_k (s_{1k} - s_{2k}) = \int_0^{T_c} u(t) [s_1(t) - s_2(t)] dt, \quad (3.17)$$

$$\sum_{k=1}^{\infty} (s_{1k}^2 - s_{2k}^2) = \int_0^{T_c} s_1^2(t) dt - \int_0^{T_c} s_2^2(t) dt.$$

Заменяя в (3.16) суммы интегралами согласно (3.17) находим

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

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

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

$$q = \frac{2}{N_0} \int_0^{T_c} u(t) [s_1(t) - s_2(t)] dt \geq \ln \frac{p_2}{p_1} + \frac{E_1 - E_2}{N_0} = I_0 \quad (3.18)$$

На практике обычно $p_1 = p_2$ и $E_1 = E_2$. Тогда алгоритм (3.18) запишется как

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

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

$$\int_0^{T_c} u(t) s_l(t) dt - \frac{E_l}{2} \geq \int_0^{T_c} u(t) s_r(t) dt - \frac{E_r}{2}, \quad l, r = 1, \dots, m; \quad r \neq l \quad (3.19)$$

Демодулятор (рисунок 3.1), реализующий алгоритм (3.19), содержит m идентичных каналов, включающих местный генератор, формирующий опорный сигнал $s_r(t)$, $r = 1, \dots, m$ множитель и интегратор. На выходе каждого канала установлено устройство, в котором из соответствующего результата интегрирования вычитается постоянное значение, определяемое энергией полезных сигналов.

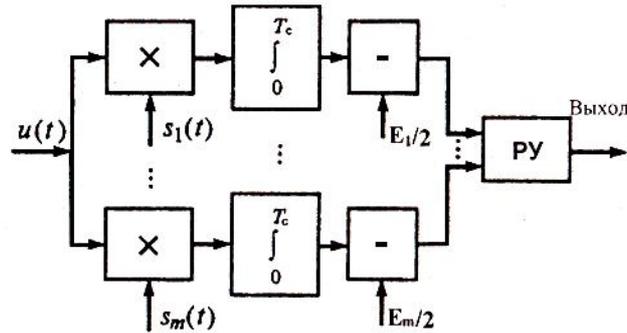


Рисунок 3.1 – Структурная схема демодулятора m детерминированных сигналов с использованием корреляторов

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

Задача построения демодулятора облегчается, если выполняется условие $E_r = E_l, r = 1, 2, \dots, m$. Такая система сигналов называется **системой с активной паузой**. При этом отпадает необходимость использовать устройства вычитания (схема на рисунке 3.1), так как в этом случае алгоритм (3.19) приобретает вид

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

Демодулятор (рисунок 3.1) называют **корреляционным**, поскольку напряжение на выходе интегратора любого i -го канала в момент окончания анализа пропорционально значению функции взаимной корреляции сигнала $u(t)$ и опорного напряжения $s_i(t)$

Другой метод реализации алгоритма (3.19) основан на использовании согласованных фильтров. Пусть имеется фильтр, согласованный с сигналом $s_l(t)$. Как известно, импульсная характеристика (реакция на воздействие в виде δ -функции) такого фильтра

$$h_i(t) = bs_l(T_c - t). \quad (3.20)$$

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

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

С учетом (3.20)

$$q_i(t) = b \int_0^t u(t-\tau)s_i(T_c - \tau)d\tau. \quad (3.21)$$

Заменив переменную интегрирования в (3.21), получим

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

Таким образом, в момент окончания полезного сигнала

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

т.е. получаем именно ту величину, которая должна быть вычислена в соответствии с алгоритмом (3.19). Следовательно, алгоритм оптимального приема может быть реализован с помощью устройства на основе фильтров $C\Phi_i$, согласованных с сигналами s_i (рисунок 3.2).

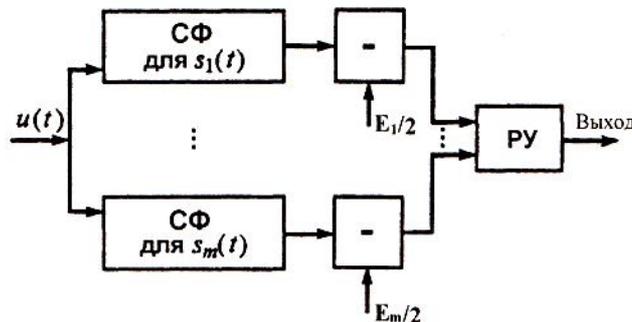


Рисунок 3.2 – Структурная схема демодулятора m детерминированных сигналов с использованием согласованных фильтров

3.2.2 Потенциальная помехоустойчивость

Двоичные системы передачи информации. Пусть сигнал на входе приемника имеет вид

$$u(t) = \lambda s_1(t) + (1-\lambda) s_2(t) + n(t) ,$$

где λ - случайная величина, принимающая значения 1 и 0 с вероятностями p_1 и p_2 соответственно, причем $p_1 + p_2 = 1$; $s_1(t)$ и $s_2(t)$ - полезные сигналы с известными параметрами; $n(t)$ - стационарный гауссовский белый шум с нулевым математическим ожиданием и корреляционной функцией $R_n(\tau) = \left(\frac{N_0}{2}\right) \delta(\tau)$.

Алгоритм работы демодулятора описывается выражением (3.18).

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

$$P_{ош} = p_1 P_{ош}(s_2 | s_1) + p_2 P_{ош}(s_1 | s_2) , \quad (3.22)$$

где условные вероятности ошибок

$$\begin{aligned}
p_{ou}(s_2 | s_1) &= \int_{q < l_0} w(q | s_1) dq, \\
p_{ou}(s_1 | s_2) &= \int_{q \geq l_0} w(q | s_2) dq,
\end{aligned}
\tag{3.23}$$

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

Найдем распределение величины q для указанных случаев. Пусть на входе приемника присутствует сигнал $s_1(t)$. Тогда с учетом (3.18)

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

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

Математическое ожидание величины q

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

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

Дисперсия величины q

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

Учитывая, что $M \{n(t_1)n(t_2)\} = \left(\frac{N_0}{2}\right) \delta(t_2 - t_1)$, и используя фильтрующее свойство δ -функции, находим $D = EN_0(1 - r_s)$. Таким образом

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

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

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

будет иметь нормальную плотность вероятности с параметрами

$$m_q = -E(1 - r_s), D_q = EN_0(1 - r_s), \text{ т. е.}$$

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

Плотности вероятности $w(q|s_1)$ и $w(q|s_2)$ представлены на рисунке 3.3.

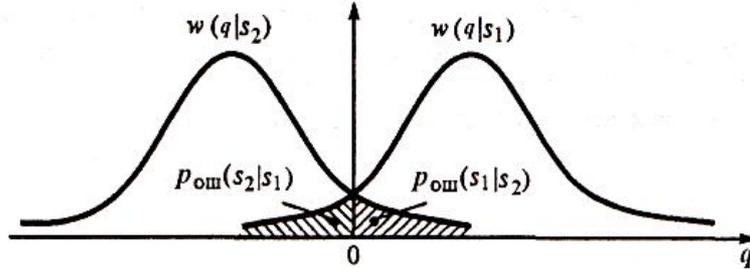


Рисунок 3.3 – Плотности вероятности сигналов на выходе оптимального демодулятора в системах с активной паузой

Теперь, используя (3.23), нетрудно определить условные вероятности ошибок:

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

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

Для $p_1 = p_2 = \frac{1}{2}$ и $E_1 = E_2 = E$ порог $l_0 = 0$. При этом

$$p_{ош}(s_2|s_1) = p_{ош}(s_1|s_2) = 1 - F\left(\sqrt{\frac{E(1-r_s)}{N_0}}\right), \quad (3.24)$$

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

Используя (3.22) и (3.24), окончательно находим

$$p_{ош} = 1 - F\left(\sqrt{\frac{E(1-r_s)}{N_0}}\right) = 1 - F(h\sqrt{1-r_s}), \quad \text{где } h^2 = \frac{E}{N_0} \quad (3.25)$$

Из (3.25) видно, что средняя вероятность ошибки зависит не только от энергии сигнала и спектральной плотности мощности шума, но и от коэффициента взаимной корреляции между сигналами, т. е. от используемой системы сигналов. Интеграл вероятности $F(x)$ является монотонно возрастающей функцией аргумента. Поэтому при одном и том же отношении $\frac{E}{N_0}$ помехоустойчивость системы оказывается тем выше, чем меньше коэффициент взаимной корреляции r_s .

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

$$p_{oui} = 1 - F\left(\sqrt{\frac{2E}{N_0}}\right) = 1 - F(\sqrt{2}h). \quad (3.26)$$

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

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

Меньшей помехоустойчивостью обладают *ортогональные* сигналы $r_s = 0$, для них

$$p_{oui} = 1 - F\left(\sqrt{\frac{E}{N_0}}\right) = 1 - F(h). \quad (3.27)$$

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

Примером ортогональных сигналов являются фазо-манипулированные сигналы с манипуляцией фазы на $\frac{\pi}{2}$

$$s_1(t) = A_0 \cos \omega_0 t, \\ s_2(t) = A_0 \cos\left(\omega_0 t + \frac{\pi}{2}\right), \quad 0 \leq t \leq T_c.$$

Ортогональные сигналы можно получить на основе частотной манипуляции. Действительно, в этом случае

$$s_1(t) = A_0 \cos(\omega_1 t - \varphi_1), \\ s_2(t) = A_0 \cos(\omega_2 t - \varphi_2), \quad 0 \leq t \leq T_c.$$

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

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

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

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

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

$$p_{ош} \approx 1 - F\left(\sqrt{\frac{1,21E}{N_0}}\right) = 1 - F(\sqrt{1,21h}).$$

Оценим помехоустойчивость системы передачи, использующей амплитудно-манипулированные сигналы

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

Алгоритм различения сигналов (5.18) в рассматриваемом случае принимает вид

$$q = \int_0^{T_c} u(t)s_1(t)dt \geq \frac{E}{2} = I_0 \quad (3.28)$$

Плотности вероятности $w(q|s_1)$ и $w(q|s_2)$ описываются гауссовским законом с параметрами $m_q = E$, $D_q = \frac{EN_0}{2}$ и $m_q = 0$, $D_q = \frac{EN_0}{2}$ соответственно (рисунок 3.4).

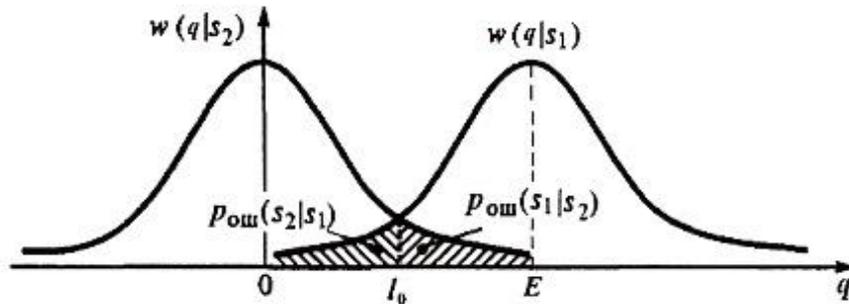


Рисунок 3.4 – Плотности вероятности сигналов на выходе оптимального демодулятора в системах с пассивной паузой

При $p_1 = p_2 = \frac{1}{2}$ средняя вероятность ошибки

$$p_{ош} = \frac{1}{2} \left[\int_{-\infty}^{I_0} w(q|s_1) dq + \int_{I_0}^{\infty} w(q|s_2) dq \right].$$

Учитывая, что порог $I_0 = \frac{E}{2}$ (3.28), находим

$$p_{ош} = 1 - F\left(0,5 \sqrt{\frac{2E}{N_0}}\right) = 1 - F\left(\frac{h}{\sqrt{2}}\right). \quad (3.29)$$

На рисунке 3.5 представлены зависимости вероятности ошибок от отношения $\frac{E}{N_0}$ для фазо- (ФМ), частотно- (ЧМ) и амплитудно-манипулированных (АМ) сигналов, рассчитанные соответственно по формулам (3.26), (3.27), (3.29) (сплошные линии).

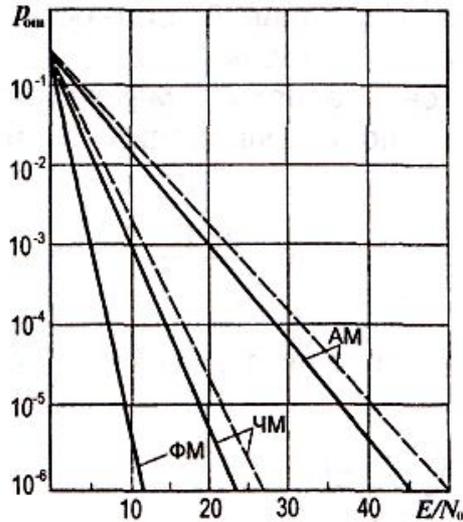


Рисунок 3.5 – Зависимость вероятности ошибки от E/N_0 для детерминированных сигналов (сплошные линии) и для сигналов со случайными начальными фазами (штриховые линии)

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

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

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

$$d = \sqrt{\int_0^T [s_1(t) - s_2(t)]^2 dt}$$

При этом формулу (3.25) можно записать в виде

$$p_{ош} = 1 - F\left(\frac{d}{\sqrt{2N_0}}\right). \quad (3.30)$$

Из (3.30) следует, что при действии в канале гауссовского белого шума вероятность ошибки зависит только от расстояния между сигналами и спектральной плотности шума. Этот вывод оказывается справедливым и для $m > 2$

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

$$p_{ош} \approx \frac{1}{\sqrt{2\pi} \sqrt{h^2(1-r_s)}} \exp\left[-\frac{h^2(1-r_s)}{2}\right], \quad (3.31)$$

которая получается при асимптотическом представлении интеграла вероятности

$$F(x) \approx 1 - \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{x^2}{2}\right). \quad (3.32)$$

Точность вычислений по формуле (3.31) не хуже 10%, если $\sqrt{h^2(1-r_s)} \geq 3$.

3.3 Синтез приемников цифровых радиосигналов со случайными параметрами

Из опыта организации и использования радиосвязи как в мирное время, так и во время военных конфликтов следует, что она является одним из основных средств обеспечения управления войсками. В современных условиях роль радиосвязи существенно возросла. Особенно велика роль радиосвязи в ВВС, так как воздушная радиосвязь является практически единственным средством оперативного обмена информацией между ВС, а также между ВС и НПУ.

Для организации авиационной радиосвязи ЛА и НПУ оборудованы соответствующими радиотехническими средствами. Важное место среди них занимают бортовые и наземные радиостанции. Современные отечественные радиостанции имеют относительно высокие ТТХ и обеспечивают надежную радиосвязь при управлении авиации на всех этапах ее применения.

3.3.1 Синтез приемников цифровых амплитудно-манипулированных радиосигналов со случайными параметрами

Практическая реализация оптимального приемного устройства, обладающего потенциальной помехоустойчивостью, сопряжена с техническими трудностями, вызванными необходимостью генерировать на приемной стороне точную копию передаваемого сигнала. Поэтому на практике часто используются более простые приемники, построенные по схеме, изображенной на рисунке 3.6. В этой схеме полосовой фильтр (ПФ), который будем считать согласованным с сигналом, реализуется фильтрующими цепями промежуточной частоты. Детектор огибающей (ДО), стоящий за фильтром, будем считать бе-

зынерционным относительно огибающей. Напряжение на его выходе совпадает с огибающей сигнала на выходе согласованного фильтра.



Рисунок 3.6 – Схема приемника АМ сигналов

При некогерентном приеме АМ сигналов их различение осуществляется сравнением в пороговом устройстве (ПУ) выходного напряжения и с некоторым пороговым напряжением H в момент окончания элементарного сигнала $t=T$. Если напряжение на выходе детектора и превышает порог H , то принимается решение о приеме символа 1, в противном случае — символа 0.

При наличии на входе суммы сигнала $s_2(t)$ с шумом на выходе детектора выделяется результирующая огибающая V смеси сигнал и шум. Известно, что амплитуда сигнала на выходе согласованного фильтра пропорциональна энергии сигнала на его входе и равна $S_b = kE$, где k — постоянный коэффициент, характеризующий усиление фильтра.

Дисперсия шума на выходе согласованного фильтра равна:

$$D = k^2 \frac{NE}{2}.$$

Огибающая смеси сигнала и узкополосного гауссовского шума распределена по закону Райса

$$p(V) = \frac{V}{D} \exp\left(-\frac{V^2 + S_b^n}{2D}\right) I_0\left(\frac{V S_b}{D}\right), \quad (3.33)$$

где I_0 — функция Бесселя нулевого порядка от мнимого аргумента.

При наличии на входе приемника одного шума ($V=0$) на выходе детектора выделяется огибающая A узкополосного гауссовского шума, распределенная по закону Релея

$$p(A) = \frac{A}{D} \exp\left(-\frac{A^2}{2D}\right). \quad (3.34)$$

Введем следующие обозначения:

$$\sigma = \sqrt{D}, \quad v = \frac{V}{\sigma}, \quad z = \frac{A}{\sigma}, \quad q = \frac{S_b}{\sigma} = \sqrt{2E/N}, \quad h = \frac{H}{\sigma} \quad (3.35)$$

Тогда выражения (3.33), (3.34) примут соответственно вид

$$p(v) = v \exp\left(-\frac{v^2 + q^2}{2}\right) I_0(qv), \quad (3.36)$$

$$p(z) = z \exp\left(-\frac{z^2}{2}\right). \quad (3.37)$$

Графики этих распределений приведены на рисунке 3.7.

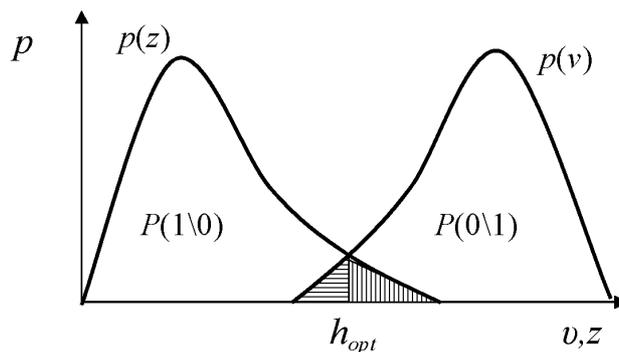


Рисунок 3.7 – Плотности вероятностей огибающих шума и смеси сигнала и шума при приеме АМ сигналов

Решение о приеме символа 1 или 0 сопровождается ошибками, условные вероятности которых находим по формулам

$$P(0|1) = P(v < h) = \int_0^h p(v) dv = \int_0^h v \exp\left(-\frac{v^2 + q^2}{2}\right) I_0(qv) dv, \quad (3.38)$$

$$P(1|0) = P(z > h) = \int_h^\infty p(z) dz = \int_h^\infty z \exp\left(-\frac{z^2}{2}\right) dz = \exp\left(-\frac{h^2}{2}\right). \quad (3.39)$$

При фиксированном значении порога h вероятности $P(0|1)$ и $P(1|0)$ равны площадям заштрихованных областей на рисунке 3.7.

Интеграл (3.38) выражается через табулированную функцию, которую называют Q -функцией:

$$Q(q, h) = \int_h^\infty v \exp\left(-\frac{v^2 - q^2}{2}\right) I_0(qv) dv. \quad (3.40)$$

Применительно к нашему случаю

$$P(0|1) = 1 - Q(q, h). \quad (3.41)$$

Окончательная формула для вероятности ошибки при некогерентном приеме АМ сигналов имеет вид:

$$P_e = 0,5 \left[1 + \exp\left(-\frac{h^2}{2}\right) - Q\left(\sqrt{\frac{2E}{N}}, h\right) \right]. \quad (3.42)$$

Отсюда видно, что вероятность ошибки зависит не только от отношения $2E/N$, но и от выбранного относительного порога h . Выражение для оптимального порога h можно найти, исследуя функцию (3.42) на минимум. Однако эту задачу можно решить проще. Из анализа рисунка 3.7 следует, что минимальное значение вероятности ошибки получается, если порог h соответствует точке пересечения кривых $p(z)$ и $p(v)$. Отклонение от этой точки в любую сторону увеличивает одну из условных вероятностей ошибок больше, чем уменьшает другую. Из условия

$$p(z) \Big|_{z=h_{opt}} = p(v) \Big|_{v=h_{opt}} \quad (3.43)$$

получим уравнение для определения оптимального значения порога

$$I_0(qh_{opt}) = \exp\left(\frac{q^2}{2}\right). \quad (3.44)$$

Нетрудно убедиться, что

$$\lim_{q \rightarrow \infty} \left[\frac{1}{2q} \ln(2\pi q h_{opt}) \right] = 0. \quad (3.45)$$

Тогда из при $q \rightarrow \infty$ получим

$$h_{opt} = \frac{q}{2}, \quad H_{opt} = h_{opt} \sigma = \frac{S_B}{2}. \quad (3.46)$$

Следовательно, при больших отношениях сигнал–шум оптимальное значение порога равно половине амплитуды сигнала на выходе фильтра. На рисунке 3.8 приведена зависимость, $H_{opt} = S_B = f(q)$, рассчитанная в соответствии с выражением (3.43)

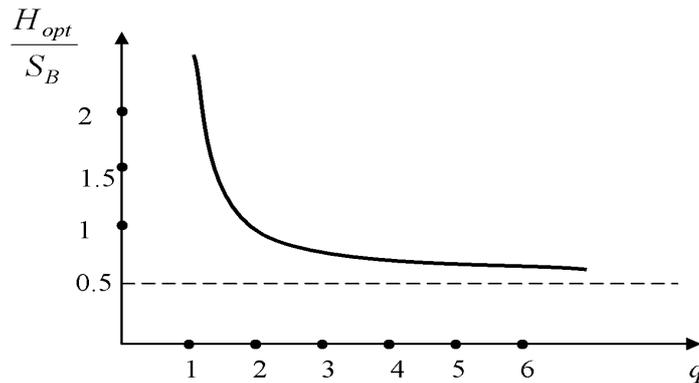


Рисунок 3.8 – Зависимость нормированного оптимального порога от отношения сигнал–шум

Отношение сигнал–шум может изменяться в процессе передачи информации, например, при взаимном перемещении ВС, замираниях сигнала и т. д. Для получения минимальной вероятности ошибки приемник АМ сигналов должен иметь систему регулирования, которая в зависимости от отношения сигнал–шум изменяла бы значение порога. Поэтому амплитудная манипуляция в системах связи с автоматической обработкой сигналов практически не применяется и используется лишь при слуховом приеме сигналов азбуки Морзе. В этом случае «решающим устройством», регулирующим порог и принимающим решение, является человек с его уникальным слуховым аппаратом.

Задавая в выражениях $P_e = 0.5 \left[1 + \exp\left(-\frac{h^2}{2}\right) - Q\left(\sqrt{\frac{2E}{N}} h\right) \right]$ и $P_e = 1 - \Phi\left(\sqrt{\frac{E}{2N}}\right)$

одинаковые значения, можно убедиться, что для обеспечения вероятности ошибки $10^{-3} \dots 10^{-6}$ некогерентный способ приема АМ сигналов требует увеличения энергии сигнала на (30...15)%. На практике из-за нестабильности частот передатчика, гетеродина приемника, эффекта Доплера приходится расширять полосу пропускания УПЧ. Поэтому помехоустойчивость реального приемника АМ сигналов оказывается еще хуже.

3.3.2 Цифровые радиосигналы с различными видами манипуляции

Амплитудная манипуляция (АМн)

Для любого момента времени t амплитудно-манипулированный сигнал можно записать в виде:

$$S(t) = A_0 \cdot \Theta(t) \cdot \cos(\omega_0 \cdot t + \phi(t)) \quad (3.47)$$

При анализе АМ сигналов на отрезке времени, равном одному тактовому интервалу T используется другая форма записи

$$\begin{aligned} S_1(t) &= 0 && \text{при } \Theta(t)=0 \text{ и } 0 \leq t \leq T \\ S_2(t) &= A_0 \Theta(t) \cos(\omega t + \phi(t)) && \text{при } \Theta(t)=1 \text{ и } 0 \leq t \leq T \end{aligned}$$

Так как $S(t)$ при $\Theta(t)=0$, то АМ сигнал называют **сигнал с пассивной паузой**.

Этюра АМ радиосигнала, ковариационная функция и спектральная плотность АМн сигнала со случайной и равномерно распределенной на интервале $[-\pi; \pi]$ начальной фазой представлены на рисунках 3.9, 3.10 и 3.11.

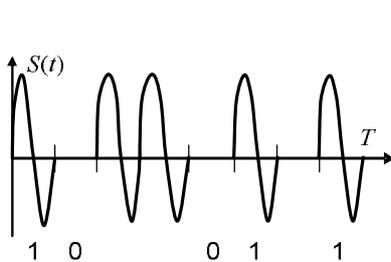


Рисунок 3.9

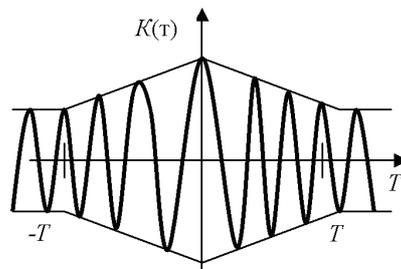


Рисунок 3.10

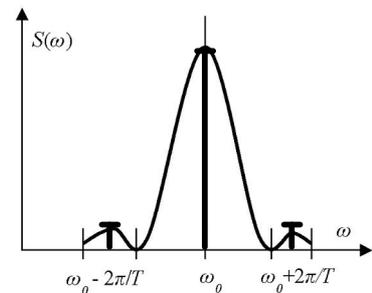


Рисунок 3.11

Спектральная плотность АМ сигнала имеет как **непрерывную**, так и **дискретную** составляющую. Непрерывная составляющая представляет собой спектральную плотность передаваемого цифрового сообщения, трансформированного в область несущей частоты ω_0 . Дискретная составляющая имеет место при постоянной начальной фазе сигнала, т.е. когда $\phi(t) = \phi_0$.

На практике в результате различных дестабилизирующих факторов начальная фаза является случайным процессом. Это приводит к «размыванию»

дискретной составляющей. Подобная особенность характерна и для других видов модуляции.

Средняя мощность АМ сигнала получается из выражения для корреляционной функции при $P_{AM} = R(0) = K(0) = \frac{A_0^2}{4}$. Эта мощность поровну распределена между дискретной и непрерывной составляющими спектральной плотности. Таким образом, на передачу полезной информации приходится лишь половина мощности, излучаемой передатчиком.

Необходимая ширина полосы частот при использовании амплитудной манипуляции, определяется равенством $\Delta f = K \times V_M$, где V_M – скорость телеграфирования, $K = 3$ для каналов связи без замирания и $K = 5$ для каналов связи с замираниями.

Фазовая манипуляция (ФМН)

Фазоманипулированный сигнал в общем случае может быть представлен следующим выражением:

$$s(t) = A_0 \cos(\omega_0 t + \Theta(t)\Phi + \phi). \quad (3.48)$$

При бинарных сигналах применяется манипуляция на $\Phi = 180^\circ$. Тогда последняя формула примет вид:

$$s(t) = A_0 \cos(\omega_0 t + \Theta(t)\pi + \phi). \quad (3.49)$$

При рассмотрении ФМ сигнала на одном тактовом интервале справедливо следующее представление:

$$\begin{aligned} s_1(t) &= A_0 \cos(\omega_0 t + \phi) && \text{при } \Theta = 0 \text{ и } 0 \leq t \leq T, \\ s_2(t) &= -A_0 \cos(\omega_0 t + \phi) && \text{при } \Theta = 1. \end{aligned} \quad (3.50)$$

ФМ сигнал является сигналом с активной паузой. Формирование ФМ сигнала может быть осуществлено с помощью схемы, приведенной на рисунке 3.12.

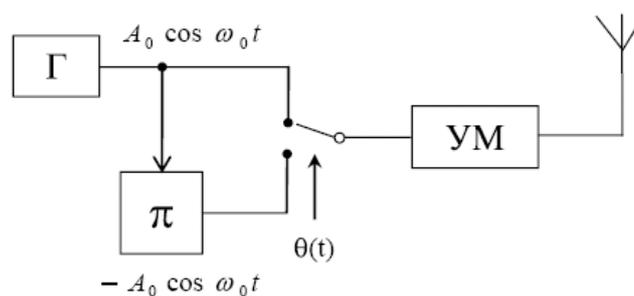


Рисунок 3.12

Временная функция, ковариационная функция и спектральная плотность ФМ сигнала представлены на рисунках 3.13, 3.14 и 3.15.

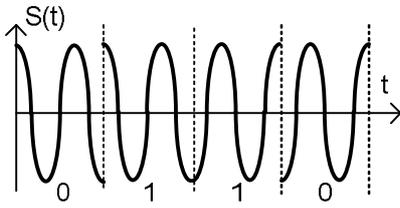


Рисунок 3.13

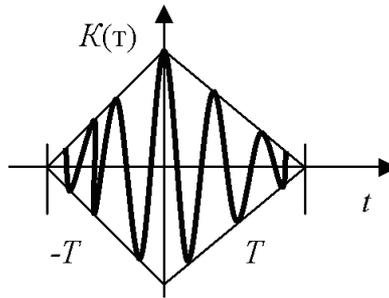


Рисунок 3.14

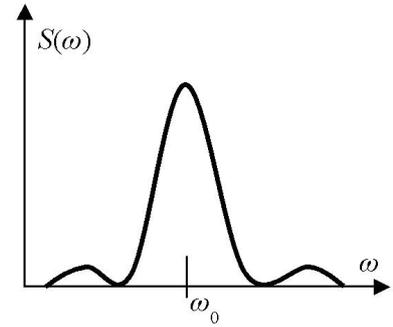


Рисунок 3.15

Дискретная составляющая в спектральной плотности ФМ отсутствует, поэтому вся мощность передатчика $P_{\text{ФМн}} = R(0) = K(0) = A_0^2/2$ расходуется на передачу полезной информации. Спектральная плотность ФМ сигнала по форме совпадает с непрерывной составляющей спектральной плотности АМ сигнала.

Частотная манипуляция (ЧМн)

Частотно манипулированный радиосигнал является сигналом с активной паузой. При передаче символа 0 передатчик излучает колебание на частоте ω_1 . При передаче символа 1 – на частоте ω_2 . Частотно-манипулированный сигнал для любого момента времени может быть записан в виде:

$$S(t) = (1 - \Theta(t))A_0 \cos(\omega_1 t + \varphi_1) + \Theta(t)A_0 \cos(\omega_2 t + \varphi_2), \quad (3.51)$$

где начальные фазы в общем случае отличаются друг от друга.

При рассмотрении сигнала на одном тактовом интервале справедлива следующая запись:

$$\begin{aligned} S_1(t) &= A_0 \cos(\omega_1 t + \varphi_1) & \text{при } \theta(t) &= 0, \\ S_2(t) &= -A_0 \cos(\omega_2 t + \varphi_2) & \text{при } \theta(t) &= 1. \end{aligned} \quad (3.52)$$

Средняя частота передачи равна

$$f_0 = \frac{1}{2}(f_1 + f_2), \quad f_i = \frac{\omega_i}{2\pi}, \quad \text{при } i = 0, 1, 2.$$

Модуль разности частот сигналов называется **частотным разносом**

$$f_p = |f_1 - f_2|,$$

а половина этой разницы называется **девиацией частоты**

$$f_d = \frac{1}{2}|f_1 - f_2|.$$

Отношение девиации к частоте манипуляции называется **индексом частотной манипуляции**.

$$\beta = \frac{f_d}{F_M} = \frac{2f_d}{V_M} = \frac{f_p}{V_M}, \quad (3.53)$$

Существует два способа осуществления манипуляции частоты:

1. На основе использования двух не связанных генераторов (рисунок 3.16).
2. Путем изменения реактивности (как правило емкости), включенной в контур автогенератора (рисунок 3.17).

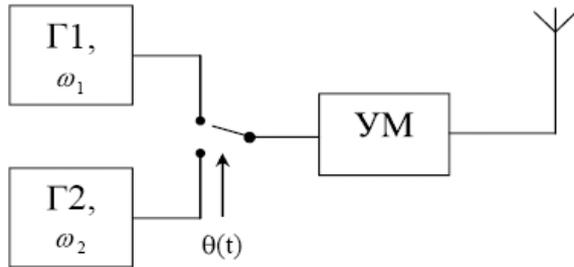


Рисунок 3.16

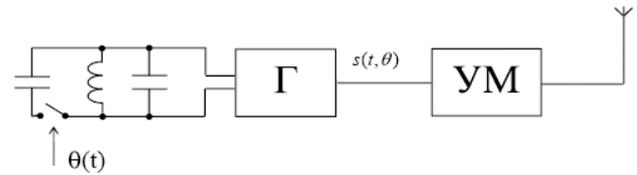


Рисунок 3.17

В данном случае получается ЧМ -сигнал с непрерывной фазой (ЧМ), который может быть записан в виде:

$$S(t) = A_0 \cos(\omega_1 t + [\omega_d \int (2\theta(\tau) - 1) d\tau] + \phi). \quad (3.54)$$

Ширина полосы частот сигнала с частотной манипуляцией при фиксированной форме импульсов зависит от скорости передачи и девиации частоты.

Ковариационная функция и спектральная плотность ЧМ радиосигнала с разрывом фазы представлены на рисунках 3.18 и 3.19.

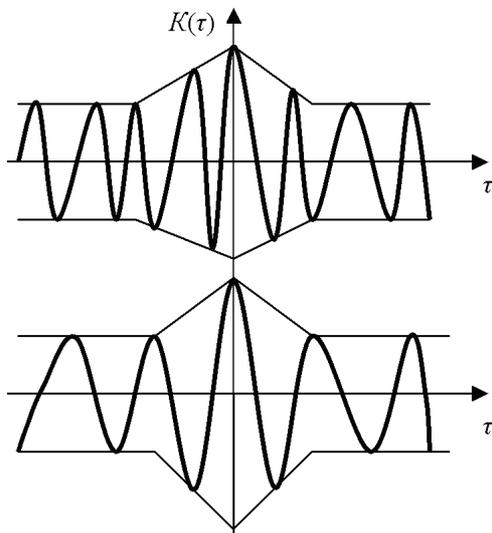


Рисунок 3.18

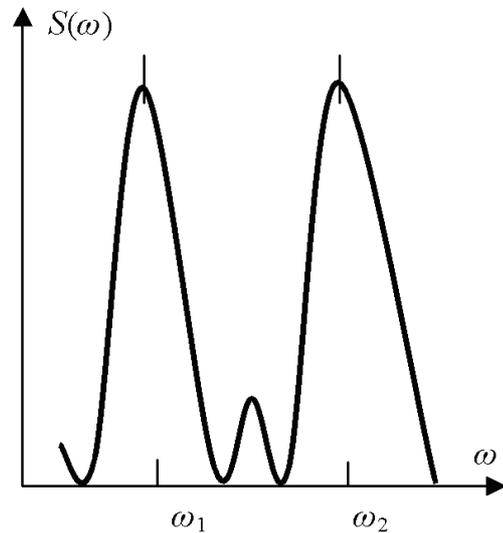


Рисунок 3.19

Здесь также наблюдается как и при АМ сигнале дискретная и непрерывная составляющие. Средняя мощность ЧМ сигнала равна

$$P_{\text{ЧМн}} = R(0) = K(0) = A_0^2 / 2 .$$

Эта мощность так же как при АМ сигнале делится поровну между дискретной и непрерывной составляющими спектральной плотности. Однако ЧМ сигнал имеет вдвое большую среднюю мощность по сравнению с АМ сигналом при одинаковых амплитудах сигналов A_0 .

Корреляционная функция ЧМн сигнала имеет более сложное представление и здесь не рассматривается.

Таким образом, можно отметить:

1. В цифровых средствах радиосвязи сообщения кодируются, каждому знаку при этом соответствует определенная кодовая комбинация.
2. Системы авиационной радиосвязи используют радиосигналы с амплитудной, фазовой и частотной манипуляцией.

ГЛАВА 4 КОДИРОВАНИЕ СИГНАЛОВ ДАННЫХ

4.1 Назначение и классификация кодов

При передаче данных по каналу с шумом всегда существует вероятность того, что принятые сигналы будут содержать ошибки. Если частота возникновения ошибок превышает некоторый допустимый уровень, который задается пользователем, то полученные данные использовать нельзя.

В теории информации доказывается, что как для неортогональных (их иногда называют симплексными), так и для ортогональных сигналов вероятность ошибки P при оптимальном приеме однозначно определяется числом сигналов m и отношением сигнал/шум $Q_{\text{вых } 0} = \mathcal{E}_0 / W_0$, где \mathcal{E}_0 — энергия элемента сигнала, а W_0 — спектральная плотность мощности белого шума. Отсюда, казалось бы, можно однозначно сделать вывод, что неограниченное повышение верности передачи сообщений (т.е. убывание вероятности ошибки P к нулю) может быть получено лишь за счет неограниченного возрастания энергетического параметра $Q_{\text{вых } 0}$ (путем увеличения мощности сигнала). В частности, если заданы средняя мощность сигнала P_c и спектральная плотность мощности белого шума W_0 , то это может происходить только при стремлении к нулю скорости передачи символов ν , поскольку $\mathcal{E}_0 = P_c T_C$, а длительность сигнала $T_C = 1/\nu$.

Первым человеком, «ошеломившим» специалистов возможностью абсолютно надежно передавать информацию по каналам связи с помехами не за счет увеличения мощности сигнала или уменьшения скорости передачи, а за счет усложнения методов модуляции-демодуляции и введения кодирования сообщений, был К. Э. Шеннон, доказавший это в известном труде «Математическая теория связи». По существу 1948 г. считается началом техники помехоустойчивого кодирования. В своей работе Шеннон нашел необходимые и достаточные условия убывания вероятности ошибки до нуля. Кроме того, он сделал это для различных моделей каналов, в частности - с ограниченной полосой частот, для фиксированного числа форм сигналов, для источников с неравновероятными или с зависимыми символами (так называемых избыточных источников), для источников непрерывных сообщений, для критериев верности приема дискретных сообщений, отличных от вероятности ошибки символа P и т. д. Все это и составило сущность теории информации.

Данная глава посвящена рассмотрению основ кодирования сообщений в цифровых каналах передачи данных.

4.1.1 Модель цифрового канала связи. Характеристика сигналов данных

Структурная схема процесса цифровой связи в общем случае представляется моделью, приведенной на рисунке 4.1.

Источник данных порождает данные в виде: двоичных символов. Обычно предполагается, что в данных «нули» и «единицы» появляются независимо и одинаковыми вероятностями. Однако в некоторых практических ситуациях это предположение неверно.

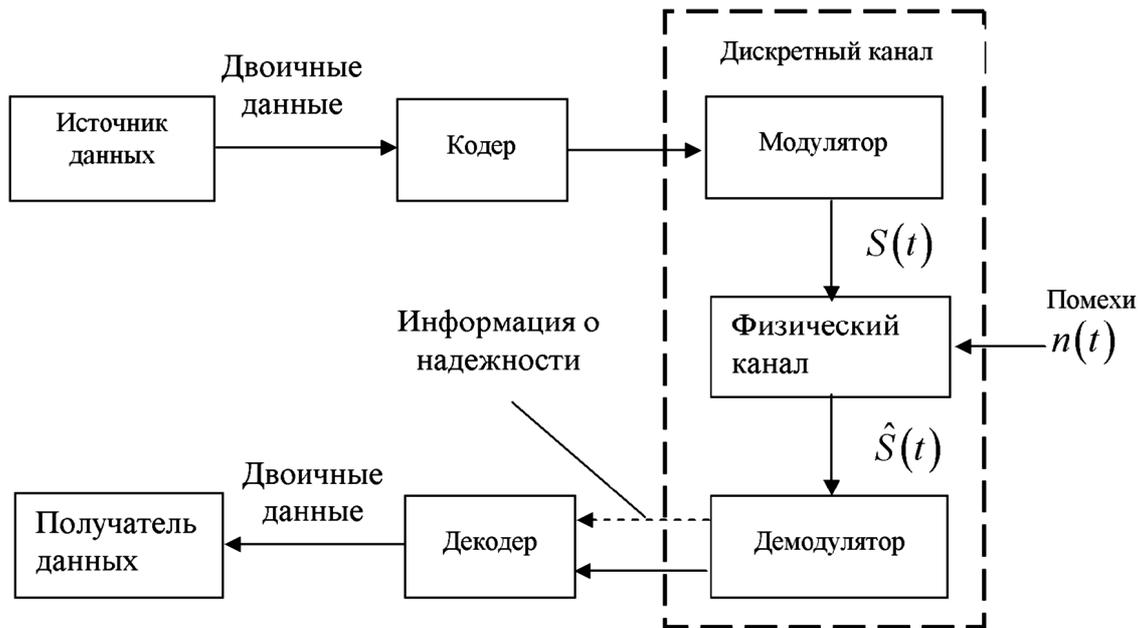


Рисунок 4.1

Кодер источника является устройством, которое преобразует наборы из k входных двоичных символов источника в наборы из n выходных символов.

Указанное преобразование должно быть взаимно однозначным. Т. е. каждому сообщению из ансамбля источника ставится в однозначное соответствие кодовая комбинация единичных элементов цифрового сигнала. Система соответствия между дискретным сообщением и кодовой комбинацией называется *первичным кодом*.

На практике значение k лежит в пределах от 3 до нескольких сотен. Число символов n в выходных блоках превышает k , что реализует принцип избыточности. Избыточные символы подчеркивают индивидуальность сообщения и выбираются из соображений минимизации вероятности потери сообщением этой индивидуальности при воздействии помех. Входные и выходные последовательности обычно состоят из двоичных символов, но иногда могут состоять из элементов некоторого алфавита большего объема.

Модулятор порождает множество непрерывных сигналов конечной длительности и реализует отображение выходных последовательностей кодера в это множество сигналов. Для схем двоичной модуляции по каждому выходному символу кодера выбирается один из двух возможных сигналов. В схемах M -ичной модуляции выходная последовательность декодера разбивается на множества по j символов (где $M = 2^j$) и каждое множество используется для выбора одного из M сигналов. В этом случае существуют различные отображения и выбор одного из них зависит от поставленных целей.

Физический канал - это вся аппаратура и вся физическая среда, через которую проходит сигнал на пути от выхода модулятора до входа демодулятора. Физический канал не обязательно представляет собой систему связи, работающую в реальном времени; он может быть системой хранения или записи данных. Обычно выходной сигнал $\hat{s}(t)$ канала является суммой входного сигнала $s(t)$, умноженного на коэффициент передачи, и случайного шума $n(t)$. Однако встречаются и более общие ситуации. Искажения могут происходить из-за сильной фильтрации или наличия нескольких путей распространения сигнала. Помехи могут вызывать замирания сигнала и приводить к колебанию его амплитуды на выходе; это эквивалентно изменению во времени параметров самого канала. Шум $n(t)$ может быть собственным шумом приемника, который моделируется аддитивным гауссовским шумом, или промышленным шумом, или, наконец, помехой, организуемой, противником.

Демодулятор или детектор - это устройство, которое на основе наблюдения принятого сигнала $\hat{s}(t)$ оценивает, какой из возможных символов был передан. Вероятность того, что эта оценка окажется правильной, зависит от отношения мощности сигнала к мощности шума в используемой полосе частот, от искажения сигнала, вызываемого фильтрацией и нелинейными эффектами, и от используемой схемы демодулятора. В системах с кодированием демодулятор часто выполняет еще одну функцию, состоящую в передаче декодеру информации о степени надежности оценки каждого символа. Эта информация может быть получена несколькими различными способами, и используемый в каждом случае подход существенно зависит от природы шума $n(t)$. Одна из возможностей возникает, когда действует помеха (например, сигнал радиолокатора), присутствие которой можно определить независимо. Информация о надежности задается в этом случае одним символом, указывающим, включен радиолокатор или нет. Другая возможность возникает, когда детектор является согласованным фильтром с отсчетом на выходе, а $n(t)$ - аддитивным гауссовским шумом. В этом случае отношение значения отсчета к пороговому значению решающей схемы является хорошим индикатором на-

дежности решения.

Декодер источника. Операция, осуществляемая декодером, обратна операции кодера. Поскольку последовательность символов, порождаемая демодулятором, может содержать ошибки, декодер должен выполнять отображение существенно более сложное, чем кодер. Чтобы сделать декодирование возможным, нужно разработать реализующие его вычислительные процедуры.

Прямое преобразование кодером и обратное - декодером должно минимизировать влияние шума (воздействующих помех) канала, т. е. минимизировать различие между входной и выходной последовательностями. *Именно для этого кодером канала во входную последовательность вводится избыточность, которая используется декодером для более точного восстановления входной последовательности на выходе канала связи.* Назначение пары кодер - декодер (кодек) источника заключается в минимальном представлении выхода источника с точки зрения избыточности.

4.1.2 Основные определения и классификация кодов

Преобразование сообщения в радиосигнал в системах связи осуществляется в виде двух операций — кодирования и модуляции. С помощью кодирования и модуляции источник сообщений согласуется с каналом связи.

В теории передачи информации *под кодированием в широком смысле понимают* преобразование сообщения в сигнал для передачи его по каналу связи. *Под кодированием в узком смысле* понимают преобразование сообщений дискретного источника для передачи их по дискретному каналу. На практике под словом «кодирование» подразумевается кодирование в узком смысле. Система кодирования — совокупность правил кодового обозначения объектов.

Сообщение может иметь как аналоговую, так и дискретную форму. Простейшим примером дискретного сообщения является текст. Любой текст представляет собой определенное число элементов: букв, цифр, знаков препинания. Их совокупность называется алфавитом источника сообщения. Поскольку число элементов в каждом алфавите конечно, то их можно пронумеровать, т. е. закодировать, и тогда передача сообщения сведется к передаче последовательности чисел.

При использовании десятичной системы счисления для образования кода требуется десять различных импульсов, например отличающихся амплитудой, длительностью и т. д. Такое представление кодов в теории связи применяют редко, поскольку для образования и его распознавания требуется сложная аппаратура. Наиболее просты арифметические действия в двоичной системе.

$$M = b_n m^n + b_{n-1} m^{n-1} + \dots + b_2 m^2 + b_1 m^1 + b_0 m^0. \quad (4.3)$$

Коэффициенты b_i принимают значения от 0 до $m-1$. Их совокупность и есть кодовая комбинация для символа a_i с номером M_i :

$$a_i \rightarrow M_i \rightarrow (b_n b_{n-1} \dots b_i b_0).$$

Код, для которого $m^n = K$, называется *примитивным* (код без избыточности).

Кодовое слово характеризуется *весом*, под которым понимается количество единиц в этом слове.

Пример: кодовое слово $X_i = 1\ 0\ 0\ 1\ 1\ 1\ 0$ имеет вес $\omega(x_i) = 4$.

Произвольное кодовое слово X_i отличается от любого другого кодового слова X_j в некоторых позициях (разрядах). Число позиций, в которых отличаются эти слова, называют *кодовым расстоянием* $d(x_i, x_j)$ (расстоянием Хэмминга) между этими словами.

Кодовое расстояние между словами X_i и X_j определяется как вес суммы по модулю "2" этих слов:

$$d(x_i, x_j) = \omega(x_i \oplus x_j).$$

Наименьшее значение $d(x_i, x_j)$ для всех $i \neq j$ называется минимальным кодовым расстоянием d_{min} . Следует отметить, что для первичных кодов $d_{min} = 1$.

Избыточность кода можно оценить соотношением

$$\rho_k = 1 - \frac{v_u (1 - \rho_u) \log K}{v_k \log m}, \quad (4.4)$$

где ρ_u – избыточность источника; v_u – скорость выдачи символов источником; v_k – скорость выдачи символов кодером.

Избыточность равномерного m -позиционного n -разрядного кода при $\rho_u = 0$

$$\rho_k = 1 - \frac{\log K}{n \log m}, \quad (4.5)$$

где $n = v_k / v_u$ – число символов кода на один символ источника.

Применяемые в технике связи коды и основные методы кодирования можно *классифицировать* по ряду специфических признаков.

По длине кодов и взаимному расположению в них символов различают равномерные и неравномерные коды. Равномерные коды имеют одинаковую длину комбинаций. Для равномерного кода число возможных комбинаций составляет m^n . Примером такого кода является пятизначный код Бодо, применяемый в телеграфии. Код Бодо содержит пять двоичных элементов ($m = 2$, $n = 5$). Число возможных кодовых комбинаций в этом коде равно $2^5 = 32$, что

позволяет кодировать все буквы русского алфавита (твердый знак не передают). Однако этого мало для передачи сообщения на русском языке, содержащего буквы, цифры, знаки препинания и условные знаки (точка, запятая, двоеточие, сложение, вычитание, умножение и т. д.). Поэтому применяют «Международный код №2» (МТК-2). В коде МТК-2 используется регистровый принцип, согласно которому одна и та же пятиэлементная кодовая комбинация может использоваться до трех раз в зависимости от положения регистра: русский, латинский, цифровой. Общее число различных знаков при этом равно 84, что достаточно для кодирования телеграммы.

Для передачи по каналам связи данных рекомендован семиэлементный код МТК-5. Коды МТК-2 и МТК-5 являются первичными (простыми). Применение равномерных кодов упрощает построение буквопечатающих устройств и не требует передачи разделительных символов между кодовыми комбинациями.

Неравномерные коды отличаются тем, что кодовые комбинации у них отличаются друг от друга не только взаимным расположением символов, но и их количеством при минимизации средней длины кодовой последовательности. Это приводит к тому, что различные комбинации имеют различную длительность. Общая идея построения неравномерных кодов подсказывается теоремой кодирования 1 Шеннона для каналов без помех. Очевидно, что средняя длина неравномерного кода будет минимизироваться тогда, когда с более вероятными сообщениями источника будут сопоставляться более короткие комбинации канальных символов. Однако проблема заключается в том, что у неравномерного кода на приемной стороне оказываются неизвестными границы этих комбинаций. Если же попытаться их выделить, используя известный способ кодирования, то декодирование может оказаться неоднозначным (действительно, если, например, букве А присвоена комбинация 1, букве Б – 0, а букве В – 10, то невозможно определить по принятой комбинации 10, передавались ли буква В или пара букв А и Б). Для того чтобы используемый код был однозначно декодируемым, он должен удовлетворять некоторым условиям. Однозначное декодирование будет обеспечено, если ни одно кодовое слово не является началом другого кодового слова. Коды, удовлетворяющие этим условиям, называются *префиксными* или *неприводимыми*. Кстати, равномерный код также является префиксным.

Наиболее известным неравномерным кодом является код Морзе, в котором символы 1 и 0 используются в двух сочетаниях – как одиночные (1 и 0) или как тройные (111 и 000). Сигнал, отражающий одну единицу, соответствует точке («•»), трем единицам – тире («—»). Символ 0 используется как знак, отделяющий точку от тире, точку от точки и тире от тире. Совокупность 000 используется как разделительный знак между кодовыми комбинациями. Заме-

тим, что код Морзе – это кодирование не с двоичным входным алфавитом (точка и тире), как кажется на первый взгляд, а с троичным алфавитом: точка, тире и пробел.

В системах связи применяются несколько алгоритмов построения неравномерных кодов с префиксным свойством. Среди них оптимальным является алгоритм Хаффмена. Здесь рассмотрим более простой алгоритм Шеннона-Фано, который в большинстве случаев приводит к тем же результатам. Упрощенно принцип создания алгоритма Шеннона-Фано заключается в следующем. Символы алфавита источника сообщения записываются в порядке убывающих вероятностей. Затем они разделяются на две части так, чтобы суммы вероятностей символов, входящих в каждую из таких частей, были примерно одинаковыми. Всем символам первой части в качестве первого символа комбинации неравномерного кода приписывается нуль, а символам второй части в качестве первого символа единица. Потом каждая из этих частей (если она содержит более одного сообщения) делится в свою очередь на две, по возможности равновероятные части и к ним применяется то же самое правило кодирования. Эта процедура повторяется до тех пор, пока в каждой из полученных частей не останется по одному сообщению. Неравномерное префиксное кодирование устраняет избыточность источника, вызванную неодинаковой вероятностью сообщений.

По признаку помехозащищенности коды, как и методы кодирования, делят на *примитивные* (первичные, простые, безызбыточные) и *помехоустойчивые* (корректирующие, избыточные).

Коды, у которых все возможные кодовые комбинации используются для передачи информации, называются примитивными или кодами без избыточности. В простых равномерных кодах превращение одного символа комбинации в другой, например 0 в 1 или 1 в 0, приводит к появлению новой разрешенной комбинации, т. е. к ошибке в принятом сообщении. Примитивное или безызбыточное кодирование применяется для согласования алфавита источника и алфавита канала. Отличительное свойство примитивного кодирования состоит в том, что избыточность дискретного источника, образованного выходом примитивного кодера, равна избыточности источника на входе кодера.

