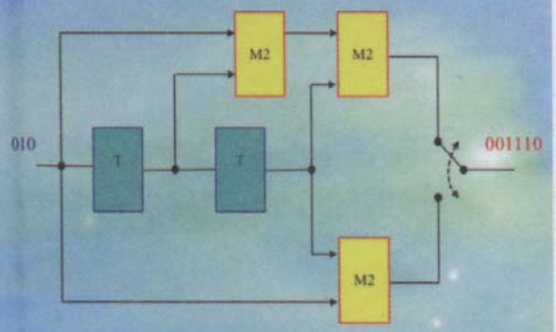
К. К. Васильев, В. А. Глушков,
А. В. Дормидонтов, А. Г. Нестеренко

ТЕОРИЯ ЭЛЕКТРИЧЕСКОЙ СВЯЗИ

$G_1(x)=1+x+x^2$



$G_2(x)=1+x^2$

Учебное пособие для ВУЗов

- сообщения, сигналы, помехи
- модуляция сигналов
- прием дискретных сообщений
- теория передачи и кодирования информации
- эффективность систем связи
- криптографическая защита информации

УДК 621.391 (075)
ББК 32.8 я 7
Т33

Утверждено редакционно-издательским советом
университета в качестве учебного пособия

Рецензенты:
Доктор технических наук, профессор Кумунжиев К.В.
Кафедра радиоэлектроники Ульяновского высшего военного инженерного училища связи

Авторы: К.К. Васильев, В.А. Глушков, А.В. Дормидонтов,
А.Г. Нестеренко

Теория электрической связи: учебное пособие / К.К. Васильев, В.А. Глушков, А.В. Дормидонтов, А.Г. Нестеренко; под общ. ред. К.К. Васильева. - Ульяновск: УлГТУ, 2008. - 452 с.
ISBN 978-5-9795-0203-8

Излагаются основные закономерности и методы анализа потенциальной помехоустойчивости и пропускной способности каналов связи. Рассматриваются параметры и характеристики сообщений, сигналов и помех, их математические модели, методы формирования и преобразования сигналов, вопросы теории передачи и кодирования сообщений, алгоритмы цифровой обработки сигналов, основные модели каналов электросвязи, принципы многоканальной связи и распределения информации, вопросы оценки эффективности систем связи и теоретико-информационные основы криптозащиты сообщений в телекоммуникационных системах.

Книга предназначена студентам, обучающимся по направлению 654400 - Телекоммуникации.

УДК 621.391 (075)
ББК 32.8 я 7

ISBN 978-5-9795-0203-8

ПРЕДИСЛОВИЕ

Пособие написано на основе опыта чтения авторами лекций по теории электрической связи. Базируясь на ранее опубликованных изданиях [8, 9, 12, 13, 14, 15], авторы обобщили и собрали воедино материалы [1, 17, 20, 22, 34], опубликованные в последние годы, дополнив их новыми информационными технологиями. Основными особенностями настоящего пособия является рассмотрение вопросов корреляционного анализа; эталонной модели взаимодействия открытых систем как единой идеологии проектирования систем связи; комплексного представления сигналов; некоторых методов модуляции, таких как квадратурная относительно-фазовая манипуляция и частотная модуляция с непрерывной фазой; методы сжатия дискретных сообщений; более детальное рассмотрение некоторых методов помехоустойчивого кодирования (коды БЧХ, Рида-Соломона, Рида-Маллера); уделено особое внимание методам приема сигналов в сложных условиях и теоретико-информационным основам криптозащиты сообщений в телекоммуникационных системах.

Настоящее учебное пособие состоит из десяти глав, в которых последовательно рассматриваются основополагающие вопросы теории электрической связи. Остановимся более подробно на содержании глав.

В первой главе пособия рассматриваются сообщения, сигналы и помехи, их математические модели, временное и спектральное представление сигналов и помех.

Во второй главе представлены методы формирования и преобразования дискретных сигналов, временные и спектральные характеристики манипулированных сигналов.

В третьей главе пособия анализируются оптимальные методы приема при различных видах передачи и способы реализации потенциальной помехоустойчивости.

В четвертой главе пособия рассматриваются основные понятия теории информации и дискретных систем обработки информации.

Пятая глава пособия посвящена анализу помехоустойчивых кодов, их классификации и сравнительным оценкам, на примере двоичных кодов, наиболее часто встречающихся в практике при проектировании и эксплуатации дискретных устройств.

В шестой главе пособия обобщены вопросы импульсных и цифровых методов передачи непрерывных сигналов.

В седьмой главе пособия анализируются методы передачи сигналов в условиях помех, решаются вопросы защиты каналов от замираний, межсимвольной интерференции и сосредоточенных помех, при передаче различных видов сообщений.

В восьмой главе пособия описываются физические явления и процессы, лежащие в основе многоканальной связи, вопросы объединения и разделения каналов, а также принципы построения систем множественного доступа и протоколы их реализации.

В девятой главе пособия рассмотрены вопросы оценки эффективности систем связи при заданной помехоустойчивости.

В десятой главе изложены классические методы шифрования и современные криптологические методы, алгоритмы, протоколы и системы как методы и средства защиты информации в компьютерных системах и системах связи.

Работа по написанию пособия распределилась следующим образом:

Главы 1 (кроме пп. 1.2.4, 1.6.2, 1.7), 2, 3 (кроме п. 3.1) написаны В. А. Глушковым, глава 4 - А. Г. Нестеренко, пп. 1.7, 3.1 - К. К. Васильевым, главы 6, 7 (кроме § 7.5), пп. 5.1, 5.2 - А. Г. Нестеренко, главы 8 (кроме пп. 8.1.2, 8.1.3), 9 - В. А. Глушковым, глава 10 - А. Г. Нестеренко, пп. 1.2.4, 1.6.2, 7.5, 8.1.2, 8.1.3 - А. В. Дормидонтовым. Глава 5 (кроме пп. 5.1, 5.2) повторяет материал учебного пособия: Васильев К. К., Новосельцев Л. Я., Смирнов В. Н. Основы теории помехоустойчивых кодов : учеб. пособие. - Ульяновск : УлГТУ, 2000.

Редактирование всех глав пособия выполнено К. К. Васильевым.

СПИСОК СОКРАЩЕНИЙ

АКФ	автокорреляционная функция;
АМн	амплитудная манипуляция;
АИМ	амплитудно-импульсная модуляция;
АЦП	аналого-цифровое преобразование;
БГШ	белый гауссовский шум;
БПФ	быстрое преобразование Фурье;
БЧХ	Боуза-Чоудхури-Хоквингема;
ВКФ	взаимно-корреляционная функция;
ВРК	временное разделение каналов;
ДДИМ	двухсторонняя длительно-импульсная модуляция;
ДИМ	длительно-импульсная модуляция;
ДК	дискретный канал;
ДМ	дельта-модуляция;
ДНК	дискретно-непрерывный канал;
ДПФ	дискретное преобразование Фурье;
ДСКБП	дискретный симметричный канал без памяти;
ИКМ	импульсно-кодовая модуляция;
КРК	кодовое разделение каналов;
КТЧ	канал тональной частоты;
МСД	многостанционный доступ;
МСИ	межсимвольная интерференция;
МЧС	многочастотный сигнал;
НВ	наиболее вероятное;
НК	непрерывный канал;
НОД	наибольший общий делитель;
НОК	наименьшее общее кратное;
ОДИМ	односторонняя длительно-импульсная модуляция;
ОСШ	отношение сигнал/шум;

ОФМн	относительная фазовая манипуляция;
ПАКФ	периодическая автокорреляционная функция;
ПВКФ	периодическая взаимно-корреляционная функция;
ППРЧ	псевдослучайное переключение рабочих частот;
ПСП	псевдослучайная последовательность;
РС	Рида-Соломона;
СКК	сигнально-кодовые конструкции;
СПМ	спектральная плотность мощности;
СФ	согласованный фильтр;
СЭС	система электрической связи;
УМ	угловая модуляция;
ФИМ	фазоимпульсная модуляция;
ФМн	фазовая манипуляция;
ФМ-ПСС	фазовая манипуляция псевдослучайными сигналами;
ФМ-ШПС	манипуляции широкополосными сигналами;
ЦАП	цифро-аналоговое преобразование;
ЦСП	цифровые системы передачи;
ЧИМ	частотно-импульсная модуляция;
ЧМн	частотная манипуляция;
ЧРК	частотное разделение каналов;
ШИМ	широтно-импульсная модуляция;

ОСНОВНЫЕ ОБОЗНАЧЕНИЯ

$A,\ B, C$ - случайные события;

$A(t) = \sqrt{S^2(t) + \tilde{S}^2(t)}$ - огибающая сигнала, мгновенная амплитуда;

$B = F \cdot T$ - база сигнала;

C - пропускная способность канала (бит/с), множество сообщений источника;

C' - пропускная способность канала (бит/символ или бит/отсчет), множество разрешенных кодовых слов;

C'' - множество возможных кодовых слов;

C_n^i - число сочетаний из n по i;

D_x - дисперсия случайной величины или процесса;

D_c - динамический диапазон;

$d(s_1, s_2)$ - расстояние между сигнальными точками

$d(C^1, C^2)$ расстояние по Хеммингу между двоичными последовательностями, минимальное расстояние по Хеммингу между комбинациями линейного блокового кода;

$E = \int_{t_1}^{t_2} S^2(t)dt$ - энергия сигнала;

$E(x)$ - ошибка в оценивании случайного параметра или процесса, шум наблюдения или квантования;

$F()$ - функция распределения вероятностей;

ΔF_c - ширина спектра сигнала (канала);

$f = 1/T$ - циклическая частота;

$\Delta f_{защ}$ - защитный интервал частот;

$G(\omega)$ - спектральная плотность мощности;

$g(t)$ - импульсная характеристика линейной цепи;

g, g' - выигрыш и обобщенный выигрыш системы модуляции;

$H(X), H(X/Y)$ - энтропия и условная энтропия дискретной случайной величины (дискретного источника);

$H'(X)$ - производительность дискретного источника;

$h(X), h(X/Y)$ - дифференциальная энтропия и условная дифференциальная энтропия непрерывной случайной величины;

$h^2 = \frac{E}{N_0}$ - отношение энергии элемента сигнала на входе демодулятора к односторонней спектральной плотности мощности белого шума;

$h_Э^2 = \frac{P}{N_0 R_И} = \frac{h^2}{R \log_2 m}$ - отношение нормированной энергии сигнала на 1 бит информации (битовой энергии) к односторонней спектральной плотности мощности белого шума;

$I(x_i)$ - количество информации в отдельно взятом единичном сообщении x_i;

$I(X,Y)$ - количество информации, переданной по дискретному каналу;

j - знак мнимой единицы, $j = \sqrt{-1}$;

$K_{СЖ}$ - коэффициент сжатия источника;

K - объем алфавита дискретного источника;

k - число информационных символов в кодой комбинации;

$m_{АМ}$ - индекс амплитудной модуляции;

$m_{УМ}$ - индекс угловой модуляции;

$m_x = M\{X\}$ - математическое ожидание случайной величины (процесса);

m - основание кода (объем алфавита кода);

N_0 - односторонняя (на положительных частотах) спектральная плотность мощности квазибелого и белого шума;

$n(t)$ - аддитивный белый гауссовский шум;

$n_{ср}$ - среднее число символов на одну букву;

n - длина (общее число символов) кодовой комбинации;

$P(t) = S^2(t)$ - мгновенная мощность сигнала $S(t)$;

$P_{ср} = \frac{1}{T}\int_{t_1}^{t_2} S^2(t)dt$ - средняя мощность сигнала;

$P(), P_x$ - безусловная вероятность события, указанного в скобках или обозначенного индексом;

$p(x_i)$ - вероятность появления символа алфавита;

$p(y_i / x_i)$ - переходные вероятности появления символа y_i при условии передачи символа x_i;

$p_b(p_Э)$ - вероятность ошибки на один информационный бит (эквивалентная вероятность ошибки)

$p_{ош}$ - вероятность ошибочного приема символа;

$R = \frac{k}{n}$ - скорость кода;

$R_И = \frac{R\log_2 m}{T}$ - максимальная производительность при $R = 1$ (информационная скорость) дискретного источника (бит/с), скорость передачи информации;

$r = n - k$ - число проверочных символов в кодовых комбинациях блочного кода;

$R(\tau)$ - нормированная функция корреляции, коэффициент корреляции;

$S = T / \tau_0$ - скважность импульсной последовательности;

$S(t)$ - случайный сигнал на выходе модулятора;

$\dot{S}(t)$ - комплексный (аналитический) сигнал;

$\tilde{S}(t)$ - преобразование Гильберта от сигнала $S(t)$;

$S_\phi = \frac{ds(t)}{dt} = tg\varphi$ - крутизна фронта колоколообразного импульса;

$\dot{S}(\Omega)$ - спектральная плотность сигнала по Фурье;

$s(t)$ - реализация случайного сигнала на выходе передатчика;

T - длительность тактового интервала, длительность финитного сигнала, знак транспонирования матрицы;

t - текущее время;

$U(t)$ - случайный сигнал на входе приемника (детектора) без учета аддитивных помех;

$u(t)$ - реализация случайного сигнала на входе приемника (детектора) без учета аддитивных помех;

$V_И$ - скорость передачи (число символов в секунду) дискретного источника (канала), число отсчетов в одну секунду непрерывного сигнала;

$W(C)$ - вес кодового слова;

W_j - функция Уолша;

$w(x,t)$ - одномерная плотность распределения вероятности случайной величины (случайного процесса);

X, Y - алфавит сообщений на входе и выходе дискретного канала;

$x(t), y(t)$ - вектор (цепочка символов) сообщений на входе и выходе дискретного канала;

α - коэффициент группирования ошибок;

$\beta(M) = \frac{\Delta f_\Sigma}{\Delta f_1}$ - спектральная цена уплотнения;

$\beta = \frac{R}{P_c / N_0}$ - коэффициент использования канала по мощности (энергетическая эффективность системы);

$\delta()$ - дельта-функция;

$\gamma = \frac{R}{\Delta F}$ - коэффициент использования канала по полосе частот (частотная эффективность системы);

Δ - шаг дискретизации непрерывного сигнала во времени;

$\eta = \frac{P_{ош}}{P_{ДКи(о)}}$ - эффективность помехоустойчивого кода;

$\eta = \frac{R}{C}$ - коэффициент использования канала по пропускной способности (информационная эффективность системы);

$\eta(M) = \frac{P_{\Sigma}}{P_1}$ - энергетическая цена уплотнения;

χ - избыточность источника, кода;

Λ - отношение правдоподобия;

$\mu_k, \mu_k(t)$ - коэффициент передачи;

υ - кодовое ограничение сверточного кода;

П - пик-фактор сообщения или сигнала (отношение максимального значения к среднеквадратическому);

$\rho(\tau)$ - нормированная периодическая автокорреляционная функция;

$\sigma_x = \sqrt{D_x}$ - среднеквадратическое отклонение

$\tau_k, \tau_k(t)$ - время задержки;

$\Delta\tau_{мл}$ - время многолучевости;

$\Phi_0(z)$ - интеграл вероятности;

$\varphi_0 = \frac{2\pi\tau_u}{T}$ - начальная фаза;

$\{\varphi_k(t)\}$ - базис разложения;

$\varphi(t) = ArgS(t)$ - мгновенная фаза сигнала;

$\omega = 2\pi f$ - угловая частота;

$\omega(t) = \frac{d\varphi(t)}{dt}$ - мгновенная частота сигнала.

ВВЕДЕНИЕ

Теория электрической связи относится к числу фундаментальных дисциплин подготовки инженеров, владеющих современными методами анализа и синтеза систем и устройств связи различного назначения и имеет цель сформировать знания основ теорий передачи и кодирования сообщений, методов передачи и приема дискретных и непрерывных сообщений, цифровых методов передачи сообщений, принципов построения многоканальных систем передачи и методов повышения эффективности систем электросвязи, а также умений использовать методы анализа систем электрической связи для количественной оценки их эффективности.

Предметом изучения дисциплины являются закономерности процессов преобразования и передачи информации в системах электросвязи.

Знания и умения по дисциплине являются составной частью общепрофессиональной подготовки к самостоятельной инженерно-эксплуатационной деятельности.

Дисциплина базируется на предшествующем изучении физики, математики, дискретной математики, теории вероятностей, математической статистики и информатики. В свою очередь «Теория электрической связи» является базовой для дисциплин «Радиопередающие устройства», «Радиоприемные устройства», «Цифровые системы передачи», «Микропроцессоры и цифровая обработка сигналов», «Средства и комплексы радиорелейной, спутниковой и электропроводной связи».

Основы современной теории электрической связи были заложены в фундаментальных работах В.А. Котельникова по теории потенциальной помехоустойчивости (1947 г.) и К. Шеннона по теории информации (1948 г.). Отдельные вопросы теории связи рассматривались в более ранних работах Х. Найквиста (1928 г.) и В.А. Котельникова (1933 г.), в которых сформулирована и доказана теорема отсчетов, в работе Р. Хартли (1928 г.), где была введена логарифмическая мера количества информации. В создании и развитии статистической теории связи большую роль сыграли работы А.Я. Хинчина (1938 г.) по корреляционной теории стационарных случайных процессов, А.Н. Колмогорова (1941 г.) и Н. Винера (1947 г.) по интерполированию и экстраполированию стационарных случайных последовательностей, А. Вальда (1950 г.) по теории ста-

тистических решений. Дальнейшее развитие теория получила в работах В.И. Сифорова, А.А. Харкевича, Л.М. Финка, Д.Р. Левина, Д.Д. Кловского, Р. Райса, Р. Галлагера, К. Хелстрома, Р. Фано и многих других отечественных и зарубежных ученых.

Классическая теория помехоустойчивости при флуктуационных помехах развита в этих работах для каналов со случайно изменяющимися параметрами и продолжает развиваться в направлении учета реальных характеристик сигналов и помех, в том числе нестационарных. Вопросы синтеза оптимальных приемников непрерывных и импульсных сигналов успешно решаются на основании теории нелинейной фильтрации. Дальнейшим шагом является разработка и применение методов построения адаптивных систем, позволяющих обеспечить высокую достоверность передачи сообщений в каналах с переменными параметрами при неполной априорной информации о сигналах и помехах.

Современная теория связи позволяет достаточно полно оценить различные системы по их помехоустойчивости и эффективности и тем самым определить, какие из них являются наиболее перспективными. Она достаточно четко указывает не только возможности совершенствования существующих систем связи, но и пути создания новых, более совершенных систем.

В настоящее время речь идет о создании систем, в которых достигаются показатели эффективности, близкие к предельным. Одновременное требование высоких скоростей и верности передачи приводит к необходимости применения систем, в которых используются многопозиционные сигналы и мощные корректирующие коды. Наиболее совершенная система связи должна быть саморегулирующей (адаптивной) системой.

Однако не следует думать, что во всех случаях необходимо стремиться к созданию сложных систем, отбрасывая простые как менее совершенные. Разработка наиболее совершенных систем передачи информации всегда должна базироваться на технико-экономическом расчете. Сложность систем не должна превосходить определенного экономически обоснованного уровня. По этой причине не следует чрезмерно усложнять системы в погоне за их максимальным совершенством. В ряде случаев более простые системы могут иметь необходимую степень совершенства, а экономически быть более целесообразными.

ГЛАВА 1. СООБЩЕНИЯ, СИГНАЛЫ И ПОМЕХИ, ИХ МАТЕМАТИЧЕСКИЕ МОДЕЛИ

1.1. Основные понятия и определения

1.1.1. Сообщение, сигнал, модуляция

Информацией называется совокупность сведений о каких-либо событиях, явлениях или предметах [6, 13, 21, 39].

Сообщение – форма представления информации, предназначенная для передачи от источника к получателю в виде текста, звука, изображения и т.д.

Например, при телеграфной передаче сообщением является текст телеграммы в виде букв или цифр. При разговоре по телефону – непрерывное изменение во времени звукового давления. В телевизионных системах сообщение представляет собой изменение во времени яркости элементов изображения.

Для передачи сообщений от источника к получателю с помощью электрической связи используют сигналы.

Сигнал это физический процесс, отображающий (несущий) передаваемое сообщение, т.е. это изменяемая физическая величина (ток, напряжение, электромагнитное поле, световые волны и т.д.).

Различают первичные и вторичные сигналы. Первичные электрические сигналы (ПЭС) возникают в результате непосредственного преобразования сообщения в электромагнитное колебание, обычно на выходе оконечных устройств. К ним относятся колебания тока микрофона, тока на выходе телеграфного аппарата и т. п. Характерным для первичных сигналов является относительно малая скорость их изменения и, следовательно, возможность передачи по низкочастотным каналам связи, например таким, как проводные. Так для передачи речи достаточен канал, пропускающий колебания от 300 до 3400 Гц. При телеграфной связи требуется полоса пропускания до нескольких сотен герц.

Для передачи сообщения по радиоканалам необходимо его «записать» на высокочастотном колебании. Такая запись осуществляется в результате модуляции (манипуляции) первичным сигналом высокочастотного колебания. В результате образуется сигнал, который будем называть вторичным. Применение высокочастот-

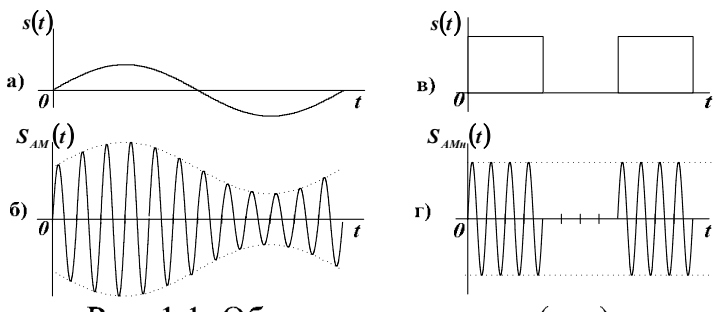


Рис. 1.1. Образцы первичных (а, в) и вторичных (б, г) сигналов

ных модулированных сигналов решает задачи использования физических свойств радиочастот, согласования геометрических размеров антенны с длиной волны колебаний, помехоустойчивости приема, частотного разнесения преобразованных ПЭС.

На (рис.1.1, а, в) показаны образцы первичных и вторичных (рис.1.1, б, г) сигналов при передаче речи (непрерывного сообщения) и телеграммы (дискретного сообщения).

Модуляция – это изменение во времени одного или нескольких параметров высокочастотного электрического колебания в соответствии с законом изменения передаваемого сообщения.

1.1.2. Основные параметры сигналов

Основными параметрами сигналов являются длительность сигнала T_c, динамический диапазон D_c и ширина спектра ΔF_c.

Всякий сигнал, рассматриваемый как временной процесс, имеет начало и конец. Поэтому длительность сигнала T_c является естественным его параметром, определяющим интервал времени, в пределах которого сигнал существует.

Динамический диапазон D_c – это отношение наибольшей мгновенной мощности сигнала $P_{c\max}$ к той наименьшей мощности $P_{c\min}$, которая необходима для обеспечения заданного качества передачи. Он выражается в децибелах [дБ]:

$$D_c = 10\lg\frac{P_{c\max}}{P_{c\min}} \text{ (дБ).}$$

Например, в радиовещании динамический диапазон часто сокращают до 30...40 дБ (1000-10000 раз) во избежание перегрузок канала.

Ширина спектра ΔF_c – этот параметр дает представление о скорости изменения сигнала внутри интервала его существования.

Спектр сигнала, в принципе, может быть неограниченным. Однако для любого сигнала можно указать диапазон частот, в пределах которого сосредоточена его основная энергия. Этим диапазоном и определяется ширина спектра сигнала. В технике связи спектр сигнала часто сознательно сокращают. Это обусловлено тем, что аппаратура и линия связи имеют ограниченную полосу пропускаемых частот. Сокращение спектра осуществляется исходя из допустимых искажений сигнала.

Например, ширина спектра телефонного сигнала:

$\Delta F_c = f_{\max} - f_{\min} = 3400 - 300 = 3100$ (Гц), а ширина спектра телевизионного сигнала при стандарте 625 строк составляет около 6 (МГц). Ширина спектра телеграфного сигнала зависит от скорости передачи и обычно принимается равной $1{,}5 \cdot V$ (Гц), где V – скорость телеграфирования в бодах, т.е. число символов, передаваемых в секунду. Так, при скорости передачи $V = 50$ Бод ширина спектра телеграфного сигнала $\Delta F_c = 75$ (Гц). Спектр модулированного сигнала (вторичного сигнала) обычно шире спектра передаваемого сообщения (первичного сигнала) и зависит от вида модуляции.

Часто вводят довольно общую и наглядную характеристику – объем сигнала:

$$V_c = T_c \cdot D_c \cdot \Delta F_c .$$

Объем сигнала V_c дает общее представление о возможностях данного множества сигналов как переносчиков сообщений. Чем больше объем сигнала, тем больше информации можно вложить в этот объем, но тем труднее передать такой сигнал по каналу связи.

1.2. Системы связи. Каналы связи

1.2.1. Структура канала электросвязи

Из приведенных ранее определений следует, что в любой системе электросвязи должны быть устройства, осуществляющие преобразования: на передаче – информация → сообщение → сигнал, на приеме – сигнал → сообщение → информация.

Кроме того, в процессе передачи сигнал подвергается и другим преобразованиям, многие из которых являются типовыми, обязательными для различных систем электросвязи, независимо от их назначения и характера передаваемых сообщений.

Рассмотрим обобщенную структурную схему системы электрической связи (СЭС) (рис.1.2.) [6, 39]. В нее входят следующие элементы.

Источник сообщения это физический объект, который формирует конкретное сообщение $x(t)$ (люди, ЭВМ, датчики). Примеры сообщений: речь, музыка, фотография, текст, рисунок.

Преобразователи сообщения в электрический сигнал (микрофон, датчик) превращают сообщение $x(t)$ в первичный сигнал $s(t)$. Например, преобразование букв текста в стандартные электрические сигналы азбуки Морзе.

Модулятор – осуществляет преобразование первичного сигнала $s(t)$ во вторичный сигнал $S(t)$, удобный для передачи в среде распространения в усло-

виях действия помех.

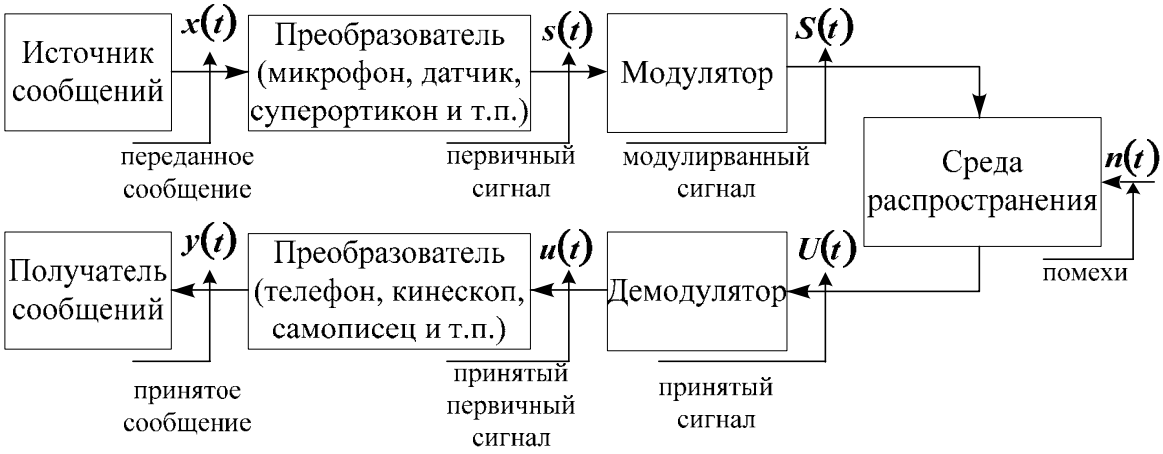


Рис. 1.2. Обобщенная структурная схема системы электрической связи

Среда распространения служит для передачи электрических сигналов от передатчика к приемнику. Это может быть кабель или волновода, в системах радиосвязи это область пространства в котором распространяются электромагнитные волны от передающей антенны к приемной.

Для каждого типа линии связи имеются сигналы, которые могут быть использованы наиболее эффективно. Например, в проводной линии применяются переменные токи невысоких частот (не более сотен кГц), в радиолинии – электромагнитные колебания высоких частот (от сотен килогерц до десятков тысяч мегагерц), а в волоконно-оптических линиях для передачи информации используют световые волны с частотами $10^{14}…10^{15}$ Гц. В среде распространения сигналы обычно значительно ослабляются (затухают) и искажаются под воздействием помех $n(t)$.

Под помехой понимается любое воздействие на сигнал, которое ухудшает достоверность воспроизведения передаваемых сообщений. В наиболее простом случае на вход демодулятора (приемника) поступает сумма сигнала $S(t)$ и помехи $n(t)$: $U(t)=S(t)+n(t)$. Такие помехи называют аддитивными.

Демодулятор это устройство, в котором из принятого сигнала $U(t)$ выделяется первичный электрический сигнал $u(t)$, который из-за действия помех может значительно отличаться от переданного $s(t)$.

Преобразователь необходим для формирования $y(t)$ сообщения из принятого первичного сигнала $u(t)$. Качество СЭС определяется степенью соответствия принятого сообщения $y(t)$ переданному сообщению $x(t)$.

Структурная схема системы электрической связи для передачи дискретных сообщений (рис. 1.3) дополнительно включает в себя кодер (декодер) ис-

точника и кодер (декодер) канала.

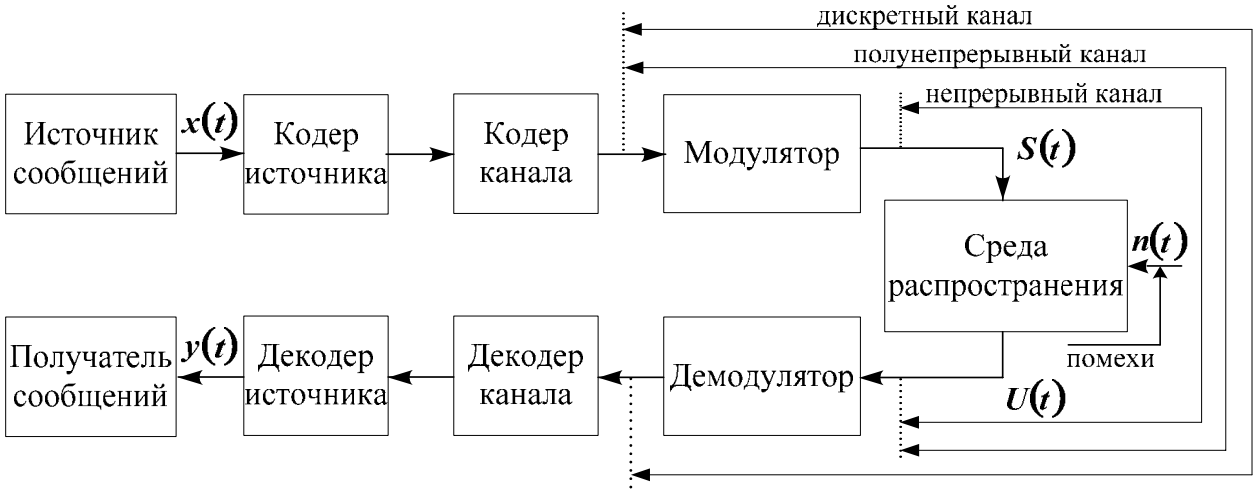


Рис. 1.3. Структурная схема системы электрической связи для передачи дискретных сообщений

Кодер источника служит для преобразования сообщений в кодовые символы с целью уменьшения избыточности источника сообщения, т.е. обеспечении минимума среднего числа символов на одно сообщение и представления в удобной форме (например, в виде двоичных чисел).

Кодер канала, предназначен для введения избыточности, позволяющей обнаруживать и исправлять ошибки в канальном декодере, с целью повышения достоверности передачи.

Декодер канала обеспечивает проверку избыточного (помехоустойчивого) кода и преобразование его в последовательность первичного электрического сигнала безызбыточного кода.

Декодер источника (ДИ) – это устройство для преобразования последовательности ПЭС безизбыточного кода в сообщение.

Принято различать две группы относительно самостоятельных устройств: кодеки и модемы. Кодеком называется совокупность кодера и декодера, которые при двухсторонней связи конструктивно объединены в одно устройство. Модемом называется конструктивно совмещенная совокупность модулятора и демодулятора.

Важнейшей частью СЭС является канал связи.

Каналом связи называется совокупность средств, обеспечивающих передачу сигнала от некоторой точки А системы до точки В. Точки А и В могут быть выбраны различным образом в зависимости от решаемой задачи построения модели, проектирования или анализа СЭС. В зависимости от вида входных и выходных символов канал связи может быть непрерывным, дискретным и полунепрерывным. В одной и той же схеме можно выделить как дискретный так и непрерывный канал, в зависимости от выбора рассматриваемых точек.

1.2.2. Линия и сеть связи

Линией связи называется физическая среда и совокупность аппаратных средств, используемых для передачи сигналов от передатчика к приемнику. В системах проводной связи это, прежде всего, кабель или волновод, в системах радиосвязи – область пространства, в котором распространяются электромагнитные волны от передатчика к приемнику. При передаче по каналу сигнал $S(t)$, может искажаться и на него могут воздействовать помехи $n(t)$. Приемное устройство обрабатывает принятый сигнал $U(t)=S(t)+n(t)$, представляющий собой сумму пришедшего искаженного сигнала $S(t)$ и помехи $n(t)$, и восстанавливает по нему сообщение $y(t)$, которое с некоторой погрешностью отображает переданное сообщение $x(t)$. Другими словами, приемник должен на основе анализа сигнала $U(t)$ определить, какое из возможных сообщений передавалось. Поэтому приемное устройство является одним из наиболее ответственных и сложных элементов системы электрической связи.

Под системой электрической связи понимают совокупность технических средств и среды распространения. В понятие система связи включаются источник и потребитель сообщений.

По виду передаваемых сообщений различают следующие системы электрической связи: системы передачи речи (телефония); системы передачи текста (телеграфия); системы передачи неподвижных изображений (фототелеграфия); системы передачи подвижных изображений (телевидение), системы телеизмерения, телеуправления и передачи данных. По назначению телефонные и телевизионные системы делят на вещательные, отличающиеся высокой степенью художественности воспроизведения сообщений, и профессиональные, имеющие специальное применение (служебная связь, промышленное телевидение и т.п.). В системе телеизмерения физические величины (температура, давление, скорость и т.п.) с помощью датчиков преобразуются в первичный электрический сигнал, поступающий на передатчик. На приемном конце переданную физическую величину или ее изменения выделяют из сигнала и используют для контроля. В системе телеуправления осуществляется передача команд для автоматического выполнения определенных действий. Нередко эти команды формируют автоматически на основании результатов измерения, переданных телеметрической системой.

Внедрение высокоэффективных ЭВМ привело к необходимости быстрого развития систем передачи данных, обеспечивающих обмен информа-

цией между вычислительными средствами и объектами автоматизированных систем управления. Этот вид электросвязи отличается высокими требованиями к скорости и верности передачи информации.

Для обмена сообщениями между многими территориально разнесенными пользователями (абонентами) создаются сети связи, обеспечивающие передачу и распределение сообщений по заданным адресам (в заданное время и с установленным качеством).

Сетью связи называют совокупность линий связи и узлов коммутации.

Классификация каналов и линий связи осуществляется:

по характеру сигналов на входе и выходе (непрерывные, дискретные, дискретно-непрерывные);

по виду сообщений (телефонные, телеграфные, передачи данных, телевизионные, факсимильные и др.);

по виду среды распространения (проводные, радио, волоконно-оптические и др.);

по диапазону используемых частот (низкочастотные (НЧ), высокочастотные (ВЧ), сверхвысокочастотные (СВЧ) и др.);

по структуре приемо-передающих устройств (одноканальные, многоканальные).

В настоящее время с целью наиболее полной характеристики каналов и линий связи могут применяться и другие классификационные признаки (по способу распространения радиоволн, способу объединения и разделения каналов, размещению технических средств, оперативному предназначению и др.)

1.2.3. Помехи и искажения в канале

При передаче сигнала по линии связи он искажается и воспроизводится с некоторой ошибкой. Причиной таких ошибок являются искажения сигналов в канале связи и помехи, воздействующие на сигнал [5, 21].

Искажения часто обусловлены известными характеристиками линии связи и тогда могут быть устранены путем надлежащей коррекции.

Помехи заранее неизвестны и поэтому не могут быть полностью устранены. Они весьма разнообразны как по своему происхождению, так и по физическим свойствам. Можно дать следующую классификацию помех по месту их возникновения:

атмосферные помехи;

промышленные помехи (индустриальные помехи);

космические помехи;

электризационные помехи;

помехи посторонних каналов связи;

внутренние шумы.

Атмосферные помехи обусловлены электрическими процессами в атмосфере и, прежде всего, грозовыми разрядами. Энергия этих помех сосредоточена, главным образом, в области ДВ и СВ.

Промышленные помехи возникают из-за резких изменений тока в электрических цепях всевозможных электроустановок. К ним относятся помехи от электротранспорта, электрических моторов, медицинских установок, систем зажигания двигателей и т.д.

Космические помехи создаются радиоизлучением внеземных источников. Они создают общий шумовой фон и в наибольшей степени проявляются на ультракоротких волнах.

Электризационные помехи, часто возникающие во время пурги или песчаной бури, создаются наэлектризованными снежными частицами или песчинками. Эти помехи возникают при скорости ветра свыше 5,5 м/с и ощутимы на частотах ниже 15 МГц.

Помехи посторонних каналов связи – обусловлены работой посторонних радиостанций. С учетом источника происхождения их называют также стационарными. Этот вид помех наиболее характерен для КВ диапазоне.

В зависимости от характера изменения во времени различают флуктуационные, импульсные (сосредоточенные во времени) и узкополосные (сосредоточенные по спектру) помехи.

Флуктуационная помеха представляет собой непрерывное колебание, меняющееся случайным образом. Часто она описывается нормальным законом распределения. Быстрое изменение во времени позволяет заменить реальные флюктуационные помехи так называемым белым шумом - процессом с постоянным спектром.

Импульсные помехи представляет собой случайную последовательность коротких сигналов обычно следующих редко, что реакция приемника на текущий импульс успевает уменьшится до нуля к моменту появления очередного импульса. Типичными примерами таких помех являются сигналы, создаваемые разрядами молний или искрением контактов в электрических двигателях.

Сосредоточенные по спектру помехи занимают сравнительно узкую полосу частот, существенно меньшую полосы частот сигнала. Чаще всего они обусловлены сигналами посторонних радиостанций, или излучениями про-

мышленных или медицинских генераторов высокой частоты различного назначения.

В зависимости от характера воздействия различают аддитивную помеху $n(t)$ суммирующуюся с полезным сигналом и мультипликативную помеху $\mu(t)$

$$U(t) = S(t) + n(t), \qquad (1.1)$$

где $S(t)$ – переданный сигнал, $n(t)$ – аддитивная помеха;

мультипликативная помеха $\mu(t)$:

$$U(t) = \mu(t) \times S(t), \qquad (1.2)$$

где $\mu(t)$ – некоторая случайная функция, отражающая изменение во времени коэффициента передачи канала связи.

В реальных системах связи часто действуют как аддитивная, так и мультипликативная помехи:

$$U(t) = \mu(t) \times S(t) + n(t). \qquad (1.3)$$

1.2.4. Эталонная модель взаимодействия открытых систем

Решение задачи передачи сообщений по системам электрической связи предъявляет к ним определенные требования. Эти требования условно можно разделить на две группы: требования к процессу передачи сообщений и требования к техническим средствам, осуществляющим этот процесс.

В числе требований к техническим средствам систем электрической связи выделяют следующие. Во-первых, система связи должна обладать способностью наращивания своих возможностей и исключения не использующихся возможностей. Системы, обладающие такой способностью, называют открытыми. Во-вторых, различные системы связи должны иметь стандартизованные и унифицированные технические устройства, что удешевляет их стоимость и эксплуатацию. В третьих, системы связи различного назначения должны обладать возможностью взаимного обмена сообщениями.

Эти требования породили необходимость единой идеологии проектирования систем связи. Международный консультативный комитет по телефонии и телеграфии в начале 80-х годов предложил такую идеологию, разработав эталонную модель взаимодействия открытых систем (ЭМВОС).

В соответствии с этой моделью процесс передачи сообщений в системах связи последовательно разбивается на принципиально различающиеся операции. Каждую ив этих операций относят к своему уровню.

Уровни строятся по принципу строгой иерархии: на высшем уровне находятся источник и получатель информации - пользователи системы связи, на нижнем - среда распространения электромагнитных волн. Высший уровень

управляет работой низшего. Каждому уровню соответствует свое техническое устройство или организационная единица системы связи пользователь или должностное лицо, обеспечивающее функционирование системы связи. В некоторых системах связи часть этих устройств может отсутствовать либо выполнять не все функции некоторого уровня.

В ЭМВОС выделяют 7 уровней: пользовательский, представительский, сеансовый, транспортный, сетевой, канальный, физический (рис. 1.4). Полную совокупность средств у одного пользователя, выполняющих операции различных уровней, называют станцией [18, 39].

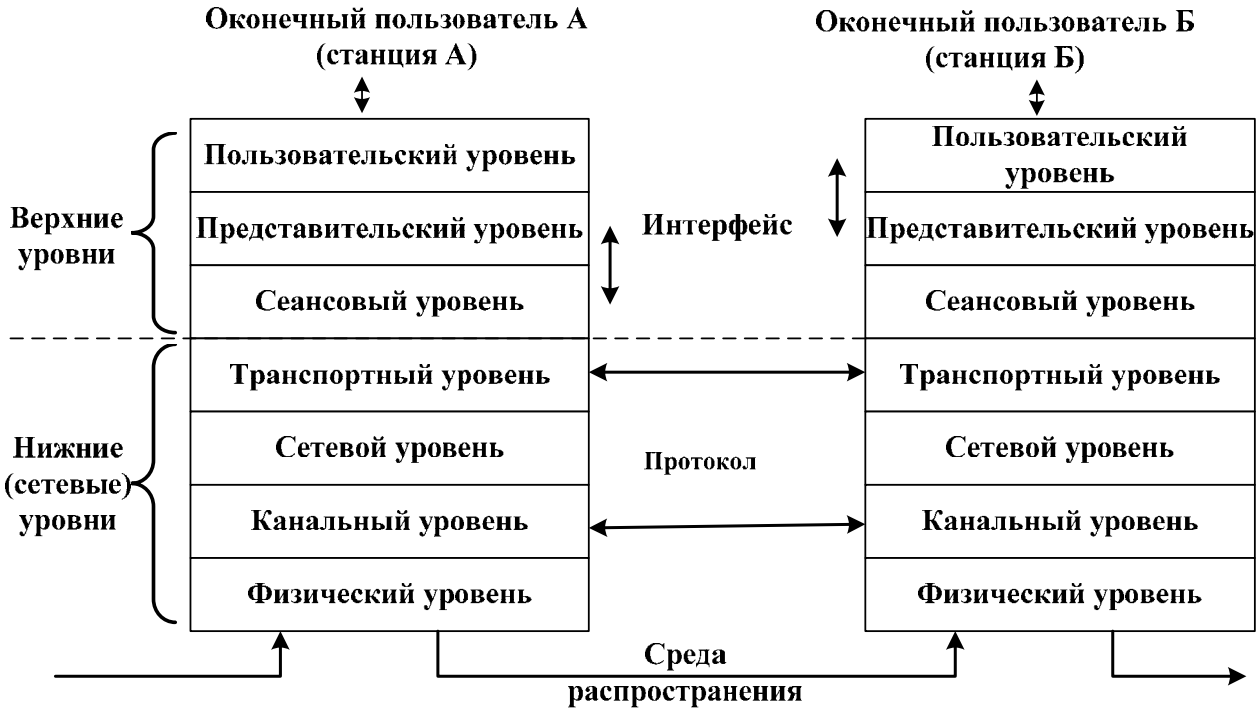


Рис. 1.4. Эталонная модель взаимодействия открытых систем (ЭМВОС)

На пользовательском уровне происходят процессы обработки информации, передаваемой системой связи. Исполнителем функций этого уровня может быть как техническое устройство (ЭВМ), так и человек.

Устройства представительского уровня преобразуют сообщения из формы представления, удобной пользователю, к форме представления, удобной системе связи, и обратно. В частности, на этом уровне происходит сжатие информации, поскольку системе связи всегда удобно, что бы сообщение занимало наименьший объем.

Устройства сеансового уровня обрамляют передаваемое сообщение служебной информацией с тем, чтобы количество топологических вариантов передачи было возможно большим. Выбор наилучшего варианта осуществляется устройствами нижних уровней. Таким образом, этот уровень отвечает за орга-

низацию сеанса связи.

На транспортном уровне принимается решение о перемещении данного сообщения к пользователю на уровне выбора необходимых сетей связи. Для этого решается задача межсетевой адресации сообщений и задача передами сообщений между сетями различного рода, называемая задачей шлюзования.

На сетевом уровне решается задача наилучшей доставки сообщения к пользователю в рамках, одной сети связи. Для этого выбирается маршрут движений сообщения подсети, решается задача внутрисетевой адресации пользователей.

Устройства канального уровня обеспечивают защиту передаваемых сообщений от искажений, которые возникают вследствие изменения параметров сигналов в процессе распространения.

Устройства физического уровня обеспечивают преобразование передаваемого сообщения, в сигналы и восстановление сообщения по принятому сигналу.

Правила, по которым взаимодействуют устройства соседних уровней одной станции, называют интерфейсом.

Правила, по которым взаимодействуют устройства одинаковых уровней у различных станций, называют протоколом.

1.2.5. Модели каналов связи и их математическое описание

Точное математическое описание любого реального канала связи обычно весьма сложное. Вместо этого используют упрощенные математические модели, которые позволяют выявить важнейшие закономерности реального канала.

Рассмотрим наиболее простые и широко используемые связи модели каналов.

Непрерывные каналы. Идеальный канал без помех вносит искажения, связанные с изменением амплитуды и временного положения сигнала и представляет собой линейную цепь с постоянной передаточной функцией, обычно сосредоточенной в ограниченной полосе частот. Допустимы любые входные сигналы, спектр которых лежит в определенной полосе частот ΔF, имеющие ограниченную среднюю мощность $P_{ср}$ [6, 32]. Эта модель используется для описания каналов малой протяженности с закрытым распространением сигналов (кабель, провод, волновод, световод и т. д.).

Канал с гауссовским белым шумом представляет собой идеальный канал, в котором на сигнал $S(t)$ накладывается помеха:

$$U(t) = \mu \cdot S(t-\tau) + n(t). \qquad (1.4)$$

Коэффициент передачи μ и запаздывание τ считаются постоянными и известными в точке приема; $n(t)$ – аддитивная помеха. Такая модель, например, соответствует радиоканалам, с приемо-передающими антеннами работающими и находящимися в пределах прямой видимости.

Гауссовский канал с неопределенной фазой сигнала

Эта модель отличается от предыдущей модели тем, что в ней запаздывание является случайной величиной. Для узкополосных сигналов выражение (1.4) при постоянном μ и случайных τ можно представить в виде [6, 32]:

$$U(t) = \mu \cdot \left[S(t) \cdot \cos\Theta - \widetilde{S}(t) \cdot \sin\Theta\right] + n(t), \qquad (1.5)$$

где $\widetilde{S}(t)$ – преобразование Гильберта от сигнала $S(t)$;

$\Theta = -\omega_0 \tau$ – случайная фаза.

Распределение вероятностей Θ предполагается заданным, чаще всего равномерным на интервале от 0 до 2π. Эта модель удовлетворительно описывает те же каналы, что и предыдущая, если фаза сигнала в них флуктуирует. Флуктуации фазы обычно вызываются небольшими изменениями протяженности канала, свойств среды, в которой проходит сигнал, а также фазовой нестабильностью опорных генераторов.

Дискретно-непрерывные каналы. Дискретно-непрерывный канал имеет дискретный вход и непрерывный выход. Примером такого канала является канал, образованный совокупностью технических средств между выходом кодера канала и входом демодулятора (см. рис. 1.3). Для его описания необходимо знать алфавит входных символов x_i, $i = 1,2,...,m$, вероятности появления символов алфавита $p(x_i)$, полосу пропускания непрерывного канала $F_{НК}$, входящего в рассматриваемый канал и плотности распределения вероятностей (ПРВ) $w\left(U(t)/x_i\right)$ появления сигнала $U(t)$ на выходе канала при условии, что передавался символ x_i.

Зная вероятности $p(x_i)$ и ПРВ $w\left(U(t)/x_i\right)$ по формуле Байеса можно найти апостериорные вероятности передачи символа x_i:

$$p\left(x_i/U(t)\right) = \frac{p(x_i) \cdot w\left(U(t)/x_i\right)}{\sum_{i=1}^{m} p(x_i) \cdot w\left(U(t)/x_i\right)},$$

Решение о переданном символе x_i обычно принимается из условия максимума $p\left(x_i\right)/\mathrm{U(t)}$.

Дискретные каналы. Примером дискретного канала без памяти может служить m-ичный канал. Канал передачи полностью описывается если заданы [20, 21] алфавит источника x_i, $i=1,2,...,m$, вероятности появления символов алфавита $p(x_i)$, скорость передачи символов $V_И$, алфавит получателя y_j, $j=1,2,...,n$ и значения переходных вероятностей $p\left(y_j/x_i\right)$ появления символа y_j при условии передачи символа x_i.

Первые две характеристики определяются свойствами источника сообщений, скорость $V_И$ – полосой пропускания непрерывного канала, входящего в состав дискретного. Объем алфавита выходных символов зависит от алгоритма работы решающей схемы; переходные вероятности $p\left(y_j/x_i\right)$ находятся на основе анализа характеристик непрерывного канала.

Стационарным называется дискретный канал, в котором переходные вероятности $p\left(y_j/x_i\right)$ не зависят от времени.

Дискретный канал называется каналом без памяти, если переходные вероятности $p\left(y_j/x_i\right)$ не зависят от того, какие символы передавались и принимались ранее.

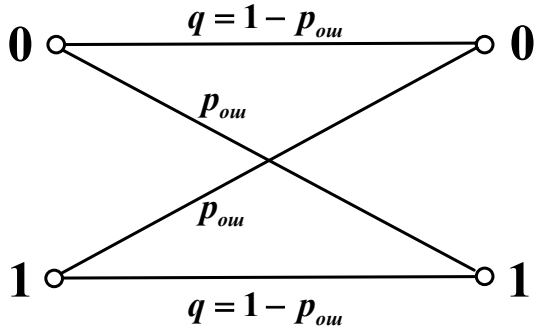


Рис. 1.5. Переходные вероятности в двоичном симметричном канале

В качестве примера рассмотрим двоичный канал (рис. 1.5). В этом случае $m=n=2$, т.е. на входе канала алфавит источника и алфавит получателя состоит из двух символов «0» и «1».

Стационарный двоичный канал называется симметричным, если алфавиты на входе и выходе совпадают. Каждый переданный кодовый символ может быть принят ошибочно с фиксированной вероятностью $p_{ош}$ и правильно с вероятностью $q=1-p_{ош}$.

Необходимо отметить, что в общем случае в дискретном канале объемы алфавитов входных и выходных символов могут не совпадать. Примером может быть канал со стиранием (рис. 1.6). Алфавит на его выходе содержит один добавочный символ по сравнению с алфавитом на входе. Этот добавочный символ (символ стирания «***?***») появляется на выходе канала тогда, когда анализируемый сигнал не удается отождествить ни с одним из передаваемых

символов. Стирание символов при применении соответствующего помехоустойчивого кода позволяет повысить помехоустойчивость.

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

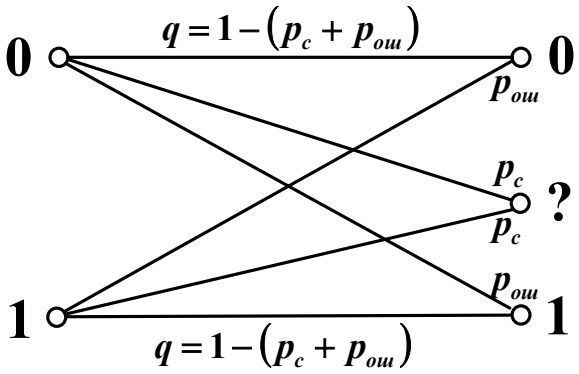


Рис. 1.6. Переходные вероятности в двоичном симметричном канале со стиранием

В постоянном симметричном канале без памяти условная вероятность ошибочного приема $(i+1)$-го, символа если i-й символ принят ошибочно, равна безусловной вероятности ошибки. В канале с памятью она может быть больше или меньше этой величины.