Примитивное кодирование используется также в целях шифрования передаваемой информации для ее защиты от несанкционированного доступа и повышения устойчивости работы устройств синхронизации систем связи. В этом случае правило кодирования выбирается так, чтобы вероятность появления на выходе кодера длинной последовательности, состоящей только из единиц или только из нулей, была минимальной. Подобный кодер часто называют скремблером (от англ. *scramble* - перемешивать).

В помехоустойчивых кодах для передачи сообщения используются не

все кодовые комбинации, а только некоторая их часть (разрешенные кодовые комбинации). Тем самым создается возможность обнаружения и исправления ошибки при неправильном воспроизведении некоторого числа символов. Корректирующие свойства кодов обеспечиваются введением в кодовые комбинации дополнительных (избыточных) символов. Отличительное свойство помехоустойчивого кодирования состоит в том, что избыточность источника, образованного выходом кодера, больше, чем избыточность источника на входе кодера.

Помехоустойчивое кодирование используется в различных системах связи, при хранении и передаче данных в компьютерных сетях, в бытовой и профессиональной аудио- и видеотехнике, использующей цифровую запись.

Особый класс образуют *статистические* коды. Эти коды являются неравномерными. Длина кодовой комбинации таких кодов, зависит от вероятности выбора соответствующей буквы алфавита, наиболее вероятным буквам сопоставляются короткие кодовые комбинации, а менее вероятным - более длинные (такое кодирование очень часто называют экономным, так как оно позволяет сократить среднюю длину кодовой комбинации). Представителями статистических (экономных) кодов являются код Шеннона - Фано и код Хаффмена.

Средняя длина кодовых комбинаций экономного кода не может быть меньше величины $\bar{n}_{мин}$, которая согласно теореме Шеннона об оптимальном кодировании в каналах без шумов $\bar{n}_{мин} = \frac{H(A)}{\log t} + \varepsilon$ (ε - сколь угодно малая величина).

Экономное кодирование, или сжатие данных, используют для повышения скорости передачи информации и приближении ее к пропускной способности канала связи. Отличительное свойство экономного кодирования состоит в том, что избыточность источника, образованного кодером, меньше, чем избыточность источника на входе кодера. Экономное кодирование применяют в компьютерах для архивации информации. Так, последние версии операционных систем содержат программы сжатия данных (динамические компрессоры и архиваторы), а новые стандарты на модемы для связи между компьютерами по телефонным сетям общего пользования включают сжатие в процедуры обработки данных.

Если экономное кодирование сокращает избыточность источника сообщений, то помехоустойчивое кодирование, наоборот, заключается в целенаправленном введении избыточности (дополнительных символов) для того, чтобы появилась возможность обнаруживать и (или) исправлять ошибки, возникающие при передаче по каналу связи.

В настоящее время разработано большое количество кодов и методов кодирования источников сообщений. Многие из них реализованы на практике, особенно для сжатия сообщений с большой избыточностью, например факсимильных, телевизионных и другой видеоинформации, где они позволяют увеличить скорость передачи сообщений в сотни раз.

Приведенная классификация не является исчерпывающей. В нее включены лишь некоторые коды и методы кодирования, которые широко используются в современных системах связи.

4.1.3 Первичные коды, используемые в системах связи ВВС

Использование первичных кодов дает сравнительно простую техническую реализацию передающих и приемных устройств цифровой информации.

В настоящее время наибольшее распространение получили такие первичные коды, как:

- код МОРЗЕ системы радиосвязи с ручным ТЛГ и слуховым приемом;
- код МТК-2 (гост 15607-70) системы буквопечатающей связи;
- код МТК-5 (гост 13052-67) система передачи данных в различных АСУ;
- код КОИ-7 (гост 13052-74) используется для обмена цифровой информацией в системах передачи данных (СПД);
- код КОИ-8 (гост 19768-74) используется для передачи обработки данных;
- код ДКОИ (двоичный код для обмена информацией) используется для передачи и обработки данных.

4.1.4 Характер распределения ошибок в каналах связи

Передача данных по каналу связи постоянно сопровождается наличием помех, что приводит к искажению передаваемых кодовых комбинаций. Поэтому в прямом виде первичные коды не могут быть использованы для передачи информации, т.к. достоверность в этом случае будет очень низкой. Помехоустойчивость системы связи в данном случае можно увеличить за счет использования помехоустойчивого кодирования.

Синтез оптимальных устройств считается уже проведенным, т.е. выбраны ресурсы оптимального приема

Важное значение при рассмотрении помехоустойчивых (корректирующих) кодов имеют статистические характеристики искажений передаваемых сообщений в реальных каналах связи.

При решении данного вопроса на первом этапе считали (модель 1), что в дискретных каналах связи ошибки происходят независимо друг от друга.

В этом случае канал связи полностью характеризуется одним параметром- вероятностью ошибки (искажения единичного элемента) P_e .

Тогда вероятность появления в n -элементной комбинации ровно i ошибок (Биномиальное распределение) равна

$$P_e(i,n) = C_n^i P_e^i (1 - P_e)^{n-i} = C_n^i P_e^i q^{n-i}.$$

Вероятность искажения кодовой комбинации (т.е. когда в кодовой комбинации возникает хотя бы одна ошибка) равна

$$P_e(\geq 1, n) = 1 - P_e(0, n) = 1 - q^n,$$

при $n P_e \ll 1$ $P_e(\geq 1, n) \approx n P_e$.

Вероятность появления в n -элементной комбинации ровно m ошибок и более, определяется выражением

$$P_e(\geq m, n) = \sum_{i=m}^n C_n^i P_e^i q^{n-i}.$$

Однако на самом деле (как следует из статистических исследований ошибок в реальных каналах связи) искажения элементов кодовых комбинаций являются зависимыми и проявляют тенденцию к группированию (пакетированию).

Основными причинами группового характера ошибок являются:

- в КВ и тропосферных радиоканалах - замирания и помехи;
- в радиорелейных каналах - импульсные помехи.

При зависимых ошибках для полной характеристики канала связи используется также P_e (вероятность ошибки) и дополнительный параметр α , отражающий степень группирования ошибок ($0 < \alpha < 1$).

Для независимых ошибок - $\alpha = 0$.

Для гипотетического канала, когда все ошибки сосредоточены в одной группе, $\alpha = 1$.

Наибольшее значение параметр α достигает

- в каналах тонального телеграфирования - 0,71...0,77;
- в радиорелейных телефонных каналах - 0,55...0,61;
- в тропосферных телефонных и КВ
- радиотелеграфных каналах - 0,32...0,45.

Вероятность появления хотя бы одной ошибки в n -элементной кодовой комбинации равна

$$P_{ош}(\geq 1, n) = n^{1-\alpha} P_e.$$

А вероятность возникновения m и более ошибок в n -элементной кодовой комбинации определяется выражением

$$P_{ош}(\geq m, n) = (n/m)^{1-\alpha} P_e.$$

Зависимость $P_{ош}$ от m, n представлена на рисунке 4.2.

На рисунке 4.2 для сравнения приведены аналогичные зависимости для $P(\geq m, n)$. Анализ приведенных зависимостей $P(\geq m, n)$ и $P_{ош}(\geq m, n)$ от m показывают, что $P(\geq m, n)$ уменьшается с увеличением m значительно быстрее, чем $P_{ош}(\geq m, n)$. Из приведенного анализа следует, что реальные каналы связи не обеспечивают требуемый уровень достоверности передачи данных (в АСУ).

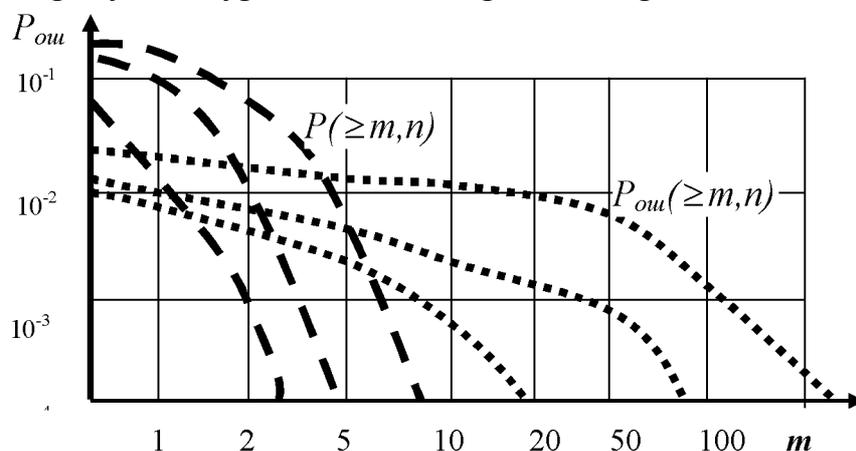


Рисунок 4.2

Таким образом, групповой характер появления ошибок существенно влияет на их распределения внутри комбинации. Для повышения помехоустойчивости СПД необходимо использовать корректирующие коды. Эти коды должны максимально отражать статистические характеристики каналов связи и иметь простую аппаратную реализацию.

4.2 Помехоустойчивые коды и их основные характеристики

4.2.1 Принципы построения избыточных помехоустойчивых кодов

Теоретической основой построения эффективных кодов является 1-ая теорема кодирования К. Шеннона, доказывающая, что для канала связи без помех всегда можно создать систему экономного кодирования дискретных сообщений, у которой среднее количество двоичных кодовых сигналов на один символ сообщения будет приближаться как угодно близко к энтропии источника сообщений.

Базой для помехоустойчивого кодирования является 2-ая теорема кодирования К. Шеннона, в которой утверждается, что для канала связи с помехами всегда можно создать систему кодирования, при которой сообщения будут переданы со сколь угодно высокой степенью верности, если только производительность источника не превышает пропускной способности канала связи.

Хотя исторически некоторые способы кодирования источников сообщений и каналов связи появились даже раньше теории Шеннона, она явилась мощным стимулом для поиска многих других, более эффективных методов, приближающихся к потенциально возможным.

Кодирование с исправлением ошибок, по существу, представляет собой метод обработки сигналов, предназначенный для увеличения надежности передачи по цифровым каналам. Хотя различные схемы кодирования очень непохожи друг на друга и основаны на различных математических теориях, всем им присущи два общих свойства. Одно из них - использование *избыточности*. Закодированные цифровые сообщения всегда содержат дополнительные (избыточные) символы, которые используют для того, чтобы сделать маловероятной потерю сообщением его индивидуальности из-за искажения символов при воздействии помех. Второе свойство состоит в *усреднении шума*. Эффект усреднения достигается за счет того, что избыточные символы зависят от нескольких информационных символов.

В двоичном канале связи с помехами, приводящими к появлению независимых ошибок символов при увеличении длины блока дискретного сообщения доля ошибочных символов в блоке стремится к средней частоте ошибок в канале, а также, что очень важно, доля блоков, число ошибок в которых существенно отличается от этого среднего значения, становится очень малой. Установлено, что при обработке символов блоками, а не одного за другим, можно уменьшить общую частоту ошибок. При фиксированной вероятности ошибки блока доля ошибок, которые нужно исправлять, уменьшается при возрастании длины блока. Сказанное свидетельствует о резервах улучшения характеристик при усреднении шума и о том, что эти резервы возрастают при увеличении длины блока. Таким образом, длинные блоковые коды эффективнее коротких.

Ключем в решении задачи исправления ошибок в символах является наличие избыточности. При построении избыточных кодов исходят из того, что из общего числа кодовых комбинаций $N_k=2^n$, которые можно образовать в элементарном двоичном коде, для передачи сообщения используется лишь их некоторая часть. Кодовые комбинации, используемые для передачи информации, называются разрешенными N_p . Остальные $(N_k - N_p)$ кодовые комбинации называются запрещенными и для передачи полезной информации не используются. Если в результате действия помех передаваемая (разрешенная) кодовая комбинация превращается в запрещенную, то ошибка будет обнаружена. Если же разрешенная кодовая комбинация превращается в другую разрешенную кодовую комбинацию, то ошибка обнаружена не будет.

При достаточно большом различии между разрешенными кодовыми

комбинациями возможно не только обнаружение, но и исправление ошибок. Избыточные коды, позволяющие обнаруживать и исправлять ошибки, называются исправляющими (корректирующими).

При исправлении ошибок в сообщении, представляемом последовательностью из n двоичных символов, очень важно, чтобы не все 2^n возможных последовательностей представляли сообщения. В самом деле, когда каждая из возможных принятых последовательностей n символов представляет некоторое сообщение, нет никаких оснований считать, что одна последовательность является более правильной, чем любая другая. Для исправления всех наборов из t или менее ошибок необходимо и достаточно, чтобы каждая последовательность, представляющая сообщение, отличалась от последовательности, представляющей любое другое сообщение, не менее чем в $(2t+1)$ местах. Например, для исправления всех одиночных или всех двойных ошибок в символах нужно, чтобы каждые две последовательности, представляющие разные сообщения, отличались не менее чем в пяти символах. Каждая принятая последовательность, содержащая два ошибочных символа и, следовательно, отличающаяся от посланной последовательности ровно в двух местах, будет всегда отличаться от всех других последовательностей, представляющих сообщения, не менее чем в трех местах. Поскольку наименьшее кодовое расстояние d_{min} для всех пар кодовых последовательностей всегда должно быть на единицу больше удвоенного числа исправляемых ошибок, то справедливо равенство

$$t = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor, \quad (4.6)$$

где $\lfloor \cdot \rfloor$ обозначает целую часть. Параметр t указывает, что все комбинации из t или менее ошибок в любой принятой последовательности могут быть исправлены.

Рассмотрим код, состоящий из четырех кодовых слов 00000, 00111, 11100 и 11011. Каждое кодовое слово используется для представления одного из четырех возможных сообщений. Поскольку код включает лишь небольшую долю всех 32 возможных последовательностей из пяти символов, мы можем выбрать кодовые слова таким образом, чтобы каждые два из них отличались друг от друга не менее чем в трех позициях. Таким образом, кодовое расстояние d_{min} равно трем и код может исправлять одиночную ошибку в любой позиции. Чтобы провести процедуру декодирования при этом коде, каждой из 28 недопустимых последовательностей нужно поставить в соответствие ближайшую к ней допустимую последовательность. Этот процесс ведет к созданию таблицы декодирования (таблица 4.1). Она строится следующим образом. Вначале под каждым кодовым словом выписываем все возможные последова-

тельности, отличающиеся от него в одной позиции. После этого осталось восемь последовательностей, каждая из которых отличается от каждого кодового слова не менее чем в двух позициях. Однако в отличие от других последовательностей эти восемь последовательностей нельзя однозначно разместить в таблице. Например, последовательность 10001 можно поместить либо в первый, либо в четвертый столбец.

При использовании этой таблицы в процессе декодирования нужно найти столбец, в котором содержится принятая последовательность, и в качестве выходной последовательности декодера взять кодовое слово, находящееся в верхней строке этого столбца.

Таблица 4.1 – Таблица декодирования для кода с четырьмя словами

<u>00000</u>	<u>11100</u>	<u>00111</u>	<u>11011</u>
10000	01100	10111	01011
01000	10100	01111	10011
00100	11000	00011	11111
00010	11110	00101	11001
00001	11101	00110	11010
10001	01101	10110	01010
10010	01110	10101	01001

Причина, по которой таблица декодирования должна строиться именно таким образом, очень проста. Вероятность появления фиксированной комбинации из i ошибок равна $P_e^i (1 - P_e)^{5-i}$ (где P_e – средняя вероятность ошибки символа). Заметим, что при $P_e < 1/2$ справедливо равенство

$$(1 - P_e)^5 > P_e (1 - P_e)^4 > P_e^2 (1 - P_e)^3 > \dots, \quad (4.7)$$

Таким образом, появление фиксированной одиночной ошибки более вероятно, чем фиксированной комбинации двух ошибок, и т. д. Это значит, что декодер, который декодирует каждую принятую последовательность в ближайшее к ней по d_{\min} расстоянию Хемминга кодовое слово, выбирает в действительности то кодовое слово, вероятность передачи которого максимальна (в предположении, что все кодовые слова равновероятны). Декодер, реализующий это правило декодирования, является *декодером максимального правдоподобия*, и в указанных предположениях он минимизирует вероятность появления ошибки декодирования принятой последовательности. В этом смысле такой декодер является оптимальным. Это понятие очень важно, поскольку декодеры максимального правдоподобия часто используются для коротких кодов. Кроме того, параметры декодера максимального правдоподобия могут

служить эталоном, с которым сравниваются параметры других, неоптимальных декодеров. Если декодирование ведется с помощью таблицы декодирования, то элементы таблицы можно расположить так, чтобы получить декодирование по максимуму правдоподобия. К сожалению, объем таблицы растет экспоненциально с ростом длины блока, так что использование таблицы декодирования для длинных кодов нецелесообразно.

Множество кодовых слов в таблице декодирования является подмножеством (первой строкой таблицы декодирования) множества всех 2^n последовательностей длиной n . В процессе построения таблицы декодирования множество всех последовательностей длиной n разбивается на непересекающиеся подмножества (столбцы таблицы декодирования). В случае, когда код исправляет t ошибок, число N_e последовательностей длиной n в каждом подмножестве удовлетворяет неравенству

$$N_e \geq 1 + n + C_n^2 + \dots + C_n^t, \quad (4.8)$$

где $C_n^i = n!/[i!(n-i)!]$ – i -й биномиальный коэффициент.

Неравенство (4.8) непосредственно следует из того, что имеется ровно n различных последовательностей, отличающихся от данной последовательности в одной позиции, C_n^2 последовательностей, отличающихся в двух позициях, и т.д. Как и в приведенном выше примере, после размещения всех последовательностей, отличающихся от кодовых в t или менее позициях, почти всегда остаются неразмещенные последовательности (отсюда неравенство в (4.8)).

Теперь можно связать избыточность кода с числом ошибок, которые им исправляются. Заметим, что число всех возможных последовательностей равно 2^n . Каждый столбец таблицы декодирования содержит N_e таких последовательностей, поэтому общее число кодовых слов должно удовлетворять неравенству

$$N_c \leq 2^n / [1 + n + C_n^2 + \dots + C_n^t]. \quad (4.9)$$

Это неравенство называется *границей Хемминга* или границей сферической упаковки. Равенство в (4.9) достигается только для так называемых *совершенных кодов*. Эти коды исправляют все наборы из t или менее ошибок и не исправляют никаких других наборов. Число известных совершенных кодов очень невелико, так что равенство в (4.9) достигается в очень редких случаях.

Процесс кодирования состоит в том, что наборы k информационных символов отображаются в кодовые последовательности, состоящие из n символов. Любое такое отображение будем называть (n, k) -кодом, хотя обычно такое название применяется только к линейным кодам (которые рассмотрим позже). Поскольку число последовательностей длиной k равно 2^k , неравенство

(4.9) можно переписать следующим образом:

$$2^k \leq 2^n / [1 + n + C_n^2 + \dots + C_n^t]. \quad (4.10)$$

Мера эффективности кода определяется отношением $R = k/n$ и называется *скоростью кода*. Доля избыточно передаваемых символов равна $(1-R)$. Отображение, возникающее при кодировании, можно задавать таблицей кодирования. Например, рассмотренный выше код с четырьмя кодовыми словами задается таблицей 4.2.

Таблица 4.2 – Таблица поиска при декодировании

Входная последовательность	Кодовая последовательность		
00	0	01	00
01	0	11	11
10	1	10	11
11	1	00	00

Часть кодовой последовательности, заключенная между штриховыми линиями, совпадает с входной последовательностью. Поэтому каждой кодовой последовательности, легко однозначно сопоставить входную последовательность. Те блоковые коды, которые обладают этим свойством, называются *систематическими кодами*. Избыточными символами для систематических кодов в таблице 4.2 являются символы на позициях 1, 4 и 5. Коды, не обладающие указанным свойством, называются *несистематическими*.

4.2.2 Классификация и основные характеристики избыточных кодов

В системах передачи данных могут применяться различные избыточные коды, классификация которых представлена на рисунке 4.3.

Существует два больших класса избыточных кодов: по способу внесения избыточности в передаваемые последовательности коды разделяются на *блочные* и *непрерывные* (или *древовидные*).

В *блочных* кодах кодовая последовательность группируется в отдельные блоки, которые кодируются независимо друг от друга. Декодирование для каждого блока также осуществляется отдельно и независимо от других блоков. При *блочном* кодировании каждому дискретному сообщению соответствует кодовое слово с определенными числом и значением символов. Важными параметрами блокового кода являются n , k , $R = k/n$ и d_{\min} . На практике значения k лежат между 3 и несколькими сотнями, а $R = 1/4 \dots 7/8$. Входные и выходные последовательности обычно состоят из двоичных символов, но иногда могут состоять из элементов некоторого алфавита большего объема.

Помимо блочных распространены рекуррентные (цепные) коды. В *цеп-*

ном коде информационные символы чередуются с проверочными, образуя последовательность

$$a_1 a_{1,2} a_2 a_{2,3} a_3 a_{3,4} a_4 \dots,$$

где a_l – l -й информационный символ, принимающий значение 0 или 1 в соответствии с передаваемым сообщением, а $a_{l,l+1}$ – проверочный символ, определяемый уравнением

$$a_{l,l+1} = a_l \oplus a_{l+1}, \quad (4.11)$$

причем суммирование осуществляется по модулю 2.



Рисунок 4.3 – Классификация кодов

В непрерывных кодах подготовленная для передачи информационная последовательность в процессе кодирования непрерывно дополняется определенным образом избыточными символами. При непрерывном (рассредоточенном) кодировании в безыбыточную последовательность информационных символов (последовательность слов простого кода) включаются избыточные символы, размещение которых сообразуется лишь с видом вносимых ошибок и может не зависеть от положения границ первоначальных слов простого кода. Значение избыточных символов при этом определяется значением некото-

рого числа информационных символов, которые могут принадлежать нескольким разным словам первоначального простого кода. Непрерывные коды характеризуются также скоростью $R = m/n$ и свободным расстоянием $d_{св}$. Точное определение $d_{св}$ более громоздко, чем определение d_{\min} для блочных кодов, однако параметр $d_{св}$, по существу, содержит ту же информацию о коде, что и d_{\min} . Типичные значения параметров непрерывных кодов таковы: $m, n = 1 \dots 8, R = 1/4 \dots 7/8, v = 2 \dots 60$.

Декодирование в непрерывных кодах, т. е. выделение информационных символов по поступающей на декодер кодовой последовательности, осуществляется также непрерывно. Определяющее различие между кодерами для кодов этих двух типов (*блочных* и *непрерывных*) состоит в наличии или отсутствии памяти. Кодер для *блочного* кода является устройством без памяти, отображающим последовательности из k входных символов в последовательности из n выходных символов. Термин «без памяти» не означает, что кодер не содержит элементов памяти, а указывает, что каждый блок из n символов зависит только от соответствующего блока из k символов и не зависит от других блоков.

Кодер для *непрерывного* кода является устройством с памятью, в которое поступают наборы из m двоичных входных символов, а на выходе появляются наборы из n двоичных выходных символов. Каждый набор n выходных символов зависит от текущего входного набора и от v предыдущих входных символов. Таким образом, память кодера должна содержать $(v + m)$ входных символов. Параметр $(v + m)$ часто называют длиной кодового ограничения данного кода и обозначают $k = v + m$ (не следует путать с параметром k для блочного кода).

Как блочные, так и непрерывные коды могут быть *систематическими* и *несистематическими*. Как было указано выше, в блочных систематических кодах *можно указать позиции размещения информационных и избыточных символов*. В противном случае код является *несистематическим*. Блочные систематические коды принято обозначать как (n, k) , где n -длина кодовой комбинации, k -количество информационных символов.

Если избыточные символы кодовой комбинации образуются путем применения к информационным символам некоторых линейных операций, то код называется *линейным*. В противном случае код является *нелинейным*.

Почти все схемы кодирования, применяемые на практике, основаны на линейных кодах.

Линейные коды образуют векторное пространство и обладают следующим важным свойством: два кодовых слова можно сложить, используя подходящее определение суммы, и получить третье кодовое слово. В случае обыч-

ных двоичных кодов эта операция является посимвольным сложением двух кодовых слов по модулю 2 (т. е. $1 + 1 = 0$, $1 + 0 = 1$, $0 + 0 = 0$). Это свойство приводит к двум важным следствиям. Первое состоит в том, что линейность существенно *упрощает процедуры кодирования и декодирования*, позволяя выразить каждое кодовое слово в виде: «линейной» комбинации небольшого числа выделенных кодовых слов, так называемых базисных векторов. Второе свойство состоит в том, что линейность существенно *упрощает задачу вычисления параметров кода*, поскольку расстояние между двумя кодовыми словами при этом эквивалентно расстоянию между кодовым словом, состоящим целиком из нулей, и некоторым другим кодовым словом. Таким образом, при вычислении параметров линейного кода достаточно рассмотреть, что происходит при передаче кодового слова, состоящего целиком из нулей. Вычисление параметров упрощается еще и потому, что расстояние Хемминга между данным кодовым словом и нулевым кодовым словом равно числу ненулевых элементов данного кодового слова. Это число часто называют весом Хемминга данного слова, и список, содержащий число кодовых слов каждого веса и называемый *спектром кода*, можно использовать для вычисления характеристик кода.

Систематические линейные блочные коды называются *групповыми* кодами, поскольку кодовые слова образуют математическую структуру, называемую *группой*. Характерным свойством групповых кодов является то, что сумма по модулю два любой пары кодовых слов также является кодовым словом этого кода.

Групповые коды составляют очень малую часть всех блочных кодов. Однако, за редкими исключениями, именно групповые коды являются блок-кодами, имеющими практическое значение. Групповые коды часто называются также *кодами с обобщенными проверками на четность*. В классе всех групповых кодов можно выделить важный большой подкласс, состоящий из так называемых *полиномиальных кодов*, которые будут рассмотрены ниже. Примерами полиномиальных кодов являются коды Боуза - Чоудхури - Хоквингема (БЧХ), коды Рида - Соломона (РС), обобщенные коды Рида - Маллера, проективно-геометрические, евклидово-геометрические и квадратично-вычетные коды.

Линейные непрерывные коды обычно называют *сверточными* кодами, поскольку операцию кодирования можно рассматривать как дискретную свертку входной последовательности с импульсным откликом кодера.

Сверточные коды являются разновидностью цепных кодов и имеют кодовые последовательности вида

$$\dots a_{(i-1),\text{пр}} a_i a_{(i+1),\text{пр}} a_{(i+2),\text{пр}} a_{(i+3),\text{пр}} a_{(i+4),\text{пр}} a_{(i+5),\text{пр}} \times \\ \times a_{(i+6),\text{пр}} a_{(i+7),\text{пр}} \dots$$

Проверочные символы такого кода формируются часто по правилу

$$a_{i,\text{пр}} = a_{i-4} + a_{i-7}; \quad a_{(i+1),\text{пр}} = a_{i-1} + a_{i-4} + a_{i-7}.$$

Суммирование ведется по модулю 2.

К *непрерывным неблочным* кодам можно условно отнести относительный код, используемый в дискретных системах связи с относительной модуляцией фазы (ОФМ).

Обширным классом линейных кодов являются *циклические коды*, у которых циклический сдвиг кодовой комбинации приводит к образованию кодовой комбинации этого кода. Среди циклических кодов в первую очередь следует выделить коды БЧХ и РС.

Тот или иной код может применяться самостоятельно или в виде: некоторого объединения с другим кодом, когда наборы символов одного избыточного кода рассматриваются как элементарные информационные символы другого избыточного кода. Такое объединение называется *каскадным кодом*.

При каскадном кодировании наблюдается существенное упрощение конструкции кодеров и декодеров по сравнению с аналогичными устройствами обычных (не каскадных) кодов той же длины и избыточности. Обычно в каскадных кодах в качестве внутреннего кода используют двоичный код (код - БЧХ), а в качестве внешнего - недвоичный (РС - код).

Различают помехоустойчивые коды с *обнаружением* и *исправлением* ошибок. Код с *обнаружением ошибок* уменьшает число неверно опознанных сообщений, позволяет «стирать» или особо отмечать сообщения, в которых установлено присутствие ошибки, а в некоторых случаях (системы с обратными информационными связями) принять меры к повторной передаче и приему неопознанных сообщений.

Простейшим из помехоустойчивых кодов является код с проверкой на четность. Этот код позволяет обнаруживать одиночную ошибку. Его суть заключается в следующем. На передающем конце канала связи устройство кодирования проводит подсчет числа логических «1» в передаваемом двоичном кодовом слове. Если сумма «1» оказывается нечетной, в конец передаваемой кодовой комбинации добавляется «1», а если нет, то «0». На приемном конце канала связи проводится аналогичный подсчет, и если контрольная сумма (число единиц в принятой кодовой комбинации) будет нечетной, то принимается решение о том, что при передаче произошло искажение информации, в противном случае принятая информация признается достоверной. В описанном способе помехоустойчивого кодирования используется один добавочный

контрольный разряд в кодовом слове. Это позволяет обнаруживать ошибку передачи только одного разряда в одном кодовом слове сообщения. При контроле на четность единственный способ получить достоверную информацию — повторная передача кодового слова. Для этого приемник формирует специальную команду, которая передается по каналу связи в обратном направлении к передатчику сообщений.

Код с *исправлением ошибок* позволяет получать верные сообщения, несмотря на наличие некоторого числа ошибок при опознании символов. *Коррекция (обнаружение или исправление)* ошибок достигается лишь при использовании в кодовом слове определенного числа избыточных символов. *Корректирующими кодами* называют коды, которые позволяют обнаруживать ошибки и исправлять ошибки и стирания, возникающие при передаче дискретных сообщений. Коды, у которых требуемая корректирующая способность достигается при минимальной избыточности, называются *оптимальными*.

Рассмотрим код, *исправляющий* одиночную ошибку. Общий принцип построения такого кода аналогичен принципу построения кода, обнаруживающего одиночную ошибку. Однако при этом из всех кодовых слов длиной n символов для использования в системе передачи необходимо отобрать лишь определенную часть, считая остальные слова запрещенными. При этом, чтобы исправить одну ошибку, все отобранные кодовые слова должны отличаться друг от друга как минимум на три символа. В этом случае одиночная ошибка переведет переданное слово в одно из запрещенных, что и позволит обнаружить ошибку, но полученное запрещенное слово отличается от переданного лишь одним символом, а от остальных разрешенных слов - не менее чем на два символа. Значит принятое слово «ближе» к действительно переданному и менее похоже на остальные разрешенные слова. Поэтому используют следующий метод принятия решения при появлении в приемнике запрещенного кодового слова: в качестве переданного слова принимается то из разрешенных, которое отличается от принятого на меньшее число символов. Если при поступлении n символов лишь один символ окажется неверным, то данный метод позволит указать среди разрешенных слов действительно переданное даже в том случае, когда принятое слово содержит одну ошибку. Это эквивалентно исправлению одиночной ошибки.

Практически пока не существует кодов, позволяющих обнаруживать и тем более исправлять все возможные ошибки. Можно построить лишь коды, обнаруживающие или исправляющие некоторое число ошибок определенного вида, наиболее вероятных или опасных.

Наконец, коды можно разбить на коды, *исправляющие случайные ошибки*, и коды, *исправляющие пакеты ошибок*. В основном мы будем иметь дело с

кодами, предназначенными для исправления случайных или независимых, ошибок. При наличии пачек ошибок часто оказывается более выгодным использовать коды, исправляющие случайные ошибки, вместе с устройством *перемежения восстановления*. Такой подход включает в себя процедуру перемешивания порядка символов в закодированной последовательности перед передачей и восстановлением исходного порядка символов после приема с тем, чтобы рандомизировать ошибки, объединенные в пакеты.

Рассмотрим основные характеристики избыточных кодов.

Обнаруживающие и исправляющие возможности избыточных кодов обусловлены введением избыточности в первичные коды, которая оценивается *коэффициентом избыточности*. Он является определяющей мерой эффективности кода, показывает степень удлинения кодовых слов для обеспечения заданной помехоустойчивости и определяется выражением вида

$$K_{II} = I - \frac{\log_2 N_p}{\log_2 N_n},$$

где N_p – количество кодовых слов (разрешенных кодовых комбинаций),
 N_n – общее количество различных кодовых комбинаций длиной n .

Например, для двоичного кода длиной n с k информационными разрядами $N_p = 2^k$; $N_n = 2^n$, отсюда $K_{II} = I - k/n = (n-k)/n = r/n$;

Часто, в качестве другой меры эффективности используется *скорость кода*, которая определяется соотношением

$$R = k/n \leftrightarrow K_{II} = I - R,$$

С учетом этих характеристик, лучшим по эффективности среди различных кодов одинаковой длины с одним и тем же значением d_{min} , следует считать код, обеспечивающий большую скорость или меньший коэффициент избыточности.

Следующими характеристиками избыточных кодов являются:

- число обнаруживаемых ошибок – S ;
- число исправляемых ошибок – t .

В общем случае d_{min} является важной характеристикой избыточных кодов по обнаружению и исправлению ошибок различной кратности.

Если при передаче x_i происходит t ошибок то $d(x_i^t, x_i) = t$; (x_i^t – искаженная кодовая комбинация).

Если минимальное хеммингово расстояние между разрешенными кодовыми комбинациями d_{min} , то код позволяет *обнаружить* ошибку, когда в кодовой комбинации число ошибочно принятых символов удовлетворяет условию $S < d_{min}$.

Если код предназначен только для *обнаружения* ошибок кратности S , то

здесь справедливо неравенство

$$d_{\min} = d(x_i^s, x_j) > S;$$

или с учетом целочисленности кодового расстояния

$$d_{\min} \geq S+1; \quad S \leq d_{\min}-1$$

Следовательно, максимальная кратность обнаруживаемых ошибок

$$S_0 = d_{\min} - 1.$$

Принятая искаженная кодовая последовательность x_i^t может быть исправлена при условии, что она отстоит от других кодовых слов на расстоянии

$$d(x_i^t, x_j) > t$$

для всех $i \neq j$.

Отсюда следует, что для исправления всех t ошибок избыточный код должен иметь минимальное кодовое расстояние, удовлетворяющее условию:

$$d_{\min} = d(x_i^t, x_i) + d(x_i^t, x_j) > 2t$$

или, учитывая целочисленность кодового расстояния, последнее неравенство можно переписать в виде:

$$d_{\min} \geq 2t+1 \text{ или } t \leq (d_{\min}-1)/2.$$

Максимальная кратность полностью исправляемых ошибок

$$t_u = \begin{cases} \frac{d_{\min} - 1}{2} & \text{при нечетном } d_{\min}, \\ \frac{d_{\min}}{2} - 1 & \text{при четном } d_{\min}. \end{cases}$$

При декодировании с исправлением ошибок и стираний могут быть исправлены $t \leq t_u$ ошибок и $t \leq t_c$ стираний, если их число удовлетворяет условию

$$t_u < \frac{d_{\min} - t_c}{2}.$$

В случае применения кода для обнаружения S ошибок и исправления t ошибок его минимальное кодовое расстояние должно удовлетворять соотношению

$$d_{\min} \geq S + t + 1,$$

где в свою очередь должно выполняться условие $S > t$.

В общем случае код с расстоянием d_{\min} исправляет произвольное число $t \leq t_c$ стираний, $t \leq t_u$ ошибок и обнаруживает произвольное число $t_u \leq t \leq t_0$ ошибок при условии, что $t_u + t_0 + t_c < d_{\min}$.

Полученные выражения определяют кратность гарантированного обна-

ружения и исправления ошибок. В действительности избыточный код способен исправлять также часть ошибок и более высокой, чем t кратности.

В качестве оценки помехоустойчивости избыточных кодов используют вероятность неисправленных ошибок $P_{ни}$.

Избыточный код, как правило, не исправляет ошибки кратности T , удовлетворяющей неравенству

$$T > t = (d_{min} - 1) / 2 .$$

Поэтому вероятность неисправленных ошибок $P_{ни}$ определяется границей

$$P_{ни} \leq P(\geq (d_{min} + 1) / 2, n)$$

и соответствует вероятности того, что в последовательности из n элементов будут искажены не менее $(d_{min} + 1) / 2$ позиций.

Значение этой границы $P(\geq (d_{min} + 1) / 2, n)$ для каналов с независимыми ошибками определяется выражением вида

$$P_{ни} \leq P(\geq (d_{min} + 1) / 2, n) = \sum_{i=\frac{d_{min} + 1}{2}}^n C_n^i P_e^i (1 - P_e)^{n-i} ,$$

а при учете группирования ошибок

$$P_{ни} \leq P_{ош}(\geq (d_{min} + 1) / 2, n) = \left(\frac{n}{(d_{min} + 1) / 2} \right)^{1-\alpha} P_e .$$

В случае использования кода только для обнаружения ошибок оперируют вероятностью необнаружения ошибок

$$P_{но} = \sum_{i=d_{min}}^n \frac{A(i)}{C_n^i} P(i, n),$$

где $A(i)$ – количество i -кратных ошибок, не обнаруживаемое кодом;

C_n^i – общее возможное количество i -кратных ошибок;

$P(i, n)$ – вероятность появления i -кратных ошибок в последовательности символов длиной n . Эта вероятность определяется каналом связи и для случая независимых ошибок определяется выражением

$$P(i, n) = C_n^i P_e^i (1 - P_e)^{n-i}$$

Для канала с зависимыми ошибками вероятность $P(i, n)$ имеет вид

$$P_{ош}(i, n) = ((n/i)^{1-\alpha} - (n/(i+1))^{1-\alpha}) P_e .$$

Для большинства кодов (особенно длинных) точное определение $A(i)$ затруднительно. Поэтому с достаточной для инженерных расчетов точностью можно полагать, что

$$A(i)/C_n^i \approx 1/2^r.$$

В этом случае

$$P_{HO} = (1/2^r) \sum_{i=d_{\min}}^n P(i, n) = (1/2^r) P(\geq d_{\min}, n).$$

Коэффициент повышения достоверности – служит для оценки повышения достоверности передачи информации за счет применения избыточных кодов. Он показывает, во сколько раз уменьшается вероятность появления ошибочных данных на выходе декодера по сравнению с вероятностью ошибочного приема последовательности в канале связи.

$$K_{ПД} = P_{ОШ} / P_{HO} \approx 2^r P(\geq l, n) / P(\geq d_{\min}, n).$$

Поскольку $P(\geq l, n) > P(\geq d_{\min}, n)$, то для $K_{ПД}$ значение 2^r является удачной оценкой снизу

$$K_{ПД} > 2.$$

Параметром, по которому различные коды, а также различные системы передачи дискретных сообщений можно сравнивать между собой по эффективности избыточного кодирования, является *эквивалентная вероятность ошибки приема элементарного символа*

$$p_{\text{э}} = 1 - (1 - p_k)^{1/k}, \quad (4.12)$$

где p_k - вероятность ошибочного декодирования кодовой комбинации данного кода; k - число информационных разрядов. Эквивалентная вероятность ошибки определяет вероятность ошибки элементарного символа в двоичном симметричном стационарном канале без памяти, в котором система с примитивным кодированием обеспечивает при передаче того же количества информации ту же вероятность ошибочного декодирования кодовой комбинации p_k , что и заданная система с избыточным кодом.

Сравнение систем связи, использующих различные коды, по эквивалентной вероятности ошибки $p_{\text{э}}$ целесообразно проводить при неизменной средней скорости передачи информации $I' = v_k(1 - \rho_k) = \frac{1}{T_c} \frac{k}{n}$ и фиксированной средней мощности сигнала P_c .

Очевидно, что корректирующий код целесообразно применять в случае, когда $p_{\text{э}} < p_0$, где p_0 - вероятность ошибки, которая имела бы место при использовании примитивного кода и неизменной скорости передачи информации. С другой стороны из двух корректирующих кодов лучше тот, который обеспечивает меньшее значение $p_{\text{э}}$.

Если фиксировать спектральную плотность мощности шума $G_{ш}$ и I' , то параметром сравнения можно сделать эквивалентное отношение сигнал/шум

$$h_{\text{э}}^2 = h^2 \frac{n}{k} = \frac{P_c}{G_{ш} I'} = \frac{1}{\beta}, \quad (4.13)$$

где β — коэффициент использования мощности сигнала. Выигрыш по эквивалентной вероятности ошибки при переходе от i -й к j -й системе кодирования

$$\alpha_{i/j} = \frac{p_{\text{э}1}}{p_{\text{э}j}}, \quad I = \text{const}, \quad h_{\text{э}}^2 = \text{const}. \quad (4.14)$$

Эффективность системы кодирования можно оценить и по энергетическому выигрышу перехода от i -й системы к j -й.

$$\eta_{pi/j} = 10 \cdot \lg \frac{h_{\text{э}i}^2}{h_{\text{э}j}^2}; \quad p_{\text{э}} = \text{const}. \quad (4.15)$$

Совокупность приведенных характеристик достаточно полно описывает свойства применяемых на практике кодов и используются при обосновании и выборе той или иной системы кодирования при разработке цифровых систем передачи данных.

4.3 Помехоустойчивые коды

4.3.1 Линейные коды

На практике широко распространены *линейные коды*. Линейный двоичный код длины n — это код, для которого сумма по модулю 2 любых разрешенных кодовых комбинаций также является разрешенной кодовой комбинацией.

Если формирование кодовой комбинации осуществляется в два этапа, причем на первом этапе образуются кодовые комбинации примитивного кода, а затем по определенному правилу к ним добавляются избыточные (контрольные, проверочные) символы, то код называется *систематическим*.

Как было отмечено выше, систематические линейные блочные коды называются *групповыми кодами* или *кодами с обобщенными проверками на четность*. В классе групповых выделяется подкласс *полиномиальных кодов*.

В начале этого параграфа рассмотрим некоторые общие свойства групповых кодов. Затем уделим основное внимание полиномиальным кодам. Для рассмотрения этих кодов придется ввести ряд элементарных понятий теории линейных векторных пространств, групп и полей.

Будем предполагать, что каждое кодовое слово группового кода разбито на две части. Первая часть, состоящая из k символов, всегда совпадает с передаваемой информационной последовательностью. Каждый из $n - k$ символов

второй части вычисляется как линейная комбинация фиксированного подмножества информационных символов. Поэтому эти символы называются символами обобщенных проверок на четность или просто *символами четности*. Коды такого типа, в которых информационные символы при кодировании не изменяются, выше назывались систематическими. Любой групповой код можно сделать систематическим на некотором множестве из k позиций, выбрав подходящее соответствие между входными последовательностями и кодовыми словами.

Групповые коды

Простым двоичным групповым кодом является $(n, n-1)$ – код, построенный с помощью одной общей проверки на четность. Например, кодовое слово $(4,3)$ -кода может быть записано в виде: вектора-столбца

$$\mathbf{a}^T = (a_1, a_2, a_3, a_1 + a_2 + a_3),$$

где a_i принимают значения 0 или 1, а $+$ означает сложение по модулю 2. Отметим, что если посимвольно прибавить к первому кодовому слову второе кодовое слово $\mathbf{b}^T = (b_1, b_2, b_3, b_1 + b_2 + b_3)$, то получим

$$\begin{aligned} \mathbf{c}^T = \mathbf{a}^T + \mathbf{b}^T = & (a_1 + b_1, a_2 + b_2, a_3 + b_3, a_1 + b_1 + a_2 + b_2 + \\ & + a_3 + b_3) = (c_1, c_2, c_3, c_1 + c_2 + c_3). \end{aligned}$$

Таким образом, проверочный символ в слове \mathbf{c} получается точно такой же процедурой, как проверочные символы в словах \mathbf{a} и \mathbf{b} . Поэтому \mathbf{c} также является кодовым словом. Этот пример показывает наиболее важное свойство групповых кодов, которое называется *замкнутостью*: сумма двух кодовых слов также является кодовым словом. Этот результат обобщается на любой групповой код.

В качестве второго примера определим кодовое слово $(6, 3)$ -кода равенством

$$\mathbf{a}^T = (a_1, a_2, a_3, a_1 + a_2, a_2 + a_3, a_1 + a_2 + a_3).$$

Если посимвольно прибавить к \mathbf{a} второе кодовое слово \mathbf{b} , то получим слово \mathbf{c} , в котором

$$\begin{aligned} c_1 &= a_1 + b_1, \\ c_2 &= a_2 + b_2, \\ c_3 &= a_3 + b_3, \\ c_4 &= a_1 + b_1 + a_2 + b_2 = c_1 + c_2, \\ c_5 &= a_2 + b_2 + a_3 + b_3 = c_2 + c_3, \end{aligned}$$

$$c_6 = a_1 + b_1 + a_2 + b_2 + a_3 + b_3 = c_1 + c_2 + c_3.$$

Таким образом, три проверочных символа в слове определяются точно так же, как в **a** и в **b**, поэтому **c** также является кодовым словом.

Из указанного свойства вытекают два важных следствия. Одно из них - существование простой процедуры для построения групповых кодов. Второе состоит в существовании связи между расстояниями в групповом коде и его весовым спектром. Этот факт позволяет значительно упростить задачу построения хороших групповых кодов и вычисления их характеристик.

Рассмотрим свойства расстояния между кодовыми словами.

Расстояние $d(a, b)$ между двумя кодовыми словами **a** и **b** определяется как число позиций, в которых эти слова различаются. Вес $w(c)$ кодового слова определяется как число ненулевых элементов этого слова. Если рассмотреть посимвольную сумму по модулю 2 двух кодовых слов, то ее ненулевые символы соответствуют несовпадающим символам двух кодовых слов. Поэтому для любых двух кодовых слов **a** и **b** имеем

$$d(\mathbf{a}, \mathbf{b}) = w(\mathbf{a} + \mathbf{b}).$$

Отсюда следует, что множество расстояний от фиксированного кодового слова до всех других кодовых слов совпадает с множеством всех весов этого кода. То есть расстояние между двумя кодовыми словами совпадает с расстоянием от нулевого кодового слова до некоторого кодового слова. Таким образом, при построении группового кода с хорошим набором расстояний нужно стремиться к тому, чтобы веса ненулевых кодовых слов были возможно большими. Кроме того, при вычислении характеристик группового кода достаточно рассматривать лишь передачу нулевого кодового слова, поскольку расстояния между любыми другими кодовыми словами будут такими же.

Своим названием групповые коды обязаны тому, что множество кодовых слов вместе с нулевым словом, снабженное операцией посимвольного сложения по модулю 2, образует математическую структуру, называемую *группой*. Основные свойства группы таковы:

- сумма двух элементов группы всегда лежит в группе (замкнутость);
- выполняется закон ассоциативности, так что $(a + b) + c = a + (b + c)$;
- группа всегда содержит единичный элемент (нулевое слово);
- каждый элемент группы обладает обратным (в случае двоичного кода каждое слово совпадает со своим обратным), для которого $a + (-a) = 0$.

Коды с обобщенными проверками на четность, определенные в этом параграфе, характеризуются всеми четырьмя указанными свойствами. Обсуж-

даемые примеры кодов были определены над «полем» из двух элементов $\{0, 1\}$ с помощью арифметики по модулю 2. Можно, однако, определить не-двоичные коды, используя конечное поле из q элементов ($q \neq 2$). Все утверждения, сформулированные для двоичных кодов, справедливы и для недвоичных кодов, к которым относится код Рида-Соломона.

Проверочная матрица. Линейный код длины n с k информационными символами и $r=n-k$ проверочными символами обозначается (n, k) . Кодовую комбинацию такого кода можно записать

$$a = a_{n,\text{пр}} a_{n-1,\text{пр}} \dots a_{k+1,\text{пр}} a_k a_{k-1} a_{k-2} \dots a_1,$$

$r = n - k$, где $\{a_i\}$ - информационные символы; $\{a_{j,\text{пр}}\}$ - проверочные символы.

Проверочные символы a_{np} формируются по правилу:

$$a_{i,np} = \sum_{l=1}^k \gamma_{l,i} a_l, \tag{4.16}$$

где $i = k + 1, k + 2, \dots, k + r$; k - число информационных символов, r - число избыточных символов. Суммирование в (4.16) осуществляется по модулю 2 для двоичного кода.

Вместо того чтобы явно выписывать уравнения для проверок на четность, как это делали в примере (6,3)-кода, более удобно использовать матричные обозначения. Очень часто линейный код задают не соотношениями (4.16), а *проверочной* и *порождающей* матрицами.

Проверочная матрица представляет собой таблицу с размерами $r \times n$ ($r = n - k$ столбцов, n строк) вида

$$\mathbf{H} = \begin{pmatrix} \gamma_{1,k+1} \dots \gamma_{1,n} \\ \dots \dots \dots \\ \gamma_{k,k+1} \dots \gamma_{k,n} \\ 100 \dots \dots 0 \\ \dots \dots \dots \\ \dots \dots \dots \\ 000 \dots \dots 1 \end{pmatrix}. \tag{4.17}$$

r

Для каждой разрешенной кодовой комбинации a матричное произведе-

ние $a\mathbf{H}=0$. Таким образом, совокупность кодовых комбинаций - это множество последовательностей a , для которых $a\mathbf{H}=0$.

Например, для приведенного выше примера проверочная матрица \mathbf{H} содержит ту же информацию о коде, что и выписанное в общем виде кодовое слово \mathbf{a} :

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad (4.18)$$

$$\mathbf{a}^T = (a_1, a_2, a_3, a_1 + a_2, a_2 + a_3, a_1 + a_2 + a_3).$$

Каждый столбец матрицы \mathbf{H} соответствует некоторому символу кодового слова: первые три столбца - информационным символам, а последние три - проверочным. Соотношение, выражаемое первой строкой, состоит в том, что четвертый символ является суммой a_1 и a_2 . Аналогично вторая строка указывает, что пятый символ является суммой a_2 и a_3 и т. д. Эта матрица записана в канонической форме. Это значит, что первые k столбцов задают информационные символы, которые входят во все уравнения, в то время как последние $n - k$ столбцов образуют единичную матрицу. Проверочную матрицу можно записывать и различными другими способами, такая возможность часто оказывается полезной.

Формализовать приведенные рассуждения можно, сказав, что последовательность \mathbf{a} является кодовым словом в том, и только в том случае, если

$$\mathbf{H}\mathbf{a} = 0. \quad (4.19)$$

В этом уравнении матричное умножение выполняется обычным образом, с тем лишь исключением, что в случае двоичных кодов используется сложение по модулю 2.

Избыточность линейного двоичного кода

$$\rho_k = 1 - \frac{\log 2k}{n} = 1 - \frac{k}{n} = \frac{r}{n}. \quad (4.20)$$

где $r = n - k$ - число проверочных символов.

«Оптимальным» является код (n, k) , обеспечивающий наименьшую вероятность ошибочного декодирования среди всех кодов той же длины n и избыточности r/n .

Совершенные коды - это коды, которые всю свою избыточность расходуют на исправление ошибок заданной кратности t . *Квазисовершенным* называется код, который исправляет некоторую часть ошибок кратности $t+1$.

Обнаружение ошибок при использовании линейных кодов основано на проверке соотношений (4.16), при этом по принятой кодовой комбинации со-

ставляются контрольные суммы по модулю 2

$$a_{i,\text{пр}}^{\text{КОНТ}} = \sum_{l=1}^k \gamma_{l,i} a_l' \quad (4.21)$$

(здесь $i = k + 1, \dots, k + r$, a_l' – l -й знак принятой кодовой комбинации a') и сопоставляются с проверочными символами принятой кодовой комбинации.

Совокупность чисел $c_{j-k} = a_{j,\text{пр}}' + a_{j,\text{пр}}^{\text{КОНТ}} \pmod{2}$ для данной кодовой комбинации называется *синдромом*:

$$c(a') = c_r, \dots, c_3, c_2, c_1. \quad (4.22)$$

Декодирование принятой кодовой комбинации может быть осуществлено с помощью проверочной матрицы \mathbf{H} . Если принята комбинация a , то синдром c можно определить равенством $\mathbf{c} = \mathbf{a}'\mathbf{H}$. Таким образом, синдром – это вектор-строка (c_1, c_2, \dots, c_r) с r компонентами (по одной для каждого проверочного символа). Если a переданная кодовая комбинация, а a' – принятая, то сумма их по модулю 2, т.е. $\text{mod}_2(a + a') = z$ называется *шумовой последовательностью*. При этом $\mathbf{c} = \mathbf{z}\mathbf{H}$.

При правильном приеме все элементы синдрома равны нулю. Отличие хотя бы одного элемента синдрома от нуля означает, что произошла ошибка. При декодировании с исправлением ошибок по виду синдрома можно определить разряд кодовой комбинации, в котором произошла ошибка.

Коды Хемминга

Для каждого фиксированного кодового слова уравнение (4.19) означает, что сумма некоторого подмножества столбцов матрицы равна 0. Например, последовательность (1 0 0 1 0 1) является кодовым словом (6, 3)-кода из предыдущего примера. При матричном умножении согласно (4.19) ненулевые элементы этой последовательности «отсеивают» первый, четвертый и шестой столбцы \mathbf{H} , т. е.

$$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Аналогичное соотношение должно выполняться для любого кодового слова. Поэтому если минимальный вес, т.е. кодовое расстояние кода, равно d , то должно существовать по крайней мере одно подмножество, состоящее из d столбцов матрицы \mathbf{H} , сумма которых равна 0. С другой стороны, не может существовать ни одного подмножества из $d - 1$ или менее столбцов, сумма которых равна 0. Если рассматривать столбцы матрицы \mathbf{H} как векторы, то можно сказать, что для кода с кодовым расстоянием d все подмножества из $d - 1$ столбцов \mathbf{H} должны быть линейно независимы. Это утверждение составляет

одну из фундаментальных теорем о групповых кодах. Оно позволяет находить кодовое расстояние группового кода, заданного матрицей \mathbf{H} , а также строить матрицы \mathbf{H} , приводящие к кодам с гарантированным кодовым расстоянием. Вновь рассматривая в качестве примера $(6, 3)$ -код, видим, что все столбцы различны, так что сумма двух столбцов никогда не равна 0. С другой стороны, существует по крайней мере одно множество из трех столбцов, например, состоящее из столбцов 1, 2 и 5, сумма которых равна 0. Таким образом, кодовое расстояние этого кода равно 3.

Возникает вопрос - насколько длинным при заданном числе проверочных символов может быть код, имеющий гарантированное кодовое расстояние 3? Поскольку для получения кодового расстояния 3 все множества из двух столбцов должны быть линейно независимыми, достаточно, чтобы все столбцы были различными и ненулевыми. Таким образом, для трех проверочных символов имеется семь различных ненулевых троек, для четырех - пятнадцать различных ненулевых четверок и т. д. Это дает семейство кодов с параметрами (n, k) вида $(2^p - 1, 2^p - 1 - p)$, где $p = n - k$. Такие коды называются кодами Хемминга, впервые они были описаны Хеммингом в 1950 г.

Коды Хемминга обладают несколькими замечательными свойствами. Прежде всего они являются примерами немногих известных совершенных кодов. Заметим, что для совершенного кода в формуле (4.10) должно выполняться равенство, которое в случае кодов Хемминга имеет вид

$$2^{n-k} = 1 + n.$$

Поскольку для этих кодов $n = 2^{n-k} - 1$, последнее равенство, очевидно, справедливо.

Другое свойство кодов Хемминга состоит в том, что они принадлежат к одному из немногих классов кодов, для которых полностью известен спектр. Определим весовую функцию кода равенством

$$A(x) = \sum_{i=0}^n A_i x^i,$$

где A_i число кодовых слов веса i . Тогда для кода Хемминга с $d = 3$ и $n = 2^p - 1$, имеем

$$A(x) = \frac{1}{n+1} \left[(1+x)^n + n(1+x)^{(n-1)/2} (1-x)^{(n+1)/2} \right]. \quad (4.23)$$

Коды Хемминга с кодовым расстоянием 3 можно превратить в коды Хемминга с расстоянием 4, добавив дополнительный проверочный символ, равный сумме всех остальных символов. Этот дополнительный проверочный символ переводит все кодовые слова веса 3 в слова веса 4, слова веса 5 в слова

веса 6 и т. д. Если, например, применить эту процедуру к (7, 4)-коду Хемминга, то получится следующая дополненная проверочная матрица:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad (4.24)$$

Левая верхняя часть этой матрицы (ограниченная штриховыми линиями) совпадает с первоначальной проверочной матрицей и состоит из различных троек. Эти новые коды с кодовым расстоянием 4, полученные добавлением одного проверочного символа к кодам Хемминга, иногда называют *расширенными* кодами Хемминга. Их параметры имеют вид $(2^p, 2^p - 1 - p)$.

Дополненная матрица \mathbf{H} в последнем примере уже не имеет канонической формы, т. е. правая часть этой матрицы уже не является единичной матрицей, что не позволяет определить каждый проверочный символ только по информационным символам. Этот недостаток можно исправить, прибавив каждую из первых трех строк матрицы к последней. В результате получим матрицу вида

$$\mathbf{H}^1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Последняя строка полученной матрицы по-прежнему задает проверку на четность. Вычислим ее действие на кодовый вектор. Если $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4$ - четыре строки матрицы \mathbf{H} , то произведение последней строки \mathbf{H}^1 на кодовый вектор \mathbf{c} имеет вид

$$\mathbf{h}_4^1 \mathbf{c} = (\mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_4) \mathbf{c} = \mathbf{h}_1 \mathbf{c} + \mathbf{h}_2 \mathbf{c} + \mathbf{h}_3 \mathbf{c} + \mathbf{h}_4 \mathbf{c} = 0,$$

поскольку $\mathbf{h}_i \mathbf{c} = 0$ для всех i .

В общем случае строки проверочной матрицы являются линейно независимыми. Это становится очевидным, если матрица записана в канонической форме, поскольку наличие единичной матрицы делает невозможным существование линейной комбинации с нулевой суммой. Из приведенного примера ясно, что существует много допустимых способов выбора строк матрицы \mathbf{H} . Любая линейная комбинация строк может служить проверочным уравнением, и любое множество из $n - k$ полученных таким способом линейно независимых уравнений может быть использовано для образования матрицы \mathbf{H} . Это

свойство используется для построения простых алгоритмов декодирования групповых кодов.

Порождающая матрица. До сих пор групповые коды описывались в терминах проверочной матрицы \mathbf{H} . Записанная в канонической форме, такая матрица позволяет выразить каждый проверочный символ в виде: линейной комбинации информационных символов. Часто полезным оказывается другое возможное описание. Напомним, что сумма по модулю 2 любых двух кодовых слов снова является кодовым словом. Несколько раз используя это свойство, получаем, что любая линейная комбинация кодовых слов (при сложении по модулю 2) также является кодовым словом. Поскольку информационные символы выбираются независимо, можно предполагать, что существуют кодовые слова, каждое из которых содержит ровно один символ 1 в информационной части кодового слова. Тогда все 2^k кодовых слов можно будет получить как 2^k возможных линейных комбинаций этих k базисных, векторов. Мы приписываем коду свойства *линейного векторного пространства*. Следующий пример показывает, что это действительно так.

Проверочная матрица (7, 4)-кода Хемминга имеет вид

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Мы хотим найти четыре различных кодовых вектора, каждый из которых содержит единственный символ 1 в первых четырех позициях. Предположим, что первый из этих векторов имеет вид

$$\mathbf{c}_1^T = (1 \ 0 \ 0 \ 0 \ p_{11} \ p_{12} \ p_{13}).$$

Умножая этот вектор на каждую из строк матрицы \mathbf{H} , находим, что p_{11}, p_{12}, p_{13} должны быть элементами ее первого столбца. Аналогично, полагая

$$\mathbf{c}_2^T = (0 \ 1 \ 0 \ 0 \ p_{21} \ p_{22} \ p_{23}),$$

получаем, что p_{21}, p_{22}, p_{23} совпадают с элементами второго столбца \mathbf{H} и т. д. Таким образом, проверочная часть каждого из четырех базисных векторов совпадает с одним из первых четырех столбцов матрицы \mathbf{H} . Если сформировать из этих четырех векторов матрицу, то получим так называемую *порождающую* матрицу кода. В канонической форме она всегда состоит из единичной матрицы порядка $k \times k$, к которой присоединена $k \times (n - k)$ - матрица проверочных символов. Проверочная часть порождающей матрицы получается из матрицы \mathbf{H} (в канонической форме) транспонированием подматрицы, образованной первыми k столбцами.

Итак, матрица, которая имеет k строк, образованных этими k линейно-независимыми кодовыми комбинациями, называется *порождающей*, или *производящей* матрицей. Размерность порождающей матрицы линейного кода $k \times n$ (k строк и n столбцов). При учете (4.21) порождающую матрицу можно записать

$$\mathbf{G} = \left\| \begin{array}{cc} & n = k + r \\ \underbrace{\left[\begin{array}{c} 100\dots 0 \\ 010\dots 0 \\ \dots \\ \dots \\ 00\dots 1 \end{array} \right]}_k & \underbrace{\left[\begin{array}{c} \gamma_{1,k+1} \dots \gamma_{1,k+r} \\ \gamma_{2,k+1} \dots \gamma_{2,k+r} \\ \dots \\ \dots \\ \gamma_{k,k+1} \dots \gamma_{k,k+r} \end{array} \right]}_r \end{array} \right\| = \left\| \mathbf{1}_k \mathbf{G}' \right\|, \quad (4.25)$$

где $\mathbf{1}_k$ - единичная матрица порядка k , а \mathbf{G}' - матрица коэффициентов размерности $k \times r$.

Таким образом, в приведенном примере

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Теперь должно быть ясно, как связаны между собой проверочная и порождающая матрицы произвольного кода, двоичного или недвоичного. Если порождающая матрица имеет вид

$$\mathbf{G} = [\mathbf{I}_k, \mathbf{P}],$$

где $\mathbf{P} - k \times (n - k)$ - матрица проверочных символов, то проверочная матрица

$$\mathbf{H} = [-\mathbf{P}^T, \mathbf{I}_{n-k}].$$

Таким образом, по порождающей матрице, записанной в канонической форме, сразу можно получить проверочную матрицу и наоборот.

Полиномиальные коды

Основные сведения о конечных полях (полях Галуа). Полиномиальные коды выделяются как подкласс в классе всех групповых кодов. Примерами полиномиальных кодов являются коды Боуза - Чоудхури - Хоквингема (БЧХ),

коды Рида – Соломона (РС), обобщенные коды Рида - Маллера, проективно-геометрические, евклидово-геометрические и квадратично-вычетные коды. Описание каждого из этих семейств кодов задается указанием алгоритма для его построения. Указанные семейства кодов пересекаются, так что некоторый конкретный код может быть одновременно кодом БЧХ и вычетным кодом или обобщенным кодом Рида - Маллера и кодом БЧХ и т. д.

Важность полиномиальных кодов обусловлена несколькими причинами. Во-первых, аппаратная реализация кодеров для таких кодов требует лишь сравнительно простых регистров сдвига с обратными связями. Во-вторых, это семейство содержит много кодов, кодовое расстояние которых близко к наилучшему, в особенности при длине блока около 100 или менее. В-третьих, существует несколько алгоритмов декодирования, позволяющих декодировать некоторые из этих кодов с помощью сравнительно несложных устройств.

В предыдущих разделах кодовое слово (n, k) - кода представлялось в виде: набора длиной n :

$$(a_0, a_1, \dots, a_{n-1}).$$

Другой способ представления того же кодового слова состоит в том, чтобы считать элементы a_0, a_1, \dots, a_{n-1} коэффициентами многочлена от x . Таким образом,

$$f(x) = a_0 + a_1x + \dots + a_{n-1}x^{n-1}.$$

Используя это обозначение, можно определить *полиномиальный код* как множество всех многочленов степени, не большей $n - 1$, содержащих в качестве множителя некоторый фиксированный многочлен $g(x)$. Многочлен $g(x)$ называется *порождающим многочленом* кода. Для того чтобы иметь возможность умножать такие кодовые многочлены, разлагать их на множители и производить над ними другие операции, нужно уметь складывать, вычитать, умножать и делить их коэффициенты. Этого легко добиться, если потребовать, чтобы все коэффициенты были элементами некоторого конечного поля.

Арифметика конечных полей

Конечное поле, называемое также *полем Галуа* и обозначаемое $GF(q)$, где q порядок поля, - это конечное множество, состоящее из q элементов, в котором определены правила для выполнения арифметических операций. Эти правила не очень отличаются от тех, которые используются при арифметических операциях с обычными числами. Основное отличие состоит в том, что в конечном поле все операции производятся над конечным числом элементов, и для учета этого обстоятельства нужно внести некоторые изменения.

Все конечные поля обладают следующими свойствами:

1. существуют две операции, используемые для комбинирования эле-

- ментов: умножение и сложение;
2. результатом умножения или сложения двух элементов поля является третий элемент, лежащий в том же поле;
 3. поле всегда содержит мультипликативную единицу 1 и аддитивную единицу 0; таким образом, $a + 0 = a$ и $a \cdot 1 = a$ для любого элемента a ;
 4. для любого элемента a существует обратный элемент по сложению ($-a$) и обратный элемент по умножению a^{-1} (если $a \neq 0$), такие что $a + (-a) = 0$ и $a \cdot a^{-1} = 1$; существование этих элементов позволяет использовать обычные обозначения для вычитания и деления;
 5. выполняются обычные правила:
 - ассоциативности $[a + (b + c) = (a + b) + c, \quad a \cdot (b \cdot c) = (a \cdot b) \cdot c]$,
 - коммутативности $[a + b = b + a, \quad a \cdot b = b \cdot a]$,
 - дистрибутивности $[a \cdot (b + c) = a \cdot b + a \cdot c]$.

Конечные поля существуют не при любом числе элементов, а только в случае, если число элементов является простым числом или степенью простого числа. В первом случае поле называется *простым*, во втором — *расширенным* соответствующего простого поля.

Для каждого допустимого значения q существует ровно одно поле. Другими словами, правила сложения и умножения, удовлетворяющие всем нужным требованиям, можно задать только одним способом. Если q - простое число, то элементами поля являются числа $0, 1, \dots, q-1$, а сложение и умножение являются обычными сложением и умножением по модулю q .

Наименьшим простым полем Галуа является поле $GF(2)$, состоящее из двух элементов 1 и 0. Здесь арифметические операции выполняются по mod 2.

Пример: таблица сложения и умножения над $GF(2)$ имеют вид:

\oplus	0	1
0	0	1
1	1	0

\otimes	0	1
0	0	0
1	0	1

Аналогично для $GF(3)$:

\oplus	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

\otimes	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Таблицы сложения и умножения для простого поля $GF(5)$ объединены в **таблице 4.3**. Для выполнения вычитания или деления следует, используя таблицы, найти соответствующий обратный элемент, а затем выполнить сложение или умножение обычным образом. Например, $3 - 4 = 3 + (-4) = 3 + 1 = 4$.

Аналогично $3/4 = 3 \cdot 4^{-1} = 3 \cdot 4 = 2$.

Таблица 4.3 – Таблицы умножения и сложения для $GF(5)$

\otimes	0	1	2	3	4		+	0	1	2	3	4
0	0	0	0	0	0		0	0	1	2	3	4
1	0	1	2	3	4		1	1	2	3	4	0
2	0	2	4	1	3		2	2	3	4	0	1
3	0	3	1	4	2		3	3	4	0	1	2
4	0	4	3	2	1		4	4	0	1	2	3

Если q является степенью простого числа (например, $q = p^m$, p - простое число, m - любое положительное число), то поле называется *расширенным*, а элементами поля являются все p^m многочленов степени $(m - 1)$ или менее, коэффициенты которых лежат в простом поле $GF(p)$. Расширенное поле $GF(p^m)$ является расширением простого поля $GF(p)$.

Например, элемент 10110 поля $GF(2^5)$ может быть представлен многочленом вида $x^4 + x^2 + x$.

Правила умножения и сложения элементов расширенного поля (многочленов) получаются из обычного умножения и сложения многочленов и последующего приведения результата по модулю некоторого специального многочлена $p(x)$ степени m . Этот многочлен обладает тем свойством, что его нельзя разложить на множители, используя только многочлены с коэффициентами из $GF(p)$. Такие многочлены называются *неприводимыми*, они аналогичны простым числам. Как и простые числа, неприводимые многочлены обычно находят методом перебора и приводятся в таблицах.

Приведение многочлена по модулю $p(x)$ эквивалентно делению на $p(x)$ и взятию остатка. С этим тесно связано понятие сравнимости:

$$a(x) \equiv b(x) \pmod{p(x)},$$

что читается как « $a(x)$ сравнимо с $b(x)$ по модулю $p(x)$ » и означает, что $a(x)$ и $b(x)$ имеют одинаковые остатки при делении на $p(x)$ или что $a(x) - b(x)$ делится на $p(x)$. Все эти понятия имеют смысл и в случае, когда a , b и p являются целыми числами.

При сложении соответствующих многочленов (элементов расширенного поля) коэффициенты при одинаковых степенях x складываются по модулю порядка расширяемого поля $GF(2)$, т. е. по $mod 2$.

При умножении результаты приводятся по модулю неприводимого многочлена $p(x)$ по правилу

$$a(x) \times c(x) \equiv b(x) \pmod{p(x)},$$

где многочлен $p(x)$ выбирается из разложения функции

$$x^{q-1} + 1 = x^{2^m-1} + 1$$

на простые множители.

В конечных полях, как для обычных чисел, можно ввести понятие логарифма. Все конечные поля обладают тем свойством, что существует по крайней мере один элемент, называемый генератором или *примитивным* элементом, обладающий тем свойством, что любой другой ненулевой элемент поля является некоторой степенью этого элемента.

Обычно в качестве примитивного элемента α выбирают $x = 0\dots 010$.

Построим поле $GF(8) = GF(2^3)$, $q=8$, $m=3$. Разложение соответствующей функции $x^{q-1}+1$ имеет вид

$$x^7+1=(x+1)(x^3+x+1)(x^3+x^2+1).$$

Многочлен $p(x) = 1 + x + x^3$ неприводим над $GF(2)$ и может быть использован для построения $GF(8)$. В этом случае элементы поля $GF(2^3)$ равны:

- нулевой (0) - в виде: двоичного числа - 000 (0_{10});
- единичный (1) - в виде: двоичного числа - 001 (1_{10});
- примитивный (α) - в виде: двоичного числа - 010 (2_{10});
- следующий (α^2) - в виде: двоичного числа - 100 (4_{10}), где многочлен имеет вид x^2 .
- следующий (α^3) - в виде: двоичного числа - 1000. Здесь для представления требуется уже четыре разряда, а в нашем распоряжении три разряда, т.е. элемент находится за пределами поля. Следовательно, необходимо осуществить операцию приведения результата по модулю многочлена $p(x)$:

$$\begin{array}{r|l} x^3 & x^3+x+1 \\ x^3+x+1 & \hline x+1 & 1 \end{array}$$

Отсюда α^3 представляется в виде: многочлена вида $x+1$, что эквивалентно $011(3_{10})$

Аналогичным образом находятся выражения для α^4 , α^5 , α^6 .

При этом $\alpha^7 = \alpha^0 = 1$ или в общем случае $\alpha^{i+7} = \alpha^i$.

В более общем случае $\alpha^{i+q-1} = \alpha^i$.

Таким образом, мы нашли все элементы поля $GF(2^3)$, как расширение поля $GF(2)$ по модулю многочлена x^3+x+1 . Элементы этого поля приведены в таблице 4.4.

Таблица 4.4

Степени	Многочлены	Двоичное представление	Десятичное представление
нулевой элемент	--	000	0
α^0	1	001	1
α^1	x	010	2
α^2	x^2	100	4
α^3	$x+1$	011	3
α^4	x^2+x	110	6
α^5	x^2+x+1	111	7
α^6	x^2+1	101	5

Показатель степени α называется *индексом*.

Из анализа данных таблицы следует, что умножение элементов можно свести к сложению соответствующим образом их индексов, а деление - к вычитанию индексов.

$$\text{Пример. } \alpha^4 * \alpha^5 = \alpha^9 = \alpha^{2+7} = \alpha^2$$

$$\alpha^4 : \alpha^5 = \alpha^{4-5} = \alpha^{-1} = \alpha^{-1+7} = \alpha^6$$

То есть видно, что операция умножения и деления в конечных полях понимаются в отличном от обычного арифметического смысле, действительно (в скобках - десятичное представление элементов поля):

$$\alpha^4(6) * \alpha^5(7) = \alpha^2(4) \quad \alpha^4(6) : \alpha^5(7) = \alpha^6. \quad (4.26)$$

Пусть $\beta_1 = 1 + x + x^2$ и $\beta_2 = 1 + x^2$ - два элемента $GF(8)$. Можно образовать их сумму

$$\beta_1 + \beta_2 = (1 + x + x^2) + (1 + x^2) = (1+1) \cdot 1 + x + (1+1) \cdot x^2 = x$$

и их произведение

$$\begin{aligned} \beta_1 \cdot \beta_2 &= (1 + x + x^2) \cdot (1 + x^2) = 1 + x + x^2 + x^2 + x^3 + x^4 = \\ &= 1 + x + x^3 + x^4. \end{aligned}$$

Для того чтобы привести это выражение по модулю $p(x)$, будем последовательно использовать равенство $p(x)=0$. Так,

$$x^3 = 1 + x, \quad x^4 = x + x^2$$

и

$$\beta_1 \cdot \beta_2 = 1 + x + (1 + x) + (x + x^2) = x + x^2.$$

Из таблицы 4.3 видно, что 2 является примитивным элементом в поле $GF(5)$, поскольку, умножая по модулю 5, имеем

$$2^1 = 2, 2^2 = 4, 2^3 = 2 \times 2 \times 2 = 4 \times 2 = 3, 2^4 = 1, 2^5 = 2^1 = 2. \quad (4.27)$$

Таким образом, способ выполнения умножения состоит в том, чтобы с помощью отображения, задаваемого соотношением (4.27), найти логарифмы, сложить их по модулю 4 и снова использовать (4.27) для нахождения антилогарифмов. Заметим, что при этом, как и для обычных чисел, нужно положить $2^0=1$.

Если так же поступить в случае поля $GF(8)$ и положить $\alpha = x$, то

$$\begin{aligned} \alpha^2 &= x^2, \\ \alpha^3 &= x^3 = 1 + x, \\ \alpha^4 &= x + x^2, \\ \alpha^5 &= x^2 + x^3 = 1 + x + x^2, \\ \alpha^6 &= x + x^2 + x^3 = 1 + x^2, \\ \alpha^7 &= x + x^3 = 1 \text{ и т.д.} \end{aligned} \tag{4.28}$$

Таким образом, можно перемножить два элемента $GF(8)$, определив их логарифмы из (4.28), сложив эти логарифмы по модулю 7 и снова используя (4.28) для нахождения антилогарифма. В этом примере элемент x оказался примитивным, так что построение таблицы логарифмов было несложным. К сожалению, это не всегда так и нахождение примитивного элемента может быть весьма нетривиальной задачей. Для практических применений полезно отметить, что x будет примитивным элементом, если неприводимый многочлен, задающий операции в поле, является также примитивным многочленом. В дальнейшем связь между примитивными многочленами и примитивными элементами станет ясной. Здесь достаточно отметить, что таблицы неприводимых многочленов обычно содержат сведения о том, какие из многочленов являются примитивными. Таким образом, на практике всегда можно выбрать неприводимый многочлен, который является также примитивным, и избежать возможных затруднений.

Существование логарифмов в конечном поле означает, что имеется представление элементов конечного поля, которое оказывается удобным для *умножения*, и другое представление, которое оказывается удобным для *сложения*. После того как установлено соответствие между этими представлениями, уже не нужно явно выписывать элементы поля в виде: многочленов. Наиболее удобный метод состоит в представлении элемента поля в виде: набора длиной t при выполнении сложения и в виде: степени примитивного элемента при выполнении умножения. Так, для $GF(8)$ представление в таблице 4.5 эквивалентно представлению, задаваемому уравнениями (4.28). Используя таблицу 4.5, легко производить арифметические операции в поле $GF(8)$.

Таблица 4.5 – Связь обычного и логарифмического представлений элементов $GF(8)$

0	$\leftrightarrow 0 0 0$
$\alpha^0 = 1$	$\leftrightarrow 1 0 0$
α^1	$\leftrightarrow 0 1 0$
α^2	$\leftrightarrow 0 0 1$
α^3	$\leftrightarrow 1 1 0$
α^4	$\leftrightarrow 0 1 1$
α^5	$\leftrightarrow 1 1 1$
α^6	$\leftrightarrow 1 0 1$

Пусть, например, требуется вычислить значение

$$f(\alpha) = (\alpha^2 + \alpha^5)\alpha + (1 + \alpha^3)\alpha.$$

Из таблицы 4.5 получаем $\alpha^2 + \alpha^5 = \alpha^3$ и $1 + \alpha^3 = \alpha$. Поэтому

$$f(\alpha) = \alpha^3 \cdot \alpha + \alpha \cdot \alpha = \alpha^4 + \alpha^2 = \alpha = (0 1 0).$$

В некоторых случаях производить многочисленные переводы логарифмического представления в представление в виде: наборов длиной m оказывается неудобным. Другой очень полезный подход состоит в использовании логарифма Зеча $z(n)$, задаваемого равенством

$$\alpha^{z(n)} = 1 + \alpha^n. \quad (4.29)$$

Используя этот метод, можно выполнять арифметические операции в конечном поле, работая только с логарифмами и не обращаясь к антилогарифмам. Сумма элементов α^m и α^n может быть вычислена в виде:

$$\alpha^m + \alpha^n = \alpha^m (1 + \alpha^{n-m}) = \alpha^{m+z(n-m)}.$$

Логарифмы Зеча для $GF(8)$ приведены в таблице 4.6. Используя эту таблицу, получаем

$$\alpha^2 + \alpha^5 = \alpha^{2+z(3)} = \alpha^3.$$

Таблица 4.6 – Логарифмы Зеча в $GF(8)$

N	$-\infty$	0	1	2	3	4	5	6
$z(n)$	0	$-\infty$	3	6	1	5	4	2

Построение полиномиальных кодов

Ранее полиномиальный (n, k) -код был определен как множество всех многочленов степени $n-1$ или меньше, делящихся на $g(x)$. Степень $g(x)$ равна $n - k$. Например, если

$$g(x) = 1 + x + x^3,$$

то множество многочленов вида

$$c(x) = (1 + x + x^3)(a_0 + a_1x + a_2x^2)$$

образует (6,3)-код. Существует ровно восемь различных многочленов в соответствии с восемью различными способами выбора коэффициентов a_0, a_1, a_2 в поле GF (2). Один из способов применения этого кода состоит в том, чтобы взять a_0, a_1, a_2 в качестве информационных символов и выполнить соответствующие умножения. Результат будет

$$c(x) = a_0 + (a_0 + a_1)x + (a_1 + a_2)x^2 + (a_0 + a_2)x^3 + a_1x^4 + a_2x^5. \quad (4.30)$$

Устройство, производящее это умножение, показано на рисунке 4.4. Хотя такой метод кодирования вполне пригоден, его недостаток состоит в том, что полученный код обычно оказывается несистематическим. (В данном примере код является систематическим в позициях 1, 5, 6; однако в общем случае это не так.)

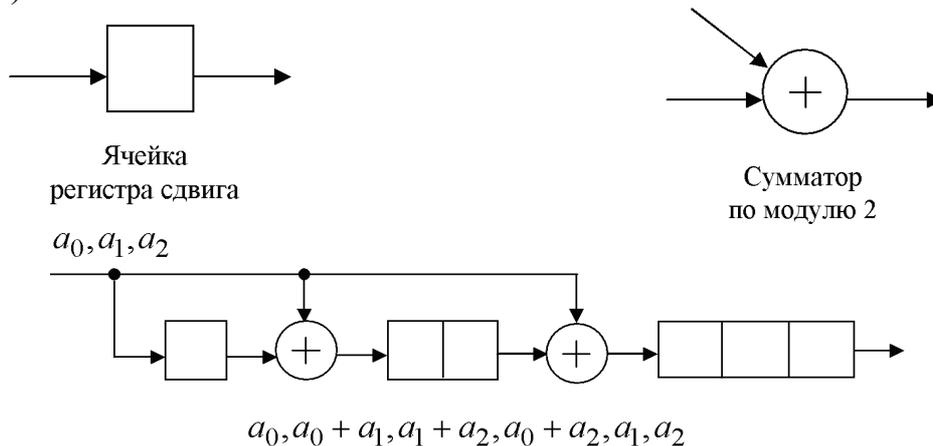


Рисунок 4.4 – Схема для умножения произвольного многочлена

Полиномиальные (n, k) -коды обладают тем свойством, что их всегда можно сделать систематическими в первых k позициях. В рассматриваемом примере положим

$$b_0 = a_0, \quad b_1 = a_0 + a_1, \quad b_2 = a_1 + a_2. \quad (4.31)$$

Решив систему уравнений относительно a_i и подставив результат в (4.30), получим кодовый многочлен вида

$$c(x) = b_0 + b_1x + b_2x^2 + (b_1 + b_2)x^3 + (b_0 + b_1)x^4 + (b_0 + b_1 + b_2)x^5.$$

Этот кодовый многочлен имеет обычный вид кода с проверками на четность, и можно заметить, что если поменять местами первый и второй прове-

рочные символы, то полученный код будет совпадать с (6, 3)-кодом с проверками. С тем же успехом можно выбрать и другое множество информационных символов. Например, выбрав последние три символа, получим

$$c(x) = (b_0 + b_2) + (b_0 + b_1 + b_2)x + (b_1 + b_2)x^2 + b_0x^3 + b_1x^4 + b_2x^5. \quad (4.32)$$

Как построить кодер для каждого из систематических вариантов полученного кода. Вначале нужно выразить элементы a_i , в виде: линейных комбинаций элементов b_i , а затем использовать схему на рисунке 4.4 для осуществления соответствующих умножений.

Второй метод, который, по существу, является объединением этих двух шагов, состоит в следующем. Предположим, что умножаем многочлен $b_0 + b_1x + b_2x^2$ на x^3 , получая $b_0x^3 + b_1x^4 + b_2x^5$. Разделив полученный многочлен на $1 + x + x^3$, найдем частное $(b_0 + b_2) + b_1x + b_2x^2$ и остаток $(b_0 + b_2) + (b_0 + b_1 + b_2)x + (b_1 + b_2)x^2$. Прибавив остаток к делимому, получим кодовое слово (4.32) в систематическом виде. Эта процедура состоит в применении алгоритма Евклида, согласно которому делимое = (частное) (делитель) + остаток.

В случае общего (n, k) -кода предварительное умножение всегда производится на x^{n-k} . Степень остатка всегда меньше степени делителя, которая равна $n-k$. Поскольку кодируемая последовательность умножалась на x^{n-k} , в делимое входят лишь члены степени, не меньшей $n-k$. Поэтому добавление остатка к частному не изменяет ни одного из членов в обоих многочленах. Выбирая в качестве делителя порождающий многочлен, можно добиться того, что результатом всегда будет кодовое слово. Схема, реализующая этот алгоритм для (6, 3)-кода, показана на рисунке 4.5.

Содержимое ячеек			Сдвиг
0	0	0	0
b_2	b_2	0	1
b_1	$b_1 + b_2$	b_2	2
$b_0 + b_2$	$b_0 + b_1 + b_2$	$b_1 + b_2$	3

Рисунок 4.5 – Схема для одновременного умножения на x^3 и деления на $1 + x + x^3$

При использовании этой схемы последовательность b_2, b_1, b_0 подается на вход по одному символу за такт. Содержимое различных ячеек после каждого

сдвига показано на рисунке 4.5. После трех сдвигов остаток (три проверочных символа) запоминается в трех ячейках регистра сдвига. Произведя длинные операции деления, читатель может удостовериться в том, что согласно этой схеме операции сдвига и вычитания повторяются на каждом шаге. При необходимости частное может быть получено в виде: последовательных членов в цепи обратной связи.

Схема, показанная на рисунке 4.5, очень проста, поскольку все арифметические операции выполняются в поле GF (2). В общем случае кода над полем GF (q) следует выполнять также операции умножения. Предположим, что порождающий многочлен кода имеет вид $g(x) = g_{n-k}x^{n-k} + \dots + g_1x + g_0$, где коэффициенты g_i лежат в GF (q). Тогда схема на рисунке 4.5 должна быть изменена следующим образом. Прежде всего, все операции сложения должны быть заменены на сложение в GF (q) и каждая ячейка регистра сдвига должна содержать элемент из GF (q). Далее, результат на выходе сумматора, формирующего член обратной связи, нужно умножить на g_{n-k}^{-1} . Наконец, для каждого ненулевого $g_i, i < n - k$, член обратной связи должен быть умножен на g_i . Сумма полученного произведения с результатом выхода предыдущей ячейки является входным значением i -й ячейки регистра (i меняется от 0 до $n-k-1$). Полученная обобщенная схема кодирования показана на рисунке 4.6. Производя деление, легко показать, что эта схема дает требуемый остаток.

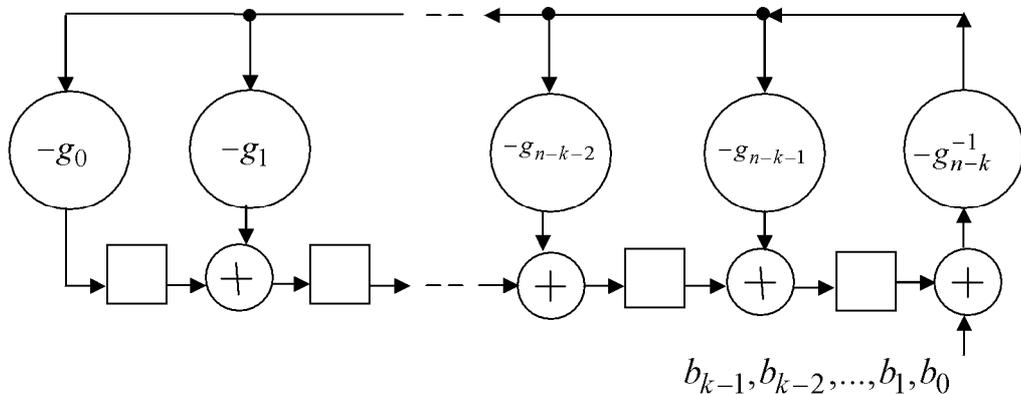


Рисунок 4.6 – Схема для умножения на x^{n-k} и деления на $g(x)$

Полезной может оказаться еще одна точка зрения. Изменим схему, изображенную на рисунке 4.5 таким образом, чтобы в самой левой ячейке появился символ 1. Производя в схеме несколько сдвигов и не подавая ничего на ее вход, будем получать содержимое регистров, показанное на рисунке 4.7.

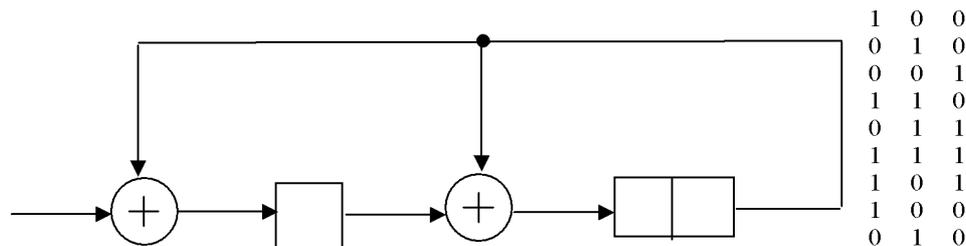


Рисунок 4.7 – Регистр обратной связи и соответствующая последовательность состояний при поступлении на вход одного символа 1

Можно заметить, что полученные тройки совпадают со столбцами проверочной матрицы, и если остановиться после семи шагов, то полученная матрица будет совпадать с проверочной матрицей (7,4)-кода Хемминга. Сделаем меньше семи шагов, получим проверочную матрицу укорочения этого кода. Заметим также, что, произведя больше семи сдвигов, получим уже встречавшиеся тройки. Полученную матрицу по-прежнему можно рассматривать как проверочную матрицу кода, однако кодовое расстояние теперь уменьшилось до 2, поскольку один и тот же столбец повторился несколько раз. Работу согласно схеме на рисунке 4.5 можно также рассматривать с помощью порождающей матрицы. Порождающая матрица (7, 4)-кода Хемминга имеет вид

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Если подать на вход символ 1 и сделать один сдвиг, то в соответствующих ячейках появится набор 1 1 0. После второго сдвига появится набор 0 1 1, после третьего 1 1 1 и после четвертого 1 0 1. Эти четыре тройки совпадают с четырьмя строками проверочной части порождающей матрицы. При вводе в регистр последовательности 1 0 0 0 (начиная с самого правого символа) получим проверочную последовательность 1 1 0. Аналогично при вводе в регистр последовательности 0 0 0 1 появятся проверочные символы 1 0 1. Кроме того, рассматриваемое устройство является линейным, так что при вводе в регистр последовательности 1 0 0 1 результатом будет 0 1 1. Эти проверочные символы возникают при сложении первой и четвертой строк порождающей матрицы. На практике схема, изображенная на рисунке 4.5, используется так, как показано на рисунке 4.8.

Кодируемая последовательность подается одновременно в канал и в кодер с включенной цепью обратной связи. Затем цепь обратной связи отключается и содержимое регистра подается в канал.

Циклические коды

При некоторых значениях n полиномиальные коды обладают свойством цикличности [1]. Это значит, что циклическая перестановка символов кодового слова вновь приводит к кодовому слову. Линейный двоичный код, которому принадлежат кодовые комбинации, полученные путем циклической перестановки символов, называется *циклическим*.

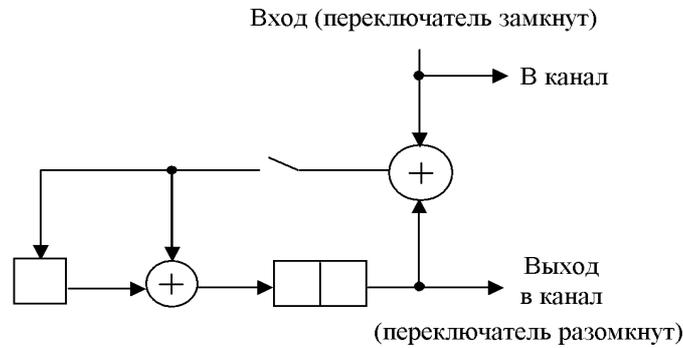


Рисунок 4.8 – Кодер для кода, порожденного многочленом $g(x) = 1 + x + x^3$ с использованием регистра сдвига с $n-k$ ячейками

Кодовый вектор a циклического кода представляют полиномом $(n-1)$ степени: $a(x) = \alpha_0 + \alpha_1x + \alpha_2x^2 + \dots + \alpha_{n-1}x^{n-1}$, где коэффициенты α_i принимают значения 0 или 1. При таком представлении кодовый вектор, полученный из \vec{a} циклической перестановкой элементов, можно рассматривать как результат умножения полинома $a(x)$ на x , если считать, что $x^n = 1$.

Причина наличия свойства цикличности может быть проиллюстрирована на примере уже рассмотренного порождающего многочлена $1 + x + x^3$. Заметим прежде всего, что $1 + x + x^3$ точно делит многочлен $1 - x^7$. Это можно проверить непосредственным делением или с помощью схемы на рисунке 4.8. Напомним, что эта схема порождает цикл длиной 7. Поэтому, если подать на вход символ 1, осуществить сдвиг семь раз и затем подать еще один символ 1 (т. е. разделить $x^7 - 1$), второй входной символ сократит содержимое регистра и в качестве остатка получится нулевой многочлен.

Предположим теперь, что взято кодовое слово $(7, 4)$ -кода, порожденного многочленом $1 + x + x^3$, и сдвинуто циклически на один элемент вправо. Это значит, что каждый символ передвинут на одно место вправо и самый правый символ переведен на левый конец. Математически это эквивалентно умножению на x и замене x^7 на 1, т. е. приведению многочлена по модулю $x^7 - 1$. Предположим, что выбрано кодовое слово, самый правый символ которого равен 1. Это кодовое слово может быть записано в виде:

$$c(x) = (1 + x + x^3)p(x),$$

где $p(x)$ - многочлен степени 3. Тогда

$$xc(x) \equiv xc(x) - (x^7 - 1) \pmod{(x^7 - 1)}.$$

Поскольку $x^7 - 1$ делится на $1 + x + x^3$, многочлен в правой части этого равенства очевидно делится на $1 + x + x^3$. Поэтому циклический сдвиг данного кодового слова снова дает кодовое слово. Заметим, что если самый правый символ кодового слова равен 0, то циклический сдвиг эквивалентен умножению на x и наше утверждение по-прежнему справедливо.

Полученный результат легко обобщить на произвольный (n, k) -код, для которого $g(x)$ делит $x^n - 1$. Наименьшее значение n , для которого $g(x)$ делит $x^n - 1$, обычно задает наибольшую полезную длину n_{\max} кода. Причина этого состоит в том, что два информационных символа, расположенные друг от друга на расстоянии $n+1$, приведут к одним и тем же проверочным символам, так что полученный код будет характеризоваться кодовым расстоянием 2. Заметим также, что наибольшее возможное значение n_{\max} равно $2^m - 1$, где m - степень порождающего многочлена. Это вытекает из того, что n_{\max} определяется как самый короткий цикл, который производится регистром сдвига, задаваемым $g(x)$, когда на вход его поступает единственный символ 1. Поскольку этот регистр содержит m ячеек, он может находиться не более чем в $2^m - 1$ различных ненулевых состояниях. Поэтому наибольшее значение $n_{\max} = 2^m - 1$. Для некоторых многочленов циклы могут быть существенно короче, чем $2^m - 1$. Читатель может легко проверить и убедиться, например, что многочлен $1 + x^3 + x^6$ порождает цикл длиной 9.

Зная порождающий полином $g(x)$ (полином наименьшей степени среди всех полиномов, соответствующих кодовым комбинациям циклического кода), можно построить все кодовые комбинации циклического кода, а также устройства кодирования и декодирования. На порождающий полином должен делиться без остатка двучлен $x^n \pm 1$. Полученный результат определяет проверочный полином

$$h(x) = \frac{x^n - 1}{g(x)}. \quad (4.33)$$

Порождающий полином $g(x)$ и проверочный полином $h(x)$ являются ортогональными, так как при $x^n - 1 = 0$ они удовлетворяют условию

$$g(x)h(x) = 0. \quad (4.34)$$

Невыполнение условия (4.34), т. е. $h(x)g(x) \neq 0$, является признаком ошибки.

Полиномиальные коды, для которых $n < n_{\max}$, как, например, рассмотренный (6,3)-код, называются *укороченными циклическими* или *псевдоциклическими* кодами. Для любого значения n всегда можно найти некоторый многочлен степени n , делящийся на $g(x)$. Используя этот многочлен, можно рассмотреть класс перестановок, получающихся при умножении кодового слова на x и приведении результата по модулю этого многочлена. Хотя эта процедура, очевидно, приводит к кодовому слову, она не находит до сих пор существенного применения. Обычно укороченные циклические коды рассматриваются как циклические путем добавления в конце кодового слова подходящего числа нулевых символов.

Коды БЧХ

Коды БЧХ представляют собой обобщенные коды Хемминга, позволяющие исправлять кратные ошибки. Эти коды проще всего описать с помощью корней порождающих многочленов.

Определение. Прimitивный код БЧХ, исправляющий t ошибок, - это блочный код длиной $n = q^m - 1$ над полем $GF(q)$, для которого элементы $\alpha^{m_0}, \alpha^{m_0+1}, \dots, \alpha^{m_0+2t-1}$ (для произвольного m_0) являются корнями порождающего многочлена $g(x)$: здесь α - примитивный элемент $GF(q^m)$.

Таким образом, порождающий многочлен кода БЧХ можно записать в виде:

$$g(x) = \text{НОК} \left[m_{m_0}(x), m_{m_0+1}(x), \dots, m_{m_0+2t-1}(x) \right], \quad (4.35)$$

где $\text{НОК} [.]$ - операция отыскания наименьшего общего кратного.

Коды с $m_0 = 1$ называются кодами БЧХ в узком смысле. Непрimitивные коды БЧХ определяются аналогично, но α заменяется на непрimitивный элемент β поля $GF(q^m)$, и длина блока становится равной порядку β . Список порождающих многочленов для примитивных и непрimitивных двоичных кодов БЧХ известен и приведен в известных источниках по кодированию.

Пример. Пусть нужно найти порождающий многочлен для кода БЧХ в узком смысле, исправляющего две ошибки и имеющего длину 7. Для этого корнями $g(x)$ должны быть элементы $\alpha, \alpha^2, \alpha^3, \alpha^4$, где α - примитивный элемент $GF(8)$. Пусть $GF(8)$ реализовано, как в таблице 4.5. Вычисление минимальных функций требуемых корней дает

$$m_1(x) = m_2(x) = m_4(x) = (x - \alpha)(x - \alpha^2)(x - \alpha^4) = x^3 + x + 1,$$

$$m_3(x) = (x - \alpha^3)(x - \alpha^6)(x - \alpha^5) = x^3 + x^2 + 1.$$

Таким образом, порождающий многочлен имеет вид

$$g(x) = (x^3 + x + 1)(x^3 + x^2 + 1) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1.$$

Этот многочлен $g(x)$ порождает тривиальный $(7, 1)$ -код с повторениями. В действительности этот код имеет кодовое расстояние $d = 7$ и может исправлять все тройные ошибки. Этот факт не является необычным. В некоторых случаях фактическая корректирующая способность кодов БЧХ превышает теоретическое значение.

Коды БЧХ доставляют большой класс легко строящихся кодов с произвольными длиной блока и скоростью. Важность этих кодов обеспечивается не только гибкостью выбора их параметров, но и тем, что при длинах блока около нескольких сотен многие из них являются оптимальными среди всех известных кодов с теми же длиной и скоростью.

Важный подкласс кодов БЧХ составляют коды с $m = m_0 = 1$. Они называются кодами Рида - Соломона. Для этих недвоичных кодов, определенных над полем $GF(q)$, длина блока составляет $n = q - 1$ и порождающий многочлен задается формулой

$$g(x) = (x - \alpha)(x - \alpha^2) \dots (x - \alpha^{2t}). \quad (4.36)$$

Заметим, что степень $g(x)$ равна $2t$, так что для исправления t ошибок требуется только $2t$ проверочных символов. Обычно выбирается значение $q = 2m$. При этом код позволяет исправлять $q = 2m$ -ичные символы, т. е. пакеты ошибок. Такие коды очень полезны в схемах двухуровневого «каскадного» кодирования.

Пример. Пусть нужно построить порождающий многочлен для кода Рида - Соломона длиной 7, исправляющего ошибки в двух символах. Снова предположим, что $GF(8)$ задается таблицей 4.5. Используя (4.36), получаем порождающий многочлен в виде:

$$g(x) = (x - \alpha)(x - \alpha^2)(x - \alpha^3)(x - \alpha^4) = x^4 + \alpha^3 x^3 + x^2 + \alpha x + \alpha^3.$$

Таким образом, $g(x)$ порождает $(7, 3)$ -код, который может исправлять все двойные ошибки. Заметим, что коэффициенты порождающего многочлена являются обычно элементами $GF(8)$.

К сожалению, о спектре кодов БЧХ в общем случае известно очень немного. В некоторых случаях, когда значение k или $n - k$ мало, перебор

позволяет найти спектр некоторых из этих кодов.

Известен также спектр (n, k) -кодов Рида - Соломона. Он имеет вид $A_0 = 1, A_j = 0$ при $1 \leq j \leq n - k$,

$$A_j = C_n^j \sum_{h=0}^{j-1-(n-k)} (-1)^h C_j^h \left(q^{j-h-(n-k)} - 1 \right) \quad (4.37)$$

при $n - k + 1 \leq j \leq n$. Важными классами групповых кодов также являются коды Голея, Рида-Маллера, коды максимальной длины, квадратично-вычетные коды. Однако их рассмотрение выходит за рамки этого пособия.

Каскадные коды

Обычный для практики метод реализации очень мощных кодов состоит в каскадировании двух или более кодов. Каскадные коды были введены Форни в качестве метода практической реализации кода с очень большой длиной блока и весьма высокой корректирующей способностью. Эти цели достигаются при наличии нескольких уровней кодирования многими различными способами. Наиболее распространенной является схема с двумя уровнями кодирования. Одним из кодов, он называется *внешним*, бывает обычно код Рида – Соломона. В качестве *внутреннего* можно выбрать один из многих различных кодов.

Основные принципы каскадного кодирования с двумя уровнями показаны на рисунке 4.9. Предположим для простоты, что рассматриваемый канал имеет двоичный вход. Внешним кодом является недвоичный код, использующий K -битовые символы. Эти символы поступают в кодер внешнего кода от источника данных, как показано на рисунке 4.9.

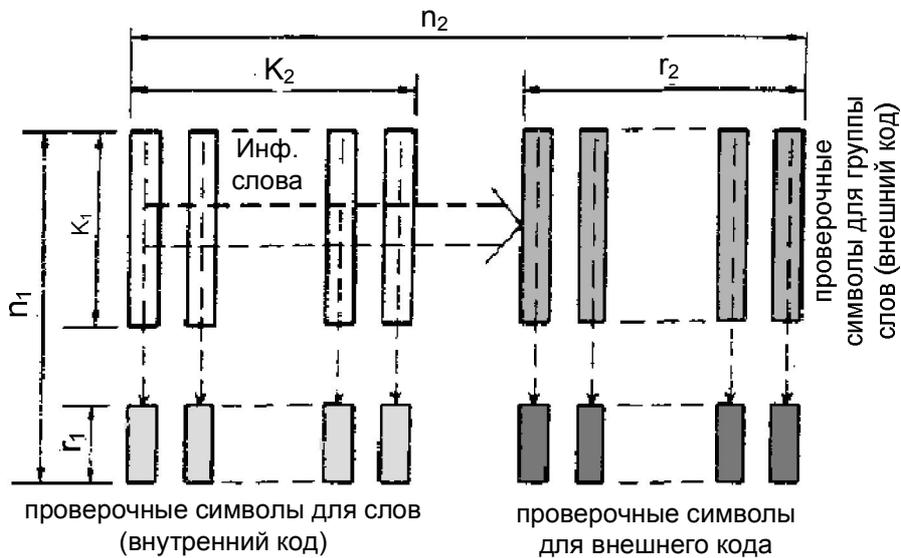
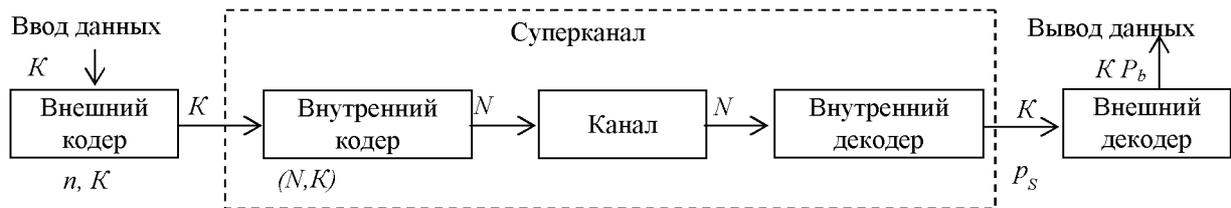


Рисунок 4.9 – Принцип каскадного кодирования

Предполагается, что внешний код является блоковым и его блок состоит из n символов, причем k из них являются информационными. Далее K -битовые символы, выходящие из кодера внешнего кода, кодируются кодером внутреннего кода. При этом добавляются $N - K$ проверочных двоичных символов, так что длина блока внутреннего кода равна N . Хотя на рисунке показано, что в канал поступают N -битовые символы, обычно для возможности осуществления двоичной передачи параллельные символы переводятся в последовательные.

На приемном конце могут использоваться как жесткие, так и мягкие решения. В любом случае результаты этих решений параллельно подаются на декодер внутреннего кода. Этот декодер дает оценку каждого K -битового символа внешнего кода со сравнительно малой вероятностью ошибки p_s . Затем декодер внешнего кода исправляет возможно большее число ошибок символов, приводя к очень низкой окончательной вероятности ошибки двоичного символа P_b .

Комбинацию внутреннего кодера, канала и внутреннего декодера иногда называют *суперканалом*. Аналогично комбинацию внешнего и внутреннего кодеров называют *суперкодером*, а комбинацию внутреннего и внешнего декодеров - *супердекодером*. Заметим, что длина каскадного кода $N^* = nN$ дво-

ичных символов, причем $K^* = kK$ символов кодового слова - информационные, и что скорость кода $R^* = rR = kN/(nN)$, где $R = K/N$ и $r = k/n$. Хотя общая длина кода равна nN , каскадирование обеспечивает такую структуру кода, что декодирование может осуществляться с помощью двух декодеров для кодов с длинами N и n соответственно. Это позволяет существенно снизить сложность по сравнению с той, которая потребовалась бы для получения той же вероятности ошибки при одном уровне кодирования. Наиболее естественно в качестве внешних выбирать коды Рида - Соломона. Эти коды наиболее широко распространены, поскольку являются кодами с максимальным расстоянием ($d = n - k + l$) и относительно просто реализуются.