Наиболее простой моделью двоичного канала с памятью является марковская модель, которая задается матрицей переходных вероятностей:

$$p = \begin{bmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{bmatrix},$$

где p_1 – условная вероятность принять $(i+1)$-й символ ошибочно, если i-й принят правильно; $1-p_1$ – условная вероятность принять $(i+1)$-й символ правильно, если i-й принят правильно; p_2 – условная вероятность принять $(i+1)$-й символ ошибочно, если i-й принят ошибочно; $1-p_2$ – условная вероятность принять $(i+1)$-й символ правильно, если i-й принят ошибочно.

Безусловная (средняя) вероятность ошибки в рассматриваемом канале должна удовлетворять уравнению:

$$p\left(x_{i+1}/x_i\right) = p_2 \cdot p_{ош}(x_i) + p_1 \cdot p_{прав}(x_i) \quad \text{или}$$

$$p\left(x_{i+1}/x_i\right) = \frac{p_1}{1+p_1+p_2}.$$

Данная модель имеет достоинство – простоту использования, не всегда адекватно воспроизводит свойства реальных каналов. Большую точность позволяет получить модель Гильберта для дискретного канала с памятью. В такой модели канал может находиться в двух состояниях S_1 и S_2. В состоянии S_1 ошибок не происходит; в состоянии S_2 ошибки возникают независимо с вероятностью p_2.

Также считаются известными вероятности перехода $p\left(S_1/S_2\right)$ из состояния S_1 в S_2 и вероятности перехода $p\left(S_2/S_1\right)$ из состояния S_2 в состояние S_1. В этом случае простую марковскую цепь образует не последовательность ошибок, а последовательность переходов:

$$p=\begin{bmatrix} 1-p\left(S_2/S_1\right) & p\left(S_2/S_1\right) \\ p\left(S_1/S_2\right) & 1-p\left(S_1/S_2\right) \end{bmatrix}.$$

При этом достаточно легко выразить безусловные вероятности нахождения канала в состояниях S_1 и S_2:

$$p(S_1)=\frac{p\left(S_1/S_2\right)}{p\left(S_1/S_2\right)+p\left(S_2/S_1\right)}, \qquad p(S_2)=\frac{p\left(S_2/S_1\right)}{p\left(S_2/S_1\right)+p\left(S_1/S_2\right)}.$$

Безусловная вероятность ошибки в этом случае может быть определена по формуле:

$$p=p_2\cdot p(S_2)=p_2\cdot\frac{p\left(S_2/S_1\right)}{p\left(S_2/S_1\right)+p\left(S_1/S_2\right)}.$$

Наиболее часто при использовании модели Гильберта для двоичного канала полагают $p_2=\frac{1}{2}$, т.е. состояние S_2 рассматривается как полный обрыв связи. Это согласуется с представлением о канале, в котором действуют коммутационные помехи.

Возможен другой подход к построению математических моделей каналов, при котором вся предыстория до некоторого фиксированного момента времени t_0 заменяется заданием некоторого начального состояния цепи. Зная характеристики цепи, начальное состояние и сигнал, действующий только на промежутке от t_0 до t_1, можно определить сигнал на выходе и новое состояние цепи в любой момент времени $t>t_0$.

Состоянием цепи называется минимальное множество величин, в которое входит n элементов, однозначно определяющих поведение цепи в момент времени t. Элементы этого множества называют переменными состояния, которые обычно рассматривают как составляющие компоненты n-мерного вектора. Для любой цепи можно записать два уравнения, позволяющих по состоянию в момент t_0 и сигналу, поступающему на вход, найти выходной сигнал и состояние в мо-

мент $t > t_0$. Эти матричные уравнения называют уравнением состояния и уравнением наблюдения.

1.3. Способы описания сигналов и помех

1.3.1. Сигнал и его математическая модель

Сигналы можно классифицировать по форме, информативности и характеристикам.

Из простых по форме сигналов в электросвязи наибольшее применение находят гармонические и импульсные сигналы.

Гармонический сигнал (рис. 1.7), записывается в виде:

$$S(t) = A_0 \cos(\omega t + \varphi_0), \qquad (1.6)$$

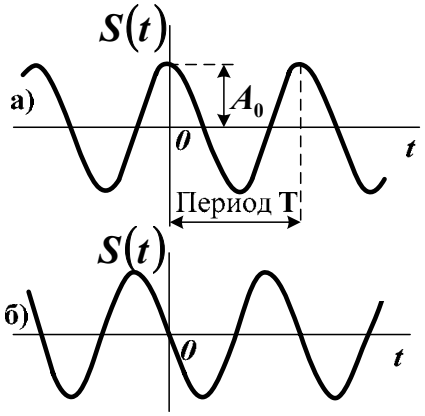


Рис. 1.7. Гармонический сигнал

где A_0 – максимальное значение (амплитуда); $\omega = 2\pi f$ – угловая частота; $f = 1/T$ – циклическая частота; $\varphi_0 = \frac{2\pi\tau_u}{T}$ – начальная фаза.

Для представленных на рис. 1.7. гармонических сигналов значения начальной фазы принимают значения: $\varphi_0 = 0$ (рис. 1.7, а); $\varphi_0 = 90^0$ (рис. 1.7, б).

Импульсными являются сигналы, отличные от нуля в течение ограниченного времени. Эти сигналы существуют лишь в пределах конечного отрезка (t_1, t_2). При этом различают видеоимпульсы (рис. 1.8, а) и радиоимпульсы (рис. 1.8, б). Если $s_B(t)$ - видеоимпульс, то соответствующий ему радиоимпульс описывается выражением: $S_P(t) = s_B(t)\cos(\omega t + \varphi_0)$ (частота ω_0 и начальная фаза φ_0 могут быть произвольными). В радиоимпульсе $s_B(t)$ называется огибающей, а функция $\cos(\omega t + \varphi_0)$ – заполнением. Параметрами видеоимпульса принято считать его амплитуду A_0, длительность τ_u, длительность фронта t_ϕ, длительность спада t_c. Происхождение термина «видеоимпульс» связано с тем, что впервые такие импульсы начали применять для описания сигналов в телевидении.

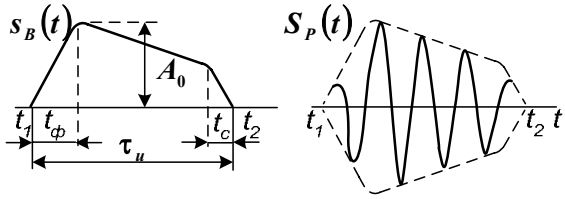


Рис. 1.8. Импульсные сигналы: а) видеоимпульс; б) радиоимпульс

В электросвязи наибольшее применение находят одиночные импульсы

или периодические последовательности импульсов, форма которых приближается к прямоугольной. Для периодической последовательности импульсов, вводится понятие скважности, определяемой как отношение периода к длительности импульса: $S = T/\tau_u$.

По информативности сигналы классифицируются на детерминированные и случайные.

Детерминированным называется сигнал, изменение которого во времени полностью предопределено заранее. Математическим описанием такого сигнала служит детерминированная функция времени $S(t)$. Это означает, что любому моменту времени t_i соответствует определенное значение функции $S(t_i)$. Детерминированные сигналы подразделяются на периодические и непериодические. Для периодического сигнала существует такой интервал времени T (период), что $S(t_i + k \cdot T) = S(t_i),\ k = 0, \pm 1, \pm 2, ...$

Случайным (или нерегулярным) сигналом называется сигнал, изменение которого во времени точно предсказать невозможно. Математическое описание подобных сигналов осуществляется с помощью случайных функций. Для случайной функции ее значение при фиксированном аргументе t_i – случайная величина.

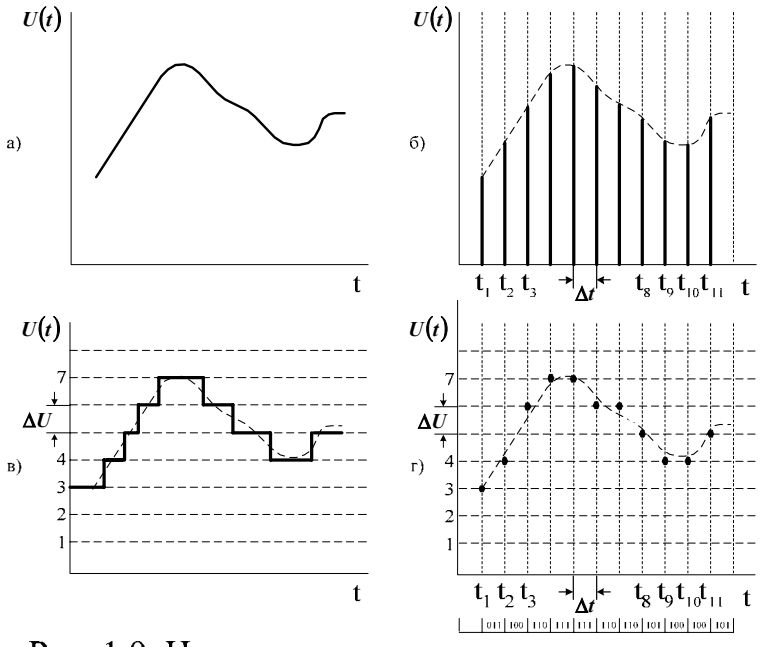


Рис. 1.9. Непрерывные и дискретные сигналы

Сигналы, связанные с передачей сообщений и воздействием помех в системах связи, относятся к разряду случайных сигналов. Такими случайными сигналами являются, например, напряжения или токи, соответствующие речи, музыке, последовательности телеграфных знаков и т.п.

По характеристикам в зависимости от области определения и области возможных значений функции различают следующие виды сигналов (рис. 1.9).

Сигналы первого вида (рис. 1.9, а), называемые непрерывными, задаются на конечном или бесконечном временном интервале и могут принимать любые значения в некотором диапазоне. Примером является сигнал на выходе микро-

фона. Такие сигналы часто называются аналоговыми.

Сигналы второго вида - непрерывные по уровню и дискретные по времени (рис. 1.9, б). Дискретизация по времени обычно выполняется путем взятия отсчетов непрерывной по времени функции $U(t)$ в определенные дискретные моменты времени $t_i, i = 1,2,\ldots$. В результате непрерывную функцию $U(t)$ заменяют совокупностью мгновенных значений $U(t_i), i = 1,2,\ldots$. Дискретизация по времени лежит в основе всех видов импульсной модуляции.

Сигналы третьего вида - дискретные (квантованные) по уровню и непрерывные по времени (рис. 1.9, в). Дискретизация значений непрерывной функции $U(t)$ по уровню называется амплитудным квантованием. В результате квантования непрерывный сигнал заменяется ступенчатой функцией. Шаг квантования ΔU (расстояние между двумя соседними разрешенными уровнями) может быть как постоянным, так и переменным. Его обычно выбирают из условий обеспечения требуемой точности восстановления непрерывного сигнала из квантованного.

Сигналы четвертого вида, называемые дискретными (рис. 1.9, г), задаются в определенные дискретные моменты и принимают определенные дискретные значения. Их можно получить, например, из непрерывных сигналов, осуществляя операции дискретизации по времени и квантования по уровню. Такие сигналы легко представить в цифровой форме, т.е. в виде чисел с конечным числом разрядов. По этой причине их называют цифровыми.

Достоинством цифровых сигналов является возможность применения кодирования для повышения помехоустойчивости.

1.3.2. Энергетические характеристики детерминированного сигнала

При передаче радиосигналов главное внимание уделяется передаче информации, а не энергии. Основными энергетическими характеристиками сигнала на интервале $t_1 \le t \le t_2$, являются мгновенная (текущая) мощность $P(t)$, энергия E и средняя мощность P_{cp}.

Величина $P(t) = S^2(t)$ определяет мгновенную мощность сигнала $S(t)$ выделяемой на единичном сопротивлении.

Энергия сигнала, рассматриваемого на интервале времени $t_1 \le t \le t_2$, является величина

$$E = \int_{t_1}^{t_2} P(t)dt = \int_{t_1}^{t_2} S^2(t)dt, \qquad (1.7)$$

а средняя мощность сигнала в том же интервале определяется по формуле

$$P_{ср} = \frac{1}{T}\int_{t_1}^{t_2} S^2(t)dt\,, \qquad (1.8)$$

где $T = t_2 - t_1$.

Такие характеристики дают определенное представление о детерминированном сигнале и достаточны для решения целого ряда задач теории связи.

1.4. Представление сигналов в виде рядов ортогональных функций

1.4.1. Разложение сигнала в системе функций

В теории и технике связи нередко приходится встречаться с достаточно сложными по своей форме сигналами и помехами. Для решения многих задач весьма полезно уметь представлять сложные сигналы в виде суммы более простых, хорошо изученных элементарных сигналов, описываемых функциями времени $\varphi_k(t)$ [6, 32]:

$$S(t) = \sum_{k=0}^{n} C_k \varphi_k(t)\,. \qquad (1.9)$$

Такое представление сложного сигнала в виде линейной комбинации заданных функций называют разложением.

Совокупность коэффициентов разложения $\{C_k\}$ называют спектром сигнала, а систему функций $\{\varphi_k(t)\}$ базисом разложения.

Произведение $C_k\varphi_k(t)$, где $\varphi_k(t)$ простейший сигнал, а C_k его амплитуда называют спектральной составляющей.

Для того чтобы разложение сигнала (1.9) было выполнимо, базис разложения $\{\varphi_k(t)\}$ должен обладать свойством ортонормированности (ортогональности и нормированности).

Две функции $S(t)$ и $\varphi(t)$ ортогональны на интервале t_1, t_2, если их скалярное произведение (интеграл от произведения)

$$\int_{t_1}^{t_2} S(t)\varphi(t) = 0\,, \qquad (1.10)$$

при том условии, что ни одна из этих функций не равна тождественно нулю при заданных свойствах.

Свойство ортогональности функций обязательно связано с интервалом их определения, т.к. на другом интервале они могут уже быть неортогональны.

Из математики известно, что, если для любой пары функций из ортогональной системы (1.11) выполняется условие

$$\int_{t_1}^{t_2} \varphi_i(t)\varphi_k(t) = \begin{cases} 1, i = k; \\ 0, i \neq k, \end{cases}, \qquad (1.11)$$

то данная система функций система функций – ортонормированна (нормированна к 1).

Особое место при решении многих задач в теории связи занимает ряд Фурье, когда в качестве простых $\varphi_k(t)$ выбирают гармонические колебания.

Представление сигнала $s(t)$ гармоническими функциями имеет следующие преимущества: простое математическое описание; инвариантность к линейным преобразованиям, т. е. если на входе линейной цепи действует гармоническое колебание, то и на выходе ее также будет гармоническое колебание, отличающееся от входного только амплитудой и начальной фазой; как и сигнал, гармонические функции периодические и имеют бесконечную длительность; техника генерирования гармонических функций достаточно проста. Если разложение входного сигнала по ортогональной системе тригонометрических функций известно, то выходной сигнал может быть получен как сумма независимо преобразованных цепью входных гармоник.

1.4.2. Представление сигналов и помех рядом Фурье

Рассмотрим спектральное разложение периодического сигнала. Будем считать, что периодический сигнал определен на бесконечном интервале и может представлен в виде ряда Фурье:

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos(k\Omega_1 t) + b_k \sin(k\Omega_1 t) \right], \qquad (1.12)$$

где $k = 1,2,\ldots$, $\Omega_1 = \dfrac{2\pi}{T} = 2\pi F_1$ - частота основной гармоники, $F_1 = \dfrac{1}{T}$;

$k\Omega_1$ ($k > 1$) – высшие гармоники; a_k (включая a_0) и b_k – коэффициенты Фурье.

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} s(t)\cos(k\Omega_1 t)dt\,, \qquad b_k = \frac{2}{T} \int_{-T/2}^{T/2} s(t)\sin(k\Omega_1 t)dt \qquad (1.13)$$

Постоянную составляющую (среднее значение) $a_0/2$ функции $s(t)$ удобно вычислять по отдельному выражению полученному из a_k при $k = 0$:

$$a_0 = \frac{2}{T} \int_{-T/2}^{T/2} s(t)dt\,, \text{ тогда } \quad \frac{a_0}{2} = \frac{1}{T} \int_{-T/2}^{T/2} s(t)dt\,, \qquad b_0 = 0 \qquad (1.14)$$

Очевидно, что если сигнал представляет собой четную функцию времени $u(t) = -u(t)$, то в тригонометрической записи ряда Фурье (1.14) остаются только косинусоидальные составляющие a_k, так как коэффициенты b_k обращаются в

нуль. Для сигнала $u(t)$ определяемого нечетной функцией времени, наоборот, в нуль обращаются коэффициенты a_k, и ряд содержит синусоидальные составляющие b_k

Часто выражение (1.15) удобно представлять в другой, эквивалентной форме ряда Фурье:

$$s(t) = A_0 + \sum_{k=1}^{\infty}\left[A_k \cos(k\Omega_1 t + \Psi_k)\right], \qquad (1.15)$$

где $A_0 = \frac{a_0}{2}$, $A_k = \sqrt{a_k^2 + b_k^2}$ - амплитуда, $\Psi_k = -arctg\frac{b_k}{a_k}$ - начальная фаза k - ой гармоники.

На рис. 1.10 приведены графики, иллюстрирующие представление периодической последовательности прямоугольных импульсов $s(t)$ конечным числом слагаемых ($k = 5$) ряда Фурье.

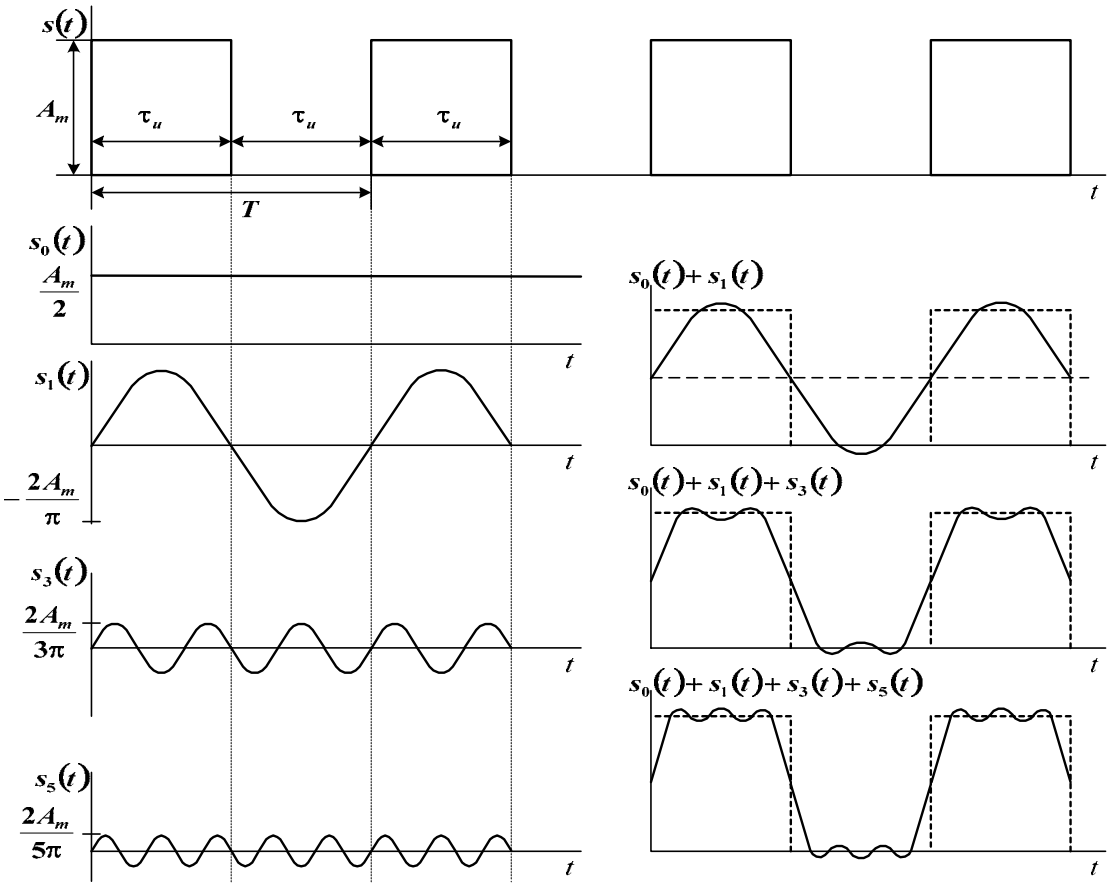


Рис. 1.10. Аппроксимация прямоугольных импульсов суммой гармоник

Для функции $s(t)$ (рис.1.10) разложение имеет вид

$$s(t) = \frac{A_m}{2} + \frac{2A_m}{\pi}\sin\Omega_1 t + \frac{2A_m}{3\pi}\sin 3\Omega_1 t + \frac{2A_m}{5\pi}\sin 5\Omega_1 t + \ldots. \qquad (1.16)$$

Периодическая последовательность прямоугольных импульсов $s(t)$ представляется как результат сложения постоянной составляющей $\frac{A_m}{2}$ и синусои-

дальных сигналов с частотами $F_1, 3F_1, 5F_1, \ldots$, причем период синусоиды с частотой F_1 совпадает с периодом последовательности импульсов $s(t)$. Для удобства F_1 можно представить в виде $F_1 = \dfrac{\Omega_1}{2\pi} = \dfrac{1}{T}$.

Совокупность всех гармонических составляющих разложения функции в ряд Фурье называется спектром функции.

Наличие отдельных гармонических составляющих спектра и величины из амплитуд можно наглядно показать с помощью спектральной диаграммы (рис.1.11), у которой горизонтальная ось служит осью частот, а вертикальная – осью амплитуд.

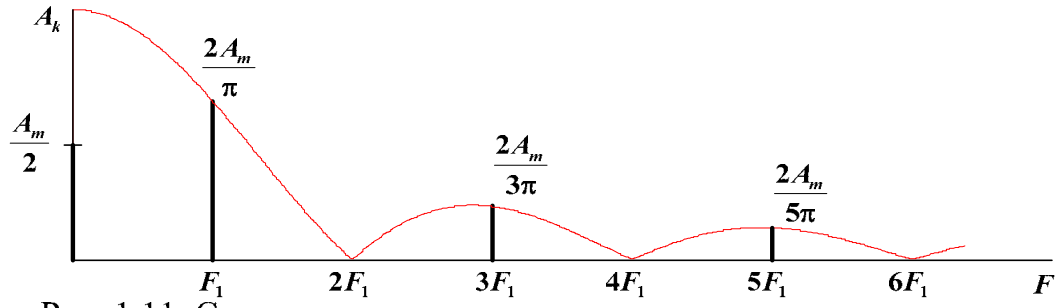


Рис. 1.11. Спектр амплитуд прямоугольных импульсов

В точках оси частот $F_1, 3F_1, 5F_1, \ldots$ отображаются амплитуды соответствующих гармонических составляющих разложения функции.

Легко заметить, что график суммы двух первых слагаемых разложения (1.16) воспроизводит форму графика функции $s(t)$ очень грубо, только в основных чертах. Учет третьего слагаемого существенно улучшает совпадение суммы с функцией $s(t)$. Таким образом, с увеличением числа учитываемых гармоник точность представления $s(t)$ возрастает.

На практике спектральные диаграммы называют более кратко – амплитудный спектр, фазовый спектр. Чаще всего интересуются амплитудным спектром (рис. 1.11). По нему можно оценить процентное содержание гармоник, наличие и уровни отдельных гармонических составляющих спектра.

Пример 1.1. Разложим в ряд Фурье периодическую последовательность прямоугольных видеоимпульсов с известными параметрами (A_m, T, τ_u) (рис. 1.12), четную относительно точки $t = 0$:

$$s(t) = \begin{cases} A_1, & -\tau_u/2 \le \tau_u < \tau_u/2 \\ A_0, & \tau_u/2 < \tau_u \le T - \tau_u/2 \end{cases}.$$

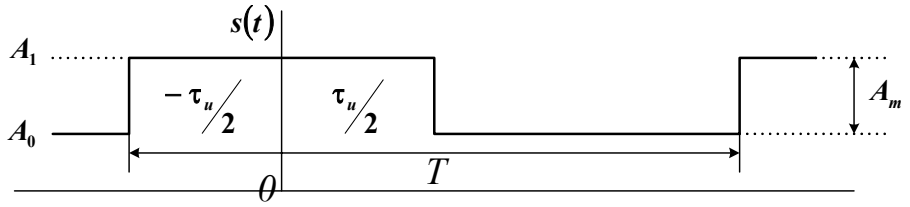


Рис.1.12. Временное представление периодической последовательности прямоугольных импульсов.

Воспользуемся для представления этого сигнала формой записи ряда Фурье в виде (1.12). Для спектрального представления последовательности прямоугольных импульсов начало отсчета целесообразно брать в середине импульса. Действительно, в этом случае и в разложении останутся только косинусоидальные составляющие, так как интегралы от нечетных функций за период равны нулю b_k=0.

По формулам (1.14) находим коэффициенты:

$$\frac{a_0}{2}=\frac{A_m\tau_u}{T}=\frac{A_m}{S},\ a_k=\frac{2A_m}{T}\int_{-t_u/2}^{t_u/2}\cos(k\Omega_1 t)dt=\frac{2A_m}{k\pi}\sin\frac{k\Omega_1\tau_u}{2},$$

позволяющие записать ряд Фурье:

$$s(t)=\frac{A_m}{S}\left[1+2\sum_{k=1}^{\infty}\frac{\sin\left(k\pi/S\right)}{k\pi/S}\cos(k\Omega_1 t)\right],$$

где $S=T/\tau_u$ - скважность импульсной последовательности.

Для построения спектральных диаграмм при конкретных числовых данных полагаем $k=0,1,2,3,...$ и вычисляем коэффициенты гармоник. Результаты расчета первых восьми составляющих спектра при $A_m=2B$, $T=20мс$, $S=T/\tau_u=2$ и 8 сведены в табл. 1.1 и построены спектральные диаграммы на рис.1.13.

Таблица 1.1. Амплитуды спектральных составляющих для периодической последовательности прямоугольных импульсов

k		0	1	2	3	4	5	6	7	8
k, Гц		0	50	100	150	200	250	300	350	400
a_k, В	S=2	1	1,27	0	0,42	0	0,25	0	0,18	0
	S=8	0,25	0,48	0,45	0,39	0,32	0,23	0,15	0,07	0

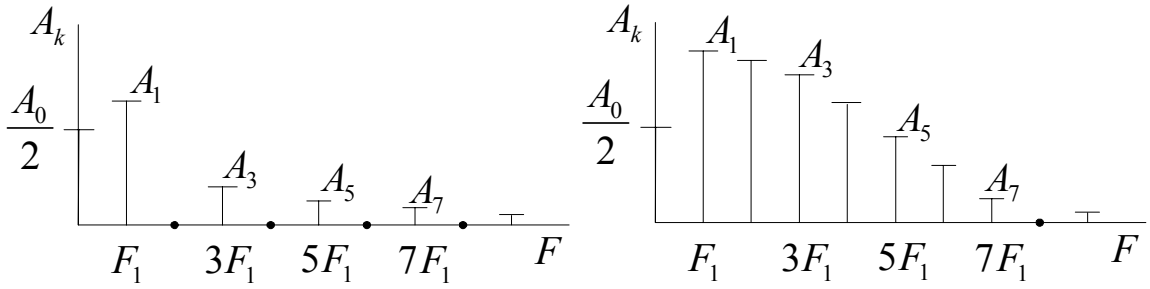


Рис. 1.13. Спектральные диаграммы периодической последовательности импульсов: а) - при скважности ***S***=2; б) - при скважности ***S***=8

Из приведенного примера следует, что с увеличением скважности увеличивается число спектральных составляющих и уменьшаются их амплитуды.

Выбор количества спектральных составляющих зависит от формы сигнала и точности его представления рядом Фурье. Плавное изменение формы сигнала потребует меньше числа гармоник при той же точности представления, чем для скачкообразного сигнала. Для приближенного представления прямоугольных импульсов на практике обычно считают, что достаточно трех - пяти гармоник.

1.4.3. Применение преобразования Фурье для непериодических сигналов

Для спектрального представления непериодических (импульсных) сигналов $s(t)$, заданных на конечном интервале $(t_1,\ t_2)$ (рис. 1.14), непосредственно воспользоваться рядом Фурье нельзя. Для гармонического разложения сигнала мысленно дополняют его такими же импульсными сигналами до периодического с некоторым интервалом T (рис. 1.14).

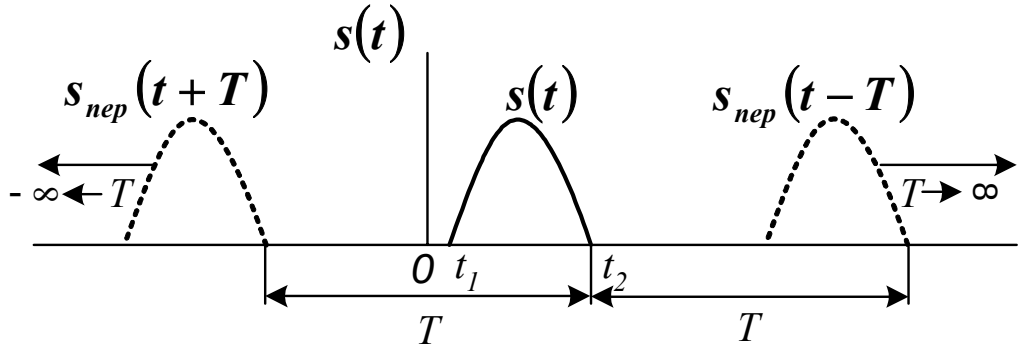


Рис. 1.14. Импульсный сигнал ***s*(*t*)** и его периодическое продолжение $s_{пер}(t+kT)$

Для того чтобы вне искусственно введенного интервала исходный сигнал был равен нулю, необходимо увеличить период повторения импульсов.

В пределе, при увеличении периода $T \to \infty$ все импульсы уйдут вправо и влево в бесконечность и периодическая последовательность вновь станет одиночным импульсом.

Для вычисления спектра удобна симметричная комплексная форма ряда Фурье, но в нем вместо суммы будет интеграл с бесконечными пределами.

$$s(t) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} \overset{*}{S}(\Omega) e^{j\Omega t} d\Omega, \qquad (1.17)$$

$$\overset{*}{S}(\Omega) = \int_{-\infty}^{\infty} s(t) e^{-j\Omega t} dt. \qquad (1.18)$$

При таком предельном переходе основная частота сигнала $\Omega_1 = {2\pi}/{T}$ стремится к нулю, бесконечно увеличивается число спектральных составляющих, частоты соседних гармоник $k\Omega_1$ и $(k+1)\Omega_1$ становятся неразличимыми, а спектр будет сплошным.

Формулы (1.17) и (1.18) называются соответственно обратным и прямым преобразованиями Фурье. Они дают взаимосвязь между сигналом $s(t)$ и его комплексной спектральной плотностью $\overset{*}{S}(\Omega)$.

Представим спектральную плотность в показательной форме:

$$\overset{*}{S}(\Omega) = S(\Omega) e^{-j\varphi(\Omega)},$$

где $S(\Omega)$ – модуль $\overset{*}{S}(\Omega)$, который называют спектральной плотностью амплитуд, или амплитудным спектром; $\varphi(\Omega)$ – аргумент $\overset{*}{S}(\Omega)$, называемый фазовым спектром сигнала. По определению, модуль $S(\Omega)$ – четная функция частоты, а аргумент $\varphi(\Omega)$ – нечетная функция.

Пример 1.2. Найти спектральную плотность прямоугольного видеоимпульса $s_в(t)$ четного относительно $t = 0$, длительностью τ_u и с амплитудой A_m (1.15, а).

Запишем аналитическое выражение для заданного видеоимпульса:

$$s(t) = \begin{cases} A_m, & -{\tau_u}/{2} \le t < {\tau_u}/{2} \\ 0, & |t| > {\tau_u}/{2} \end{cases}.$$

Тогда спектральную плотность импульса находим по формуле (1.18):

$$\overset{*}{s}(\Omega) = A_m \cdot \int_{-t_u}^{t_u} e^{-j\Omega t} dt = \frac{A_m}{-j\Omega} \left[\exp\left(-\frac{j\Omega\tau_u}{2} \right) - \exp\left(\frac{j\Omega\tau_u}{2} \right) \right].$$

Это выражение с учетом формулы Эйлера $\sin\alpha = \dfrac{\left(e^{j\alpha} - e^{-j\alpha}\right)}{2j}$ можно перепи-

сать в виде:

$$\overset{*}{s}(\Omega) = A_m \tau_u \frac{\sin\left(\Omega\tau_u / 2\right)}{\left(\Omega\tau_u / 2\right)}. \qquad (1.19)$$

Отсюда следует, что спектральная плотность прямоугольного видеоимпульса, четного относительно t=0, вещественная. Фазовый спектр

$$\psi(\Omega) = \begin{cases} 0, & при\ \overset{*}{s} > 0 \\ -\pi, & при\ \overset{*}{s} < 0 \end{cases}. \qquad (1.20)$$

Рассчитанные по формулам (1.19) и (1.20) амплитудный и фазовый спектры прямоугольного видеоимпульса изображены на рис. 1.15.

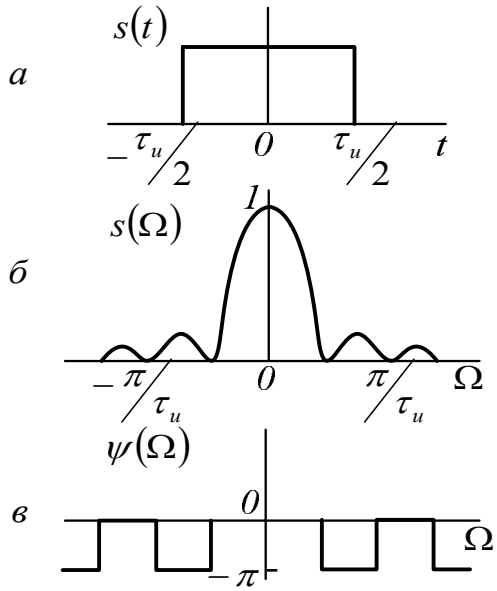


Рис. 1.15. Прямоугольный видеоимпульс и его амплитудный (б) и фазовый (в) спектры

Следует отметить, что нули амплитудного спектра определяются длительностью импульса. При удлинении импульса расстояние между нулями $s(\Omega)$ сокращается, что равносильно сужению спектра. При укорочении (сжатии) импульса, наоборот, расстояние между нулями функции $\overset{*}{s}(\Omega)$ увеличивается, спектр расширяется.

Спектральный метод является одним из основных при расчетах линейных электрических цепей. Знание спектра сигнала позволяет правильно рассчитать и установить полосу пропускания усилителей, фильтров и других элементов каналов связи. Это необходимо для осуществления неискаженной передачи сигнала для обеспечения разделения сигналов и ослабления помех.

1.5. Теорема Котельникова

В 1933 году В.А. Котельниковым доказана теорема отсчетов [6, 32], имеющая важное значение в теории связи: непрерывный сигнал $s(t)$ с ограниченным спектром можно точно восстановить (интерполировать) по его отсчетам $s(k\Delta t)$, взятым через интервалы $\Delta t = \frac{1}{(2F)}$, где F – верхняя частота спектра сигнала.

В соответствии с этой теоремой сигнал $s(t)$ можно представить рядом Котельникова [6, 32]:

$$s(t)=\sum_{k=-\infty}^{\infty} s\left(\frac{k}{2F}\right)\frac{\sin 2\pi F\left[t-\frac{k}{2F}\right]}{2\pi F\left[t-\frac{k}{2F}\right]}. \qquad (1.21)$$

Таким образом, сигнал $s(t)$, можно абсолютно точно представить с помощью последовательности отсчетов $s\left(\frac{k}{2F}\right)$, заданных в дискретных точках $\frac{k}{2F}$ (рис.1.16).

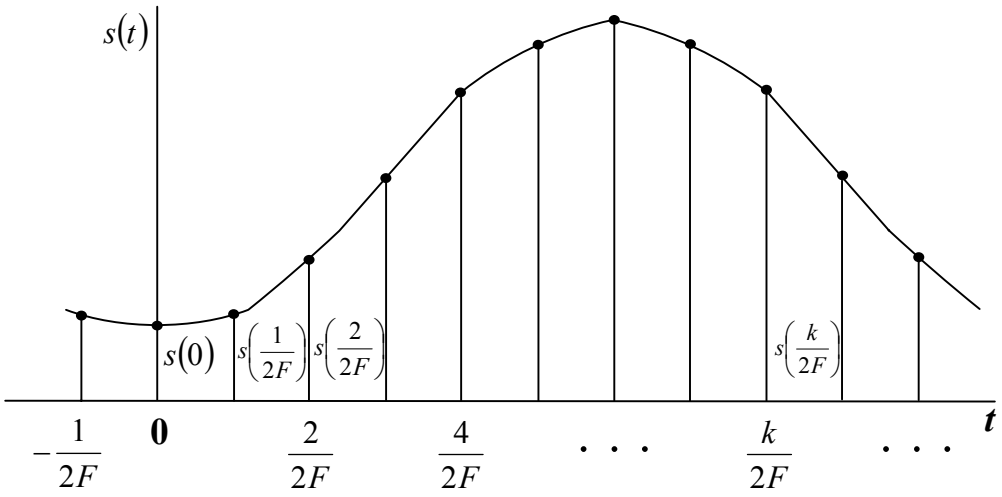


Рис. 1.16. Сигнал и его отсчеты

Функции

$$\psi(t)=\frac{\sin 2\pi F\left[t-\frac{k}{2F}\right]}{2\pi F\left[t-\frac{k}{2F}\right]} \qquad (1.22)$$

образуют ортогональный базис в пространстве сигналов, характеризующихся ограниченным спектром:

$$\Phi(f)=0 \text{ при } |f|>F. \qquad (1.23)$$

Обычно для реальных сигналов можно указать диапазон частот, в пределах которого сосредоточена основная часть его энергии и которым определяется ширина спектра сигнала. В ряде случаев спектр сознательно сокращают. Это обусловлено тем, что аппаратура и линия связи должны иметь минимальную полосу частот. Сокращение спектра выполняют, исходя из допустимых искажений сигнала. Например, при телефонной связи хорошая разборчивость речи и узнаваемость абонента обеспечиваются при передаче сигналов в полосе час-

тот $\Delta F = 0{,}3...3{,}4\,[\text{кГц}]$. Увеличение ΔF приводит к неоправданному усложнению аппаратуры и повышению затрат. Для передачи телевизионного изображения при стандарте в 625 строк полоса частот, занимаемая сигналом, составляет около 6 МГц.

Из вышесказанного следует, что процессы с ограниченными спектрами могут служить адекватными математическими моделями многих реальных сигналов.

Функция вида $\dfrac{\sin 2\pi F\left[t-\dfrac{k}{2F}\right]}{2\pi F\left[t-\dfrac{k}{2F}\right]}$ называется функцией отсчетов (рис.1.17).

Она характеризуется следующими свойствами. Если $k=0$, функция отсчетов имеет максимальное значение при $t=0$, а в моменты времени $t=\dfrac{i}{2F}$ $(i=1,2,...)$ она обращается в нуль; ширина главного лепестка функции отсчетов на нулевом уровне равна $\dfrac{1}{F}$, поэтому минимальная длительность импульса, который может существовать на выходе линейной системы с полосой пропускания F, равна $\dfrac{1}{F}$; функции отсчетов ортогональны на бесконечном интервале времени.

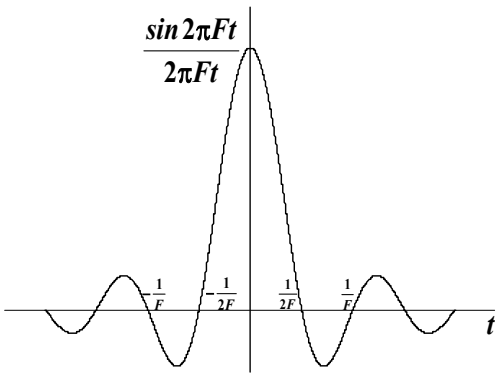


Рис. 1.17. Функция отсчётов

На основании теоремы Котельникова может быть предложен следующий способ дискретной передачи непрерывных сигналов:

Для передачи непрерывного сигнала $s(t)$ по каналу связи с полосой пропускания F определим мгновенные значения сигнала $s(t)$ в дискретные моменты времени $t_k=\dfrac{1}{2F}$, $(k=0,1,2,...)$. После этого передадим эти значения по каналу связи каким - либо из возможных способов и восстановим на приемной стороне переданные отсчеты. Для преобразования потока импульсных отсчетов в непрерывную функцию пропустим их через идеальный ФНЧ с граничной частотой F.

Можно показать, что энергия сигнала находится по формуле [6, 32]:

$$E=\int_{-\infty}^{\infty}s^2(t)dt=\frac{1}{2F}\sum_{k=-\infty}^{\infty}s^2\left(\frac{k}{2F}\right). \qquad (1.24)$$

Для сигнала, ограниченного во времени, выражение (1.24) преобразуется к виду:

$$E = \int_{1}^{2FT} s^2(t)dt = \frac{1}{2F}\sum_{k=1}^{2FT} s^2\left(\frac{k}{2F}\right). \qquad (1.25)$$

Выражение (1.25) широко применяется в теории помехоустойчивого приема сигналов, но является приближенным, т.к. сигналы не могут быть одновременно ограничены по частоте и времени.

1.6. Пространство сигналов

Для решения ряда задач теории связи целесообразно сигналы представить векторами или точками в некотором функциональном пространстве – пространстве сигналов.

1.6.1. Линейное пространство

Пространством сигналов называется множество сигналов, обладающих общим свойством и отличающихся друг от друга, каким либо параметром (расстоянием).

При анализе ансамблей сигналов, их преобразований в системах связи и методов приема в условиях воздействия помех широко используют понятия линейного пространства. Приведем необходимые и достаточные условия того, чтобы пространство линейным [6, 32]:

сумма любых двух элементов пространства, является элементом пространства, и выполняются равенства: $x + y = y + x$ и $x + (y + z) = (x + y) + z$;

существует нулевой элемент пространства, такой, что $x + 0 = x$ для всех x;

для любого элемента пространства существует противоположный элемент, такой, что $x + (-x) = 0$;

новый элемент пространства можно получить, умножив элемент пространства на скаляр, при этом должны выполняться следующие равенства: $1 \cdot x = x$, $\alpha(\beta x) = (\alpha\beta)x$, $\alpha(x + y) = \alpha x + \alpha y$, $(\alpha + \beta)x = \alpha x + \beta x$.

К линейным пространствам можно отнести совокупность векторов в трехмерном пространстве, совокупность сигналов, имеющих конечную энергию, и т.п.

1.6.2. Представление сигнала в многомерном пространстве

Если обозначить точки $a = (a_1, a_2, a_3)$ $b = (b_1, b_2, b_3)$ для трехмерного пространства, то расстояние между ними (рис.1.18):

$$d(a,b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}. \qquad (1.26)$$

По аналогии с трехмерным геометрическим пространством можно определить расстояние между элементами *n* - мерного пространства:

$$d(a,b)=\sqrt{\sum_{k=1}^{n}(a_k-b_k)^2}\ . \qquad (1.27)$$

Рассмотрим два сигнала $s_1(t)$ и $s_2(t)$, имеющих ограниченную полосу частот F. В соответствии с теоремой Котельникова эти сигналы могут быть представлены разложениям по ортогональным функциям $\left\{s_1\left(\frac{k}{2F}\right)\right\}$ и $\left\{s_2\left(\frac{k}{2F}\right)\right\}$, количество отсчетов теоретически бесконечно. На основании этого сигналы можно представить точками в бесконечномерном пространстве.

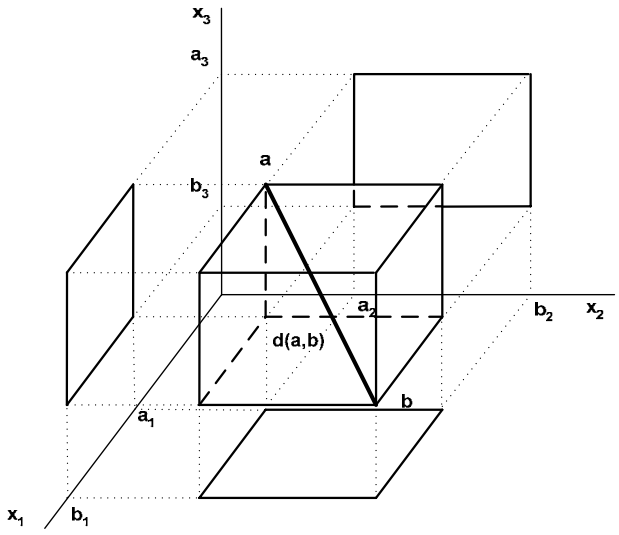


Рис. 1.18. Геометрическое представление элементов трехмерного пространства

При этом расстояние между двумя сигналами

$$d(s_1,s_2)=\sqrt{\sum_{k=-\infty}^{\infty}\left[s_1\left(\frac{k}{2F}\right)-s_2\left(\frac{k}{2F}\right)\right]^2}\ . \qquad (1.28)$$

Обозначим через $E_1=\int_0^{\infty}s_1^2(t)dt$ и $E_2=\int_0^{\infty}s_2^2(t)dt$ – энергии первого и второго сигналов, а через $E_{12}=\int_0^{\infty}s_1(t)\cdot s_2(t)dt$ взаимную энергию сигналов. Преобразуем выражение (1.28):

$$d(s_1,s_2)=\sqrt{2F\cdot\frac{1}{2F}\cdot\sum_{k=-\infty}^{\infty}\left[s_1\left(\frac{k}{2F}\right)-s_2\left(\frac{k}{2F}\right)\right]^2}=$$
$$=\sqrt{2F\cdot\int_0^{\infty}[s_1(t)-s_2(t)]^2dt}=\sqrt{2F\left(E_1-2E_{1,2}+E_2\right)}=\sqrt{2FE_0}\ , \qquad (1.29)$$

где $E_0=E_1+E_2-2E_{1,2}$ – эквивалентная энергия сигналов.

Заметим, что расстояние между сигналами в бесконечномерном пространстве определяется шириной полосы частот F, занимаемых сигналами и эквивалентной энергией E_0. Величина $d(s_1,s_2)$ показывает удаленность двух сигналов друг от друга, а следовательно, - их различимость для системы связи. В общем случае расстояние определяет помехоустойчивость двоичных каналов связи.

1.7. Сигналы как случайные процессы

Случайные сигналы и помехи относятся к случайным явлениям природы, изучением основных закономерностей которых занимается теория вероятностей. Все случайные явления, изучаемые в теории вероятностей, можно разбить на три типа: случайные события, случайные величины и случайные процессы. Каждый из этих типов случайных явлений имеет свои особенности и характеристики.

Для математического описания сигналов и помех необходимо решить две задачи. К какому типу случайных явлений отнести случайный сигнал (помеху) в конкретной ситуации и как определить необходимые вероятностные характеристики?

Напомним важнейшие понятия теории вероятностей, необходимые для описания случайных сигналов и помех.

Случайные события. Случайное событие – это всякий факт, который в результате опыта может произойти или не произойти. Это и передача текста без ошибок, и работа канала связи без повреждений не менее T часов, и превышение помехой заданного уровня и т. д. Обозначаются случайные события начальными прописными буквами латинского алфавита: A, B, C.

Напомним, что вероятность $P(A_1 \cdot A_2 \cdot \ldots \cdot A_n)$ произведения n случайных событий $A_1, A_2, \ldots, A_n$ равна произведению условных вероятностей этих событий:

$$P(A_1 \cdot A_2 \ldots A_n) = P(A_1)P(A_2 / A_1)P(A_3 / A_1A_2)\ldots P(A_n / A_1A_2\ldots A_{n-1}). \qquad (1.30)$$

Для n независимых событий условные вероятности $P(A_i / A_1A_2\ldots A_{i-1})$ появления события A_i равны безусловным $P(A_i), i = 2,3,\ldots,n$. Поэтому вероятность произведения n независимых событий определяется по формуле $P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$.

Сумма $A + B$ двух совместных событий может быть представлена как сумма $A\overline{B} + \overline{A}B + AB$ трех несовместных. С учетом очевидных соотношений $P\left(A\overline{B} + \overline{A}B + AB\right) = P\left(A\overline{B}\right) + P\left(\overline{A}B\right) + P(AB)$, $A = AB + A\overline{B}$ и $B = AB + \overline{A}B$ можно найти формулу для вероятности суммы двух совместных событий в виде

$$P(A + B) = P(A) + P(B) - P(AB). \qquad (1.31)$$

Однако уже для суммы трех совместных событий A, B и C подобная формула будет содержать семь слагаемых. Поэтому для вычисления вероятно-

сти $P(C)$ суммы, $C=\sum_{i=1}^{n} A_i$ большого числа слагаемых обычно переходят к противоположному событию $\overline{C}=P\prod_{i=1}^{n}\overline{A_i}$:

$$P\left(\sum_{i=1}^{n} A_i\right)=1-P\left(\prod_{i=1}^{n}\overline{A_i}\right). \qquad (1.32)$$

Эта формула упрощается, если события $A_1, A_2,..., A_n$ совместны, но независимы. Тогда

$$P\left(\sum_{i=1}^{n} A_i\right)=1-\prod_{i=1}^{n}(1-P(A_i)). \qquad (1.33)$$

Приведенное выражение (1.33) часто встречается в расчетах надежности системы параллельно соединенных устройств. Действительно, система с параллельным соединением элементов работает безотказно, когда работает хотя бы один из ее элементов (устройств). При независимом функционировании каждого из элементов $A_1, A_2,...A_n$ с вероятностями безотказной работы $P(A_1), P(A_2),..., P(A_n)$ соответственно по формуле (1.33) находим вероятность безотказной работы всей системы.

Предположим теперь, что событие A может произойти одновременно с одним из несовместных событий (гипотез) $H_1, H_2,..., H_n$, образующих полную группу. Событиями $H_1, H_2,..., H_n$ часто являются взаимоисключающие предположения об условиях проведения эксперимента, результатом которого может быть случайное событие A. Например, две гипотезы H_1 и H_2 можно связать с передачей сообщений «0» или «1» по каналу связи с помехами, а случайное событие A с превышением выходным напряжением приемника порогового уровня.

В подобных схемах заданы вероятности гипотез $P(H_i)$ и условные вероятности появления события $P(A/H_i)$, когда справедливы предположения H_i, $i=1,2,...,n$. Безусловную вероятность события $P(A)$ можно найти с помощью формулы полной вероятности:

$$P(A)=\sum_{i=1}^{n}P(H_i)P(A/H_i). \qquad (1.34)$$

Если стало известно, что в результате испытания событие A произошло, то условная вероятность гипотезы H_i (апостериорная вероятность гипотезы H_i) определяется по формуле Байеса:

$$P(H_i/A)=P(H_1)\frac{P(A/H_i)}{P(A)}. \qquad (1.35)$$

Возможность переоценки вероятностей гипотез после проведения эксперимента может быть показана на примере приема двоичных сигналов. Допустим, что вероятности передачи сигналов «0» и «1» одинаковы: $P(A/H_1)=0.1,\quad P(A/H_2)=0.8$, а вероятности превышения порогового уровня при передаче сигналов «0» и «1» значительно отличаются, скажем, $P(H_1)=P(H_2)=0.5$. В результате наблюдения установлено превышение порогового уровня (т.е. произошло событие A). Очевидно, предпочтение после получения такой информации следует отдать гипотезе H_2 (передача сигнала «1»). Количественно охарактеризовать это "предпочтение" позволяет формула Байеса. Действительно, расчет по формуле (1.35) с учетом (1.34) дает следующий результат: $P(H_2/A)=0.89, P(H_1/A)=0.11$.

Большую роль при анализе цифровых систем обработки сигналов играет следующая схема. Пусть n раз при постоянных условиях повторяется один и тот же опыт, с которым связано случайное событие A, имеющее вероятность p. При этом предполагается, что исход каждого опыта не зависит от результатов других опытов. Тогда вероятность $P_n(k)$ того, что в этой последовательности n опытов событие A появится ровно k раз (безразлично в каком порядке) находится по формуле Бернулли:

$$P_n(k)=C_n^k p^k q^{n-k}, k=0,1,...,n, \qquad (1.36)$$

где $q=1-p,\ C_n^k=n!/k!(n-k)!$. Правая часть формулы имеет вид общего члена разложения бинома Ньютона: $(p+q)^n=\sum C_n^k p^k q^{n-k}$. Поэтому совокупность чисел $P_n(k),\quad k=0,1...n$, называют биноминальным распределением вероятностей.

Так как числа $P_n(k),\quad k=0,1...n$, являются вероятностями попарно несовместных событий, то вероятность $P_n(m_1 \le k \le m_2)$ того, что число появления события A в n опытах будет заключено в пределах от m_1 до m_2, определяется с помощью суммирования:

$$P_n(m_1 \le k \le m_2) = \sum_{k=m_1}^{m_2} P_n(k) = \sum_{k=m_1}^{m_2} C_n^k p^k q^{n-k} . \qquad (1.37)$$

На практике часто встречаются задачи, когда число испытаний n велико и вычисления по формуле Бернулли затруднены. Для этих случаев применяются приближенные методы расчета. При малых $p \to 0$ и ограниченных значениях $\lambda = np$ используется формула Пуассона:

$$P_n(k) = C_n^k p^k q^{n-k} \cong (\lambda^k / k!)\exp(-\lambda). \qquad (1.38)$$

По этой формуле для любых $n >> 1$ легко выполняются расчеты с помощью таблиц распределения Пуассона [42] или на ЭВМ.

Если p фиксировано, а n и k стремятся к бесконечности при ограниченном отношении $(k-np)/\sqrt{npq}$, то может быть использована асимптотическая формула Лапласа:

$$P_n(k) \cong \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k-np)^2}{2npq}\right). \qquad (1.39)$$

Когда p не слишком близко к нулю или единице, формула (1.39) может быть достаточно точна уже при n порядка нескольких десятков. Сумма вероятностей (1.37) при этом хорошо аппроксимируется следующим выражением:

$$P_n(m_1 \le k \le m_2) \cong \Phi_0\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi_0\left(\frac{m_1 - np}{\sqrt{npq}}\right), \qquad (1.40)$$

где $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-0,5t^2} dt$ – функция Лапласа [40].

Следует подчеркнуть, что применение приближенных асимптотических соотношений всегда должно сопровождаться контролем величины погрешности. Для этого могут использоваться точные формулы, специальные аналитические методы [40] или результаты экспериментов.

Случайные величины. Величина, которая принимает то или иное значение, заранее неизвестно какое именно, называется случайной. Число ошибок в тексте, число занятых каналов многоканальной системы связи, уровень помехи в канале, мощность сигнала на выходе линии связи – это все примеры случайных величин (СВ).

Будем обозначать СВ прописными буквами латинского алфавита X, Y, Z, а значения, которые они принимают, – строчными буквами x, y, z.

СВ делятся на дискретные и непрерывные. Дискретная случайная величина X может принимать только конечное множество значений $x_1, x_2, ..., x_n$, непрерывная – любые значения x из некоторого интервала, даже бесконечного.

Для математического описания СВ вводятся следующие неслучайные основные статистические характеристики [5, 13, 32, 39]:

Будем рассматривать множество всех случайных исходов, возможных при данном испытании. Предположим, что каждому исходу ω этого испытания соответствует число X. Тогда множество исходов отображается в некоторое числовое множество. Такое отображение, т.е. числовая функция $X(\omega)$, построенная на множестве исходов эксперимента, называется случайной величиной (СВ). Примерами СВ могут быть число единиц в последовательности n двоичных символов, значение напряжения на выходе приемника в фиксированный момент времени и т.д.

Если число n возможных исходов $x_1, x_2, ..., x_n$ конечно или счетно, то СВ X называется дискретной. Дискретная СВ может быть описана с помощью задания всех вероятностей $p_i, i = 1, 2, ..., n$, с которыми СВ принимает значения $x_1, x_2, ..., x_n$, т.е. $p_i = P(X = x_i),\ i = 1, 2, ..., n$. Сумма этих вероятностей равна единице. Вместо набора $\{p_i\}_{i=1}^{n}$ вероятностей свойства СВ могут быть заданы с помощью функции распределения

$$F(x) = P(X < x). \tag{1.41}$$

Как следует из определения, $F(-\infty) = 0,\quad F(\infty) = 1,\quad P(a \le x \langle b) = F(b) - F(a)$. Кроме того, $F(x)$ является неубывающей функцией. Для дискретных СВ эта

функция имеет ступенчатый вид, причем каждая «ступенька» величиной p_i расположена в точке с абсциссой x_i.

Другим важным классом является СВ, для которых функция распределения $F(x)$ непрерывна. Если $F(x)$ дифференцируема, то ее производная

$$w(x) = dF(x)/dx \tag{1.42}$$

называется плотностью распределения вероятностей (ПРВ) непрерывной случайной величины. Поскольку

$$dF/dx = \lim_{\Delta x \to 0}(F(x+\Delta x) - F(x))/\Delta x = \lim_{\Delta x \to 0} P(x \le X \langle x+\Delta x)/\Delta x,$$

то ПРВ можно рассматривать как предел отношения вероятности попадания случайной величины на отрезок $(x; x+\Delta x)$ к длине Δx этого отрезка при $\Delta x \to 0$. Очевидно, $P(a \le X \le b) = \int_a^b w(x)dx$, т.е. вероятность попадания СВ на отрезок $[a, b]$ численно равно площади под графиком ПРВ. В отличие от дискретных непрерывные СВ принимают несчетное множество значений. Вероятность того, что непрерывная СВ примет любое конкретное значение, например a, равна нулю.

Важнейшими числовыми характеристиками СВ являются математическое ожидание

$$m_x = M\{X\} = \int_{-\infty}^{\infty} x w(x) dx, \tag{1.43}$$

дисперсия

$$D_x = M\{(X - m_x)^2\} = \int_{-\infty}^{\infty} (x - m_x)^2 w(x) dx \tag{1.44}$$

и среднее квадратическое отклонение $\sigma_x = \sqrt{D_x}$. Обобщением числовых характеристик являются начальные моменты распределения СВ

$$m_k = M\{X^k\} = \int_{-\infty}^{\infty} x^k w(x) dx \tag{1.45}$$

и центральные моменты

$$\mu_k = M\{(X - m_x)^k\} = \int_{-\infty}^{\infty} (x - m_x)^k w(x) dx. \tag{1.46}$$

Напомним, что $m_1 = m_x, \mu_1 = 0, \mu_2 = D_x$, а числа μ_3/σ_x^3 и $\mu_4/\sigma_x^4 - 3$ называются коэффициентами асимметрии и эксцесса. Ряд часто встречающихся в статистической радиотехнике распределений и соответствующих числовых характеристик СВ приведены в табл. 1.2.

Таблица 1.2

Название закона распределения	Плотность распределения вероятностей $w(x)$	Моменты
1	2	3
Нормальный	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$	$m_1 = a, \mu_2 = \sigma^2,$ $\mu_3 = 0, \mu_4 = 3\sigma^4$
Релея	$\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right), x \geq 0$	$m_1 = \sigma\sqrt{\pi/2}, m_2 = 2\sigma^2,$ $\mu_2 = \frac{4-\pi}{2}\sigma^2, \mu_3 \cong 0{,}63\sigma^3,$ $\mu_4 \cong 2{,}7\sigma^4$
1	2	3
Равномерный	$\frac{1}{b-a}, a \leq x \leq b$	$m_1 = \frac{a+b}{2}, \mu_2 = \frac{(b-a)^2}{12},$ $\mu_3 = 0, \mu_4 = \frac{1}{80}(b-a)^4$
Экспоненциальный	$\lambda e^{-\lambda x}, x \geq 0$	$m_1 = 1/\lambda, m_2 = 2/\lambda^2,$ $\mu_2 = 1/\lambda^2, \mu_3 = 2/\lambda^3,$ $\mu_4 = 9/\lambda^4$
Логарифмически-нормальный	$\frac{1}{x\sqrt{2\pi}\sigma}\exp\left(-\frac{(\ln x - a)^2}{2\sigma^2}\right), x > 0$	$m_1 = \exp\left(a + 0{,}5\sigma^2\right),$ $\mu_2 = \exp\left(2a + \sigma^2\right)\left(\exp\left(\sigma^2\right) - 1\right)$
Гамма	$\frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, x \geq 0, \beta > 0$	$m_1 = \alpha\beta, m_2 = \alpha(\alpha+1)\beta^2,$ $\mu_2 = \alpha\beta^2, \mu_3 = 2\alpha\beta^3,$ $\mu_4 = 3(\alpha+2)\alpha\beta^4$
Вейбулла	$\alpha\beta x^{\alpha-1}\exp\left(-\beta x^\alpha\right), x \geq 0$	$m_1 = \Gamma\left(1+\frac{1}{\alpha}\right)\beta^{-1/\alpha}$ $\mu_2 = \left(\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right)\beta^{-2/\alpha}$

Системы случайных величин. В тех случаях, когда с каждым исходом ω эксперимента связана пара чисел $X_1(\omega)$ и $X_2(\omega)$, соответствующее отображение

$(X_1(\omega), X_2(\omega))$ называется двумерной СВ или системой двух СВ и обозначается (X_1, X_2). Например, если случайный сигнал $X(t)$ на выходе радиоприемного устройства наблюдается в два момента времени t_1 и t_2, то упорядоченная пара возможных значений сигнала $X_1 = X(t_1)$ и $X_2 = X(t_2)$ представляет собой двумерную СВ (X_1, X_2).

Двумерную СВ (X_1, X_2) можно рассматривать как случайную точку или как случайный вектор на координатной плоскости. При этом каждому конкретному исходу опыта (x_1, x_2) ставится в соответствие точка плоскости с координатами x_1 и x_2.

1.7.1. Характеристики случайного процесса

Помехи в системах связи описываются методами теории случайных процессов.

Функция называется случайной, если в результате эксперимента она принимает тот или иной вид, заранее неизвестно, какой именно. Случайным процессом называется случайная функция времени. Конкретный вид, который принимает случайный процесс в результате эксперимента, называется реализацией случайного процесса.

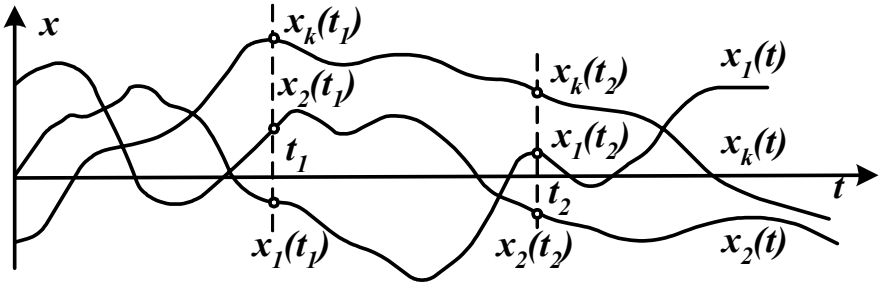


Рис. 1.19. Реализация случайного процесса ***X(t)***

На рис. 1.19 показана совокупность нескольких (трех) реализаций случайного процесса $x^{(1)}(t)$, $x^{(2)}(t)$, $x^{(3)}(t)$. Такая совокупность называется ансамблем реализаций. При фиксированном значении момента времени $t = t_1$ в первом

эксперименте получим конкретное значение $x^{(1)}(t_1)$, во втором – $x^{(2)}(t_1)$, в третьем – $x^{(3)}(t_1)$.

Случайный процесс носит двойственный характер. С одной стороны, в каждом конкретном эксперименте он представлен своей реализацией – неслучайной функцией времени. С другой стороны, случайный процесс описывается совокупностью случайных величин.

Действительно, рассмотрим случайный процесс $X(t)$ в фиксированный момент времени $t = t_1$ Тогда $X(t_1)$ в каждом эксперименте принимает одно значение $x(t_1)$, причем заранее неизвестно, какое именно. Таким образом, случайный процесс, рассматриваемый в фиксированный момент времени $t = t_1$ является случайной величиной. Если зафиксированы два момента времени t_1 и t_2, то в каждом эксперименте будем получать два значения $x(t_1)$ и $x(t_2)$. При этом совместное рассмотрение этих значений приводит к системе $(X(t_1), X(t_2))$ двух случайных величин. При анализе случайных процессов в N моментов времени приходим к совокупности или системе N случайных величин $(X(t_1),..., X(t_N))$.

Математическое ожидание, дисперсия и корреляционная функция случайного процесса. Поскольку случайный процесс, рассматриваемый в фиксированный момент времени, является случайной величиной, то можно говорить о математическом ожидании и дисперсии случайного процесса:

$$m(t) = M\{X(t)\},\ D(t) = M\{X(t) - m(t))^2\}.$$

Так же, как и для случайной величины, дисперсия характеризует разброс значений случайного процесса относительно среднего значения $m(t)$. Чем больше $D(t)$, тем больше вероятность появления очень больших положительных и отрицательных значений процесса. Более удобной характеристикой является среднее квадратичное отклонение (СКО) $\sigma(t) = \sqrt{D(t)}$, имеющее ту же размерность, что и сам случайный процесс.

Если случайный процесс описывает, например, изменение дальности до объекта, то математическое ожидание – средняя дальность в метрах; дисперсия

измеряется в квадратных метрах, а СКО – в метрах и характеризует разброс возможных значений дальности относительно средней.

Среднее значение и дисперсия являются очень важными характеристиками, позволяющими судить о поведении случайного процесса в фиксированный момент времени. Однако, если необходимо оценить «скорость» изменения процесса, то наблюдений в один момент времени недостаточно. Для этого используют две случайные величины $(X(t_1), X(t_2))$, рассматриваемые совместно. Так же, как и для случайных величин, вводится характеристика связи или зависимости между $X(t_1)$ и $X(t_2)$. Для случайного процесса эта характеристика зависит от двух моментов времени t_1 и t_2 и называется корреляционной функцией: $R(t_1,t_2) = M\{(X(t_1) - m(t_1)(X(t_2) - m(t_2))\}$.

Стационарные случайные процессы. Многие процессы в системах управления протекают однородно во времени. Их основные характеристики не изменяются. Такие процессы называются стационарными. Точное определение можно дать следующим образом. Случайный процесс $X(t)$ называется стационарным, если любые его вероятностные характеристики не зависят от сдвига начала отсчета времени. Для стационарного случайного процесса математическое ожидание, дисперсия и СКО постоянны: $m(t) = m$, $D(t) = D = \sigma^2$.

Корреляционная функция стационарного процесса не зависит от начала отсчета t, т.е. зависит только от разности $\tau = t_2 - t_1$ моментов времени:

$$R(\tau) = M\{(X(t_1) - m) \cdot (X(t_1 - \tau) - m)\}.$$

Корреляционная функция стационарного случайного процесса имеет следующие свойства:

$$1)\ \mathrm{R}(\tau = 0) = \sigma^2; \qquad 2)\ \mathrm{R}(\tau) = \mathrm{R}(-\tau); \qquad 3)\ R(\tau \to \infty) = 0.$$

Часто корреляционные функции процессов в системах связи имеют вид, показанный на рис. 1.20.

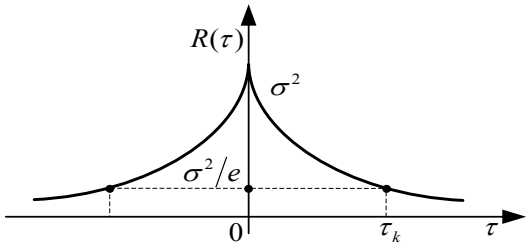


Рис. 1.20. Корреляционные функции процессов

Интервал времени τ_k, на котором корреляционная функция, т.е. величина связи между значениями случайного процесса, уменьшается в М раз, называется интервалом или временем корреляции случайного процесса. Обычно $M = 10$ или $M = e$. Можно сказать, что значения случайного процесса, отличающиеся по времени на интервал корреляции, слабо связаны друг с другом.

Таким образом, знание корреляционной функции позволяет судить о скорости изменения случайного процесса.

Другой важной характеристикой является энергетический спектр случайного процесса. Он определяется как преобразование Фурье от корреляционной функции:

$$G(\omega) = \int_{-\infty}^{\infty} R(\tau) e^{-j\omega\tau} d\tau .$$

Очевидно, справедливо и обратное преобразование:

$$R(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) e^{j\omega\tau} d\omega .$$

Энергетический спектр показывает распределение мощности случайного процесса, например помехи, на оси частот.

При анализе САУ очень важно определить характеристики случайного процесса на выходе линейной системы при известных характеристиках процесса на входе САУ. Предположим, что линейная система задана импульсной переходной характеристикой $h(\tau)$. Тогда выходной сигнал в момент времени t_1 определяется интегралом Дюамеля:

$$x(t_1) = \int_{-\infty}^{\infty} h(\tau_1) g(t_1 - \tau_1) d\tau_1 ,$$

где $g(t)$ – процесс на входе системы. Для нахождения корреляционной функции $R_x(t_1,t_2)=M\{x(t_1)x(t_2)\}$ запишем $x(t_2)=\int_{-\infty}^{\infty}h(t_2)g(t_2-\tau_2)d\tau_2$ и после перемножения найдем математическое ожидание

$$R_x(t_1,t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\tau_1)h(\tau_2)M\{g(t_1-\tau_1)g(t_2-\tau_2)\}d\tau_1 d\tau_2 .$$

Таким образом, связь между корреляционными функциями входного и выходного случайных процессов устанавливается с помощью следующего двойного интеграла:

$$R_x(t_1,t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\tau_1)h(\tau_2)R_g(t_1-\tau_1,t_2-\tau_2)d\tau_1 d\tau_2 .$$

Для стационарных процессов корреляционные функции зависят только от разности аргументов $t_1-t_2=u$, $(t_1-\tau_1)-(t_2-\tau_2)=v$ и поэтому

$$R_x(u)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\tau_1)h(\tau_2)R_g(v)d\tau_1 d\tau_2 .$$

Более простое соотношение можно найти для энергетических спектров $G_g(\omega)$ и $G_x(\omega)$ входного и выходного сигналов при известной передаточной функции $W(j\omega)$ линейной системы. Действительно, найдем преобразование Фурье от левой и правой частей последнего равенства. Получим следующее выражение:

$$G_x(\omega)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\tau_1)\,h(\tau_2)\,R_g(v)e^{-j\omega u}d\tau_1\,d\tau_2\,du .$$

После замены переменной $v=t_1-t_2-(\tau_1-\tau_2)=u-\tau_1-\tau_2$ или $u=v+\tau_1-\tau_2$ тройной интеграл преобразуется в произведение

$$R_x(u)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\tau_1)h(\tau_2)R_g(v)d\tau_1 d\tau_2$$

$$G_x(\omega)=\left(\int_{-\infty}^{\infty}h(\tau_1)e^{-j\omega\tau_1}d\tau_1\right)\left(\int_{-\infty}^{\infty}h(\tau_2)e^{j\omega\tau_2}d\tau_2\right)\left(\int_{-\infty}^{\infty}R_g(v)e^{j\omega v}dv\right).$$

Поскольку преобразование Фурье от импульсной характеристики дает передаточную функцию, находим окончательно связь между энергетическими спектрами процессов на входе и на выходе линейной системы:

$$G_x(\omega) = W(j\omega)W(-j\omega)G_g(\omega) = |W(j\omega)|^2 G_g(\omega).$$

Часто помехи в системах управления имеют очень широкий спектр. В таких случаях их удобно представить в виде так называемого белого шума – процесса с постоянным энергетическим спектром: $G(\omega) = N_0$. Корреляционная функция белого шума $R(\tau) = N_0\delta(\tau)$, где $\delta(t)$ – импульсная дельта-функция. Это означает, что даже очень близкие по времени значения белого шума не связаны друг с другом.

1.7.2. Флуктуационный шум

Примером случайного процесса является флуктуационный шум, наиболее характерный для большинства каналов электросвязи. Для количественных расчетов воздействия флуктуационного шума на сигнал необходимо знать основные вероятностные характеристики. Поскольку шум образуется как сумма большого числа отдельных независимых колебаний, он, согласно центральной предельной теореме представляет собой стационарный эргодический случайный процесс с гауссовским (нормальным) распределением вероятности.

ПРВ гауссовского процесса описывается формулой [6, 32]:

$$w(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right], \qquad (1.47)$$

в которую входят два числовых параметра m и σ^2, имеющие смысл математического ожидания и дисперсии: $m = M(X)$, $\sigma^2 = D(X)$. График плотности вероятности $w(x)$ представляет собой колоколообразную кривую с единственным максимумом в точке $x = m$ (рис. 1.21). Из графика видно, что с уменьшением σ кривая все более локализуется в окрестности точки $x = m$. Для флуктуационного шума обычно $M(X) = 0$.

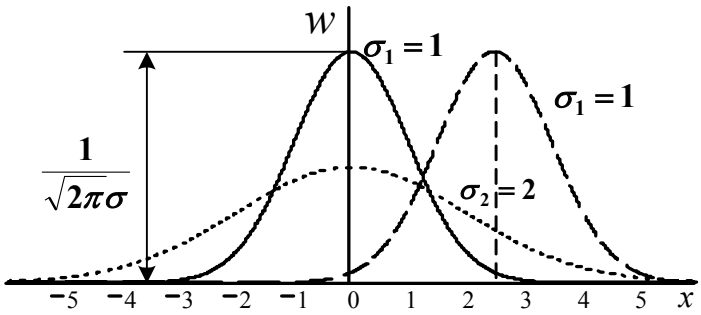


Рис. 1.21. Гауссовское распределение вероятностей:

Функция распределения вероятности для гауссовского случайного процесса:

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]dx\,.$$

После замены переменных $y=\dfrac{(x-m)}{\sigma}$ эта функция приводится к виду:

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x-m)/\sigma}\exp\left[-y^2\big/2\right]dy=0{,}5+\Phi_o\left(\frac{x-m}{\sigma}\right),\qquad(1.48)$$

где

$\Phi_o(z)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^x\exp\left[-y^2\big/2\right]dy$ - интеграл вероятности.

Функция $\Phi_0(z)$ табулирована в математических справочниках. Заметим, что $\Phi_0(-z)=-\Phi_0(z)$, $\Phi_0(0)=0$, $\Phi_0(\infty)=0{,}5$. Для приближенных вычислений можно воспользоваться приближенным выражением:

$$\Phi_o(z)\approx 0{,}5-0{,}65\exp\left[-0{,}44(z+0{,}75)^2\right]\qquad(1.49)$$

Пример 1.3. Вычислим вероятность того, что мгновенное значение флуктуационного шума с нулевым средним и дисперсией $\sigma^2=9[B^2]$ превысит уровень $x_0=6[B]$.

Исходя из определения функции распределения вероятности (1.43), вероятность превышения случайным процессом уровня x_0

$$p(X>x_o)=1-p(X\le x_o)=1-F(x_o).$$

Подставляя значение $F(x_0)$ для гауссовского случайного процесса, получаем:

$$p(X>x_o)=1-0{,}5-\Phi_o[(x_o-m)/\sigma]=0{,}5-\Phi_o[(x_o-m)/\sigma].$$

Для заданных числовых значений и $m=0$, воспользовавшись таблицами или приближенной формулой (1.49) для $\Phi_0(z)$, получаем:

$$p(X>6)\approx 2{,}33\cdot 10^{-2}.$$

Спектральная плотность мощности $G_x(f)$ флуктуационного шума зависит

от физической природы его образования, а также участка канала связи, где он рассматривается. Обычно спектральная плотность мощности $G_x(f)$ флуктуационного шума постоянна в широком диапазоне частот, т. е. можно приближенно считать, что: $G(f) = N_o$ при $0 \le f \le \infty$. В этом случае шум называют *белым*. Это название дано по аналогии с белым светом, имеющим все частотные компоненты.

1.8. Комплексное представление сигналов и помех

Ранее было изучено представление детерминированных сигналов рядами ортогональных функций. Такая модель оказывается особенно полезной при анализе прохождения сигналов через линейные радиотехнический устройства.

Вместе с тем при анализе нелинейных преобразований сигналов и, в частности, модуляции и демодуляции, требуется иной подход. Этот подход основывается на понятии аналитического сигнала.

1.8.1. Понятие аналитического сигнала

Многие формулы гармонического анализа записываются значительно проще и некоторые задачи решаются легче, если использовать в качестве элементарных функций не обычные действительные синусоиды, а экспоненциальные функции мнимого аргумента. Действительно, по формуле Эйлера [6, 21]:

$$\cos(\omega t + \varphi) = \frac{1}{2}\left[\exp\{j(\omega t + \varphi)\} + \exp\{-j(\omega t + \varphi)\}\right]. \qquad (1.50)$$

Этой записи можно дать геометрическую трактовку, пользуясь представлением комплексных чисел в виде точек или векторов на плоскости.

Выражение $\exp j(\omega t + \varphi)$ представляет в данном случае вектор единичной длины, проведенный под углом $\omega t + \varphi$ к действительной оси (рис.1.22). При изменении времени t этот вектор, единичной длины, меняет положение, вращаясь в положительном направлении с угловой скоростью ω.

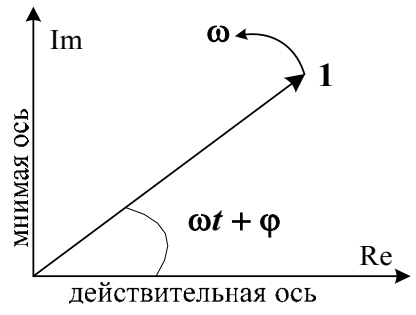


Рис. 1.22. Геометрическая трактовка экспоненциальной функции мнимого аргумента

Изобразить синусоиду в форме (1.50), это значит представить ее суммой двух векторов, длина каждого из которых равна $1/2$, расположенных в любой момент времени симметрично относительно действительной оси и вращающихся в разных направлениях с угловыми скоростями ω и $-\omega$ (рис.1.23).

В момент $t = 0$ они занимают положения под углами φ и $-\varphi$ относительно действительной оси. Геометрическая сумма векторов всегда совпадает по направлению с действительной осью и представляет действительную функцию времени $\cos(\omega t + \varphi)$.

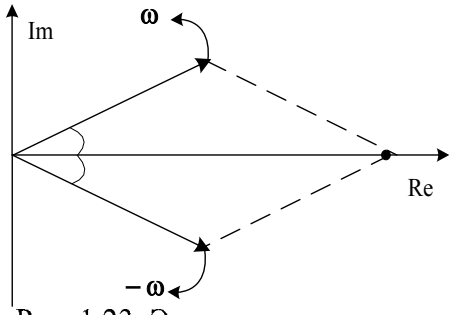


Рис. 1.23. Экспоненциальное представление элементарной функции

При представлении косинусоиды в виде $\cos(\omega t + \varphi) = \mathrm{Re}[\exp j(\omega t + \varphi)]$ можно ограничиться одним вращающимся в положительном направлении вектором и представить косинусоиду его проекцией на действительную ось.

В этом случае нет необходимости вводить отрицательные частоты. Длина вектора представляет амплитуду косинусоиды, а угол, образуемый им в данный момент с действительной осью, - полную фазу $(\omega t + \varphi)$.

Проекция этого вектора на мнимую ось равна $\mathrm{Im}[\exp j(\omega t + \varphi)] = \sin(\omega t + \varphi)$, т.е. представляет ту же косинусоиду, сдвинутую по фазе на $\pi/2$ (рис. 1.24).

Многие сигналы в системах электросвязи можно представлять в виде:

$$S(t) = A(t)\cos[\omega t + \varphi(t)], \qquad (1.51)$$

т.е. как «квазигармоническую» функцию с переменными «амплитудой» и «начальной фазой». Такой сигнал можно интерпретировать геометрически как проекцию на действительную ось вращающегося вектора, но при этом изменяющего свою длину и угловую скорость. Для описания свойств сигнала представленного в форме (1.51) вводят понятие комплексного аналитического сигнала.

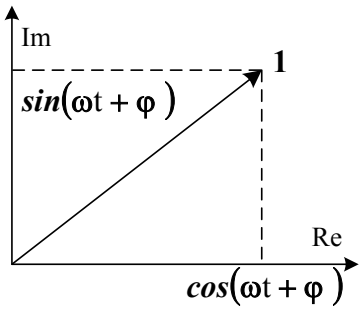


Рис. 1.24. Проекции единичного вектора на действительную и мнимую оси

Рассмотрим отрезок сигнала на некотором интервале времени $0 < t < T$.

Его можно представить на этом интервале рядом Фурье в экспоненциальной форме [6, 21]:

$$S(t) = \sum_{-\infty}^{\infty} \dot{S}_k \cdot e^{jk\omega_o t}, \omega_0 = \frac{2\pi}{T}. \qquad (1.52)$$

Пользуясь геометрическим представлением синусоиды, можно представить сигнал $S(t)$ в виде суммы вращающихся векторов, каждый из которых имеет вид:

$$\dot{S}_k \cdot e^{jk\omega_o t} = \left|\dot{S}_k\right| \cdot e^{j\varphi_k} \cdot e^{jk\omega_o t} = \left|\dot{S}_k\right| \cdot e^{j(k\omega_o t + \varphi_k)}.$$

Векторы с индексами $k > 0$ вращаются в положительном направлении, а с $k < 0$ в отрицательном. Пара таких векторов с индексами k и $-k$ образует одну действительную косинусоиду.

Поэтому предполагая среднее состояние сигнала нулевым ($S_0 = 0$)косинусоида может быть представлена проекцией на действительную ось одного вектора, вращающегося, например, в положительном направлении. Вместо (1.68) можно взять проекцию (т.е. действительную часть) суммы векторов, вращающихся только в положительном направлении, увеличив их величину вдвое:

$$S(t) = \mathrm{Re}\sum_{k=1}^{\infty} 2\dot{S}_k \cdot e^{jk\omega_o t}, 0 < t < T\,. \qquad (1.53)$$

Ряд в правой части (1.53) представляет собой комплексную функцию времени, которую обозначим $\dot{S}(t)$ и будем называть комплексным или аналитическим сигналом:

$$\dot{S}(t) = \sum_{k=1}^{\infty} 2\dot{S}_k \cdot e^{jk\omega_o t}, 0 < t < T\,. \qquad (1.54)$$

Его геометрическим представлением является вектор, образующийся при суммировании элементарных векторов $S_{\mathrm{k}}, \mathrm{k} = 1,2,\ldots$. Так как элементарные векторы вращаются с разными угловыми скоростями $k\omega_0$, то их взаимная конфигурация со временем изменяется. Поэтому их векторная сумма (рис. 1.25) представляет собой вектор с переменной длиной, вращающийся с переменной угловой скоростью.

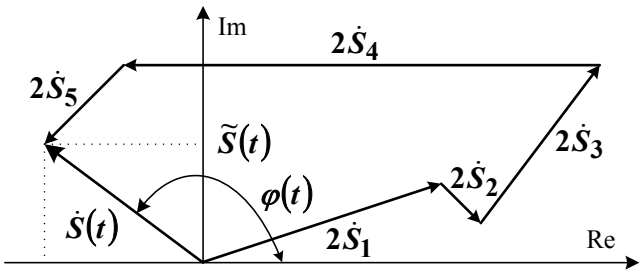


Рис. 1.25. Геометрическое представление комплексного сигнала

Исходный сигнал (1.51) является действительной частью аналитического сигнала.

Учитывая выражение (1.19) для комплексных коэффициентов ряда Фурье получим:

$$\begin{aligned} S(t) &= \mathrm{Re}\,\dot{S}(t) = \sum_{k=1}^{\infty} \mathrm{Re}\left[(a_k - jb_k)e^{jk\omega_0 t}\right] = \\ &= \sum_{k=1}^{\infty} (a_k \cos k\omega_0 t + b_k \sin k\omega_0 t), 0 < t < T, \end{aligned} \qquad (1.55)$$

что является обычным разложением сигнала в ряд Фурье в тригонометрической форме.

Мнимая часть аналитического сигнала представляет собой некоторую функцию времени, однозначно определяемую исходным сигналом $S(t)$.

Ее обозначают $\widetilde{S}(t)$ и называют сигналом, сопряженным по Гильберту с $S(t)$ [6, 20]:

$$\widetilde{S}(t) = \operatorname{Im}\dot{S}(t) = \sum_{k=1}^{\infty} \operatorname{Im}\left[(a_k - jb_k)e^{jk\omega_0 t}\right] = \\ = \sum_{k=1}^{\infty} \left(-b_k \cos k\omega_0 t + a_k \sin k\omega_0 t\right), 0 < t < T. \tag{1.56}$$

Отсюда видно, что сопряженный сигнал можно получить из исходного, повернув начальные фазы всех его составляющих на $-\pi/2$ или, другими словами, заменив в ряде Фурье (1.71) cos на sin, а sin на $-\cos$.

В соответствии с (1.54), (1.55) и (1.56) аналитический сигнал может быть выражен через реальный и сопряженный сигналы следующим образом:

$$\dot{S}(t) = S(t) + j\widetilde{S}(t). \tag{1.57}$$

Исходя из этого, аналитический сигнал в момент времени *t* может быть представлен точкой на комплексной плоскости, если по оси абсцисс откладывать значения реального сигнала $S(t)$, а по оси ординат - сопряженного с ним сигнала $\widetilde{S}(t)$ (рис. 1.26).

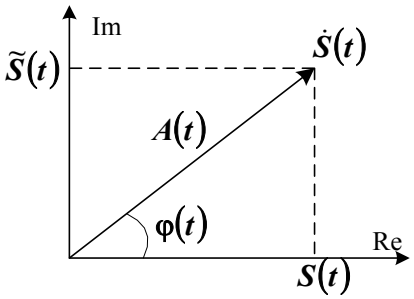


Рис. 1.26. Представление аналитического сигнала точкой

В качестве примера рассмотрим гармоническое колебание:

$$S(t) = A \cdot \cos \omega_0 t, \omega_0 = \frac{2\pi}{T}. \tag{1.58}$$

Такой сигнал представляется только одним членом ряда Фурье, так как

$$S_0 = \frac{a_0}{2} = \frac{1}{T}\int_0^T A\cos\omega_0 t dt = \frac{A}{\omega_{0T}} \sin\omega_0 t\Big|_0^{T=2\pi/\omega_0} = 0, \tag{1.59}$$

$$a_k = \begin{cases} a_1 = A, k = 1 \\ 0, k > 0 \end{cases}, \tag{1.60}$$

$$b_k = 0, \text{ при любых } k. \tag{1.61}$$

Справедливость (1.60) и (1.61) вытекает из ортогональности функций $\cos k\omega_0 t$ и $\sin k\omega_0 t$.

В соответствии с (1.56), (1.60), (1.61) сопряженный сигнал:

$$\widetilde{S}(t) = A\sin\omega_0 t. \tag{1.62}$$

Тогда аналитический сигнал $\dot{S}(t)$, соответствующий реальному сигналу (9), можно записать следующим образом:

$$\dot{S}(t) = A\cos\omega_0 t + jA\sin\omega_0 t. \tag{1.63}$$

Точки, отображающие реальный и сопряженный сигналы (рис.1.28), совершают в данном случае гармонические колебания по оси абсцисс и ординат относительно точки 0 по законам, соответственно, косинуса и синуса.

Длина вектора, соединяющего начало координат на рис.1.28 с точкой $\dot{S}(t)$, отображающей аналитический сигнал [6, 32],

$$A(t)=\sqrt{S^2(t)+\widetilde{S}^2(t)}=\sqrt{A^2\cos^2\omega_0 t+A^2\sin^2\omega_0 t}=A=const\,. \qquad (1.64)$$

Угол между вектором $\dot{S}(t)$ и осью абсцисс

$$\varphi(t)=arctg\frac{\widetilde{S}(t)}{S(t)}+\begin{cases}0,\widetilde{S}\ge 0\\ \pi,\widetilde{S}<0\end{cases}. \qquad (1.65)$$

В рассматриваемом случае для верхней полуплоскости получаем:

$$\varphi(t)=arctg\frac{A\cdot\sin\omega_0 t}{A\cdot\cos\omega_0 t}=arcg\left[tg(\omega_0 t)\right]=\omega_0 t\,. \qquad (1.66)$$

Таким образом, с течением времени точка, отображающая аналитический сигнал $\dot{S}(t)$, соответствующий гармоническому колебанию (1.58), равномерно вращается по окружности с радиусом $A(t)$ с угловой скоростью ω_0. Параметры $A(t)$ и $\varphi(t)$ в данном случае определяют амплитуду и фазу синусоидального сигнала.

Для других сигналов, отличных от гармонических, точка $\dot{S}(t)$ перемещается на комплексной плоскости по более сложной траектории, отличающейся от круговой.

1.8.2. Огибающая, мгновенная фаза и мгновенная частота узкополосного случайного процесса

Комплексный сигнал (1.57) можно представить в форме [6]:

$$\dot{S}(t)=A(t)e^{j\varphi(t)}=A(t)\cos\varphi(t)+jA\sin\varphi(t), \qquad (1.67)$$

где $A(t)=\sqrt{S^2(t)+\widetilde{S}^2(t)}$ называется огибающей сигнала, (1.68)

а $\varphi(t)=Arg\,\dot{S}(t)=arctg\frac{\widetilde{S}(t)}{S(t)}+\begin{cases}0,\widetilde{S}(t)\ge 0\\ \pi,\widetilde{S}(t)<0\end{cases}$ мгновенной фазой сигнала.

Здесь: $S(t)=A(t)\cos\varphi(t);\qquad \widetilde{S}(t)=A(t)\sin\varphi(t)$ (1.69)

Функция $\varphi(t)$ называется мгновенной фазой сигнала.

Производная от мгновенной фазы сигнала по времени называется мгновенной частотой сигнала:

$$\omega(t)=\frac{d\varphi(t)}{dt}=\frac{1}{2\pi}\cdot\frac{\widetilde{S}'(t)\cdot S(t)-S'(t)\cdot\widetilde{S}(t)}{S^2(t)+\widetilde{S}^2(t)}. \qquad (1.70)$$

Например, для гармонического сигнала [6]:

$$\omega(t)=\frac{\cos\omega_0 t\cdot\cos\omega_0 t\cdot\omega_0\cdot A^2+\sin\omega_0 t\cdot\sin\omega_0 t\cdot\omega_0\cdot A^2}{A^2\cdot t+A^2\cdot\sin^2\omega_0 t}=\omega_0=const\,.$$

В общем случае мгновенная частота изменяется во времени.

Из (1.68) следует, что $A(t)\geq S(t)$, причем равенство достигается в моменты времени, когда $\widetilde{S}(t)=0$. В этих точках производная $A(t)$ совпадает с производной сигнала $S(t)$:

$$A'(t)=\frac{S(t)\cdot S'(t)+\widetilde{S}(t)\cdot\widetilde{S}'(t)}{\sqrt{S^2(t)+\widetilde{S}^2(t)}};\ A'(t)=S'(t)\quad\text{при}\quad\widetilde{S}(t)=0\,. \tag{1.71}$$

Следовательно, при $\widetilde{S}(t)=0$ огибающая $A(t)$ касается сигнала $S(t)$.

Функция $\cos\varphi(t)$ называется высокочастотным заполнением сигнала.

Процесс формирования сигнала на основе огибающей $A(t)$ и фазы $\varphi(t)$ показан на рис. 1.27.

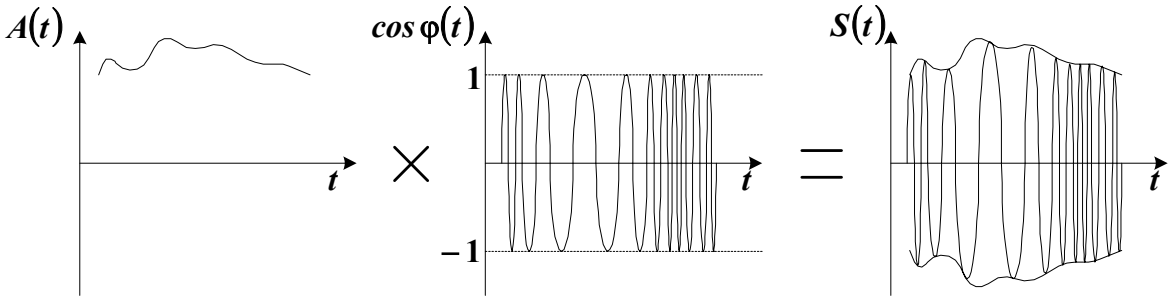


Рис. 1.27. Временное представление огибающей и высокочастотного заполнения

Если мгновенная частота колеблется вокруг среднего значения $\omega_{ср}$, то можно записать:

$$\begin{aligned}&\varphi(t)=\omega_{ср}t+\Theta(t);\\&S(t)=A(t)\cos\left[\omega_{ср}t+\Theta(t)\right],\end{aligned} \tag{1.72}$$

где $\Theta(t)$ – называется мгновенной начальной фазой сигнала.

Выражение (1.72) удобно для описания узкополосных сигналов. В этом случае основная часть спектра амплитуд сосредоточена в относительно узкой, по сравнению с $\omega_{ср}$, полосе частот. При этом $A(t)$ и $\varphi(t)$ изменяются медленно по сравнению с $\cos\omega_{ср}t$. Такие сигналы называются квазигармоническими. У случайных сигналов и помех $A(t)$, $\varphi(t)$, $\omega(t)$, $\omega_{ср}(t)$ и $\Theta(t)$ являются случайными функциями времени.

Контрольные вопросы

1. Что такое информация, сообщение, сигнал? Что общего и в чем отличие между этими понятиями?

2. Что такое линия связи, канал связи?

3. Какие радиотехнические устройства обязательно входят в систему электросвязи?

4. Что понимается под аддитивными и мультипликативными помехами?

5. Перечислите известные Вам источники помех. В чем состоит существенное отличие помех от искажений?

6. Постройте спектральную диаграмму разложения в ряд Фурье однополярной периодической последовательности прямоугольных видеоимпульсов с известными параметрами?

7. Как график автокорреляционной функции сигнала характеризует полосу частот, занимаемую сигналом?

8. Источник выдает первичный сигнал $s(t)$, представляющий собой непрерывный стационарный случайный процесс, мгновенные значения которого в интервале $0 \div 4$ (В) распределены по равномерному закону, а мощность постоянна в полосе частот от $0{,}3$ до $3{,}4$ (кГц):

запишите аналитическое выражение и постройте графики плотности вероятности и функции распределения мгновенных значений сигнала $s(t)$;

определите математическое ожидание $M(x)$ и дисперсию $D(x)$ сигнала $s(t)$;

определить величину энергетического спектра сигнала.

ГЛАВА 2. МЕТОДЫ ФОРМИРОВАНИЯ И ПРЕОБРАЗОВАНИЯ СИГНАЛОВ

2.1. Модуляция сигналов

Формирование модулированных сигналов (модуляция) предполагает взаимодействие двух сигналов: управляющего модулирующего и вспомогательного несущего. Суть управляющего воздействия модулирующего сигнала $s_c(t)$ заключается в том, что некоторые параметры γ несущего колебания изменяются в соответствии с модулирующим колебанием.

В системах связи в качестве управляющих колебаний используются разнообразные первичные электрические сигналы (ПЭС): телефонные, телеграфные, телевизионные и др..

В качестве несущих широко применяются гармонические сигналы, собственная частота которых ω_0 значительно превосходит верхнюю частоту Ω_{max} спектра модулирующего колебания. Это означает, что по отношению к несущему колебанию модулирующее колебание медленно изменяет свои значения во времени. Медленность изменения $s_c(t)$ подчеркивает, что на период модулирующего колебания приходятся тысячи, сотни тысяч и более периодов несущего колебания. При этом, с одной стороны, обеспечивается достаточно полное отображение модулирующего колебания в несущем колебании, а с другой, обусловливается узкополосность спектра модулированного колебания.

Таким образом, для передачи информации, содержащейся в ПЭС, используется вспомогательное несущее колебание, выполняющее роль переносчика сообщения

$$S_н(t) = A\cos(\omega_н t + \varphi_н), \qquad \omega_н = 2\pi f_н . \tag{2.1}$$

Обычно полагают $f_н >> k \cdot F_1$, где F_1 – наивысшая гармоника ПЭС.

Процесс изменения одного или нескольких параметров высокочастотного (несущего) колебания в соответствии с первичным (модулирующим) сигналом называется модуляцией. Дискретную модуляцию обычно называют манипуля-

цией.

При модуляции информационными параметрами несущего колебания $S_н(t) = A\cos(\omega_н t + \varphi_н)$ могут быть амплитуда A, частота $\omega_н$ или фаза $\varphi_н$, которые изменяются в соответствии с модулирующим сигналом $s_c(t)$, поэтому различают амплитудную модуляцию (АМ), частотную модуляцию (ЧМ) и фазовую модуляцию (ФМ).

В модулируемых колебаниях изменяемые параметры имеют вид:

при амплитудной модуляции – $A(t) = A + \Delta A(t) = A + as_c(t)$;

при частотной модуляции – $\omega_н(t) = \omega_0 + \Delta\omega(t) = \omega_0 + as_c(t)$;

при фазовой модуляции – $\varphi_н(t) = \varphi_0 + \Delta\varphi(t) = \varphi_0 + as_c(t)$;

где $\Delta A_m(t)$, $\Delta\omega(t)$, $\Delta\varphi(t)$ – приращения, пропорциональные модулирующему колебанию $s_c(t)$; a – коэффициент пропорциональности.

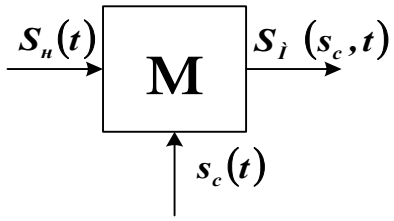


Рис. 2.1. Обобщенная схема модулятора

Устройство для получения результирующего (модулированного) сигнала $S_M(S_c, t)$ называется модулятором, (рис. 2.1), на один вход которого подается несущее (модулируемое) колебание $S_н(t)$, на второй вход первичный (модулирующий) сигнал $s_c(t)$.

2.1.1. Амплитудная модуляция гармонической несущей

Амплитудная модуляция – процесс изменения амплитуды несущего колебания, соответствующего изменению непрерывного информационного сигнала [21, 32, 39].

При амплитудной модуляции мгновенная амплитуда несущего колебания:

$$A(t) = A_0 + as_c(t), \qquad (2.2)$$

где A_0 – амплитуда несущей; a – коэффициент пропорциональности, выбираемый так, чтобы амплитуда $A(t)$ всегда была положительной. Частота и фаза несущего гармонического колебания при АМ остаются неизменными.

Для математического описания АМ сигнала в (2.2) вместо коэффициента a, зависящего от конкретной схемы модулятора, вводится индекс модуляции:

$$m_{AM} = \left(A_{\max} - A_{\min}\right) \Big/ \left(A_{\max} + A_{\min}\right), \qquad (2.3)$$

т.е. отношение разности между максимальным и минимальным значениями амплитуд АМ сигнала к сумме этих значений. Для симметричного модулирующего сигнала $s_c(t)$ АМ сигнал также симметричный, т.е. $A_{\max} = A_{\min} = 2\Delta A$. Тогда индекс модуляции равен отношению максимального приращения амплитуды, к амплитуде несущей.

$$m_{AM} = \Delta A \Big/ A_0 . \qquad (2.4)$$

Физически индекс модуляции характеризует собой глубину амплитудной модуляции и может изменяться в пределах $0 \le m_{AM} \le 1$.

Таким образом для любого АМ сигнала справедливо:

$$S_{AM}(s_c, t) = A_0\left[1 + m_{AM} s_c(t)\right]\cos\left(\omega_0 t + \varphi_0\right). \qquad (2.5)$$

Амплитудная модуляция гармоническим колебанием. В простейшем случае модулирующий сигнал является гармоническим колебанием с частотой $\Omega << \omega_0$. При этом выражение

$$S_{AM}(s_c, t) = A_0\left[1 + m_{AM}\cos\Omega t\right]\cos\left(\omega_0 t + \varphi_0\right), \qquad (2.6)$$

соответствует однотональному АМ сигналу, представленному на рис. 2.26.

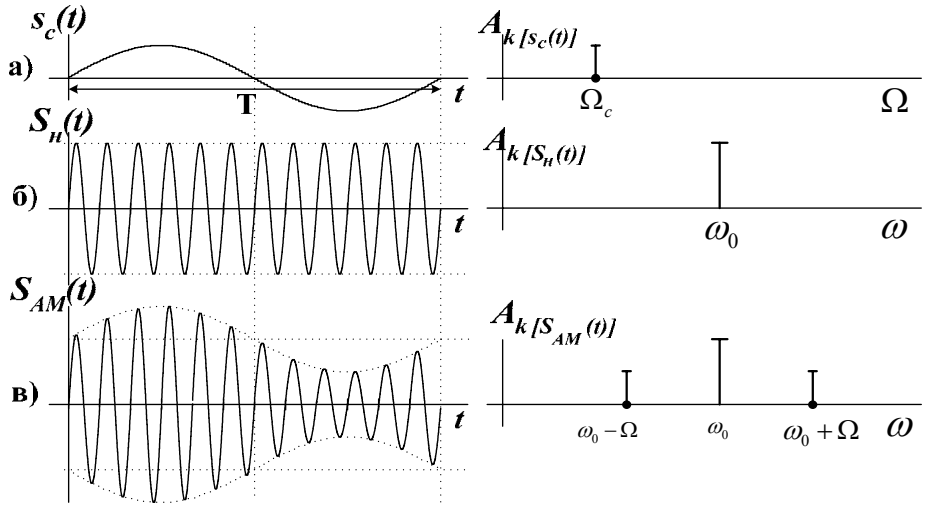


Рис. 2.2. Временные и спектральные диаграммы процесса формирования АМ гармонического колебания

Однотональный АМ сигнал можно представить в виде суммы трех гармонических составляющих с частотами: ω_0 – несущей; $\omega_0 + \Omega$ – верхней боковой и $\omega_0 - \Omega$ – нижней боковой:

$$S_{AM}(s_c, t) = A_0 \cos(\omega_0 t + \varphi_0) + \frac{A_0 m_{AM}}{2} \cos[(\omega_0 + \Omega)t + \varphi_0] + \\ + \frac{A_0 m_{AM}}{2} \cos[(\omega_0 - \Omega)t + \varphi_0] \quad . \qquad (2.7)$$

Спектральная диаграмма однотонального АМ сигнала, построенная по (2.7), симметрична относительно несущей частоты ω_0 (рис. 2.2,в). Амплитуды боковых колебаний с частотами $\omega_0 - \Omega$ и $\omega_0 + \Omega$ одинаковы и даже при $m_{AM} = 1$ не превышают половины амплитуды несущего колебания A_0.

Гармонические модулирующие сигналы и соответственно однотональный АМ сигнал на практике встречаются редко. В большинстве случаев модулирующие первичные сигналы $s_c(t)$ являются сложными функциями времени (рис.2.3,а). Любой сложный сигнал $s_c(t)$ можно представить в виде конечной или бесконечной суммы гармонических составляющих, воспользовавшись рядом или интегралом Фурье. Каждая гармоническая составляющая сигнала $s_c(t)$ с частотой Ω_i приведет к появлению в АМ сигнале двух боковых составляющих с частотами $\omega_0 \pm \Omega_i$.