Далее будет использоваться внешний код Рида - Соломона с декодированием лишь при наличии ошибок. Это значит, что декодер внутреннего кода в качестве решений будет выдавать лишь символы, не указывая ни стираний, ни достоверности каждого символа. Предполагая, что вероятность ошибки символа на входе декодера РС равна p_s , можно получить следующую оценку для вероятности ошибки двоичного символа на его выходе:

$$P_b < \frac{2^{K-1}}{2^K - 1} \cdot \sum_{j=t+1}^n \frac{j+t}{n} C_n^j p_s^j (1-p_s)^{n-j}, \quad (4.38)$$

(при коде РС, исправляющем t ошибок символов). Множитель $2^{K-1}/(2^K - 1)$ учитывает среднее число ошибок в информационных двоичных символах на одну ошибку K -битового символа. Для оценки характеристик конкретной каскадной схемы следует оценить p_s для принятого внутреннего кода, а затем вычислить P_b с помощью (4.38).

Кратко охарактеризуем три подхода, при которых внутренними кодами служат ортогональные, короткие блочные и сверточные коды.

Системы, использующие код Рида-Соломона и ортогональный код

Сложность внешнего декодера кода РС определяется объемом алфавита (K двоичных символов), длиной n и кодовым расстоянием d . При заданном внешнем коде можно пытаться найти наиболее эффективный внутренний код, т.е. такой (N, K) -код, который при заданном K позволяет работать при возможно меньшем значении отношения E_b / N_0 энергии информационного символа к спектральной мощности шума. Таким кодом является упоминавшийся ранее симплексный $(2^K - 1, K)$ -код. Однако почти оптимальными получаются характеристики при биортогональном $(2^{K-1}, K)$ -коде и ортогональном $(2^K, K)$ -коде. Эти коды могут быть реализованы либо как двоичные блочные

коды, либо с помощью M -ичной модуляции (например, (2^K) -ичной ФМ, обеспечивающей ортогональность сигналов). В любом случае сложность декодера растет экспоненциально с ростом K , поскольку для каждого кодового слова нужен отдельный коррелятор или согласованный фильтр. Таким образом, интерес представляют лишь относительно небольшие значения K ($K \leq 10$).

Вероятность ошибки двоичного символа для каскадной системы легко найти, приняв вероятность ошибки кодового слова для этих кодов при фазовой модуляции, за вероятность ошибки символа p_s кода РС в (4.38).

Улучшение характеристик за счет увеличения K (и длины кода РС) приводит к убыванию общей скорости кода как $R^* \approx K2^{-K}$, и это является одним из наиболее серьезных недостатков либо из-за недостаточности ширины полосы частот, либо из-за трудностей построения хороших демодуляторов для очень низких значений E_s / N_0 , требуемых при этих скоростях.

Результаты оценки вероятности ошибки символа на выходе декодера в случае некогерентного приема M ортогональных сигналов свидетельствуют о том, что весьма эффективным является использование в некогерентных системах алфавитов большого объема.

Очевидное свойство и недостаток внутренних кодов этого класса состоит в их очень низких скоростях. Поэтому они могут быть полезны только при использовании в очень широкополосных каналах (т. е. когда требуемые скорости передачи данных намного меньше той, которую обеспечивает полоса частот канала). Такими являются каналы систем космической связи, где скорости поступления данных весьма малы. В настоящее время весьма актуально их использование в широкополосных системах связи.

Системы, использующие код Рида-Соломона и короткий блочный код

Получить разумные скорости кодирования, лежащие в интервале $1/2 \leq R^* \leq 3/4$, и приближающиеся к системам с биортогональными кодами характеристики декодирования по максимуму правдоподобия, можно применив в качестве внутренних кодов короткие блочные коды. Другое преимущество, которое часто остается незамеченным, состоит в том, что такие системы весьма эффективны при очень высоких скоростях поступления данных.

Коды РС хороши при высоких скоростях поступления данных, поскольку при их декодировании не нужно выполнять операции со скоростью, равной скорости поступления данных. Вместо этого операции нужно выполнять либо со скоростью поступления символов кода РС (что составляет малую долю от скорости поступления данных), либо несколько раз при обработке кодового слова (намного меньше числа двоичных символов в кодовом слове). Кроме того, легко построить высокоскоростную реализацию коротких блочных кодов

с кодовыми расстояниями 4 или менее. Структура такой каскадной системы облегчает распараллеливание операций при декодировании, снижая, таким образом, требования к быстродействию логических элементов.

Системы, использующие код Рида-Соломона и сверточный код

В качестве внутренних можно применять также сверточные коды, получая при этом высокие характеристики при сравнительно высоких общих скоростях кодирования. Поскольку сверточный код можно применять при умеренно высоких вероятностях ошибки, наиболее перспективным является использование кодов с малой длиной кодового ограничения и алгоритма декодирования Витерби (с мягкими решениями демодулятора).

Структурная схема системы, внутренним кодом которой является сверточный код, не сильно отличается от показанной на рисунке 4.9. Поступающие из внешнего кодера K -битовые символы должны быть вновь переведены в последовательные для того, чтобы можно было осуществить сверточное кодирование. Однако при этом следует сохранять блоковую синхронизацию, т. е. совпадение начала блока внешнего кода и каждого символа кода РС. После декодирования данные вновь должны быть переведены в K -битовые символы, поступающие на вход декодера РС. Поскольку корреляция ошибок на выходе декодера может вызвать ухудшение параметров декодера, следует применить перемежение. Это перемежение должно быть таким, чтобы никакие два символа на выходе декодера, расстояние между которыми меньше глубины декодирования, не принадлежали одному кодовому слову кода РС.

ГЛАВА 5 СИСТЕМЫ РАДИОСВЯЗИ С РЕТРАНСЛЯЦИЕЙ И МНОГОСТАНЦИОННЫМ ДОСТУПОМ

5.1 Системы авиационной наземной и воздушной радиосвязи

Все системы авиационной радиосвязи можно разделить на две большие группы:

- системы ближней связи, иногда называемые системами прямой видимости,
- системы дальней радиосвязи.

Такое разделение обусловлено использованием определенных диапазонов частот и предназначением той или иной системы радиосвязи. В настоящее время для организации наземной и воздушной связи используются, как правило, гектометровый, декаметровый, метровый, дециметровый и сантиметровый диапазоны волн. В указанных диапазонах можно обеспечить необходимые дальности связи, скорость передачи, а также применить сравнительно малогабаритные самолетные антенны.

Широкое использование для авиационной связи метровых, дециметровых и сантиметровых волн объясняется возможностью построения в этом диапазоне систем с высокой пропускной способностью, применения для передачи информации широкополосных сигналов, а также сравнительно низким уровнем помех. В последнее время наряду с обычными системами радиосвязи широкое применение получили системы наземной радиосвязи, использующие эффект дальнего распространения волн указанных диапазонов за счет рассеяния на локальных неоднородностях тропосферы и ионосферы или отражения от ионизированных следов метеоров. Для увеличения дальности связи и обеспечения связи с низколетящими самолетами широко используются различные методы ретрансляции сигналов через наземные радиорелейные станции, самолетные (вертолетные) и спутниковые ретрансляторы.

5.1.1 Системы радиосвязи прямой видимости

Системы радиосвязи прямой видимости обеспечивают ближнюю связь (в пределах прямой видимости) самолетов с наземными пунктами управления, связь между самолетами, находящимися в воздухе, а также связь между наземными радиостанциями. Они работают, как правило, в метровом и дециметровом диапазонах волн.

Как известно, дальность прямой видимости с учетом сферической формы Земли определяется по формуле

$$D = 3,75(\sqrt{h_1 + h_2}), \quad (5.1)$$

где D — дальность прямой видимости в километрах; h_1 и h_2 — высота подъема антенн передающей и приемной радиостанций в метрах. В рассматриваемых системах связь осуществляется наземным лучом, который вследствие тропосферной рефракции в определенной степени огибает земную поверхность. Поэтому дальность связи достигается несколько больше дальности прямой видимости и определяется по формуле:

$$D_{\text{макс}} = 4,12 \left(\sqrt{h_1 + h_2} \right). \quad (5.2)$$

Входящие в эту формулу величины имеют такой же смысл и размерность, что и в случае (5.1).

Таким образом, для обеспечения связи на заданную дальность необходимо определенным образом выбрать высоту приемной и передающей антенн, а также рассчитать мощность передатчика и другие параметры системы так, чтобы обеспечить заданное отношение сигнал — шум на входе приемника. Если осуществляется связь самолета с наземной станцией, то $D_{\text{макс}}$ определяется высотой полета самолета и высотой подъема наземной антенны. В случае связи между самолетами, находящимися в воздухе, h_1 и h_2 имеют значение высот полета самолетов, участвующих в связи.

Дальность наземной радиосвязи для практически используемых антенных систем не превышает 40...50 км. Для увеличения дальности, особенно наземной связи, радиостанции обычно размещают на возвышенностях, искусственных холмах и т. п.

Максимальные значения дальности связи $D_{\text{макс}}$ в системах прямой видимости, рассчитанные по формуле (5.2) приведены в таблице 5.1.

Таблица 5.1 – Максимальные значения дальности связи $D_{\text{макс}}$ в системах прямой видимости рассчитанные по формуле (5.2)

Высота полета самолета h_1 , м	$D_{\text{макс}}$, км	
	с наземной станцией при $h_2=10$ м	между самолетами $h_1 = h_2$
100	54	82
1000	144	262
5000	301	580
10000	425	823
15000	515	1002
20000	591	1178
25000	661	1290

Практически мощность самолетных передатчиков выбирается из условия обеспечения дальности связи до 500...600 км. На самолетах дальней и во-

енно-транспортной авиации кроме радиостанций ближней связи устанавливаются радиостанции дальней связи, обеспечивающие радиосвязь на полный радиус действия самолета.

В авиационных системах связи прямой видимости используются, как правило, поднятые над Землей антенны. В наземных радиостанциях для достижения необходимой дальности связи антенны размещаются на мачтах на высоте от десяти метров и более, антенны самолетных УКВ радиостанций в принципе относятся к группе поднятых антенн.

Важной особенностью этих систем связи является также то, что они обеспечивают достаточно надежную передачу информации с большой скоростью (до нескольких тысяч бит/с), что весьма важно для осуществления быстрой цифровой связи. Имеется возможность использования в этих системах широкополосных сигналов, многоканальных способов передачи информации, а также перестройки приемно-передающей аппаратуры в широком диапазоне частот.

Сравнительно малая масса и габариты УКВ аппаратуры, возможность применения антенн с небольшим аэродинамическим сопротивлением или не выступающих антенн обеспечили широкое использование этих систем для воздушной радиосвязи. В состав бортового радиосвязного оборудования любого самолета и вертолета, как правило, входит радиостанция метрового или дециметрового диапазона, обеспечивающая ближнюю связь самолета с наземными пунктами и между самолетами.

Особенности систем радиосвязи прямой видимости:

1. Высокая устойчивость связи, поскольку характер распространения радиоволн слабо зависит от времени суток и года, метеорологических условий, особенно при работе на частотах не выше 3 ГГц. Поглощение в тропосфере сравнительно невелико. Значительные поглощения в осадках, а также молекулярные поглощения наблюдаются лишь в сантиметровом и в более коротковолновых диапазонах.

2. Основным видом помех радиосвязи в УКВ диапазоне космические шумы и внутренние шумы аппаратуры, а также помехи, создаваемые другими радиостанциями.

Сравнительно слабая подверженность помехам объясняется еще и тем, что на такую систему действуют только помехи, источники которых находятся в пределах прямой видимости с приемным устройством данной системы.

Помехозащищенность может быть улучшена также тем, что в рассматриваемых системах возможно применение антенн направленного действия, особенно в системах наземной связи. В системах связи прямой видимости слабо проявляется мультипликативная помеха, поскольку связь ведется в основном прямым лучом, хотя при связи между самолетами этот вид помехи

оказывает заметное влияние на качество связи, поскольку на вход приемника действует помимо прямого луча отраженный от земной поверхности сигнал.

3. Обеспечение достаточно надежной передачи информации с большой скоростью (до нескольких тысяч бит/с), что весьма важно для осуществления быстродействующей цифровой связи. Имеется возможность использования в этих системах широкополосных сигналов, многоканальных способов передачи информации, а также перестройки приемо-передающей аппаратуры в широком диапазоне частот.

4. Сравнительно малые масса и габариты УКВ аппаратуры, возможность применения антенн с небольшим аэродинамическим сопротивлением или не выступающих антенн обеспечили широкое использование этих систем для воздушной радиосвязи. В состав бортового радиосвязного оборудования любого самолета и вертолета, как правило, входит радиостанция метрового и дециметрового диапазона, обеспечивающая ближнюю связь самолета с наземными пунктами и между самолетами.

Основным недостатком систем радиосвязи прямой видимости является ограниченная дальность действия. Поэтому на линиях большой протяженности возникает необходимость производить ретрансляцию сигналов. В этом случае ретранслятор обычно устанавливается на борту летательного аппарата, совершающего полет в заданном районе.

5.1.2 Системы коротковолновой авиационной радиосвязи

Коротковолновые системы радиосвязи широко используются в авиации для обеспечения дальней воздушной и наземной связи. Применение частот диапазона 2...30 МГц обусловлено свойством дальнего и сверхдальнего распространения волн этого диапазона за счет отражения от ионосферы. Благодаря этому, обеспечивается связь практически на любые земные расстояния. При этом создание и эксплуатация КВ систем связи оказывается значительно экономичнее других систем дальней связи (тропосферных, метеорных, радиорелейных и т. п.). Коротковолновые радиостанции еще долго будут оставаться основным средством дальней воздушной радиосвязи.

Основные характеристики и принцип КВ систем определяются особенностями распространения радиоволн этого диапазона. Дальнее распространение ДКВ происходит за счет ионосферных волн, т. е. путем однократного или многократного отражения от ионосферы. Как известно, основную роль в распространении ДКВ играет ионосферный слой F2 как слой отражающий. Ионосферные слои D и E являются поглощающими. Поэтому условия распространения коротких волн зависят прежде всего от состояния этих областей ионосферы.

Особенностью распространения ДКВ является прежде всего то, что интенсивность отражения и поглощения в ионосфере существенно зависит от времени суток и года, а также от рабочей частоты и протяженности трассы. Для осуществления радиосвязи в этом диапазоне волн должны одновременно выполняться следующие условия: используемая частота должна быть меньше максимально применимой частоты (МПЧ) для заданной трассы и состояния отражающего слоя; используемая частота должна быть такой, чтобы поглощение радиоволн в областях D и E не было чрезмерно большим.

Первое условие ограничивает диапазон используемых частот сверху для данной линии и определенного времени суток и года. Оно является критическим, так как при невыполнении его радиоволны вообще не будут отражаться от слоя F2.

Второе условие ограничивает диапазон используемых частот снизу, так как с уменьшением частоты поглощение в ионосфере возрастает. Это условие не является критическим, поскольку потери при распространении могут быть компенсированы увеличением мощности передатчика.

Поэтому очень важным при связи в ДКВ диапазоне является выбор оптимальных частот с учетом резкой зависимости состояния ионосферы от времени суток и года, а также фазы 11-летнего периода солнечной активности. Эта задача дополнительно усложняется тем, что на протяженных трассах (более 4000 км) имеет место неоднократное отражение от участков ионосферы, расположенных в различных земных широтах с различными условиями освещенности. Особенно это относится к протяженным трассам, проходящим вдоль земных параллелей.

С помощью ионосферных карт составляется суточный график МПЧ для заданной трассы, а также аналогичный график изменения ОРЧ с учетом того, что последняя составляет обычно 0,7 ... 0,8 от МПЧ. Суточный график ОРЧ имеет вид, подобный изображенному на рисунке 5.1.

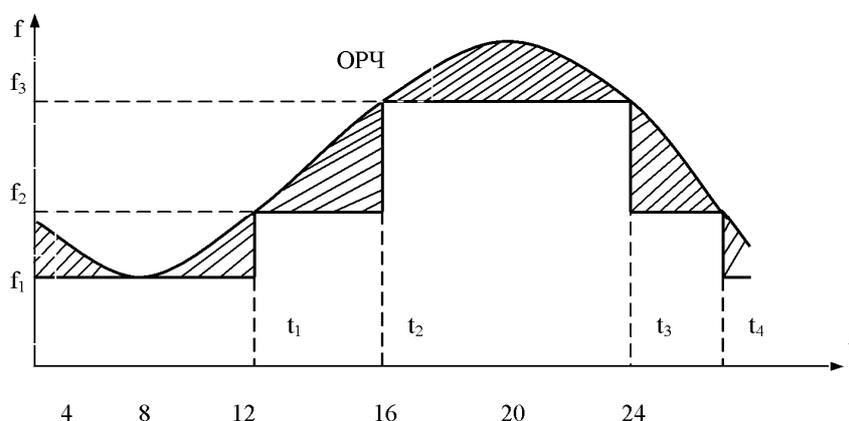


Рисунок 5.1 – К определению суточного набора рабочих частот

Практически сложно производить непрерывную перестройку приемника и передатчика системы радиосвязи в строгом соответствии с изменением ОРЧ. Поэтому для поддержания круглосуточной связи рекомендуется использовать лишь несколько рабочих частот, которые совпадают с ОРЧ только в определенные моменты времени. Для определения указанных частот на графике (рисунок 5.1) плавную кривую суточного изменения ОРЧ заменяют ломаной линией такой конфигурации, при которой заштрихованная область будет обладать наименьшей площадью. В приведенном на рисунке примере суточный набор составляет три частоты (f_1, f_2, f_3). Выбранные таким образом частоты будут наиболее близки к оптимальным. Рекомендуемое время перехода с одной частоты на другую в течение суток определяется абсциссами ступенек (t_1, t_2, t_3, t_4) ломаной линии.

Особенность связи на ДКВ состоит также в том, что может иметь место зона молчания – некоторая кольцевая область вокруг работающего передатчика, в которой отсутствует прием сигналов. Существование зоны молчания объясняется тем, что, с одной стороны, короткие волны как земные сильно поглощаются в почве; с другой стороны, пространственные лучи отражаются от ионосферы только в том случае, если они падают на нее под углами, превышающими некоторый критический угол. Таким образом, внутренний радиус зоны молчания определяется условиями распространения земных волн, внешний радиус – условиями распространения ионосферных волн. При этом величина внешнего радиуса в значительной степени зависит от времени суток и от рабочей частоты. С наступлением темноты внешний радиус увеличивается. С возрастанием частоты он также возрастает. Однако путем рационального выбора частот связи можно избавиться от зоны молчания.

Следующей особенностью работы систем КВ связи является то, что этот диапазон сильно перегружен большим количеством действующих радиостанций. Так, по данным Международного комитета регистрации частот (МКРЧ), число зарегистрированных рабочих частот в ДКВ диапазоне по всем странам мира доходит до 750 тысяч. В это число не входят многочисленные частоты, используемые военными и специальными службами. Таким образом, на каждую работающую КВ радиостанцию приходится участок диапазона 20...30 Гц, что значительно меньше реально необходимой полосы частот и приводит к перекрытию полос, занимаемых действующими радиостанциями.

Характерной особенностью КВ систем связи является то, что они в значительной степени подвержены вредному действию мультипликативной помехи, которая возникает в процессе распространения радиоволн в ионосфере и проявляется в виде флюктуации амплитуды и фазы принимаемого сигнала. Наблюдаются быстрые и медленные замирания.

Быстрые замирания объясняются преимущественно интерференцией множества лучей, приходящих в точку приема, фаза которых вследствие непостоянства ионизированных слоев непрерывно и случайным образом изменяется. Амплитуда сигнала при этом изменяется в десятки и даже в сотни раз. Период замираний колеблется от десятых долей секунд до нескольких секунд.

Медленные замирания сигнала с периодом в несколько минут и более связаны с изменением поглощения радиоволн в среде распространения, изменением неоднородностей ионосферы, а также обусловлены интерференцией обыкновенного и необыкновенного компонентов сигнала. Медленные изменения уровней сигнала могут превышать соответствующие флюктуации при быстрых замираниях. Часовые, суточные, сезонные периоды изменений уровней сигнала чаще всего зависят от изменения освещенности ионосферы и поглощения в среде распространения.

5.1.3 Особенности радиосвязи с ВС на малых высотах.

Ретрансляция сигналов

Основным недостатком систем радиосвязи прямой видимости является ограниченная дальность действия. Поэтому на линиях большой протяженности возникает необходимость производить ретрансляцию сигналов. Классическим примером системы связи с ретрансляцией является система радиорелейной связи, используемая в авиации для организации наземной радиосвязи на дециметровых и более коротких волнах.

Ретрансляция сигналов широко используется также в сетях воздушной радиосвязи с целью увеличения дальности связи за пределы прямой видимости, обеспечения связи низколетящих самолетов между собой и с наземными пунктами управления. В этом случае ретранслятор обычно устанавливается на борту воздушного судна, совершающего полет в заданном районе. Ретрансляционная аппаратура обеспечивает прием, преобразование, усиление и последующую передачу радиосигналов при симплексной и дуплексной двусторонней радиосвязи между конечными пунктами радиолинии. Управление бортовой аппаратурой ретрансляторов производится дистанционно с наземных пунктов или оператором на борту самолета.

В настоящее время широко используются для организации дальней связи на метровых и более коротких волнах системы, работа которых основана на эффекте дальнего распространения радиоволн за счет тропосферного и ионосферного рассеяния или отражения от ионизированных следов метеоров.

Таким образом, можно выделить основные положения:

- системы радиосвязи прямой видимости обеспечивают ближнюю связь (в пределах прямой видимости) самолетов с наземными пунктами

управления, связь между самолетами, находящимися в воздухе, а также связь между наземными радиостанциями. Они работают, как правило, в метровом и дециметровом.

- основные характеристики и принцип КВ систем определяются особенностями распространения радиоволн этого диапазона.

5.2 Общие сведения о многоканальной связи

5.2.1 Многоканальные системы связи

Любая информация передается от передатчика к приемнику через физическую среду, в качестве которой в системах электросвязи используются радиолинии и кабельные линии связи. Высокая стоимость линейных сооружений обуславливает необходимость их наиболее эффективного использования. Поэтому в настоящее время получили широкое применение *многоканальные системы связи (МКС), обеспечивающие одновременную и независимую передачу большого числа абонентов с использованием одной линии связи.*

Упрощенная структурная схема МКС приведена на рисунке 5.2.

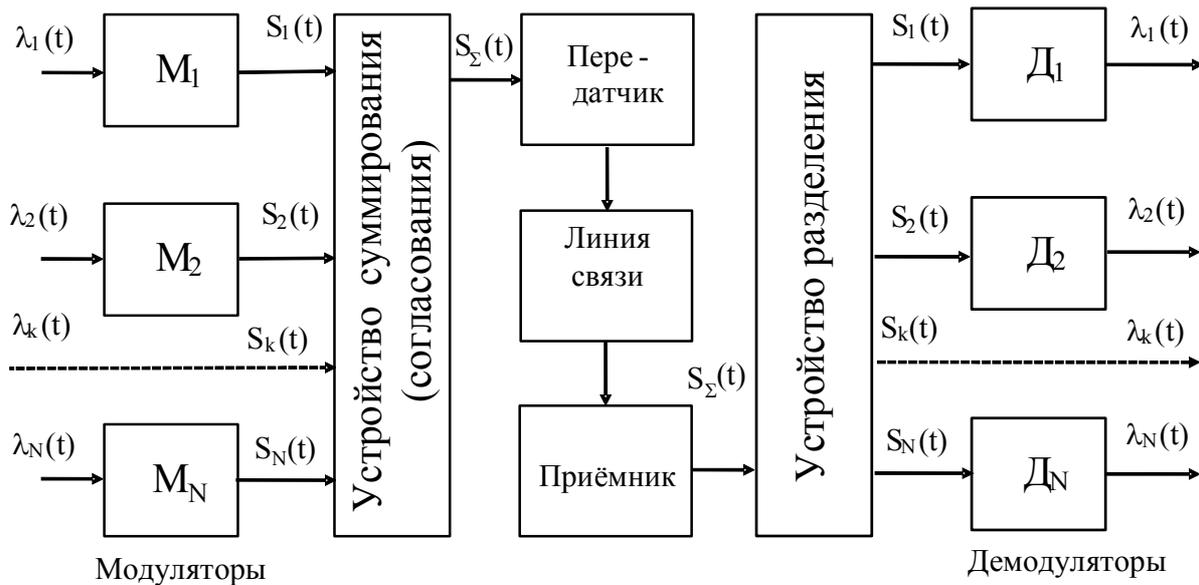


Рисунок 5.2 – Упрощенная структурная схема МКС

Первичные сигналы $\lambda_k(t)$, подлежащие передаче от N источников информации, могут обладать одинаковыми характеристиками. Поэтому на передающей стороне производится их преобразование с помощью преобразователей M_k , в результате чего каждый из канальных информационных сигналов $S_k(t)$, приобретает определенный признак, который используется на приемной стороне для разделения сигналов по соответствующим каналам. Это преобразование в общем виде можно представить выражением

$$S_k(t) = M_k \{ \lambda_k(t) \}, \quad (5.3)$$

где M_k – оператор преобразователя k -го канала.

Преобразование может заключаться в частотном, фазовом, временном и других изменениях первичного сигнала. В устройстве суммирования (УС) производится сложение сигналов, в результате чего получается групповой сигнал вида

$$S_{\Sigma}(t) = \sum_{k=1}^N S_k(t), \quad (5.4)$$

где N – количество каналов в МКС. Для передачи этого сигнала в передающем устройстве (ПРД) осуществляется его дополнительное преобразование в зависимости от физической природы линии связи (ЛС). Например, для радиолинии на выходе передатчика необходимо сформировать радиосигнал, удобный для передачи по радиоканалу, для проводной ЛС – достаточно с помощью ПРД усилить групповой сигнал по мощности и т.д. В приемнике (ПРМ) производится преобразование линейного сигнала в групповой сигнал $S_{\Sigma}(t)$, а затем с помощью устройства разделения (УР) осуществляется распределение сигналов $S_k(t)$ по соответствующим каналам и их преобразование с помощью демодуляторов D_k в первичный сигнал $\lambda_k(t)$, который в определенной степени может отличаться от исходного сигнала $\lambda_k(t)$ из-за действия помех в МКС. Для обеспечения дуплексной связи необходимо в общем случае иметь двойной комплект устройств, приведенных на рисунке 5.2.

Многоканальные системы связи отличаются друг от друга методом разделения каналов (сигналов). Известно большое разнообразие этих методов, однако на практике наибольшее распространение получили МКС с линейным разделением сигналов, в которых реализуются принципы частотного, временного или фазового разделения сигналов. Рассмотрим разделения сигналов в МКС.

5.2.2 Методы линейного разделения сигналов

Преобразование первичного сигнала, описываемое выражением (5.3), осуществляется путем модуляции сигналом $\lambda_i(t)$ некоторого переносчика $\psi_i(t)$. На приемной стороне производится разделение группового сигнала по правилу

$$L_k \{ S_{\Sigma}(t) \} = S_k(t), \quad (5.3)$$

где L_k — оператор выделения i -го сигнала. При линейном разделении сигналов оператор L_k является линейным, поэтому можно записать

$$L_k \{S_\Sigma(t)\} = L_k \left\{ \sum_{k=1}^N S_k(t) \right\} = \sum_{k=1}^N L_k \{S_k(t)\}. \quad (5.4)$$

Тогда условие разделения сигналов (5.3) можно записать в виде

$$L_k \{S_i(t)\} = \begin{cases} S_k(t) & \text{при } i = k \\ 0 & \text{при } i \neq k \end{cases} \quad (5.5)$$

Оно выполняется, если канальные сигналы $S_i(t)$ являются линейно независимыми. Для формирования таких сигналов необходимо использовать линейно независимые переносчики. Если к тому же переносчики представляют собой систему ортогональных функций (корреляционная функция которых равна нулю), корреляционная функция которых удовлетворяет в интервале времени (t_1, t_2) условию

$$\int_{t_1}^{t_2} \Psi_i(t) \Psi_k(t) dt = \begin{cases} 1 & \text{при } i = k, \\ 0 & \text{при } i \neq k, \end{cases} \quad (5.6)$$

где $\Psi(t)$ - функция переносчика сигнала.

В этом случае выделение сигнала $S_i(t)$ из группового сигнала $S_\Sigma(t)$ может быть выполнено путем перемножения последнего с переносчиком данного канала и интегрирования в интервале ортогональности

$$S_k(t) = \int_{t_1}^{t_2} S_\Sigma(t) \Psi_k(t) dt. \quad (5.7)$$

Требованию ортогональности удовлетворяет целый ряд сигналов, которые могут быть использованы в МКС в качестве переносчиков информационных сигналов. Так, например, ортогональными являются электрические колебания с неперекрывающимися по частоте спектрами, периодические последовательности импульсов, неперекрывающиеся по времени, гармонические колебания с одинаковыми частотами и начальными фазами, отличающимися на $\pi/2$, и т. д. На базе указанных переносчиков практически реализуются МКС соответственно с частотным, временным или фазовым разделением каналов.

Возможно использование и других более сложных переносчиков (кратные несущие, ортогональные полиномы и др.), которые могут обеспечить разделение сигналов по форме. Однако такие МКС достаточно сложны и только в настоящее время стали находить практическое применение. Системы с неортогональными линейно независимыми переносчиками обладают значительно меньшей взаимной защищенностью между каналами, чем системы с ортогональными сигналами, и поэтому нашли применения в военных МКС повышенной скрытности связи.

В авиационной связи наиболее широко применяются МКС с частотным или временным разделением каналов.

5.2.3 Временное разделение каналов

В многоканальных системах с временным разделением каналов каналные сигналы передаются в строгой очередности без перекрытия по времени. В качестве переносчиков первичных сигналов $\lambda_i(t)$ используются периодические последовательности импульсов $\psi_i(t)$, ортогональные по времени.

Структурная схема МКС с временным разделением каналов приведена на рисунке 5.3. В передающей части системы на модуляторы каналов M подаются информационные сигналы $\lambda_i(t)$ и переносчики $\psi_i(t)$. Последние формируются устройством, состоящим из генератора тактовых импульсов (ГТИ) и линии задержки (ЛЗ). В модуляторах осуществляется один из видов импульсной модуляции, в результате которой формируются каналные сигналы $S_i(t)$. После линейного сложения каналных сигналов образуется групповой сигнал $S_{\Sigma}(t)$.

Период следования тактовых импульсов Δ для передачи аналоговых сигналов со спектром, ограниченным верхней частотой F должен выбираться на основании теоремы отсчетов ($\Delta < 1/2F$). Для ограничения входных сигналов по верхней граничной частоте используются фильтры нижних частот (ФНЧ). Групповой сигнал поступает в передатчик ПРД, где осуществляется формирование линейного сигнала (радиосигнала), который по линии связи ЛС передается на приемник ПРМ. В результате определенной обработки на выходе приемника выделяется групповой сигнал. Распределение сигналов по соответствующим каналам осуществляется с помощью специальных электронных коммутаторов (ЭК). Коммутатор соответствующего канала отпирается только на время импульса, принадлежащего данному каналу. Это достигается за счет синхронной работы формирователей функций $\psi_i(t)$ на передающей и приемной сторонах путем передачи в конце каждого периода дискретизации специального синхронизирующего импульса, который образуется на выходе схемы формирования импульса синхронизации (ФИС). Синхронизирующий импульс (или группа импульсов), имея специфические признаки, на приемной стороне выделяется с помощью селектора импульса синхронизации (СИС), поступает в линию задержки приемника, обеспечивая формирование импульсов коммутации каналов $\psi_i(t)$. Канальные сигналы поступают на соответствующие детекторы (Д), где преобразуются в первичные сигналы.

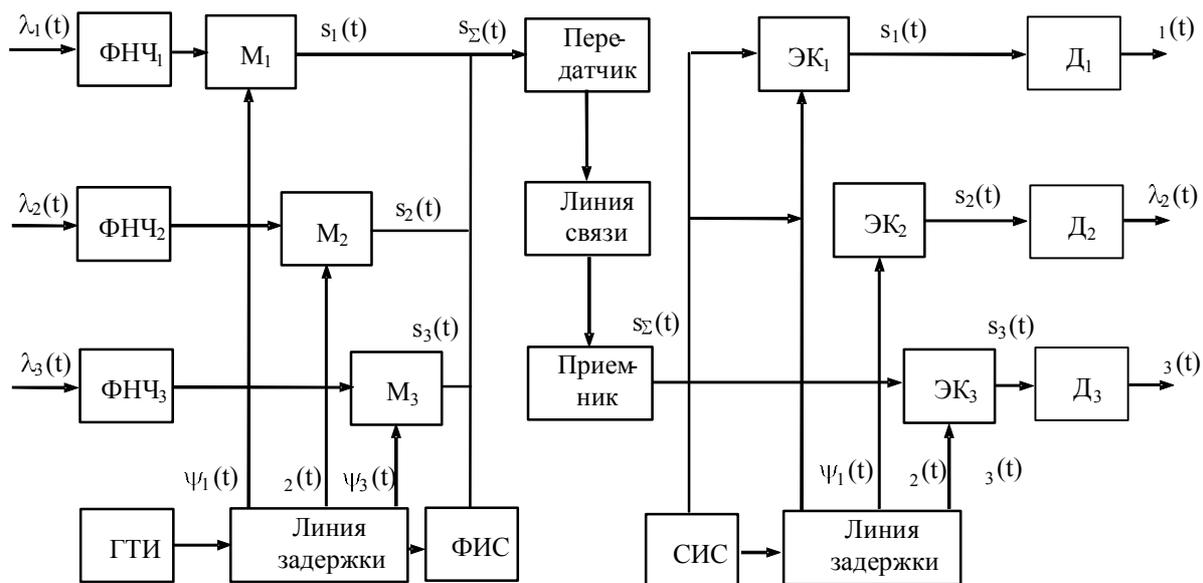


Рисунок 5.3 - Структурная схема МКС с временным разделением каналов

Технология временного разделения каналов согласно международной классификации носит название **мультиплексирования с разделением времени (Time Division Multiplexing, TDM)**. Аппаратура TDM-сетей - мультиплексоры, коммутаторы, демультиплексоры - работает в режиме разделения времени, поочередно обслуживая в течение цикла своей работы все абонентские каналы. Каждому соединению выделяется один квант времени цикла работы аппаратуры, называемый также тайм-слотом. Длительность тайм-слота зависит от числа абонентских каналов, обслуживаемых мультиплексором TDM или коммутатором (рисунок 5.4).

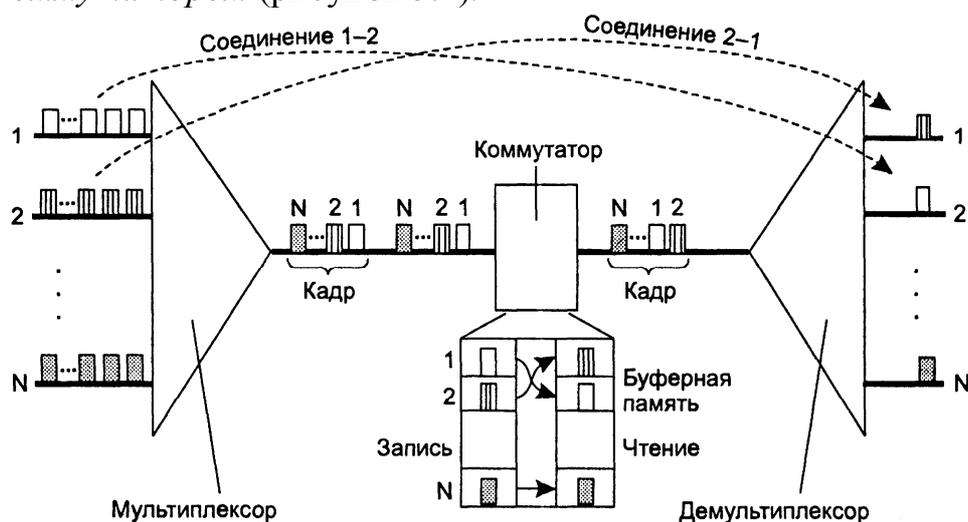


Рисунок 5.4 – Принцип формирования группового сигнала при временном разделении каналов связи

Мультиплексор принимает информацию по N входным каналам от конечных абонентов. В каждом цикле мультиплексор выполняет следующие действия:

- прием от каждого канала очередного байта данных;
- составление из принятых байтов уплотненного кадра, называемого также обоймой;
- передача уплотненного кадра на выходной канал с битовой скоростью, равной $N \cdot 64$ Кбит/с.

Порядок байт в обойме соответствует номеру входного канала, от которого этот байт получен. Количество обслуживаемых мультиплексором абонентских каналов зависит от его быстродействия.

Демультимплексор выполняет обратную задачу - он разбирает байты уплотненного кадра и распределяет их по своим нескольким выходным каналам, при этом он считает, что порядковый номер байта в обойме соответствует номеру выходного канала.

Коммутатор принимает уплотненный кадр по скоростному каналу от мультиплексора и записывает каждый байт из него в отдельную ячейку своей буферной памяти, причем в том порядке, в котором эти байты были упакованы в уплотненный кадр.

Однажды выделенный номер тайм-слота остается в распоряжении соединения «входной канал-выходной слот» в течение всего времени существования этого соединения, даже если передаваемый трафик является пульсирующим и не всегда требует захваченного количества тайм-слотов. Существует модификация техники TDM, называемая статистическим разделением канала во времени (Statistical TDM, STDM). Эта техника разработана специально для того, чтобы с помощью временно свободных тайм-слотов одного канала можно было увеличить пропускную способность остальных.

В МКС с временным разделением каналов для уменьшения взаимного влияния между каналами в групповом сигнале предусматривается защитный интервал τ между импульсами соседними каналами. Тогда длительность информационного импульса в случае АИМ можно определить по формуле $\tau = \Delta / (N+1)q$, где N — число каналов с учетом канала синхронизации, q — скважность импульсной последовательности группового сигнала. При передаче по такой системе речевых сигналов со спектром, ограниченным частотой 3,4 кГц, частота дискретизации выбирается обычно 8 кГц на основании рекомендации МККТТ, а скважность $q=2$. Тогда длительность рабочего импульса в двенадцатиканальной системе будет 5 мкс.

5.2.3 Частотное разделение каналов

В многоканальных системах с частотным разделением каналов каналные сигналы отличаются друг от друга положением своих спектров на оси частот. Для формирования таких сигналов используется модуляция ортогональных переносчиков $\psi_i(t)$ в канальных модуляторах M (рисунок 5.5).

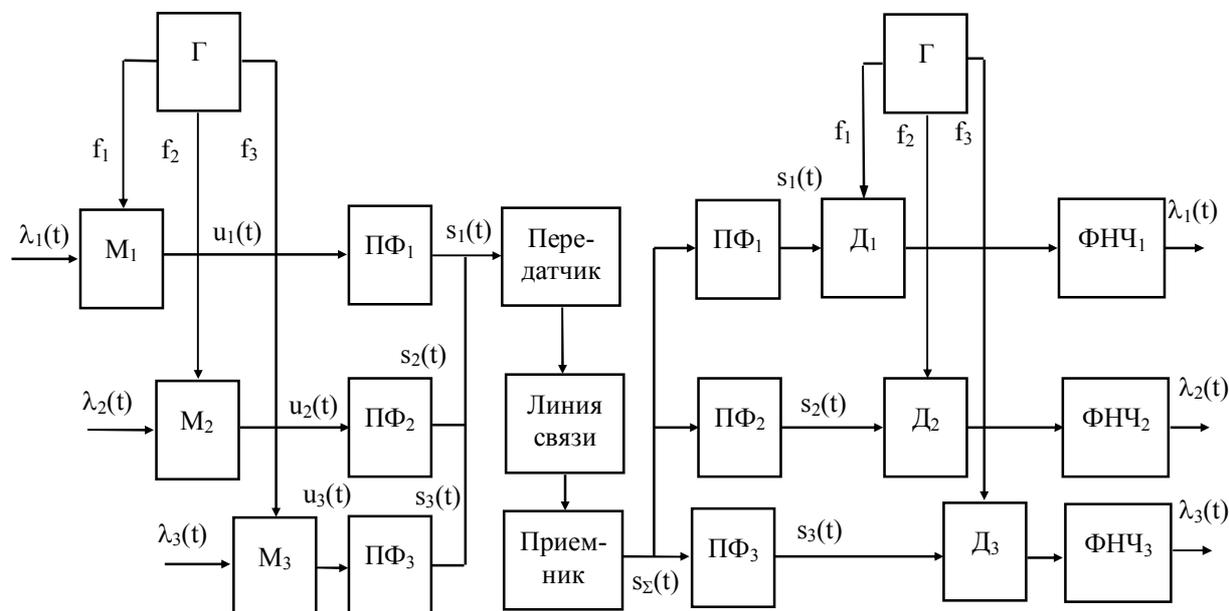


Рисунок 5.5 - Структурная схема МКС с частотным разделением каналов

Обычно МКС с частотным разделением каналов используются для передачи аналоговых сигналов. При этом наиболее часто в индивидуальных каналах применяется однополосная модуляция колебаний с соответственно подобранными частотами f_i , которые создаются генератором Γ . Такой способ модуляции обеспечивает минимальную полосу частот группового сигнала $S_{\Sigma}(t)$. Подавление несущих достигается в модуляторах M , которые, как правило, строятся по балансной схеме, а выделение одной боковой полосы осуществляется с помощью полосовых фильтров ПФ.

На приемной стороне сигналы разделяются полосовыми фильтрами ПФ, затем производится их детектирование в канальных демодуляторах D . Опорные колебания создаются с помощью генератора Γ . На выходе демодуляторов включены фильтры нижних частот ФНЧ, которые подавляют побочные колебания, возникающие в процессе детектирования. В результате указанных преобразований на выходе каналов получают соответствующие первичные сигналы $\lambda_i(t)$, которые в связи с наличием в системе помех могут в определенной степени отличаться от передаваемых сигналов.

В международной классификации частотное разделение каналов полу-

чила наименование - **техника частотного мультиплексирования каналов (FDM)** была разработана для телефонных сетей, но применяется она и для других видов сетей, например сетей кабельного телевидения.

Для разделения абонентских каналов характерна техника модуляции высокочастотного несущего синусоидального сигнала низкочастотным речевым сигналом (рисунок 5.6). Эта техника подобна технике аналоговой модуляции при передаче дискретных сигналов модемами, только вместо дискретного исходного сигнала используются непрерывные сигналы, порождаемые звуковыми колебаниями. В результате спектр модулированного сигнала переносится в другой диапазон, который симметрично располагается относительно несущей частоты и имеет ширину, приблизительно совпадающую с шириной модулирующего сигнала.

Если сигналы каждого абонентского канала перенести в свой собственный диапазон частот, то в одном широкополосном канале можно одновременно передавать сигналы нескольких абонентских каналов.

На входы FDM-коммутатора поступают исходные сигналы от абонентов телефонной сети. Коммутатор выполняет перенос частоты каждого канала в свой диапазон частот. Обычно высокочастотный диапазон делится на полосы, которые отводятся для передачи данных абонентских каналов (рисунок 5.6). Чтобы низкочастотные составляющие сигналов разных каналов не смешивались между собой, полосы делают шириной в 4 кГц, а не в 3,1 кГц, оставляя между ними страховой промежуток в 900 Гц. В канале между двумя FDM-коммутаторами одновременно передаются сигналы всех абонентских каналов, но каждый из них занимает свою полосу частот. Такой канал называют **уплотненным**.

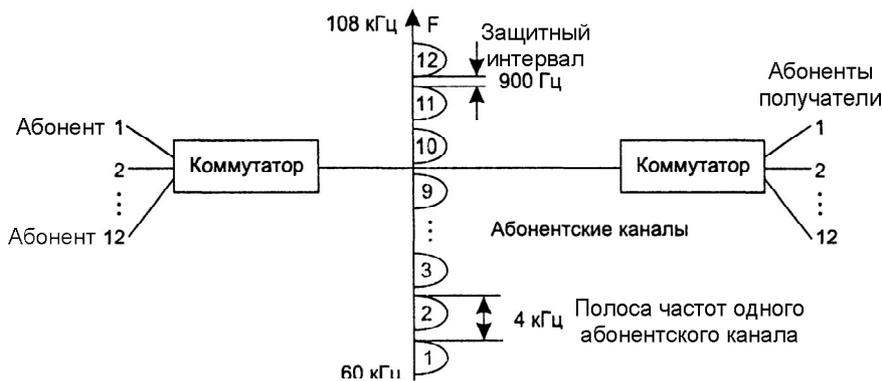


Рисунок 5.6 - Принцип формирования группового сигнала при частотном разделении каналов

Выходной FDM-коммутатор выделяет модулированные сигналы каждой

несущей частоты и передает их на соответствующий выходной канал, к которому непосредственно подключен абонентский телефон.

Реальные сигналы, подлежащие передаче, имеют неограниченный спектр. Кроме того, в результате преобразования сигналов могут появиться побочные колебания. Для снижения влияния одного канала на другой помимо использования высококачественных полосовых канальных фильтров предусматривается защитный частотный интервал ΔF_3 между соседними каналами. В связи с этим ширина полосы частот, приходящаяся на стандартный канал ТЧ, принимается равной 4 кГц.

Важным элементом аппаратуры уплотнения являются полосовые фильтры, которые сравнительно легко могут быть построены в диапазоне 50...120 кГц на базе высокодобротных пьезоэлектрических или магнитострикционных резонаторов.

5.2.4 Фазовое разделение каналов

К классу ортогональных сигналов относятся два гармонических колебания с одинаковыми частотами и начальными фазами, отличающимися друг от друга на $\pi/2$.

$$\psi_1(t) = \cos \omega t \quad \text{и} \quad \psi_2(t) = \cos(\omega t + \pi/2) = \sin \omega t. \quad (5.8)$$

Используя эти колебания в качестве ортогональных переносчиков, можно построить двухканальную систему связи с фазовым разделением каналов. Упрощенная схема такой системы показана на рисунке 5.7.

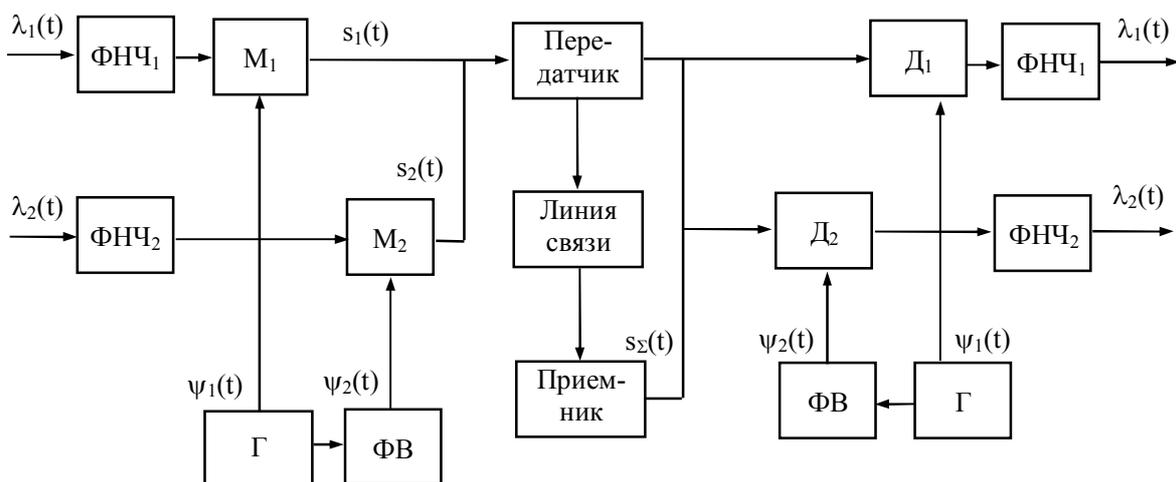


Рисунок 5.7 - Структурная схема многоканальной системы связи с фазовым разделением каналов

Гармонические колебания формируемые с помощью генератора Г и фазовращателя ФВ, поступают на модуляторы. В результате модуляции этих колебаний первичными сигналами образуются каналные сигналы

$$s_1(t) = \lambda_1(t) \cos \omega t \quad \text{и} \quad s_2(t) = \lambda_2(t) \sin \omega t. \quad (5.9)$$

Спектр каждого каналного сигнала содержит две боковые полосы относительно несущей частоты. Эти спектры могут перекрываться по частоте, если спектры передаваемых сигналов находятся в одной частотной области. Это имеет место всегда при передаче по обоим каналам речевых сигналов. Очевидно, действие сигналов $\lambda_i(t)$ может совпадать и по времени. Тем не менее, разделение этих сигналов на приемной стороне возможно благодаря взаимной ортогональности переносчиков $\psi_i(t)$. После линейного сложения каналных сигналов формируется групповой сигнал

$$s_{\Sigma}(t) = \lambda_1(t) \cos \omega t + \lambda_2(t) \sin \omega t, \quad (5.10)$$

который через тракт ПРД, ЛС, ПРМ поступает на синхронные детекторы Д, каналных сигналов. На детекторы также поступают опорные напряжения, которые должны быть синфазны с соответствующими колебаниями, генерируемыми на передающей стороне В результате детектирования в первом канале получается сигнал

$$\begin{aligned} u_{\text{дл}} &= u_{\Sigma}(t) \cos \omega t = \hat{\lambda}_1(t) \cos^2 \omega t + \hat{\lambda}_2(t) \sin \omega t \cos \omega t = \\ &= \frac{1}{2} \left[\hat{\lambda}_1(t) + \hat{\lambda}_1(t) \cos 2\omega t + \hat{\lambda}_2(t) \sin 2\omega t \right]. \end{aligned} \quad (5.11)$$

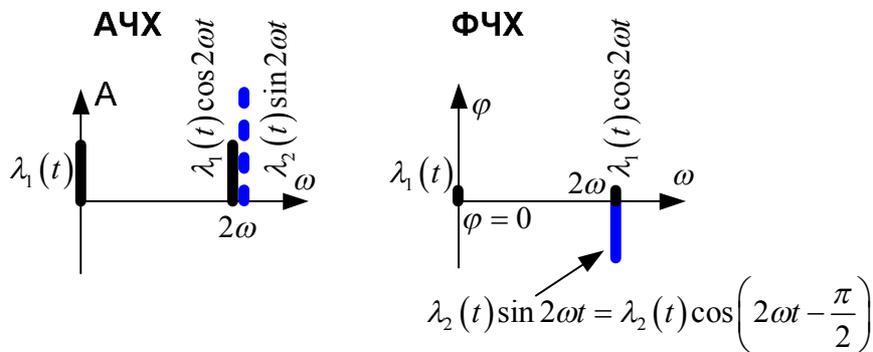


Рисунок 5.8 - АЧХ и ФЧХ МКС с фазовым разделением каналов

Видно, что первичный сигнал $\lambda_i(t)$ можно выделить из (5.11) с помощью фильтра нижних частот ФНЧ, поскольку частота несущего колебания всегда значительно больше максимальной частоты в спектре модулирующего сигнала. Аналогично выделяется сигнал на выходе второго канала. Влияние каналов друг на друга при правильном выборе параметров системы отсутствует.

5.2.5 Кодовое разделение каналов

Кодовое разделение каналов (КРК) применяется в тех случаях, когда канальные сигналы $S_i(t)$ передаются одновременно и в общей полосе частот. Разделение сигналов осуществляется по их форме, например, по виду манипулирующей ПСП в ШПС с бинарной ФМн или по расположению частотных элементов в ДИС. МКС с КРК также называют **адресными системами**, т.к. сигналы абонента (код их формирования) выполняют роль его адреса.

В таких МКС чаще всего используются квазиортогональные ШПС, сформированные на основе M-последовательностей, последовательностей Баркера и др. Достоинством систем с КРК является возможность совмещения операций разделения сигналов и их демодуляции в одном устройстве, которое также может быть и оптимальным (M-канальный корреляционный приемник).

В международной классификации частотное разделение каналов получила наименование - **многоканальный доступ с кодовым разделением каналов (CDMA)**. В отличие от других технологий радиосвязи, в которых имеющийся частотный спектр разбивается на узкополосные каналы и временные интервалы, в системе CDMA сигналы распределяются в широкой полосе частот. Таким способом система CDMA обеспечивает более эффективное использование имеющегося частотного спектра, обеспечивая значительное увеличение пропускной способности.