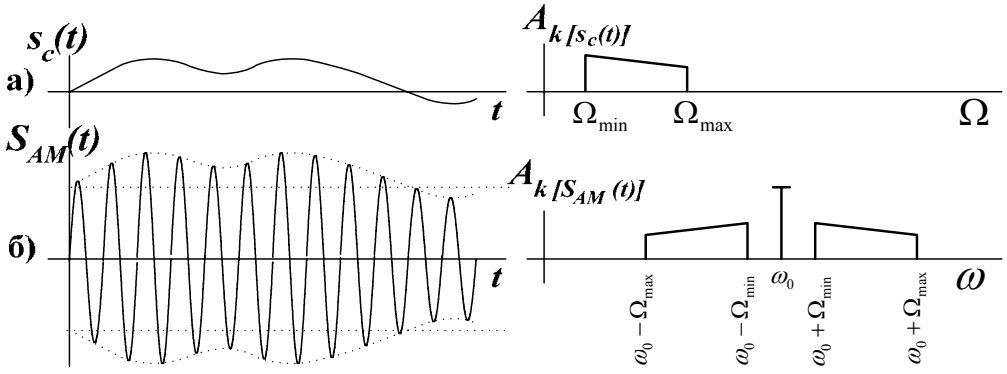


Рис. 2.3. Временные и спектральные диаграммы АМ сигнала

Множеству гармонических составляющих в модулирующем сигнале с частотами $\Omega_i, i = 1,2,...,N$ будет соответствовать множество боковых составляющих с частотами $\omega_0 \pm \Omega_i, i = 1,2,...,N$. Для наглядности такое преобразование спек-

тра при АМ показано на рис. 2.3,б. Спектр сложномодулированного АМ сигнала, помимо несущего колебания с частотой ω_0, содержит группы верхних и нижних боковых колебаний, образующих соответственно верхнюю боковую полосу и нижнюю боковую полосу АМ сигнала.

При этом верхняя боковая полоса частот является масштабной копией спектра информационного сигнала, сдвинутого в область высоких частот на величину ω_0. Нижняя боковая полоса частот также повторяет спектральную диаграмму сигнала $s_c(t)$, но частоты в ней располагаются в зеркальном порядке относительно несущей частоты ω_0.

Ширина спектра АМ сигнала $\Delta\omega_{AM}$ равна удвоенному значению наиболее высокой частоты $\Omega_{\max}$ спектра модулирующего низкочастотного сигнала, т. е. $\Delta\omega_{AM} = 2\Omega_{\max}$.

Наличие двух боковых полос обусловливает расширение занимаемой полосы частот примерно в два раза, по сравнению со спектром информационного сигнала. Мощность, приходящаяся на колебание несущей частоты, постоянна. Мощность, заключенная в боковых полосах, зависит от индекса модуляции и увеличивается с увеличением глубины модуляции. Однако даже в крайнем случае, когда $m_{AM} = 1$, только $1/3$ всей мощности колебания приходится на две боковые полосы.

2.1.2. Балансная и однополосная модуляция гармонической несущей

Балансная модуляция

Анализ спектрального состава АМ сигнала показал, что первичный модулирующий сигнал находит свое отображение лишь в составляющих боковых полос спектра АМ сигнала. В процессе отображения первичного сигнала в модулированном колебании составляющая спектра частоты ω_0 выполняет лишь роль своеобразного начала отсчета для частот боковых спектральных составляющих. Поэтому ее можно исключить из спектра передаваемого сигнала и

восстановить па приемном конце.

Если модулированное колебание не содержит составляющей несущей частоты ω_0, то модуляцию называют балансной (БМ). Такой вид модуляции целесообразен с энергетической точки зрения, поскольку на несущую приходится $2/3$ всей мощности модулированного колебания. При прочих равных условиях высвободившаяся мощность позволит реализовать большую дальность связи, либо при прежней дальности улучшить ее качество.

Однополосная модуляция

Балансная модуляция позволяет более рационально распределить энергию сигнала, однако ширина спектра $\Delta\Omega_{БМ}$ остается такой же, как и для обычной амплитудной модуляции. В то же время симметрия спектра АМ сигнала означает, что верхняя боковая полоса и нижняя боковая полоса каждая в отдельности, полностью отображают модулирующее колебание. При этом вторая боковая полоса не несет никакой дополнительной информации, вдвое расширяя спектр. Вид модуляции, при котором в спектре амплитудно-модулированного колебания сохраняется лишь одна боковая полоса (верхняя или нижняя), называется однополосной модуляцией.

2.2. Методы угловой модуляции

При фазовой и частотной модуляции сигнал имеет постоянную амплитуду и может быть записан в следующем виде:

$$S_{ФМ(ЧМ)}(t) = A_0 \cos\varphi(t). \qquad (2.8)$$

В отсутствие модуляции аргумент гармонического колебания мгновенная (полная) фаза $\varphi(t) = \omega_0 t$ изменяется с постоянной скоростью ω_0, т.е. является линейной функцией времени. И фазовая, и частотная модуляция предполагают зависимость изменения фазы $\varphi(t)$ от информационного сигнала $s_c(t)$. Эта общность позволяет объединить оба вида модуляции одним названием – угловая модуляция.

При угловой модуляции линейность изменения $\varphi(t)$ нарушается и в каж-

дый момент времени ***t*** скорость изменения $\varphi(t)$ определяется мгновенной частотой $\omega(t)$, причем:

$$\omega(t)=\frac{d\varphi(t)}{dt};\qquad\qquad \varphi(t)=\int_0^t \omega(t)dt\ .$$

2.2.1. Принципы частотной и фазовой (угловой) модуляции

Фазовая модуляция – процесс изменения мгновенной фазы несущего колебания пропорционально изменению непрерывного информационного сигнала:

$$\varphi(t)=\omega_0 t+\Delta\varphi(t)=\omega_0 t+as_c(t). \qquad (2.9)$$

Таким образом

$$S_{ФМ}(t)=A_0\cos[\omega_0 t+as_c(t)]. \qquad (2.10)$$

Максимальное отклонение фазы называется индексом модуляции:

$$a|s_c(t)|_{\max}=m_{ФМ}\ . \qquad (2.11)$$

Если модуляция осуществляется гармоническим колебанием (тональная модуляция) $s_c(t)=A_{0\Omega}\cos\Omega t$ с частотой Ω, то

$$S_{ФМ}(t)=A_0\cos(\omega_0 t+aA_{0\Omega}\cos\Omega t)=A_0\cos(\omega_0 t+m_{ФМ}\cos\Omega t).$$

Заметим, что индекс модуляции $m_{ФМ}=aA_{0\Omega}$ пропорционален амплитуде модулирующего колебания.

На рис. 2.4 показано, как изменяются мгновенная частота и фаза при тональной фазовой модуляции.

Информационный однотональный сигнал $s_c(t)=A_{0\Omega}\cos\Omega t$ (рис.2.4,а) модулирует несущее колебание $s_н(t)$ (рис.2.4,б), при этом закон изменения мгновенной фазы несущего колебания $\varphi(t)=\omega_0 t+as_c(t)$ повторяет закон изменения $s_c(t)$ «косинус» (рис.2.4,в), т.е. на линейное изменение фазы (пунктир на рисунке) накладывается переменное приращение $\Delta\varphi(t)=as_c(t)$, а закон изменения мгновенной частоты несущего колебания $\omega(t)$ (рис.2.4,г) определяется производной:

$$\omega(t)=\frac{d\varphi(t)}{dt}=\frac{d}{dt}(\omega_0 t+aA_{0\Omega}\cos\Omega t)=\omega_0-aA_{0\Omega}\sin\Omega t\ .$$

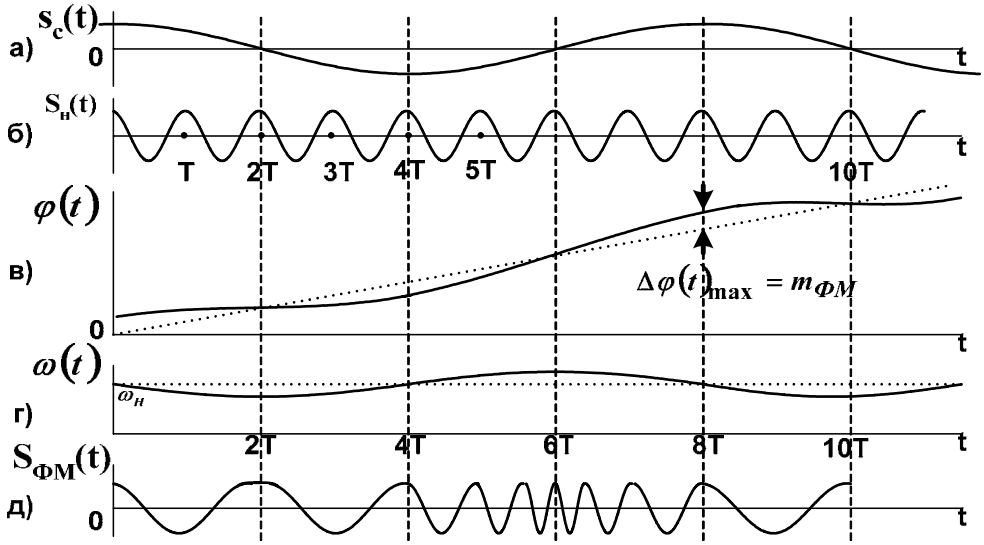


Рис. 2.4. Временные диаграммы процесса формирования ФМ сигналов

Фазомодулированное колебание (рис.2.4,д) построено на основании графика $\omega(t)$; в моменты времени $t = 2T$ и $t = 10T$ сигнал $S_{ФМ}(t)$ имеет минимальную, а в момент $t = 6T$ максимальную мгновенную частоту.

Частотная модуляция – процесс изменения мгновенной частоты несущего колебания в соответствии с изменением информационного сигнала:

$$\omega(t) = \omega_0 + aS_c(t).$$

Рассмотрим наиболее простой способ однотональной частотной модуляции.

На рис. 2.5 изображены временные диаграммы изменения мгновенной частоты и фазы для однотональной частотной модуляции.

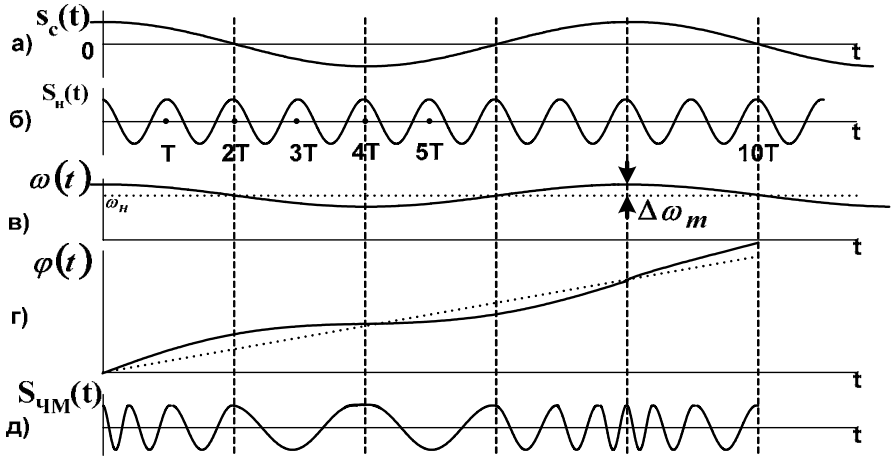


Рис. 2.5. Временные диаграммы процесса формирования ЧМ сигналов

Информационный однотональный сигнал $s_c(t) = A_0 \cos \Omega t$ (рис.2.5,а) модулирует несущее колебание $s_н(t)$ (рис.2.5,б), при этом закон изменения мгновенной частоты несущего колебания $\omega(t) = \omega_0 + aA_0 \cos \Omega t$ повторяет закон изменения $s_c(t)$ (рис.2.5,в). Здесь $aA_{0\Omega} = \Delta\omega_m$ – девиация частоты, пропорциональная амплитуде модулирующего колебания $A_{0\Omega}$. Девиацией частоты называется максимальное отклонение частоты от среднего значения ω_0:

$$a|s_c(t)|_{\max} = \Delta\omega_m \qquad (2.12)$$

Отношение девиации частоты $\Delta\omega_m$ к частоте модулирующего колебания Ω называется индексом частотной модуляции:

$$m_{ЧМ} = \frac{\Delta\omega_m}{\Omega}. \qquad (2.13)$$

В моменты времени $t = 0$, $t = 8T$ мгновенная частота максимальна, в момент $t = 4T$ – минимальна. Закон изменения мгновенной фазы несущего колебания $\varphi(t)$ (рис.2.5,г) определяется интегрированием

$$\varphi(t) = \int_0^t \omega(t)dt \qquad \varphi(t) = \omega_0 t + m_{ЧМ} \sin \Omega t .$$

Учитывая связь частоты и фазы, выражение для частотно-модулированного сигнала запишется следующим образом:

$$S_{ЧМ}(t) = A_0 \cos\left[\int_0^t \omega(t)dt\right] = A_0 \cos\left[\omega_0 t + a\int_0^t s_c(t)dt\right]. \qquad (2.14)$$

Для тональной частотной модуляции формула (2.14) принимает вид

$$S_{ЧМ}(t) = A_0 \cos(\omega_0 t + m_{ЧМ} \sin \Omega t). \qquad (2.15)$$

Сравнение выражений (2.10) и (2.14) показывает, что при ФМ приращение фазы пропорционально модулирующему колебанию $s_c(t)$, а при ЧМ - интегралу от $s_c(t)$. Если сначала проинтегрировать $s_c(t)$, а затем этим колебанием модулировать несущую по фазе, то получится ЧМ сигнал. Такой способ формирования ЧМ сигнала применяется практически. Подобным же образом, если продифференцировать $s_c(t)$ и это колебание использовать для модуляции частоты, то получим ФМ сигнал.

2.2.2. Спектр сигналов угловой модуляции

Сигналы с угловой модуляцией, как и при АМ, могут быть представлены в виде суммы гармонических колебаний. Сравнительно просто это можно сделать для тональной модуляции. При тональной модуляции спектры ФМ и ЧМ одинаковы, если $m_{ФМ} = m_{ЧМ} = m$, поэтому будем рассматривать только спектр ЧМ сигнала.

Преобразуем (2.15) по формуле косинуса суммы двух аргументов:

$$\begin{aligned} S_{ЧМ}(t) &= A_0 \cos(\omega_0 t + m\sin\Omega t) = A_0 \cos(\omega_0 t)\cdot\cos(m\sin\Omega t) - \\ &- A_0 \sin(\omega_0 t)\cdot\sin(m\sin\Omega t) \end{aligned} . \qquad (2.16)$$

Из теории бесселевых функций [21, 32] известны следующие соотношения:

$$\begin{aligned} \cos(m\sin x) &= J_0(m) + 2\sum_{k=1}^{\infty} J_{2k}(m)\cos 2kx; \\ \sin(m\sin x) &= 2\sum_{k=1}^{\infty} J_{2k-1}(m)\sin(2k-1)x, \end{aligned} \qquad (2.17)$$

где $J_k(m)$ – функция Бесселя k-го порядка от аргумента m. Подставляя (2.17) в (2.16), выполняя обычные алгебраические преобразования и раскрывая произведение тригонометрических функций, получаем:

$$\begin{aligned} S_{ЧМ}(t) &= A_0 J_0(m)\cos(\omega_0 t) + \sum_{k=1}^{\infty} A_0 J_k(m)\cos(\omega_0 + k\Omega)t + \\ &+ \sum_{k=1}^{\infty} (-1)^k A_0 J_k(m)\cos(\omega_0 - k\Omega)t \end{aligned} . \qquad (2.18)$$

Таким образом, спектр даже для однотональной угловой модуляции является довольно сложным. В формуле (2.18) первый член – гармоническая составляющая с частотой несущей. Группа гармонических составляющих с частотами $(\omega_0 + k\Omega), k = 1,2,...,$ определяет верхнюю боковую полосу частот, а группа составляющих с частотами $(\omega_0 - k\Omega), k = 1,2,...,$ нижнюю боковую полосу частот. Число верхних и нижних гармоник боковых частот теоретически бесконечно. Боковые гармонические колебания расположены симметрично относительно ω_0 на расстоянии Ω. Амплитуды всех компонент спектра, в том числе и с частотой ω_0, пропорциональны значениям функций Бесселя $J_k(m)$.

Формулу (2.18) можно представить в более компактном виде. Действительно учитывая $(-1)^k J_k(m) = J_k(m)$, получаем:

$$S_{ЧМ}(t) = A_0 \sum_{k=-\infty}^{\infty} J_k(m)\cos(\omega_0 + k\Omega)t\,. \qquad (2.19)$$

Для построения спектральных диаграмм необходимо знание функций Бесселя $J_k(m)$ при различных значениях k и m. Эти сведения имеются в математических справочниках [21, 32]. На рис. 2.6 приведены графики функций Бесселя при $k = 0,1,...,7$. Значения функций Бесселя, отсутствующих на графиках, можно найти по рекуррентной формуле:

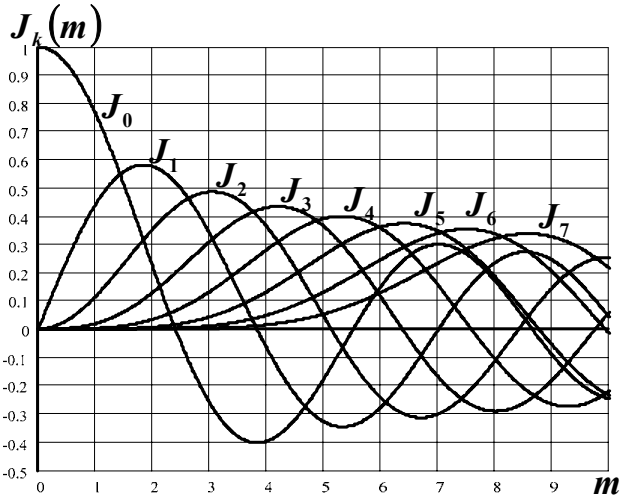


Рис. 2.6. Графики функций Бесселя

$$J_{k+1}(m) = \left(\frac{2k}{m}\right) J_k(m) - J_{k-1}(m)\,.$$

Пример 2.1. Задано аналитическое выражение модулированного сигнала $S(t) = 10\cos\left(2\cdot 10^6 t + 3\cos 10^4 t\right)$. Построить спектральную диаграмму этого сигнала.

Из математического уравнения сигнала следует, что это однотональная угловая модуляция с индексом $m = 3$. Спектральные составляющие сигнала определяем из уравнения (2.18), приняв $k = 0,1,2,3,...$, до тех пор, пока амплитуда составляющих не будет заданной, например меньше 2% от A_0. По результатам расчетов построена спектральная диаграмма (рис. 2.7).

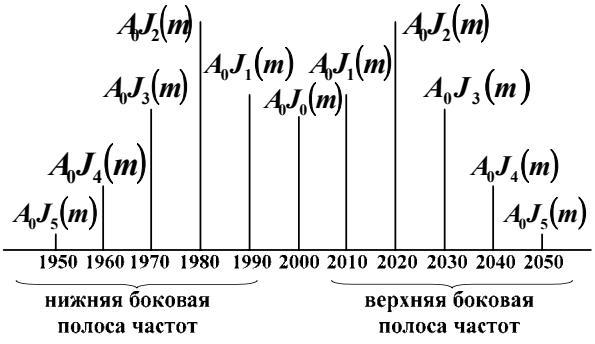


Рис. 2.7. Спектральная диаграмма сигналов с однотональной угловой модуляцией при m=3

Анализ графиков функций Бесселя показывает, чем больше порядок k функции Бесселя, тем при больших аргументах m наблюдается ее максимум, однако при $k > m$ значения функций Бесселя оказываются малой

величиной. Следовательно, малыми будут и соответствующие составляющие спектра; ими можно пренебречь. Поэтому ширину спектра сигналов с угловой модуляцией можно приближенно определить по формуле:

$$\Delta\omega_{УМ} \approx 2(m+1)\Omega, \qquad (2.20)$$

где Ω – частота модулирующего сигнала. Для передачи модулированного сигнала с высокой точностью иногда считают, что надо учитывать спектральные составляющие с уровнем не менее 1% от уровня несущей. Тогда ширина спектра с угловой модуляцией $\Delta\omega_{УМ} \approx 2\left(m+\sqrt{m}+1\right)F$ [21, 32, 39].

Если $m < 0{,}6$, то ширина спектра угловой модуляции соизмерима с шириной спектра амплитудной модуляции. Если $m >> 1$ то при угловой модуляция из (2.20) и (2.14) следует, что ширина полосы частот примерно равна удвоенной девиации частоты.

2.3. Формирование и детектирование модулированных сигналов

2.3.1. Формирование и детектирование сигналов амплитудной и однополосной амплитудной модуляции

Устройства формирования и демодуляции радиосигналов могут быть различными в зависимости от применяемых активных элементов, способа подачи на них несущей и модулирующего сигнала.

Рассмотрим основные принципы построения модуляторов для АМ сигналов. При входном сигнале относительно небольшой мощности одним из методов формирования АМ сигналов $S_{АМ}(t)$ является операция перемножения двух колебаний: информационного сигнала $s_{ПЭС}(t)$ и несущего колебания $S_н(t)$, где в качестве перемножителя может использоваться специальная микросхема.

Операция амплитудного детектирования противоположна амплитудной модуляции. Если детектирование АМ колебания производится без опорного напряжения, то в качестве преобразующего используется нелинейный элемент (НЭ). Второй необходимый элемент детектора – фильтр низких частот (ФНЧ),

который осуществляет подавление спектральных составляющих кратных несущей частоте (рис. 2.8).

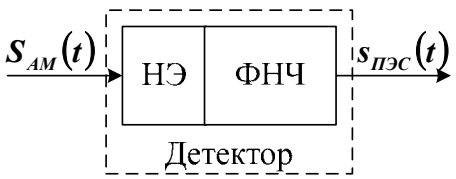


Рис. 2.8. Амплитудный детектор

Нелинейный элемент амплитудного детектора выполняется обычно на полупроводниковом диоде. В зависимости от амплитуды АМ сигнала и степени нелинейности характеристик НЭ возможны два режима детектирования: квадратичное детектирование при малых амплитудах входного сигнала и линейное детектирование – в режиме больших амплитуд.

Наряду с амплитудным, используется синхронное детектирование, которое основано на перемножении АМ сигнала и колебаний опорного генератора, совпадающего по частоте и фазе с несущей АМ сигнала, с последующим выделением низкочастотных составляющих с помощью ФНЧ. Главная трудность при синхронном детектировании заключается в получении синфазного с несущей опорного колебания. Такое колебание формируется с помощью системы фазовой автоподстройки частоты (ФАПЧ) (рис.2.9).

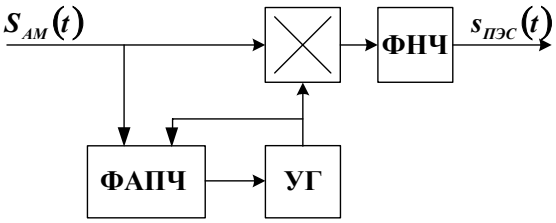


Рис. 2.9. Синхронный детектор

В системе ФАПЧ осуществляется сравнение фазы входного сигнала и управляемого генератора (УГ). При наличии отклонений осуществляется изменение фазы УГ до тех пор, пока не обеспечится синфазность входного сигнала и УГ.

При синхронном детектировании сигналов с балансной и однополосной модуляцией возникают принципиальные трудности в получении синфазного опорного напряжения. Это связано с тем, что в спектре этих сигналов несущая отсутствует. Находят применение два технических решения.

В первом случае вместе с балансной или однополосной модулированными сигналами передается так называемый пилот-сигнал, представляющий собой остаток несущей. Пилот-сигнал используется в приемнике для системы ФАПЧ опорного генератора.

Второе решение заключается в том, что для детектирования используется

высокостабильный генератор несущей, отличающейся по частоте от передаваемой несущей.

В результате несинхронности несущего и опорного колебании в системах передачи с однополосной модуляцией спектр восстановленного сигнала смещается. Эти искажения снижают качество передачи первичных сигналов. Однако, как показывают экспериментальные исследования, небольшой частотный сдвиг заметного влияния на качество не оказывает. При телефонной связи абонент практически не замечает сдвига частот до 10 ... 20 Гц. При передаче радиовещательных программ допустимым является сдвиг частот до 2 Гц. Примерно такой же сдвиг(±1Гц) не сказывается на качестве факсимильной связи. Тем не менее отсюда следуют весьма жесткие требования к стабильности генераторного оборудования систем связи с ОМ.

2.3.2. Формирование и детектирование сигналов угловой модуляции

Частотная модуляция (ЧМ) является основным видом модуляции в современных системах передачи информации СВЧ диапазона, в том числе системах спутниковой радиосвязи и телевидения. При ЧМ обеспечивается высокая помехоустойчивость и высокое качество передачи информации, допускается возможность одновременной работы в общем канале связи большого числа корреспондентов и реализуется более полное использование по энергетическим показателям радиопередающего устройства в силу постоянства амплитуды сигнала по сравнению с амплитудной модуляцией.

Способы осуществления частотной и фазовой модуляции можно разделить на две группы: прямые и косвенные (рис. 2.10).

Прямой метод при ЧМ означает непосредственное воздействие на автогенератор или, точнее, - на колебательную систему, определяющую частоту колебаний. Косвенный метод ЧМ состоит в преобразовании фазовой модуляции в частотную [5, 21].

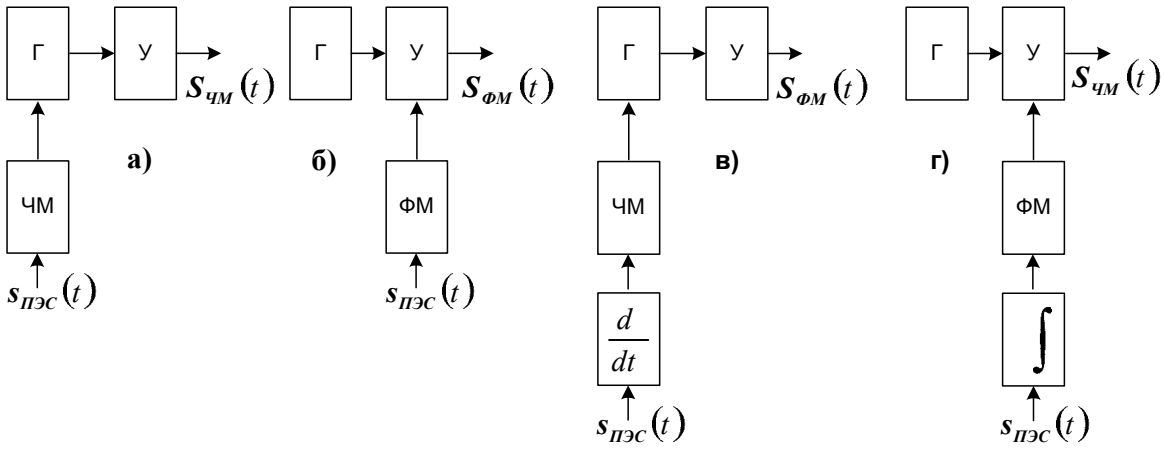


Рис. 2.10. Прямые (а,б) и косвенные (в,г) методы частотной и фазовой модуляции

Прямой метод при ФМ означает воздействие на высокочастотный усилитель или умножитель частоты, т. е. на электрические цепи, определяющие фазу высокочастотных колебаний. Косвенный метод ФМ заключается в преобразовании частотной модуляции в фазовую.

Для преобразования фазовой модуляции в частотную на входе фазового модулятора включается интегратор (рис. 2.10, г). Для преобразования частотной модуляции в фазовую на входе частотного модулятора включается дифференцирующая цепь (рис. 2.10, в).

В зависимости от характера преобразований различают частотно-амплитудные, частотно-фазовые и частотно-импульсные детекторы.

В частотно-амплитудных детекторах изменение частоты сигнала преобразуется в изменение амплитуды которое затем выделяется амплитудным детектором. Для того, чтобы на выходе детектора не возникли искажения за счет возможных изменений амплитуды входного напряжения, перед детектированием обычно производят ограничение.

В частотно-фазовых детекторах изменение частоты преобразуется в изменение фазового сдвига между двумя напряжениями с последующим фазовым детектированием.

Фазовые детекторы преобразуют входной фазомодулированный сигнал в выходное напряжение, изменяющееся по закону модулирующего сигнала. Выявить фазовый сдвиг в ФМ сигнале можно путем сравнения с когерентным не-

модулированным колебанием, которое называют опорным.

Структурная схема фазового детектора аналогична схеме синхронного детектора (рис.2.9). Все фазовые детекторы различаются по типу используемого перемножителя, наличию или отсутствию ограничителя и методам создания опорного напряжения. В качестве перемножителей можно использовать любые нелинейные или параметрические элементы – диоды, транзисторы, дифференциальные и операционные усилители с управляемой обратной связью, специальные аналоговые перемножители, ключевые схемы и др.

2.4. Манипуляция сигналов

При дискретном изменении управляющего колебания модулируемые параметры несущей будут изменяться скачком. В этом случае вместо термина «модуляция» применяется термин «манипуляция», а само колебание называется манипулированным. В частности манипуляция – это модуляция несущего колебания посылками постоянного тока прямоугольной формы.

Дискретное манипулирующее колебание может иметь вид униполярных (рис.2.12,б) или биполярных (рис.2.12,в) прямоугольных импульсов. Для описания двух возможных состояний широко используются термины «посылка» и «пауза». Эти состояния обозначают обычно символами +1 и -1 или 1 и 0.

2.4.1. Временные и спектральные характеристики амплитудно-манипулированных сигналов

Амплитудной манипуляцией (АМн) называется процесс изменения амплитуды несущего (высокочастотного, манипулируемого) колебания в соответствии с законом изменения амплитуды дискретного информационного (первичного электрического, манипулирующего) сигнала.

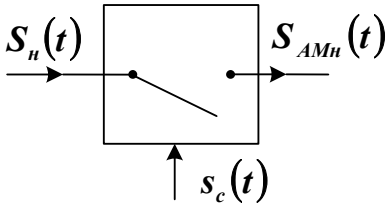


Рис. 2.11. Структурная схема амплитудного модулятора

Структурную схему получения АМн сигнала можно представить как ключ, управляемый пер-

вичным сигналом $s_c(t)$, на вход которого поступает несущий сигнал $S_н(t)$ (рис. 2.11). При этом первичный сигнал можно представить в виде отрезка ряда Фурье:

$s_c(t) = \sum_{k=1}^{\infty} A_k \cdot \cos(\Omega_k t + \varphi_k)$ – сигнал (рис. 2.12,а), а несущий сигнал

$S_н(t) = A_m \cdot \cos(\omega_н t + \varphi_0)$ (рис. 2. 12,б).

Амплитудно-манипулированный сигнал имеет вид последовательности радиоимпульсов с прямоугольной огибающей (рис. 2. 12, в). Единичные элементы с длительностью интервалов $\tau_И$, соответствующих символам кодовой комбинации (1 и 0 или +1 и-1), преобразуются к виду [21, 39]:

$$S_{АМн}(t) = \frac{1}{2} A_m \cdot [1 + x_c(t)] \cdot \cos(\omega_н t + \varphi_0), \qquad (2.21)$$

где $x_c(t)$ – нормированная функция, повторяющая закон изменения $S_c(t)$ (рис. 2. 12, а) и принимающая значения ± 1.

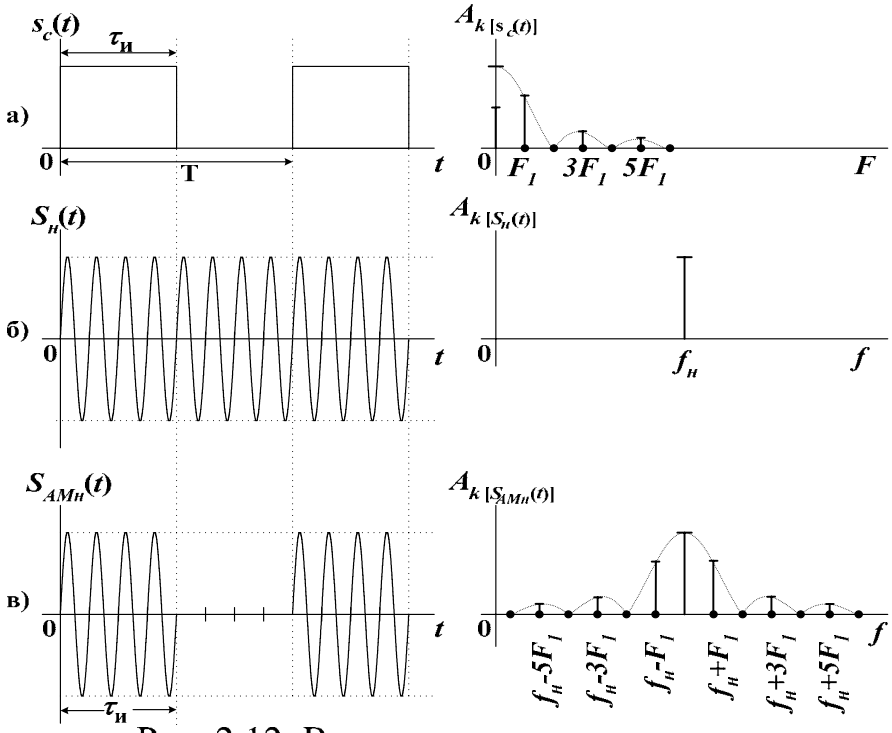


Рис. 2.12. Временные и спектральные характеристики формирования АМн сигнала

Спектральный состав периодической последовательности АМн сигналов определяется следующим выражением [21, 39]:

$$S_{AMн}(t)=\frac{A_m\cdot\tau_u}{T}\left[1+2\sum_{k=1}^{\infty}A_k\cdot\cos(k2\pi F_1t)\right]\cdot\cos(2\pi f_нt)=\frac{A_m\cdot\tau_u}{T}\cdot\cos(2\pi f_нt)+$$
$$+\frac{A_m\cdot\tau_u}{T}\sum_{k=1}^{\infty}\left|\frac{\sin(k\pi F_1\tau_u)}{k\pi F_1\tau_u}\right|\times\left[\cos(f_н+kF_1)2\pi t+\cos(f_н-kF_1)2\pi t\right]. \qquad (2.22)$$

Спектр модулированного сигнала содержит в своем составе:

составляющую с амплитудой $\frac{A_m\cdot\tau_u}{T}$ на несущей частоте $f_н$ и две симметричные боковые полосы с частотами составляющих $(f_н+kF_1)$; $(f_н-kF_1)$ и амплитудами $\frac{A_m\cdot\tau_u}{T}\cdot\left|\frac{\sin(k\pi F_1\tau_u)}{k\pi F_1\tau_u}\right|$.

Для периодических сигналов – спектр дискретный, а при случайном следовании кодовых символов (непериодических сигналов) – спектр становится сплошным.

Ширина спектра АМн колебания: $\Delta F_{AMн}=2kF_1$,

где k – номер учитываемой гармоники;

$F_1=1/T$ – частота первой гармоники информационного сигнала.

В реальных каналах ширину спектра берут с учетом третьей или пятой гармоники, например при необходимости передать цифровой сигнал со скоростью $V=50$ Бод, ширина спектра $\Delta F_{AMн}=2\cdot5\cdot F_1=5\cdot V=250$ Гц.

В настоящее время двоичная амплитудная манипуляция используется в низкоскоростных системах передачи информации, в многоканальных системах связи с временным разделением, в радиолокационных системах, а также в ряде оптических систем.

2.4.2. Временные и спектральные характеристики частотно-манипулированных сигналов

При частотной манипуляции (ЧМн) частота высокочастотного колебания изменяется скачком на величину $\pm\Delta f_m$ относительно несущей $f_н$ (рис. 2.13). Таким образом, на выходе ЧМн вырабатываются колебания на частотах f_1 и f_2. Разность частот $f_2-f_1=\Delta f_{сдв}$ называют частотным сдвигом. Максимальное от-

клонение частоты Δf_m от несущей называют девиацией.

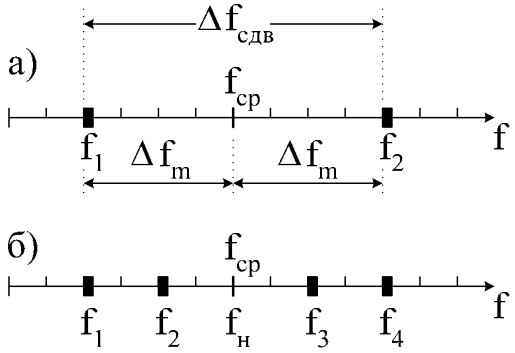


Рис. 2.13. Параметры сигналов ЧМн

Отношение девиации частоты Δf_m к частоте манипулирующего колебания F называется индексом частотной манипуляции. Индекс ЧМн прямо пропорционален девиации и обратно пропорционален частоте информационного сигнала: $m_{ЧМн} = \frac{\Delta f_m}{F}$.

Различают частотную манипуляцию: с разрывом фазы и без разрыва фазы. Общий вид ЧМн сигнала с разрывом фазы можно представить в виде суммы двух АМн сигналов с разными несущими частотами f_1 и f_2. Технически такой вид манипуляции реализуется с помощью двух генераторов (рис. 2.14), которые управляются ключом под воздействием информационного сигнала: $S_{ЧМн}(t) = S_{1АМн}(t) + S_{2АМн}(t)$.

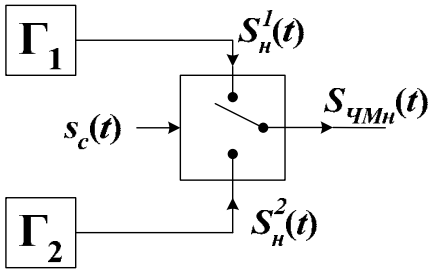


Рис. 2.14. Структурная схема формирования ЧМн колебаний с разрывом фазы

Это представление позволяет спектр колебания $S_{ЧМн}(t)$ найти как результат наложения двух спектров колебаний АМн, который будет иметь вид [32]:

$$S_{ЧМн}(t) = \frac{A_m \cdot \tau_u}{T} \cdot \cos(2\pi f_1 t) + A_m \cdot \left(1 - \frac{\tau_u}{T}\right) \cdot \cos(2\pi f_2 t) + \frac{A_m \cdot \tau_u}{T} \cdot \sum_{k=1}^{\infty} \frac{\sin(k\pi F_1 \tau_u)}{k\pi F_1 \tau_u} \times$$
$$\times \left[\cos(f_1 + kF_1)2\pi t + \cos(f_1 - kF_1)2\pi t - \cos(f_2 + kF_1)2\pi t - \cos(f_2 - kF_1)2\pi t\right]. \quad (2.23)$$

Первое слагаемое определяет составляющую на частоте f_1, второе - на частоте f_2. Формирование ЧМн сигнала с разрывом фазы показано на рис. 2.15.

Из рис. 2.15 видно, что ширина спектра ЧМн сигнала отличается от спектра сигнала АМн на величину $2\Delta f_m$: $\Delta F_{ЧМн} = 2kF_1 + 2\Delta f_m$, где k – номер учитываемой гармоники.

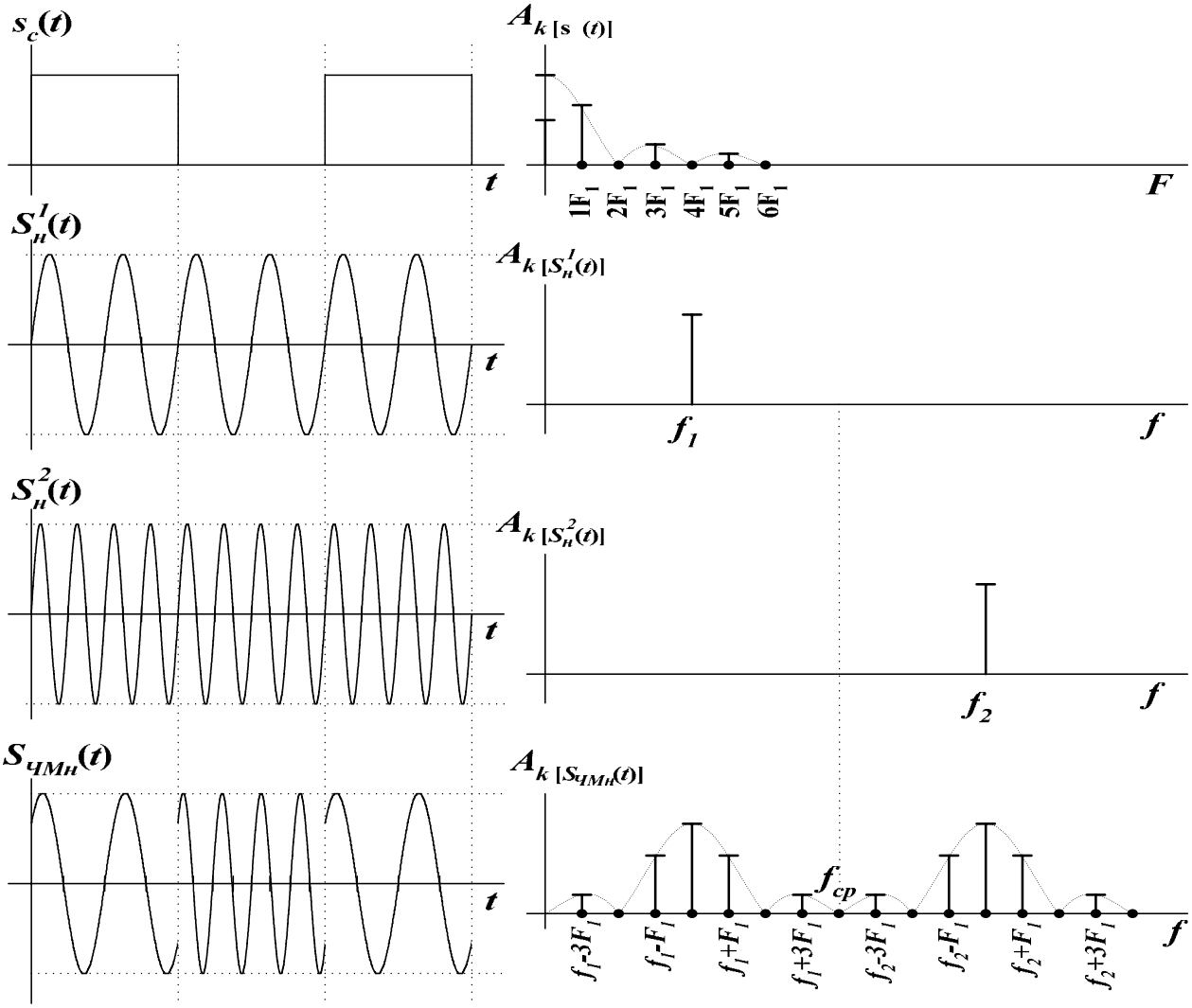


Рис. 2.15. Временные и спектральные характеристики формирования ЧМн сигнала с разрывом фазы

Например при необходимости передать цифровой сигнал со скоростью $V = 75$ бит/с, $\Delta f_m = 250$ Гц, $k = 3$, ширина спектра $\Delta F_{ЧМн} = 2 \cdot 3 \cdot \frac{75}{2} + 2 \cdot 250 = 725$ Гц.

Общий вид ЧМн сигнала без разрыва фазы (рис.2.16) можно записать в виде [32]: $S_{ЧМн}(t) = A_m \cos[\omega_н t + \Delta\varphi(t)]$,

где $\Delta\varphi(t)$ – приращение фазы, обусловленное приращением частоты $\Delta\omega(t)$.

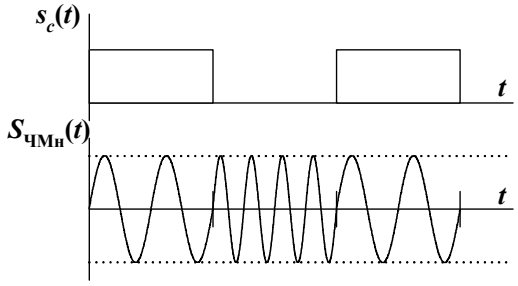


Рис. 2.16. Временные характеристики формирования ЧМн колебаний без разрыва фазы

Этот вид манипуляции предполагает использовать один источник колебаний (рис. 2.17.), частота которого изменяется посредством управляемой реактивности (в этом случае фаза изменяется непрерывно – без разрыва).

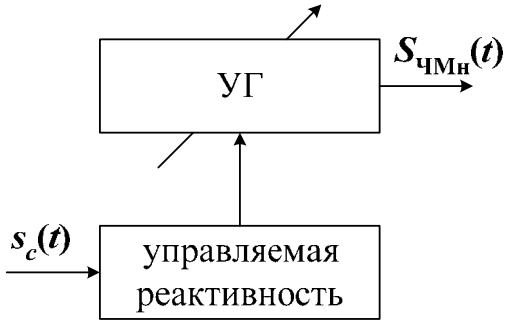


Рис. 2.17. Структурная схема формирования
ЧМн колебаний без разрыва фазы

Спектральный состав ЧМн сигнала без разрыва фазы можно получить, раскрывая выражение для $S_{ЧМн}(t)$;

$$S_{ЧМн}(t) = A_m\left[\cos\omega_н t \cdot \cos\Delta\varphi(t) - \sin\omega_н t \cdot \sin\Delta\varphi(t)\right].$$

Из этой формулы следует, что для нахождения спектра ЧМн сигнала необходимо определить спектр функций $\cos\Delta\varphi(t)$ и $\sin\Delta\varphi(t)$ разложив их в ряд Фурье:

$$S_{ЧМн}(t) = \frac{2A_m \cdot m}{\pi} \cdot \sum_{k=-\infty}^{\infty} \frac{\sin\left[0{,}5\pi(m+k)\right]}{\left(m^2-k^2\right)} \cdot \cos\left(f_н + kF_1\right)2\pi t\,. \qquad (2.24)$$

Из спектральной характеристики (рис. 2.24) видно, что для спектра при $m_{ЧМн} << 1$ энергия колебания находится вблизи $f_н$. Спектр ограничен несущей и двумя боковыми частотами, а ширина спектра равна ширине спектра АМн сигнала [21, 32, 39]:

$$\begin{aligned} S_{ЧМн}(t) &= A_m \cdot \frac{\sin(0{,}5\pi m)}{(0{,}5\pi m)} \cdot \cos\omega_н t + \\ &+ \frac{2A_m \cdot m}{\pi} \cdot \sum_{k=2,4,\ldots}^{\infty} \frac{\sin(0{,}5\pi m)}{\left(m^2-k^2\right)} \cdot \cos(\omega_н + k\Omega)t + \cos(\omega_н - k\Omega)t - \\ &- \frac{2A_m \cdot m}{\pi} \cdot \sum_{k=1,3,\ldots}^{\infty} \frac{\cos(0{,}5\pi m)}{\left(m^2-k^2\right)} \cdot \sin(\omega_н + k\Omega)t + \sin(\omega_н - k\Omega)t. \end{aligned} \qquad (2.25)$$

По мере увеличения индекса частотной модуляции энергия концентрируется вблизи частот f_1 и f_2. На рис. 2.18 приведены спектры колебаний при различных $m_{ЧМн}$.

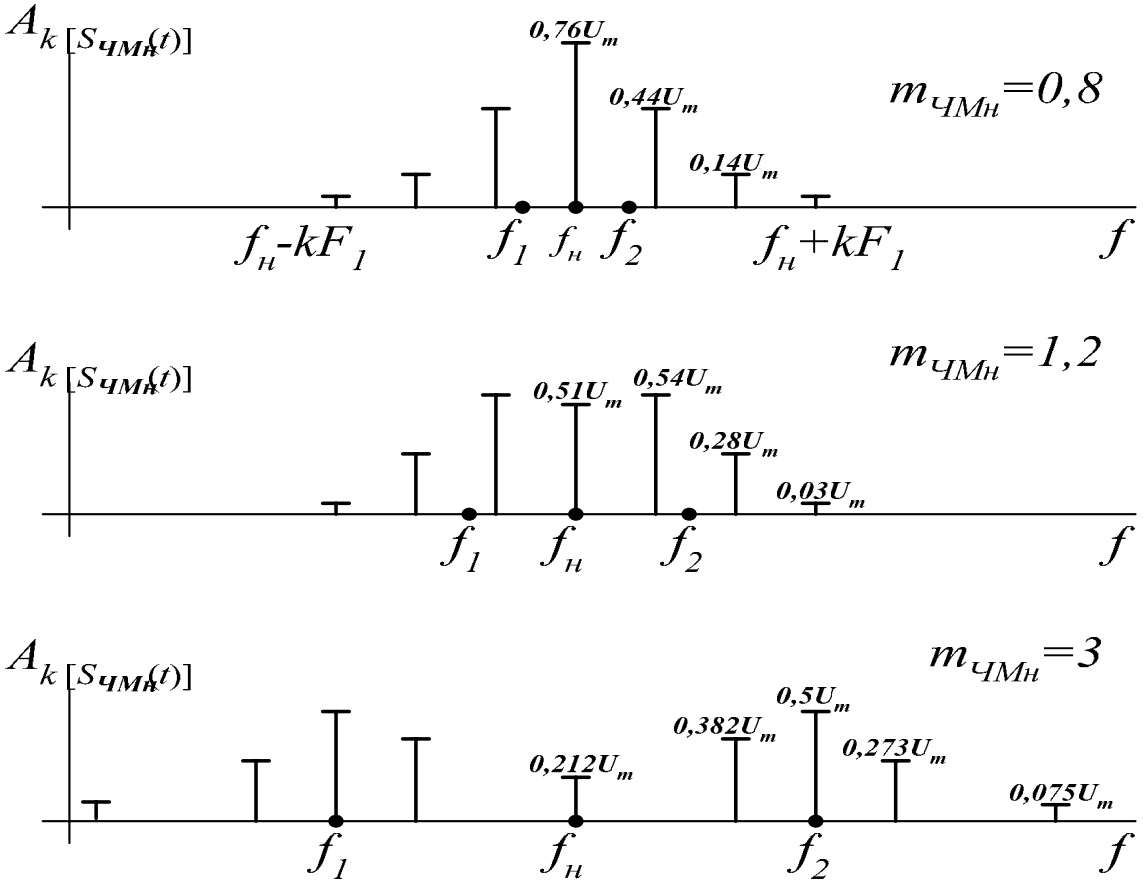


Рис. 2.18. Спектральные характеристики ЧМн сигнала без разрыва фазы для различных индексов модуляций

Ширина спектра определяется по общей формуле [21, 32, 39]:

$$\Delta F_{ЧМн} = 2(\Delta f_m + \Delta F) = 2F(m+2) = 2\Delta f_m(1+{}^{2}\!/_{m}), \tag{2.26}$$

либо по формулам для различных $m_{ЧМн}$:

$$\Delta F_{ЧМн} = \begin{cases} (1{,}3 \cdot m_{ЧМн} + 1{,}4) \cdot V; & 2 \le m_{ЧМн} \le 8 \\ (1{,}1 \cdot m_{ЧМн} + 1{,}6) \cdot V; & 8 \le m_{ЧМн} \le 20 \end{cases}, \tag{2.27}$$

где V – скорость телеграфирования в бодах.

2.4.3. Фазовая (относительно-фазовая) манипуляция сигналов

В настоящее время разработано несколько вариантов двухпозиционной (бинарной) и многопозиционной фазовой манипуляции. В радиосистемах передачи информации наиболее часто применяются двоичная, четырех позиционная и восьми позиционная фазовая манипуляция (ФМн). Данные сигналы обеспечивают высокую скорость передачи, применяются в радиосвязи, в системах фа-

зовой телеграфии, при формировании сложных сигналов.

Временные и спектральные характеристики фазоманипулированных сигналов

Наиболее простой является бинарная ФМн, при которой изменение фазы несущего колебания происходит скачком в определенные моменты первичного сигнала (рис. 2.25, а) на 0 или 180°; при этом его амплитуда и частота несущей остаются неизменными.