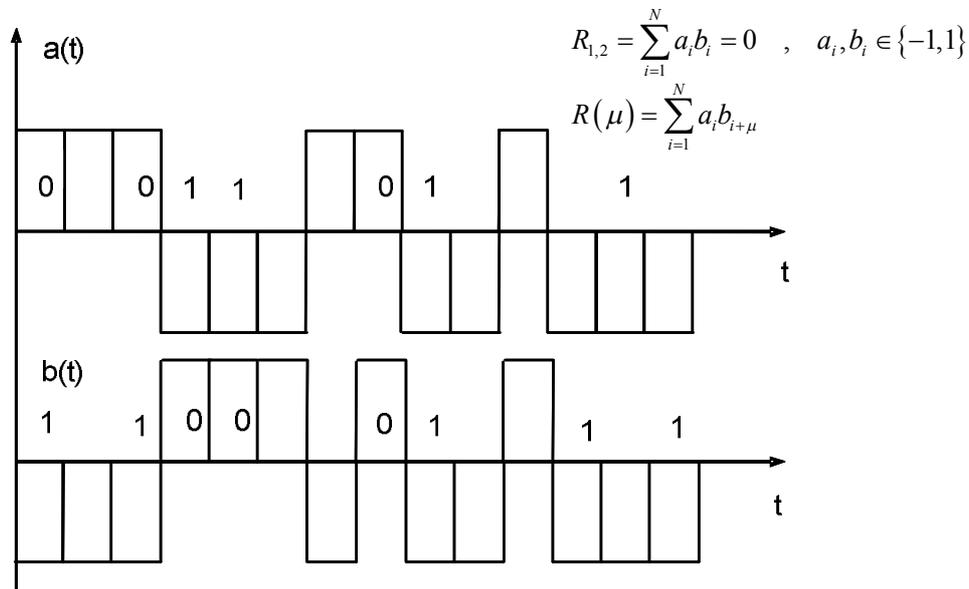


Рисунок 5.9 - Принцип кодового разделения каналов с ортогональными сигналами

Технология CDMA манипулирует тремя параметрами информационно-

го сигнала — частотой, временем и энергией. Канал CDMA является широкополосным (минимальная ширина полосы частот 1,23 МГц). Частотный канал разбивается на несколько десятков логических, передача и приём по которым осуществляется в одной полосе частот одновременно. На передаче сигналы от различных источников обрабатываются каждый своим кодом и объединяются в широкополосный сигнал с распределенной энергией, а на приемной стороне разделяются с помощью аналогичных кодов. Основная идея заключается в том, что в одной и той же полосе частот создаются сигналы, практически не влияющие друг на друга. В отличие от FDMA и TDMA, где энергия сигнала концентрируется на выбранных частотах или временных интервалах, сигналы CDMA распределены в непрерывном частотно-временном пространстве. Одна и та же полоса частот используется всеми каналами одновременно.

Замысел кодового уплотнения каналов на примере линии "вниз", т. е. от базовой станции (БС) к абонентской (АС), сводится к следующему:

- N информационным потокам, предназначенным для N абонентов, присваивается своя кодовая псевдослучайная последовательность (ПСП);
- кодовые последовательности некоррелированы друг с другом;
- бинарные информационные потоки модулируются собственной ПСП;
- каналные широкополосные сигналы складываются в суммирующем устройстве;
- модуляция несущей результирующим сложным широкополосным сигналом и излучение радиосигнала в пространство.

На приемной стороне в абонентской станции:

- известна "своя" кодовая последовательность;
- осуществляется перенос сигнала с радиочастоты в область низких частот;
- низкочастотный импульсный сигнал поступает на вход коррелятора, на второй вход которого синхронно поступает кодирующая ПСП; коррелятор, состоящий из перемножителя и интегратора, вычисляет взаимокорреляционную функцию двух сигналов;
- отклик на выходе коррелятора происходит только тогда, когда в уплотненном сложном сигнале присутствует "своя" ПСП, в противном случае на выходе наблюдается только шум.

5.2.6 Методы комбинированного разделения каналов

В случае комбинированного разделения каналов используется одновре-

менное разделение каналов по нескольким параметрам например по времени, частоте и коду. Такое разделение повышает возможное число квазиортогональных каналов в системе и снижает уровень взаимных помех. В современных помехозащищённых системах передачи используются именно комбинированные способы разделения каналов.

5.2.7 Помехи в многоканальных системах связи

Оценка помехоустойчивости МКС, использующих для передачи сигналов линии радиосвязи, может быть произведена на основе следующей методики: каждый канал связи в МКС можно рассматривать как самостоятельную одноканальную систему связи с определенным видом используемой в ней модуляции сигналов. Мощность радиопередающего устройства МКС в определенной пропорции распределяется по каналам системы. Аналогичные расчеты можно сделать и для МКС с кабельными линиями связи.

Многоканальным системам связи присущи специфические виды помех, так называемые перекрестные помехи. Перекрестные помехи обусловлены тем, что в реальных МКС всегда имеет место проникновение энергии из одного канала в другой.

Основной причиной перекрестных помех в системах с частотным разделением каналов является то, что передаваемый информационный сигнал имеет неограниченный частотный спектр и в то же время полосовые фильтры, применяемые для разделения этих сигналов, обладают амплитудно-частотной характеристикой, отличной от прямоугольной формы. Это приводит к определенному перекрытию спектров передаваемых сигналов.

В МКС с временным разделением каналов вследствие наличия инерционных элементов в тракте передачи рабочие импульсы претерпевают искажения (увеличивается длительность их переднего фронта и среза), что приводит к перекрытию соседних рабочих импульсов группового сигнала по времени и, как результат, к появлению перекрестной помехи.

Предусматриваемые защитные интервалы не полностью устраняют перекрестные помехи. Чрезмерное увеличение этих интервалов приводит к неоправданному расширению полосы частот группового сигнала, что может вызвать снижение помехоустойчивости системы по отношению к внешним помехам и внутренним шумам.

Практически любой тракт передачи вносит как линейные, так и нелинейные искажения. Для МКС с частотным разделением каналов особенно опасны нелинейные искажения, возникающие в тракте группового сигнала, поскольку они связаны с появлением побочных составляющих, которые попадают в определенные каналы данной системы и вызывают перекрестные по-

мехи.

Пусть амплитудная характеристика тракта группового сигнала (рисунок 5.10) аппроксимируется выражением

$$u_{\text{вых}}(t) = a_1 u_{\text{вх}}(t) + a_2 u_{\text{вх}}^2(t) + a_3 u_{\text{вх}}^3(t) + \dots + a_N u_{\text{вх}}^N(t), \quad (5.12)$$

Ограничимся для простоты рассмотрения процесса образования перекрестных помех тем, что нелинейность определяется вторым членом полинома. Найдем побочные колебания, образующиеся вследствие нелинейности в тракте двухканальной системы с частотным разделением сигналов

$$u_1(t) = u_1 \cos \omega_1 t, \quad u_2(t) = u_2 \cos \omega_2 t. \quad (5.13)$$

Предположим, что модуляция отсутствует и амплитуды сигналов одинаковы. Входное напряжение будет равно групповому сигналу

$$u_{\text{вх}}(t) = u_{\Sigma}(t) = u_m (\cos \omega_1 t + \cos \omega_2 t). \quad (5.14)$$

После подстановки (5.14) в (5.12) и соответствующих преобразований получим

$$u_{\text{вых}}(t) = a_1 u_m \cos \omega_1 t + a_1 u_m \cos \omega_2 t + a_2 u_m^2 + \frac{a_2 u_m^2}{2} \cos 2\omega_1 t + \frac{a_2 u_m^2}{2} \cos 2\omega_2 + \quad (5.15) \\ + a_2 u_m^2 \cos(\omega_1 + \omega_2)t + a_2 u_m^2 \cos(\omega_2 - \omega_1)t.$$

Из выражения (5.15) видно, что даже в самом простом случае (отсутствие модуляции, минимальное число каналов, аппроксимация двумя членами) имеет место целый ряд побочных колебаний постоянная составляющая, вторые гармоники сигналов, комбинационные составляющие.

Перекрестные помехи резко возрастает с увеличением количества каналов и при наличии модуляции канальных сигналов, а также с учетом в выражении (5.12) слагаемых высших порядков. Следует отметить, что применение высококачественных фильтров в каналах не приводит к существенному снижению этого вида перекрестных помех, так как неизбежно определенные составляющие от нелинейных искажений оказываются в полосе пропускания данного канала.

При небольшом числе каналов можно избавиться от некоторых составляющих перекрестных помех путем специального выбора значений несущих частот. Наиболее эффективным способом снижения рассматриваемого вида помех является обеспечение линейности тракта группового сигнала. С этой целью при конструировании аппаратуры рабочий участок амплитудной характеристики каждого усилителя и других узлов группового тракта выбирается с малой нелинейностью, широко применяются в схемах глубокие отрицательные обратные связи. Поэтому амплитудную характеристику реального группового тракта вполне можно аппроксимировать полиномом (5.12) не выше третьей степени. Вместе с этим необходимо соответствующим образом выби-

ратель уровни несущих колебаний.

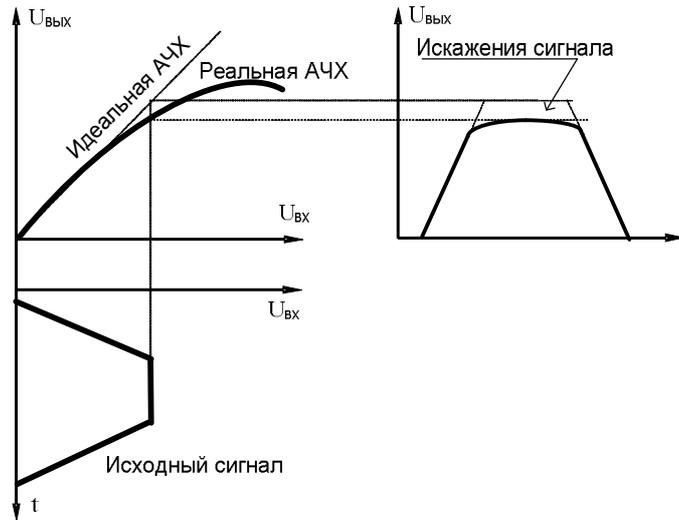


Рисунок 5.11 - Нелинейные искажения импульсных сигналов

В МКС с временным разделением каналов нелинейность тракта группового сигнала проявляется в искажении формы импульса, однако длительность импульса остается неизменной (рисунок 5.11). Поэтому такие искажения не вызывают перекрытия во времени соседних импульсов и не создают перекрестных помех. Для таких систем наиболее опасными являются искажения, вызванные нелинейностью амплитудно-частотной и фазочастотной характеристик группового тракта, так как они приводят к искажению формы импульса с изменением его длительности.

На рисунке 5.12 пунктиром показаны каналные неискаженные импульсы трех каналов. Сплошной линией представлен импульс первого канала, искаженный в связи с наличием в тракте амплитудно-частотных фазочастотных искажений.

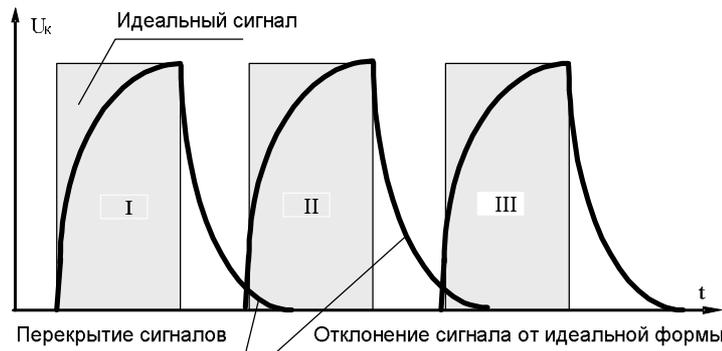


Рисунок 5.12 - Нелинейные искажения импульсных сигналов

Такое перекрытие во времени каналных импульсов приводит к перекре-

стным помехам. Для уменьшения этих помех увеличивают защитный временной интервал между каналами и применяют различные рода коррекции в групповом тракте. Широкое применение в цифровых системах находит регенеративный принцип восстановления формы рабочих импульсов. Это снижает уровень перекрестных помех, повышает помехоустойчивость системы и обеспечивает устойчивую работу оконечной аппаратуры.

5.3 *Авиационно-космические системы радиосвязи*

5.3.1 *Спутниковые системы связи с воздушными судами и их особенности*

Процесс совершенствования систем и средств авиационной радиосвязи в передовых зарубежных странах ориентирован на:

- обеспечение эффективности управления,
- расширение оперативных возможностей,
- интеграцию систем связи с системами автоматизированного боевого управления войсками.

В этой связи на базе новых телекоммуникационных и компьютерных технологий осуществляется:

- интеграция систем и сетей связи всех видов вооруженных сил (от локальных до глобальных) в единую военную глобальную сеть передачи информации с интеграцией услуг;
- разработка и начало реализации концепции создания взаимодействующих местных, национальных и глобальных информационных инфраструктур вооруженных сил в целом и ВВС в частности.

Развитие и совершенствование сопряженных сетей и систем персональной подвижной связи (местного, регионального и глобального охвата) и широкополосных магистральных (на базе волоконно-оптических кабелей и ИСЗ) трансконтинентальных и трансокеанских линий связи с гигабитными потоками обмена информацией.

В целом систему связи можно представить следующей структурой (рис. 5.13):

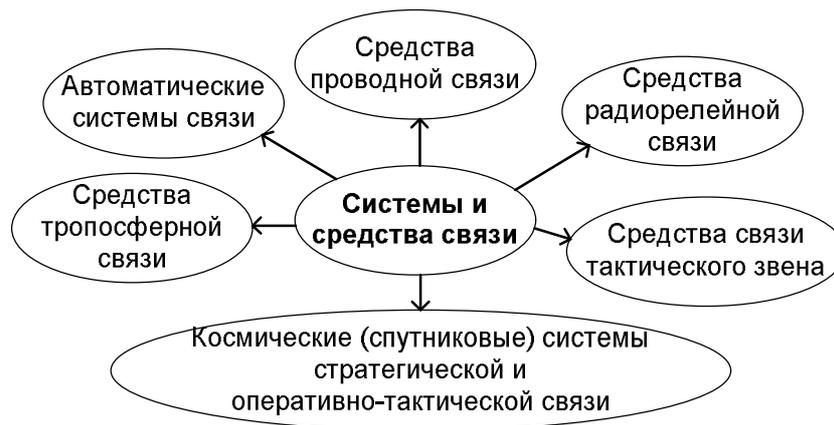


Рисунок 5.13

Здесь КСС занимает одно из ведущих мест в организации требуемого уровня связи. Они обеспечивают широкий спектр услуг связи, из которых следует выделить три основных класса:

- системы пакетной передачи данных;
- системы речевой (радиотелефонной) связи;
- системы для определения местоположения (координат) потребителей.

При радиотелефонной связи в спутниковых системах используют цифровую передачу сообщений с учетом общепринятых международных стандартов. При этом задержка сигнала на трассе распространения не должна превышать 0,3с и обслуживание абонентов должно быть непрерывным и происходить в реальном масштабе времени.

При необходимости абонент может определить свое местоположение (координаты) на Земле при помощи специальной навигационной аппаратуры.

Эта аппаратура по сигналам спутников персональной связи и (или) шлюзовых станций определяет координаты потребителя с достаточной точностью. Спутниковые системы связи обеспечивают достаточно высокую точность определения местоположения (координат) абонента, которая составляет около 300 м (независимо от времени суток, погодных условий и т. п.)².

Для реализации перечисленных услуг спутниковая сеть должна отвечать следующим требованиям:

- спутники должны оснащаться высокоточной системой ориентации для удержания луча их антенны в заданном направлении.
- количество спутников в системе должно быть достаточным для обеспечения сплошного и непрерывного покрытия зоны обслуживания.

² Здесь не рассматривается стандартная навигационная аппаратура GPS систем ГЛОНАСС/НАВСТАР, обеспечивающая очень высокую точность определения координат потребителя

Для обеспечения достаточного количества каналов связи должны применяться многолучевые антенные системы, работающие на высоких частотах³.

Для обеспечения непрерывности радиотелефонной связи через спутник, оснащенный многолучевыми антенными системами, требуется большое количество узловых (шлюзовых) станций с дорогим коммуникационным оборудованием.

Значительный прогресс в развитии спутниковых систем связи достигнут благодаря:

- обработки сигнала на борту спутника-ретранслятора,
- созданию перспективных сетевых протоколов обмена информацией,
- применение недорогих портативных пользовательских терминалов с малым энергопотреблением.
- микроминиатюризации функциональных узлов коммуникационного оборудования.
- применение методов многостанционного доступа с кодовым разделением каналов (CDMA), (основан на использовании широкополосных сигналов).

Главным преимуществом спутниковой связи заключается в отсутствии ограничений по привязке к конкретной местности Земли. Поэтому последнее обстоятельство является определяющим фактором использования ССС в ВВС, когда противоборство конфликтующих сторон сопровождается высокой динамикой изменения их дислокации.

Таким образом, спутниковая связь направлена на качественное изменение всей индустрии телекоммуникаций. Она, в частности, откроет новые возможности для управления ВВС на всех их иерархических уровнях.

Структура спутниковых систем связи

В состав любой спутниковой системы связи (рисунок 5.14) входят следующие составляющие: космический сегмент, состоящий из нескольких спутников ретрансляторов; пользовательский (абонентский) сегмент, осуществляющий связь при помощи персональных спутниковых терминалов; наземные сети связи, с которыми через интерфейсы связи сопрягаются шлюзовые станции космической связи.

³ Более 1,5 ГГц, но при этом значительно усложняет конструкцию антенн КА.

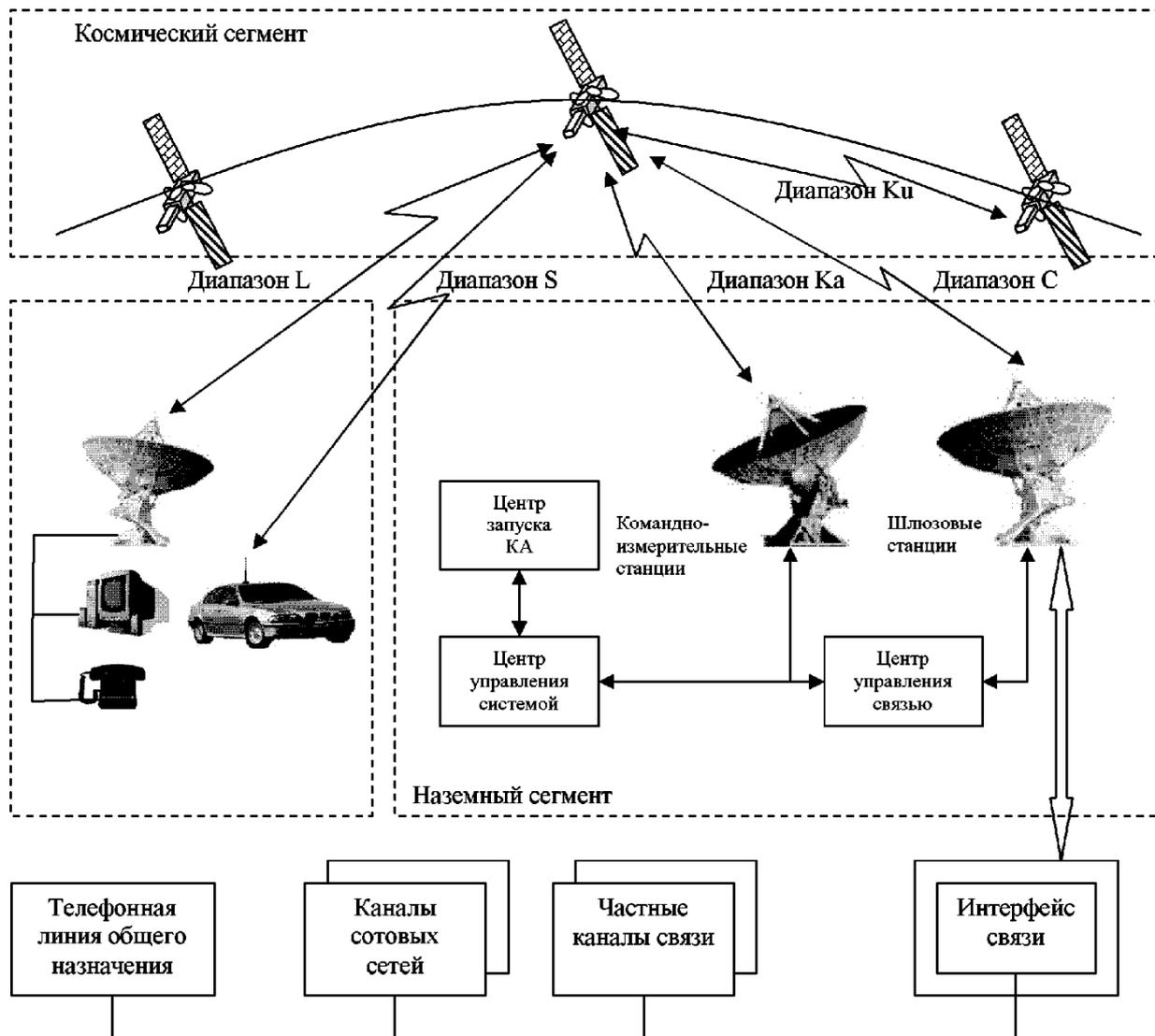


Рисунок 5.14 - Состав спутниковой системы связи

Космический сегмент. Космический сегмент включает в себя несколько спутников-ретрансляторов, которые образуют космическую группировку. Спутники-ретрансляторы, как правило, размещаются равномерно на определенных орбитах.

В состав любого связного КА входят следующие основные элементы:

- Центральный процессор
- Радиоэлектронное оборудование БРТК
- Антенные системы
- Система ориентации и стабилизации
- Двигательная установка
- Система электропитания (аккумуляторы и солнечные батареи)

К каждому КА предъявляются вполне определенные требования по жесткости, расходованию электроэнергии ориентации, стабилизации и т.д. Число спутников в орбитальной группировке является результатом компромисса между стоимостью и желаемым объемом услуг связи с одной стороны, и простотой и ценой персонального спутникового терминала — с другой. Эти сведения надо учитывать при ориентации на ту или другую систему связи.

Наземный сегмент включает в себя:

- центр управления системой;
- центр запуска КА;
- центр управления связью;
- шлюзовые станции.

Центр управления системой осуществляет:

- слежение за КА;
- расчет их координат сверку и коррекцию времени;
- диагностику работоспособности бортовой аппаратуры;
- передачу служебной (командной) информации и т.д.

Указанные функции управления выполняются на основе телеметрической (ТМ) информации, поступающей от каждого КА орбитальной группировки. Для управления космической группировкой в различных режимах работы КА используют как штатные каналы связи (с учетом перекрестной спутниковой связи), так и отдельные территориально разнесенные командно измерительные станции. Благодаря этому центр управления системой позволяет обеспечить с достаточно высокой оперативностью:

- контроль запуска и точность вывода КА на заданную орбиту,
- контроль состояния каждого КА,
- контроль и управление орбитой отдельного КА,
- контроль и управление КА в нештатных режимах работы.

Центр запуска КА определяет:

- программу запуска,
- осуществляет сборку ракеты-носителя, ее проверку,
- установку полезной нагрузки (КА)
- проведение предстартовых проверок и испытаний.

После запуска ракеты-носителя производят траекторные измерения на активном участке полета, которые транслируются в центр управления системой, где для формирования промежуточной орбиты корректируются расчетные траекторные данные.

Центр управления связью. Центр управления связью планирует использование ресурса спутника, координируя эту операцию с центром управления системой. Центр управления связью осуществляет через национальные шлюзовые станции анализ и контроль связи, а также управление.

Шлюзовая станция (шлюз) состоит из нескольких приемопередающих комплексов (обычно не менее трех), в каждом из которых имеется параболическая антенна. Применение нескольких приемопередающих комплексов позволяет практически без нарушения связи переходить последовательно от одного КА к другому. В состав шлюзовой станции включены быстродействующие ЭВМ, в которых имеется банк данных персональных терминалов. Шлюзовые станции в своем составе имеют коммутационное оборудование (интерфейсы связи) для соединения с различными наземными системами связи.

Основной задачей любой шлюзовой станции является:

- организация дуплексной телефонной связи,
- передача факсимильных сообщений,
- передача больших объемов данных.

Персональный пользовательский сегмент. Персональные спутниковые терминалы подвижной связи работают в диапазонах частот 137—900 и 1970—2520 МГц, которые практически не отличаются от диапазона частот сотовой связи (450—1800 МГц). Средняя мощность передатчика невелика и составляет, например, для спутникового терминала системы Iridium 15...400 мВт.

5.3.2 Принципы построения перспективных систем спутниковой связи с многостанционным доступом

Экспериментальные запуски первых советских низкоорбитальных спутников "Стрела-1" начались в 1964 году. После проведения испытаний такие спутники стали запускаться регулярно по 8 штук за один пуск на ракетах-носителях "Космос-3", а позже - "Космос-3М". Более крупный аппарат "Стрела-2" впервые был выведен на орбиту год спустя - в 1965 году. Он был тяжелее "Стрелы-1", однако "Стрела-2" имел при этом и более высокую информативную пропускную способность, чем его предшественник. С конца 60-х годов запуски обоих вариантов "Стрел" стали проводиться в СССР регулярно с целью создания орбитальных группировок и перехода от экспериментальной к штатной эксплуатации систем низкоорбитальной связи. Впоследствии эти аппараты прошли модернизацию, на них изменилась элементная база. "Стрела-1М" и "Стрела-2М" запускались вплоть до начала 90-х годов. С 1985 года началось развертывание более современной системы спутниковой связи на низких орбитах - "Гонец", использовавшей более современный спутник "Стрела-3".

Создание системы "Гонец" предусматривает два этапа.

- На первом этапе будет осуществлено развертывание системы, ныне обозначаемой как "Гонец-Д1".

- На втором этапе развертывания - после 1998 года - система "Гонец" ко всем прочим услугам пакетной связи должна быть оборудована системой межспутниковой связи.

Система "Гонец-ДИМ" должна обеспечивать подвижную помехозащищенную и конфиденциальную персональную связь потребителей, оснащённых маломассогабаритными терминалами, в глобальной зоне обслуживания, а также определение местоположения подвижных объектов.

Система предназначена для решения следующих задач:

- обеспечение связью государственных структур (преимущественно с труднодоступными районами и районами со слаборазвитой инфраструктурой связи);
- обеспечение передачи данных о состоянии и местоположении объектов, в том числе подвижных, и, как следствие, мониторинг грузопотоков (морские и речные суда, автомобильный и железнодорожный транспорт);
- экологический и промышленный мониторинг - автоматический сбор данных с обслуживаемых и необслуживаемых объектов (электростанции, нефтепроводы, газопроводы и др.), с различного типа датчиков охранной или пожарной сигнализации, экологических датчиков и др.;
- срочная межрегиональная доставка информации, связь в чрезвычайных ситуациях (в том числе телемедицина);
- сбор гидрометеорологической информации.

Возможности системы предусматривают (рисунок 5.15):

- передачу в подчиненные части сигналов и команд боевого управления с автоматическим документированным подтверждением их получения;
- передачу телеграмм;
- сбор данных в формализованном виде;
- анализ поступающей информации о проводимых в частях мероприятиях в особый период (по срокам исполнения).

Для системы "Гонец" разработаны абонентские терминалы трех различных назначений.

- ***стационарные абонентские терминалы.*** Они обеспечивают передачу данных со скоростью до 9,6 кбит/с, а на втором этапе развертывания системы "Гонец" - дополнительно и полудуплексную речевую связь.
- ***терминал подвижной связи.*** Он предназначен для установки на транспортных средствах. Такие терминалы также обеспечивают передачу данных со скоростью до 9,6 кбит/с, а на втором этапе допол-

нительно и полудуплексную речевую связь. Кроме того, по желанию заказчика такие терминалы могут быть оснащены дополнительной платой для определения местоположения по сигналам глобальных навигационных систем GPS (США) или ГЛОНАСС (Россия). Терминалы, оснащенные подобными платами, могут использоваться автономно. Так будет реализовываться контроль за местоположением и состоянием транспортных средств.

- **необслуживаемые терминалы**, которые предназначены для автономной работы в сетях дистанционного мониторинга и обеспечивают передачу данных со скоростью 1200 бит/с.

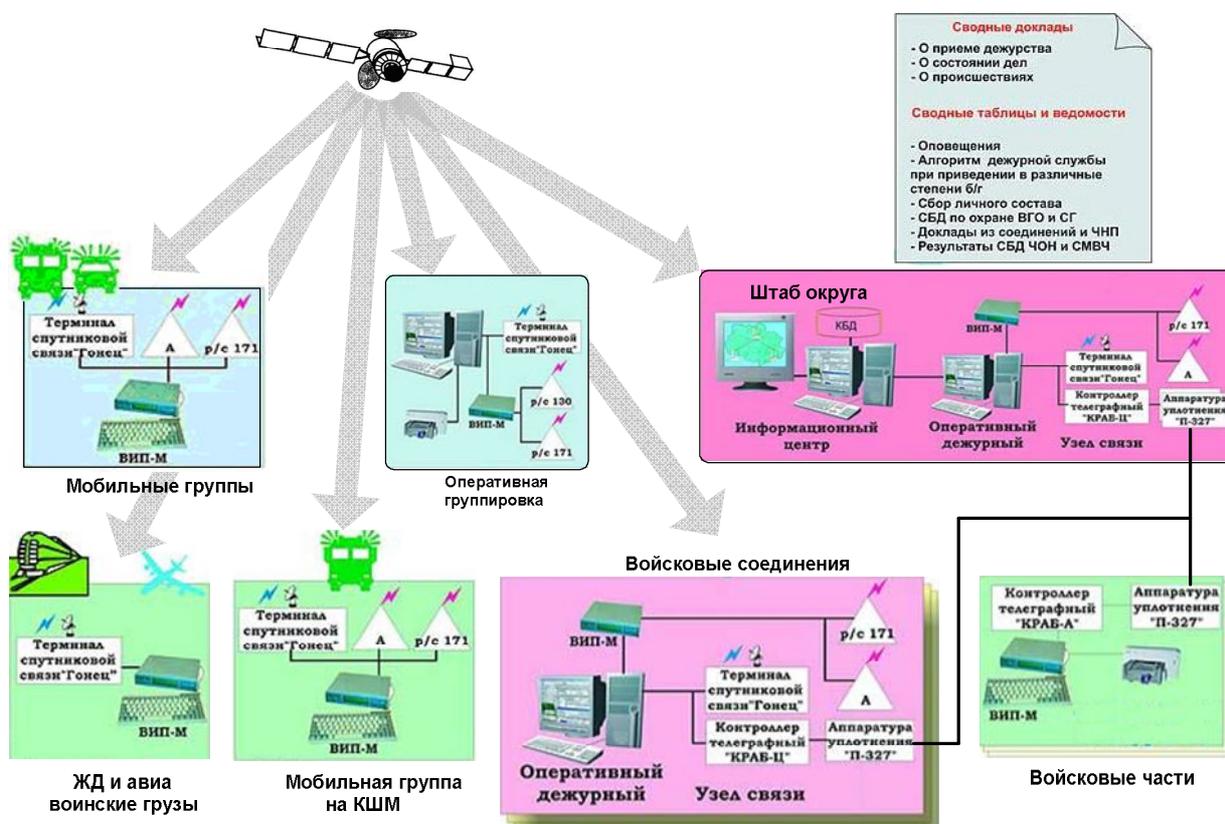


Рисунок 5.15 – Структура доведения сообщений в спутниковой системе связи «Гонец»

5.3.3 Использование космических ретрансляторов

Существенно увеличить территорию охвата сетью можно, если ретранслятор сети установить на возможно большей высоте. Помимо высотных наземных сооружений для этой цели можно использовать и воздушные объекты: аэростаты, вертолеты, самолеты. Однако стоимость поддержания этих объектов в воздухе (за исключением аэростатов) высока и поэтому их можно ис-

пользовать лишь ограниченные периоды времени.

Особый интерес представляет установка ретрансляторов на искусственных спутниках Земли (ИСЗ). ВС континентального и глобального масштабов на основе спутниковой радиосвязи существуют уже многие годы. Как правило, ретрансляторы таких ВС устанавливаются на геостационарных (высота орбиты 36 тыс. км) или орбитальных спутниках, находящихся на высокоэллиптической орбите (высота апогея порядка 40 тыс. км, перигея – 200...400 км), так как геостационарный спутник обеспечивает круглосуточную связь между абонентами, находящимися примерно на одной трети поверхности Земли, а высокоэллиптический обеспечивает периодическую связь на ту же площадь с периодами связи до 11 ч.

Однако значительная высота орбиты и ограниченная мощность передатчика ретранслятора ИСЗ требуют существенной эквивалентной изотропно излучаемой мощности (ЭИИМ), которая определяется как произведение мощности передатчика и коэффициента усиления антенны, а также большой добротности станции, определяемой отношением коэффициента усиления антенны к эквивалентной шумовой температуре приемника. Например, коэффициент затухания энергии радиоволны длиной $\lambda = 0,1$ м при ее распространении до геостационарной орбиты ($r = 36 \cdot 10^6$ м)

$$A = 16 \pi^2 r^2 / \lambda^2 = 2,04 \cdot 10^{19} \text{ (193 дБ)}.$$

Коэффициент усиления антенны зависит от апертуры антенны: чем больше апертура (например, диаметр параболической антенны), тем больше коэффициент усиления

$$G = \frac{4\pi S_3}{\lambda^2},$$

где S_3 – эффективная поверхность (площадь апертуры) антенны.

ЭИИМ передатчиков земных станций (ЗС) 50...95 дБ, в то время как добротность приемной части станций, имеющих антенны диаметром более 10 м, составляет 20 – 31,7 дБ. Именно от размеров антенны земных станций в основном зависят размеры и стоимость всей спутниковой системы связи.

Высокая стоимость спутниковой системы связи требует максимального использования спутниковых ретрансляторов в течение всего срока службы. Для снижения стоимости передачи единицы информации информацию, поступающую от различных источников, перед передачей через спутниковый канал связи «уплотняют» в групповом тракте и передают через ИСЗ на другую станцию, где производится разуплотнение, и информация передается получателям по наземным линиям связи. Таким образом, большинство спутниковых систем связи до последнего времени, по сути, были магистральными групповыми трактами передачи типа «точка – точка» с определенным числом земных станций, являющихся узлами более широкой сети.

В последние годы появилась потребность в организации непосредственной связи между абонентами ВС через ИСЗ. Эта задача обычно решается с использованием геостационарных ИСЗ и спутников на высокоэллиптических орбитах с использованием методов множественного доступа.

Стоимость спутниковой ВС можно существенно уменьшить, а количество абонентов ВС увеличить, если вместо высокоорбитальных спутников использовать ИСЗ на средне- и низковысотных орбитах.

Количество информации, передаваемое в канале связи, $I = \nu t$, где ν – скорость передачи; t – время передачи. Скорость передачи связана с полосой пропускания канала связи Π и равна $m \Pi$, где m – коэффициент использования полосы пропускания, зависящий от вида модуляции (манипуляции). Тогда

$$I = m \Pi t.$$

Определим зависимость Π от высоты орбиты ИСЗ при заданных других параметрах. Допустим земная станция 1 передает одноканальное сообщение на ИСЗ 2 (рисунок 5.16). Мощность сигнала на входе ретранслятора ИСЗ 2

$$P_{C_{21}} = P_1 G_1 G_2 A,$$

где P_1 – мощность земной станции; G_1, G_2 – реальные коэффициенты усиления антенн ЗС и бортового ретранслятора с учетом потерь в фидерных устройствах.

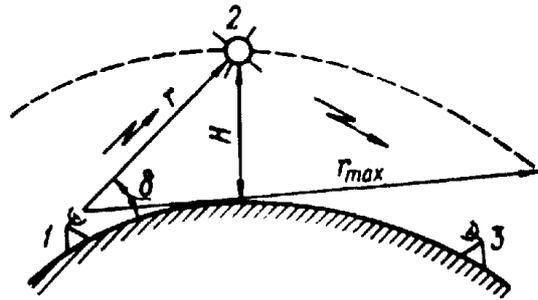


Рисунок 5.16 – Ретрансляция пакетов через космический ретранслятор

Можно записать

$$P_{C_{21}} = \frac{P_1 \alpha_{12}}{r_{12}^2},$$

где $\alpha_{12} = \alpha_1 \alpha_2 = \frac{S_{\theta_1} S_{\theta_2}}{\lambda^2}$ – коэффициент, зависящий от суммарной площади антенн земной станции и ИСЗ.

На выходе ретранслятора 2 ИСЗ кроме сигнала имеется и мощность шумов $P_{Ш_{21}}$, складывающаяся из мощности внешних и внутренних шумов

$$P_{Ш_{21}} = K T_{ВН_{21}} \Pi + K T_{0_2} \Pi = K \Pi (T_{ВН_{21}} + T_{0_2}) = K T_{\Sigma} \Pi$$

K – постоянная Больцмана; $T_{ВН_{21}}$ – эквивалентная температура внешних шумов, наведенных на входе ретранслятора 2; T_{0_2} – эквивалентная температура

внутренних шумов ретранслятора; $T_{\Sigma 2}$ – суммарная шумовая температура ретранслятора.

Мощность суммарного сигнала на выходе ретранслятора, имеющего коэффициент усиления K_P ,

$$P_{\Sigma 23} = K_P (P_{C_{21}} + P_{Ш_{21}}) = P_{C_{23}} + P_{Ш_{23}} = \frac{K_P P_1 \alpha_{21}}{r_{12}^2} + K_P K T_{\Sigma 2} \Pi$$

где $P_{C_{23}}$ $P_{Ш_{23}}$ – мощности сигнала и шума соответственно на выходе ретранслятора.

Мощность сигнала на входе приемной станции 3

$$P_{C_{32}} = \frac{P_{C_{23}} \alpha_{23}}{r_{23}^2} = \frac{K_P P_1 \alpha_{12} \alpha_{23}}{r_{12}^2 r_{23}^2}$$

и мощность шумов

$$P_{Ш_{32}} = \frac{P_{Ш_{23}} \alpha_{23}}{r_{23}^2} + K \Pi (T_{ВН_{23}} + T_{03}) = \frac{K_P \alpha_{23} K T_{\Sigma 2} \Pi}{r_{23}^2} + K T_{\Sigma 3} \Pi,$$

где α_{23} – эквивалентная площадь антенн ИСЗ и земных станций (ЗС); $T_{ВН_{23}}$ – эквивалентная температура внешних шумов на входе приемника ЗС; T_{03} – собственная температура шумов приемника ЗС; $T_{\Sigma 3}$ – суммарная шумовая температура приемника ЗС.

Первое слагаемое в предыдущей формуле представляет собой мощность шумов, наведенных ретранслятором на входе приемника ЗС, второе – мощность шумов, наведенных другими источниками, находящимися в диаграмме направленности антенны ЗС 3, и мощность собственных шумов приемника ЗС 3.

Соотношение сигнал/шум в приемнике ЗС 3

$$h^2 = P_{C_{32}} / P_{Ш_{32}} = \frac{K_P P_1 \alpha_{12} \alpha_{23}}{r_{12}^2 (K_P K T_{\Sigma 2} \Pi \alpha_{23} + r_{23}^2 K T_{\Sigma 3} \Pi)}.$$

Считаем, что шумовая полоса Π соответствует полосе пропускания приемника, тогда

$$\Pi = \frac{K_P P_1 \alpha_{12} \alpha_{23}}{h^2 (r_{12}^2 K_P \alpha_{23} K T_{\Sigma 2} + r_{12}^2 r_{23}^2 K T_{\Sigma 3})}.$$

Учитывая, что $K T_{\Sigma 2} = N_2$ – спектральная плотность шума на выходе ретранслятора, $K T_{\Sigma 3} = N_3$ – спектральная плотность шума на входе приемника ЗС 3, окончательно запишем

$$\Pi = \frac{K_P P_1 \alpha_{12} \alpha_{23}}{h^2 (r_{12}^2 K_P \alpha_{23} N_2 + r_{12}^2 r_{23}^2 N_3)}.$$

Если земные станции расположены на небольшом расстоянии друг от друга, то $r_{12} \approx r_{23} = r$

$$\Pi = \frac{K_p P_1 \alpha_{12} \alpha_{23}}{h^2 (r^2 K_p \alpha_{23} N_2 + r^4 N_3)}$$

Таким образом, при малом коэффициенте усиления ретранслятора ($K_p \leq 1$) полоса пропускания при заданном значении h^2 пропорциональна мощности земного передатчика, произведению эффективной площади антенн обоих участков линии связи и обратно пропорциональна четвертой степени наклонной дальности r , что соответствует режиму пассивной ретрансляции, т. е. ретрансляции за счет отражения от спутника, применявшейся в начале развития спутниковой связи. В этом режиме определяющей является шумовая характеристика земного приемника с учетом шумов, попадающих на его антенну извне (N_3). Если коэффициент усиления ретранслятора большой $K_p \geq 1$ (обычно 1010...1012) и уровни шумов ретранслятора и земного приемника равны ($N_2 \approx N_3$), полоса пропускания определяется суммой членов знаменателя, содержащих r во второй и четвертой степенях. На рисунке 5.13 показана зависимость отношения полосы пропускания канала связи через ИСЗ, находящегося на более низкой орбите, к полосе пропускания канала связи через ИСЗ на геостационарной орбите ($r = 36\,106$ м) от высоты орбиты ИСЗ при неизменных других параметрах устройств линии связи (кривая 1).

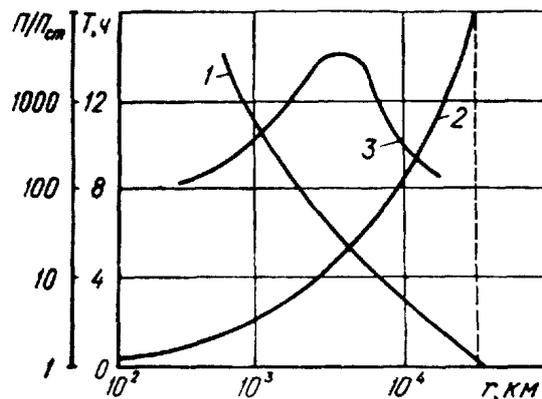


Рисунок 5.13 – Зависимость относительной полосы пропускания (1), среднего времени радиовидимости (2) и количества переданной информации (3) от высоты орбиты ИСЗ

Резкое увеличение полосы пропускания при уменьшении высоты орбиты позволяет теоретически реализовать любую заданную полосу пропускания и, следовательно, скорость передачи. С другой стороны, при заданных соотношении сигнал/шум (т. е. достоверности передачи) и скорости передачи можно определить характеристики земных приемопередающих устройств и ретранслятора. Следует, однако, заметить, что существует предел увеличения полосы пропускания. Он связан с максимальной наклонной дальностью (r_{\max}) от ЗС до ИСЗ, которая для самой низкой орбиты ($H = 200$ км) при угле возвышения

$\delta = 0$ равна $r_{\max} = 1608$ км. На более высоких орбитах при $\delta = 0$ максимальную наклонную дальность определяют по следующей формуле

$$r_{\max} = \sqrt{H(H + 2R_0)},$$

где $R_0 = (6200 \text{ км})$ – радиус Земли.

Следовательно, если использовать всю часть орбиты ИСЗ, находящуюся над горизонтом, то значение r для определения характеристик выбирают как $r = r_{\max}$. Если же использовать орбиты при $\delta > 0$, то значение r выбирают в пределах $H \leq r < r_{\max}$.

Однако при уменьшении H и увеличении δ уменьшается время видимости из одной точки на поверхности Земли и тем более из двух и более точек, поскольку для организации связи через ИСЗ необходима одновременная видимость ИСЗ из этих точек.

Выразить аналитически среднее время сеанса связи T_{cp} между двумя пунктами при произвольных параметрах орбиты не представляется возможным. Такие расчеты выполняют обычно численными или графическими методами.

На рисунке 5.13 показана зависимость среднего времени видимости спутника из одной точки на широте 45° за сутки от высоты орбиты ИСЗ (кривая 2) при угле наклона плоскости орбиты и экватора $i = 83^\circ$. Произведение $m \Pi T_{cp}$ определяет количество информации, которое можно передать через спутник в среднем за сутки (кривая 3 на рисунке 5.13). Из графика видно, что максимальное количество информации можно передать при использовании ИСЗ на орбитах с высотой 1000 – 5000 км. Необходимо отметить, что существует ограничение при использовании низких орбит, связанное со сложностью выделения слишком широкой полосы частот для функционирования ВС в диапазоне ниже 10 – 15 ГГц. Использование же более высоких частот в настоящее время сопряжено с техническими сложностями, повышающими стоимость наземной и бортовой аппаратуры. Например, для создания необходимой ЭИИМ на сверхвысоких частотах необходимо применять антенны с большим коэффициентом усиления, что, в свою очередь, потребует прецизионной системы слежения за спутником, стоимость которой может быть слишком высока, чтобы подобные системы могли использовать абоненты массовых ВС. Поэтому при проектировании ВС на основе спутниковой системы связи имеет смысл построить кривую стоимости передачи одного бита информации от высоты орбиты спутника, выбранного диапазона частот и выделенной полосы, по которой можно определить технико-экономические параметры сети. Для указанных орбит и ограниченной скорости передачи информации (до 100 кбит/с) в диапазоне частот до 0,5 – 1 ГГц возможно использование простой приемопередающей аппаратуры вплоть до переносной.

Использование низко- и средневысоких орбит связано с эффектом Доплера, т. е. изменение частоты принимаемого сигнала и знака этого изменения пропорционально относительной скорости спутника по отношению к земной станции. Причем, чем выше применяемая частота, тем больше абсолютное отклонение частоты.

Для предотвращения ухода частоты за пределы полосы пропускания приемника необходимо либо расширять полосу пропускания приемника на величину максимально возможного сдвига частоты за счет эффекта Доплера (что приводит к увеличению уровня шумов и снижению достоверности передачи), либо применять автоматическую подстройку частоты гетеродина приемника под изменяющуюся частоту сигнала. Для этого можно использовать сигнал, пропорциональный расстройке частоты сигнала радиомаяка спутника и частоты эталонного генератора земной станции.

Увеличить длительность сеансов и среднего времени связи можно за счет использования нескольких ИСЗ на смещенных во времени или пространстве орбитах: после обмена данными через один спутник возможен обмен через другой. Увеличение территории охвата возможно за счет ретрансляции сигналов между ИСЗ при функционировании сети в реальном времени, а также путем записи сообщения в память ЭВМ на борту ИСЗ на одном участке орбиты и передачи ее на Землю на другом по команде с Земли (так называемая космическая почта).

Следует отметить, что один ИСЗ или система ИСЗ могут обслуживать множество региональных сетей, находящихся в различных географических зонах, а также переносить или передавать информацию из одной сети в другую, используя память бортовой ЭВМ или связь между ИСЗ. Структура и алгоритмы взаимодействия в таких ВС во многом сходны с алгоритмами взаимодействия наземных сотовых сетей. Очевидно, что с точки зрения увеличения достоверности передаваемой информации предпочтительными являются спутниковые ретрансляторы с обработкой (регенерацией) сигналу на борту ИСЗ.

В последние годы проведены эксперименты по использованию ИСЗ на средневысоких орбитах для определения характеристик земных станций, работающих в ВС. Передача данных велась через спутники серии «Радио» (международный регистрационный индекс РС5, РС6, РС7, РС8), находящиеся на круговой орбите высотой 1500 км. Период обращения спутников равен 120 мин, угол наклона плоскости орбиты к экватору $i = 83$, смещение за период 30° , диаметр пятна видимости на поверхности Земли около 8000 км, максимальное время видимости спутника из одной точки при орбите, проходящей через зенит, 28 мин, среднее время радиовидимости за сутки из точки на широте 45° около 3,5 ч. Четыре спутника на смещенных орбитах позволяют уве-

личить среднее время видимости приблизительно в 4 раза. Полоса пропускания ретранслятора спутника 40 кГц, что при частотной манипуляции позволяет производить обмен данными со скоростью до 30 кбит/с.

Мощность передатчика бортового ретранслятора 2 Вт, мощность передатчика земной станции 20 Вт. При работе с использованием зенитных орбит применялись ненаправленные антенны с коэффициентом усиления 1, при использовании «горизонтных» орбит – антенны направленного действия с усилением около 14 дБ, что в значительной мере компенсировало потери энергии сигнала за счет увеличения наклонной дальности до 6500 км. Качество приема сигналов определялось уровнем помех в месте приема и составляло $5 \cdot 10^{-2} \dots 10^{-5}$ ош./бит.

На спутниках имелась специализированная ЭВМ с памятью, что позволяло записывать информацию и переносить ее из одной географической зоны в другую (космическая почта). Эксперименты подтвердили правильность теоретических предпосылок и возможность построения недорогих региональных ВС с большим количеством абонентов в каждой, обслуживаемых одним или несколькими ИСЗ. Подобные работы ведутся и за рубежом.

ГЛАВА 6 СИСТЕМЫ ПЕРЕДАЧИ ДАННЫХ

6.1 Общие сведения о системах передачи данных

Главным направлением совершенствования управления на современном этапе является применение автоматизированных систем, обеспечивающих сбор, обработку и передачу информации, проведение оперативных расчетов и т. д. Их широкое внедрение в практику позволит достигнуть резкого повышения оперативности, качества, устойчивости и скрытности управления.

Одно из важнейших мест в автоматизированных системах отводится средствам связи. Они должны осуществлять, автоматический обмен информацией и обеспечивать функционирование системы управления в условиях сложной обстановки.

В автоматизированных системах управления в зависимости от их назначения находят применение различные виды связи: телефонная, телеграфная, факсимильная и главным образом передача данных (ПД) или телекодированная связь. Главное отличие ПД от телефонной, телеграфной и других видов связи заключается в том, что при ней отправителями и получателями сообщений являются различного рода человеко-машинные системы. Важной особенностью систем ПД являются повышенные требования в отношении достоверности передачи, что, как правило, обусловлено большой важностью передаваемой информации и трудностью логического контроля ее в процессе передачи и приема.

6.1.1 Назначение и структура систем передачи данных (СПД)

Системы ПД, обеспечивающие обмен данными между двумя объектами системы управления при выполнении заданных требований по достоверности, надежности и времени передачи являются компонентами более сложных структур – сетей передачи данных.

Все известные системы ПД по используемым в них способам повышения достоверности можно разделить на две основные группы (рисунок 6.1):

- системы без обратной связи,
- системы с обратной связью.

В **системах без обратной связи** информация от отправителя к получателю передается по симплексным каналам связи, состояние которых известно только по априорным сведениям, заведомо неточным. Поэтому энергетические соотношения в системе ПД приходится устанавливать в расчете на неблагоприятные случаи. Поскольку сведения о состоянии канала связи являются ориентировочными, методы повышения достоверности приходится выбирать достаточно сложными, что, естественно, усложняет аппаратуру. Некото-

рое снижение требований к аппаратуре ПД может быть достигнуто за счет повторения передачи и передачи информации по параллельным каналам. Однако это требует, во-первых, известного увеличения энергетических ресурсов и, во-вторых, усложняет аппаратуру.

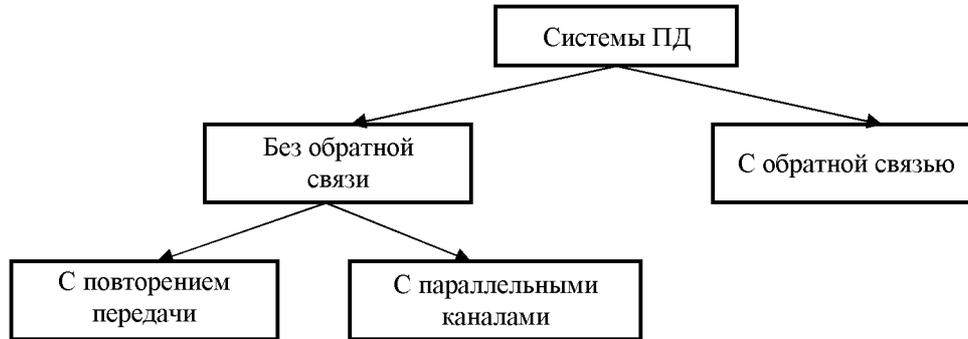


Рисунок 6.1 – Классификация систем передачи данных

Для системы ПД с обратной связью необходимы не симплексные, а дуплексные и полудуплексные каналы связи, что является недостатком таких систем. Однако за счет обратной связи можно существенно повысить достоверность ПД. В наиболее распространенном варианте в прямом канале применяется код, обнаруживающий ошибки. Сведения об обнаруженных ошибках посылаются на передающую сторону, откуда организуется их повторная передача. При этом отпадает необходимость в сложных методах кодирования и, кроме того, устраняются потери информации, которые могут происходить в системах без обратной связи, вследствие отбраковки кодовых комбинаций с обнаруженными ошибками.

В качестве основных характеристик систем ПД, позволяющих производить их оценку и сравнение между собой, используется достоверность, скорость передачи и время задержки сообщений.

Достоверность - степень соответствия принятых сообщений переданным. Возможные несоответствия между ними могут быть вследствие ошибок при вводе информации, воздействия помех в канале связи и т. д. Основными источниками ошибок в принятой информации, как известно, являются каналы связи.

Количественно достоверность наиболее часто оценивается коэффициентом необнаружения ошибки на знак сообщения, каждому из которых соответствует кодовая комбинация первичного кода. Если $A_{но}$ — число знаков, принятых с необнаруженной ошибкой, а $A_{общ}$ — общее количество переданных знаков в заданном интервале времени, то коэффициент необнаружения ошибок

$$K_{но} = \frac{A_{но}}{A_{общ}} . \quad (6.1)$$

Если время передачи выбирать достаточно большим, то $K_{но} \approx P_{но}$, где $P_{но}$ – вероятность необнаружения ошибки.

Известно, системы ПД в зависимости от характеристик используемых незащищенных каналов связи должны обеспечивать градации достоверности передачи информации, приведенные в таблице 6.1.

Таблица 6.1

Максимальное значение коэффициента ошибок на бит в незащищенном канале	Коэффициент необнаруженной ошибки для градации достоверности		
	1	2	3
10^{-2}	10^{-4}	$10^{-6} \dots 10^{-7}$	$10^{-8} \dots 10^{-9}$
10^{-3}	10^{-5}	$10^{-7} \dots 10^{-8}$	$10^{-9} \dots 10^{-10}$

Для характеристики систем передачи данных в основном используются следующие виды скоростей передачи: скорость передачи битов данных, скорость передачи слов (знаков) данных, скорость передачи единичных элементов цифрового сигнала данных (скорость манипуляций или телеграфирования).

Скорость передачи битов данных — скорость передачи символов данных, выраженная числом битов данных, переданных в единицу времени. Единицей измерения этой скорости является бит/с.

Скорость передачи слов (знаков) данных определяется числом слов (знаков) данных, переданных в единицу времени.

Скорость передачи единичных элементов цифрового сигнала данных определяется числом единичных элементов цифрового сигнала данных, переданных в единицу времени. Размерность этой скорости C^{-1} .

Требования по достоверности и скорости передачи противоречивы. В общем случае, чем выше достоверность, тем меньше скорость ПД и наоборот. Поэтому при проектировании систем ПД одной из основных задач является обеспечение максимальной скорости ПД при заданной достоверности.

Время задержки сообщений определяется как время от момента ввода сообщений до момента его выдачи на устройство отображения. Время задержки, как правило, является случайной величиной. Поэтому для оценки системы ПД с точки зрения задержки информации используют функцию распределения времени задержки, т. е. вероятность задержки сообщения из L знаков на время меньше заданного τ_3 : $P[t_3(L) < \tau_3]$.

Системы передачи данных без обратной связи. Структурная схема передающей и приемной частей системы ПД без обратной связи приведена на

рисунке 6.2. Функционирование изображенной системы опишем применительно к сеансному режиму работы, при котором время ожидания сеанса связи значительно превышает его длительность. Сеансный режим работы является наиболее характерным для систем воздушной радиосвязи. Сообщения в таких системах передаются в виде кодограмм.



Рисунок 6.2 – Структурная схема системы ПД без обратной связи

Знаки передаваемого сообщения с помощью устройства ввода преобразуются в кодовые комбинации первичного кода, единичные элементы которых записываются в ячейках памяти передающего накопителя. Из передающего накопителя кодовые комбинации подаются на передающий распределитель, который уже в последовательном коде пересылает их в устройство защиты от ошибок (УЗО) передатчика. Основной смысл преобразований, производимых в передающем накопителе и распределителе, состоит в трансформации временного масштаба сообщений, которая необходима для согласования скоростей ввода информации и ее передачи по каналу связи, а также для возможного добавления в УЗО к кодовым комбинациям первичного кода проверочных элементов, а в некоторых случаях еще и других дополнительных служебных символов.

В УЗО производятся преобразования, предусмотренные принятым в сис-

теме алгоритмом повышения достоверности, т. е. помехоустойчивое кодирование, размножение кодовых комбинаций (если будет осуществляться повторение передачи), добавление служебных знаков и т. д. Далее следует каналообразующая аппаратура, т. е. модулятор и передатчик радиостанции, формирующие высокочастотные сигналы.

Согласование работы отдельных устройств передающей части системы ПД осуществляется устройством управления и синхронизации. От этого же устройства в передаваемые последовательности единичных элементов кодовых комбинаций дополнительно вводится специальная синхронизирующая информация, необходимая на приемной стороне для определения границ единичных элементов и кодовых комбинаций.

В приемной части рассматриваемой системы ПД выполняются преобразования, обратные описанным. Принятые сигналы данных после усиления, преобразования и демодуляции в каналообразующей аппаратуре (приемник радиостанции) подаются в решающее устройство, которое с некоторой вероятностью ошибки принимает решение о том, какой двоичный символ был передан (0 или 1). С выхода решающего устройства сформированные двоичные последовательности поступают в УЗО приемника. Здесь в соответствии с принятым алгоритмом повышения достоверности производится определение или исправление ошибок.

Принятые двоичные последовательности после исключения дополнительных элементов в виде кодовых комбинаций первичного кода через приемный распределитель поступают на элементы памяти приемного накопителя и далее на устройство вывода и отображения данных. Приемный распределитель и накопитель необходимы для согласования скоростей передачи и вывода данных.

Системы ПД без обратной связи находят применение в случаях, когда обратный канал отсутствует или его применение является нежелательным по соображениям скрытности и помехоустойчивости связи. Отсутствие обратного канала является основным преимуществом таких систем ПД. К недостаткам систем без обратной связи следует отнести несогласованность вводимой в информацию избыточности (в силу ее постоянства) с характером помех в канале связи.

6.1.2 Структура сообщений в СПД

Передача сообщений в сеансных системах ПД производится при помощи кодограмм, длительность которых определяется объемом памяти передающего и приемного накопителей. Кодограмма, как правило, состоит из служебных частей в начале и конце и информационной части, содержащей информацию в

формализованном или неформализованном виде.

Применение формализованных структур имеет особое значение при обработке информации на ЭВМ. При использовании формализованных структур кодограмма представляет собой последовательность цифр и букв или только цифр. Основными показателями, определяющими формализованную структуру (кодограмму), является разрядность элемента информации и принятый метод кодирования. Элемент информации содержит две части: постоянную (признак частного объема информации) и переменную (содержание частного объема информации). Разрядность элемента информации определяется количеством знаков, необходимых для передачи постоянной и переменной частей.

Метод кодирования определяется способом обозначения в кодограмме признака элемента информации, т. е. его постоянной части. Наиболее распространенными являются два метода кодирования

- признаковый,
- позиционный.

При признаковом методе кодирования в кодограмме указывается сначала признак элемента информации, а потом его значение. Этот метод кодирования достаточно прост и гарантирует наибольшую надежность. Его удобно использовать при разработке формализованной структуры. Но так как в кодограмме, выполненной по этому методу, указываются признаки элементов информации, объем кодограмм увеличивается. При позиционном методе кодирования каждому элементу информации в кодограмме отводится постоянное место: Это позволяет не включать в кодограмму признак элемента, а ограничиваться указанием только его значения, так как функцию признака в этом случае выполняет заданная позиция. Этот метод кодирования эффективен, когда все или большинство позиций формализованной структуры постоянно заполнены информацией. В противном случае он приводит к значительной избыточности сообщений, так как вместо отсутствующих сведений необходимо передавать нули.

На практике наиболее эффективно применение комбинированного (признако-позиционного) метода кодирования, сочетающего в себе достоинства обоих рассмотренных методов. В этом случае в кодограммах применяется позиционный метод кодирования, и все они классифицируются по типам или форматам. Помимо информации каждая такая кодограмма включает специальный служебный знак (признак), однозначно определяющий тип кодограммы.

Как известно, в сеансных системах ПД приемная часть аппаратуры находится в режиме ожидания приема информации. В связи с этим для данного вида связи возникает задача синхронизации аппаратуры по единичным элементам (тактам) и кодовым комбинациям (группам единичных элементов). По

этой причине в кодограмму должны дополнительно входить комбинация тактовой синхронизации (КТС) и комбинация групповой синхронизации (КГС). Согласно изложенному на рисунке 6.3 приведена структура кодограммы.

Служебная часть кодограммы может содержать сведения об адресе получателя, подписи отправителя, времени передачи кодограммы и т. д. Если длительность кодограммы является постоянной, то служебная часть в конце кодограммы (конец кодограммы) может не передаваться.

КТС	КГС	Служебная часть	Информационная часть	Конец кодограммы
-----	-----	-----------------	----------------------	------------------

Рисунок 6.3 – Структура кодограмм в сеансных системах ПД без обратной связи

Таким образом, целесообразно подчеркнуть:

- В качестве основных характеристик систем ПД, позволяющих производить их оценку и сравнение между собой, используется достоверность, скорость передачи и время задержки сообщений.
- Служебная часть кодограммы может содержать сведения об адресе получателя, подписи отправителя, времени передачи кодограммы и т.д.

6.2 Сети передачи данных

6.2.1 Назначение и структура сетей передачи данных

Сеть связи представляет собой совокупность узлов, линий (каналов) связи и абонентских терминалов и решает задачу по обеспечению связи в соответствии с заданным адресом при выполнении требований по своевременности, достоверности и безопасности связи.

Сети связи различного назначения характеризуются следующими основными признаками:

- упорядоченной структурой;
- единой целью функционирования всех элементов;
- единым принципом функционирования всех элементов сети;

Как правило сеть связи может быть представлена в виде совокупности взаимосвязанных, но функционально обособленных подсистем, работа которых подчинена цели функционирования всей сети.

Абонентские терминалы (оконечная аппаратура абонентов) предназначены для ввода, вывода, отображения, обработки и хранения информации. Как правило, они используются непосредственно пользователем.

Узлы связи решает задачи маршрутизации, коммутации каналов (со-

общений, пакетов), а в ряде случаев и хранение информации.

Маршрутизация - процедура выбора оптимального пути (последовательности линий передач, соединяющих источник с адресатом) прохождения информации при передаче абоненту. Маршрутизация осуществляется устройствами, называемыми **маршрутизаторами**.

Коммутация - процесс соединения каналов связи, либо процесс передачи сообщения (пакета) по выбранному в процессе маршрутизации пути. Долговременная коммутация, часто называемая кроссированием, выполняется на кроссах путем физического соединения требуемых каналов. Оперативная (ручная или автоматическая) коммутация осуществляется на тех узлах связи, где установлены специальные устройства - коммутаторы.

Все сети связи делятся на:

- **Первичные сети связи** создаются на территории страны, области, района или города и представляют собой совокупность средств связи, образующих множество каналов с различной пропускной способностью. Они являются основой для создания вторичных сетей связи.
- **Вторичные сети связи** делятся:
 - по виду передаваемой информации:
 - сети передачи данных,
 - телеграфные сети,
 - телефонные сети и др.;
 - по принадлежности:
 - сети общего пользования (например, городская телефонная сеть, сотовые сети связи)
 - ведомственные (сети связи различных организаций);
 - по способу распределения информации:
 - коммутируемые (с ручной или автоматической коммутацией)
 - некоммутируемые.

СПД различаются по многим признакам, основными из которых являются:

- назначение,
- структура,
- тип используемых каналов связи,
- методы коммутации и т. д.

По назначению СПД делятся на сети:

- **Сети сбора данных** предназначены для сбора данных от территориально удаленных объектов в один центральный пункт управления. Передача данных в сетях этого класса осуществляется по однопользовательным каналам связи.

- **Сети распределения данных** предназначены для распределения и ПД от центрального пункта к удаленным объектам. В этих сетях ПД также осуществляется по однонаправленным каналам связи.
- **Сети сбора и распределения данных** сочетают в себе функции по сбору и распределению данных между взаимодействующими пунктами по двунаправленным каналам связи.
- **Сети обмена данными** предназначены для обеспечения двустороннего информационного обмена в реальном масштабе времени между взаимодействующими оконечными объектами сети по двунаправленным каналам связи.

6.2.2 Структура информационных сетей

Структура любой системы - это фиксированная совокупность элементов и связей между ними.

Структура СПД - совокупность узлов (центров управления, коммутации), линий (каналов) связи между ними и особенностей взаимодействия этих элементов между собой.

Под особенностями взаимодействия понимается, например, наличие старших и подчиненных узлов, основных и резервных линий связи (пунктов управления) и т.д.

*Для описания связей абонентов в сети используется понятие **топология**, под которой принято считать совокупность элементов и связей между ними, «очищенных» от всех свойств, кроме свойств существования и связности.*

Обычно топология задается графом $G(A, N)$, где A - множество вершин графа, т.е. элементов сети (пользователи, узлы, центры коммутации и т. п.), а N - множество его ребер, соответствующих линиям связи. Каждое ребро имеет длину, которая эквивалентна некоторым параметрам его использования, например, ее геометрической длине, пропускной способности, общей загрузке и т.п. в случае, когда учитываются направления ребер задается ориентированный граф (орграф). Пример геометрического представления орграфа показан на рисунке 6.4.

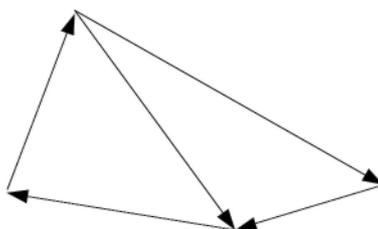


Рисунок 6.4 - Орграф

Существуют три основных вида структур (топологий) СПД:

1. **Полносвязная сеть** - в которой узлы соединены линиями (каналами) по принципу «каждый с каждым» (рисунок 6.5.а).

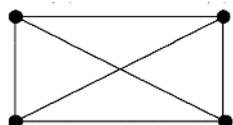


Рисунок 6.5.а - Полносвязная структура сети



Рисунок 6.5.б - Древоподобная структура сети

Полносвязная топология на практике при большом количестве абонентских терминалов практически не используется из-за необходимости организации значительного числа каналов и низкой эффективности их использования, так как большую часть времени каналы простаивают. Тем не менее следует отметить, что такая сеть проста в управлении, обладает высокой живучестью, своевременностью, пропускной способностью.

2. **Древоподобная сеть** - узлы соединяются между собой минимальным числом линий (каналов) без образования замкнутых путей. Между любыми двумя узлами только один путь (рисунок 6.5.б).

В большинстве случаев сетью управляет узел на самом высоком уровне иерархии. Однако, практический интерес представляет распределенный подход к иерархической сети, при котором в системе подчиненных узлов определяются такие, которые обеспечивают непосредственное управление устройствами, находящимися ниже в иерархии.

Древоподобная структура может иметь ряд частных случаев (рисунок 6.6):

- а) линейная,
- б) звездообразная,
- в) радиальная,
- г) радиально-узловая



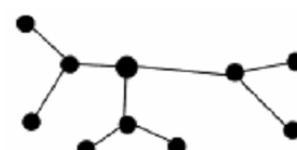
а) линейная



б) звездообразная



в) радиальная



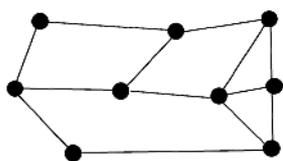
г) радиально-узловая

Рисунок 6.6 - Частные случаи древоподобной структуры сети

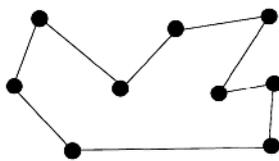
Топология «звезда» - одна из наиболее распространенных. В сетях с

такой топологией вся информационная нагрузка исходит из центрального узла, который полностью управляет устройствами, подсоединенными к нему. Центральный узел отвечает за маршрутизацию трафика через себя в другие компоненты. Он также отвечает за локализацию неисправности, которая является относительно простой в звездообразной сети, поскольку решение проблемы обусловлено возможностью локализации линии. Подобно иерархической структуре, звездообразная сеть также подвержена проблемам, связанными с наличием одного центрального узла.

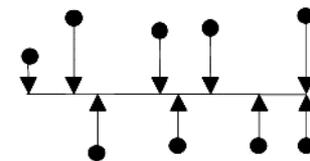
3. **Структура типа «сетка»** - каждый узел соединен с несколькими ближайшими узлами так, что образуются замкнутые пути (рисунок 6.7, а).



а.) Структура сети типа «сетка»



б.) Кольцевая структура сети



в.) Многоточечное подключение

Рисунок 6.7

Топология типа «сетка» нашла широкое применение. Ее привлекательность заключается в высокой надежности и относительной устойчивости к перегрузкам. Благодаря множественности путей трафик может быть направлен в обход отказавших или занятых узлов. Это достигается большей сложностью и дороговизной сети.

Частным случаем структуры типа «сетка» является

- а) **кольцевая структура** (рисунок 6.7,б). При кольцевой топологии в большинстве случаев данные распространяются только в одном направлении, причем только одна станция принимает сигнал, а затем при необходимости передает его следующей станции в кольце.

Кольцевая топология привлекательна, т. к. перегрузки характерные для иерархической и звездообразной конфигураций, здесь достаточно редки. Так же следует отметить простоту организации кольцевой сети. Недостатком является то, что имеется только один канал, соединяющий все компоненты в кольцо. Если отказывает канал между двумя узлами, наступает отказ всей сети.

- б) **многоточечное подключение** (рисунок 6.7,в) - все абонентские терминалы подключаются к одному каналу связи. При этом канал связи используется поочередно одной из пар абонентских терминалов для обмена информацией между собой.

Данная структура нашла широкое распространение в локаль-

ных сетях. Это обусловлено относительно простым для управления трафиком, поскольку шина допускает, чтобы каждое сообщение принималось всеми станциями (одна единственная станция работает в широковещательном режиме на несколько станций).

6.3 Методы коммутации в сетях передачи данных

Коммутируемые сети связи получили широкое распространение, так как по сравнению с некоммутируемыми сетями они являются более экономичными. В них для передачи информации в любом направлении требуется меньшее количество каналов. Некоммутируемая сеть для передачи информации в любом направлении должна быть полносвязной.

Принято различать сети, построенные с использованием:

- выделенных каналов;
- коммутацией каналов,
- коммутацией сообщений,
- коммутацией пакетов.

6.3.1 Выделенные каналы в сетях передачи данных

Выделенные (арендуемые) каналы - арендуемые в интересах ВС РФ у телекоммуникационных компаний, которые владеют каналами дальней связи (таких, например, как «РОСТЕЛЕКОМ»), у телефонных компаний, которые обычно сдают в аренду каналы в пределах города или региона.

Использовать выделенные линии можно двумя способами.

- **Первый вариант** состоит в построении с их помощью территориальной сети определенной технологии, в которой арендуемые выделенные линии служат для соединения промежуточных, территориально распределенных коммутаторов пакетов.
- **Второй вариант** - соединение выделенными линиями только объединяемых локальных сетей или конечных абонентов другого типа, без установки транзитных коммутаторов пакетов, работающих по технологии глобальной сети (рисунок 6.8). Второй вариант является наиболее простым с технической точки зрения, так как основан на использовании маршрутизаторов или удаленных мостов в объединяемых локальных сетях и отсутствии протоколов глобальных технологий, таких как X.25 или frame relay. По глобальным каналам передаются те же пакеты сетевого или канального уровня, что и в локальных сетях.

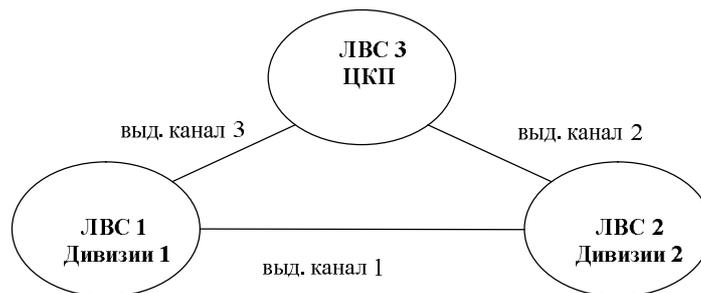


Рисунок 6.8

6.3.2 Метод коммутации каналов в сетях передачи данных

Метод коммутации каналов предполагает необходимость установления физического соединения между абонентами, обменивающимися информацией, как это делается при обычной телефонной связи. При этом на время соединения другие абоненты не могут пользоваться каналом. Если соединение в данный момент не может быть установлено из-за занятости соответствующего направления, то абоненту дается отказ. Вызовы, заканчивающиеся отказом, перегружают сеть. Реальная производительность систем связи с коммутацией каналов составляет 20...30% от пропускной способности.

При коммутации методом коммутации каналов процесс обслуживания на узлах коммутации каналов включает следующие стадии:

- *установление соединения между источником и адресатом с использованием определенной системы сигнализации;*
- *обмен информацией по выделенному при установлении соединения каналу;*
- *разрыв установленного соединения и освобождение канала после окончания обмена информацией.*

Основным недостатком метода является:

- **в случае, когда сообщение короткое, эффективность использования канала очень низкая, так как в основном ресурс сети затрачивается на организацию необходимого соединения.**
- **канал простаивает в периоды организации физического соединения.**
- **канал простаивает в периоды естественных пауз между сообщениями.**

Достоинства метода коммутации каналов:

- **простота оборудования и алгоритма управления,**
- **простота структуры кодограмм,**
- **малая избыточность служебной информации в кодограммах,**
- **малое время задержки сообщений и т. п.**

Вследствие указанных недостатков метод коммутации каналов в сетях

ПД применяется редко (при передаче длинных сообщений в адрес малого числа абонентов, а также при необходимости обмена информацией в реальном масштабе времени).

6.3.3 Метод коммутации сообщений в сетях передачи данных

В сетях связи с коммутацией сообщений любой свободный канал связи выделяется абоненту лишь на время, необходимое для передачи одного сообщения. После окончания передачи этого сообщения канал освобождается, а для передачи нового сообщения выделяется любой свободный в данный момент канал.

В сетях ПД с коммутацией сообщений не требуется предварительного установления всего соединения между объектами (абонентами), обменивающимися информацией. Сообщение, поступившее от отправителя на коммутационный центр, становится в очередь и обслуживается в соответствии с присвоенным ему приоритетом. При этом имеется возможность передавать сообщение по любым свободным участкам сети и хранить сообщение на коммутационных центрах, если какой-либо из последующих участков сети оказывается в данный момент занятым.

В сетях ПД с коммутацией сообщений информация от абонентов в виде кодограмм произвольной длины передается на ближайший узел коммутации сообщений. Кодограмма передается поэтапно от узла к узлу до тех пор, пока не поступит на узел коммутации, к которому подключен абонент-получатель сообщения.

На каждом транзитном участке выполняются следующие основные операции:

- прием сообщения с проверкой на наличие ошибок и при положительном результате отправка кодограммы в накопительное устройство принимающего узла;
- хранение нескольких поступивших кодограмм в накопительном устройстве до момента выдачи на передачу;
- передача кодограмм в соответствии с протоколом обслуживания (например, в порядке поступления, в обратном порядке, с учетом приоритетов и т. п.) следующему узлу в направлении адресата.

Достоинства метода коммутации сообщений по сравнению с методом коммутации каналов:

- Возможность свободного выбора скорости и способа передачи сообщений от отправителя в независимости от скорости и способа работы технических средств получателя, в то время как при коммутации каналов они должны быть идентичными.

- Возможность передачи циркулярных и многоадресных сообщений, что представляет значительные трудности при использовании коммутации каналов.
- Обслуживание объектов со значительно меньшей вероятностью отказов, что значительно увеличивает удобство пользования сетью со стороны абонентов различных категорий.

Метод обеспечивает более эффективное использование каналов благодаря исключению простоя каналов в периоды пауз между сообщениями и пауз при организации физического соединения, что позволяет повысить производительность каналов сети до 95% пропускной способности.

6.3.4 Метод коммутации пакетов в сетях передачи данных

Коммутация пакетов - техника коммутации абонентов, которая была специально разработана для эффективной передачи разнородного трафика.

Эксперименты по созданию первых компьютерных сетей на основе техники **коммутации каналов** показали, что этот вид коммутации не позволяет достичь высокой общей пропускной способности сети. Суть проблемы заключается в пульсирующем характере трафика, который генерируют типичные сетевые приложения. Коэффициент пульсации трафика отдельного пользователя сети, равный отношению средней интенсивности обмена данными к максимально возможной, может составлять 1:50 или 1:100. Если для описанной сессии организовать коммутацию канала между компьютером пользователя и сервером, то большую часть времени канал будет простаивать. В то же время коммутационные возможности сети будут использоваться - часть тайм-слотов или частотных полос коммутаторов будет занята и недоступна другим пользователям сети.

*При коммутации пакетов все передаваемые пользователем сети **сообщения** разбиваются в исходном узле на сравнительно небольшие части, называемые **пакетами**.*

Напомним, что **сообщением** называется логически завершенная порция данных - запрос на передачу файла, ответ на этот запрос, содержащий весь файл, и т. п. Сообщения могут иметь произвольную длину, от нескольких байт до многих мегабайт.

Пакеты обычно тоже могут иметь переменную длину, но в узких пределах, например от 46 до 1500 байт. Каждый пакет снабжается заголовком, в котором указывается адресная информация, необходимая для доставки пакета узлу назначения, а также номер пакета, который будет использоваться узлом назначения для сборки сообщения (рисунок 6.9).

Пакеты транспортируются в сети как независимые информационные

блоки. Коммутаторы сети принимают пакеты от конечных узлов и на основании адресной информации передают их друг другу, а в конечном итоге - узлу назначения.



Рисунок 6.9 - Разбиение сообщения на пакеты

Коммутаторы пакетной сети отличаются от коммутаторов каналов тем, что они имеют внутреннюю буферную память для временного хранения пакетов, если выходной порт коммутатора в момент принятия пакета занят передачей другого пакета (рисунок 6.10).



Рисунок 6.10 - Сглаживание пульсаций трафика в сети с коммутацией пакетов

В этом случае пакет находится некоторое время в очереди пакетов в буферной памяти выходного порта, а когда до него дойдет очередь, то он передается следующему коммутатору. Это позволяет сглаживать пульсации трафика на магистральных связях между коммутаторами и тем самым использовать их наиболее эффективным образом для повышения пропускной способности сети в целом.

Пропускная способность сети коммутацией пакетов

Одним из отличий метода коммутации пакетов от метода коммутации каналов является неопределенность пропускной способности соединения ме-

жду двумя абонентами.

В методе коммутации каналов после образования составного канала пропускная способность сети при передаче данных между конечными узлами известна - это пропускная способность канала. Данные после задержки, связанной с установлением канала, начинают передаваться на максимальной для канала скорости (рисунок 6.11, а). Время передачи сообщения в сети с коммутацией каналов $T_{к.к.}$ равно сумме задержки распространения сигнала по линии связи $t_{а.р.}$ и задержки передачи сообщения $t_{з.п.}$.

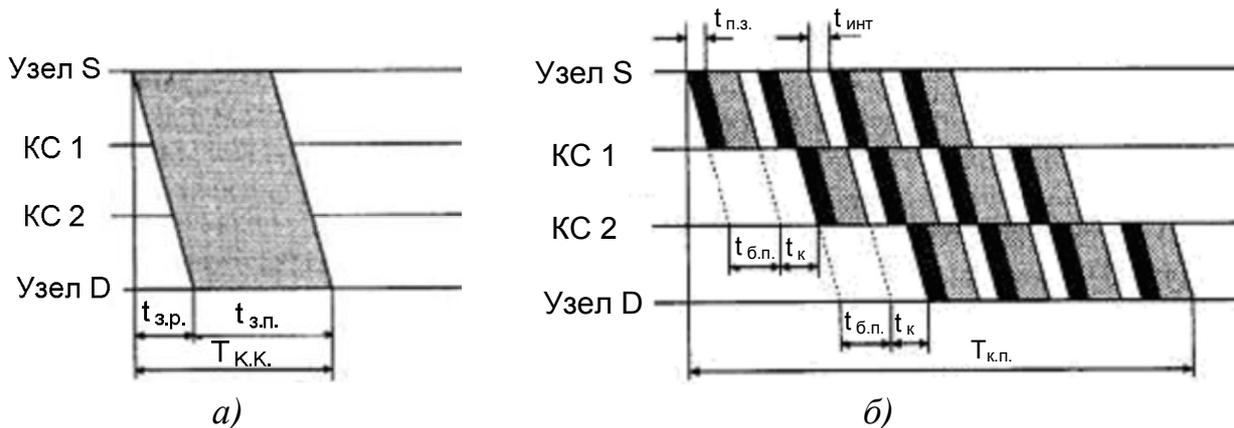


Рисунок 6.11 - Задержки передачи данных в сетях с коммутацией каналов (а) и пакетов (б)

В сети с коммутацией пакетов наблюдается принципиально другая картина. Процедура установления соединения в этих сетях, если она используется, занимает примерно такое же время, как и в сетях с коммутацией каналов, поэтому будем сравнивать только время передачи данных.

На рисунке 6.11, б показан пример передачи в сети с коммутацией пакетов. Предполагается, что в сеть передается сообщение того же объема, что и сообщение, иллюстрируемое рисунком 6.11, а, однако оно разделено на пакеты, каждый из которых снабжен заголовком. Время передачи сообщения в сети с коммутацией пакетов обозначено на рисунке $T_{к.п.}$.

При передаче этого сообщения, разбитого на пакеты, по сети с коммутацией пакетов возникают дополнительные временные задержки. Во-первых, это задержки в источнике передачи, который, помимо передачи собственно сообщения, тратит дополнительное время на передачу заголовков $t_{п.з.}$, плюс к этому добавляются задержки $t_{инт}$, вызванные интервалами между передачей каждого следующего пакета (время на формирование очередного пакета стеком протоколов).

Во-вторых, дополнительное время тратится в каждом коммутаторе. Здесь задержки складываются из времени буферизации пакета $t_{б.п.}$ (коммута-

тор не может начать передачу пакета, не приняв его полностью в свой буфер) и времени коммутации t_k . Время буферизации равно времени приема пакета с битовой скоростью протокола. *Время коммутации складывается из времени ожидания пакета в очереди и времени перемещения пакета в выходной порт.* Если время перемещения пакета фиксировано и обычно невелико (от нескольких микросекунд до нескольких десятков микросекунд), то *время ожидания пакета в очереди колеблется в очень широких пределах и заранее неизвестно, так как зависит от текущей загрузки сети пакетами.*

Неопределенная пропускная способность сети с коммутацией пакетов - это плата за ее общую эффективность при некотором ущемлении интересов отдельных абонентов. *На эффективность работы сети существенно влияют размеры пакетов. Слишком большие размеры пакетов приближают сеть с коммутацией пакетов к сети с коммутацией каналов, поэтому эффективность сети при этом падает. Слишком маленькие пакеты заметно увеличивают долю служебной информации, так как каждый пакет несет с собой заголовков фиксированной длины, а количество пакетов, на которые разбиваются сообщения, будет резко расти при уменьшении размера пакета.*

В сети пакетной радиосвязи могут использоваться два режима передачи пакетов между двумя конечными узлами:

- **дейтаграммный**, при его использовании коммутатор может изменить маршрут какого-либо пакета в зависимости от состояния сети - работоспособности каналов и других коммутаторов, длины очередей пакетов в соседних коммутаторах и т. п. Данный метод рассмотрен выше.
- **передача пакетов по виртуальному каналу**. В этом случае перед тем, как начать передачу данных между двумя конечными узлами, должен быть установлен виртуальный канал, который представляет собой единственный маршрут, соединяющий эти конечные узлы. Виртуальный канал устанавливается при передаче в сеть специального пакета - запроса на установление соединения. Этот пакет проходит через коммутаторы и «прокладывает» виртуальный канал. Это означает, что коммутаторы запоминают маршрут для данного соединения и при поступлении последующих пакетов данного соединения отправляют их всегда по проложенному маршруту.

6.4 Методы повышения достоверности передачи информации в сетях передачи данных

Основными методами повышения достоверности в системах ПД без обратной связи являются:

- *передача с повторением сообщений,*

- передача по параллельным каналам,
- помехоустойчивое кодирование.

6.4.1 Передача с повторением сообщений

Метод повышения достоверности систем ПД за счет *многократного повторения передачи* сообщений является самым старым, но он не потерял своего значения до сих пор, особенно для каналов связи с длинными (охватывающими несколько кодовых комбинаций) пакетами ошибок.

Передача с повторением сообщений использует три основных алгоритма повышения достоверности:

- **Алгоритм поэлементного сравнения** заключается в многократном повторении передачи элементов сообщения. При этом за принятый элемент кодовой комбинации принимается элемент, имеющий наибольшее число совпадений.
- **Алгоритм покомбинационного сравнения** заключается в многократном повторении передачи комбинации. Сложно реализуется на практике и дает худшие результаты по помехоустойчивости. Например, при $v = 3$ вероятность необнаружения ошибки в кодовой комбинации будет определяться выражением $P_{но}(v = 3) = 3n^2P_e^2$ т. е. она в n раз больше, чем при поэлементном сравнении кодовых комбинаций.
- **Алгоритм последовательного анализа кодовой комбинации** предусматривает использование кодов, обнаруживающих ошибки.

Общий недостаток систем ПД, использующих многократное повторение, состоит в значительном снижении скорости передачи данных, так как при v -кратном повторении вводится v -кратная избыточность.

6.4.2 Передача сообщений по параллельным каналам

Метод повышения достоверности систем ПД за счет **передачи по параллельным каналам** используется, прежде всего, по тактическим соображениям:

- повышается живучесть системы связи,
- скорость ПД не снижается по сравнению с одноканальной системой, так как многократная повторение передачи ведется не по одному каналу, а по множеству каналов.
- при достаточном разнесении каналов, потоки ошибок в них оказываются некоррелированными.

Недостаток метода состоит в увеличении расхода сил и средств связи.

В системах ПД без обратной связи с параллельными каналами могут применяться следующие алгоритмы обработки данных.

- **Поэлементное сравнение кодовых комбинаций по критерию большинства.** Алгоритм трудно реализуем (необходимо выравнивать время задержки сигналов в каналах) и требует не менее трех каналов.
- **Покомбинационное сравнение.** Имеет недостатки первого алгоритма и уступает ему в достоверности.
- **Алгоритм, основанный на применении обнаруживающих кодов.** Используется наиболее часто, так как при этом в большинстве случаев можно ограничиться только двумя каналами. В этом случае прием ведется по двум каналам, информация поступает к получателю из какого-нибудь одного из них до тех пор, пока в этом канале не будет обнаружена ошибка. Как только это произойдет, данные начинают поступать к получателю из второго канала, пока в нем, в свою очередь, не будет обнаружена ошибка, после чего осуществляется переход к первому каналу и т. д. В том случае, когда при декодировании в обоих каналах одновременно обнаружатся ошибки, происходит стирание информации (рисунок 6.12).

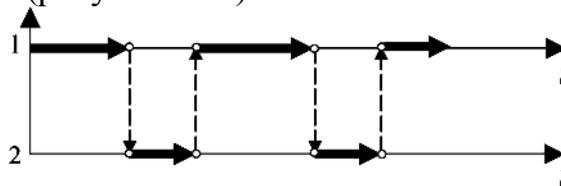


Рисунок 6.12 - Передача данных по двум каналам

6.4.3 Метод использования помехоустойчивого кодирования

Метод повышения достоверности систем ПД за счет использования помехоустойчивого кодирования заключается в использовании избыточного кодирования как самостоятельного метода повышения достоверности.

ГЛАВА 7 ПРИНЦИПЫ ПОСТРОЕНИЯ СОВРЕМЕННЫХ СЕТЕЙ СВЯЗИ

7.1 Международные стандарты построения сетей связи

Существуют две концепции создания вычислительных сетей. Первая основана на том, что распределенные процессы взаимодействуют между собой непосредственно, что, естественно требует, чтобы каждый из них содержал в себе средства, позволяющие правильно понимать действия других процессов. Под процессом понимается программа, или их совокупность, или человек-оператор совместно с терминалом. Недостатки такой концепции – излишняя громоздкость и отсутствие гибкости. Действительно, невозможно создать компактную сетевую операционную систему для всех возможных типов взаимодействующих процессов. Кроме того, включение в сеть новых процессов, отличающихся от действующих, каждый раз потребует модификации системы. Поэтому распространение получила другая концепция, предусматривающая взаимодействие процессов не непосредственно друг с другом, а через некоторую функциональную среду, в которой выполняются определенные и единые для всех процессов правила их взаимодействия в сети. В этом случае на каждый процесс возлагается понимание действия только функциональной среды.

Начиная с первых попыток создания вычислительных сетей стала очевидной необходимость стандартизации функциональной среды. В начале 80-х годов ряд международных организаций по стандартизации - ISO, ITU-T и некоторые другие - разработали модель, которая сыграла значительную роль в развитии сетей. Эта модель *называется эталонной моделью взаимодействия открытых систем (Open System Interconnection, OSI, ЭМВОС) или моделью OSI.*

7.1.1 Эталонная модель взаимодействия открытых систем - Open System Interconnection (OSI)

*В основе эталонной моделью взаимодействия открытых систем ЭМВОС (OSI) лежат стандартные, четко определенные алгоритмы обмена данными и процедур взаимодействия абонентов независимо от их типа, географической удаленности, принципов построения и получивших свое оформление в виде протоколов взаимодействия, имеющих многоуровневую иерархическую архитектуру. В соответствии с многоуровневой концепцией построения сети в модели ЭМВОС процесс взаимодействия различных пользователей делится на семь функциональных слоев, называемых **уровнями**. Каждый уро-*

вень выполняет определенные задачи сети и обеспечивает выполнение функции смежному более высокому уровню.

Базовыми понятиями модели ЭМВОС являются открытость систем, архитектура сети, межуровневый интерфейс, протокол.

Под **системой** в ЭМВОС понимается совокупность одной или нескольких ЭВМ со своим программным обеспечением (ПО), абонентских пунктов, терминалов, людей-операторов, выполняющих функции сбора, хранения, обработки и ввода-вывода информации через сеть. По существу, "система" представляет собой оконечное оборудование пользователя сети, поэтому часто используется термин "абонентская система".

Открытость систем означает возможность взаимодействия абонентских систем между собой с помощью определенных процедур, хотя системы могут быть реализованы на основе различных технических и программных средств. Другими словами, в открытых системах возможно подключение к сети любого абонента независимо от типа оконечного оборудования, лишь бы интерфейс "абонент – сеть" соответствовал определенному стандарту.

Архитектура сети – это общее описание сети, включающее принципы логической, информационной (функциональной) и физической организации сети, определяющее ее элементы и характер их функционального взаимодействия.

В модели OSI (рисунок 7.2) средства взаимодействия делятся на семь уровней: прикладной, представительный, сеансовый, транспортный, сетевой, канальный и физический. Каждый уровень имеет дело с одним определенным аспектом взаимодействия сетевых устройств.

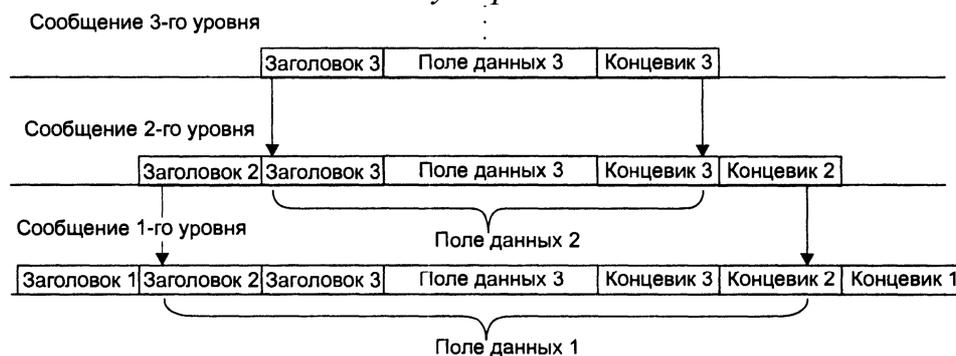


Рисунок 7.1 - Вложенность сообщений различных уровней

После формирования сообщения прикладной уровень направляет его вниз по стеку представителю уровню. Протокол представительного уровня на основании информации, полученной из заголовка прикладного уровня, выполняет требуемые действия и добавляет к сообщению собственную слу-

жебную информацию - заголовок представительного уровня, в котором содержатся указания для протокола представительного уровня машины-адресата. Полученное в результате сообщение передается вниз сеансовому уровню, который в свою очередь добавляет свой заголовок, и т. д. (Некоторые реализации протоколов помещают служебную информацию не только в начале сообщения в виде заголовка, но и в конце, в виде так называемого «концевика».) Наконец, сообщение достигает нижнего, физического уровня, который собственно и передает его по линиям связи машине-адресату. К этому моменту сообщение «обрастает» заголовками всех уровней (рисунк 7.1).

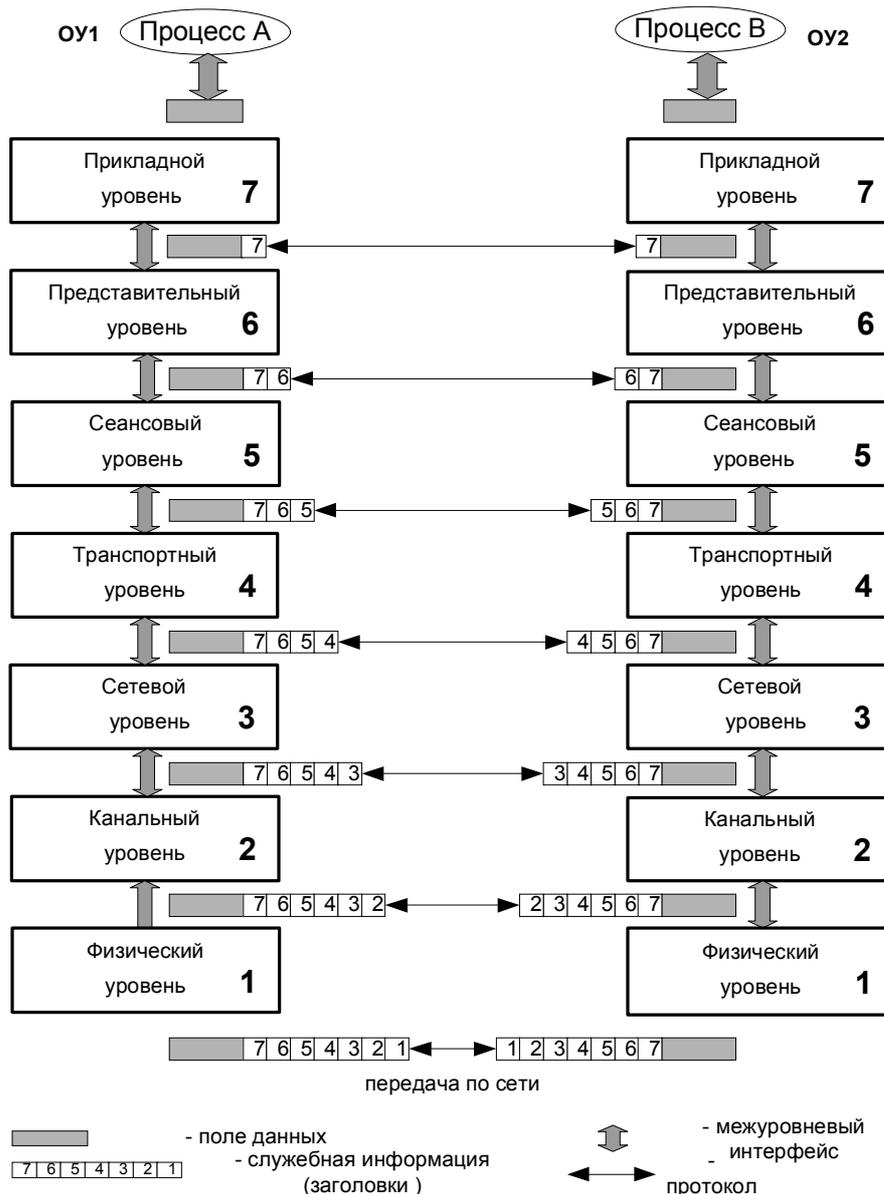


Рисунок 7.2 - Модель взаимодействия открытых систем ISO/OSI

В модели OSI различаются два основных типа протоколов.

В протоколах с установлением соединения (connection-oriented) перед обменом данными отправитель и получатель должны сначала установить соединение и, возможно, выбрать некоторые параметры протокола, которые они будут использовать при обмене данными. После завершения диалога они должны разорвать это соединение. Телефон - это пример взаимодействия, основанного на установлении соединения.

Вторая группа протоколов - протоколы без предварительного установления соединения (connectionless). Такие протоколы называются также дейтаграммными протоколами. Отправитель просто передает сообщение, когда оно готово. Опускание письма в почтовый ящик - это пример связи без предварительного установления соединения. При взаимодействии компьютеров используются протоколы обоих типов.

7.1.2 Уровни модели ЭМВОС (OSI)

Физический уровень

Физический уровень (Physical layer) имеет дело с передачей битов по физическим каналам связи, таким, например, как коаксиальный кабель, витая пара, оптоволоконный кабель или цифровой территориальный канал. К этому уровню имеют отношение характеристики физических сред передачи данных, такие как полоса пропускания, помехозащищенность, волновое сопротивление и другие. На этом же уровне определяются характеристики электрических сигналов, передающих дискретную информацию, например, крутизна фронтов импульсов, уровни напряжения или тока передаваемого сигнала, тип кодирования, скорость передачи сигналов. Кроме этого, здесь стандартизируются типы разъемов и назначение каждого контакта.

Примером протокола физического уровня может служить спецификация 10Base-T технологии Ethernet, которая определяет в качестве используемого кабеля неэкранированную витую пару категории 3 с волновым сопротивлением 100 Ом, разъем RJ-45, максимальную длину физического сегмента 100 метров, манчестерский код для представления данных в кабеле, а также другие характеристики среды и электрических сигналов.

Канальный уровень

Одной из задач канального уровня (Data Link layer) является проверка доступности среды передачи. Другой задачей канального уровня является реализация механизмов обнаружения и коррекции ошибок. Для этого на канальном уровне биты группируются в наборы, называемые кадрами (frames). Канальный уровень обеспечивает корректность передачи каждого кадра, помещая специальную последовательность бит в начало и конец каждого

кадра, для его выделения, а также вычисляет контрольную сумму, обрабатывая все байты кадра определенным способом и добавляя контрольную сумму к кадру. Когда кадр приходит по сети, получатель снова вычисляет контрольную сумму полученных данных и сравнивает результат с контрольной суммой из кадра. Если они совпадают, кадр считается правильным и принимается. Если же контрольные суммы не совпадают, то фиксируется ошибка. Канальный уровень может не только обнаруживать ошибки, но и исправлять их за счет повторной передачи поврежденных кадров. Необходимо отметить, что функция исправления ошибок не является обязательной для канального уровня, поэтому в некоторых протоколах этого уровня она отсутствует, например, в Ethernet и frame relay.

В протоколах канального уровня, используемых в локальных сетях, заложена определенная структура связей между компьютерами и способы их адресации. Хотя канальный уровень и обеспечивает доставку кадра между любыми двумя узлами локальной сети, он это делает только в сети с совершенно определенной топологией связей, именно той топологией, для которой он был разработан. К таким типовым топологиям, поддерживаемым протоколами канального уровня локальных сетей, относятся общая шина, кольцо и звезда, а также структуры, полученные из них с помощью мостов и коммутаторов. Примерами протоколов канального уровня являются протоколы Ethernet, Token Ring, FDDI, 100VG-AnyLAN.

Сетевой уровень

Сетевой уровень (Network layer) служит для образования единой транспортной системы, объединяющей несколько сетей, причем эти сети могут использовать совершенно различные принципы передачи сообщений между конечными узлами и обладать произвольной структурой связей. Функции сетевого уровня достаточно разнообразны.

Протоколы канального уровня локальных сетей обеспечивают доставку данных между любыми узлами только в сети с соответствующей типовой топологией, например топологией иерархической звезды. Это очень жесткое ограничение, которое не позволяет строить сети с развитой структурой, например, сети, объединяющие несколько сетей предприятия в единую сеть, или высоконадежные сети, в которых существуют избыточные связи между узлами. Можно было бы усложнять протоколы канального уровня для поддержания петлевидных избыточных связей, но принцип разделения обязанностей между уровнями приводит к другому решению. Чтобы с одной стороны сохранить простоту процедур передачи данных для типовых топологий, а с другой допустить использование произвольных топологий, вводится дополнительный сетевой уровень.

На сетевом уровне сам термин сеть наделяют специфическим значени-

ем. В данном случае под сетью понимается совокупность компьютеров, соединенных между собой в соответствии с одной из стандартных типовых топологий и использующих для передачи данных один из протоколов канального уровня, определенный для этой топологии.

Внутри сети доставка данных обеспечивается соответствующим канальным уровнем, а вот доставкой данных между сетями занимается сетевой уровень, который и поддерживает возможность правильного выбора маршрута передачи сообщения даже в том случае, когда структура связей между составляющими сетями имеет характер, отличный от принятого в протоколах канального уровня. Сети соединяются между собой специальными устройствами, называемыми маршрутизаторами.

Маршрутизатор - это устройство, которое собирает информацию о топологии межсетевых соединений и на ее основании пересылает пакеты сетевого уровня в сеть назначения. Чтобы передать сообщение от отправителя, находящегося в одной сети, получателю, находящемуся в другой сети, нужно совершить некоторое количество транзитных передач между сетями, или хопов (от hop - прыжок), каждый раз выбирая подходящий маршрут. Таким образом, маршрут представляет собой последовательность маршрутизаторов, через которые проходит пакет.

Проблема выбора наилучшего пути называется **маршрутизацией**, и ее решение является одной из главных задач сетевого уровня. Эта проблема осложняется тем, что самый короткий путь не всегда самый лучший. Часто критерием при выборе маршрута является время передачи данных по этому маршруту; оно зависит от пропускной способности каналов связи и интенсивности трафика, которая может изменяться с течением времени. Некоторые алгоритмы маршрутизации пытаются приспособиться к изменению нагрузки, в то время как другие принимают решения на основе средних показателей за длительное время. Выбор маршрута может осуществляться и по другим критериям, например надежности передачи.

В общем случае функции сетевого уровня шире, чем функции передачи сообщений по связям с нестандартной структурой, которые мы сейчас рассмотрели на примере объединения нескольких локальных сетей. *Сетевым уровнем решаются также задачи согласования разных технологий, упрощения адресации в крупных сетях и создания надежных и гибких барьеров на пути нежелательного трафика между сетями.*

Сообщения сетевого уровня принято называть пакетами (packets). При организации доставки пакетов на сетевом уровне используется понятие «номер сети». В этом случае адрес получателя состоит из старшей части - номера сети и младшей - номера узла в этой сети. Все узлы одной сети должны иметь одну и ту же старшую часть адреса, поэтому термину «сеть» на сетевом уров-

не можно дать и другое, более формальное определение: *сеть* - это совокупность узлов, сетевой адрес которых содержит один и тот же номер сети.

На сетевом уровне определяются два вида протоколов. Первый вид - **сетевые протоколы (routed protocols)** - реализуют продвижение пакетов через сеть. Однако часто к сетевому уровню относят и другой вид протоколов, называемых **протоколами обмена маршрутной информацией или просто протоколами маршрутизации (routing protocols)**. С помощью этих протоколов маршрутизаторы собирают информацию о топологии межсетевых соединений. Протоколы сетевого уровня реализуются программными модулями операционной системы, а также программными и аппаратными средствами маршрутизаторов.

На сетевом уровне работают протоколы еще одного типа, которые отвечают за отображение адреса узла, используемого на сетевом уровне, в локальный адрес сети. Такие протоколы часто называют протоколами разрешения адресов - Address Resolution Protocol, ARP. Иногда их относят не к сетевому уровню, а к канальному, хотя тонкости классификации не изменяют их сути. Примерами протоколов сетевого уровня являются протокол межсетевого взаимодействия IP стека TCP/IP и протокол межсетевого обмена пакетами IPX стека Novell.

Транспортный уровень

На пути от отправителя к получателю пакеты могут быть искажены или утеряны. Хотя некоторые приложения имеют собственные средства обработки ошибок, существуют и такие, которые предпочитают сразу иметь дело с надежным соединением. **Транспортный уровень (Transport layer)** обеспечивает приложениям или верхним уровням стека - прикладному и сеансовому - передачу данных с той степенью надежности, которая им требуется. Модель OSI определяет пять классов сервиса, предоставляемых транспортным уровнем. Эти виды сервиса отличаются качеством предоставляемых услуг: срочностью, возможностью восстановления прерванной связи, наличием средств мультимплексирования нескольких соединений между различными прикладными протоколами через общий транспортный протокол, а главное - способностью к обнаружению и исправлению ошибок передачи, таких как искажение, потеря и дублирование пакетов.