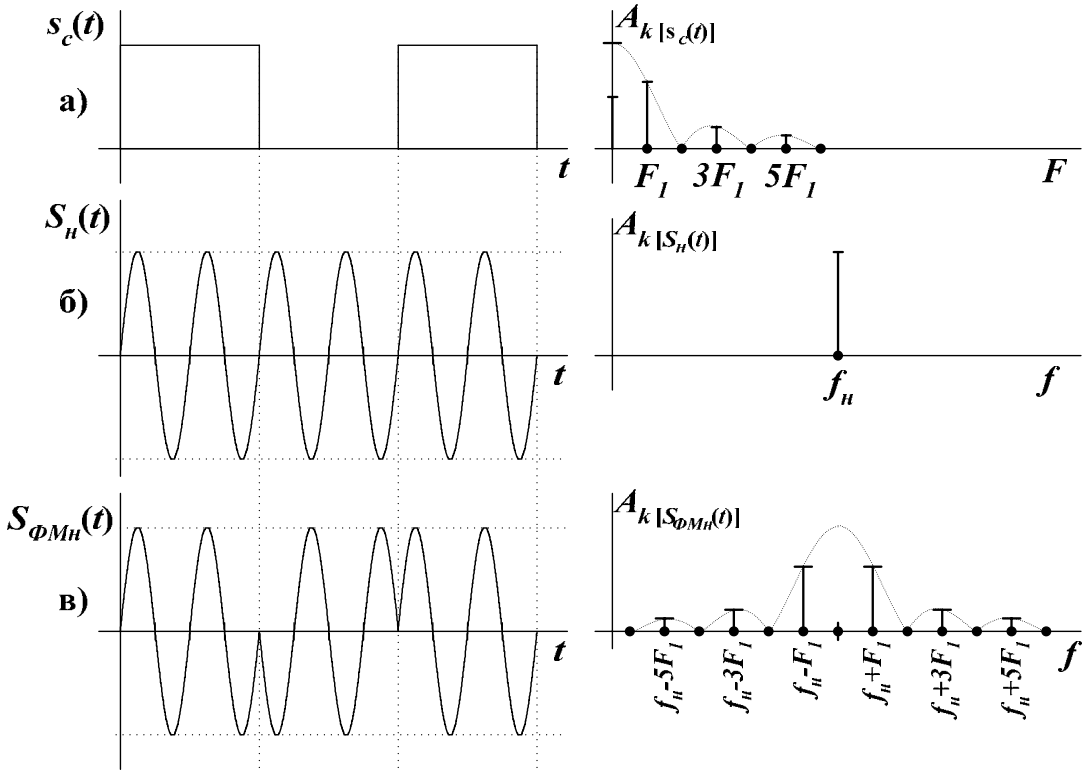


Рис. 2.19. Временные и спектральные характеристики ФМн сигнала

ФМн сигнал имеет вид последовательности радиоимпульсов (отрезков гармонических колебаний) с прямоугольной огибающей (рис. 2.19, в) [32, 39]:

$$S_{ФМн} = A_m \cdot \cos\left[\omega_н t + \left(1 + x_c(t)\right) \cdot \Delta\varphi_m\right], \qquad (2.28)$$

где $x_c(t)$ – нормированная функция, принимающая значения -1 и 1, и повторяющая изменения информационного сигнала (рис. 2.19, а); $\Delta\varphi_m$ – девиация фазы (максимальное отклонение фазы от начальной).

Величина $\Delta\varphi_m$ может быть любой, однако, для лучшего различения двух сигналов на приеме целесообразно, чтобы они максимально отличались друг от

друга по фазе, т.е. на 180^o ($\Delta\varphi_m = \pi$).

Таким образом, одни из ФМн колебаний будут синфазны с колебаниями несущей, а другие противоположны по фазе на 180^o.

Такой сигнал можно представить в виде суммы двух АМн сигналов, с противофазными несущими 0^o и 180^o: $S_{ФМн}(t) = S^1_{АМн}(t) + S^2_{АМн}(t)$.

Структурная схема модулятора в этом случае реализуется с помощью двух самостоятельных источников колебаний (генераторов) с разными начальными фазами, выходы которых управляются информационным сигналом с помощью ключа (рис. 2.20).

Спектр ФМн колебания находится суммированием спектров колебаний $S^1_{АМн}(t)$ и $S^2_{АМн}(t)$ [21, 32, 39]:

$$S_{ФМн}(t) = A_m\left(\frac{2\tau_u}{T} - 1\right)\cdot\cos(2\pi f_н t) + 2A_m\frac{\tau_u}{T}\cdot\sum_{k=1}^{\infty}\frac{\sin(k\pi F_1\tau_u)}{(k\pi F_1\tau_u)}\times$$
$$\times\left[\cos(f_н + kF_1)2\pi t + \cos(f_н - kF_1)2\pi t\right]. \qquad (2.29)$$

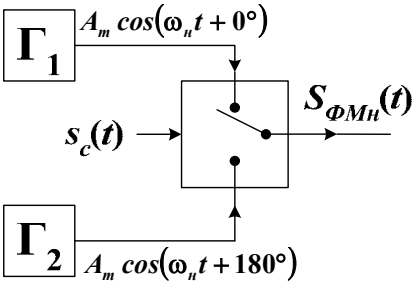


Рис. 2.20. Структурная схема формирования ФМн колебаний

Из формулы следует, что спектр колебаний ФМн в общем случае содержит несущее колебание, верхнюю и нижнюю боковые полосы, состоящие из спектральных составляющих частот $(2\pi f_н \pm k2\pi F_1)t$.

Анализ спектров ФМн сигналов (рис. 2.21) при различных значениях $\Delta\varphi_m$ показывает, что при изменении $\Delta\varphi_m$ от 0 до π происходит перераспределение энергии сигнала между несущим колебанием и боковыми составляющими, а при $\Delta\varphi_m = \pi$ вся энергия сигнала содержится только в боковых полосах. Из рис. 2.21 следует, что спектр амплитуд ФМн сигнала содержит те же составляющие, что и спектр АМн сигнала, а для скважности $\frac{T}{\tau_И} = 2$ составляющая на несущей частоте отсутствует. Амплитуды боковых составляющих ФМн сигнала в 2 раза больше, чем АМн сигнала.

Это объясняется наложением 2-х спектров - спектра ФМн сигнала и не-

сущей. На интервале, где колебания синфазны, суммарная амплитуда удваивается, а где фазы противоположны, компенсируется, в результате для нахождения спектра ФМн достаточно определить спектр АМн колебания.

Равенство полос частот АМн и ФМн сигнала предполагает также и равенство максимально возможных скоростей модуляции. Большая амплитуда спектральных составляющих ФМн сигнала по сравнению с АМн обусловливает большую помехоустойчивость.

При ФМн начальная фаза является информационным параметром, и в алгоритмах работы фазового демодулятора с целью получения сведений о начальной фазе должны формироваться и храниться образцы вариантов передаваемого сигнала, достаточно точно совпадающие с ним по частоте и начальной фазе. Но на приеме нет признаков по которым можно точно установить однозначное соответствие между переданными двоичными символами и образцами сигнала на входе демодулятора, в результате возможно явление так называемой «обратной работы».

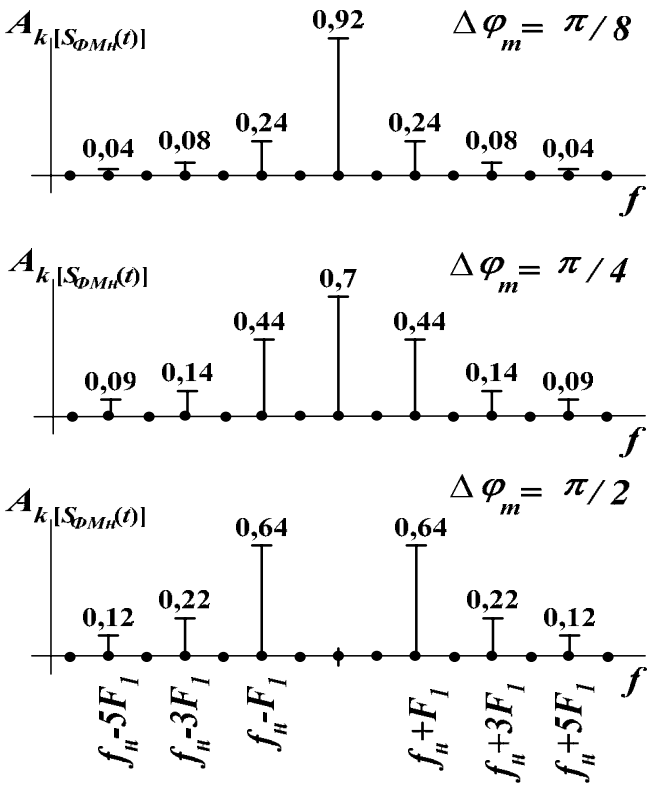


Рис. 2.21. Спектры сигналов фазовой манипуляции при различных значениях девиации фазы

Неопределенность начальной фазы объясняется с одной стороны тем, что в канале связи к переданной фазе добавляется произвольный и неизвестный фазовый сдвиг. С другой стороны, фаза сигнала всегда приводится к интервалу 2π и сигналы, различающиеся по фазе на 2π, для приемника одинаковы.

Данное свойство неоднозначности решения характерно именно для ФМн. При АМн сигнал, прошедший канал связи, также отличается от переданного, однако если на выходе модулятора сигналу с большей амплитудой соответство-

вал некоторый двоичный символ, то и на входе демодулятора варианту сигнала с большей амплитудой будет соответствовать тот же самый символ – неоднозначность отсутствует. При ЧМн ситуация аналогична. Если одна из двух частот больше другой на выходе модулятора, то после всех преобразований в канале она останется больше и на входе демодулятора.

Временные характеристики сигналов с относительной фазовой манипуляцией

Неоднозначность характерная для ФМн сигналов, устранена в системах относительно-фазовой манипуляции (ОФМн). У такого метода манипуляции информация заложена не в абсолютном значении начальной фазы, а в разности начальных фаз соседних посылок, которая остается неизменной и на приемной стороне. Для передачи первого двоичного символа в системах с ОФМн необходима одна дополнительная посылка сигнала, передаваемая перед началом передачи информации и играющая роль отсчетной.

Процесс формирования сигнала с ОФМн можно свести к случаю формирования сигнала с ФМн путем перекодирования передаваемой двоичной последовательности. Алгоритм перекодировки прост: если обозначить $s_c^n = \pm 1$ как информационный символ, подлежащий передаче на n-м единичном элементе сигнала, то перекодированный в соответствии с правилами ОФМн символ $s_{отн}^n$ определяется следующим рекуррентным соотношением: $s_{отн}^n(t) = s_c^n(t) \cdot s_{отн}^{n-1}(t)$. Для получения сигнала с ОФМн достаточно умножить полученный (перекодированный) сигнал $s_{отн}^n$ на несущее колебание. Структурная схема модулятора для ОФМн (рис. 2.22) содержит генератор несущего колебания, перемножитель (ФМ) и перекодирующее устройство (относительный кодер) состоящий из перемножителя и элемента памяти.

Демодулятор сигнала с ОФМн содержит фазовый детектор, состоящий из перемножителя и ФНЧ, на который подается опорное колебание, совпадающее с одним из вариантов принимаемого сигнала. Дальнейшее вычисление разности фаз и определение переданного ПЭС осуществляется перемножением сигналов

на выходе детектора, задержанных друг относительно друга на длительность единичного интервала.

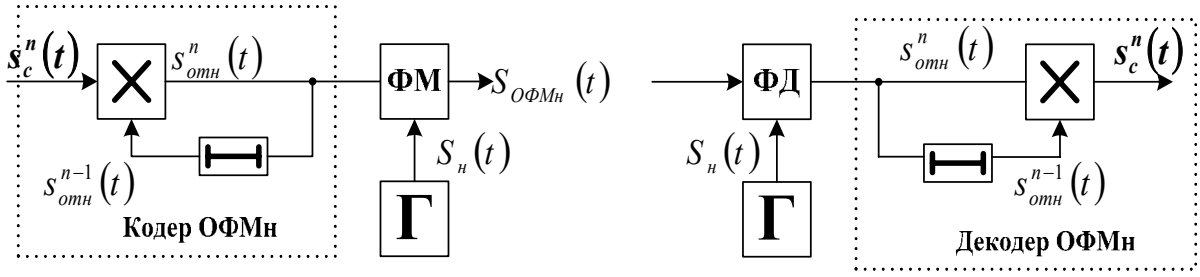


Рис. 2.22. Модулятор и демодулятор ОФМн

На рис. 2.23 представлены временные и спектральные диаграммы формирования сигналов ОФМн: а) непериодический информационный сигнал; б) информационный сигнал в относительном коде; в) несущее колебание; г) сигнал ОФМн на выходе модулятора.

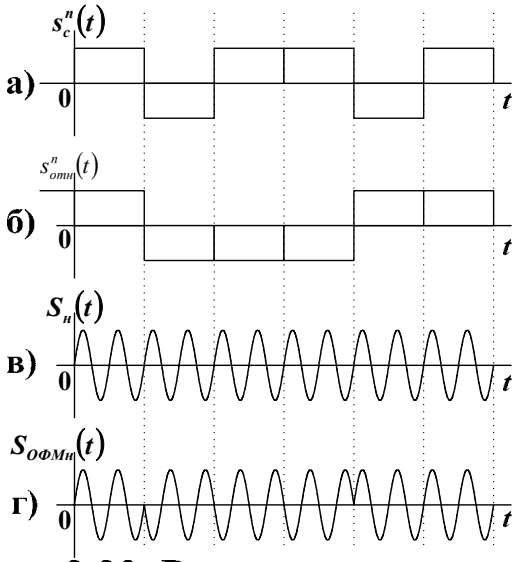


Рис. 2.23. Временные диаграммы формирования сигналов ОФМн

Алгоритмы демодуляции сигналов с ОФМн в сравнении с ФМн иллюстрируются временными диаграммами на рис. 2.24 и 2.25.

На рис. 2.25 представлены временные диаграммы демодуляции сигналов ОФМн и ФМн при однократной ошибке в принятом радиосигнале, в качестве исходного информационного взят сигнал рис. 2.24,а: а) сигнал с ОФМн на выходе модулятора; б) сигнал с ОФМн на входе демодулятора, в принятый сигнал специально введена ошибка для 3 посылки; в) опорное колебание; г) принятый информационный сигнал, на выходе относительного декодера; д) принятый

информационный сигнал, на выходе демодулятора; е) принятый информационный сигнал, на выходе демодулятора в случае отсутствия ошибки.

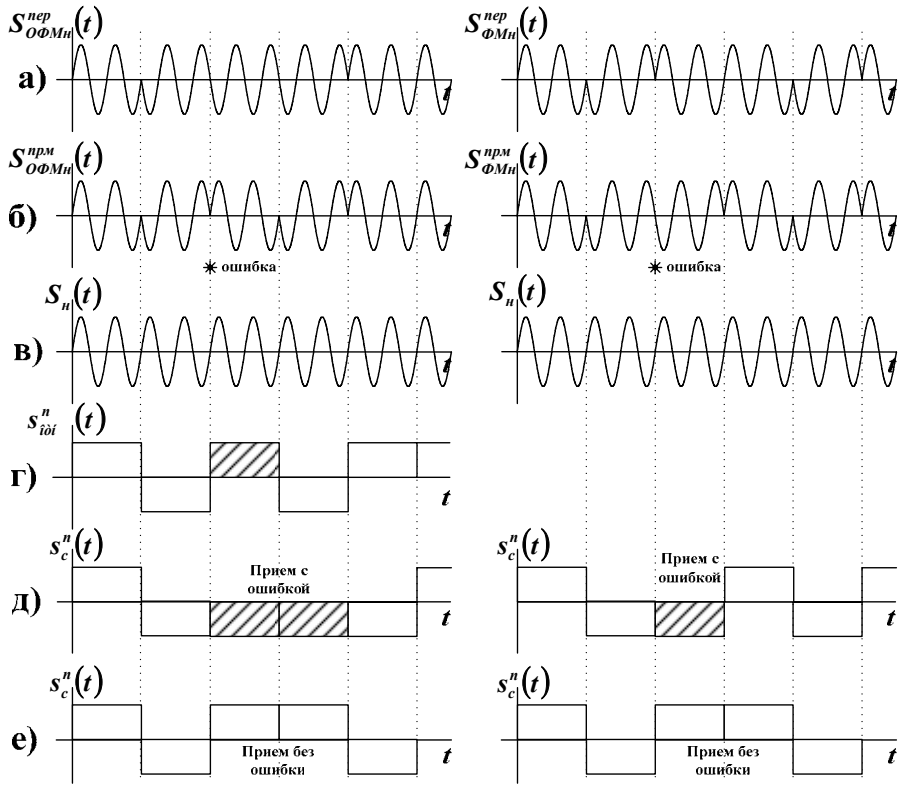


Рис. 2.24. Временные диаграммы демодуляции сигналов ОФМн и ФМн при одной ошибке в принятом радиосигнале

Случай возникновения скачка фазы в опорном колебании представлен на рис. 2.25. При этом в опорное колебание специально введен скачок фазы на 180^{o} между 2 и 3 посылками.

Это дает возможность проиллюстрировать появление ошибок в системах с ФМн и ОФМн. В системе с ФМн, после изменения полярности опорного колебания, все последующие символы ошибочные (обратная работа), причем ошибка будет оставаться до следующего скачка фазы опорного колебания. В системе с ОФМн скачкообразное изменение полярности опорного колебания приводит к одиночной ошибке, что и определяет преимущества сигналов с ОФМн.

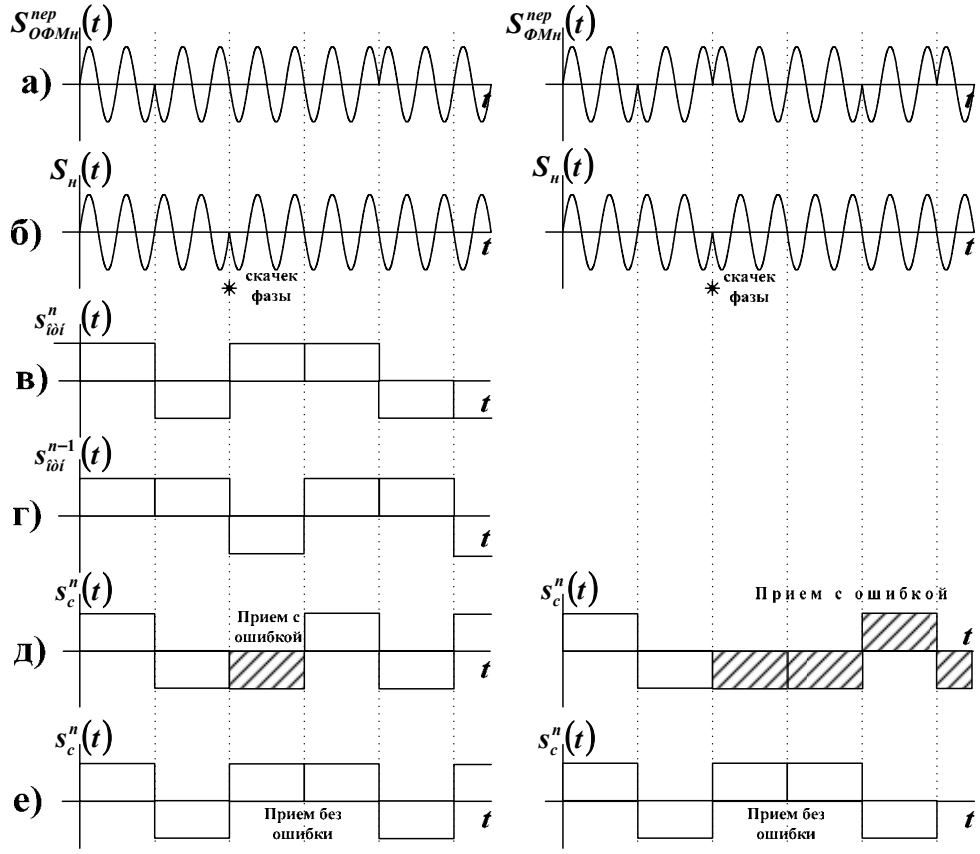


Рис. 2.25. Временные диаграммы демодуляции сигналов с ОФМн и ФМн при изменении полярности опорного колебания

Однако следует отметить недостатки систем с ОФМн, которые следует учитывать при выборе методов модуляций:

необходимость передачи отсчетной посылки в начале сеанса связи;

увеличение вероятности ошибки примерно вдвое;

появление двойных ошибок в цифровом потоке, что усложняет кодек при использовании корректирующих кодов;

сложность построения модема для ОФМн по сравнению с модемом для ФМн.

Для реализации системы с ФМн необходима передача специального синхросигнала (маркерного сигнала), соответствующему одному из символов, например 0. Другой путь реализации ФМн – применение специальных кодов с избыточностью, позволяющих обнаруживать ошибки типа инвертирования всех символов. Все это ведет к определенным потерям: энергетическим, скоростным и аппаратурным, и при выборе метода модуляции ФМн или ОФМн необходимо учитывать их достоинства и недостатки.

2.5. Системы связи с многопозиционной относительной фазовой манипуляцией

2.5.1. Принцип формирования сигнала с многократной относительной фазовой манипуляцией

Важным параметром на выходе модулятора является число значений модулируемого параметра (m) выходного сигнала. Это число называется позиционностью сигнала или способа модуляции. Так, m-позиционная фазовая модуляция означает, что каждый элемент сигнала на выходе модулятора имеет одну из m выбранных начальных фаз. Если все m вариантов сигнала равновероятны, то производительность модулятора как источника информации на входе непрерывного канала связи прямо пропорционально двоичному логарифму числа m: $N = \log_2 m$. Эту величину называют кратностью модуляции: она показывает, сколько двоичных единиц информации содержится в каждом элементе сигнала при данном способе модуляции или во сколько раз (крат) увеличится информационная емкость данной системы по сравнению с двухпозиционной (однократной) системой при той же длительности элементарного сигнала. Наиболее часто позиционность выбирают так, чтобы она равнялась целой степени числа два, тогда кратность N – целое число.

Например, N-кратная фазовая манипуляция означает, что в каждом элементарном сигнале на выходе модулятора содержится N бит информации, а фаза сигнала на входе непрерывного канал имеет $m = 2^N$ допустимых значений. Если длительность элементарного сигнала в модуляторе равна T, то скорость формирования элементов (скорость модуляции) составляет $1/T$ элементов. Количество кодовых символов в единицу времени B показывает скорость формирования информации на выходе модулятора. При равновероятных сигналах

$$B = N/T = \log_2 m/T = V \cdot \log_2 m \text{ [бит/с]}. \qquad (2.30)$$

В зависимости от числа уровней модулирующего сигнала различают двухуровневую и многоуровневую манипуляции. Четырехпозиционная (двухуровневая) модуляция ФМн (ДФМн) предполагает передачу двух двоичных символов одновременно (рис. 2.26). В табл. 2.1 приведены допустимые значения начальных фаз для ФМн-2 и ФМн-4.

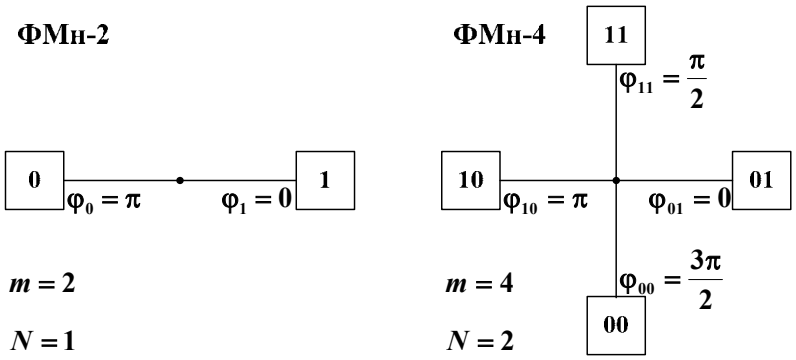


Рис. 2.26. Векторное представление ФМн, m - позиционных сигналов

Таблица 2.1. Допустимые значения начальных фаз для ФМн-2 и ФМн-4

ФМн-2	ФМн-4
$\varphi_i = 0;\pi$	$\varphi_i = 0;\pi/2;\pi;3\pi/2$, или $\left(\pi/4;3\pi/4;5\pi/4;7\pi/4\right)$

Ширина спектра ОФМн-m радиосигнала, определяемая длительностью радиоимпульса (T) зависит от скорости передачи информации (B) и числа уровней манипуляции (N):

$$\Delta F_{ОФМн} = \frac{B}{\log_2 N}. \qquad (2.31)$$

Очевидно, при увеличении числа уровней манипуляции полоса частот необходимая для ОФМн радиосигнала уменьшается. Так, при ОФМн-4 полоса частот вдвое меньше, чем при ОФМн-2 при одинаковой скорости передачи информации. Для двоичных сигналов $(m = 2)$ длительность радиоимпульса T равна длительности единичного элемента ПЭС T_c, а ширина спектра радиосигнала ОФМн-2 пропорциональна скорости передачи цифровой информации: $\Delta F_{ОФМн} = B = V = \frac{1}{T}$ [бод].

В случае многоуровневой манипуляции $(m > 2)$ длительность T сигнала оказывается равной $T = T_c \cdot \log_2 m$, что приводит к соответствующему сокращению в $\log_2 m$ полосы занимаемых частот при передаче одного и того же объема данных.

2.5.2. Квадратурная относительно-фазовая манипуляция (КОФМ)

Формирование модулируемого цифрового сигнала удобно пояснить на основе квадратурного представления сигналов. Смысл его заключается в представлении гармонического колебания с произвольной фазой линейной комбинацией синусоидального и косинусоидального колебания, что вытекает из тригонометрического равенства: $\sin(\omega t + \varphi) = \cos\varphi \sin\omega t + \sin\varphi \cos\omega t$.

Модулятор может быть выполнен по схеме, представленной на рис.2.27,а. Преобразователь кода (Пр) преобразует входной сигнал в два параллельных сигнала каждый из которых модулирует по фазе на 180^o синфазную и квадратурную составляющие.

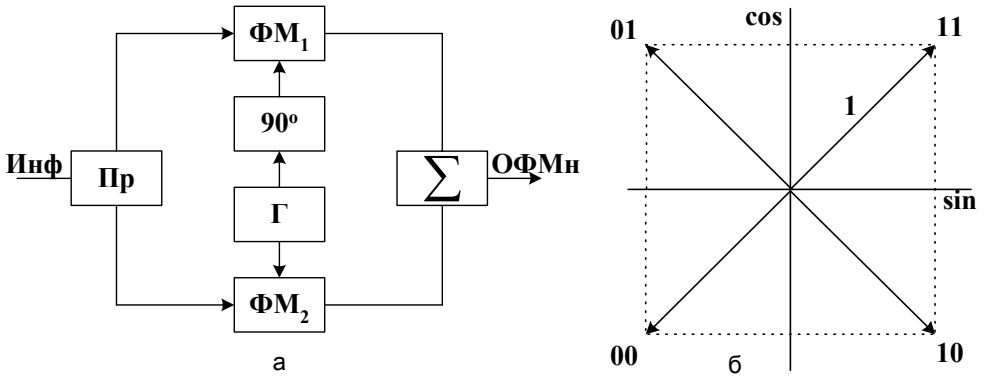


Рис. 2.27. а) квадратурная схема формирования сигнала КОФМ ;
б)векторная диаграмма 4-х позиционного сигнала

Манипуляция осуществляется в двух каналах на несущих, которые имеют относительный угловой сдвиг 90° ($\sin\omega t$ и $\cos\omega t$ – базисные функции разложения), т.е. находящихся в квадратуре (откуда и название метода модуляции). Именно в силу специфики формирования последовательности сигналов метод ОФМн при $m = 4$ часто называют квадратурной ОФМн (КОФМ).

Сигнальные векторы (s_i) получаются суммированием базисных векторов при умножении их на определенные коэффициенты. Если длина сигнального вектора равна **1**, то коэффициенты равны $\pm\frac{1}{\sqrt{2}}$.

Таким образом, в качестве манипулирующих сигналов используют сигналы, отличающиеся по структуре от исходных передаваемых двоичных сигна-

лов, для формирования которых используется специальное кодирующее устройство - кодер модулятора.

Рассмотрим подробнее один из возможных методов формирования сигналов с двукратной ОФМн-4 манипуляцией (m=4) по квадратурной схеме (рис. 2.27, а), на примере сигнала ФМн-4 при которой формируются четыре элементарных сигнала (S_i), каждый из которых характеризуется своей фазой (φ_i) [5]:

$$S_i = \cos\varphi_i \sin\omega t + \sin\varphi_i \cos\omega t\,,\quad 0 \le t \le T\,,\ i = 0,1,2,3,\ldots\ . \tag{2.32}$$

Метод ОФМн можно рассматривать как обычную фазовую манипуляцию на 180° при условии предварительного перекодирования исходного сообщения:

$$x(t) = x_1(t) + x_2(t). \tag{2.33}$$

Поэтому для простоты будем считать, что в сообщениях, представленных функциями $x_1(t)$ и $x_2(t)$ в (2.33), перекодирование произведено, и для передачи исходного сообщения необходимо лишь осуществить ФМн высокочастотных колебаний на 180°.

Исходная последовательность двоичных информационных символов разделяется на последовательности четных x_{2k}, и нечетных символов x_{2k+1} с длительностью элементов $T = T_c \cdot \log_2 m$. Так, например, исходная последовательность двоичных элементов длительностью T_c с помощью кодера модулятора преобразуется в совокупность 2-х $(m = 4)$ или 3-х $(m = 8)$ последовательностей двоичных элементов длительностью $2T_c$ или $3T_c$ соответственно. Тогда передаваемое сообщение $x(t)$ (рис.2.28,а), можно представить в виде суммы четных $\boldsymbol{x_{2k}}$ (рис.2.28,б), и нечетных x_{2k+1} (рис. 2.28,в) составляющих:

$$x(t) = x_{2k}(t) + x_{2k+1}(t - T_c). \tag{2.34}$$

Для экономии полосы занимаемых частот осуществим раздельно фазовую модуляцию сообщениями x_{2k} и x_{2k+1} двух квадратурных составляющих одного и того же колебания $\sin\omega_0 t$. При этом последовательность передаваемых сигналов $S(t)$ представляется в виде [5, 13, 15]:

$$S(t) = S_{2k}(t) + S_{2k+1}(t - T_c), \tag{2.35}$$

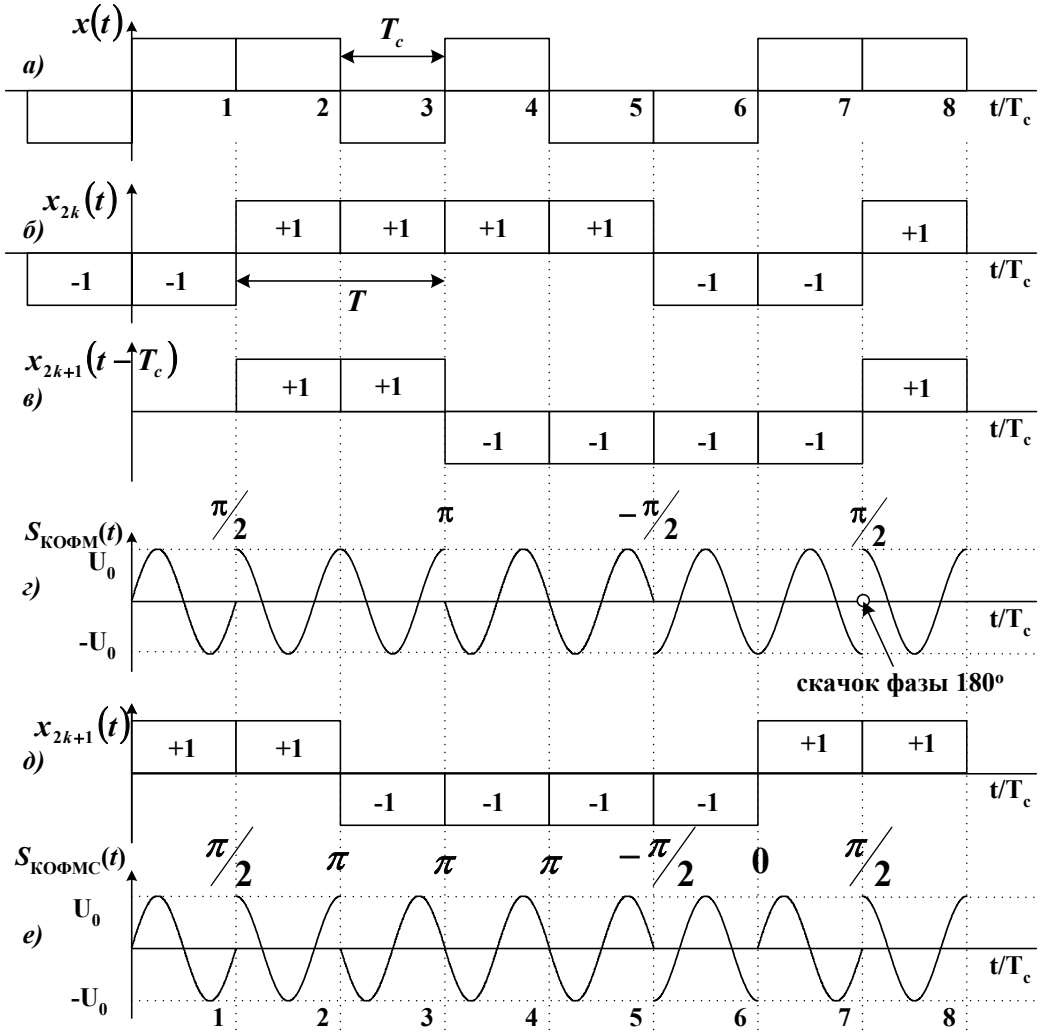


Рис. 2.28. Диаграммы формирования сигналов КОФМ и КОФМС

где $S_{2k}(t)=\left(A_0/\sqrt{2}\right)\cdot x_{2k}(t)\cos\left(\omega_0 t+\pi/4\right)$,

$S_{2k+1}(t-T_c)=\left(A_0/\sqrt{2}\right)\cdot x_{2k+1}(t-T_c)\sin\left(\omega_0 t+\pi/4\right)$.

Комбинации двоичных элементов полученных последовательностей x_{2k} и x_{2k+1} используются при кодировании фазового сдвига при ОФМн. Значения начальной фазы φ колебания $S(t)$ (рис. 2.28, г) при различных сочетаниях передаваемых символов x_{2k} и x_{2k+1} приведены в табл. 2.2.

Таблица 2.2. Значения начальной фазы колебания $S(t)$

x_{2k}	-1	+1	+1	-1
x_{2k+1}	+1	+1	-1	-1
φ	0	$\pi/2$	π	$-\pi/2$

При одновременной смене символов в обоих каналах модулятора в сигнале КОФМ происходят скачки начальной фазы на 180° (как, например, в момент $t = 7T_c$ на (рис. 2.28, г). При прохождении последовательности таких сигналов через узкополосные фильтры в моменты скачков фазы колебания на 180° возникает глубокая паразитная амплитудная модуляция огибающей сигнала (в ней появляются провалы огибающей до нуля). Это приводит к увеличению пик–фактора сигнала и, как следствие, к дополнительным искажениям при нелинейных режимах усиления, может увеличить энергию боковых полос и увеличить помехи в соседних каналах.

Для снижения уровня такой паразитной амплитудной модуляции при $m = 4$ разработана модификация метода КОФМн, называемая квадратурной относительной фазовой модуляцией со сдвигом (КОФМС). В этом случае колебание $S(t)$, и отличие от (2.35), формируется в виде [5, 13]:

$$S(t) = S_{2k}(t) + S_{2k+1}(t), \tag{2.36}$$

где $$S_{2k}(t) = \left(A_0 \big/ \sqrt{2} \right) \cdot x_{2k}(t) \cos\left(\omega_0 t + \pi/4\right),\ S_{2k+1}(t) = \left(A_0 \big/ \sqrt{2} \right) \cdot x_{2k+1}(t) \sin\left(\omega_0 t + \pi/4\right). \tag{2.37}$$

Как следует из соотношений (2.33) и (2.37), знак любой из функций $x_{2k}(t)$ или $x_{2k+1}(t)$ может меняться лишь в те моменты, когда значение другой функции сохраняется неизменным. Такой сдвиг по времени моментов возможной смены знака модулирующих последовательностей приводит к существенному отличию результирующего колебания $S(t)$ (рис. 2.28, е) при КОФМС по сравнению с КОФМ.

Заметим, что скачки начальной фазы φ колебания $S(t)$ возможны лишь на $\pm \pi/2$ (рис. 2.28, е) что снижает паразитную амплитудную модуляцию при прохождении сигнала через полосовые цепи. Длительность радиосигнала T КОФМС равна длительности исходного информационного символа T_c, т.е. вдвое меньше, чем при КОФМ. Однако это не приводит к расширению спектра последовательности $S(t)$ по сравнению с использованием КОФМ. Последнее объясняется тем, что ширина спектра колебания $S(t)$ определяется шириной

спектра квадратурных составляющих $S_{2k}(t)$ и $S_{2k+1}(t)$ в (2.37), которая остается той же, что и при КОФМ (2.35).

При приеме сигналов как с КОФМ, так и с КОФМС можно воспользоваться тем, что составляющие $S_{2k}(t)$ и $S_{2k+1}(t)$ суммарной последовательности $S(t)$ сдвинуты на 90° по фазе высокочастотного заполнения, а сообщения $x_{2k}(t)$ или $x_{2k+1}(t)$ независимы.

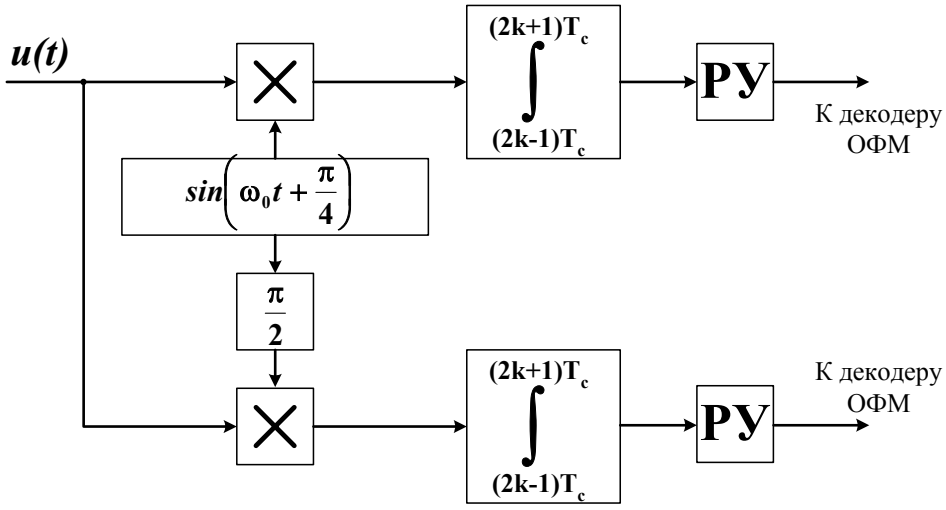


Рис. 2.29. Структурная схема демодулятора сигналов с КОФМ и КОФМС

В этих условиях при наличии в демодуляторе генераторов непрерывных колебаний $\sin(\omega_0 t + \pi/4)$ и $\cos(\omega_0 t + \pi/4)$ легко осуществить раздельный прием каждой из составляющих $S_{2k}(t)$ и $S_{2k+1}(t)$. Действительно, рассмотрим устройство на рис. 2.29. При поступлении на вход колебания $S(t)$ вида (2.35) вклад составляющей $S_{2k+1}(t)$ в выходном напряжении интегратора верхнего канала в моменты $t = (2k+1)T_c$ оказывается пренебрежимо малым. Таким образом, верхний канал схемы на рис. 2.29 представляет собой демодулятор двоичных сигналов с ОФМн, содержащих информацию о сообщении $x_{2k}(t)$. Нижний канал выполняет функции демодулятора двоичных сигналов с ОФМн, составляющих последовательность $S_{2k+1}(t)$ и содержащих информацию о сообщении $x_{2k+1}(t)$.

При КОФМС нет скачков фазы на 180° т.к. текущее изменение фазы происходит в моменты смены знака любой из функций $x_{2k}(t)$ или $x_{2k+1}(t)$. При этом

значение другой функции сохраняется неизменным, что позволяет избежать глубокой паразитной модуляции огибающей.

2.5.3. Принцип частотной модуляции с непрерывной фазой

Частотная модуляция с непрерывной фазой (ЧМНФ) является частным случаем частотной модуляции с минимальным сдвигом. В этом случае фаза манипулируемого колебания изменяется непрерывно и не имеет скачков на границах радиоимпульсов. При ЧМНФ для передачи $+1$ и -1, как и при обычной двоичной ЧМн, используются две частоты, однако разность частот выбирается такой, чтобы за время длительности элемента T фаза несущего колебания изменялась ровно на $\pi/2$.

Как отмечалось ранее, ширина спектра модулированного сигнала $S(t)$ определяется видом квадратурных составляющих $S_{2k}(t)$ и $S_{2k+1}(t)$. Поэтому ширину спектра сигнала с КОФМС можно сократить, если ввести вспомогательную амплитудную модуляцию этих квадратурных составляющих, позволяющую уменьшать значение огибающих колебаний $S_{2k}(t)$ и $S_{2k+1}(t)$ в моменты скачков фазы этих колебаний на 180°, Вспомогательную амплитудную модуляцию квадратурных составляющих удобно осуществить по гармоническому закону [5, 13]:

$$S_{2k+1}(t)=A_0 x_{2k+1}(t)\sin\left(\pi t/2T_c\right)\sin\omega_0 t\,,\qquad S_{2k}(t)=A_0 x_{2k}(t)\cos\left(\pi t/2T_c\right)\cos\omega_0 t\,. \qquad (2.38)$$

Функции $A_0 x_{2k+1}(t)\sin\left(\pi t/2T_c\right)$, $S_{2k+1}(t)$, $A_0 x_{2k}(t)\cos\left(\pi t/2T_c\right)$ и $S_{2k}(t)$ показаны для информационной последовательности $x(t)$ изображенной на рис. 2. 30,а. Как следует из (2.38) и рис. 2.30, д, знак функции $x_{2k+1}(t)$ может меняться лишь в моменты равенства нулю огибающей квадратурной составляющей $S_{2k+1}(t)$, причем огибающая квадратурной составляющей $S_{2k}(t)$ в эти моменты времени достигает максимального значения. Соответственно, функция $x_{2k}(t)$ может изменять свой знак лишь в моменты равенства нулю огибающей квадратурной со-

ставляющей $S_{2k}(t)$. Этим обеспечивается непрерывность фазы суммарного колебания $S(t)$ в моменты смены информационных символов, причем на каждом i-м интервале времени $[iT_c,(i+1)T_c]$ колебание $S(t)$ имеет постоянную огибающую и одну из двух возможных частот $\omega_0 \pm \pi/2T_c$. Действительно, как следует из (2.38), на рассматриваемом i-м интервале времени:

$$S(t) = A_0 \cos(\omega_0 t + \varphi(t)), \qquad (2.39)$$

где $\varphi(t) = b_i(t)\pi t/2T_c + \varphi_i$; $b_i = -x_{2k+1}(t)x_{2k}(t)$; фаза φ_i принимает значения **0** или π, причем значение $\varphi_i = \pi$ только тогда, когда одна из функций $x_{2k}(t)$ или $x_{2k+1}(t)$ примет значение -1.

Таким образом, при условии (2.38) колебание $S(t)$**)** представляет собой последовательность ЧМн сигналов с непрерывной фазой. В отличие от обычной двоичной ЧМн, когда разнос частот выбирается кратным $1/T_c$**,** в данном случае разнос частот существенно меньше и равен $1/2T_c$, что и обусловило название этого метода – частотная модуляция с минимальным сдвигом (ЧММС).

Закон изменения фазы $\varphi(t)$ колебания (2.39) и сам вид колебания $S(t)$ для последовательности информационных символов, изображенной на рис. 2.30,а, показаны на рис. 2.30,е,ж для $\omega_0 = 2\pi/T_c$. Как следует из (2.39), мгновенная частота колебания $S(t)$ на *i*-м интервале времени зависит не от значения x_i передаваемого i-го информационного символа, а от знака произведения $x_{i-1} \cdot x_i$ (табл. 2.3).

Таблица 2.3. Соответствие значения произведения квадратурных составляющих и величин мгновенных фазы и частоты:

Произведение квадратурных составляющих:	$x_{i-1} \cdot x_i = +1$	$x_{i-1} \cdot x_i = -1$
Приращение мгновенной фазы: $\Delta\varphi$	$-\frac{\pi}{2}$	$\frac{\pi}{2}$
Значение мгновенной частоты $\omega(t)$:	$\omega_1(t) = \frac{3}{2}\pi$	$\omega_2(t) = \frac{5}{2}\pi$

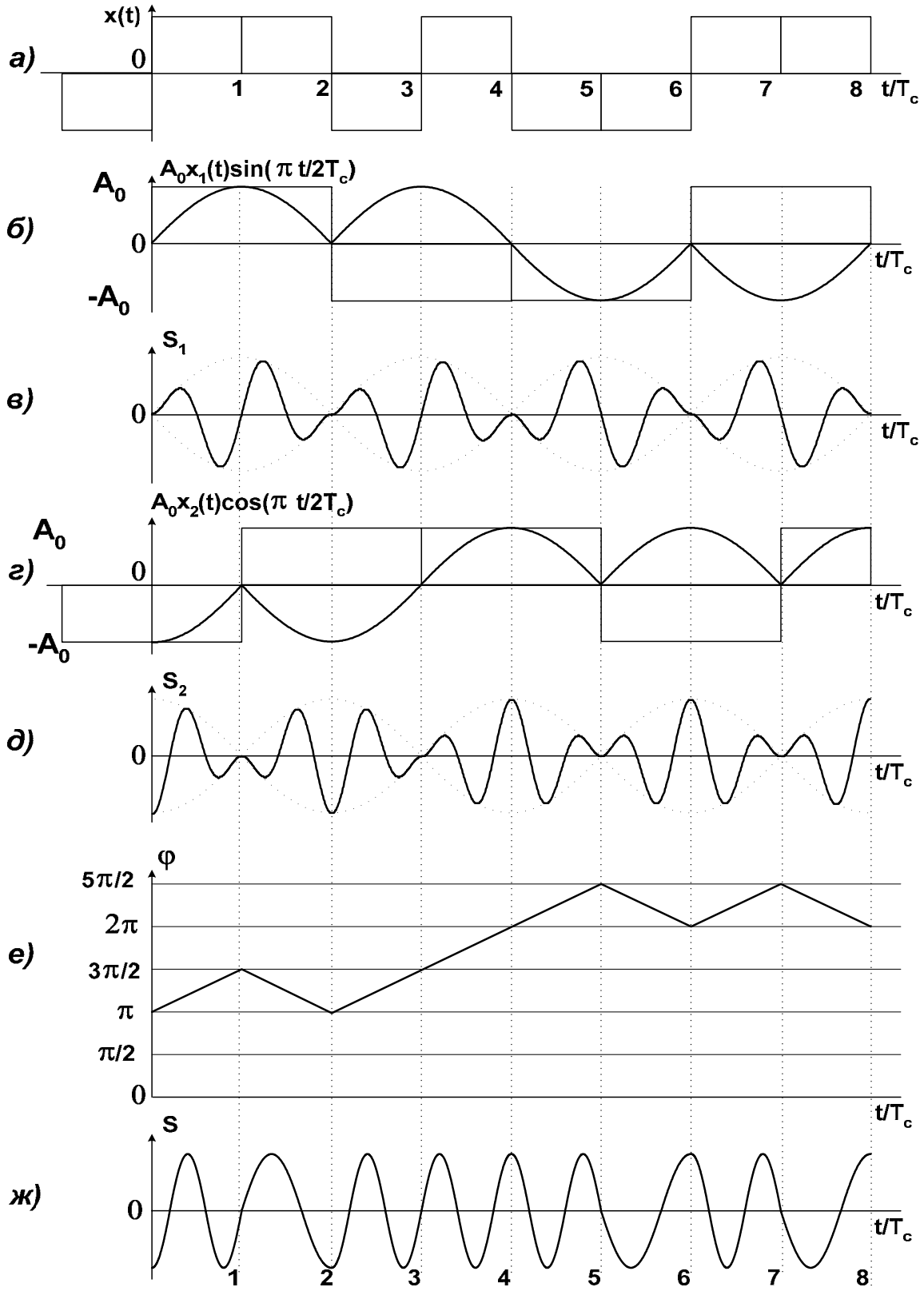


Рис. 2.30. Диаграммы формирования частотно-модулированных сигналов с непрерывной фазой

Рассмотренный метод формирования сигналов с ЧММС называется квадратурным. Возможен и другой метод формирования таких сигналов, когда частота каждого сигнала на интервале длительности T_c определяется непосредст-

венно передаваемым в этот момент символом сообщения.

Такой метод называют прямым. Если на интервале $[0,T_c]$ сигнал

$$S(t)=A_0\cos\left(\omega_0 t+{}^{x_0\pi t}\!/_{2T_c}\right), \tag{2.40}$$

то в общем случае на интервале $[kT_c,(k+1)T_c]$ он будет иметь вид:

$$S(t)=A_0\cos(\omega_0 t+\varphi(t)),$$
$$\text{где } \varphi(t)=\frac{\pi}{2}\sum_{i=0}^{k-1}x_i+x_k\frac{\pi}{2T_c}(t-kT_c). \tag{2.41}$$

Спектр последовательности сигналов с ЧММС не зависит от того, как вводится полезная информация в квадратурные составляющие сигнала: раздельно или путем непосредственной манипуляции частоты сигнала в соответствии с передаваемыми информационными символами.

Сигналы с ЧММС являются частным случаем ЧМ – сигналов с непрерывной фазой (ЧМНФ). На интервале $[0,T_c]$ такой сигнал представляется в виде:

$$S(t)=A_0\cos\left(\omega_0 t+{}^{x_0 m\pi t}\!/_{T_c}\right), \tag{2.42}$$

а в общем случае на интервале $[kT_c,(k+1)T_c]$:

$$S(t)=A_0\cos\left[\omega_0 t+x_k\frac{\pi m}{T_c}(t-kT_c)+\pi m\sum_{i=0}^{k-1}x_i\right], \tag{2.43}$$

где m – индекс модуляции.

Контрольные вопросы

1. Представьте временные характеристики передаваемого сообщения, состоящего из последовательности двоичных символов вида 0101011010 в униполярной форме. В форме относительного кода. В биимпульсной форме.

2. Для каких целей используется модуляция в системах электрической связи?

3. Как вычислить ширину спектра сигналов АМн, ЧМн и ФМн?

4. Задано аналитическое выражение для модулированного сигнала $S(t)=12\cos(6{,}28\cdot10^6 t+5\cos 12{,}56\cdot10^4 t)$. Определите мощность модулированного

сигнала, максимальные девиации частоты и фазы, постройте спектральную диаграмму этого сигнала. Можно ли определить, какой сигнал здесь представлен: ЧМ или ФМ?

5. Представьте временные характеристики сигналов многопозиционной амплитудной и частотной модуляции.

6. Какое свойство функций Бесселя позволяет считать спектр сигналов угловой модуляции ограниченным?

7. Как изменится ширина спектра однотональных АМн и ЧМн сигналов, если частоту модулирующего сигнала увеличить вдвое?

ГЛАВА 3. ПОМЕХОУСТОЙЧИВОСТЬ ПРИЕМА ДИСКРЕТНЫХ СООБЩЕНИЙ

Проблема помехоустойчивости является одной из важнейших проблем современной радиотехники. В классической работе основоположника статистической теории связи В.А. Котельникова «Теория потенциальной помехоустойчивости» была сформулирована и решена задача статистического синтеза оптимальных приемных устройств, проанализированы системы связи при различных видах модуляции и определена предельная помехоустойчивость, которая может быть достигнута при заданном способе передачи информации. В дальнейшем теория развивалась не только для гауссовского канала, но и для каналов с переменными параметрами, пространственно-временных каналов, каналов с сосредоточенными и импульсными помехами и др. В настоящее время статистической теории связи посвящена многочисленная литература [5, 21, 32, 39]. При этом результаты теории оказались весьма конструктивными и способствовали быстрому развитию современной техники связи.

3.1. Критерии качества и правила приема дискретных сообщений

3.1.1. Понятие о помехоустойчивости систем электрической связи

Качество передаваемой информации принято оценивать достоверностью передачи сообщений, т.е. степенью соответствия принятого сообщения переданному. Количественная мера достоверности и помехоустойчивости зависит от характера передаваемых сообщений (текст, речь, музыка, изображение и т. д.).

Для оценки качества приема дискретных сообщений применяется вероятность ошибочного приема переданного символа $p_{ош}$. Значение вероятности ошибки должно быть достаточно малым. Так для систем радиорелейной связи

$p_{ош}$ имеет порядок 10^{-6}, для тропосферной связи 10^{-5}, для радиосвязи 10^{-3}.