Выбор класса сервиса транспортного уровня определяется, с одной стороны, тем, в какой степени задача обеспечения надежности решается самими приложениями и протоколами более высоких, чем транспортный, уровней, а с другой стороны, этот выбор зависит от того, насколько надежной является система транспортировки данных в сети, обеспечиваемая уровнями, расположенными ниже транспортного - сетевым, канальным и физическим. Так, например, если качество каналов передачи связи очень высокое и вероят-

ность возникновения ошибок, не обнаруженных протоколами более низких уровней, невелика, то разумно воспользоваться одним из облегченных сервисов транспортного уровня, не обремененных многочисленными проверками, квитированием и другими приемами повышения надежности. Если же транспортные средства нижних уровней изначально очень ненадежны, то целесообразно обратиться к наиболее развитому сервису транспортного уровня, который работает, используя максимум средств для обнаружения и устранения ошибок, - с помощью предварительного установления логического соединения, контроля доставки сообщений по контрольным суммам и циклической нумерации пакетов, установления тайм-аутов доставки и т. п.

Как правило, все протоколы, начиная с транспортного уровня и выше, реализуются программными средствами конечных узлов сети - компонентами их сетевых операционных систем. В качестве примера транспортных протоколов можно привести протоколы TCP и UDP стека TCP/IP и протокол SPX стека Novell.

Протоколы нижних четырех уровней обобщенно называют сетевым транспортом или транспортной подсистемой, так как они полностью решают задачу транспортировки сообщений с заданным уровнем качества в составных сетях с произвольной топологией и различными технологиями. Остальные три верхних уровня решают задачи предоставления прикладных сервисов на основании имеющейся транспортной подсистемы.

Сеансовый уровень

Сеансовый уровень (Session layer) обеспечивает управление диалогом: фиксирует, какая из сторон является активной в настоящий момент, предоставляет средства синхронизации. Последние позволяют вставлять контрольные точки в длинные передачи, чтобы в случае отказа можно было вернуться назад к последней контрольной точке, а не начинать все с начала. На практике немногие приложения используют сеансовый уровень, и он редко реализуется в виде отдельных протоколов, хотя функции этого уровня часто объединяют с функциями прикладного уровня и реализуют в одном протоколе.

Представительный уровень

Представительный уровень (Presentation layer) имеет дело с формой представления передаваемой по сети информации, не меняя при этом ее содержания. За счет уровня представления информация, передаваемая прикладным уровнем одной системы, всегда понятна прикладному уровню другой системы. С помощью средств данного уровня протоколы прикладных уровней могут преодолеть синтаксические различия в представлении данных или же различия в кодах символов, например кодов ASCII и EBCDIC. На этом уровне может выполняться шифрование и дешифрование данных, благодаря кото-

рому секретность обмена данными обеспечивается сразу для всех прикладных служб. Примером такого протокола является протокол Secure Socket Layer (SSL), который обеспечивает секретный обмен сообщениями для протоколов прикладного уровня стека TCP/IP.

Прикладной уровень

Прикладной уровень (Application layer) - это в действительности просто набор разнообразных протоколов, с помощью которых пользователи сети получают доступ к разделяемым ресурсам, таким как файлы, принтеры или гипертекстовые Web-страницы, а также организуют свою совместную работу, например, с помощью протокола электронной почты. Единица данных, которой оперирует прикладной уровень, обычно называется сообщением (message).

Существует очень большое разнообразие служб прикладного уровня. Приведем в качестве примера хотя бы несколько наиболее распространенных реализации файловых служб: NCP в операционной системе Novell NetWare, SMB в Microsoft Windows NT, NFS, FTP и TFTP, входящие в стек TCP/IP.

7.2 Протоколы обмена данными в сетях связи

Сетевой протокол – это распределенный алгоритм, выполняемый на нескольких компьютерах, и обеспечивающий унифицированный обмен данными. Основной и первичной задачей сетевых протоколов было обеспечение передачи данных:

- собственно организация сервиса передачи данных;
- организация маршрутизации;
- сигналинг, мониторинг, управление;
- прикладные протоколы: электронная почта терминальный доступ и т.п.

сейчас понятие сетевого протокола в целом сохраняется, но просматривается явная тенденция к усложнению протоколов; число протоколов стремительно растет, причем наибольшие «темпы роста» показывают протоколы прикладного уровня.

7.2.1 Протоколы канального уровня

Канальный уровень предназначен для обеспечения взаимодействия сетей на физическом уровне и контроля за ошибками, которые могут возникнуть. Полученные с физического уровня данные он упаковывает в дейтаграммы, проверяет на целостность, если нужно исправляет ошибки и отправляет на сетевой уровень. Канальный уровень может взаимодействовать с одним или

несколькими физическими уровнями, контролируя и управляя этим взаимодействием.

Канальный уровень разделяется на 2 подуровня:

- *MAC (Media Access Control) регулирует доступ к разделяемой физической среде,*
- *LLC (Logical Link Control) обеспечивает обслуживание сетевого уровня. На этом уровне работают коммутаторы, мосты.*

В программировании канальный уровень представляет драйвер сетевой платы, в операционных системах имеется программный интерфейс взаимодействия канального и сетевого уровней между собой, это не новый уровень, а просто реализация модели для конкретной ОС. Термин "*топология сети*" на канальном уровне относится к пути, по которому данные перемещаются по сети.

Рассмотрим различные типовые архитектуры канального уровня OSI соответствующие основным видам топологий.

Frame relay

Frame relay (англ. *«ретрансляция кадров»*, FR) — протокол канального уровня сетевой модели OSI. Служба коммутации пакетов Frame Relay в настоящее время широко распространена во всём мире. Максимальная скорость, допускаемая протоколом FR — 34.368 мегабит/сек (каналы E3). Коммутация: точка-точка. Топология: звезда.

Frame Relay был создан в начале 1990-х в качестве замены протоколу X.25 для быстрых надёжных каналов связи. В отличие от X.25, рассчитанного на линии с достаточно высокой частотой ошибок, FR изначально ориентировался на физические линии с низкой частотой ошибок.

Frame relay обеспечивает множество независимых виртуальных каналов (Virtual Circuits, VC) в одной линии связи, идентифицируемых в FR-сети по идентификаторам подключения к соединению (Data Link Connection Identifier, DLCI), но не имеет средств коррекции и восстановления. Вместо средств управления потоком включает функции извещения о перегрузках в сети. Возможно назначение минимальной гарантированной скорости (CIR) для каждого виртуального канала.

Token Ring

Token ring — *«маркерное кольцо»*, архитектура кольцевой сети с маркерным доступом. Тип сети, в которой все абоненты схематически объединены в кольцо. По кольцу от абонента к абоненту (станции сети) передается специальный блок данных, называемый **маркером** (англ. *token*).

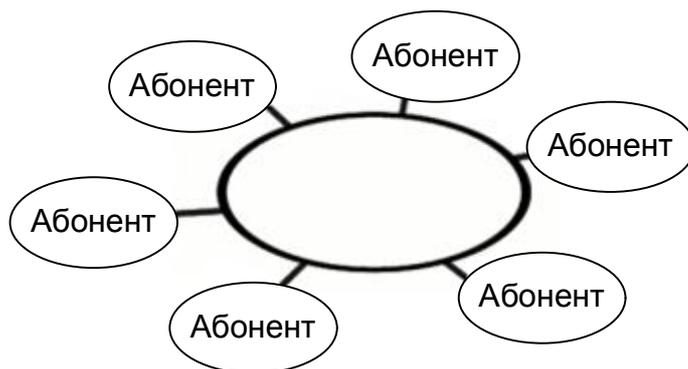


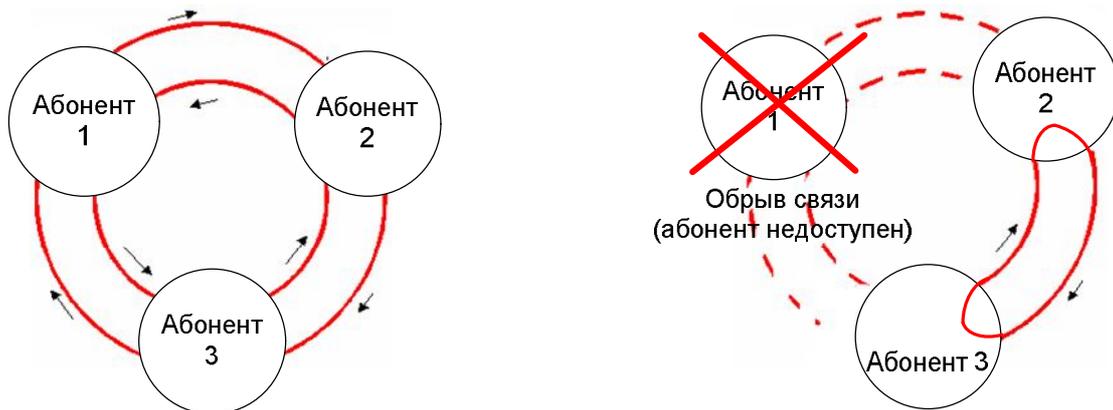
Рисунок 7.3 – Структура сети с маркерным доступом

Сети с передачей маркера перемещают вдоль сети небольшой блок данных, называемый маркером. Владение этим маркером гарантирует право передачи. Если узел, принимающий маркер, не имеет информации для отправки, он просто переправляет маркер к следующей конечной станции. Каждая станция может удерживать маркер в течение определенного максимального времени. Информационный блок циркулирует по кольцу, пока не достигнет предполагаемой станции назначения, которая копирует информацию для дальнейшей обработки. Информационный блок продолжает циркулировать по кольцу; он окончательно удаляется после достижения станции, отославшей этот блок. Станция отправки может проверить вернувшийся блок, чтобы убедиться, что он был просмотрен и затем скопирован станцией назначения.

Данная технология предлагает вариант решения проблемы коллизий, которая возникает при работе локальной сети. В технологии Ethernet, такие коллизии возникают при одновременной передаче информации несколькими рабочими станциями, находящимися в пределах одного сегмента, то есть использующих общий физический канал данных. Однако, в настоящее время по надежности Ethernet не уступает Token Ring и существенно выше по производительности.

FDDI

FDDI (англ. *Fiber Distributed Data Interface* — распределённый волоконный интерфейс данных) — стандарт передачи данных в локальной сети, протянутой на расстоянии до 200 километров. Стандарт основан на протоколе Token Ring. Кроме большой территории, сеть FDDI способна поддерживать несколько тысяч пользователей.



а.) Нормальный режим функционирования сети FDDI б.) Функционирования сети FDDI при обрыве канала связи

Рисунок 7.4 - Функционирование сети FDDI

В качестве среды передачи данных в FDDI рекомендуется использовать оптоволоконный кабель, однако можно использовать и медный кабель, в таком случае используется сокращение CDDI (Copper Distributed Data Interface). В качестве топологии используется схема двойного кольца, при этом данные в кольцах циркулируют в разных направлениях. Одно кольцо считается основным, по нему передаётся информация в обычном состоянии; второе — вспомогательным, по нему данные передаются в случае обрыва на первом кольце. Для контроля за состоянием кольца используется сетевой маркер, как и в технологии *Token Ring*.

Поскольку дублирование повышает надёжность системы, данный стандарт с успехом применяется в магистральных каналах связи.

Технология Ethernet

Ethernét (*эзернет*, от лат. *aether* — эфир) — пакетная технология компьютерных сетей, преимущественно локальных. Технология Ethernet является классической реализацией топологии «общая шина».

Стандарты Ethernet определяют проводные соединения и электрические сигналы на физическом уровне, формат кадров и протоколы управления доступом к среде — на канальном уровне модели OSI. Ethernet в основном описывается стандартами IEEE 802.3. Ethernet стал самой распространённой технологией построения сетей.

В стандарте первых версий Ethernet в качестве передающей среды использовался коаксиальный кабель, в дальнейшем появилась возможность использовать витую пару, оптический кабель и радиоканал.

Метод управления доступом — множественный доступ с контролем несущей и обнаружением коллизий (CSMA/CD, Carrier Sense Multiple Access

with Collision Detection), скорость передачи данных 10 Мбит/с, размер пакета от 72 до 1526 байт, описаны методы кодирования данных. Количество узлов в одном разделяемом сегменте сети ограничено предельным значением в 1024 рабочих станции.

Для адресации абонента в Ethernet сети используется **MAC адрес**. Это уникальный шестнадцатиричный серийный номер, назначаемый каждому сетевому устройству Ethernet, для идентификации его в сети. MAC адреса имеют длину 6 байт и обычно записываются шестнадцатиричным числом в виде 12:34:56:78:90:AB (двоеточия могут отсутствовать, но их наличие делает адрес более читабельным).

CSMA/CD (*Carrier-Sense Multiple Access with Collision Detection* — *множественный доступ с контролем несущей и обнаружением коллизий*) - технология множественного доступа к общей передающей среде в локальной сети с контролем коллизий. Если во время передачи фрейма рабочая станция обнаруживает другой сигнал, занимающий передающую среду, она останавливает передачу, посылает jam signal и ждет в течение случайного промежутка времени (известного как «backoff delay» и находимого с помощью алгоритма *truncated binary exponential backoff*), перед тем как снова отправить фрейм.



Рисунок 7.5 - Пример работы механизма CSMA/CD

Механизм передачи CSMA/CD (Carrier Sense Multiple Access with Collision Detection) решает следующие задачи:

- Контроль среды передачи (радиосреды либо кабеля). Когда абонент собирается передавать данные в сеть методом CSMA/CD, он должен сначала проверить, передает ли в это же время по этому же кабелю

свои данные другой абонент. Другими словами проверить состояние носителя: занят ли он передачей других данных.

- Множественный доступ. Это означает, что несколько абонентов могут начать передачу данных в сеть одновременно.
- Обнаружение конфликтов. Это главная задача метода CSMA/CD. Когда абонент готов передавать данные, он проверяет состояние носителя. Если кабель занят, абонент не посылает сигналы, а ожидает когда освободится носитель. Если же абонент не слышит передаваемых данных в носителе, то он начинает передавать данные сам. Так же может случиться, что прослушивают носитель сразу два абонента готовые к передаче. И когда носитель освобождается, оба узла начинают передавать данные одновременно. В этом случае в носителе происходит смешение сигналов, из-за чего теряются данные. Такая ситуация называется конфликтом или **коллизией** (рисунок 7.5).

Во время коллизии прием данных невозможен, поэтому оба абонента ожидают в течении случайного промежутка времени, после которого посылают сигналы заново. Случайные промежутки времени для разных абонентов создаются автоматически, для того чтобы не повторить коллизию. Абонент, у которого промежуток времени меньше, начнет передавать данные раньше. Таким образом, происходит "захват" носителя для передачи данных.

Обнаружение коллизий используется для улучшения производительности CSMA с помощью прерывания передачи сразу после обнаружения коллизии и снижения вероятности второй коллизии во время повторной передачи. Методы обнаружения коллизий зависят от используемого оборудования, но на электрических шинах, таких как Ethernet, коллизии могут быть обнаружены сравнением передаваемой и получаемой информации. Если она различается, то другая передача накладывается на текущую (возникла коллизия) и передача прерывается немедленно. Посылается jam signal («пробка»), что вызывает задержку передачи всех передатчиков на произвольный интервал времени, снижая вероятность коллизии во время повторной попытки (рисунок 7.6).

Обнаружение коллизий используется для улучшения производительности CSMA с помощью прерывания передачи сразу после обнаружения коллизии и снижения вероятности второй коллизии во время повторной передачи. Методы обнаружения коллизий зависят от используемого оборудования, но на электрических шинах, таких как Ethernet, коллизии могут быть обнаружены сравнением передаваемой и получаемой информации. Если она различается, то другая передача накладывается на текущую (возникла коллизия) и передача прерывается немедленно. Посылается jam signal («пробка»), что вызывает задержку передачи всех передатчиков на произвольный интервал времени, снижая вероятность коллизии во время повторной попытки (рисунок 7.6).

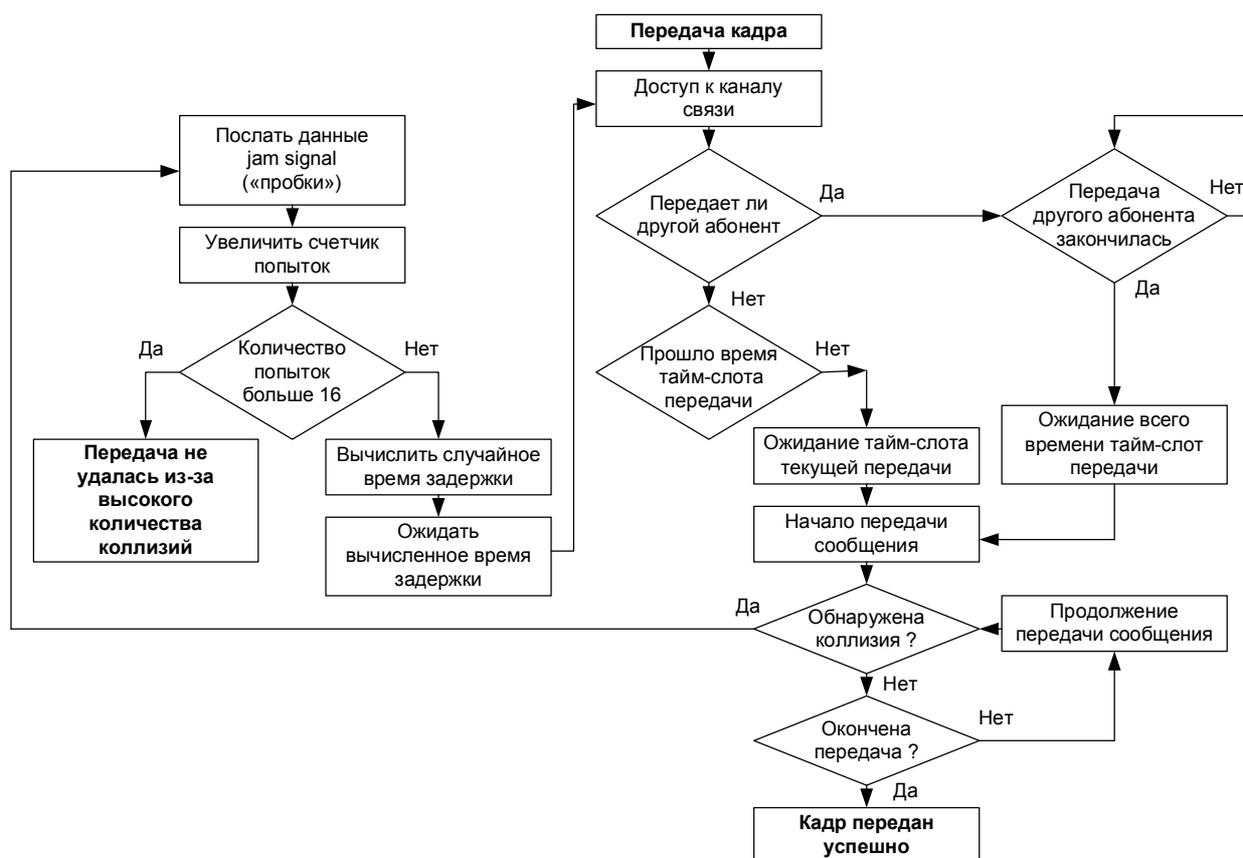


Рисунок 7.6 - Алгоритм механизма CSMA/CD

7.2.2 Протоколы сетевого уровня

В стандартной модели взаимодействия открытых систем в функции сетевого уровня входит решение следующих задач:

- передача пакетов между конечными узлами в составных сетях;
- выбор маршрута передачи пакетов, наилучшего по некоторому критерию;
- согласование разных протоколов канального уровня, используемых в отдельных подсетях одной составной сети.

Протоколы сетевого уровня реализуются, как правило, в виде программных модулей и выполняются на **конечных узлах**, а также на промежуточных узлах - **маршрутизаторах**, называемых **шлюзами**.

Основная идея введения сетевого уровня состоит в следующем. *Сеть в общем случае рассматривается как совокупность нескольких сетей и называется **составной сетью** или **интерсетью** (internetwork или internet). Сети, входящие в составную сеть, называются **подсетями** (subnet), составляющими сетями или просто сетями.*

Все узлы в пределах одной подсети взаимодействуют, используя единую для них технологию. Так, в составную сеть, показанную на рисунке, входит несколько сетей разных технологий: локальные сети Ethernet, Fast Ethernet, Token Ring, FDDI и глобальные сети frame relay, X.25, ISDN. Каждая из этих технологий достаточна для того, чтобы организовать взаимодействие всех узлов в своей подсети, но не способна построить информационную связь между произвольно выбранными узлами, принадлежащими разным подсетям.

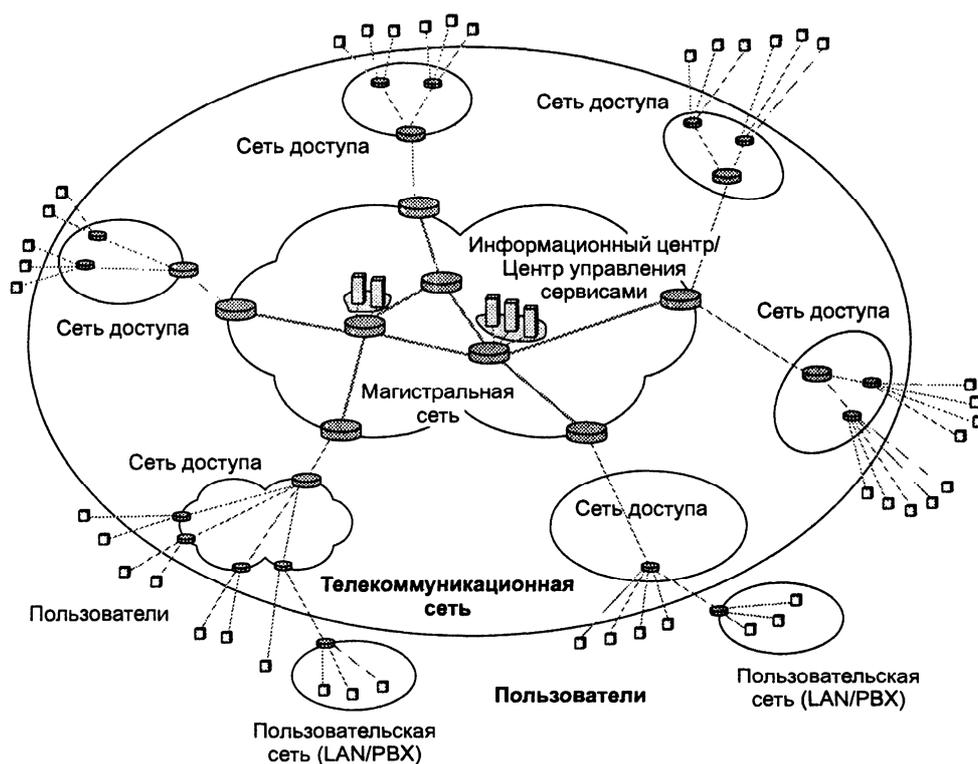


Рисунок 7.7 – Составная сеть

TCP/IP (Transport Control Protocol/Internet Protocol) — это профессиональный коммуникационный протокол, который позволяет связать различные вычислительные или информационные системы в единую глобальную информационную сеть. В настоящее время использование протокола TCP/IP планируется в перспективной системе ОСНОД.

Функции коммутации выполняют маршрутизаторы по сетевому адресу (позывному) корреспондента. При этом используется весь совокупный коммуникационный ресурс сети и информация может следовать к корреспонденту различными путями и каналами.

Для персональной идентификации каждая TCP/IP станция имеет свой уникальный (неповторяемый) IP адрес.

Адресация на сетевом уровне

Протокол TCP/IP является на настоящий момент наиболее распространенным протоколом сетевого и транспортного уровней. Поэтому реализацию конечных протоколов данных уровней OSI рассмотрим на их примере.

В стеке TCP/IP используются три типа адресов:

- *локальные (называемые также аппаратными или MAC-адресами),*
- *символьные доменные имена*
- *IP-адреса.*

В терминологии TCP/IP под локальным адресом (MAC-адресом) понимается такой тип адреса, который используется средствами базовой технологии для доставки данных в пределах сети, являющейся элементом более глобальной интeрсети.

Если подсетью интeрсети является локальная сеть, то локальный адрес - это MAC - адрес. MAC - адрес сетевого адаптера и сетевого интерфейса маршрутизаторов. MAC - адреса назначаются производителями оборудования и являются уникальными, так как управляются централизованно. Для всех существующих технологий локальных сетей MAC - адрес имеет формат 6 байт, например 11-A0-17-3D-BC-01.

Символьные доменные имена. Символьные имена в IP-сетях называются доменными и строятся по иерархическому признаку. Составляющие полного символьного имени в IP-сетях разделяются точкой и перечисляются в следующем порядке:

- сначала простое имя конечного узла,
- затем имя группы узлов (например, имя организации),
- затем имя более крупной группы (поддомена) и так до имени домена самого высокого уровня (например, домена объединяющего организации по географическому принципу: RU - Россия, UK - Великобритания, SU - США),

Примеров доменного имени может служить имя base2.sales.zil.ru. Между доменным именем и IP-адресом узла нет никакого алгоритмического соответствия, поэтому необходимо использовать дополнительные таблицы или службы, чтобы узел сети однозначно определялся как по доменному имени, так и по IP-адресу. В сетях TCP/IP используется специальная распределенная служба **Domain Name System (DNS)**, которая устанавливает это соответствие на основании создаваемых администраторами сети таблиц соответствия. Поэтому доменные имена называют также **DNS-именами**,

IP-адреса представляют собой основной тип адресов, на основании которых сетевой уровень передает пакеты между сетями. IP-адрес состоит

из двух частей: номера сети и номера узла, при этом в качестве признака номера сети выступает маска сети.

В последнее время поле номера сети в адресе называется **сетевым префиксом**, так как первая порция каждого IP-адреса идентифицирует номер сети. Все станции определенной сети используют один и тот же сетевой префикс, но при этом должны иметь уникальные номера устройств. Аналогично две любые станции, расположенные в разных сетях, должны иметь различные сетевые префиксы, но при этом они могут иметь одинаковые номера устройств.

IP-адрес имеет длину 4 байта и обычно записывается в виде четырех чисел, представляющих значения каждого байта в десятичной форме и разделенных точками, например,

128.10.2.30 - традиционная десятичная форма представления адреса,

10000000 00001010 00000010 00011110 - двоичная форма представления этого же адреса.

***Маска** - это число, которое используется в паре с IP-адресом; двоичная запись маски содержит единицы в тех разрядах, которые должны в IP-адресе интерпретироваться как номер сети.*

Поскольку номер сети является цельной частью адреса, единицы в маске также должны представлять непрерывную последовательность.

Для стандартных классов сетей маски имеют следующие значения:

- класс А - 11111111. 00000000. 00000000. 00000000 (255.0.0.0);
- класс В - 11111111. 11111111. 00000000. 00000000 (255.255.0.0);
- класс С - 11111111. 11111111. 11111111. 00000000 (255.255.255.0).

В масках количество единиц в последовательности, определяющей границу номера сети, не обязательно должно быть кратным 8, чтобы повторять деление адреса на байты. Пусть, например, для IP-адреса 129.64.134.5 указана маска 255.255.128.0, то есть в двоичном виде:

- IP-адрес 129.64.134.5 - 10000001. 01000000. 10000110. 00000101
- Маска 255.255.128.0 - 11111111. 11111111. 10000000. 00000000

Снабжая каждый IP-адрес маской, можно отказаться от понятий классов адресов и сделать более гибкой систему адресации. Например, если рассмотренный выше адрес 185.23.44.206 ассоциировать с маской 255.255.255.0, то номером сети будет 185.23.44.0, а не 185.23.0.0, как это определено системой классов.

Таблица 7.1 - Зарезервированные IP-адреса

IP - адрес		Примечания
Все биты установлены в «0»		Данное устройство
Номер сети	Все биты номера равны 0	Данная IP-сеть
Все биты равны 0	Номер устройства	Устройство в данной IP-сети
Все биты установлены в «1»		Все устройства в данной IP-сети
Номер сети	Все биты номера равны 1	Все устройства в указанной сети
127 (десятичное)	Что-нибудь (обычно 1)	Адрес обратной связи

При адресации необходимо учитывать те ограничения, которые вносятся особым назначением некоторых IP-адресов. Так, ни номер сети, ни номер узла не может состоять только из одних двоичных единиц или только из одних двоичных нулей. Отсюда следует, что максимальное количество узлов, приведенное в таблице для сетей каждого класса, на практике должно быть уменьшено на 2.

Особый смысл имеет IP-адрес, первый сокет которого равен 127. Он используется для тестирования программ и взаимодействия процессов в пределах одной машины т. е. образуется как бы «петля». Данные не передаются по сети, а возвращаются модулям верхнего уровня как только что принятые. Поэтому в IP-сети запрещается присваивать машинам IP-адреса, начинающиеся со 127.

Отображение доменных имен на IP адреса

Для идентификации узлов аппаратное и программное обеспечение в сетях TCP/IP полагается на IP-адреса, поэтому для доступа к сетевому ресурсу в параметрах программы вполне достаточно указать IP-адрес. Однако пользователи обычно предпочитают работать с символьными именами. Следовательно, в сетях TCP/IP должны существовать символьные имена абонентов и механизм для установления соответствия между символьными именами и IP-адресами.

Для эффективной организации именованя узлов в больших сетях естественным является применение **иерархических составных имен**. В стеке TCP/IP применяется доменная система имен, которая имеет иерархическую древовидную структуру, допускающую использование в имени произвольного количества составных частей (рисунок 7.8).

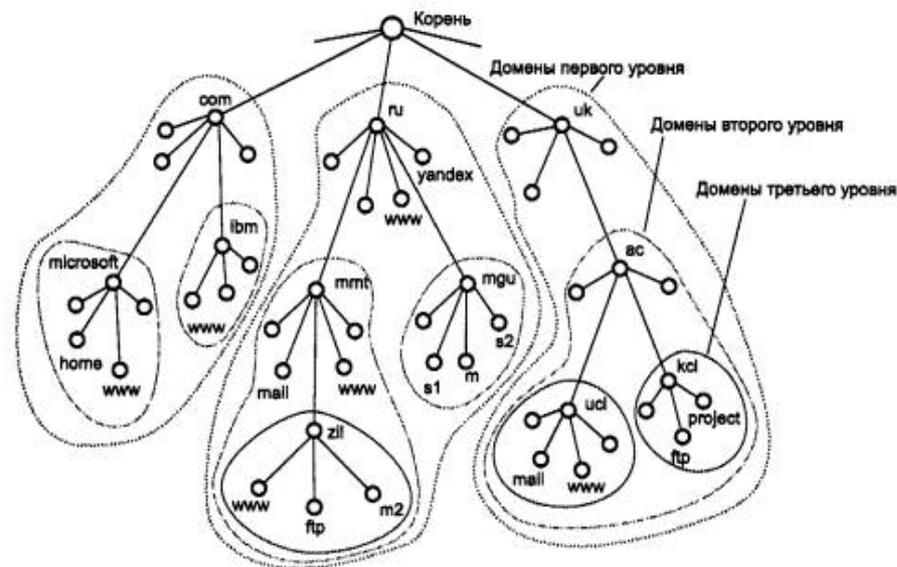


Рисунок 7.8 - Пространство доменных имен

Иерархия доменных имен аналогична иерархии имен файлов, принятой во многих популярных файловых системах. **Дерево имен** начинается с **корня**, обозначаемого здесь точкой (.). Затем следует старшая символьная часть имени, вторая по старшинству символьная часть имени и т. д. Младшая часть имени соответствует конечному узлу сети.

В отличие от имен файлов, при записи которых сначала указывается самая старшая составляющая, затем составляющая более низкого уровня и т. д., запись доменного имени начинается с самой младшей составляющей, а заканчивается самой старшей. Составные части доменного имени отделяются друг от друга точкой.

Необходимо отметить, что абоненты входят в домен в соответствии со своими составными именами, при этом они могут иметь совершенно различные IP-адреса, принадлежащие к различным сетям и подсетям. Например, в домен mgu.ru могут входить хосты с адресами 132.13.34.15, 201.22.100.33, 14.0.0.6. Доменная система имен реализована в сети Internet, но она может работать и как автономная система имен в крупной корпоративной сети, использующей стек TCP/IP.

*Соответствие между доменными именами и IP-адресами может устанавливаться как средствами локального хоста, так и средствами централизованной службы. В качестве централизованной специальной службы используется - **система доменных имен (Domain Name System, DNS)**.*

DNS - это централизованная служба, основанная на распределенной базе отображений «**доменное имя - IP-адрес**». Служба DNS использует в своей работе протокол типа «клиент-сервер». В нем определены DNS-серверы и

DNS-клиенты:

- **DNS-серверы** поддерживают распределенную базу отображений,
- **DNS-клиенты** обращаются к серверам с запросами о разрешении доменного имени в IP-адрес.

Каждый DNS-сервер кроме таблицы отображений имен содержит ссылки на DNS-серверы своих поддоменов. Эти ссылки связывают отдельные DNS-серверы в единую службу DNS. Ссылки представляют собой IP-адреса соответствующих серверов. Для обслуживания корневого домена выделено несколько дублирующих друг друга DNS-серверов, IP-адреса которых являются широко известными

Существуют две основные схемы разрешения DNS-имен.

В первом варианте работу по поиску IP-адреса координирует DNS-клиент:

- DNS-клиент обращается к корневому DNS-серверу с указанием полного доменного имени;
- DNS-сервер отвечает, указывая адрес следующего DNS-сервера, обслуживающего домен верхнего уровня, заданный в старшей части запрошенного имени;
- DNS-клиент делает запрос следующего DNS-сервера, который отсылает его к DNS-серверу нужного поддомена, и т. д., пока не будет найден DNS-сервер, в котором хранится соответствие запрошенного имени IP-адресу. Этот сервер дает окончательный ответ клиенту.

В такой схеме взаимодействия клиент сам итеративно выполняет последовательность запросов к разным серверам имен. Так как эта схема загружает клиента достаточно сложной работой, то она применяется редко.

Во втором варианте реализуется рекурсивная процедура:

- DNS-клиент запрашивает локальный DNS-сервер, то есть тот сервер, который обслуживает поддомен, к которому принадлежит имя клиента;
- если локальный DNS-сервер знает ответ, то он сразу же возвращает его клиенту; это может соответствовать случаю, когда запрошенное имя входит в тот же поддомен, что и имя клиента, а также может соответствовать случаю, когда сервер уже узнавал данное соответствие для другого клиента и сохранил его в своем кэше;
- если же локальный сервер не знает ответ, то он выполняет итеративные запросы к корневому серверу и т. д. точно так же, как это делал клиент в первом варианте; получив ответ, он передает его клиенту, который все это время просто ждал его от своего локального DNS-сервера.

В этой схеме клиент перепоручает работу своему серверу, поэтому схема называется косвенной или рекурсивной. Практически все DNS-клиенты используют рекурсивную процедуру.

Все адреса *TCP/IP* общих радиостанций начинаются с номера 44. Например адрес *TCP/IP* станции *RW3AH* 44.178.1.7 расшифровывается так: 44 - радиолюбительская *TCP/IP* сеть; 178 -Россия; 1 - Москва; 7 - персональный номер.

IP адреса выдаются *TCP/IP* координаторами для каждой сети. Кроме *IP* адреса используются также *даменные имена*. В настоящее время в соответствии с соглашением, принятым в пакетной *TCP/IP* любительской радиосети, в качестве *доменного имени* используют позывной или комбинацию позывного с аббревиатурой, показывающей специфику *TCP/IP* станции. Например: *hostname RW3AH* будет «*rwSah*»; *hostname Gateway RA3APW* будет «*gw.ra3apw*»; *hostname BBS VE3JF* будет «*bbs.ve3jf*».

Маршрутизация

Важнейшей задачей сетевого уровня является маршрутизация - передача пакетов между двумя конечными узлами в составной сети.

В сложных составных сетях почти всегда существует несколько альтернативных маршрутов для передачи пакетов между двумя конечными узлами.

Маршрут - это последовательность маршрутизаторов, которые должен пройти пакет от отправителя до пункта назначения. Задачу выбора маршрута из нескольких возможных решают **маршрутизаторы**, а также конечные узлы.

Маршрут выбирается на основании имеющейся информации о текущей конфигурации сети, а также на основании указанного критерия выбора маршрута. В качестве критерия выступает задержка прохождения маршрута пакетом или средняя пропускная способность маршрута (протоколы *OSPF*, *IGRP*, *BGP*). Часто также используется весьма простой критерий, учитывающий только количество пройденных в маршруте промежуточных маршрутизаторов (протокол *RIP*).

7.2.3 Протоколы транспортного и вышестоящих уровней

Транспортный уровень (Transport layer) обеспечивает приложениям или верхним уровням стека - прикладному и сеансовому - передачу данных с той степенью надежности, которая им требуется. Модель OSI определяет различные классы сервиса, предоставляемых транспортным уровнем:

- срочность,
- возможностью восстановления прерванной связи,

- наличием средств мультиплексирования нескольких соединений между различными прикладными протоколами через общий транспортный протокол,
- способность к обнаружению и исправлению ошибок передачи, таких как искажение, потеря и дублирование пакетов.

Как правило, все протоколы, начиная с транспортного уровня и выше, реализуются программными средствами конечных узлов сети - компонентами их сетевых операционных систем.

Протокол TCP

Transmission Control Protocol (TCP) (протокол управления передачей) — один из основных сетевых протоколов, предназначенный для управления передачей данных в сетях и подсетях TCP/IP. Выполняет функции протокола транспортного уровня модели OSI.

TCP — это транспортный механизм, предоставляющий поток данных, с предварительной установкой соединения, контролирующей достоверность получаемых данных, осуществляет повторный запрос данных в случае потери данных и устраняет дублирование при получении двух копий одного пакета. В отличие от UDP, гарантирует, что приложение получит данные точно в такой же последовательности, в какой они были отправлены, и без потерь.

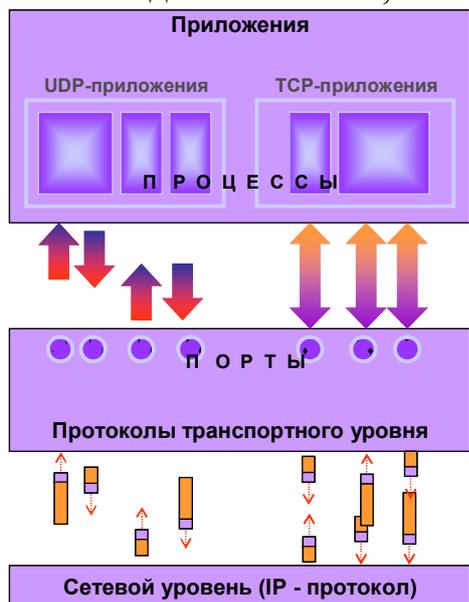


Рисунок 7.9 - Доступ приложений через порты к среде передачи

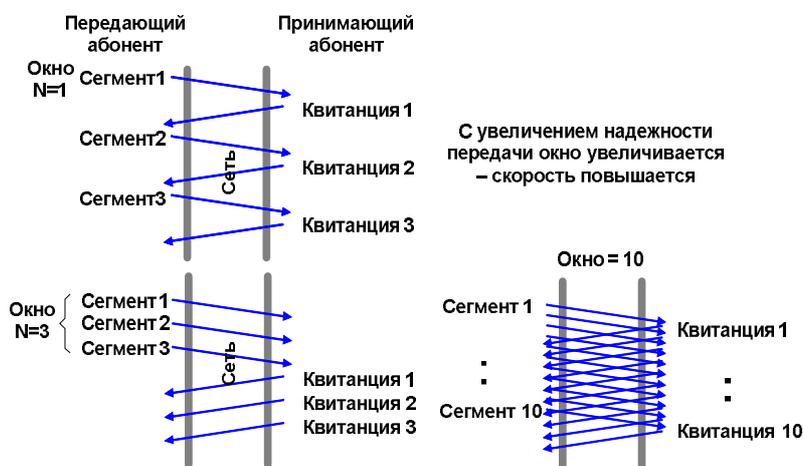


Рисунок 7.10 – Адаптивная регулировка скорости передачи к надежности соединения за счет механизма «плавающего окна»

Приложения не формируют IP-пакеты - для этого разработчикам приложений пришлось бы разбираться с неспецифичными для них сетевыми задачами и приложения стали бы зависимы от типа сети. Транспортный уровень принимает потоки данных или сообщения, «упаковывает» данные приложений в IP-пакеты и передает в сеть:

- сервис негарантированной доставки единичных сообщений обеспечивает транспортный протокол UDP
- потоковый транспортный сервис с надежной доставкой обеспечивает протокол TCP

Основные протоколы верхних уровней модели OSI

Кратко рассмотрим основные протоколы, выполняемые на *TCP/IP* станции:

TELNET - протокол эмуляции терминала. Этот протокол позволяет осуществить терминальное соединение с удаленной *TCP/IP* станцией и очень похож на соединение одной пакетной станции с другой (пользователя с *BBS*).

FTP (File Transfer Protocol) - протокол передачи файлов, позволяющий организовать пересылку *ASCII* и бинарных файлов (в том числе и программ) между *TCP/IP* станциями.

SMTP (Simple Mail Transfer Protocol) - протокол передачи почтовых сообщений, который позволяет переслать почтовое сообщение на указанный почтовый адрес.

POP (Post Office Protocol) - разновидность почтового протокола. Для работы *SMTP* протокола необходимо, чтобы удаленная станция работала 24 часа в сутки. Это не всегда удобно. Использование *POP* протокола позволяет обеспечить хранение почты, предназначенной абоненту, на почтовом сервере (*Mailbox*), а при появлении станции абонента в эфире - автоматически пересылать информацию ему.

PING (Packet InterNet Groper) - протокол, служащий для диагностики канала. Протокол инициирует посылку пакета-запроса на удаленную станцию, ждет ответа и показывает время соединения с удаленной станцией.

7.3 Архитектура пакетных радиосетей и принципы их построения

В радиосвязи пакетные сети строятся с использованием сетевого протокола X.25. Доработанный с учетом особенностей ведения радиосвязи протокол X.25, получил название AX.25 и обеспечивает следующие возможности:

- *установление логического канала для передачи данных с одним или несколькими абонентами и отслеживание его состояния;*
- *обеспечение безошибочности передачи данных по установленно-*

му каналу, методом подтверждения правильного приема пакетов информации и повтора искаженных;

- возможность согласования средствами протокола скорости передачи и скорости приема на другой стороне соединения.

Особенность пакетных радиосетей заключается в том, что один и тот же радиоканал используется для передачи данных всеми пользователями сети в режиме множественного доступа. Протокол АХ.25 предусматривает множественный доступ в радиоканал связи с контролем несущей. Все пользователи (станции) сети считаются равноправными. Прежде чем начать передачу радиомодем проверяет свободен ли радиоканал или нет. Если канал занят, то передача своих данных радиомодемом откладывается до момента его освобождения. Если радиомодем обнаруживает радиоканал свободным, то он сразу же начинает передачу своей информации. Очевидно, что в тот же самый момент может начать передачу и любой другой пользователь данной радиосети. В этом случае происходит наложение (конфликт) сигналов двух радиомодемов, в результате чего их данные с высокой вероятностью серьезно искажаются под воздействием взаимных помех. Радиомодем-передатчик узнает об этом получив отрицательное подтверждение на переданный пакет данных от радиомодема-получателя или в результате превышения времени тайм-аута. В такой ситуации он обязан будет повторить передачу этого пакета по уже описанному алгоритму.

По сути, протокол АХ.25 при доступе к каналу связи реализует радиовersion алгоритма CSMA/CD (рисунок 7.6).

Для обеспечения обмена данными по протоколу АХ.25 используется устройство, называемое *TNC (Terminal Node Controller)*. Это, как правило, отдельный прибор, соединенный с радиостанцией и компьютером, содержащий оборудование для ведения связи в соответствии с протоколом АХ.25. Кроме того, он содержит средства для преобразования цифровых сигналов в аналоговые сигналы, пригодные для передачи радиостанцией в эфир и переключает радиостанцию с передачи на прием, следуя логике протокола. Поэтому иногда *TNC* называют радиомодемом.

Дальнейшее развитие логики пакетного радио вызвало необходимость распределения работающих станций на несколько одновременно функционирующих радиодиапазонов с сохранением возможности взаимодействия находящихся на них систем. Таким образом, возникла система узлов, одновременно работающих на нескольких частотных диапазонах и позволяющих станции, работающей только на одном из них, устанавливать соединение с корреспондентом, находящимся на других.

Использование пакетных сетей связи в радиосвязи

Наибольшего совершенства и гибкости пакетное радио достигло после привнесения в его практику протоколов компьютерных сетей TCP/IP, которые используют АХ.25 в качестве транспортной среды, передавая собственные данные «поверх» него.

В пакетной радиосвязи на базе типовых радиостанций применяются два способа модуляции для коротких и ультракоротких волн. На КВ используется однополосная модуляция для формирования канала тональной частоты в радиоканале. Для передачи данных применяется частотная модуляция поднесущей в полосе частот телефонного канала 0,3...3,4 кГц. Значение частоты поднесущей может быть различной, а разнос частот всегда равен 200 Гц. В таком режиме обеспечивается скорость передачи, равная 300 бит/с. В Европе обычно используется частота 1850 Гц для передачи «0» и 1650 Гц для передачи «1».

В УКВ диапазоне чаще работают на скорости 1200 бит/с при использовании частотной модуляции с разносом поднесущих частот 1000 Гц. Принято, что «0» соответствует частота 1200 Гц, а «1» - 2200 Гц. Реже в диапазоне УКВ применяют относительную фазовую манипуляцию (ОФМ). В этом случае достигаются скорости передачи 2400, 4800, а иногда 9600 и 19200 бит/с.

Функции коммутации выполняют маршрутизаторы по сетевому адресу (позывному) корреспондента. При этом используется весь совокупный коммуникационный ресурс сети и информация может следовать к корреспонденту различными путями и каналами.

Наиболее важным достоинством TCP/IP систем является то, что для установления связи с корреспондентами пользователю не требуется знать маршрут до него, т.к. система сама выбирает оптимальный маршрут и устанавливает соединение.

Для персональной идентификации каждая TCP/IP станция имеет свой уникальный (неповторяемый) IP адрес.

В соответствии с принятым в профессиональной сети *Internet* стандартом, IP адрес состоит из четырехбайтной последовательности, где каждый байт дает определенную информацию к какой сети и подсети относится данная станция. IP адрес имеет иерархическую структуру, то есть крайний левый байт имеет максимальную значимость (определяет всю сеть), а крайний правый минимальную (определяет конкретную станцию в сети). Такое построение необходимо для маршрутизации информации как внутри одной сети, так и между разными компьютерными сетями. Все адреса TCP/IP станций начинаются с номера 44. Например адрес TCP/IP станции RWЗАН 44.178.1.7 расшифровывается так: 44 - радиолюбительская TCP/IP сеть; 178 -Россия; 1 - Мо-

сква; 7 - персональный номер.

IP адреса выдаются *TCP/IP* координаторами для каждой сети. Кроме *IP* адреса используются также понятия «*hostname*» и «*domain name*». В настоящее время в соответствии с соглашением, принятым в пакетной *TCP/IP* любительской радиосети, в качестве *hostname* используют позывной или комбинацию позывного с аббревиатурой, показывающей специфику *TCP/IP* станции. Например: *hostname* RW3AH будет «*rw3ah*»; *hostname* Gateway RA3APW будет «*gw.ra3apw*»; *hostname* BBS VE3JF будет «*bbs.ve3jf*».

Понятие «*domain*» определяет группу *TCP/IP* станций логически соединенных в единую сеть. *Domain* имя подобно *IP* адресу, части имени также разделены друг от друга точками, что также соответствует иерархической структуре имени, однако высшую иерархию имеет крайняя правая часть имени. Для радиолобительской *TCP/IP* сети *domain* именем является «*ampr.org*», где «*org*» (*org* - сокращение от *organization*) имеет высшую иерархию, а «*ampr*» имеет более низкую иерархию и указывает на принадлежность к радиолобительской *TCP/IP* сети (44.XXX.YYY.ZZZ). Полное *IP* имя {*FQDN* - *Fully Qualified Domain Name*) образуется путем соединения *hostname* и *domain* имени. Так, полное *IP* имя для RW3AH будет выглядеть так «*rw3ah.ampr.org*». Если станция имеет нескольких пользователей (например, если станция *TCP/IP Mailbox*), то в этом случае к полному *IP* имени добавляется зарегистрированное на этой станции имя и знак «@». Например, если RW3AG хочет получать (и отправлять) почту со станции «*gw.ra3apw.ampr.org*», то его почтовый адрес будет таким «*rw3ag@gw.ra3apw.ampr.org*».

ГЛАВА 8 КРИПТОГРАФИЧЕСКИЕ МЕТОДЫ ЗАЩИТЫ ДАННЫХ

8.1 Основы криптографической защиты данных

8.1.1 Понятие криптосистемы

Криптография представляет собой совокупность методов преобразования данных, направленных на то, чтобы сделать эти данные бесполезными для противника. Такие преобразования позволяют решить две главные проблемы защиты данных: **проблему конфиденциальности** (путем лишения противника возможности извлечь информацию из канала связи) и **проблему целостности** (путем лишения противника возможности изменить сообщение так, чтобы изменился его смысл, или ввести ложную информацию в канал связи).

Обобщенная схема криптографической системы, обеспечивающей шифрование передаваемой информации, показана на рисунке 8.1.

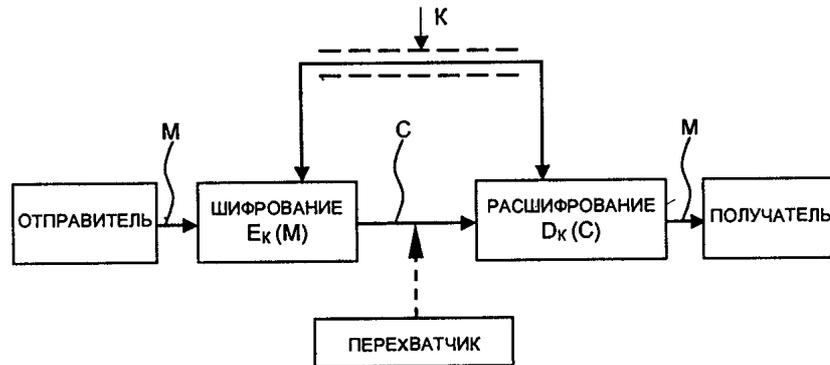


Рисунок 8.1 - Обобщенная схема криптосистемы

Отправитель генерирует **открытый текст** исходного сообщения M , которое должно быть передано законному **получателю** по незащищенному каналу. За каналом следит **перехватчик** с целью перехватить и раскрыть передаваемое сообщение. Для того чтобы перехватчик не смог узнать содержание сообщения M , отправитель шифрует его с помощью обратимого преобразования E_k и получает **шифртекст** (или **криптограмму**) $C = E_k(M)$, который отправляет получателю.

Законный получатель, приняв шифртекст C , расшифровывает его с помощью обратного преобразования $D = E_k^{-1}$ и получает исходное сообщение в виде открытого текста M :

$$D_k(C) = E_k^{-1}(E_k(M)) = M$$

Преобразование E_k выбирается из семейства криптографических преобразований, называемых **криптоалгоритмами**. Параметр, с помощью которого выбирается отдельное используемое преобразование, называется **криптогра-**

фическим ключом K .

Криптосистема имеет разные варианты реализации: набор инструкций, аппаратные средства, комплекс программ компьютера, которые позволяют зашифровать открытый текст и расшифровать шифртекст различными способами.

Формально, **криптографическая система** - это однопараметрическое семейство $(E_k)_{k \in \bar{K}}$ обратимых преобразований вида:

$$E_k : \bar{M} \rightarrow \bar{C}$$

из пространства \bar{M} сообщений открытого текста в пространство \bar{C} шифрованных текстов. Параметр K (ключ) выбирается из конечного множества \bar{K} , называемого **пространством ключей**.

Шифр (в соответствии со стандартом ГОСТ 28147-89) - совокупность обратимых преобразований множества открытых данных на множество зашифрованных данных, задаваемых ключом и алгоритмом криптографического преобразования.

Ключ - это конкретное секретное состояние некоторых параметров алгоритма криптографического преобразования данных, обеспечивающее выбор только одного варианта из всех возможных для данного алгоритма.

Криптостойкость - основная характеристика шифра является, которая определяет его стойкость к раскрытию методами криптоанализа. Обычно эта характеристика определяется интервалом времени, необходимым для раскрытия шифра.

К шифрам, используемым для криптографической защиты информации, предъявляется ряд требований:

- достаточная криптостойкость (надежность закрытия данных);
- простота процедур шифрования и расшифрования;
- незначительная избыточность информации за счет шифрования;
- нечувствительность к небольшим ошибкам шифрования и др.