Появление ошибок в системах связи происходит из-за действия помех различных видов. Способность систем связи различать (восстанавливать) сигналы с заданной достоверностью при наличии помех называется помехоустойчивостью. Различают потенциальную и реальную помехоустойчивость. Под потенциальной помехоустойчивостью понимают предельно достижимую помехоустойчивость при заданных сигналах и помехах. Реальная помехоустойчивость систем связи с учетом конкретного выполнения элементов передающего и приемного тракта, линии связи, кодека, модема всегда меньше теоретической.

3.1.2. Задача оптимального приема

Рассмотрим сначала наиболее простую задачу теории обнаружения сигналов. Допустим, что некоторый объект, интересующий наблюдателя, может находиться в одном из двух состояний S_0 или S_1. Такими состояниями могут быть, например, наличие или отсутствие цели в зоне действия РЛС, передача сигнала «0» или «1» по каналу связи, работоспособность или отказ устройства и др. В каждом конкретном эксперименте объект находится в состоянии S_0 или в состоянии S_1 с вероятностями p_0 и p_1 $(p_0 + p_1 = 1)$ соответственно. Поскольку действительное состояние объекта наблюдателю не известно, то можно лишь выдвинуть предположение (гипотезу) H_0 о том, что объект находится в состоянии S_0 и альтернативное предположение H_1.

В зависимости от состояния S_0 или S_1 объекта результаты $\bar{y} = (y_1\, y_2 \ldots y_n)^T$ эксперимента имеют плотность распределения вероятностей (ПРВ) $w(\bar{y}/H_0)$ или ПРВ $w(\bar{y}/H_1)$. На основе анализа наблюдений $y_1, y_2, \ldots, y_n$ необходимо определить, в каком именно состоянии находится объект.

В этих терминах задача состоит в том, чтобы на основе наблюдении $\bar{y}$ проверить справедливость гипотезы H_0. Любое правило проверки гипотезы каждому конкретному результату эксперимента $\bar{y}$ должно поставить в соответст-

вие определенное решение. Но это означает, что при заданном правиле решения среди всех возможных исходов $\bar{y} \in G$ можно выделить область G_0, где принимается гипотеза H_0. Если же результат наблюдения $\bar{y} \notin G_0$, то принимается решение H_1.

Так, например, если производится только одно наблюдение y_1 на отрезке $[a, b]$, то для конкретного значения y_1 должно быть принято либо решение H_0, либо H_1. Таким образом, множество $G\{y_1 : a \le y_1 \le b\}$ всех точек отрезка (всех возможных исходов эксперимента) разбивается на две области G_0 и G_1 (рис.3.1,а).

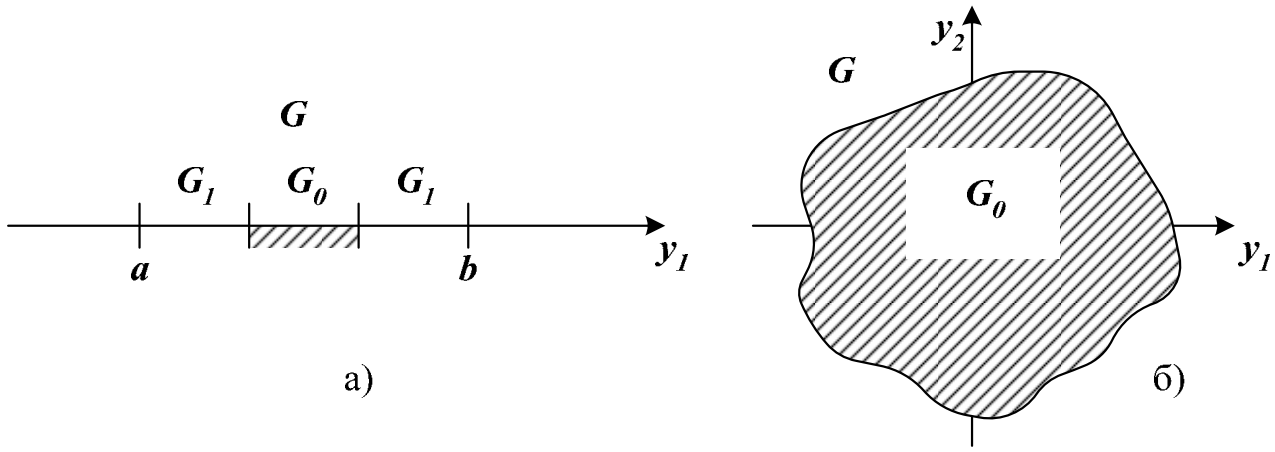


Рис. 3.1. Допустимая и критическая области

Если $y_1 \in G_0$, то принимается решение о справедливости гипотезы H_0; если же $y_1 \in G_1$, то предпочтение отдается гипотезе H_1. Когда производится два наблюдения y_1, y_2, множество всех исходов эксперимента представляется точками плоскости G (рис.3.1,б). Поскольку каждому исходу (y_1, y_2) соответствует конкретное решение H_0 или H_1, то все множество G так же, как и в одномерном случае, должно быть разделено на два подмножества G_0 и G_1 (рис.3.1,б).

Очевидно, в общем случае выборки $\bar{y} = (y_1, y_2, ..., y_n)^T$ произвольного объема n-мерная область G всех возможных исходов опытов разбивается на две подобласти G_0 и G_1. Область G_0, где принимается гипотеза H_0, называют допустимой областью. Вторую область G_1, отклонения гипотезы H_0, называют

критической.

Таким образом, построение оптимального правила проверки гипотезы H_0 может трактоваться как нахождение наилучшего разбиения пространства G всех возможных результатов эксперимента на две области G_0 и G_1 или, что в данном случае то же самое, как выбор наилучшей допустимой области.

Для того чтобы выяснить, что следует понимать под наилучшим разбиением, необходимо ввести критерий качества правила принятия решения. Поскольку состояние объекта заранее не известно, а прием сигналов затруднен помехами, то при использовании любого правила решения возможны ошибки. С этой точки зрения после принятия решения возможны четыре ситуации, схематично изображенные на рис.3.2.

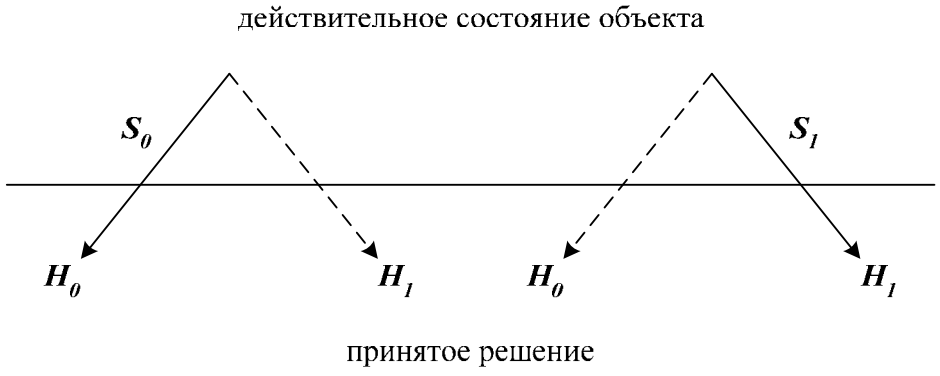


Рис. 3.2. Правильные (сплошные линии) и ложные (пунктир) решения

Две из них соответствуют правильным решениям (сплошные линии) и две – ошибочным (пунктир). Ошибка, в результате которой принимается решение H_1 при нахождении объекта в состоянии S_0, называется ошибкой первого рода. Другая ошибка – ошибкой второго рода.

В задачах обнаружения цели состояние S_0 и гипотеза H_0 соответствуют отсутствию цели, и ошибка первого рода обычно называется ложной тревогой. Ошибка второго рода состоит в принятии неверного решения об отсутствии цели, когда цель присутствует, и называется пропуском цели.

Используя формулу (1.45), нетрудно записать следующие выражения для вероятности ошибки 1 рода [6]:

$$p_F = \int_{G_1} \ldots \int w\left(\bar{y}/H_0\right) d\bar{y}\,, \tag{3.1}$$

и вероятности ошибки 2 рода:

$$p_M = \int_{G_0} \ldots \int w\left(\bar{y}/H_1\right) d\bar{y}\,, \tag{3.2}$$

где $d\bar{y} = dy_1\, dy_2 \ldots dy_n$.

Вместо p_M можно использовать вероятность противоположного события, т.е. вероятность правильного решения. Очевидно:

$$p_D = 1 - p_M = \int_{G_1} \ldots \int w\left(\bar{y}/H_1\right) d\bar{y}\,. \tag{3.3}$$

Для заданного размера выборки невозможно одновременно сделать сколь угодно малыми вероятности ошибок первого и второго рода. Например, чтобы уменьшить вероятность ложной тревоги p_F, следует уменьшить размер критической области G_1, но тогда увеличивается размер допустимой области G_0 и возрастает вероятность ошибки второго рода (3.2). Поэтому «разумный» критерий оптимальности должен быть построен на основе какого-либо компромисса между вкладом двух типов возможных ошибок в общую характеристику или общие показатели системы обнаружения.

Одним из возможных способов построения критерия оптимальности может быть байесовский подход, общая методология которого рассматривалась в предыдущем разделе применительно к задачам оценивания параметров. Точно так же основой байесовского подхода к проблемам обнаружения является введение функции потерь, которая приписывает каждой из четырех возможных ситуаций (рис.3.2) определенную плату. При этом обычно правильным решениям соответствует нулевой размер штрафа. Ошибке первого рода поставим в соответствие плату R_0, а ошибке второго рода – плату размером R_1. Тогда средние потери составят величину

$$\bar{R} = p_0 R_0 P_F + p_1 R_1 P_M\,, \tag{3.4}$$

которая и принимается как критерий качества обнаружения. При этом обнаружитель, для которого средние потери R минимальны, называется оптимальным байесовским обнаружителем.

Подставляя выражения (3.1) и (3.2) в формулу (3.4), получим следующую связь средних потерь с видом критической области:

$$\overline{R} = p_0 R_0 \int_{G_1} \ldots \int w\left(\overline{y}/H_0\right) d\overline{y} + p_1 R_1 \int_{G_0} \ldots \int w\left(\overline{y}/H_1\right) d\overline{y} =$$

$$= p_1 R_1 - \int_{G_1} \ldots \int \left(p_1 R_1 w\left(\overline{y}/H_1\right) - p_0 R_0 w\left(\overline{y}/H_0\right)\right) d\overline{y} . \qquad (3.5)$$

Очевидно, потери минимальны, если интеграл

$$\int_{G_1} \ldots \int \left(p_1 R_1 w\left(\overline{y}/H_1\right) - p_0 R_0 w\left(\overline{y}/H_0\right)\right) d\overline{y} \qquad (3.6)$$

достигает максимального значения.

Какие же точки пространства G возможных исходов эксперимента следует включить в область G_1 для максимизации выражения (3.6)? Простой анализ показывает, что при наблюдении $\overline{y}$ следует проверить – положительным или отрицательным окажется подынтегральное выражение (3.6). Если

$$p_1 R_1 w\left(\overline{y}/H_1\right) - p_0 R_0 w\left(\overline{y}/H_0\right) \geq 0 \qquad (3.7)$$

то такую точку $\overline{y}$ следует отнести к критической области G_1. Действительно, после добавления такой точки вместе с некоторой окрестностью к области G_1 возрастает интеграл (3.6) по этой области и, следовательно, уменьшаются средние потери (3.5). Таким образом, неравенство (3.7) определяет все точки критической области G_1. Но это, в свою очередь, означает, что для наблюдений, удовлетворяющих неравенству (3.7), следует принимать верной гипотезу H_1, а для остальных точек – гипотезу H_0. Переписывая неравенство (3.7), определяющее критическую область, в форме

$$\Lambda \geq \Lambda_0 , \qquad (3.8)$$

где $\Lambda = w\left(\overline{y}/H_1\right) / w\left(\overline{y}/H_0\right)$ – отношение правдоподобия; $\Lambda_0 = \dfrac{p_0 R_0}{p_1 R_1}$; можно заметить, что формула (3.8) определяет алгоритм обработки входных данных

$\bar{y}$. Действительно, оптимальный обнаружитель должен формировать на основе наблюдений $\bar{y}$ отношение правдоподобия Λ и производить сравнение этого отношения с пороговым уровнем Λ_0. Если $\Lambda \geq \Lambda_0$, то выносится решение в пользу гипотезы H_1. При $\Lambda < \Lambda_0$ принимается, что справедлива гипотеза H_0. Так же, как и при оценивании параметров, можно вместо отношения правдоподобия сравнивать с пороговым уровнем любую монотонную функцию $f(\Lambda)$, например, $\ln \Lambda$. При этом достаточно изменить величину порога обнаружения и положить, что $\Lambda_0^{/} = f(\Lambda_0)$.

Рассмотрим пример решения задачи последетекторного обнаружения радиосигнала по совокупности независимых наблюдений $y_1, y_2, \ldots, y_n$. При отсутствии сигнала эти наблюдения подчиняются закону распределения Релея:

$$w(\bar{y}/H_0) = \prod_{i=1}^{n} w(y_i/H_0) = \prod_{i=1}^{n} \frac{y_i}{\sigma^2} \exp\left(-\frac{y_i^2}{2\sigma^2}\right). \tag{3.9}$$

Появление полезного сигнала вызывает увеличение параметра σ^2 в $(1+q)$ раз, где q – отношение сигнал/шум. При этом

$$w(\bar{y}/H_1) = \prod_{i=1}^{n} \frac{y_i}{\sigma^2(1+q)} \exp\left(-\frac{y_i^2}{2\sigma^2(1+q)}\right). \tag{3.10}$$

Для нахождения оптимального алгоритма обнаружения составим отношение правдоподобия

$$\Lambda = \prod_{i=1}^{n} \frac{1}{1+q} \exp\left(\frac{q\, y_i^2}{2\sigma^2(1+q)}\right) = \frac{1}{(1+q)^n} \exp\left(\frac{q}{2\sigma^2(1+q)} \sum y_i^2\right)$$

и будем сравнивать его с порогом $\Lambda_0 = p_0 R_0 / p_1 R_1$, зависящим от априорных вероятностей наличия p_1 и отсутствия p_0 полезного сигнала и стоимостей R_1 и R_0 ошибок. После логарифмирования можно записать оптимальную процедуру обнаружения в виде сравнения с пороговым значением $\Lambda_0^{/} = \left(2\sigma^2(1+q)/q\right)\ln\left(\Lambda_0(1+q)^n\right)$ суммы квадратов наблюдений, т.е.

$$\sum_i^n y_i^2 \begin{cases} \geq \Lambda_0^{/} & \text{сигнал есть,} \\ < \Lambda_0^{/} & \text{сигнала нет.} \end{cases} \tag{3.11}$$

Одним из существенных недостатков байесовского правила обнаружения сигналов является большое количество априорной информации о потерях и вероятностях состоянии объекта, которая должна быть в распоряжении наблюдателя. Этот недостаток наиболее отчетливо проявляется при анализе радиолокационных задач обнаружения цепи, когда указать априорные вероятности наличия цели в заданной области пространства и потери за счет ложной тревоги или пропуска цели оказывается весьма затруднительным. Поэтому в подобных задачах вместо байесовского критерия обычно используется критерий Неймана-Пирсона. Согласно этому критерию выбирается такое правило обнаружения, которое обеспечивает минимальную величину вероятности пропуска сигнала (максимальную вероятность правильного обнаружения) при условии, что вероятность ложной тревоги не превышает заданной величины F_0. Таким образом, оптимальное, в смысле критерия Неймана-Пирсона, правило обнаружения минимизирует

$$P_M = \int_{G_0} \ldots \int w\left(\bar{y}/H_1\right) d\bar{y} \tag{3.12}$$

при дополнительном ограничении

$$\int_{G_1} \ldots \int w\left(\bar{y}/H_0\right) d\bar{y} = F_0 . \tag{3.13}$$

Для поиска оптимальной процедуры обработки данных преобразуем задачу на условный экстремум (3.12) при условии (3.13) к задаче на безусловный экстремум. С этой целью воспользуемся методом множителей Лагранжа [39]. Введем множитель Лагранжа λ и запишем функцию Лагранжа

$$J = P_M + \lambda\left(\int_{G_1} \ldots \int w\left(\bar{y}/H_0\right) d\bar{y} - F_0 \right). \tag{3.14}$$

После преобразований, аналогичных выводу формулы (3.5), соотношение (3.14) можно переписать в виде:

$$J = 1 - \lambda F_0 - \left(\int_{G_1} \ldots \int \left(w\left(\bar{y}/H_1\right) - \lambda w\left(\bar{y}/H_0\right) \right) d\bar{y} \right).$$

Сравнение полученного выражения с формулой (3.5) показывает, что минимум функции Лагранжа достигается, если в качестве критической области выбрать совокупность точек $\bar{y}$, удовлетворяющих неравенству

$$\Lambda = w(\bar{y}/H_1) / w(\bar{y}/H_0) \geq \lambda . \qquad (3.15)$$

При этом множитель λ, являющийся пороговым значением, должен находиться из условия (3.13) равенства вероятности ложной тревоги заданной величине F_0.

Из сравнения (3.15) и (3.8) можно заключить, что оптимальное, в смысле критерия Неймана-Пирсона, правило обнаружения отличается от байесовского лишь величиной порогового уровня, с которым производится сравнение отношения правдоподобия.

В качестве примера построения обнаружителя (3.15) рассмотрим задачу проверки гипотезы H_0:

$$w(y_i/H_0) = \left(1/\sqrt{2\pi}\sigma\right)\exp\left(-y_i^2/2\sigma^2\right),\ i = 1, 2, \ldots, n ,$$

при альтернативе

$$w(y_i/H_1) = \left(1/\sqrt{2\pi}\sigma\right)\exp\left(-(y_i - a)^2/2\sigma^2\right),\ i = 1, 2, \ldots, n .$$

Такая задача возникает в тех случаях, когда появление полезного сигнала вызывает изменение среднего значения нормального шума на величину a. При независимых отсчетах $y_1, y_2, \ldots, y_n$ входного процесса отношение правдоподобия может быть записано в виде

$$\Lambda = \frac{w(\bar{y}/H_1)}{w(\bar{y}/H_0)} = \prod_{i=1}^{n} \frac{w(y_i/H_1)}{w(y_i/H_0)} = \exp\left(-\frac{na^2}{2\sigma^2} + \frac{a}{\sigma^2}\sum_{i=1}^{n} y_i\right).$$

После логарифмирования получаем следующий алгоритм обнаружения сигнала:

$$T = \sum_{i}^{n} y_i \begin{cases} \geq \mathrm{T}_0 & \text{сигнал есть,} \\ < \mathrm{T}_0 & \text{сигнала нет.} \end{cases} \qquad (3.16)$$

причем пороговый уровень T_0 выбирается из условия

$$P(T \geq T_0/H_0) = F_0 . \qquad (3.17)$$

Поскольку сумма T нормальных случайных величин (СВ) подчиняется нормальному закону распределения, то при отсутствии сигнала можно записать следующее выражение для условной ПРВ $w(T/H_0)=\left(1/\sqrt{2\pi}\sigma\right)\exp\left(-T^2/2n\sigma^2\right)$. С учетом формул табл.1 соотношение (3.17) перепишется в форме $\Phi_0\left(T_0/\sqrt{n\sigma^2}\right)=0{,}5-F_0$. Из этого равенства по таблицам функции Лапласа $\Phi_0(x)$ [40] можно определить величину порогового уровня T_0. Так, при $F_0=10^{-2}$ получим $T_0/n\sigma^2=2{,}32$; при $F_0=10^{-3}$, $T_0/n\sigma^2=3{,}09$.

3.1.3. Вычисление вероятностей ошибок

Рассмотрим методы анализа помехоустойчивости систем обнаружения сигналов, т.е. методы расчета вероятности ложной тревоги (3.1) и вероятности пропуска сигнала (3.2) (или вероятности правильного обнаружения (3.3)). Подобные расчеты являются обязательным этапом проектирования систем обнаружения, осуществляемым после синтеза оптимального алгоритма.

Как было показано в п.3.1, оптимальный по нескольким критериям качества алгоритм обнаружения сигналов состоит в сравнении с порогом отношения правдоподобия. Для независимых отсчетов $y_1, y_2, ..., y_n$ входного процесса такой алгоритм может быть записан в форме произведения

$$\Lambda=\prod_{i=1}^{n}\frac{w(y_i/H_1)}{w(y_i/H_0)}\begin{cases}\geq\Lambda_0\rightarrow H_1,\\ <\Lambda_0\rightarrow H_0.\end{cases} \tag{3.18}$$

После логарифмирования (3.18) процедура обработки приводится к виду:

$$z=\ln\Lambda=\sum_{i=1}^{n}l_i(y_i)>z_0, \tag{3.19}$$

где $l_i(y_i)=\ln\frac{w(y_i/H_1)}{w(y_i/H_0)}$; $z_0=\ln\Lambda_0$. Таким образом, для расчета вероятностей

$$P_F=\int_{z_0}^{\infty}w(z/H_0)dz,\quad P_D=\int_{z_0}^{\infty}w(z/H_1)dz \tag{3.20}$$

необходимо найти ПРВ $w(z/H_0)$ и $w(z/H_1)$ и вычислить интегралы (3.20). Поскольку для расчета (3.20) при известных ПРВ $w(z/H_0)$ и $w(z/H_1)$ могут эффективно использоваться численные методы интегрирования, то, как правило,

наиболее трудоемким является определение ПРВ суммы z СВ l_i, $i=1,2,...,n$, полученных, вообще говоря, нелинейным преобразованием $l_i(y_i)$.

Условные законы распределения каждого слагаемого $l_i(y_i)$ находят с помощью формулы (1.36). Для рассматриваемой задачи выражение (1.36) перепишется в виде:

$$w(l_i/H_{0,1}) = w(y_i/H_{0,1})\left|\frac{dy_i}{dl_i}\right|,\ i=1,2,...,n. \qquad (3.21)$$

Заметим, что в правой части (3.21) необходимо заменить y_i на функцию $y_i = y_i(l_i)$, полученную в результате решения уравнения $l_i = l_i(y_i)$ относительно y_i.

После нахождения ПРВ (3.21) требуется определить законы распределения суммы $z=\sum_{i=1}^{n} l_i$ независимых СВ l_i, $i=1,2,...,n$. Для этого используется либо точный подход, основанный на вычислении характеристических функций (1.39), либо приближенный, базирующийся на центральной предельной теореме теории вероятностей.

Точный расчет характеристик обнаружения осуществляется следующим образом. Вначале находятся характеристические функции слагаемых:

$$g_{l_i}(v/H_0) = \int_{-\infty}^{\infty} w(l_i/H_0)e^{iVl_i}dl_i; \qquad g_{l_i}(v/H_1) = \int_{-\infty}^{\infty} w(l_i/H_1)e^{iVl_i}dl_i. \qquad (3.22)$$

Затем характеристические функции суммы z определяются как произведения характеристических функции слагаемых:

$$g_z(v/H_0) = \prod_{i=1}^{n} g_{l_i}(v/H_0)\ , \qquad g_z(v/H_1) = \prod_{i=1}^{n} g_{l_i}(v/H_1). \qquad (3.23)$$

Наконец, с помощью обратного преобразования Фурье (1.39) вычисляются искомые ПРВ.

$$w(Z/H_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g_z(v/H_0)e^{-ivz}dv, \quad w(Z/H_1) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g_z(v/H_1)e^{-ivz}dv. \quad (3.24)$$

В качестве примера проведем расчет характеристик алгоритма обнаружения (3.11) $\left(z = \sum_{i=1}^{n} y_i^2 > z_0 \right)$, синтезированного для релеевских ПРВ (3.9), (3.10).

При наличии полезного сигнала ПРВ слагаемых $l_i\left(y_i\right) = y_i^2$ могут быть найдены с помощью формулы (3.21)

$$w\left(l_i / H_1\right) = \lambda e^{-\lambda l_i},\ l_i > 0,\ i = 1,2,...,n,$$

где $\lambda = 1/2\sigma^2\left(1+q\right)$. Характеристические функции имеет один и тот же вид $g_{l_i}\left(v/H_1\right) = \int_0^{\infty} \lambda e^{-\lambda l_i} e^{ivl_i} dl_i = \frac{\lambda}{\lambda - iv}$ для всех слагаемых. Поэтому легко находится характеристическая функция суммы z независимых СВ

$$g_z\left(v/H_1\right) = g_{l_i}^n\left(v/H_1\right) = \lambda^n / \left(\lambda - iv\right)^n .$$

Интеграл в обратном преобразовании Фурье (3.24)

$$w\left(z/H_1\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda^n}{\left(\lambda - iv\right)^n} e^{-ivz} dv = \frac{\lambda^n z^{n-1}}{\left(n-1\right)!} e^{-\lambda z}$$

наиболее просто вычисляется с помощью вычетов. Интегрируя последнее выражение еще раз с помощью таблиц [36], получаем следующую расчетную формулу для вероятности правильного обнаружения

$$P_D = \int_{z_0}^{\infty} w\left(z/H_1\right) dz = 1 - \frac{1}{\left(n-1\right)!} \int_0^{\lambda z_0} x^{n-1} e^{-x}\, dx = 1 - \frac{\Gamma\left(n;\ \lambda z_0\right)}{\left(n-1\right)!}\ , \qquad (3.25)$$

где $\Gamma\left(n;\ \lambda z_0\right)$ – неполная гамма-функция, табулированная, например, в [40]; $\lambda = 1/2\sigma^2\left(1+q\right)$.

При отсутствии полезного сигнала изменяется лишь параметр λ, но все приведенные преобразования остаются справедливыми. Поэтому вероятность ложной тревоги также находится по формуле (3.25), если положить, что $q = 0$:

$$P_F = 1 - \frac{\Gamma\left(n;\ z_0/2\sigma^2\right)}{\left(n-1\right)!}. \qquad (3.26)$$

В радиолокационных задачах обнаружения полученные формулы (3.25) и (3.26) обычно используются следующим образом. По заданной вероятности ложной тревоги P_F из соотношения (3.26) определяют порог обнаружения

$z_0/2\sigma^2$. При этом удобно использовать широко распространенные таблицы распределения χ^2[42], поскольку

$$1-\frac{\Gamma(n;\beta)}{(n-1)!}=\frac{1}{2^{(2n/2)}\Gamma(2n/2)}\int_0^{\infty}x^{\frac{2n}{2}-1}e^{-\frac{x}{2}}dx=P(2\beta;2n),$$

где $P(a;b)$ – табулированная функция (распределение χ^2 [42]). После определения $z_0/2\sigma^2$ формула (3.25) позволяет рассчитать характеристики обнаружения, т.е. зависимость вероятности правильного обнаружения P_D от величины отношения сигнал/шум q. Такие характеристики приведены на рис.3.3 для двух значений вероятностей ложной тревоги $P_F=10^{-2}$ и $P_F=10^{-3}$ при $n=10$. Соответствующие значения порога обнаружения $z_0/2\sigma^2=18{,}5$ и $z_0/2\sigma^2=22{,}5$ находятся по формуле (3.26).

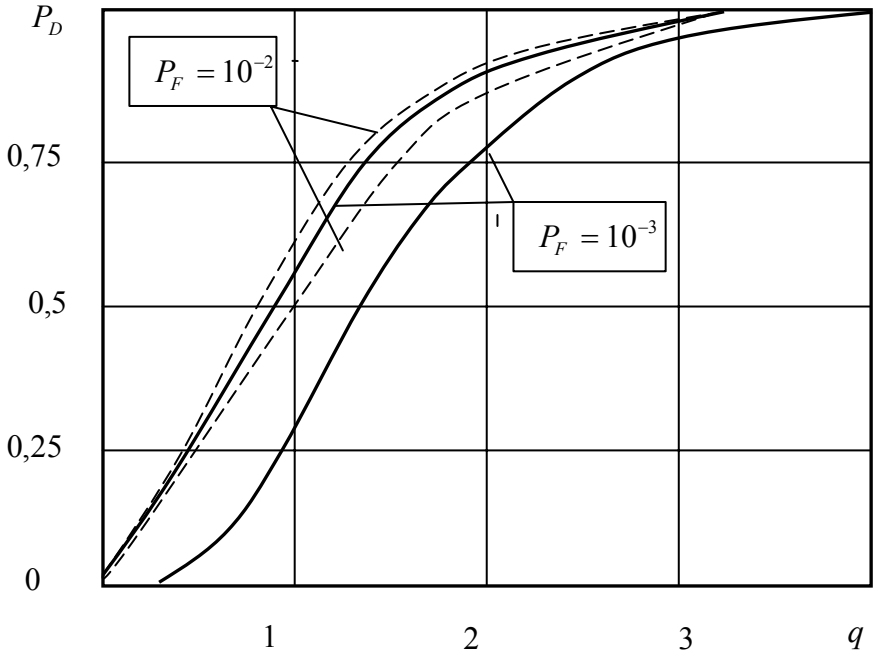


Рис. 3.3. Характеристики обнаружения сигналов

С помощью характеристик обнаружения можно по заданным значениям P_D и P_F определить необходимую величину порогового сигнала q, обеспечивающую требуемое качество обнаружения.

Рассмотренный метод дает возможность рассчитывать точные характеристики обнаружения сигналов. Однако во многих задачах возникают значительные, а иногда и непреодолимые, математические трудности, связанные, чаще

всего, с нахождением обратного преобразования Фурье (3.24). В подобных ситуациях используют приближенный метод расчета характеристик, заключающийся в следующем. Если n велико и дисперсии $M\{(l_i - M\{l_i\})^2\}$ ограничены, то распределение суммы большого числа независимых СВ $z = \sum_{i=1}^{n} l_i(y_i)$ согласно центральной предельной теореме приближается к нормальному [29, 40]:

$$w(z/H_0) \cong \frac{1}{\sqrt{2\pi}\sigma_{Z0}} e^{-\frac{(z-m_{Z0})^2}{2\sigma_{Z0}^2}}, \qquad w(z/H_1) \cong \frac{1}{\sqrt{2\pi}\sigma_{Z1}} e^{-\frac{(z-m_{Z1})^2}{2\sigma_{Z1}^2}}, \tag{3.27}$$

где m_{Z0}, σ_{Z0}^2 и m_{Z1}, σ_{Z1}^2 – условные математические ожидания и дисперсии z, когда справедливы гипотезы H_0 и H_1 соответственно. Параметры (3.27) обычно могут быть вычислены достаточно просто, поскольку

$$m_{Z0} = \sum_{i=1}^{n} M\{l_i/H_0\}, \qquad m_{Z1} = \sum_{i=1}^{n} M\{l_i/H_1\}$$

$$\sigma_{Z0}^2 = \sum_{i=1}^{n} \left(M\{l_i^2/H_0\} - M^2\{l_i/H_0\}\right), \qquad \sigma_{Z1}^2 = \sum_{i=1}^{n} \left(M\{l_i^2/H_1\} - M^2\{l_i/H_1\}\right)$$

причем, с учетом основной теоремы (1.37) о математическом ожидании,

$$M\{l_i/H_{0,1}\} = \int_{-\infty}^{\infty} l_i(y_i) w(y_i/H_{0,1}) dy_i, \quad M\{l_i^2/H_{0,1}\} = \int_{-\infty}^{\infty} l_i^2(y_i) w(y_i/H_{0,1}) dy_i. \tag{3.28}$$

После выполнения указанных преобразований искомые вероятности

$$P_F \cong \int_{z_0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{Z0}} e^{-\frac{(z-m_{Z0})^2}{2\sigma_{Z0}^2}} dz = \frac{1}{2} - \Phi_0\left(\frac{z_0 - m_{Z0}}{\sigma_{Z0}}\right), \qquad P_D \cong \frac{1}{2} - \Phi_0\left(\frac{z_0 - m_{Z1}}{\sigma_{Z1}}\right) \tag{3.29}$$

находятся по таблицам функции Лапласа. Рассматривая в качестве примера правило $z = \sum_{i=1}^{n} y_i^2 \gtrless z_0$ обнаружения релеевского сигнала, запишем последовательно

$$M\{l_i = y_i^2/H_1\} = 1/\lambda, \lambda = 1/2\sigma^2(1+q), \qquad M\{l_i^2/H_1\} = 2/\lambda^2, m_{Z1} = n/\lambda, \sigma_{Z1}^2 = n/\lambda^2,$$

$$P_F \cong 0.5 - \Phi_0\left(Z_0/2\sigma^2\sqrt{n}\right) - \sqrt{n}, \qquad P_D = 0{,}5 - \Phi_0\left(\frac{z_0}{2\sigma^2(1+q)\sqrt{n}} - \sqrt{n}\right). \tag{3.30}$$

Полагая $n = 10, P_F = 10^{-2}$ и $P_F = 10^{-3}$, по таблицам функции Лапласа [40] находим $Z_0/2\sigma^2 = 17{,}33$ и $Z_0/2\sigma^2 = 19{,}76$ соответственно. Сравнивая теперь эти зна-

чения с пороговыми уровнями $Z_0/2\sigma^2 = 18{,}5$ и $Z_0/2\sigma^2 = 22{,}5$, рассчитанными с помощью точного соотношения (3.26), видим, что погрешность выше при меньшей вероятности ложной тревоги.

Для оценки применимости метода аппроксимации нормальным распределением в рассматриваемом примере на рис.3.3 нанесены пунктирные кривые, найденные с помощью приближенной формулы (3.30). Анализ приведенных зависимостей показывает, что приближенный метод приводит к значительным погрешностям при вероятности ложной тревоги $P_F < 10^{-3}$. Вместе с тем погрешность при $P_F < 10^{-2}$ во многих задачах может считаться допустимой. Кроме того, следует отметить, что приведенные погрешности соответствуют относительно малому значению $n = 10$, принятому в данной задаче. При обработке большего числа наблюдений погрешности за счет нормальной аппроксимации заметно уменьшаются, и при $n > 30 \div 100$ точность приближенного метода, как правило, становится удовлетворительной.

К сожалению, в общем случае нельзя дать достаточно надежную аналитическую оценку погрешности, возникающей при замене действительного распределения суммы $z = \sum_{i=1}^{n} l_i(y_i)$ нормальным. Поэтому при использовании приближенного метода расчета характеристик обнаружения необходимо применять те или иные приемы обеспечения достаточной степени уверенности в справедливости найденных результатов. Одним из таких приемов является метод статистического моделирования [6]. Суть метода заключается в формировании с помощью ЭВМ последовательности N псевдослучайных выборок $\bar{y}^{(1)}, \bar{y}^{(2)}, ..., \bar{y}^{(N)}$ с ПРВ $w(y_i/H_0)$, где $\bar{y}^{(j)} = \left(y_1^{(j)} y_2^{(j)} ... y_n^{(j)}\right)^T$, вычислении для каждой выборки суммы $z^{(j)} = \sum_{i=1}^{n} l_i\left(y_i^{(j)}\right),\ j = 1,2,...,N$, и построении на основе случайных чисел $\left\{z^{(j)}\right\},\ j = 1,2,...,N$, гистограммы $w^*(z/H_0)$, аппроксимирующей искомую ПРВ $w(z/H_0)$. Совершенно аналогично формируется гистограмма $w^*(z/H_1)$, позволяющая дать оценку P_D^* вероятности правильного обнаружения P_D. При этом

погрешности оценивания вероятностей P_F и P_D зависят лишь от величин P_D или P_F и числа N экспериментов, т.е., в принципе, могут быть сделаны сколь угодно малыми при достаточно больших объемах вычислений на ЭВМ.

Действительно, рассмотрим оценку P_D, в качестве которой используется частота $P_D^* = k/N$, где k – число превышений суммой z порогового уровня z_0 в серии из N опытов. Поскольку k подчиняется биномиальному закону распределения (1.7) с параметром $p = P_D$, то дисперсия ошибки оценивания вероятности правильного обнаружения определяется следующим образом:

$$M\left\{\left(P_D^* - P_D\right)^2\right\} = M\left\{\left(\frac{k - NP_D}{N}\right)^2\right\} = \frac{1}{N^2} M\left\{\left(k - NP_D\right)^2\right\} = P_D\left(1 - P_D\right)/N .$$

Аналогично и $M\left\{\left(P_F^* - P_F\right)^2\right\} = P_F\left(1 - P_F\right)/N$. Итак, задавая погрешности оценивания P_D или P_F, можно с помощью этих формул определить необходимое число N повторений эксперимента.

Метод статистического моделирования во многих случаях требует проведения очень большого числа экспериментов и, следовательно, значительного машинного времени. Например, при $P_F = 10^{-4}$, $\sqrt{M\left\{\left(P_F^* - P_F\right)^2\right\}}\Big/P_F = 0{,}01$ получаем $N = 10^8$, иобщее количество $10^8 \cdot n$ формируемых на ЭВМ псевдослучайных чисел, а также операций по вычислению $l_i\left(y_i^{(j)}\right)$ весьма велико. Для современных ЭВМ решение задач статистического моделирования часто требует десятков или сотен часов непрерывной работы. Поэтому анализ помехоустойчивости радиосистем требует в сложных случаях искусного сочетания аналитических методов и экспериментов на ЭВМ.

3.2. Оптимальная демодуляция при когерентном приеме сигналов

3.2.1. Оптимальные алгоритмы приема при полностью известных сигналах

Методы приема, для реализации которых необходимо точное знание на-

чальных фаз приходящих сигналов, называют когерентными.

Предположим что анализируемый сигнал ограничен во времени интервалом $[0,T]$ и передается в условиях воздействия гауссовского аддитивного белого шума $n(t)$, со спектральной плотностью N_0. Это значит, что при передаче символа $\boldsymbol{x_k}$ принимаемое напряжение имеет вид:

$$U(t) = S_k(t) + n(t), \quad 0 \le t \le T . \tag{3.31}$$

Определим в этих условиях алгоритм работы оптимального демодулятора, основанного на правиле максимального правдоподобия:

$$\Lambda_k(\vec{U}) > \Lambda_r(\vec{U}), \qquad r = 1,2,...m, \qquad r \ne k \tag{3.32}$$

Рассмотрим вначале левую часть неравенства, где определим знаменатель, а затем числитель функции, где

$$\Lambda_k(\vec{U}) = \frac{w\left(\vec{U}/x_k\right)}{w\left(\vec{U}/x_0\right)}, \qquad k = 1,2,...m . \tag{3.33}$$

Многомерная ПРВ при независимых наблюдениях равна произведению одномерных ПРВ [5, 32, 39]:

$$w\left(\vec{U}/x_0\right) = \prod_{i=1}^{2FT} w\left(U_i/x_0\right) = \prod_{i=1}^{2FT} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{U_i^2}{2\sigma^2}} = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{2FT}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{2FT} U_i^2} . \tag{3.34}$$

Поскольку: $\sum_{i=1}^{2FT} U_i^2 = 2F\int_0^T U^2(t)dt$, $\qquad \sigma^2/F = N_0$, то

$$w\left(\vec{U}/x_0\right) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{2FT}} \cdot e^{-\frac{1}{N_0}\int_0^T U^2(t)dt} . \tag{3.35}$$

Многомерная ПРВ в числителе (3.33) определяется аналогично:

$$w\left(\vec{U}/x_k\right) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^{2FT}} \cdot e^{-\frac{1}{N_0}\int_0^T [U(t)-S_k(t)]^2 dt} . \tag{3.36}$$

Подставляя (3.35) и (3.36) в (3.33), находим:

$$\Lambda_{k/0}(\vec{U}) = \frac{w\left(\vec{U}/x_k\right)}{w\left(\vec{U}/x_0\right)} = e^{-\frac{1}{N_0}\int_0^T [U(t)-S_k(t)]^2 dt} \cdot e^{\frac{1}{N_0}\int_0^T U^2(t)dt} , \tag{3.37}$$

Учитывая, что энергия k-го сигнала на выходе канала:

$$E\left(S_k(t)\right) = E_k = \int_0^T S_k^2(t)dt,$$

получим следующее выражение:

$$\Lambda_{k/0}\left(\vec{U}\right) = e^{\frac{2}{N_0}\int_0^T U(t)S_k(t)dt} \cdot e^{-\frac{E_k}{N_0}}. \qquad (3.38)$$

Правило принятия решения (3.32) можно записать в логарифмической форме:

$$\ln \Lambda_{k/0}\left(\vec{U}\right) > \ln \Lambda_{r/0}\left(\vec{U}\right). \qquad (3.39)$$

Подставив (3.37) в (3.39) и умножив обе части неравенства на N_0, получим:

$$\int_0^T \left[U(t) - S_k(t)\right]^2 dt < \int_0^T \left[U(t) - S_r(t)\right]^2 dt, \qquad (3.40)$$

Проделав аналогичные операции с (3.38) и (3.39), и учитывая $P_k = E_k/T$, получим:

$$\frac{1}{T}\int_0^T U(t)S_k(t)dt - \frac{P_k}{T} > \frac{1}{T}\int_0^T U(t)S_r(t)dt - \frac{P_r}{T}. \qquad (3.41)$$

Полученные неравенства представляют собой оптимальное правило принятия решения, согласно которому следует принять решение о передаче символа x_k (сигнала $S_k(t)$), если неравенство выполняется для всех $k \neq r$.

3.2.2. Реализация алгоритмов оптимального когерентного приема

Реализация полученных алгоритмов оптимального когерентного приема может быть представлена в виде функциональных схем, состоящих из m ветвей обработки входного напряжения $U(t)$ в соответствии с правилами (3.40) и (3.41) и устройств сравнения, определяющих номер k-ой ветви в момент $t = T$.

Функциональные узлы могут быть реализованы на аналоговой или цифровой элементной базе.

Генераторы опорных сигналов должны формировать сигналы, совпадающие с соответствующими реализациями сигналов, поступающих на вход демодулятора из линии связи.

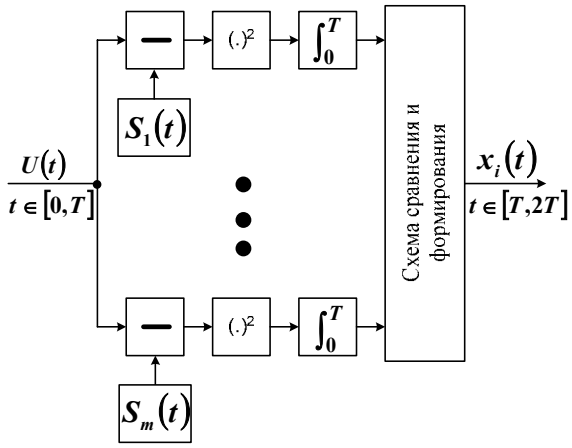


Рис. 3.4. Схема с квадраторами

Выполнение неравенства (3.40) означает, что принимаемое напряжение «ближе» всего к образцу сигнала $S_k(t)$. Функциональная схема оптимальной обработки в соответствии с правилом (3.40) называется схемой с квадраторами (рис.3.4). В момент окончания обработки $t = T$ производится сравнение выходов m ветвей интегрирования квадратов разностей входного сигнала $U(t)$ и образцов сигнала $S_k(t)$, $k = 1,2,...m$ и делается выбор номера k-й ветви по минимуму напряжения. Таким образом принимается решение о том, что передавался k-й сигнал и в момент времени $t = T$ формируется соответствующее сообщение $x_k(t)$.

Функциональная схема оптимальной обработки в соответствии с правилом (3.41) объединяющая генератор опорного сигнала $S_i(t)$, перемножитель и интегратор называется схемой на корреляторах (рис. 3.5).

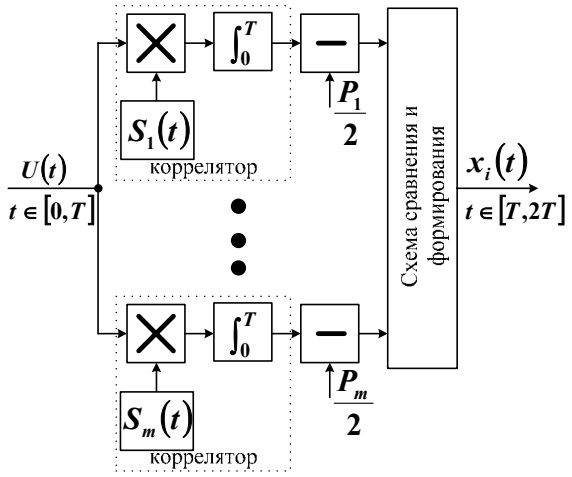


Рис. 3.5. Схема на корреляторах

При корреляционном способе используются образцы сообщения, хранящиеся в памяти приемного устройства.

В соответствии с правилом (3.41) входное напряжение $U(t)$ в пределах интервала наблюдения $[0,T]$ перемножается со всеми эталонными реализациями $S_i(t)$ и результат интегрируется на промежутке $[0,T]$. В момент $t = T$ из значений интегралов вычитаются слагаемые $P_i/2$, $i = 1,2,...m$ и выбирается наибольший результат.

Наиболее трудно выполнимым требованием в рассмотренных алгоритмах приема сигналов является обеспечение точного фазирования опорных напряжений $S_i(t)$ и точного совпадения формы. В то же время есть возможность построения оптимального демодулятора с помощью замены коррелятора линейным фильтром с теми же свойствами. Напряжение на выходе любого линейного фильтра в момент времени $t = T$ определяется интегралом Дюамеля [5, 39]:

$$U_{вых}(T) = \int_0^T U(t) \cdot g(T - t)dt , \qquad (3.42)$$

где $g(t)$ – импульсная переходная характеристика фильтра.

Напряжение на выходе коррелятора

$$Z_i(T) = \frac{1}{T}\int_0^T U(t) \cdot S_i(t)dt . \qquad (3.43)$$

Сравнение (3.42) и (3.43) показывает, что напряжения $Z_i(t)$ и $U_{вых}(T)$ совпадают, если импульсная реакция фильтра удовлетворяет условию:

$$g_i(t) = \frac{1}{T} S_i(T - t) . \qquad (3.44)$$

Линейный фильтр обладающий импульсной реакцией вида (3.44), называется фильтром согласованным с сигналом $S_i(t)$.

Таким образом, функциональная схема оптимальной обработки на корреляторах (рис. 3.5) заменяется схемой на согласованных фильтрах (рис. 3.6).

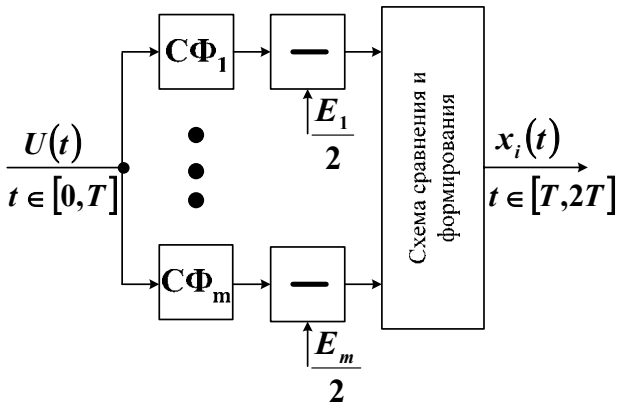


Рис. 3.6. Схема на согласованных фильтрах

В отличие от схемы на корреляторах, в схеме на СФ не нужны генераторы опорных напряжений $S_i(t)$, точно сфазированные с приходящим сигналом. Однако в схеме на СФ должна обеспечена высокая точность выбора момента отсчета $t = T$, что создает трудности для практической реализации.

Из-за неравномерности АЧХ и нелинейности ФЧХ форма напряжения на

выходе СФ значительно отличается от формы входного сигнала. Однако при приеме дискретных сигналов возможная их форма заранее известна и требуется определить лишь номер переданной реализации.

3.3. Помехоустойчивость приема сигналов с известными параметрами

Помехоустойчивость дискретного канала связи определяется вероятностью ошибочного приема $p_{ош}$ сигналов. Рассмотрим определение $p_{ош}$ для двоичного канала связи *m=2*, по которому передается один из двух сигналов $S_1(t)$ или $S_2(t)$. Перепишем правило принятия решения в виде:

$$\text{Если} \quad 2\int_0^T U(t)\cdot[S_1(t)-S_2(t)]dt-[E_1-E_2]>0\,, \tag{3.45}$$

то принимается решение в пользу $S_1(t)$.

Пусть на вход схемы оптимальной обработки поступает сигнал $U(t)=S_1(t)+n(t)$. В этом случае правильное решение принимается, если неравенство выполняется, и ошибочное, если

$$2\int_0^T S_1(t)\cdot[S_1(t)-S_2(t)]dt+2\int_0^T n(t)\cdot[S_1(t)-S_2(t)]dt-E_1+E_2<0\,. \tag{3.46}$$

Раскроем скобки первого интеграла, и представим энергию первого и второго сигналов, а также взаимную корреляционную функцию соответственно следующим образом:

$$E_1=\int_0^T S_1^2(t)dt\,,\qquad E_2=\int_0^T S_2^2(t)dt\,,\qquad E_{12}=\int_0^T S_1(t)\cdot S_2(t)dt\,.$$

неравенство (3.46) принимает вид

$$E_0+2\int_0^T n(t)\cdot[S_1(t)-S_2(t)]dt<0\,, \tag{3.47}$$

где $E_0=E_1+E_2-2E_{12}$ суммарная (эквивалентная) энергия двух сигналов.

Последнее слагаемое в правой части неравенства (3.47) гауссовская случайная величина с нулевым математическим ожиданием (средним) и дисперси-

ей $\sigma^2 = 2N_0$ таким образом $\eta = E_0 + 2\int_0^T n(t)\cdot[S_1(t) - S_2(t)]dt < 0$ будет иметь гауссовское распределение со средним E_0 и дисперсией σ^2 (рис. 3.7).

Вероятность ошибок (рис. 3.7) для рассматриваемого двоичного канала связи, определяется площадью, ограниченной ПРВ $w(\eta)$ и осью абсцисс для всех $\eta \le 0$.

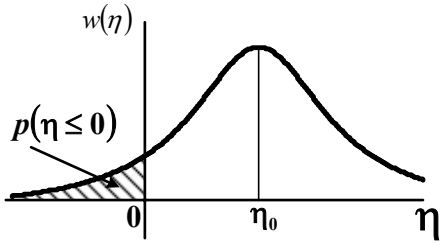


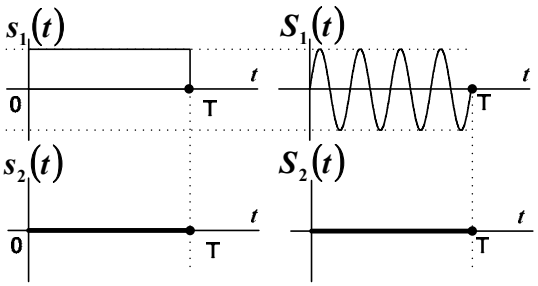
Рис. 3.7. График плотности распределения вероятности

Формула, характеризующая вероятность ошибочного приема $S_1(t)$ (т.е. принятия решения о передаче $S_2(t)$, когда передавался $S_1(t)$), будет следующей:

$$p_{ош} = \int_{-\infty}^{-0,5E_0} w(\eta)d\eta = 1 - F\left(\sqrt{\frac{E_0}{2N_0}}\right) = 1 - F\left(\sqrt{\frac{E_1 + E_2 - 2E_{12}}{2N_0}}\right). \qquad (3.48)$$

Из (3.48) следует, что вероятность ошибочного приема элементов двоичного сообщения тем меньше, чем больше эквивалентная энергия E_0 и чем меньше спектральная плотность мощности помех N_0.

Оценим влияние структуры передаваемых сигналов на вероятность их ошибочного приема. Если сигналы близки по форме $S_1(t) \approx S_2(t)$,то $E_1 = E_2 = E_{12}$, $E_0 = 0$, и вероятность ошибки максимальна $p_{ош} = 0{,}5$. Такие сигналы разделить невозможно, надо использовать сигналы $S_1(t)$ и $S_2(t)$ значительно отличающихся друг от друга. Рассмотрим несколько видов сигналов, применяющихся в системах связи. Поэтому применение сигналов, близких по форме, нецелесообразно. Наиболее простым является сигнал с пассивной паузой амплитудно - модулированный: $S_1(t) = S(t)$, $S_2(t) = 0$ (рис. 3.8).

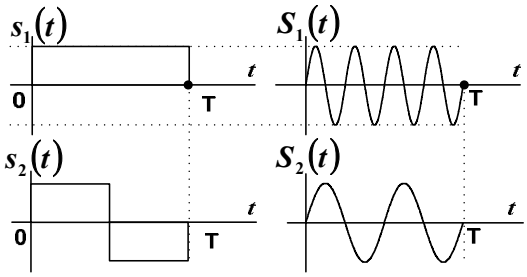


3.8. Сигналы с пассивной паузой

Тогда энергия первого сигнала равна E: $E_1 = E$, а энергия второго и взаимная энергия сигналов равны нулю: $E_2 = E_{12} = 0$, тогда $E_0 = E$.

Вероятность ошибки определяется выражением:

$$p_{ош} = 0.5 - \Phi_0\left(\sqrt{\frac{E}{2N_0}}\right). \qquad (3.49)$$

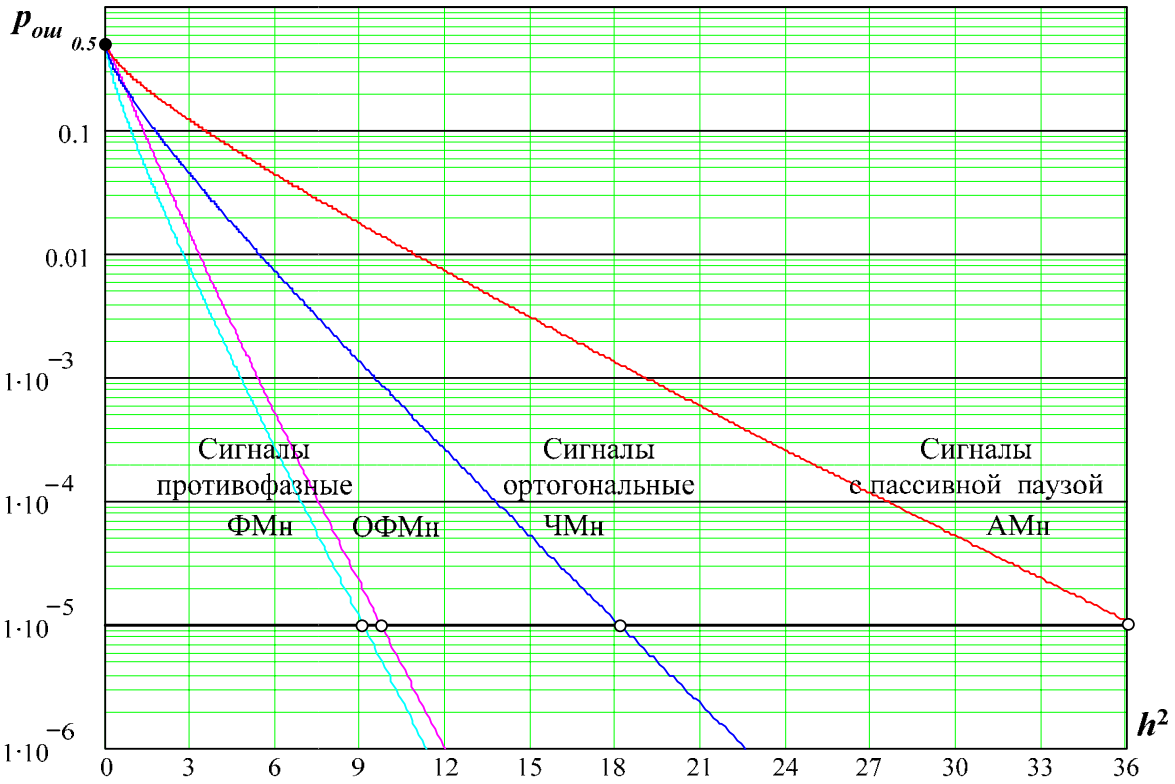


3.9. Ортогональные сигналы

Определим $p_{ош}$ для ортогональных сигналов $S_1(t)$ и $S_2(t)$ (рис. 3.9). Пусть $E_1 = E_2 = E$. Согласно условию ортогональности $E_{12} = 0$, тогда $E_0 = 2E$. При этом вероятность ошибки:

$$p_{ош} = 0.5 - \Phi_0\left(\sqrt{\frac{E}{N_0}}\right). \qquad (3.50)$$

На рис. 3.10 представлены кривые зависимости вероятности ошибок от отношения сигнал/шум для сигналов: АМн, ЧМн, ФМн, ОФМн.



3.10. Зависимость вероятности ошибки от величины отношения сигнал/шум для сигналов АМн, ЧМн, ФМн, ОФМн

Для противоположных сигналов: $S_1(t) = S(t)$, а $S_2(t) = -S(t)$ (рис. 3.11)

Тогда $E_1 = E_2 = E$, $E_{12} = -E$, $E_0 = 4E$.

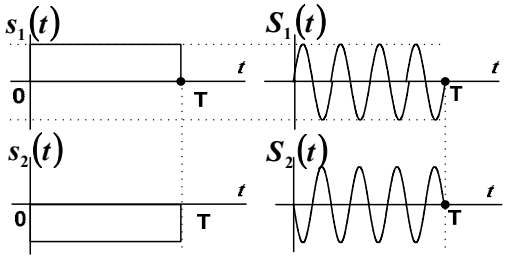
Вероятность ошибки определяется выражением [21, 32, 39]:

$$p_{ош} = 0.5 - \Phi_0\left(\sqrt{\frac{2E}{2N_0}}\right). \qquad (3.51)$$

В этом случае расчет вероятности ошибки для ОФМн сигналов производится по формуле:

$$p_{ОФМн}^{КГ} := 2 \cdot p_{ФМн}^{КГ} \cdot \left(1 - p_{ФМн}^{КГ}\right). \qquad (3.52)$$

Сравнение различных по форме сигналов (рис. 3.9 – 3.11) показывает, что помехоустойчивость противофазных сигналов может быть получена при затратах энергии в 2 раза меньшей, чем при ортогональных сигналах и в 4 раза меньшей, чем при передаче сообщений сигналами с пассивной паузой.



3.11. Противоположные (противофазные) сигналы

3.4. Прием сигналов с неопределенной фазой

Когерентный прием дискретных сигналов основан на точном знании фазы возможных реализаций переданного сигнала на входе демодулятора. Однако при передаче сигналов по каналам радиосвязи фаза принимаемого сигнала обычно является случайной величиной, принимающей значения в пределах от 0 до 2π.

3.4.1. Оптимальный некогерентный прием дискретных сигналов

Способ когерентной обработки сигналов относится к идеализированным условиям, когда неизвестно, какая из заданных реализаций была передана. Форма реализаций, момент прихода и мощность достоверно известны.

Изменение параметров канала связи (μ, τ), изменение режима передающего устройства и ряд других факторов приводят к тому, что некоторые параметры реализаций сигнала делаются случайными и могут быть оценены с неко-

торой погрешностью. Такая ситуация особенно характерна для радиоканалов. Получение алгоритмов оптимальной обработки сигналов со случайными параметрами — проблема значительно более сложная. При ее решении пользуются двумя приемами.

Первый прием характерен для ситуации, когда некоторый случайный параметр γ меняется медленно и за время действия одной реализации его можно считать неизменным. Тогда правило обработки, и схема приемника сохраняются, а параметр для каждого следующего решения экстраполируется по множеству его предыдущих значений. Обычно это удается выполнить с помощью введения автоматических регулировок. Так, например, при неизвестной амплитуде сигнала вводится автоматическая регулировка порогового уровня.

Второй прием применяется в условиях, если параметр γ меняется быстро и за время действия реализации T он может значительно изменяться. Для получения правила оптимальной обработки в этих условиях составляют функцию правдоподобия $\Lambda_{1,2}(\gamma)$, как это делалось ранее. Поскольку функция правдоподобия оказывается зависящей от параметра $W(\gamma)$, ее усредняют. Для усреднения необходимо знать плотность распределения вероятностей $W(\gamma)$ случайного параметра γ. Тогда правило оптимальной обработки принимает вид

$$\int_{-\infty}^{\infty} W(\gamma)\,\Lambda_{1,2}(\gamma)d\ \geq C_0\,,$$

где величина порога C_0 зависит от выбранного критерия.

Оптимальный прием сигнала в условиях, когда начальная фаза сигнала является величиной случайной, называют оптимальным некогерентным приемом.

Обычно предполагают, что в пределах периода начальная фаза имеет равновероятное распределение, так что плотность распределения постоянна и равна

$$W(\varphi)=\frac{1}{2\pi}\qquad(\varphi\in[0,2\pi]).$$

При этом правило оптимальной обработки имеет вид:

$$\ln I_0\left(\frac{2TV_1(u)}{N_0}\right)-\frac{P_1T}{N_0}\geq \ln I_0\left(\frac{2TV_2(u)}{N_0}\right)-\frac{P_2T}{N_0}\,, \qquad (3.53)$$

I_0 — модифицированная функция Бесселя нулевого порядка; $V(u)$ — огибающая некоторого процесса $Z(u)$, зависящего от принимаемого сигнала.

Если неравенство (3.53) выполняется, то принимается решение пользу реализации $s_1(t)$.

Прежде чем перейти к правилам обработки сигнала, остановимся несколько подробнее на смысле функции $V(u)$. Рассмотрим реализацию сигнала на входе приемника:

$$c_1(t)=\mu s_1(t).$$

Если в ее спектре все составляющие сдвинуты по фазе на — 90°, то получится реализация $\widetilde{c}_1(t)$, ортогональная с $c_1(t)$ и однозначно ней связанная, которую называют сопряженной реализацией с $c_1(t)$.

Образуем взаимокорреляционные функции принимаемого сигала с $c_1(t)$ и $\widetilde{c}_1(t)$:

$$\left.\begin{aligned} Z_1(u)&=\frac{1}{T}\int_0^T u(t)c_1(t)dt;\\ \widetilde{Z}_1(u)&=\frac{1}{T}\int_0^T u(t)c_1(t)dt. \end{aligned}\right\} \qquad (3.54)$$

Тогда

$$V_1(u)=\sqrt{Z_1^2(u)+\widetilde{Z}_1^2(u)}\,. \qquad (3.55)$$

На рис. 3.12 приведена функциональная схема устройства оптимальной обработки принимаемых сигналов при случайной фазе. Она содержит генераторы реализаций $c_1(t)$ и $c_2(t)$, от которых формируются сопряженные с ними реализации $\widetilde{c}_1(t)$ и $\widetilde{c}_2(t)$. Затем путем перемножения, интегрирования, возведения в квадрат и суммирования формируются V_r^2. Результаты суммирования поступают в блок нелинейного преобразования, который осуществляет умножение на

коэффициенты, извлечения корня, взятие функции Бесселя и логарифмирование.

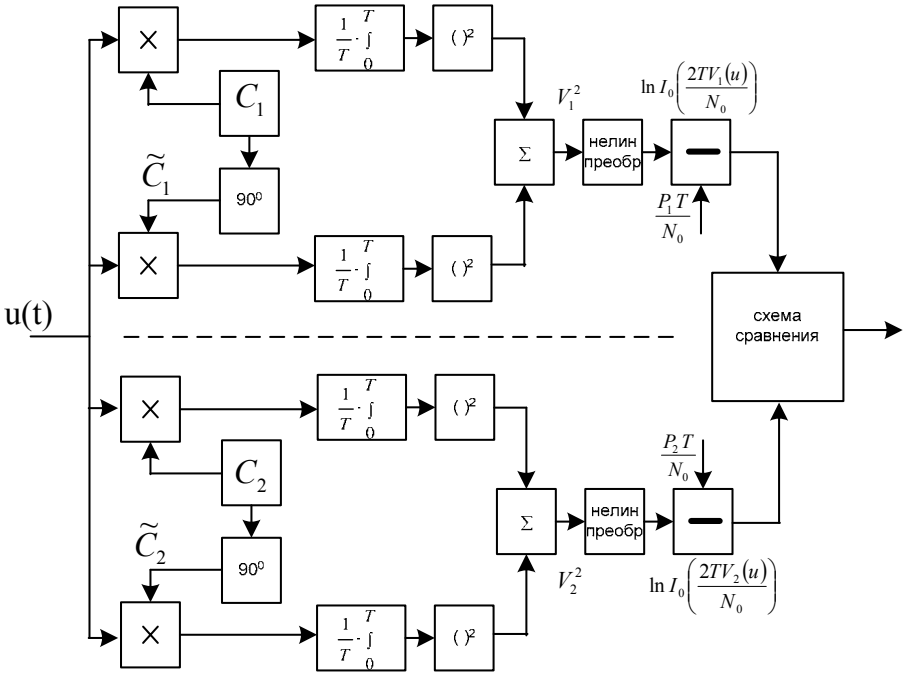


Рис. 3.12. Схема устройства оптимальной обработки принимаемых сигналов при случайной фазе

Колебание с выхода нелинейного блока поступает на пороговое устройство, где осуществляется вычитание порогового уровня, равного $\frac{P_rT}{N_0}$. После этого результаты сравниваются, и принимается решение в соответствии с правилом (3.53). Заметим, что, как и в схемах оптимальной когерентной обработки, пороговые устройства будут отсутствовать, если работа идет с активной паузой и $P_1 = P_2 = P_c$. Однако при этом правило оптимальной обработки упрощается еще более. Действительно, в силу монотонности линейных операций правило выполняется тогда, когда

$$\frac{2V_1T}{N_0} \geq \frac{2V_2T}{N_0}$$

и, следовательно,

$$V_1 \geq V_2 . \qquad (3.56)$$

Таким образом, отпадает необходимость в блоке нелинейного преобразования.

Более просто выглядит схема устройства оптимальной обработки, если удается создать согласованные с реализациями $c_1(t)$ и $c_2(t)$ фильтры. На рис. 3.13 приведена такая схема с пороговыми устройствами.

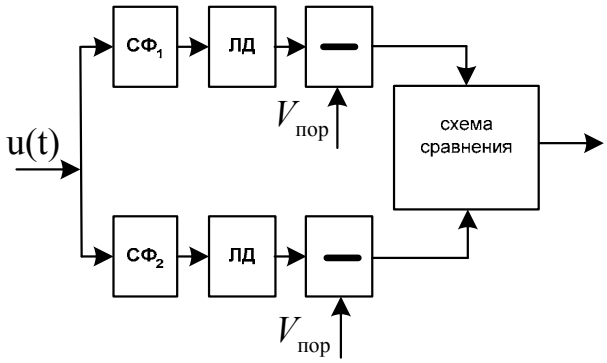


Рис. 3.13. Схема с пороговыми устройствами

В случае $P_1 = P_2 = P_c$ пороговые устройства отсутствуют, и детекторная характеристика может быть любой. Важно лишь, чтобы эти характеристики для обоих детекторов были одинаковыми и монотонно нарастающими. В этом случае входной сигнал $u(t)$ поступает на согласованные фильтры, отклики фильтров детектируются и в момент отсчета результаты сравниваются. Решение принимается в пользу той реализации, у которой огибающая на выходе согласованного фильтра оказалась больше.

Вероятность ошибки при оптимальном некогерентном приеме, естественно, больше, чем при когерентном:

$$P_{\text{ош}} = \frac{1}{2} e^{-\frac{h^2}{2}}. \qquad (3.57)$$

Рассмотрим примеры некогерентного приема двоичных радиосигналов, т. е. таких сигналов, которые могут быть представлены отрезками гармонических колебаний. При этом будем считать, что значение начальной фазы сигнала, действующего на входе приемника, неизвестно.

Амплитудная манипуляция

Для АМн, как уже отмечалось, реализации сигналов записывают следующим образом:

$$c_1(t) = \mu A \cos(\omega_c t + \varphi);$$

$$c_2 t = 0,$$

где φ — случайная величина, а $t \in [0, T]$.

Правило оптимальности обработки (8.34) принимает вид

$$\ln I_0\left(\frac{2TV_1(u)}{N_0}\right) \geq \frac{P_1 T}{N_0}. \qquad (3.58)$$

Если неравенство (3.58) выполняется, то решение принимается в пользу реализации $s_1(t)$.

Правило (3.58) может быть реализовано в виде корреляционного приемника или в виде приемника на согласованных фильтрах. Структурные схемы таких устройств изображены в верхней части схем рис. 3.12 и 3.13, отделенной пунктирной линией.

Таким образом, после выделения огибающей и нелинейного ее преобразования производится сравнение результатов с пороговым уровнем. Если результат нелинейного преобразования превышает порог, то принимается решение о том, что передавалась реализация $s_1(t)$. Заметим, что пороговый уровень определяется мощностью сигнала и спектральной плотностью мощности шума. И то и другое может быть оценено в реальных условиях лишь с определенной точностью и к тому же имеет тенденцию к изменению во времени.

На рис. 3.14 изображены условные плотности распределения огибающих при передаче $s_1(t)$ и $s_2(t)$. Они подчиняются обобщенному закону Рэлея при передаче $s_1(t)$ или просто закону Рэлея, если передается $s_2(t)$ и фактически речь идет о законе распределения огибающей белого шума.

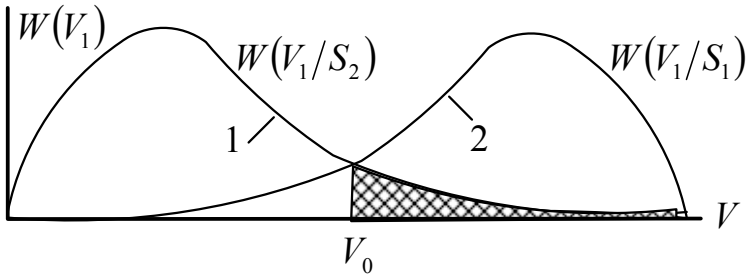


Рис. 3.14 Условные плотности распределения огибающих

при передаче $s_1(t)$ и $s_2(t)$

Пороговый уровень на практике выбирают по точке пересечения кривых распределения, для которой

$$W(V_1 / S_1) = W(V_1 / S_2).$$

Это достаточно близко к теоретическому значению уровня, вытекающему из правила (3.58), а при $q_1 = q_2$ такой выбор порога является оптимальным.

Вероятность ошибки определяется как сумма площадей, ограниченных осью абсцисс, кривой *1* (для $V_1 \rangle V_0$) и кривой *2* (для $V_1 \langle V_0$). На рис. 3.14 эти площади заштрихованы. Приближенно можно пользоваться и формулой (3.57).

Частотная манипуляция

При ЧМн реализации сигнала записываются в виде:

$$\begin{aligned} s_1(t) &= A\cos(\omega_1 t + \varphi_1) \quad ; \\ s_2(t) &= A\cos(\omega_2 t + \varphi_2) \quad , \end{aligned} \tag{3.59}$$

где φ_1 и φ_2 — случайные фазы, а $t \in [0, T]$.

Рассматриваемый случай отличается от АМн тем, что работа осуществляется с активной паузой. Что касается реализации оптимальных фильтров или устройств выделения огибающей V корреляционным способом, то они такие же, как при АМн.

Структурные схемы оптимальной обработки принимаемых сигналов с неизвестной начальной фазой при ЧМн полностью совпадают со схемами, изображенными на рис. 3.12 и 3.13. Генераторы Γ_1 и Γ_2 (см. рис. 3.12) генерируют косинусоиды с нулевыми начальными фазами и с частотами соответственно ω_1 и ω_2. В схеме, приведенной на рис. 3.13, оптимальные фильтры должны быть согласны с такими реализациями. Принцип работы подобных схем описан выше. Отметим лишь один существенный момент. В системах с активной паузой различие форм реализаций существенно влияет на вероятность ошибки. Минимум вероятности ошибки имел место при противоположных реализациях, когда

$s_1(t) = -s_2(t)$. При частотной телеграфии создать противоположные реализации невозможно, однако можно добиться их ортогональности, если выбрать

$$\omega_1 - \omega_2 = \frac{2\pi}{T} k ,$$

или

$$f_1 - f_2 = \frac{k}{T} = \frac{k\, u_k}{2}, \qquad (3.60)$$

где $k = 1,2,3,...$; $u_k = 2/T$ — скорость работы в бодах, равная числу символов, передаваемых по каналу связи в одну секунду.

Относительно-фазовая манипуляция

Естественно, что фазовую телеграфию, обеспечивающую минимум вероятности ошибки, трудно осуществить при случайной фазе, так как значение фазы реализации сигнала несет информацию передаваемом сообщении. При ОФМн информация заключена в значении фазы, а в ее изменении при переходе от одной реализации к другой. Для того чтобы принять решение о том, какой из символов передавался, необходимо анализировать принимаемый сигнал не в течение времени T, равного длительности одной реализации, а за промежуток времени, в два раза больший. Возможны два взаимно исключающих случая.

1. За промежуток времени ($-T \le t \le T$) начальная фаза реализаций принимаемого сигнала не изменилась и, следовательно, можно записать:

$$\sigma_1(t) = \mu\, A\cos(\omega_c t + \varphi) \qquad -T \le t \le T , \qquad (3.61)$$

где φ — случайная фаза.

2. В момент времени $t = 0$ произошла смена фазы и, следовательно, для принимаемых реализаций сигнала справедливо условие

$$\sigma_1(t) = \begin{cases} \mu\, A\cos(\omega_c t + \varphi) & (-T \le t \le 0); \\ \mu\, A\cos(\omega_c t + \varphi + \pi) & (0 < t \le T). \end{cases} \qquad (3.62)$$

Отсюда следует, что ОФМн можно представить, как некоторый вид работы с активной паузой, когда реализации сигнала $\sigma_1(t)$ и $\sigma_2(t)$ описываются выражениями (3.61) и (3.62). Таким образом, можно синтезировать устройство обра-

ботки принимаемых сигналов по общим правилам некогерентного приема, считая, что длительность реализаций равна $2T$ (хотя отсчеты производятся через промежуток времени равный T).

Правило (3.53) для ОФМн принимает вид

$$V_1 \geq V_2 \quad \text{или} \quad X_a\, X_b + Y_a\, Y_b \geq 0\,. \tag{3.63}$$

Здесь

$$\begin{aligned} X_a &= \int_{-T}^{0} u(t)\cos\omega_c t\, dt\,; \quad X_b = \int_{0}^{T} u(t)\cos\omega_c t\, dt\,; \\ Y_a &= \int_{-T}^{0} u(t)\sin\omega_c t\, dt\,; \quad Y_b = \int_{0}^{T} u(t)\sin\omega_c t\, dt\,. \end{aligned} \tag{3.64}$$

Из (3.64) видно, что X_a отличается от X_b, так же как и Y_a от Y_b, лишь сдвигом по времени на величину T, так как $u(t)$ берется из разных интервалов времени. Следовательно, величины X_a и Y_a, как результат формирования X и Y за предыдущий интервал времени длительностью T , должны запоминаться схемой.

На рис. 3.15 изображена функциональная схема устройства оптимальной обработки, построенная по правилу (3.63).

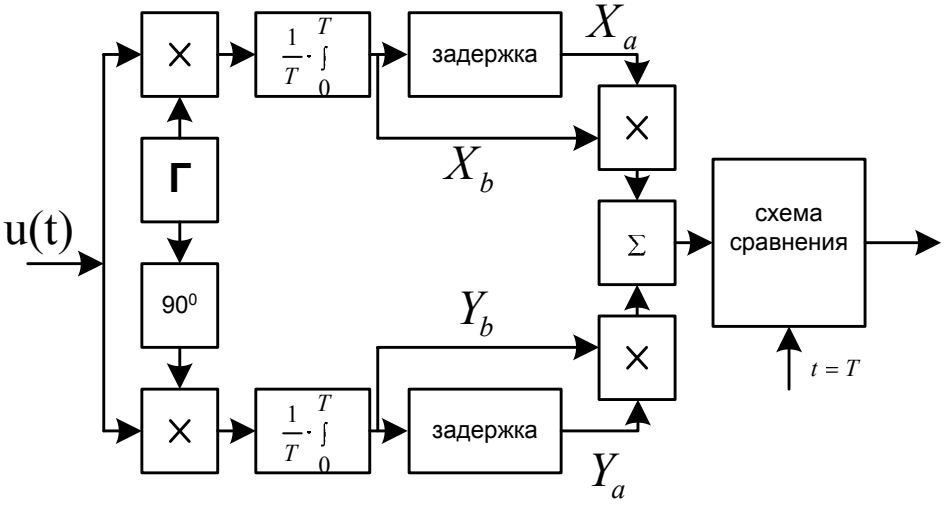


Рис. 3.15. Функциональная схема устройства оптимальной обработки

Именно такая схема реализована в аппаратуре передачи дискретной информации, обеспечивающей в стандартном телефонном канале 20 телеграфных каналов со скоростью работы в каждом до 120 Бод. Она содержит генератор

колебания косинусоидальной формы, фазовращатель на угол 90°, позволяющий получить синусоидальные колебания, сдвинутые друг относительно друга на 90°, два высокочастотных перемножителя, две линии задержки для формирования X_a и Y_a, сумматор и устройства определения знака результата суммирования. Отсчет производится через интервалы времени, равные *T.* Если результат суммирования больше нуля, то принимается решение, что передавалась реализация $s_1(t)$, в противном случае решение принимается в пользу $s_2(t)$.

Следует отметить, что в данной аппаратуре задержка сигналов осуществляется с помощью запоминающих устройств.

Правило (3.63) может быть реализовано с помощью согласованного фильтра. На рис.3.16 приведена функциональная схема такого устройства.

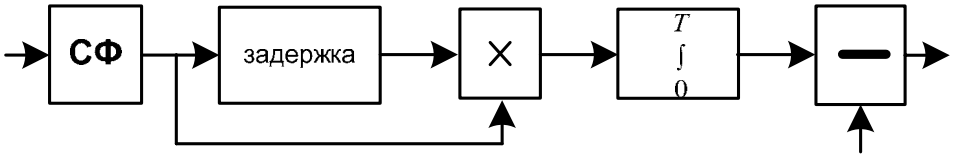


Рис.3.16 Схема реализации с помощью согласованного фильтра

Согласованный фильтр строится на реализацию сигнала длительностью *T* косинусоидальной формы. Текущий отклик фильтра и отклик с выхода линии задержки перемножаются и интегрируются на интервале [0, *T*].

3.4.2. Помехоустойчивость оптимального некогерентного приема

Для нахождения вероятности ошибки при оптимальном некогерентном приеме двоичных сигналов с ЧМн необходимо найти вероятность $A_1 > A_2$, при приходе сигнала $C_1(t)$ (или вероятность неравенства $A_1 < A_2$ при приходе сигнала $C_2(t)$).

Выражение для вероятности ошибки при оптимальном некогерентном приеме сигналов ЧМн в отсутствии замираний амплитуд $(U_m = const)$ [5, 21, 32]:

$$p_{ош} = \frac{1}{2} \cdot e^{-\frac{h^2}{2}}, \qquad (3.65)$$

где, $h^2 = E/N_0$ – отношение энергии элемента сигнала к спектральной плотно-

сти мощности шума.

Точное выражение для вероятности ошибки при оптимальном некогерентном приеме сигналов с АМн в явном виде получить не удается. Однако при больших уровнях полезного сигнала $h^2 >> 1$ хорошее приближение дает формула:

$$p_{\text{ош}} = \frac{1}{2} \cdot e^{-\frac{h^2}{4}}, \qquad (3.66)$$

Сравнение (3.65) и (3.66) показывает, что применение сигналов АМн приводит к проигрышу в мощности сигнала примерно в 2 раза по сравнению с сигналами ЧМн.

Вероятность ошибки при некогерентном приеме сигналов ОФМн:

$$p_{ош} = \frac{1}{2} \cdot e^{-h^2} \quad . \qquad (3.67)$$

Анализ соотношений (3.67), (3.65) и (3.66) показывает, что ОФМн имеет двукратный выигрыш в мощности сигнала по сравнению с ЧМн и четырехкратный выигрыш по сравнению с АМн. Необходимо также отметить, что применение некогерентного приема дает приводит к увеличению вероятности ошибки, эквивалентному полутора–двукратному уменьшению мощности сигнала при когерентном приеме.

На рис. 3.17 представлены кривые зависимости вероятности ошибок от отношения энергии сигнала к спектральной плотности мощности помех для сигналов: АМн, ЧМн и ОФМн.

Некогерентный прием обладает худшей помехоустойчивостью по сравнению с когерентной обработкой, так как он основан на знании лишь части параметров приходящих сигналов. Достоинством некогерентного приема является простота реализации, ухудшение помехоустойчивости при этом компенсируется соответствующим увеличением мощности.

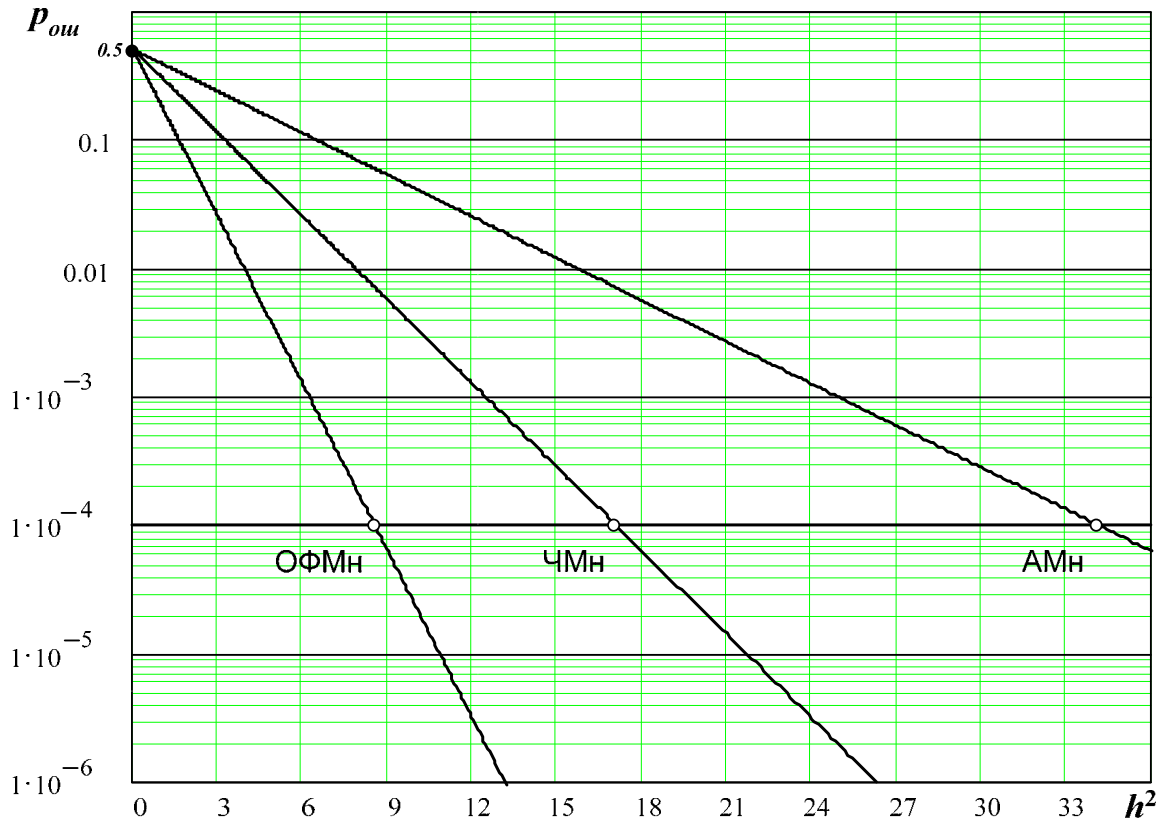


Рис. 3.17. Зависимости вероятностей ошибок для сигналов с АМн, ЧМн и ОФМн

3.5. Подоптимальные методы приема

3.5.1. Причины применения неоптимальных методов приема

Ранее были рассмотрены методы оптимального когерентного и оптимального некогерентного приема сигналов. Однако реализация оптимальных методов приема сигналов возможна лишь при абсолютно синхронной работе модулятора и демодулятора. Например, при оптимальном когерентном приеме требуется совпадение фаз приходящего сигнала и сигнала, формируемого в демодуляторе. В оптимальном когерентном демодуляторе на согласованных фильтрах момент отсчета T должен совпадать с моментом пикового значения выходного напряжения и с концом длительности элемента сигнала.

При оптимальном некогерентном приеме необходимы сведения о начале и конце передачи каждой посылки.

В реальных системах связи синхронизация передатчика и приемника может достигаться применением в модуляторе и демодуляторе высокостабильных

опорных генераторов или путем передачи сведений о фазе сигналов и моментах отсчетов $t_i, t_i + T$ по отдельному каналу связи.

Наряду с оптимальными методами обработки принимаемых сигналов в настоящее время находят широкое распространение квазиоптимальные (подоптимальные) правила приема. Это происходит по разным причинам. Во-первых, кроме статистических, существуют другие критерии, например, стоимость, габариты, вес и т. д., которые могут существенно повлиять на принцип построения всего комплекса связи в целом. Во-вторых, оптимальные методы обработки иногда бывает трудно реализовать. Например, трудно построить согласованный фильтр для достаточно сложных сигналов или учесть некоторые параметры, необходимые для разработки оптимальных методов. В-третьих, некоторые неоптимальные методы мало уступают оптимальным, но отличаются большей простотой, универсальностью и экономичностью.

Конечно, при любом неоптимальном построении приемного устройства вероятность ошибки повышается и для того, чтобы сохранить качество связи, приходится повышать мощность передатчика или уменьшать дальность связи.

Величина η, показывающая, во сколько раз необходимая мощность сигнала P_c в одной системе больше мощности сигнала в другой системе, называется энергетическим проигрышем первой системы относительно второй. Обычно энергетический проигрыш выражают в децибелах [6]:

$$\eta[\partial Б] = 10 \lg \frac{P_{c1}}{P_{c2}}. \qquad (3.68)$$

3.5.2. Квазиоптимальные методы приема

Методы обработки сигналов, у которых энергетический проигрыш (3.64) относительно оптимальных не превышает нескольких единиц децибел, принято называть квазиоптимальными (подоптимальными, субоптимальными). Как правило, квазиоптимальные методы обработки отличаются тем, что вместо оптимальных фильтров применяют более простые полосовые фильтры а операция интегрирования заменяется фильтрацией.

В современных системах связи наибольшее применение находят схемы квазиоптимального приема, использующие сигналы ОФМн. Среди методов приема ОФМн сигналов наибольшее распространение получили методы сравнения фаз, обеспечивающие некогерентный прием, и методы сравнения полярностей при когерентном приеме.

Структурная схема демодулятора ОФМн сигналов по методу сравнения фаз представлена на рис. 3.18. В фазовом детекторе (ФД) производится сравнение фаз принятого $U_i(t)$ и предыдущего $U_j(t-\tau_{задерж})$ сигналов. Формирование выходных сигналов $y(t)$ после ФД осуществляется так же, как и в схеме приема ФМн сигналов. Так как в этой схеме в качестве опорного напряжения для ФД используется принятый сигнал, то появление обратной работы принципиально исключается.

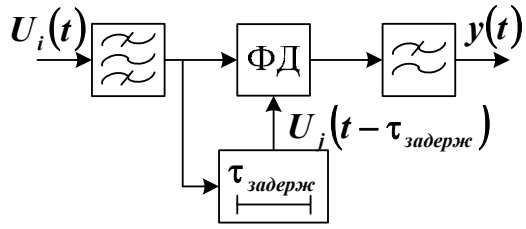


Рис. 3.18. Структурная схема приемника ОФМн сигналов по методу сравнения фаз

Демодулятор ОФМн сигналов по методу сравнения полярностей функционально состоит из двух частей: когерентного демодулятора ФМн сигналов и относительного декодера, или схемы сравнения полярностей (рис. 3.19). Принимаемый сигнал сначала обрабатывается когерентным демодулятором ФМн и, конечно, на его выходе наблюдается обратная работа. Однако относительный декодер (линия задержки (ЛЗ) и устройство сравнения (УС)) устраняет ее. Это происходит потому, что в УС сравниваются полярности настоящей и предыдущей посылок и вырабатывается выходной сигнал по следующему правилу: если полярности совпадают вырабатываются положительное напряжение; если полярности соседних посылок разные – отрицательное. Обратная работа изменяет полярность как настоящей, так и предыдущей посылок и поэтому не сказывается на полярности сигнала на выходе УС.

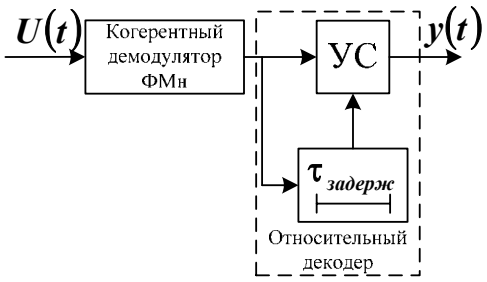


Рис. 3.19. Структурная схема приемника ОФМн сигналов по методу сравнения полярностей

Отношение мощности сигнала к мощности шума на выходе полосового

фильтра рассчитывается по формуле [6]:

$$\left(\frac{P_c}{P_ш}\right)_{вых} = P_c \frac{P_{с\,вх}}{N_0 \Delta F} = \frac{P_{с\,вх} T}{N_0 \Delta F T} = \frac{h^2}{\Delta F T},$$

где ΔF - полоса пропускания фильтра. Поскольку на выходе оптимального фильтра отношение равно h^2:

$$h^2 = \left(\frac{P_c}{P_ш}\right)_{ОПТ} = \frac{P_c \cdot T}{N_0}, \text{ то}$$

энергетический проигрыш квазиоптимального метода

$$\eta = \Delta F T .$$

Чем уже полоса пропускания фильтра ΔF, тем меньше энергетический проигрыш квазиоптимальной схемы. Однако чрезмерное уменьшение ширины полосы пропускания фильтра приводит к увеличению длительности переходных процессов в устройстве; действие предыдущей посылки начинает сказываться на последующей. Вероятность ошибки увеличивается. Оптимальная с точки зрения помехоустойчивости ширина полосы пропускания $\Delta F_{опт} \cong \frac{2}{T}$. Поэтому энергетический проигрыш в лучшем случае составляет около 2 дБ.

Учитывая энергетический проигрыш $\Delta F T$, который получается при замене согласованных фильтров полосовыми фильтрами, получим следующие оценки вероятностей ошибок в подоптимальных схемах [6]

$$p_{ош} = \frac{1}{2}\exp\left(-\frac{P_c \cdot T}{4 N_0 \Delta F T}\right) = \frac{1}{2}\exp\left(-\frac{P_c}{4 N_0 \Delta F}\right) \text{ для АМн;}$$

$$p_{ош} = \frac{1}{2}\exp\left(-\frac{P_c}{2 N_0 \Delta F}\right) \text{ для ЧМн;} \qquad p_{ош} = \frac{1}{2}\exp\left(-\frac{P_c}{N_0 \Delta F}\right) \text{ для ОФМн.}$$

Сравнивая помехоустойчивость оптимального и квазиоптимального приема дискретных сигналов видим, что переход к квазиоптимальному приему сигналом ЧМн приводит к увеличению вероятности ошибки вследствие уменьшения h^2 в $\Delta F T$ раз.

Контрольные вопросы

1. Как ставится задача обнаружения сигнала на фоне помех?

2. В чем сущность критерия среднего риска?

3. В чем состоит сущность критерия максимального правдоподобия?

4. Нарисуйте и поясните по схеме работу оптимального корреляционного приемника.

5. Нарисуйте и поясните по схеме работу оптимального приемника на согласованных фильтрах.

6. Определите значение отношения сигнал/шум обеспечивающего вероятность ошибки $p^*_{ош} = 10^{-5}$, для рассмотренных методов манипуляции.

7. Сравните потенциальную помехоустойчивость систем с АМн, ЧМн, ФМн и ОФМн.

8. В чем состоит сущность квазиоптимальных алгоритмов приема?

ГЛАВА 4. ТЕОРИЯ ПЕРЕДАЧИ ИНФОРМАЦИИ

4.1. Информационные характеристики источника сообщений

Из определения информации как совокупности неизвестных для получателя сведений следует, что в общем случае дать оценку количества информации довольно затруднительно, так как всякое подлежащее передаче сообщение имеет свое содержание, свой смысл и определенную ценность для получателя. Одно и то же сообщение может давать одному получателю много информации, другому мало. Однако содержательная сторона несущественна для теории и техники связи и поэтому не учитывается при определении количественной меры информации.

4.1.1. Количественное определение информации

В основу измерения количества информации положены вероятностные характеристики передаваемых сообщений, которые не связаны с конкретным содержанием сообщений, а отражают степень их неопределенности. Естественно, что чем меньше вероятность сообщения, тем больше информации оно несет.

Количество информации $I(x_i)$ в отдельно взятом единичном сообщении x_i определяется величиной, обратной вероятности появления сообщения $p(x_i)$ и вычисляется в логарифмических единицах [2]:

$$I(x_i) = \log_b\left(\frac{1}{p(x_i)}\right) = -\log_b\left(p(x_i)\right). \qquad (4.1)$$

Логарифмическая мера, впервые предложенная в 1928 г. английским ученым Р. Хартли, обладает свойством аддитивности, что соответствует нашим интуитивным представлениям об информации. Кроме того, при $p(x_i)=1$ количество информации, вычисленное по (4.1), равно нулю, что соответствует принятому определению информации.

Если источник выдает зависимые сообщения $x_i = x_1,...,x_m$, то они характеризуются условными вероятностями $p\left(\frac{x_i}{x_1,...,x_m}\right)$. И в этом случае количество информации вычисляется по формуле (4.1) с подстановкой в нее условных вероятностей сообщений.

Единицы измерения количества информации.

Выбор основания логарифмов в формуле (4.1) определяет единицы измерения количества информации. При использовании десятичного логарифма $(b=10)$ информация измеряется в десятичных единицах – дитах. В случае использования натуральных логарифмов единицей измерения является натуральная единица – нат.

Более удобно в системах, работающих с двоичными кодами, использовать основание логарифма $b=2$, и тогда информация измеряется в двоичных единицах – дв.ед. Весьма часто вместо двоичных единиц используется эквивалентное название – бит, возникшее как сокращенная запись английских слов binary digit (двоичная цифра). 1 бит это количество информации, которое передается единичным символом сообщения, вероятность передачи которого $p(x_i)=0{,}5$:

$$I(x_i)=-\log_2(0{,}5)=-\log_2\left(\frac{1}{2}\right)=-\log_2\left(2^{-1}\right)=1\ (\textit{бит}).$$

В настоящее время термин бит в информатике, вычислительной и импульсной технике употребляется не только как единица количества информации, но и для обозначения числа двоичных символов 0 и 1, поскольку они обычно равновероятны и каждый из них несет 1 бит информации.

Количество информации в сообщении, составленном из ***n*** символов, определяется по формуле [2]:

$$I(x,n)=-n\cdot\sum_{i=1}^{m}p(x_i)\cdot\log_2 p(x_i),$$

где $i=\{1,...,m\}$ – номер символа x_i из алфавита источника;

$p(x_i)$ – вероятность передачи i-го символа.

4.1.2. Энтропия и производительность дискретного источника сообщений

Энтропия источника сообщений

Для большинства реальных источников сообщения имеют разные вероятности. Например, в тексте буквы А, О, Е встречаются сравнительно часто, а Щ, Ы – редко. Согласно экспериментальным данным, для букв русского алфавита

характерны безусловные вероятности, сведенные в табл. 4.1.

Таблица 4.1

Безусловные вероятности букв русского алфавита

буква	вероятность	буква	вероятность	буква	вероятность
пробел	0,175	М	0,026	Ч	0,012
О	0,090	Д	0,025	Й	0,010
Е	0,072	П	0,023	Х	0,009
А	0,062	У	0,021	Ж	0,007
И	0,062	Я	0,018	Ю	0,006
Т	0,053	Ы	0,016	Ш	0,006
Н	0,053	З	0,016	Ц	0,004
С	0,045	Ь,Ъ	0,014	Щ	0,003
Р	0,040	Б	0,014	Э	0,003
В	0,038	Г	0,013	Ф	0,002
Л	0,035	К	0,028		

При разных вероятностях сообщения несут различное количество информации $I(x_i)$. При решении большинства практических задач необходимо знать среднее количество информации, приходящееся на один элемент сообщения. Это среднее количество информации при общем числе элементов сообщения источника *n* и числе символов алфавита *m* равно:

$$H(X) = \frac{I(x,n)}{n} = -\sum_{i=1}^{m} p(x_i) \cdot \log_2 p(x_i) \text{ (бит/сообщение).} \qquad (4.2)$$

Величину $H(X)$ называют энтропией источника сообщений. Термин «энтропия» заимствован из термодинамики, где она характеризует среднюю неопределенность состояния системы молекул вещества. В теории информации этот термин введен в 1948 г. американским ученым К. Шенноном [49] и далее более строго определен советскими математиками А.Я. Хинчиным [46, 47, 48]и А.Н. Колмогоровым [27]. Физически энтропия $H(X)$ выражает среднюю неопределенность состояния источника сообщений и является объективной информационной характеристикой источника. Энтропия всегда положительна и принимает максимальное значение при равновероятных сообщениях [2]:

$$H_{\max}(X) = -\sum_{i=1}^{m} \frac{1}{m} \cdot \log_2\left(\frac{1}{m}\right) = \log_2(m). \qquad (4.3)$$

Минимальное значение энтропии $H_{\min}(X) = 0$ соответствует случаю, когда одна из вероятностей $p(x_i) = 1$, а остальные равны нулю, т.е. имеется полная оп-

ределенность.

Для источника с зависимыми сообщениями энтропия тоже вычисляется как математическое ожидание количества информации на один элемент этих сообщений. Следует заметить, что полученное в этом случае значение энтропии будет меньше, чем для источника независимых сообщений. Это следует из того, что при наличии зависимости сообщений неопределенность выбора уменьшается и, соответственно, уменьшается энтропия. Так, в тексте после сочетания "чт" вероятнее всего, что третьей буквой будет "о" и маловероятно появление в качестве третьей буквы "ж" или "ь". В среднем, сочетание "что" несет меньше информации, чем эти буквы в отдельности.

Наиболее широкое применение в дискретных системах передачи информации получили двоичные источники. Двоичные источники характеризуются передачей только двух возможных сообщений. Причем, если вероятность передачи одного из них $p(x_1)$, то вероятность передачи другого $p(x_2) = 1 - p(x_1)$.

Определим энтропию двоичного источника. Из формулы (4.2) получим:

$$H(X) = -\sum_{i=1}^{2} p(x_i) \cdot \log_2 p(x_i) = -p(x_1) \cdot \log_2 p(x_1) - p(x_2) \cdot \log_2 p(x_2) = \\ = -p(x_1) \cdot \log_2 p(x_1) - [1 - p(x_1)] \cdot \log_2 [1 - p(x_1)]. \qquad (4.4)$$

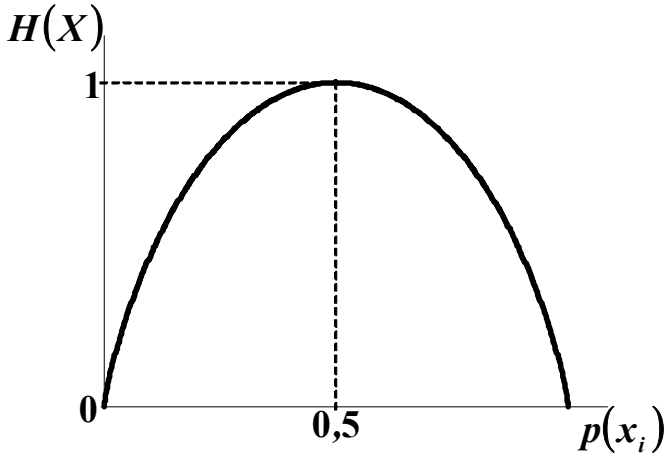


Рис. 4.1. Зависимость энтропии от вероятности символов

График зависимости (4.4) представлен на рис. 4.1. Как следует из графика, энтропия двоичного источника изменяется в пределах от нуля до единицы. Энтропия равна нулю, когда вероятность передачи одного из символов равна нулю или единице, т.е. передается только одно сообщение. Получение же одного единственно возможного сообщения никакой новой информации не дает. Энтропия двоичного источника будет максимальна, если существует наибольшая неопределенность, т.е. $p(x_2) = p(x_1) = 0{,}5$. При этом $H(X) = \log_2 m = 1$.

Избыточность источника сообщений

Избыточными в источнике являются сообщения, которые несут малое, иногда нулевое, количество информации. Наличие избыточности означает, что часть сообщений можно и не передавать по каналу связи, а восстановить на приеме по известным статистическим связям. Так и поступают при передаче телеграмм, исключая из текста союзы, предлоги, знаки препинания, поскольку они легко восстанавливаются по смыслу телеграммы на основании известных правил построения фраз.

Количественно избыточность оценивается коэффициентом избыточности:

$$\chi = \frac{H_{\max}(X) - H(X)}{H_{\max}(X)} = 1 - \frac{H(X)}{H_{\max}(X)}, \qquad (4.5)$$

где $H(X)$ – энтропия источника; $H_{\max}(X) = \log_2 m$ – максимальная энтропия источника с алфавитом из m сообщений.

Избыточность при передаче сообщений имеет свои положительные и отрицательные стороны. Увеличение избыточности приводит к увеличению времени передачи сообщений, излишней загрузке каналов связи. За определенный промежуток времени по каналу передается меньшее количество информации, чем это возможно; поэтому одной из задач теории информации и техники кодирования является задача сокращения избыточности.

Однако при увеличении избыточности появляется возможность повышения помехоустойчивости передачи сообщений. Так, избыточность текста позволяет исправлять отдельные ошибки или восстанавливать пропущенные буквы или даже слова в телеграмме. У русского и всех европейских языков избыточность с учетом всех статистических зависимостей букв примерно одинакова $\chi = 0{,}5$. Она сформировалась в результате длительной, общественной практики на основе требований исправления искажения слов и фраз под воздействием различных мешающих факторов. Для систем связи устанавливается компромиссное значение избыточности, которое обеспечивает заданную скорость и надежность передачи сообщений.

Производительность источника сообщений

Для источников сообщений с фиксированной скоростью важным параметром является его производительность $H'(X)$, определяемая выражением:

$$\mathrm{H}'(\mathrm{X}) = \frac{1}{\mathrm{T}} \cdot \mathrm{H}(\mathrm{X}) \ [бит/с],$$

где T – интервал времени для передачи элементарного сообщения.

Физический смысл производительности – количество информации, выдаваемое источником в среднем за единицу времени (одну секунду) его непрерывной работы.

4.2. Пропускная способность дискретного канала

4.2.1. Количество информации переданной по дискретному каналу

Основной задачей систем связи является передача информации от источника к получателю. Решение этой задачи сопряжено с определенными трудностями, связанными не только с представлением информации в виде сообщения, пригодного для восприятия и обработки, но и с преобразованием данного сообщения в сигнал, пригодный для передачи по линии связи. Представим графически процесс передачи информации по каналу связи (рис.4.2) и учтем при этом возможное влияние помех.

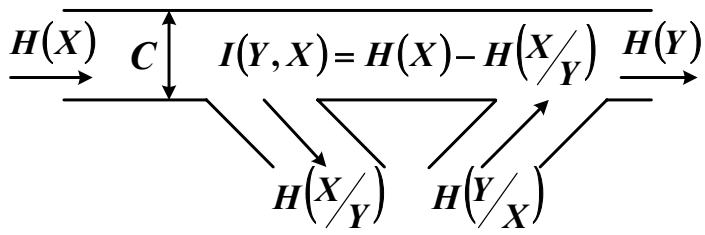


Рис. 4.2. Иллюстрация передачи информации по каналу связи

Пусть дискретный канал определяется:

$X = \{x_i\}$ – алфавитом источника сообщений;

$Y = \{y_j\}$ – алфавитом получателя сообщений;

$H(X/Y)$ – «потерями» информации;

$H(Y/X)$ – ложной информацией, создаваемой помехами;

$I(Y,X)$ – количеством информации, переданной по каналу.

Условная энтропия характеризует среднюю степень неопределенности принимаемых сигналов, обусловленную действием помех.