В той или иной мере этим требованиям отвечает:

- **Шифрование перестановкой** заключается в том, что символы шифруемого текста переставляются по определенному правилу в пределах некоторого блока этого текста. При достаточной длине блока, в пределах которого осуществляется перестановка, и сложном неповторяющемся порядке перестановки можно достигнуть приемлемой для простых практических приложений стойкости шифра.
- **Шифрование заменой (подстановкой)** заключается в том, что символы шифруемого текста заменяются символами того же или другого алфавита в соответствии с заранее обусловленной схемой замены.

- **Шифрование гаммированием** заключается в том, что символы шифруемого текста складываются с символами некоторой случайной последовательности, именуемой **гаммой-шифром**. Стойкость шифрования определяется в основном длиной (периодом) неповторяющейся части гаммы шифра. Поскольку с помощью ЭВМ можно генерировать практически бесконечную гамму шифра, то данный способ является одним из основных для шифрования информации в автоматизированных системах.
- **Шифрование аналитическим преобразованием** заключается в том, что шифруемый текст преобразуется по некоторому аналитическому правилу (формуле). Например, можно использовать правило умножения вектора на матрицу, причем умножаемая матрица является ключом шифрования (поэтому ее размер и содержание должны храниться в секрете), а символами умножаемого вектора последовательно служат символы шифруемого текста. Другим примером может служить использование так называемых однонаправленных функций для построения криптосистем с открытым ключом.

8.1.2 Симметричные и асимметричные криптосистемы

В общем случае шифрование может быть симметричным или асимметричным относительно преобразования расшифрования, что определяет два класса криптосистем:

- **симметричные (одноключевые) криптосистемы;**
- **асимметричные (двухключевые) криптосистемы (с открытым ключом).**

Схема симметричной криптосистемы с одним секретным ключом была показана на рисунке 8.1. В ней используются одинаковые секретные ключи в блоке шифрования и блоке расшифрования.

Обобщенная схема асимметричной криптосистемы с двумя разными ключами K_1 и K_2 показана на рисунке 8.2. В этой криптосистеме один из ключей является открытым, а другой - секретным.

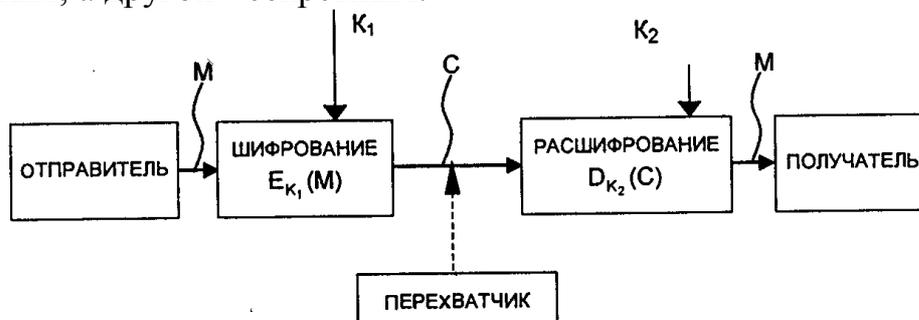


Рисунок 8.2 - Обобщенная схема асимметричной криптосистемы с открытым ключом

В симметричной криптосистеме секретный ключ надо передавать отправителю и получателю по защищенному каналу распространения ключей, например такому, как курьерская служба. На рис. 1.1 этот канал показан "экранированной" линией. В асимметричной криптосистеме передают по незащищенному каналу только открытый ключ, а секретный ключ сохраняют на месте его генерации.

8.1.3 Понятие криптоанализа

Криптоанализ - это наука о раскрытии исходного текста зашифрованного сообщения без доступа к ключу. Успешный анализ может раскрыть исходный текст или ключ. Он позволяет также обнаружить слабые места в криптосистеме, что, в конечном счете, ведет к тем же результатам.

Фундаментальное правило криптоанализа (впервые сформулированное голландцем А. Керкхоффом еще в XIX веке) **заключается в том, что стойкость шифра (криптосистемы) должна определяться только секретностью ключа.** Иными словами, правило Керкхоффа состоит в том, что весь алгоритм шифрования, кроме значения секретного ключа, известен криптоаналитику противника. Это обусловлено тем, что криптосистема, реализующая семейство криптографических преобразований, обычно рассматривается как открытая система. Такой подход отражает очень важный принцип технологии защиты информации: защищенность системы не должна зависеть от секретности чего-либо такого, что невозможно быстро изменить в случае утечки секретной информации.

Другое почти общепринятое допущение в криптоанализе состоит в том, что криптоаналитик имеет в своем распоряжении шифртексты сообщений.

Криптоаналитическая атака - любая попытка со стороны перехватчика расшифровать шифртекст C для получения открытого текста M или зашифровать свой собственный текст M' для получения правдоподобного шифртекста C' , не имея подлинного ключа.

На рисунке 8.3 показан поток информации в криптосистеме в случае активных действий перехватчика. Активный перехватчик не только считывает все шифртексты, передаваемые по каналу, но может также пытаться изменять их по своему усмотрению.

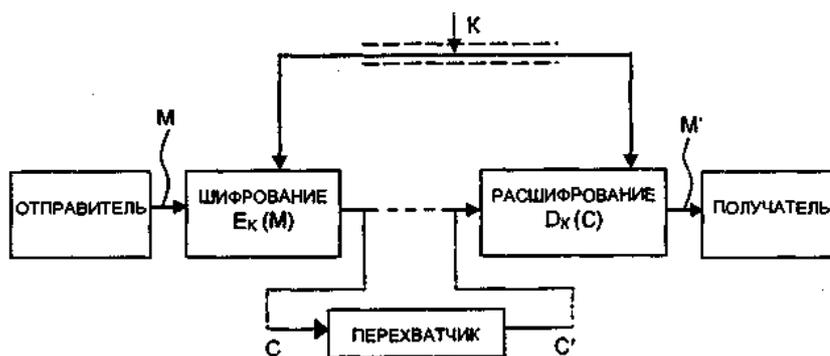


Рисунок 8.3 - Поток информации в криптосистеме при активном перехвате сообщений

Если предпринятые криптоаналитические атаки не достигают поставленной цели и криптоаналитик не может, не имея подлинного ключа, вывести M из C или C' из M' , то полагают, что такая криптосистема является **криптостойкой**.

8.2 Симметричные криптосистемы

Большинство средств защиты информации базируется на использовании криптографических шифров и процедур шифрования – расшифрования осуществляющихся в рамках некоторой криптосистемы. **Характерной особенностью симметричной криптосистемы является применение одного и того же секретного ключа как при шифровании, так и при расшифровании сообщений.**

Как открытый текст, так и шифртекст образуются из букв, входящих в конечное множество символов, называемых алфавитом. Примерами алфавитов являются конечное множество всех заглавных букв, конечное множество всех заглавных и строчных букв и цифр и т. п. В общем виде некоторый алфавит Σ можно представить так:

$$\Sigma = \{a_0, a_1, \dots, a_{m-1}\}$$

Объединяя по определенному правилу буквы из алфавита Σ можно создать новые алфавиты:

- алфавит Σ^2 , содержащий m^2 биграмм $a_0a_0, a_0a_1, \dots, a_{m-1}a_{m-1}$
- алфавит Σ^3 , содержащий m^3 триграмм $a_0a_0a_0, a_0a_0a_1, \dots, a_{m-1}a_{m-1}a_{m-1}$.

В общем случае, объединяя по n букв, получаем алфавит Σ^n содержащий m^n n -грамм. Например, английский алфавит

$$\Sigma = \{ABCDEFGHIH \dots WXYZ\}$$

объемом $m = 26$ букв позволяет сгенерировать посредством операции конкатенации алфавит из $26^2 = 676$ биграмм:

AA, AB, ... ,XZ, ZZ

алфавит из $26^3=17576$ триграмм:

AAA, AAB, ... ,XZZ, ZZZ и т.д.

При выполнении криптографических преобразований полезно заменить буквы алфавита целыми числами 0, 1, 2, 3, Это позволяет упростить выполнение необходимых алгебраических манипуляций.

Например, можно установить взаимно однозначное соответствие между русским алфавитом

$$\Sigma = \{АБВГД \dots ЮЯ\}$$

и множеством целых

$$\bar{Z}_{32} = \{0, 1, 2, 3, \dots, 31\};$$

между английским алфавитом

$$\Sigma_{англ} = \{АВСDEF \dots YZ\}$$

и множеством целых

$$\bar{Z}_{26} = \{0, 1, 2, 3, \dots, 25\}$$

(см. таблицы 8.1 и 8.2).

В дальнейшем будет обычно использоваться алфавит

$$\bar{Z}_m = \{0, 1, 2, 3, \dots, m-1\},$$

содержащий m "букв" (в виде чисел).

Замена букв традиционного алфавита числами позволяет более четко сформулировать основные концепции и приемы криптографических преобразований. В то же время в большинстве иллюстраций будет использоваться алфавит естественного языка.

Таблица 8.1 Соответствие между русским алфавитом и множеством целых $\bar{Z}_{32} = \{0, 1, 2, 3, \dots, 31\}$;

Буква	Число	Буква	Число	Буква	Число	Буква	Число
А	0	И	8	Р	16	Ш	24
Б	1	Й	9	С	17	Щ	25
В	2	К	10	Т	18	Ъ	26
Г	3	Л	11	У	19	Ы	27
Д	4	М	12	Ф	20	Ь	28
Е	5	Н	13	Х	21	Э	29
Ж	6	О	14	Ц	22	Ю	30
З	7	П	15	Ч	23	Я	31

Таблица 8.2 - Соответствие между английским алфавитом и множеством целых $\bar{Z}_{26} = \{0, 1, 2, 3, \dots, 25\}$

Буква	Число	Буква	Число	Буква	Число
A	0	J	9	S	18
B	1	K	10	T	19
C	2	L	11	U	20
D	3	M	12	V	21
E	4	N	13	W	22
F	5	O	14	X	23
G	6	P	15	Y	24
H	7	Q	16	Z	25
I	8	R	17		

Текст с n буквами из алфавита Z_m можно рассматривать как n -грамму

$$\bar{x} = (x_0, x_1, x_2, \dots, x_{n-1}),$$

где $\bar{x}_i \in \bar{Z}_m$, $0 \leq i < n$, для некоторого целого $n = 1, 2, 3, \dots$. Через $\bar{Z}_{m,n}$ будем обозначать множество n -грамм, образованных из букв множества \bar{Z}_m .

Криптографическое преобразование E представляет собой совокупность преобразований

$$E = \{E^{(n)} : 1 \leq n < \infty\}$$

$$E^{(n)} : \bar{Z}_{m,n} \rightarrow \bar{Z}_{m,n}$$

Преобразование $E^{(n)}$ определяет, как каждая n -грамма открытого текста $\bar{x} \in \bar{Z}_{m,n}$ заменяется n -граммой шифртекста \bar{y} , т. е.

$$\bar{y} = E^{(n)}(\bar{x})$$

причем

$$\bar{x}, \bar{y} \in \bar{Z}_{m,n}$$

при этом обязательным является требование взаимной однозначности преобразования $E^{(n)}$ на множестве $\bar{Z}_{m,n}$.

Криптографическая система может трактоваться как семейство криптографических преобразований

$$\bar{E} = \{E_k : K \in \bar{K}\},$$

помеченных параметром K , называемым ключом.

Множество значений ключа образует ключевое пространство \bar{K} . Далее рассматриваются традиционные (классические) методы шифрования, отличающиеся симметричной функцией шифрования. К ним относятся шифры перестановки, шифры простой и сложной замены, а также некоторые их модификации и комбинации. Следует отметить, что комбинации шифров перестановок и шифров замены образуют все многообразие применяемых на практике симметричных шифров.

8.2.1 Шифры перестановки

При шифровании перестановкой символы шифруемого текста переставляются по определенному правилу в пределах блока этого текста. Шифры перестановки являются самыми простыми и, вероятно, самыми древними шифрами.

С начала эпохи Возрождения (конец XIV столетия) начала возрождаться и криптография. Наряду с традиционными применениями криптографии в политике, дипломатии и военном деле появляются и другие задачи - защита интеллектуальной собственности от преследований инквизиции или заимствований злоумышленников. В разработанных шифрах перестановки того времени применяются шифрующие таблицы, которые в сущности задают правила перестановки букв в сообщении.

В качестве ключа в шифрующих таблицах используются:

- размер таблицы;
- слово или фраза, задающие перестановку;
- особенности структуры таблицы.

Одним из самых примитивных табличных шифров перестановки является простая перестановка, для которой ключом служит размер таблицы. Этот метод шифрования сходен с шифром скитала. Например, сообщение

ТЕРМИНАТОР ПРИБЫВАЕТ СЕДЬМОГО В ПОЛНОЧЬ

записывается в таблицу поочередно по столбцам. Результат заполнения таблицы из 5 строк и 7 столбцов показан на рисунке 8.4.

Т	Н	П	В	Е	Г	Л
Е	А	Р	А	Д	О	Н
Р	Т	И	Е	Ь	В	О
М	О	Б	Т	М	П	Ч
И	Р	Ы	С	О	О	Ь

Рисунок 8.4 - Заполнение таблицы из 5 строк и 7 столбцов

После заполнения таблицы текстом сообщения по столбцам для формирования шифртекста считывают содержимое таблицы по строкам. Если шифртекст записывать группами по пять букв, получается такое шифрованное сообщение:

ТНПВЕ ГЛЕАР АДОНР ТИЕЬВ ОМОБТ МПЧИР ЫСООЬ

Естественно, отправитель и получатель сообщения должны заранее условиться об общем ключе в виде размера таблицы. Следует заметить, что объединение букв шифртекста в 5-буквенные группы не входит в ключ шифра и

осуществляется для удобства записи несмыслового текста. При расшифровке действия выполняются в обратном порядке.

Несколько большей стойкостью к раскрытию обладает метод шифрования, называемый одиночной перестановкой по ключу. Этот метод отличается от предыдущего тем, что столбцы таблицы переставляются по ключевому слову, фразе или набору чисел длиной в строку таблицы.

Применим в качестве ключа, например, слово
ПЕЛИКАН,

а текст сообщения возьмем из предыдущего примера. На рисунке 8.5 показаны две таблицы, заполненные текстом сообщения и ключевым словом, при этом левая таблица соответствует заполнению до перестановки, а правая таблица — заполнению после перестановки.

Ключ
→

П	Е	Л	И	К	А	Н
7	2	5	3	4	1	6
Т	Н	П	В	Е	Г	Л
Е	А	Р	А	Д	О	Н
Р	Т	И	Е	Ь	В	О
М	О	Б	Т	М	П	Ч
И	Р	Ы	С	О	О	Ь

А	Е	И	К	Л	Н	П
1	2	3	4	5	6	7
Г	Н	В	Е	П	Л	Т
О	А	А	Д	Р	Н	Е
В	Т	Е	Ь	И	О	Р
П	О	Т	М	Б	Ч	М
О	Р	С	О	Ы	Ь	И

До перестановки После перестановки

Рисунок 8.5

В верхней строке левой таблицы записан ключ, а номера под буквами ключа определены в соответствии с естественным порядком соответствующих букв ключа в алфавите. Если бы в ключе встретились одинаковые буквы, они бы были пронумерованы слева направо. В правой таблице столбцы переставлены в соответствии с упорядоченными номерами букв ключа.

При считывании содержимого правой таблицы по строкам и записи шифртекста группами по пять букв получим зашифрованное сообщение:

ГНВЕП ЛТООА ДРНЕВ ТЕЬИО РПОТМ БЧМОР СОЫЬИ

Для обеспечения дополнительной скрытности можно повторно зашифровать сообщение, которое уже прошло шифрование. Такой метод шифрования называется *двойной перестановкой*. В случае двойной перестановки столбцов и строк таблицы перестановки определяются отдельно для столбцов и отдельно для строк. Сначала в таблицу записывается текст сообщения, а потом поочередно переставляются столбцы, а затем строки. При расшифровании порядок перестановок должен быть обратным.

Пример выполнения шифрования методом двойной перестановки показан на рисунке 8.6.

	4	1	3	2
3	П	Р	И	Л
1	Е	Т	А	Ю
4	В	О	С	Ь
2	М	О	Г	О

	1	2	3	4
3	Р	Л	И	П
1	Т	Ю	А	Е
4	О	Ь	С	В
2	О	О	Г	М

	1	2	3	4
1	Т	Ю	А	Е
2	О	О	Г	М
3	Р	Л	И	П
4	О	Ь	С	В

Исходная таблица

Перестановка столбцов

Перестановка строк

Рисунок 8.6 - Пример выполнения шифрования методом двойной перестановки

Если считать шифртекст из правой таблицы построчно блоками по четыре буквы, то получится следующее:

ТЮАЕ ООГМ РЛИП ОЬСВ

Ключом к шифру двойной перестановки служит последовательность номеров столбцов и номеров строк исходной таблицы (в нашем примере последовательности 4132 и 3142 соответственно).

Число вариантов двойной перестановки быстро возрастает при увеличении размера таблицы:

- для таблицы 3x3 - 36 вариантов;
- для таблицы 4x4 - 576 вариантов;
- для таблицы 5x5 - 14400 вариантов.

Однако двойная перестановка не отличается высокой стойкостью и сравнительно просто "взламывается" при любом размере таблицы шифрования.

8.2.2 Система шифрования Цезаря

Шифр Цезаря является частным случаем шифра простой замены (одноалфавитной подстановки). Свое название этот шифр получил по имени римского императора Гая Юлия Цезаря, который использовал этот шифр при переписке с Цицероном (около 50 г. до н.э.).

При шифровании исходного текста каждая буква заменялась на другую букву того же алфавита по следующему правилу. Заменяющая буква определялась путем смещения по алфавиту от исходной буквы на K букв. При достижении конца алфавита выполнялся циклический переход к его началу. Цезарь использовал шифр замены при смещении $K=3$. Такой шифр замены можно задать таблицей подстановок, содержащей соответствующие пары букв открытого текста и шифртекста. Совокупность возможных подстановок для $K=3$ показана на рисунке 8.7.

Одноалфавитные подстановки ($K = 3, m = 26$)

A → D	J → M	S → V
B → E	K → N	T → W
C → F	L → O	U → X
D → G	M → P	V → Y
E → H	N → Q	W → Z
F → I	O → R	X → A
G → J	P → S	Y → B
H → K	Q → T	Z → C
I → L	R → U	

Рисунок 8.7 - Совокупность возможных подстановок для $K = 3$

Например, послание Цезаря:

VENI VIDI VICI

(в переводе на русский означает "Пришел, Увидел, Победил"), направленное его другу Аминтию после победы над понтийским царем Фарнаком, сыном Митридата, выглядело бы в зашифрованном виде так:

YNQL YLGL YLFL

Достоинством системы шифрования Цезаря является простота шифрования и расшифрования.

К недостаткам системы Цезаря следует отнести следующие:

- подстановки, выполняемые в соответствии с системой Цезаря, не маскируют частот появления различных букв исходного открытого текста;
- сохраняется алфавитный порядок в последовательности заменяющих букв; при изменении значения K изменяются только начальные позиции такой последовательности;
- число возможных ключей K мало;
- шифр Цезаря легко вскрывается на основе анализа частот появления букв в шифртексте.

Криптоаналитическая атака против системы одноалфавитной замены начинается с подсчета частот появления символов: определяется число появлений каждой буквы в шифртексте. Затем полученное распределение частот букв в шифртексте сравнивается с распределением частот букв в алфавите исходных сообщений, например в английском. Буква с наивысшей частотой появления в шифртексте заменяется на букву с наивысшей частотой появления в английском языке и т.д. Вероятность успешного вскрытия системы шифрования повышается с увеличением длины шифртекста.

8.2.3 Аффинная система подстановок Цезаря

В данном преобразовании буква, соответствующая числу t , заменяется на букву, соответствующую числовому значению $(at + b)$ по модулю m .

Следует заметить, что преобразование $E_{a,b}(t)$ является взаимно однозначным отображением на множестве \bar{Z}_m только в том случае, если наибольший общий делитель чисел a и m , обозначаемый как $НОД(a, m)$, равен единице, т.е. a и m должны быть взаимно простыми числами.

Например, пусть $m = 26$, $a = 3$, $b = 5$. Тогда, очевидно, $НОД(3, 26) = 1$, и мы получаем следующее соответствие между числовыми кодами букв:

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
3t+5	5	8	11	14	17	20	23	0	3	6	9	12	15	18	21	24	1	4	7	10	13	16	19	22	25	2

Преобразуя числа в буквы английского языка, получаем следующее соответствие для букв открытого текста и шифртекста:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
F	I	L	O	R	U	X	A	D	G	J	M	P	S	V	Y	B	E	H	K	N	Q	T	W	Z	C

Исходное сообщение **HOPE** преобразуется в шифртекст **AVYR**

Достоинством аффинной системы является удобное управление ключами - ключи шифрования и расшифрования представляются в компактной форме в виде пары чисел (a, b) . Недостатки аффинной системы аналогичны недостаткам системы шифрования Цезаря.

8.2.4 Система Цезаря с ключевым словом

Система шифрования Цезаря с ключевым словом является одноалфавитной системой подстановки. Особенностью этой системы является использование ключевого слова для смещения и изменения порядка символов в алфавите подстановки.

Выберем некоторое число k , $0 < k < 25$, и слово или короткую фразу в качестве *ключевого слова*. Желательно, чтобы все буквы ключевого слова были различными. Пусть выбраны слово **DIPLOMAT** в качестве ключевого слова и число $k = 5$.

Ключевое слово записывается под буквами алфавита, начиная с буквы, числовой код которой совпадает с выбранным числом k :

0	1	2	3	4	5					10					15					20				25	
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
			D	I	P	L	O	M	A	T															

Оставшиеся буквы алфавита подстановки записываются после ключевого слова в алфавитном порядке:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
V	W	X	Y	Z	D	I	P	L	O	M	A	T	B	C	E	F	G	H	J	K	N	Q	R	S	U

Теперь мы имеем подстановку для каждой буквы произвольного сообщения.

тографии Леон Батист Альберти, который также был известным архитектором и теоретиком искусства. Его книга "Трактат о шифре", написанная в 1566 г., представляла собой первый в Европе научный труд по криптологии. Кроме шифра многоалфавитной замены, Альберти также подробно описал устройства из вращающихся колес для его реализации. Криптологи всего мира почитают Л. Альберти основоположником криптологии.

8.2.6 Одноразовая система шифрования

Почти все применяемые на практике шифры характеризуются как условно надежные, поскольку они могут быть в принципе раскрыты при наличии неограниченных вычислительных возможностей. Абсолютно надежные шифры нельзя разрушить даже при использовании неограниченных вычислительных возможностей. Существует единственный такой шифр, применяемый на практике, - одноразовая система шифрования. Характерной особенностью одноразовой системы шифрования является одноразовое использование ключевой последовательности.

Одноразовая система шифрует исходный открытый текст

$$\bar{X} = (x_0, x_1, \dots, x_{n-1})$$

в шифртекст

$$\bar{Y} = (y_0, y_1, \dots, y_{n-1})$$

посредством подстановки Цезаря

$$Y_i = (X_i + K_i) \bmod m, \quad 0 \leq i < n,$$

где K_i - i -й элемент случайной ключевой последовательности.

Ключевое пространство \bar{K} одноразовой системы представляет собой набор дискретных случайных величин из \bar{Z}_m и содержит m^n значений. Процедура расшифрования описывается соотношением

$$X_i = (Y_i - K_i) \bmod m,$$

где K_i - i -й элемент той же самой случайной ключевой последовательности.

Одноразовая система изобретена в 1917 г. американцами Дж. Моборном и Г. Вернамом. Для реализации этой системы подстановки иногда используют одноразовый блокнот. Этот блокнот составлен из отрывных страниц, на каждой из которых напечатана таблица со случайными числами (ключами) K_j . Блокнот выполняется в двух экземплярах: один используется отправителем, а другой - получателем. Для каждого символа X_j сообщения используется свой ключ K_j из таблицы только один раз. После того как таблица использована, она должна быть удалена из блокнота и уничтожена. Шифрование нового сообщения начинается с новой страницы.

Этот шифр абсолютно надежен, если набор ключей K_i действительно

случаен и непредсказуем. Если криптоаналитик попытается использовать для заданного шифртекста все возможные наборы ключей и восстановить все возможные варианты исходного текста, то они все окажутся равновероятными. Не существует способа выбрать исходный текст, который был действительно послан. Теоретически доказано, что одноразовые системы являются нераскрываемыми системами, поскольку их шифртекст не содержит достаточной информации для восстановления открытого текста.

Казалось бы, что благодаря данному достоинству одноразовые системы следует применять во всех случаях, требующих абсолютной информационной безопасности. Однако возможности Применения одноразовой системы ограничены чисто практическими аспектами. Существенным моментом является требование одноразового использования случайной ключевой последовательности. Ключевая последовательность с длиной, не меньшей длины сообщения, должна передаваться получателю сообщения заранее или отдельно по некоторому секретному каналу. Это требование не будет слишком обременительным для передачи действительно важных одноразовых сообщений, например, по горячей линии Вашингтон-Москва. Однако такое требование практически неосуществимо для современных систем обработки информации, где требуется шифровать многие миллионы символов.

8.2.7 Шифрование методом гаммирования

Под *гаммированием* понимают процесс наложения по определенному, закону гаммы шифра на открытые данные.

Гамма шифр - это псевдослучайная последовательность, выработанная по заданному алгоритму для зашифрования открытых данных и расшифрования зашифрованных данных.

Процесс зашифрования заключается в генерации гаммы шифра и наложении полученной гаммы на исходный открытый текст обратимым образом, например с использованием операции сложения по модулю 2.

Следует отметить, что перед зашифрованием открытые данные разбивают на блоки $T_0^{(i)}$ одинаковой длины, обычно по 64 бита. Гамма шифр вырабатывается в виде последовательности блоков $\Gamma_w^{(i)}$ аналогичной длины.

Уравнение зашифрования можно записать в виде

$$T_w^{(i)} = \Gamma_w^{(i)} \oplus T_0^{(i)}, i = 1 \dots M,$$

где $T_w^{(i)}$ - i -й блок шифртекста;
 $\Gamma_w^{(i)}$ - i -й блок гаммы шифра;
 $T_0^{(i)}$ - i -й блок открытого текста;
 M - количество блоков открытого текста.

Процесс расшифрования сводится к повторной генерации Гаммы шиф-

ра и наложению этой гаммы на зашифрованные данные. Уравнение расшифрования имеет вид

$$T_0^{(i)} = \Gamma_u^{(i)} \oplus T_u^{(i)}$$

Получаемый этим методом шифр-текст достаточно труден для раскрытия, поскольку теперь ключ является переменным. По сути дела гамма шифра должна изменяться случайным образом для каждого шифруемого блока. Если период гаммы превышает длину всего шифруемого текста и злоумышленнику неизвестна никакая часть исходного текста, то такой шифр можно раскрыть только прямым перебором всех вариантов ключа. В этом случае криптостойкость шифра определяется длиной ключа.

8.2.8 Стандарт шифрования данных DES

Стандарт шифрования данных DES (Data Encryption Standard) опубликован в 1977 г. Национальным бюро стандартов США. Стандарт DES предназначен для защиты от несанкционированного доступа к важной, но не секретной информации в государственных и коммерческих организациях США. Алгоритм, положенный в основу стандарта, распространялся достаточно быстро, и уже в 1980 г. был одобрен Национальным институтом стандартов и технологий США (НИСТ). С этого момента DES превращается в стандарт не только по названию (Data Encryption Standard), но и фактически. Появляются программное обеспечение и специализированные микро-ЭВМ, предназначенные для шифрования и расшифрования информации в сетях передачи данных.

К настоящему времени DES является наиболее распространенным алгоритмом, используемым в системах защиты коммерческой информации. Более того, реализация алгоритма DES в таких системах становится признаком хорошего тона.

Основные достоинства алгоритма DES:

- *используется только один ключ длиной 56 бит;*
- *зашифровав сообщение с помощью одного пакета программ, для расшифровки можно использовать любой другой пакет программ, соответствующий стандарту DES;*
- *относительная простота алгоритма обеспечивает высокую скорость обработки;*
- *достаточно высокая стойкость алгоритма.*

Алгоритм DES использует комбинацию подстановок и перестановок. DES осуществляет шифрование 64-битовых блоков данных с помощью 64-битового ключа, в котором значащими являются 56 бит (остальные 8 бит - проверочные биты для контроля на четность). Дешифрование в DES является операцией, обратной Шифрованию, и выполняется путем повторения опера-

ций шифрования в обратной последовательности. Обобщенная схема процесса шифрования в алгоритме DES показана на рисунке 8.9, в.

Матрица начальной перестановки IP

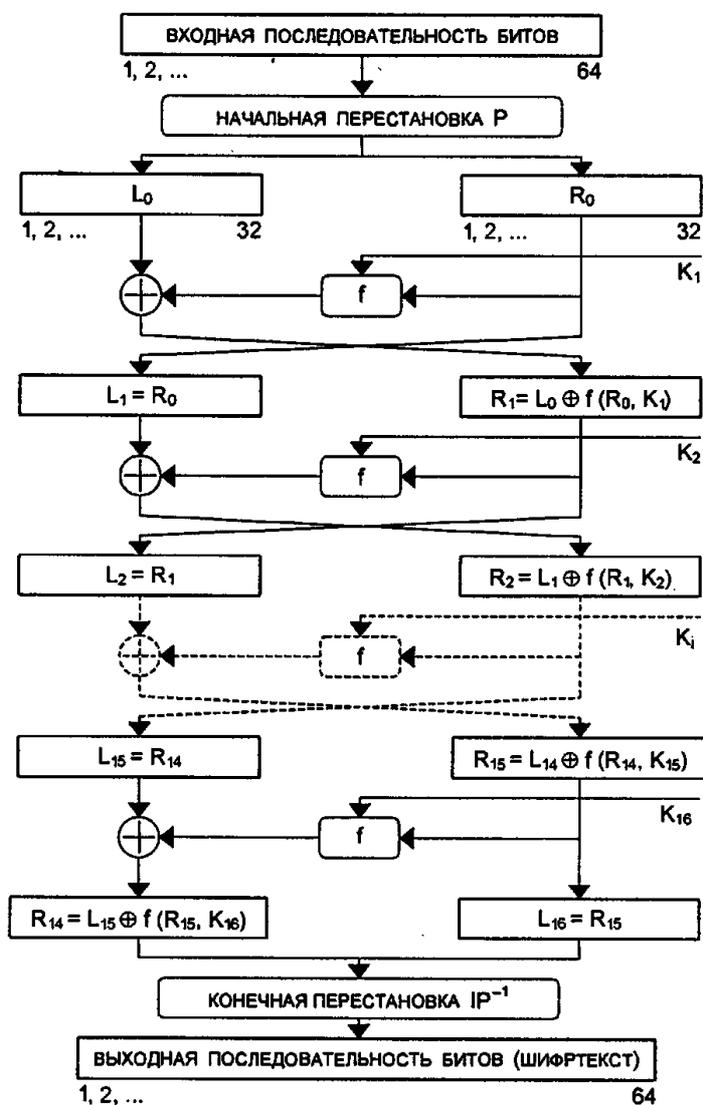
68	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

а.)

Матрица обратной перестановки IP⁻¹

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

б.)



в.)

Рисунок 8.9 – Таблицы перестановки для алгоритма DES

Процесс шифрования заключается в начальной перестановке битов 64-битового блока, шестнадцати циклах шифрования и, наконец, в конечной перестановке битов (рисунок 8.9, в). Следует сразу отметить, что все приводимые таблицы (на рисунке 8.9, а,б) являются стандартными и должны включаться в реализацию алгоритма DES в неизменном виде. Все перестановки и коды в таблицах подобраны разработчиками таким образом, чтобы максимально затруднить процесс расшифровки путем подбора ключа.

При описании алгоритма DES (рисунок 8.9, в) применены следующие обозначения:

- L и R - последовательности битов (левая (left) и правая (right));
- LR - конкатенация последовательностей L и R, т.е. такая последовательность битов, длина которой равна сумме длин L и R; в последовательности LR биты последовательности R следуют за битами последовательности L;
- \oplus - операция побитового сложения по модулю 2.

Пусть из файла исходного текста считан очередной 64-битовый (8-байтовый) блок **T**. Этот блок **T** преобразуется с помощью *матрицы начальной перестановки* **IP** (рисунок 8.9, а).

Биты входного блока **T** (64 бита) переставляются в соответствии с матрицей **IP**: бит 58 входного блока **T** становится битом 1, бит 50-битом 2 и т.д. Эту перестановку можно описать выражением $T_0 = IP(T)$. Полученная последовательность битов **T**₀ разделяется на две последовательности: **L**₀ - левые или старшие биты, **R**₀ - правые или младшие биты, каждая из которых содержит 32 бита.

Затем выполняется итеративный процесс шифрования, состоящий из 16 шагов (циклов). Пусть **T**_{*i*}- результат *i*-й итерации:

$$T_i = L_i R_i$$

Где $L_i = t_1 t_2 \dots t_{32}$ (первые 32 бита); $R_i = t_{33} t_{34} \dots t_{64}$ (последние 32 бита).

Тогда результат *i*-й итерации описывается следующими формулами:

$$L_i = R_{i-1}, \quad i = 1, 2, \dots, 16; \quad R_i = L_{i-1} \oplus f(R_{i-1}, K_i), \quad i = 1, 2, \dots, 16.$$

Функция **f** называется функцией шифрования. Ее аргументами являются последовательность **R**_{*i-1*}, получаемая на предыдущем шаге итерации, и 48-битовый ключ **K**_{*i*}, который является результатом преобразования 64-битового ключа шифра **K**. (Подробнее функция шифрования **f** и алгоритм получения ключа **K**_{*i*} описаны ниже.) На последнем шаге итерации получают последовательности **R**₁₆ и **L**₁₆ (без перестановки местами), которые конкатенируются в 64-битовую последовательность **R**₁₆**L**₁₆

По окончании шифрования осуществляется восстановление позиций битов с помощью матрицы обратной перестановки **IP**⁻¹ (рисунок 8.9, б).

Процесс расшифрования данных является инверсным по отношению к процессу шифрования. Все действия должны быть выполнены в обратном порядке. Это означает, что расшифровываемые данные сначала переставляются в соответствии с матрицей **IP**⁻¹, а затем над последовательностью битов **R**₁₆**L**₁₆ выполняются те же действия, что и в процессе шифрования, но в обратном порядке:

$$R_{i-1} = L_i, \quad i = 1, 2, \dots, 16; \quad L_{i-1} = R_i \oplus f(L_i, K_i), \quad i = 1, 2, \dots, 16.$$

В настоящее время блочный алгоритм DES считается относительно безопасным алгоритмом шифрования. Он подвергался тщательному криптоанализу в течение 20 лет, и самым практичным способом его взламывания является метод перебора всех возможных вариантов ключа. Ключ DES имеет длину 56 бит, поэтому существует 2^{55} возможных вариантов такого ключа. Если предположить, что суперкомпьютер может испытать миллион вариантов ключа за секунду, то потребуется 2285 лет для нахождения правильного ключа. Если бы ключ имел длину 128 бит, то потребовалось бы 10^{25} лет (для сравнения: возраст Вселенной около 10^{10} лет).

8.3 Концепция криптосистемы с открытым ключом

Эффективными системами криптографической защиты данных являются асимметричные криптосистемы, называемые также криптосистемами с открытым ключом. В таких системах для зашифрования данных используется один ключ, а для расшифрования - другой ключ (отсюда и название - асимметричные). Первый ключ является открытым и может быть опубликован для использования всеми пользователями системы, которые зашифровывают данные. Расшифрование данных с помощью открытого ключа невозможно.

Для расшифрования данных получатель зашифрованной информации использует второй ключ, который является *секретным*. Разумеется, ключ расшифрования не может быть определен из ключа зашифрования.

Обобщенная схема асимметричной криптосистемы с открытым ключом показана на рисунке 8.10.

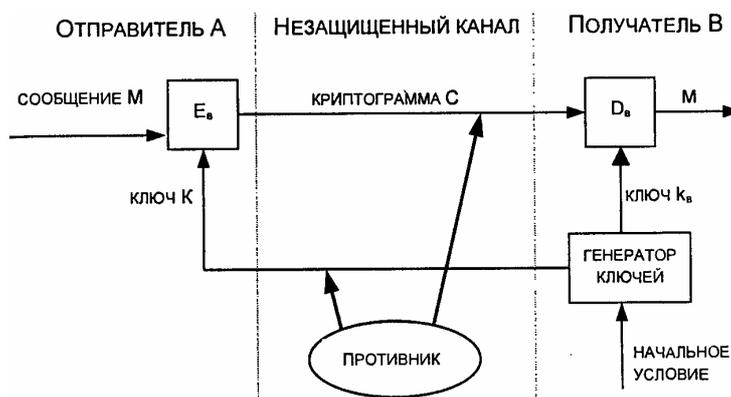


Рисунок 8.10 - Асимметричная криптосистема с открытым ключом

В этой асимметричной криптосистеме применяют два различных ключа:

- K_e - открытый ключ отправителя А;
- k_e - секретный ключ получателя В.

Генератор ключей целесообразно располагать на стороне получателя B (чтобы не пересылать секретный ключ k_e по незащищенному каналу). Значения ключей K_e и k_e зависят от начального состояния генератора ключей.

Раскрытие секретного ключа k_e по известному открытому ключу K_e должно быть вычислительно неразрешимой задачей.

Характерные особенности асимметричных криптосистем:

1. Открытый ключ K_e и криптограмма C могут быть отправлены по незащищенным каналам, т.е. противнику известны K_e и C .
2. Алгоритмы шифрования и расшифрования:

$$\begin{aligned} E_B : M &\rightarrow C, \\ E_B^{-1} : C &\rightarrow M \end{aligned}$$

являются открытыми.

Защита информации в асимметричной криптосистеме основана на секретности ключа k_e .

У. Диффи и М. Хеллман сформулировали требования, выполнение которых обеспечивает безопасность асимметричной криптосистемы:

1. Вычисление пары ключей (K_e, k_e) получателем B на основе начального условия должно быть простым.
2. Отправитель A , зная открытый ключ K_e и сообщение M , может легко вычислить криптограмму

$$C = E_{K_e}(M) .$$

3. Получатель B , используя секретный ключ k_e и криптограмму C , может легко восстановить исходное сообщение

$$M = E_{k_e}^{-1}(C) = E_{k_e}^{-1}(C) = E_{k_e}^{-1}[E_{K_e}(M)] .$$

4. Противник, зная открытый ключ K_e , при попытке вычислить секретный ключ k_e наталкивается на непреодолимую вычислительную проблему.

Противник, зная пару (K_e, C) , при попытке вычислить исходное сообщение M наталкивается на непреодолимую вычислительную проблему.

8.3.1 Однонаправленные функции

Концепция асимметричных криптографических систем с открытым ключом основана на применении однонаправленных функций.

Неформально однонаправленную функцию можно определить следующим образом. Пусть X и Y - некоторые произвольные множества. Функция

$$f : X \rightarrow Y$$

является однонаправленной, если для всех $x \in X$ можно легко вычислить функцию

$$y = f(x), \text{ где } y \in Y.$$

И в то же время для большинства $y \in Y$ достаточно сложно получить значение $x \in X$, такое, что $f^{-1}(x) = y$ (при этом полагают, что существует по крайней мере одно такое значение x).

Основным критерием отнесения функции f к классу однонаправленных функций является отсутствие эффективных алгоритмов обратного преобразования $Y \rightarrow X$.

В качестве первого примера однонаправленной функции рассмотрим целочисленное умножение. Прямая задача - вычисление произведения двух очень больших целых чисел P и Q , т.е. нахождение значения

$$N = P \cdot Q,$$

является относительно несложной задачей.

Обратная задача-разложение на множители большого целого числа, т.е. нахождение делителей P и Q большого целого числа $N = P \cdot Q$, является практически неразрешимой задачей при достаточно больших значениях N . По современным оценкам теории чисел при целом $N \approx 2^{664}$ и $P \approx Q$ для разложения числа N потребуется около 10^{23} операций, т.е. задача практически неразрешима на современных ЭВМ.

Следующий характерный пример однонаправленной функции - это модульная экспонента с фиксированными основанием и модулем. Пусть A и N -целые числа, такие, что $1 \leq A < N$. Определим множество Z_N :

$$Z_N = \{0, 1, 2, \dots, N-1\}.$$

Тогда модульная экспонента с основанием A по модулю N представляет собой функцию

$$f_{A,N} : Z_N \rightarrow Z_N,$$

$$f_{A,N}(x) = A^x \bmod N,$$

где x - целое число, $1 < x < N-1$; операция $i \bmod j$ - остаток от целочисленного деления i на j .

Существуют эффективные алгоритмы, позволяющие достаточно быстро вычислить значения функции $f_{A,N}(x)$.

Если $y = A^x$, то естественно записать $x = \log_A(y)$.

Поэтому задачу обращения функции $f_{A,N}(x)$ называют задачей нахождения дискретного логарифма или задачей дискретного логарифмирования.

Задача дискретного логарифмирования формулируется следующим образом. Для известных целых A , N , y найти целое число x , такое, что

$$A^x \bmod N = y.$$

Алгоритм вычисления дискретного логарифма за приемлемое время пока не найден. Поэтому модульная экспонента считается однонаправленной функцией.

По современным оценкам теории чисел при целых числах $A \approx 2^{664}$ и

$N \approx 2^{64}$ решение задачи дискретного логарифмирования (нахождение показателя степени x для известного y) потребует около 10^{26} операций, т.е. эта задача имеет в 10^3 раз большую вычислительную сложность, чем задача разложения на множители. При увеличении длины чисел разница в оценках сложности задач возрастает.

Следует отметить, что пока не удалось доказать, что не существует эффективного алгоритма вычисления дискретного логарифма за приемлемое время. Исходя из этого, модульная экспонента отнесена к однонаправленным функциям условно, что, однако, не мешает с успехом применять ее на практике.

Вторым важным классом функций, используемых при построении криптосистем с открытым ключом, являются так называемые **однонаправленные функции с "потайным ходом"** (с лазейкой). Дадим неформальное определение такой функции.

Функция

$$f : X \rightarrow Y$$

относится к классу однонаправленных функций с **"потайным ходом"** в том случае, если она является однонаправленной и, кроме того, возможно эффективное вычисление обратной функции, если известен "потайной ход" (секретное число, строка или другая информация, ассоциирующаяся с данной функцией).

В качестве примера однонаправленной функции с "потайным ходом" можно указать используемую в криптосистеме RSA модульную экспоненту с фиксированными модулем и показателем степени. Переменное основание модульной экспоненты используется для указания числового значения сообщения M либо криптограммы C .

8.3.2 Криптосистема шифрования данных RSA

Алгоритм RSA предложили в 1978 г. три автора: Р. Райвест (Rivest), А. Шамир (Shamir) и А. Адлеман (Adleman). Алгоритм получил свое название по первым буквам фамилий его авторов. Алгоритм RSA стал первым полноценным алгоритмом с открытым ключом, который может работать как в режиме шифрования данных, так и в режиме электронной цифровой подписи.

Надежность алгоритма основывается на трудности факторизации больших чисел и трудности вычисления дискретных логарифмов.

Введем следующие понятия:

- простое число - делится только на 1 и на само себя;
- взаимно простым- не имеют ни одного общего делителя, кроме 1;
- результат операции $i \bmod j$ - остаток от целочисленного деления i на j .

В криптосистеме RSA открытый ключ K_e , секретный ключ k_e , сообщение M и криптограмма C принадлежат множеству целых чисел

$$Z_N = \{0, 1, 2, \dots, N-1\},$$

где N - модуль: $N = P \cdot Q$.

Здесь P и Q - случайные большие простые числа. Для обеспечения максимальной безопасности выбирают P и Q равной длины и хранят в секрете.

Множество Z_N с операциями сложения и умножения по модулю N образует арифметику по модулю N .

Открытый ключ K_e выбирают случайным образом так, чтобы выполнялись условия:

$$1 < K_e \leq \varphi(N), \text{НОД}(K_e, \varphi(N)) = 1, \varphi(N) = (P-1)(Q-1),$$

где: $\varphi(N)$ - функция Эйлера; **НОД** - наибольший общий делитель.

Функция Эйлера $\varphi(N)$ указывает количество положительных целых чисел в интервале от 1 до N , которые взаимно просты с N .

Второе из указанных выше условий означает, что открытый ключ K_e и функция Эйлера $\varphi(N)$ должны быть взаимно простыми.

Далее, вычисляют секретный ключ k_e , такой, что:

$$k_e \cdot K_e \bmod \varphi(N) = 1$$

или

$$k_e = \left(\frac{1}{K_e} \right) \bmod ((P-1)(Q-1)).$$

Это можно осуществить, так как получатель B знает пару простых чисел (P, Q) и может легко найти $\varphi(N)$. Заметим, что k_e и N должны быть взаимно простыми.

Открытый ключ K_e используют для шифрования данных, а секретный ключ k_e - для расшифрования.

Преобразование шифрования определяет криптограмму C через пару (открытый ключ K_e , сообщение M) в соответствии со следующей формулой:

$$C = E_{K_e}(M) = (M^{K_e}) \bmod N$$

В качестве алгоритма быстрого вычисления значения C используют ряд последовательных возведений в квадрат целого M и умножений на M с приведением по модулю N .

Обращение функции $C = M^{K_e} \bmod N$, т.е. определение значения M по известным значениям C, K_e и N , практически не осуществимо при $N \approx 2^{512}$.

Однако обратную задачу, т.е. задачу расшифрования криптограммы C , можно решить, используя пару (секретный ключ k_e , криптограмма C) по следующей формуле:

$$M = E_{k_e}^{-1}(C) = (C^{k_e}) \bmod (N)$$

Таким образом, получатель B , который создает криптосистему, защищает два параметра:

1) секретный ключ k_e

2) пару чисел (P, Q) , произведение которых дает значение модуля N .

С другой стороны, получатель B открывает значение модуля N и открытый ключ K_e .

Противнику известны лишь значения K_e и N . Если бы он смог разложить число N на множители P и Q , то он узнал бы "потайной ход" - тройку чисел $\{P, Q, K_e\}$, вычислил значение функции Эйлера

$$\varphi(N) = (P-1)(Q-1)$$

и определил значение секретного ключа k_e . Однако, как уже отмечалось, разложение очень большого N на множители вычислительно не осуществимо (при условии, что длины выбранных P и Q составляют не менее 100 десятичных знаков).

ЗАКЛЮЧЕНИЕ

Общее направление развития телекоммуникаций в конце XX - начале XXI века характеризуется следующими понятиями: универсализация, интеграция и интеллектуализация - в части технических средств и в сетевом плане; глобализация и персонализация - в части услуг.

Одной из важнейших тенденций рубежа веков является бурный рост обмена информацией. Еще несколько лет назад подавляющая часть сетевого трафика приходилась на телефонию. Объем трафика в течение многих лет рос медленно и предсказуемо. Взрывной рост обмена мультимедийной информацией в таких сетях, как Интернет, радикально изменил эту ситуацию.

Резко возросла потребность в радиосистемах передачи цифровой информации, обеспечивающих существенно более высокую скорость передачи и пропускную способность по сравнению с существующими, в том числе в интересах управления авиацией. При этом выполнение перечисленных требований к системам передачи информации должно происходить во все более усложняющихся условиях их функционирования. Во многих реальных каналах связи простым выбором сигналов (их ансамбля, структуры, вида модуляции) не удастся получить значений пропускной способности, близких к предельным, и минимальной вероятности ошибки, определенных В. А. Котельниковым и К.Шенноном для каналов с аддитивной флуктуационной помехой.

Прогресс в области создания перспективных авиационных систем радиосвязи основан на разработке и освоении новых телекоммуникационных технологий, а также дальнейшем развитии и совершенствовании еще не исчерпавших свой потенциал существующих.

В каналах с комплексом помех и искажений качественно новые результаты могут быть достигнуты при построении систем связи на основе применения широкополосных сложных сигналов при кодовом разделении каналов и помехоустойчивых корректирующих кодов, исправляющих ошибки большой кратности при сравнительно малой избыточности. Этому способствует бурное развитие микроэлектронной техники, позволяющее создавать приемлемые по габаритам, надежности и стоимости устройства формирования и обработки сложных сигналов и кодеров-декодеров для кодов с высокой исправляющей способностью.

Рассмотренные соображения должны учитываться при выборе оптимальных методов построения перспективных систем связи, работающих в реальных - весьма сложных условиях.

Дальнейшее развитие систем связи базируется как на достижениях микроэлектроники, так и на широком внедрении средств компьютерной техники, которое определяет алгоритмы и программы работы соответствующего обо-

рудования; компьютеры и процессоры - объединенные элементы вычислительной техники). Последнее позволяет обеспечить интеграцию сетей связи путем создания единой информационной сети на основе первичного канала связи, способного передавать все виды информации; при этом создается единая система контроля, управления, коммутации и распределения информации и каналообразования.

ЛИТЕРАТУРА

1. Авиация ПВО России и научно-технический прогресс: Боевые комплексы и системы вчера, сегодня, завтра [Текст]: / Под ред. Е. А. Федосова - М.: Дрофа, 2001.— 816 с. ISBN 5—7107—133—7.
2. Авиационные радиосвязные устройства [Текст]: Учебник для слушателей и курсантов инженерных вузов ВВС / под ред. В.И. Тихонова. Изд-во ВВИА им. проф. Н.Е. Жуковского, 1987. – 442 с.
3. Васильев, В.И. Системы связи [Текст]: учеб. пособие для вузов / В.И. Васильев, А.П. Буркин, В.А. Свириденко – М.: Высшая школа, 1987, – 280 с.: ил.
4. Кларк Дж., Кейн Дж. Кодирование с исправлением ошибок в системах цифровой связи [Текст]: [пер. с англ.] С.И. Гельфанда / под ред. Б.С. Цыбакова – М.: Радио и связь, 1987. – 392 с.: ил.
5. Кловский, Д.Д., Теория передачи сигналов в задачах [Текст]: учебное пособие для вузов / Д.Д. Кловский, В.А. Шилкин. – М.: Связь, 1978. – 252 с. ил.
6. Лялюк, И.Н. Системы связи, АСУ и разведки Вооруженных сил США [Текст]: учебное пособие / И.Н. Лялюк. – М.: Изд-во ВАГУ, 2000. – 110 с.
7. Набатов, О.С. Связь в автоматизированных системах управления воздушным движением [Текст]: учеб. пособие для вузов / О.С. Набатов, Н.С. Вдовиченко. – М.: «Транспорт». 1984. – 287 с.: ил. - Библиогр.: с. 281-282.
8. Нефедов, В. И. Основы радиоэлектроники и связи [Текст]: учебник для вузов / В. И. Нефедов — 2-е изд., перераб. и доп. — М.: Высш. шк., 2002, - 510 с.: ил.
9. Основы организации связи в частях и подразделениях общевойсковых соединений [Текст]: учебник для вузов войск связи / под ред. В.В.Панкина — СПб.: Издат-во ВУС, 2002. - 320 с.: ил.
10. Передача дискретных сообщений [Текст]: учебник для вузов / В.П. Шувалов [и др.]; под ред. В.П. Шувалова. – М.: Радио и связь, 1990, - 464 с.: ил. - Библиогр.: с. 451-455.
11. Радиосистемы передачи информации [Текст]: учебное пособие для вузов / В.А. Васин [и др.]; под ред. И.Б. Федорова и В.В. Калмыкова. – М.: «Горячая линия – Телеком», 2005. – 472 с.: ил. - Библиогр.: с. 467-469.
12. Радиоприемные устройства [Текст]: учебник для вузов / Н.Н. Фомин [и др.]; под ред. Н.Н. Фомина. – Изд. 3-е, стереотип. – М.: «Горячая линия. – Телеком», 2007, – 520 с.: ил.
13. Радиоэлектронное оборудование [Текст]: учебник для вузов ВВС / В. А. Ефимов [и др.]; - М.: ВВИА им. проф. Н.Е. Жуковского, 2004, - 228 с.: ил. - Библиогр.: с. 227-228.
14. Рачков, В.Е. Топология пакетных радиосетей и способы ее реализации на

базе существующей системы связи позиционного района соединения РВСН. [Текст]: / В.Е. Рачков. – Ставрополь: СВИС РВ, 2005.- 83 с.

15. Средства и организация связи [Текст]: монография / В.Е. Ополовкин [и др.]; Часть 1. Основы передачи сообщений, средства проводной связи – М.: Изд-во МО РФ, 2007. - 144 с.: ил. - Библиогр.: с. 141-142.
16. Романец Ю. В., Тимофеев П. А., Шаньгин В. Ф. Защита информации в компьютерных системах и сетях. - М.: Радио и связь. 2001. - 376 с.