При сопряжении входа канала с любым источником двоичной информации на вход могут поступать двоичные символы x_1 и x_2 с вероятностями $p(x_1)$ и $(1-p(x_1))$ соответственно (рис.4.3). На выходе канала появляются двоичные символы y_1 и y_2. Обозначим вероятность ошибки при передаче любого символа через $p_{ош}$. Тогда, $p\left(y_1/x_1\right)=p\left(y_2/x_2\right)=1-p_{ош}$, а $p\left(y_1/x_2\right)=p\left(y_2/x_1\right)=p_{ош}$. В общем случае, для m-ичного дискретного канала [5]:

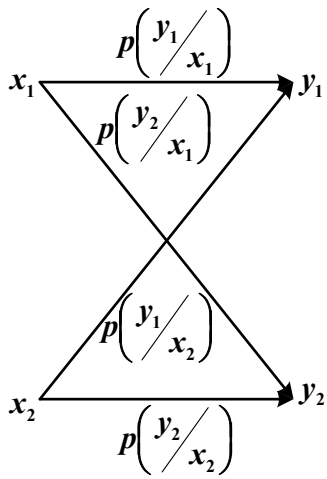


Рис. 4.3. Вероятностные характеристики передачи информации

$$p\left(y_j/x_i\right)=\begin{cases}\dfrac{p_{ош}}{m-1} & \text{при } i\neq j,\\ 1-p_{ош} & \text{при } i=j.\end{cases}.$$

Тогда «шумовая» энтропия будет определяться выражением [5]:

$$H\left(Y/X\right)=\sum_{i=1}^{m}p(x_i)\sum_{j=1}^{m}p\left(y_j/x_i\right)\cdot\log_2\frac{1}{p\left(y_j/x_i\right)}=$$
$$=-(1-p_{ош})\cdot\log_2(1-p_{ош})-p_{ош}\cdot\log_2\frac{p_{ош}}{m-1}. \qquad (4.6)$$

Из полученного выражения следует, что энтропия определяемая только помехой не зависит от вероятности $p(x_i)$ появления символов на входе канала.

Если символы на входе канала выбираются независимо от предыдущих символов с одинаковыми вероятностями, то энтропия выходных символов достигает своего максимального значения, равного: $H(Y)=\log_2 m$.

Таким образом, количество информации, переданной по каналу, это разность между энтропией на выходе и энтропией шума:

$$I(Y,X)=H(Y)-H\left(Y/X\right)=\log_2 m+(1-p_{ош})\cdot\log_2(1-p_{ош})+p_{ош}\cdot\log_2\frac{p_{ош}}{m-1}. \qquad (4.7)$$

Количество информации, переданной по каналу связи, обладает следующими основными свойствами:

$I(Y,X) \geq 0$, причем $I(Y,X) = 0$ тогда и только тогда, когда входные и выходные сообщения в канале взаимно независимы;

$I(Y,X) \leq H(X)$, причем $I(Y,X) = H(X)$ тогда и только тогда, когда входная последовательность определяется однозначно по выходной последовательности, например, когда в канале нет помех;

$I(Y,X) = I(X,Y) = H(Y) - H(Y/X)$ следует из того, что количество информации не изменится, если входную и выходную последовательность поменять местами.

4.2.2. Пропускная способность дискретного канала

Пропускной способностью канала, рассчитанной на один входной символ, называется максимальное количество информации, которое может быть передано по каналу, причем максимум ищется по всем возможным источникам X, имеющим различные (произвольные) вероятностные характеристики:

$$C' = \max_{X} I(Y,X) \ [\textit{бит/символ}].$$

Часто более удобно пользоваться пропускной способностью канала, рассчитанной не на один входной символ, а на единицу времени:

$$C = \frac{1}{T} \cdot C' \ [\textit{бит/с}].$$

Величину C называют пропускной способностью канала в единицу времени или просто пропускной способностью.

Пропускная способность канала обладает следующими основными свойствами:

$C \geq 0$, $C = 0$ тогда и только тогда, когда вход и выход канала статистически независимы;

$C \leq \frac{1}{T} \cdot \log_2 m$ для канала без помех.

Из определения пропускной способности следует, что характеристика описывает свойства канала, по которому передается информация от определенного источника. Очевидно, никакой источник не способен передать по каналу

количество информации большее пропускной способности и данная характеристика описывает потенциальные возможности канала по передаче информации.

4.2.3. Пропускная способность симметричного дискретного канала без памяти

Пропускная способность дискретного канала, по которому передается m дискретных сигналов с учетом (4.7) вычисляется по формуле [6]:

$$C = V_И \cdot \left[\log_2 m + (1 - p_{ош}) \cdot \log_2 (1 - p_{ош}) + p_{ош} \cdot \log_2 \frac{p_{ош}}{m-1} \right], \qquad (4.8)$$

где $V_И = \frac{1}{T}$ – скорость модуляции, бод; T – длительность сигнала; $p_{ош}$ – вероятность ошибки в канале. Заметим, что пропускная способность дискретного канала без помех при ($p_{ош} = 0$):

$$C_{ДК} = V_И \cdot [\log_2 m].$$

В частности пропускная способность двоичного канала ($m = 2$):

$$C_{ДК} = V_И \cdot [1 + (1 - p_{ош}) \cdot \log_2 (1 - p_{ош}) + p_{ош} \cdot \log_2 p_{ош}]. \qquad (4.9)$$

Зависимость отношения $C/V_И$ от вероятности ошибки $p_{ош}$, рассчитанная по формуле (4.9), показана на рис. 4.4.

Как следует из графика, при $p_{ош} = 0{,}5$ пропускная способность двоичного канала равна нулю ($C = 0$). Этот случай называют обрывом канала. Действительно вероятность ошибки $p_{ош} = 0{,}5$ можно получить и без передачи информации по каналу связи. А при $p_{ош} = 1$ пропускная способность такая же, как и при $p_{ош} = 0$ (канал без помех). Это объясняется тем, что при $p_{ош} = 1$ достаточно заменить нули на единицы и единицы на нули, чтобы абсолютно правильно восстановить переданный сигнал.

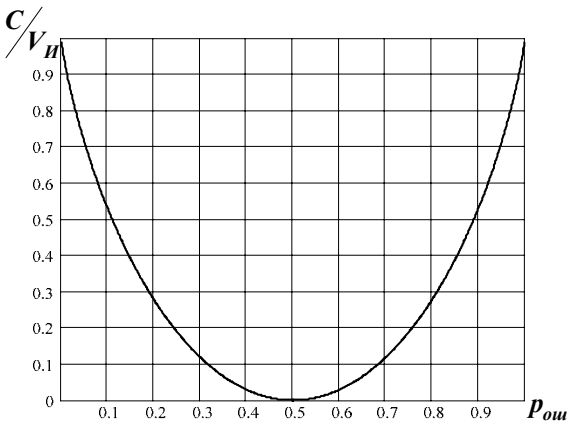


Рис. 4.4. Пропускная способность дискретного канала

Определим пропускную способность двоичного телеграфного канала, если скорость передачи в нем 1000 бит/с и вероятность ошибки 10^{-3} и сделаем вывод о том насколько отличается пропускная способность этого канала от идеального. Согласно формуле (4.9), при заданных параметрах

$$C_{дк} = 1000 \cdot [1 + 0{,}001 \cdot \log_2 0{,}001 + (1 - 0{,}001) \cdot \log_2 (1 - 0{,}001)] = 989 \ [бит/с].$$

Для идеального канала при $p_{ош} = 0$ получаем $C_{ДК} = V_И = 1000$ бит/с. Сравнение этих величин показывает, что ошибки в канале привели к уменьшению пропускной способности на 11 бит/с (т.е. потери составили 1,1%).

4.3. Методы сжатия дискретных сообщений

4.3.1. Условия существования оптимального неравномерного кода

При передаче сообщения осуществляется его преобразование в сигнал, пригодный для передачи по каналу связи. При этом необходимо согласовывать источник с каналом путем определения правила, по которому каждому элементу сообщения ставится в соответствие некоторый код, преобразуемый далее в сигнал.

В настоящее время существует два основных направления развития теории кодирования. В одном из них рассматриваются задачи повышения достоверности передачи в каналах с помехами, решаемые применением помехоустойчивых кодов, которые позволяют обнаруживать или исправлять ошибки. Такое кодирование называется помехоустойчивым. При этом избыточность кодовой последовательности выше, чем избыточность источника сообщений. Благодаря этому и оказывается возможным обнаружение и исправление ошибок передачи.

Другое направление теории кодирования связано с вопросами устранения избыточности при передаче сообщений в каналах без помех. Цель кодирования при этом состоит в таком преобразовании сообщения, при котором избыточность кодовой последовательности должна стать меньше, чем избыточность сообщений источника. В результате появляется возможность увеличения скоро-

сти передачи информации или снижаются требования к пропускной способности канала. Это особенно важно в случае, когда источники сообщений имеют большую избыточность, например, источники речевых сообщений, изображений и т.д.

Процесс кодирования с целью уменьшения избыточности источника сообщений носит название согласования источника с каналом или сжатия источника (экономного кодирования, энтропийного кодирования).

Избыточность (4.5) равна нулю только в том случае, когда элементы сообщения появляются на выходе источника с равными вероятностями $p(x_i)=\frac{1}{m}$ ($i=1,2,3,...,m$) и независимо друг от друга $p(x_i,x_j)=p(x_i)\cdot p(x_j)$. Если же $H(X)<\log_2 m$, то оказывается возможным построение кодов, имеющих меньшую избыточность, чем источник сообщений.

Покажем это на простейшем примере.

Пусть источник имеет алфавит из четырех символов *А*, *Б*, *В*, *Г* с вероятностями $p(A)=0{,}5$; $p(Б)=0{,}25$; $p(B)=p(Г)=0{,}125$.

Энтропия такого источника:

$$H(X)=-\sum_{i=1}^{4}p(x_i)\cdot\log_2 p(x_i)=-0{,}5\cdot\log_2 0{,}5-0{,}25\cdot\log_2 0{,}25-$$
$$-0{,}125\cdot\log_2 0{,}125-0{,}125\cdot\log_2 0{,}125=1{,}75.$$

Для передачи по каналу будем использовать равномерное кодирование, например, $A\to 00$, $Б\to 01$, $B\to 10$, $Г\to 11$. Тогда среднее число двоичных символов в сообщении, приходящихся на один символ источника, равно двум. Поскольку это на 12,5% больше энтропии источника, то используемый код не является оптимальным.

Рассмотрим теперь неравномерный код: $A\to 0$; $Б\to 10$; $B\to 110$; $Г\to 111$. В этом случае среднее число двоичных символов, приходящихся на один символ источника в сообщении,

$$n_{ср}=\sum_{i=1}^{4}p(x_i)\cdot n_i=0{,}5\cdot 1+0{,}25\cdot 2+0{,}125\cdot 3+0{,}125\cdot 3=1{,}75.$$

Таким образом, среднее число двоичных символов, приходящихся на

один символ источника, равно энтропии источника, т.е. для указанного источника неравномерный код оказывается более экономичным, чем равномерный.

Важно отметить, что при кодировании неравномерным кодом должна обеспечиваться возможность однозначного декодирования символов сообщения. Например, для рассмотренного источника, нецелесообразно применять код: $А \to 0$; $Б \to 1$; $В \to 10$; $Г \to 11$, поскольку прием последовательности 10 может означать передачу символа $В$, или двух символов $Б$ и $А$. Неоднозначно также декодирование символов 11. Для однозначного декодирования неравномерные коды должны удовлетворять условию префиксности: никакое более короткое слово не должно являться началом более длинного слова. Неравномерные коды, удовлетворяющие этому условию, называют префиксными.

Неравномерные коды позволяют в среднем уменьшить число двоичных символов на единичное информационное сообщение. Однако им присущ существенный недостаток: при возникновении ошибки она распространяется на все последующие элементы сообщения. Возникает ошибка синхронизации, приводящая к резкому ухудшению достоверности приема. Этот недостаток отсутствует в равномерных кодах. При кодировании равномерными кодами используется одно и то же число двоичных символов – блок; поэтому такие коды называют блоковыми.

4.3.2. Прямая и обратная теоремы кодирования источника неравномерными кодами

Прямая теорема кодирования состоит в том, что для любого однозначно декодируемого кода среднее число символов в двоичном кодовом слове всегда не меньше энтропии источника сообщений $n_{ср} \geq H(X)$, и существует однозначно декодируемый код, для которого выполняется неравенство $n_{ср} < H(X)+1$.

Обратная теорема кодирования утверждает, что невозможно построить однозначно декодируемый код, для которого выполнялось бы неравенство $n_{ср} < H(X)$.

Из этих теорем следует, что невозможно закодировать сообщение таким образом, чтобы средняя длина кодовых слов была меньше энтропии сообщения. Кроме того, существует кодирование, при котором средняя длина кодового слова незначительно отличается от энтропии источника сообщений. Среднее число символов кода на сообщение можно уменьшить, если кодировать не каждый символ сообщения, а блоки по n символов из алфавита X. Используя кодирование блоков, можно получить среднее число символов на сообщение, сколь угодно мало отличающееся от энтропии источника, но при этом возрастает сложность кодирования.

4.3.3. Показатели эффективности сжатия

Наряду с коэффициентом избыточности (4.5), часто используется коэффициент сжатия источника:

$$K_{сж.и} = \frac{H_{\max}(X)}{H(X)} = \frac{1}{1-\chi}.$$

Коэффициент сжатия источника показывает, во сколько раз можно уменьшить количество двоичных символов для представления единичного символа источника с энтропией $H(X)$ по сравнению со случаем, когда все сообщения источника передаются равновероятно.

Например, для источника, рассмотренного в п. 4.3.1, коэффициент сжатия

$$K_{сж.и} = \frac{\log_2 m}{H(X)} = \frac{\log_2 4}{1{,}75} = \frac{2}{1{,}75} = 1{,}14,$$

т.е. скорость передачи информации по каналу связи при использовании экономичного кодирования может быть в 1,14 раза больше, чем при равномерном кодировании.

4.3.4. Кодирование источника дискретных сообщений методом Шеннона-Фано

Кодирование методом Шеннона – Фано рассмотрим на примере. Пусть алфавит источника содержит шесть элементов $\{А, Б, В, Г, Д, Е\}$, появляющихся

с вероятностями $p(A)=0{,}15$, $p(Б)=0{,}25$, $p(B)=0{,}1$, $p(Г)=0{,}13$, $p(Д)=0{,}25$, $p(E)=0{,}12$. Энтропия такого источника:

$$H(X)=-0{,}15\cdot\log_2 0{,}15-0{,}25\cdot\log_2 0{,}25-0{,}1\cdot\log_2 0{,}1-$$
$$-0{,}13\cdot\log_2 0{,}13-0{,}25\cdot\log_2 0{,}25-0{,}12\cdot\log_2 0{,}12=2{,}492.$$

Алгоритм построения сжимающего кода Шеннона – Фано заключается в следующем.

1. Все m символов дискретного источника располагаются в порядке убывания вероятностей их появления (табл. 4.2).

Таблица 4.2

Построение кода Шеннона-Фано

Элемент сообщения	Вероятность элемента	Деление сообщения на группы и подгруппы	Код
Б	0,25	1 1	11
Д	0,25	1 0	10
А	0,15	0 1 1	011
Г	0,13	0 1 0	010
Е	0,12	0 0 1	001
В	0,1	0 0 0	000

2. Образованный столбец символов делится на две группы таким образом, чтобы суммарные вероятности каждой группы мало отличались друг от друга.

3. Верхняя группа кодируется символом «1», а нижняя – «0».

4. Каждая группа делится на две подгруппы с близкими суммарными вероятностями; верхняя подгруппа кодируется символом «1», а нижняя – «0».

5. Процесс деления и кодирования продолжается до тех пор, пока в каждой подгруппе не окажется по одному символу сообщения источника.

6. Записывается код для каждого символа источника; считывание кода осуществляется слева направо.

При использовании простейшего равномерного кода для кодирования шести элементов алфавита источника потребуется по три двоичных символа на каждую букву сообщения. Если же используется код Шеннона – Фано, то сред-

нее число символов на одну букву

$$n_{ср} = \sum_{i=1}^{6} p(x_i) \cdot n_i = 0{,}25 \cdot 2 + 0{,}25 \cdot 2 + 0{,}15 \cdot 3 + 0{,}13 \cdot 3 + 0{,}12 \cdot 3 + 0{,}1 \cdot 3 = 2{,}5 ,$$

что меньше, чем при простейшем равномерном коде и незначительно отличается от энтропии источника.

4.3.5. Кодирование источника дискретных сообщений методом Хаффмена

Рассмотрим еще один подход к кодированию, предложенный Хаффменом [6], на примере источника сообщений, заданного в табл. 4.3.

Таблица 4.3

Построение кода Хаффмена

Элемент сообщения	Вероятность элемента	Деление сообщения на группы и подгруппы	Код
Б	0,25	0	10
Д	0,25	1	01
А	0,15	1	111
Г	0,13	0	110
Е	0,12	1	001
В	0,1	0	000

(0,53) 1 (1) 0 (0,47) 1 (0,28) 1 0 (0,22) 1 0

Алгоритм построения сжимающего кода Хаффмена включает в себя следующие действия.

1. Все m символов дискретного источника располагаются в таблице в порядке убывания вероятностей.

2. Два символа, имеющих наименьшие вероятности, объединяются в один блок, а их вероятности суммируются.

3. Ветви скобки, идущей к большей вероятности, присваивается символ «1», а идущей к меньшей – символ «0».

4. Операции 2 и 3 повторяются до тех пор, пока не сформируется один

блок с вероятностью единица.

5. Записывается код для каждого символа источника; при этом считывание кода осуществляется справа налево.

Среднее число символов на одну букву для полученного кода

$$n_{ср} = \sum_{i=1}^{6} p(x_i) \cdot n_i = 0{,}25 \cdot 2 + 0{,}25 \cdot 2 + 0{,}15 \cdot 3 + 0{,}13 \cdot 3 + 0{,}12 \cdot 3 + 0{,}1 \cdot 3 = 2{,}5 .$$

Таким образом, для данного примера кодирование методами Хаффмена и Шеннона–Фано приводит к одинаковой эффективности. Однако опыт кодирования показывает, что код Хаффмена часто оказывается экономичнее кода Шеннона–Фано.

Рассмотренные методы построения сжимающих кодов широко известны и имеют практическое применение. Длина кодовой комбинации таких кодов зависит от вероятности выбора соответствующей буквы алфавита: наиболее вероятным буквам сопоставляются короткие кодовые комбинации, а менее вероятным – более длинные.

4.4. Пропускная способность непрерывного канала

4.4.1. Постановка задачи передачи дискретных сообщений в непрерывном канале

Вместо последовательностей символов для дискретного канала, в непрерывном канале осуществляется передача последовательности непрерывных величин с дискретным или непрерывным временем. В первом случае эти последовательности можно представить в виде импульсов различной величины, появляющихся в определенные моменты времени, а во втором случае как непрерывные функции времени.

Каналы с дискретным или непрерывным временем считаются заданными, если определено множество сигналов $x(t)$, которые можно подавать на вход, множество сигналов $y(t)$, на выходе, а также условные вероятностные характеристики появления сигнала $y(t)$ на выходе, если на вход был подан сигнал $x(t)$.

4.4.2. Количество информации, переданной по непрерывному каналу

Рассмотрим непрерывный источник с дискретным временем, в котором амплитуды импульсов статистически независимы друг от друга. Предположим, что в канале действует аддитивная помеха $n(t)$ с широким спектром, не зависящая от очередных и предыдущих импульсов. Тогда на выходе получим последовательность импульсов с амплитудами $y(t_1), y(t_2),..., y(t_k)$, статистически не зависящими друг от друга.

Свойства источника непрерывного сигнала будут определяться ПРВ $w(x_i)$, $i = 1,2,...,k$ входных (информационных) случайных величин $x(t_i)$, а воздействие помехи будет определяться условными ПРВ $w\left(y_i / x_i\right)$ выходных СВ $y(t_i)$ при заданных входных СВ $x(t_i)$ (рис. 4.5).

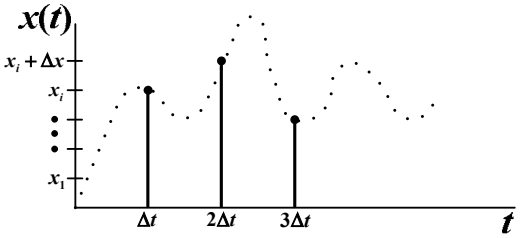


Рис. 4.5. Представление непрерывной функции дискретными отсчетами

Разделим области определения величин x и y на малые отрезки длиной Δx и Δy. Вероятность $p(x_i \le x < x_i + \Delta x)$ того, что значение x лежит на некотором отрезке $x_i \le x < x_i + \Delta x$, приблизительно равна $w(x_i)\Delta x$. Аналогично, $p(y_j \le y < y_j + \Delta y) \approx w(y_j)\Delta y$, а совместная вероятность этих двух событий будет $w(x_i, y_j)\Delta x \Delta y$. При такой дискретизации количество информации, переданное по каналу и рассчитанное на один импульс, приближенно находится по формуле:

$$I(Y,X)_{\Delta y}^{\Delta x} = \sum_i \sum_j w(x_i, y_j)\Delta x \Delta y \log \frac{w(x_i, y_j)\Delta x \Delta y}{w(x_i)\Delta x w(y_j)\Delta y}.$$

Устремив Δx и Δy к нулю, перейдем к непрерывному каналу. При этом двойная сумма преобразуется в двойной интеграл, а количество передаваемой информации

$$I(Y,X) = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} I(Y,X)_{\Delta y}^{\Delta x} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} w(x,y) \log \frac{w(x,y)}{w(x)w(y)} dx dy . \qquad (4.10)$$

Отметим следующие свойства количества информации, передаваемой в непрерывном канале:

$I(Y,X)\geq 0$, причем $I(Y,X)=0$ тогда и только тогда, когда вход и выход канала статистически независимы, т.е. $w\left(y/x\right)=w(y)$;

$I(Y,X)=I(X,Y)$ – свойство симметрии;

$I(Y,X)=\infty$, если помехи в канале отсутствуют, т.е. $y=x$, $n=0$.

Можно показать, что энтропия источника неограниченно возрастает, когда его алфавит переходит от дискретного к непрерывному. Для этого разделим область определения непрерывного сигнала $\boldsymbol{x}$ на отрезки Δx (рис. 4.5), и превратим сигнал в дискретный, положив вероятность появления $\boldsymbol{x_i}$, равной $w(x_i)\Delta x$. Энтропия такого дискретного сигнала

$$H_{\Delta x}(X)=\sum_i w(x_i)\Delta x\log\frac{1}{w(x_i)\Delta x}.$$

Устремим теперь Δx к нулю для перехода к энтропии непрерывного сигнала [6]:

$$H(X)=\lim_{\Delta x\to 0}H_{\Delta x}(X)=\lim_{\Delta x\to 0}\sum_i w(x_i)\Delta x\log\frac{1}{w(x_i)}+$$

$$+\lim_{\Delta x\to 0}\log\frac{1}{\Delta x}\sum_i w(x_i)\Delta x=\int_{-\infty}^{+\infty}w(x)\log\frac{1}{w(x)}dx+\lim_{\Delta x\to 0}\log\frac{1}{\Delta x}.$$

Первое слагаемое

$$\int_{-\infty}^{+\infty}w(x)\log\frac{1}{w(x)}dx=h(X),$$

представляет собой так называемую дифференциальную энтропию сигнала (или дифференциальную энтропию распределения $w(x)$). Второе слагаемое стремится к бесконечности совершенно независимо от природы и распределения вероятностей сигнала. Таким образом, при переходе от дискретных значений x к непрерывным энтропия сигнала неограниченно возрастает.

По аналогии с дискретным каналом количество информации, переданной по непрерывному каналу можно представить в следующей форме:

$$I(Y,X) = h(X) - h\left(X/Y\right) = h(Y) - h\left(Y/X\right), \tag{4.11}$$

где $h\left(X/Y\right)$ – условная дифференциальная энтропия сигнала x при известном сигнале y.

Второе равенство следует из второго свойства количества передаваемой информации (симметрии). Полученное выражение по форме напоминает (4.7), а дифференциальная энтропия играет здесь роль обычной энтропии дискретных сигналов. Однако свойства дифференциальной энтропии существенно отличаются от свойств обычной энтропии. Так, например, $h(X)$ и $h\left(X/Y\right)$ могут быть отрицательными.

Дифференциальная энтропия $h(X)$ уже не представляет собой среднее количество информации, выдаваемое источником сигнала (для непрерывного сигнала оно бесконечно). Аналогично $h\left(X/Y\right)$ не представляет собой количество информации, потерянной в канале, поскольку эта величина тоже бесконечна. Поэтому дифференциальную энтропию следует понимать лишь формально, как некоторую вспомогательную величину полезную при расчетах.

Если помеха аддитивная $y = x + n$, то нетрудно показать, что

$$h\left(Y/X\right) = \int_{-\infty}^{+\infty} w(n) \log \frac{1}{w(n)} dn = h(N), \tag{4.12}$$

где $w(n)$ – ПРВ помехи; $h(N)$ – дифференциальная энтропия помехи.

Подставляя (4.12) в (4.11), находим

$$I(Y,X) = h(Y) - h\left(Y/X\right) = \int_{-\infty}^{+\infty} w(y) \log \frac{1}{w(y)} dy - \int_{-\infty}^{+\infty} w(n) \log \frac{1}{w(n)} dn, \tag{4.13}$$

Найдем дифференциальную энтропию гауссовской помехи с нулевым средним и дисперсией σ^2 при отсутствии корреляции между значениями помехи. Согласно (4.12),

$$h(N) = \int_{-\infty}^{+\infty} w(n) \log \left[\sqrt{2\pi\sigma^2} \cdot \exp\left(\frac{n^2}{2\sigma^2} \right) \right] dn = \log\left(\sqrt{2\pi\sigma^2}\right) \int_{-\infty}^{+\infty} w(n) dn + \frac{\log e}{2\sigma^2} \int_{-\infty}^{+\infty} n^2 w(n) dn .$$

Учитывая $\int_{-\infty}^{+\infty} w(n) dn = 1$ и $\int_{-\infty}^{+\infty} n^2 w(n) dn = \sigma^2$, получим:

$$h(N) = \log\left(\sqrt{2\pi\sigma^2}\right) + \frac{1}{2}\log e = \log_2\left(\sqrt{2\pi e\sigma^2}\right). \qquad (4.14)$$

Дифференциальная энтропия принятого сигнала $y = x + n$ с гауссовским нормальным законом распределения вероятности [6]:

$$h(Y) = -\int_{-\infty}^{\infty} w(y)\log w(y)dy = \log_2\sqrt{2\pi e\sigma_y^2}\,. \qquad (4.15)$$

где $\sigma_y^2 = \sigma_c^2 + \sigma^2$; σ_c^2 – дисперсия сигнала. Подставляя (4.14) и (4.15) в (4.13) получим выражение для определения количества информации, переданной по непрерывному каналу:

$$I(Y,X) = h(X) - h\left(X/Y\right) = \frac{1}{2}\log_2\left(\frac{\sigma_c^2 + \sigma^2}{\sigma^2}\right). \qquad (4.16)$$

Полученное выражение показывает, что пропускная способность гауссовского канала с дискретным временем определяется отношением дисперсии сигнала к дисперсии помехи. Нередко величину $\frac{\sigma_c^2}{\sigma^2} = h^2$ называют отношением сигнал/шум. Чем больше это отношение, тем выше пропускная способность. Последнее вполне естественно, так как если дисперсия сигнала меньше дисперсии помехи или сравнима с ней, то по принятому сигналу трудно судить с определенностью, какое значение сигнала было подано на вход канала.

4.4.3. Пропускная способность непрерывного канала

Пусть сигнал $y(t)$ на выходе канала представляет собой сумму полезного сигнала $x(t)$ и шума $n(t)$, т.е. $y(t) = x(t) + n(t)$, причем $x(t)$ и $n(t)$ статистически независимы. Допустим, что канал имеет ограниченную полосу пропускания шириной ΔF_{HK}. Тогда в соответствии с теоремой Котельникова (см. п. 1.5) функции $y(t)$, $x(t)$ и $n(t)$ можно представить совокупностями отсчетов y_i, x_i, и n_i, $i = 1,2,...,L$, где $L = 2\Delta F_{HK}T$. При этом статистические свойства сигнала $x(t)$ можно описать многомерной ПРВ $w(x_1, x_2,..., x_L) = w(x)$, а свойства шума – ПРВ $w(n_1, n_2,..., n_L) = w(n)$.

Пропускная способность непрерывного канала определяется следующим

образом:

$$C = \lim_{T\to\infty} \frac{1}{T} \max_{w(x)} I(X,Y),$$

где $I(X,Y)$ – количество информации о какой-либо реализации сигнала $x(t)$ длительности T, которое в среднем содержит реализация сигнала $y(t)$ той же длительности T, а максимум ищется по всем возможным распределениям $w(x)$.

Когда сигнал на входе канала имеет нормальное распределение и отсчеты независимы величина $h(X)$ максимизируется [6]. Поэтому пропускная способность гауссовского канала с дискретным временем, рассчитанная на единицу времени, с учетом (4.16) может быть записана в виде

$$C = V_И \cdot I(Y,X) = \frac{V_И}{2} \log_2\left(\frac{\sigma_c^2 + \sigma^2}{\sigma^2}\right) = \frac{V_И}{2} \log_2\left(1 + h^2\right). \qquad (4.17)$$

Полученное выражение показывает, что пропускная способность гауссовского канала с дискретным временем определяется числом импульсов, передаваемых в секунду, и отношением сигнал/шум (h).

С учетом взаимосвязи скорости передачи информации и полосы частот непрерывного канала от (4.17) можно перейти к формуле Шеннона, которая устанавливает связь пропускной способности гауссовского канала с полосой пропускания непрерывного канала и отношением мощности сигнала к мощности помехи:

$$C = \Delta F_{НК} \log_2\left(1 + h^2\right). \qquad (4.18)$$

График отношения $\frac{C}{\Delta F_{НК}} = \log_2\left(1 + h^2\right)$ изображен на рис. 4.6. Заметим, что при малом отношении $h^2 << 1$

$$C \cong \Delta F_{НК} \cdot 1{,}442 \cdot h^2,$$

а пропускная способность канала связи прямо пропорциональна этому отношению.

При большом отношении $h^2 >> 1$ в (4.18) можно пренебречь единицей и считать, что

$$\frac{C}{\Delta F_{НК}} \approx \log_2\left(h^2\right),$$

т.е. зависимость пропускной способности непрерывного канала от отношения сигнал/шум логарифмическая.

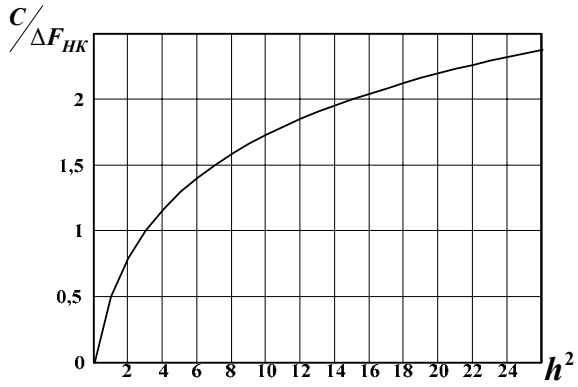


Рис. 4.6. Пропускная способность непрерывного канала

Пропускная способность канала, как предельное значение скорости безошибочной передачи информации, является одной из основных характеристик любого канала.

Определим пропускную способность стандартного канала тональной частоты, имеющего границы эффективно передаваемых частот 0,3...3,4 кГц, среднюю мощность сигнала на выходе 56 мкВт при средней мощности помехи 69000 пВт.

Согласно (4.18), при заданных параметрах

$$C_{НК} = 3{,}1\cdot 10^3 \cdot \log_2\left(\frac{56\cdot 10^{-6}}{69\cdot 10^{-12}}\right) = 3{,}0\cdot 10^4 \ \ [бит/с].$$

Для непрерывных каналов справедлива теорема Шеннона, согласно которой сообщения дискретного источника могут быть закодированы и переданы по непрерывному каналу так, что вероятность ошибочного декодирования принятого сигнала $y(t)$ будет меньше наперед заданной положительной величины $p^*_{ош}$, если производительность источника $H'(X)$ меньше пропускной способности C непрерывного канала.

Для типовых непрерывных каналов многоканальной связи основные технические характеристики и пропускная способность, вычисленная по формуле Шеннона (4.18), при отношении сигнал/шум 20 дБ, приведены в табл. 4.4.

Зная пропускную способность канала и информационные характеристики сообщений (табл. 4.5), можно определить, какие сообщения (первичные сигналы) можно передавать по заданному каналу.

Таблица 4.4

Характеристики типовых каналов многоканальной связи

Наименование канала	Границы передаваемых частот, Гц	Пропускная способность, бит/с
Тональной частоты	300...3400	$20{,}64 \cdot 10^3$
Предгрупповой широкополосный	$12{,}3 \cdot 10^3 ... 23{,}4 \cdot 10^3$	$73{,}91 \cdot 10^3$
Первичный широкополосный	$60{,}6 \cdot 10^3 ... 107{,}7 \cdot 10^3$	$313{,}6 \cdot 10^3$
Вторичный широкополосный	$312{,}3 \cdot 10^3 ... 551{,}4 \cdot 10^3$	$1{,}59 \cdot 10^6$
Третичный широкополосный	$812{,}3 \cdot 10^3 ... 2043{,}7 \cdot 10^3$	$8{,}2 \cdot 10^6$

Таблица 4.5

Производительность источников сообщений

Вид сообщения	Характер сообщения	Параметры АЦП		Производительность, бит/с
		f_d, Гц	$N = \log_2 L$	
Телеграфные, 50 Бод	дискретные	–	–	30…50
Телефонные	непрерывные	$8 \cdot 10^3$	8	$64 \cdot 10^3$
Звукового вещания: первого класса высшего класса	 непрерывные непрерывные	 $24 \cdot 10^3$ $32 \cdot 10^3$	 13	 $240 \cdot 10^3$ $416 \cdot 10^3$
Факсимильные, 120 строк/с: полутоновые штриховые	 непрерывные дискретные	 $2{,}93 \cdot 10^3$ –	 4 –	 $11{,}72 \cdot 10^3$ $2{,}93 \cdot 10^3$
Передача данных, 2400 Бод	дискретные	–	–	$2{,}4 \cdot 10^3$
Телевизионные	непрерывные	$13 \cdot 10^6$	16	$208 \cdot 10^6$

Например, первичный сигнал телевизионного вещания имеет $H'(X) = 208 \cdot 10^6 бит/с$ (табл. 4.5) и поэтому не может быть передан ни по одному из типовых непрерывных или цифровых каналов без потери качества. Следовательно, для передачи сигнала телевизионного вещания требуется создание специальных каналов с более высокой пропускной способностью или снижение скорости цифрового потока.

Контрольные вопросы

1. Какие свойства дискретного сообщения позволяют представить его с помощью сигнала имеющего меньшую избыточность?

2. Определите пропускную способность двоичного телеграфного канала,

если скорость передачи в нем 125 бит/с и вероятность ошибки 10^{-4}. Насколько отличается пропускная способность этого канала от идеального?

3.Аппаратура формирует шесть сообщений (символов) алфавита ***A*** со следующими вероятностями $p(a_1) = 0{,}22$; $p(a_2) = 0{,}31$; $p(a_3) = 0{,}18$; $p(a_4) = 0{,}09$; $p(a_5) = 0{,}05$; $p(a_6) = 0{,}15$. Постройте сжимающий код методом Шеннона-Фано (Хаффмена).

4. Почему дифференциальная энтропия источника непрерывных сообщений отличается от энтропии источника дискретных сообщений?

5. Определите пропускную способность стандартного канала тональной частоты, имеющего границы эффективно передаваемых частот 0,3…3,4кГц, среднюю мощность сигнала на выходе 32 мкВт при средней мощности помехи 87000 пВт.

ГЛАВА 5. ТЕОРИЯ КОДИРОВАНИЯ СООБЩЕНИЙ

5.1. Помехоустойчивое кодирование: блочные и непрерывные коды

Решение задачи выбора (отыскания) кода, оптимального по тому или иному критерию, составляет суть теории кодирования. Заметим, что современные методы кодирования не позволяют близко подойти к потенциальной пропускной способности канала связи при одновременно высокой верности передачи. Однако грамотный выбор кода позволяет, во многих случаях, значительно снизить вероятность ошибочного приема при скорости передачи порядка $10 \div 50\%$ пропускной способности канала.

В настоящее время повышение достоверности передачи в каналах с помехами, осуществляется с помощью кодов, позволяющих обнаруживать или исправлять ошибки. Такое кодирование называется помехоустойчивым. При этом избыточность кодовой последовательности выше, чем избыточность источника сообщений**.** Благодаря этому и оказывается возможным обнаружение и исправление ошибок приема.

5.1.1. Постановка задачи помехоустойчивого кодирования

Кодирование называется процесс преобразования сообщений в комбинации из дискретных сигналов. Основными задачами, решаемыми кодированием в процессе передачи сообщений, являются:

согласование источника сообщений с каналом по объемам алфавитов;

повышение скорости передачи информации по каналу за счет устранения избыточности в последовательности сообщений;

повышение помехоустойчивости передачи информации.

Первые две задачи решаются в кодере источника сообщений. Третья задача решается в кодере канала.

Для постановки задачи помехоустойчивого кодирования обратимся к структурной схеме канала электрической связи (рис. 5.1).

В общем случае в системе электрической связи можно передавать самые

различные по физической природе сообщения: цифровые данные, полученные от ЭВМ, речь, тексты телеграмм, команды управления, результаты измерений различных физических величин. Естественно, что все эти сообщения предварительно должны быть преобразованы в электрические колебания, сохраняющие все свойства исходных сообщений, а затем унифицированы, т.е. представлены в форме, удобной для последующей передачи.

Множество возможных дискретных сообщений источника $\boldsymbol{C}$, должно при этом обладать следующими свойствами:

оно должно быть конечным;

все сообщения равновероятны.

Тем самым обеспечивается максимальная энтропия источника (т.е. кодирование источника выполнено наилучшим образом).

Под источником информации на рис. 5.1 понимается устройство, в котором выполнены все названные ранее операции.

Рис. 5.1. Структурная схема системы электрической связи

Для более экономного использования линии связи, а также для уменьшения влияния различных помех и искажений передаваемая от источника информация может быть в дальнейшем преобразована с помощью кодирующего устройства. Это преобразование, как правило, состоит из ряда операций, включающих учет статистики поступающей информации для устранения избыточности (статистическое кодирование) реализуемое в кодере источника, а также введение дополнительных элементов для уменьшения влияния помех и искажений (помехоустойчивое кодирование) – кодер канала.

В результате ряда преобразований на выходе кодирующего устройства образуется последовательность элементов, которая с помощью модулятора преобразуется в форму, удобную для передачи по линии связи. Среда распространения – это среда, по которой происходит передача сигналов от передатчика (модулятора) к приемнику (демодулятору).

На вход демодулятора, кроме сигналов, прошедших среду, попадают также различные помехи. Демодулятор выделяет из смеси сигнала и помех последовательность, которая должна соответствовать последовательности на выходе кодирующего устройства. Однако из-за действия помех, влияния среды, погрешностей различных преобразований полное соответствие получить невозможно. Поэтому такая последовательность вводится в декодирующее устройство, которое выполняет операции по ее преобразованию в последовательность, соответствующую переданной. Полнота этого соответствия зависит от ряда факторов: корректирующих возможностей кодированной последовательности, уровня сигнала и помех, а также их статистики, свойств декодирующего устройства. Сформированная в результате декодирования последовательность поступает к получателю информации. Естественно, что при проектировании систем передачи информации всегда стремятся обеспечить такие условия работы, чтобы отличие информации, получаемой от источника, от информации, передаваемой получателю, было невелико и не превышало некоторой допустимой величины. В данном случае основным показателем качества передачи является достоверность передачи информации – степень соответствия принятого сообщения переданному [2, 21].

Основной принцип построения линейных кодов – отыскание процедур, которые позволяют получать все разрешенные кодовые комбинации C' путем конечного числа несложных линейных преобразований исходных разрешенных комбинаций C, а при декодировании для обнаружения и исправления ошибок получать информацию об ошибках по результатам конечного числа относительно простых преобразований над символами получаемых кодовых комбинаций C'' [23].

Основная задача оптимального построения корректирующего кода заключается в том, что из всех возможных кодовых комбинаций C'' применяется лишь некоторая часть. Используемые при передаче кодовые комбинации C' обычно называются разрешенными, а остальные – запрещенными. Следовательно, если под действием помехи передаваемая кодовая комбинация перехо-

дит в запрещенную, то такую ошибку можно обнаружить.

Чтобы получаемый код обладал наилучшей корректирующей способностью и минимальной вероятностью некорректируемых ошибок, необходимо выполнить следующие пять условий:

множество кодовых слов C'' должно быть конечным;

любому сообщению из множества C источника соответствует кодовое слово из множества C'';

множество C'', должно иметь большую размерность, чем C;

в качестве кодового слова соответствующего сообщению используется не все множество возможных кодовых слов C'', а лишь некоторая его часть C' – множество разрешенных кодовых слов;

в канал должны передаваться только разрешенные кодовые слова из множества C'.

Правило, по которому сообщению из множества C ставится в соответствие кодовое слово из множества C'', называется алгоритмом кодирования или кодом, корректирующим ошибки.

Рассмотрим следующий пример. Пусть множество C составляет 2 сообщения: 0 и 1, а множество C'' составляет 8 кодовых слов: 000, 001, 010, 100, 011, 101, 110, 111.

В качестве разрешенных (множество C') выбраны только 2 кодовые комбинации: 000 и 111.

При равновероятной передаче сообщений по каналам с независимыми ошибками, когда вероятность появления ошибок с увеличением кратности уменьшается, для минимизации средней вероятности ошибочного декодирования необходимо в первую очередь исправлять однократные ошибки как наиболее часто встречающиеся, затем двукратные и т.д. При этом декодер из множества кодовых комбинаций C'' выделяет кодовые комбинации $\widetilde{C}$, которые отличаются от C' в меньшем числе символов. Соответственно декодер принимает решение оптимальное по критерию максимума правдоподобия [5, 23].

Правило, по которому кодовому слову из множества C'' ставится в соот-

ветствие сообщение из множества $\widetilde{C}$, называется алгоритмом декодирования.

Избыточный код дает возможность обнаружить, в каких принятых символах имеются ошибочные символы. Кроме того, при разумном выборе кода вероятность не обнаруживаемой ошибки (т.е. ошибки, которая переводит разрешенную кодовую комбинацию в другую разрешенную кодовую комбинацию) может быть весьма малой.

5.1.2. Классификация кодов

Классификация рассматриваемых в данной главе методов кодирования приведена на рис. 5.2. Эта классификация не является исчерпывающей, в нее включены лишь некоторые методы, которые широко используются в современных системах связи.

Коды можно разделить на две самостоятельные группы. К первой относятся коды, использующие все возможные комбинации – неизбыточные коды. В литературе их еще называют простыми, или первичными. Ко второй группе относятся коды, использующие лишь определенную часть всех возможных комбинаций, такие коды называются избыточными. Оставшаяся часть комбинаций используется для обнаружения или исправления ошибок, возникающих при передаче сообщений. В этих кодах количество разрядов кодовых комбинаций можно условно разделить на определенное число разрядов, предназначенных для информации (информационные разряды), и число разрядов, предназначенных для коррекции ошибок (проверочные разряды).

Обе группы кодов, в свою очередь, подразделяются на равномерные и неравномерные. Равномерные коды – это коды, все кодовые комбинации которых содержат постоянное количество разрядов. Неравномерные коды содержат кодовые комбинации с различным числом разрядов. Ввиду того что неравномерные избыточные коды не нашли применения на практике из-за сложности их технической реализации, в дальнейшем их рассматривать не будем.

Все корректирующие (избыточные) коды делятся на два больших класса: блочные и непрерывные коды (рис. 5.2).

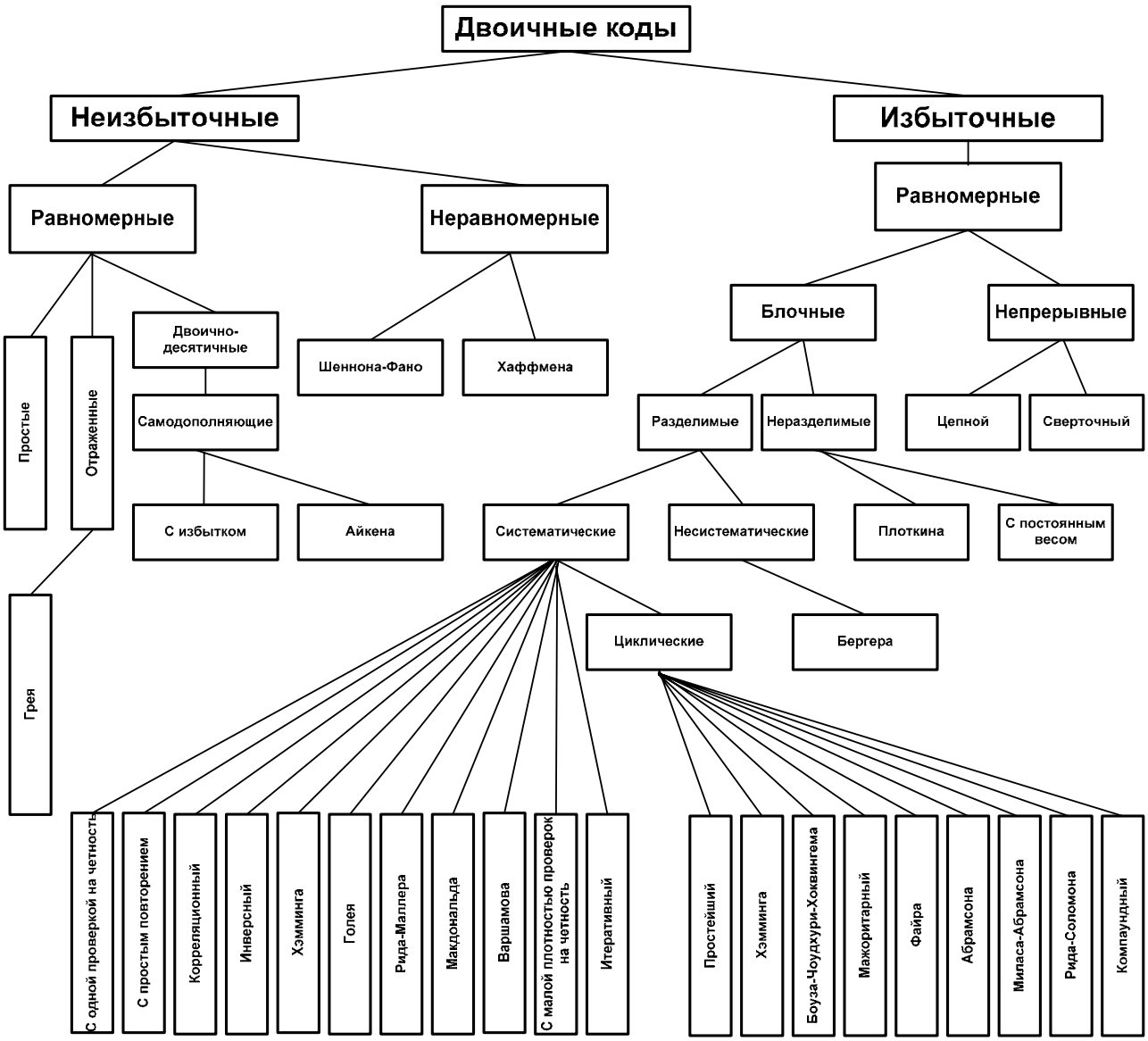


Рис. 5.2. Структурная классификация кодов

При кодировании блочным кодом последовательность k элементов данных от источника сообщений принимается за блок (сообщение). Каждому возможному блоку из k информационных символов ставится в соответствие кодовый блок (слово) длиной n. Код называется (n,k) – кодом, $n \geq k$. Кодовый блок в канале связи искажается шумом и декодируется независимо от других кодовых блоков.

В разделимых кодах всегда можно выделить информационные символы, содержащие передаваемую информацию, и контрольные (проверочные) символы, которые являются избыточными и служат исключительно для коррекции ошибок. Неразделимые коды не имеют четкого разделения кодовой комбина-

ции на информационные и проверочные символы. К ним относятся коды с постоянным весом и коды Плоткина [2].

Разделимые блочные коды, в свою очередь, делятся на несистематические и систематические. Наиболее многочисленный класс разделимых кодов составляют систематические коды. Основная их особенность в том, что проверочные символы образуются как линейные комбинации информационных символов. К систематическим кодам относятся коды с проверкой на четность, коды с повторением, корреляционный, инверсный, коды Хэмминга, Голея, Рида-Маллера, Макдональда, Варшамова, с малой плотностью проверок на четность, итеративный код [2].

В несистематических кодах проверочные символы представляют собой суммы подблоков с l разрядами, на которые разделена последовательность информационных символов. К этим кодам относятся коды Бергера.

Разновидностью систематических кодов являются циклические коды. Кроме всех свойств систематического кода, циклические коды имеют следующее свойство: если некоторая кодовая комбинация принадлежит коду, то получающаяся путем циклической перестановки символов новая комбинация также принадлежит данному коду. К наиболее известным циклическим кодам относятся простейшие коды, коды Хэмминга, Боуза-Чоудхури-Хоквингема, мажоритарные, коды Файра, Абрамсона, Миласа-Абрамсона, Рида-Соломона, компаундные коды.

Отличительной особенностью непрерывных кодов является то, что первичная последовательность символов, несущих информацию, непрерывно преобразуется по определенному закону в другую последовательность, содержащую избыточное число символов. Здесь процессы кодирования и декодирования не требуют деления кодовых символов на блоки.

5.1.3. Основные характеристики и свойства блочных кодов

При блочном кодировании сообщениям источника ставится в соответствие M_P разрешенных кодовых слов длиной n, где $M_P = m^k$, m – основание кода,

k – количество информационных символов. Общее число возможных кодовых комбинаций (алфавит) определятся $M = m^n$. Блочный код обозначается (n,k). Скоростью блочного кода называется отношение длины последовательности информационных символов к общей длине кодового слова блочного кода: $R = \frac{k}{n}$. Будем в дальнейшем полагать, что источник может выдавать одно из двух сообщений: $c_i = 0$ или 1 (т.е. $m = 2$) [2].

Весом кодового слова $C = (c_0, c_1, c_2, ..., c_{n-1})$ называется количество ненулевых бит в нем:

$$W(C) = \sum_{i=0}^{n-1} c_i \text{ , где } \qquad c_i \neq 0 .$$

Расстоянием Хэмминга между двумя кодовыми словами называется число одноименных разрядов, в которых эти слова отличаются друг от друга:

$$d(C^1, C^2) = \sum_{i=0}^{n-1} (c_i^1 \oplus c_i^2),$$

где $\oplus$ – операция сложения по модулю два;

$$C^1 = (c_0^1, c_1^1, c_2^1, ..., c_{n-1}^1);$$

$$C^2 = (c_0^2, c_1^2, c_2^2, ..., c_{n-1}^2).$$

Пользуясь расстоянием Хэмминга как метрикой, можно определить корректирующие способности кода.

Можно показать, что для обнаружения в кодовом слове произвольной комбинации из s ошибок, необходимо и достаточно, чтобы расстояние Хэмминга для любых двух разрешенных кодовых слов было на **1** больше, чем число s обнаруживаемых ошибок:

$$d_{\min}(C^1, C^2) = s + 1 .$$

Для исправления t ошибок, необходимо и достаточно, чтобы $d_{\min}(C^1, C^2) = 2t + 1$. Исправление ошибок происходит по правилу: если принята запрещенная комбинация, то считается переданной ближайшая разрешенная комбинация. При этом будут исправлены все ошибки кратности: $t \leq (d-1)/2$.

Для того чтобы исправить t и обнаружить s ошибок в кодовом слове, не-

обходимо и достаточно, чтобы $d_{\min}\left(C^1, C^2\right) = 2t + s + 1$.

5.2. Эффективность помехоустойчивого кодирования

При проектировании систем передачи информации оценка достоверности обмена информацией определяется вероятностью искажения двоичного символа передаваемого сообщения $p_{ош}$.

Для двоичной последовательности, содержащей n символов, при безызбыточном кодировании вероятность правильного приема последовательности:

$$p_{пр}(n) = \left(1 - p_{ош}\right)^n,$$

а вероятность ошибки в принятой последовательности:

$$p_{ош}(n) = 1 - \left(1 - p_{ош}\right)^n. \qquad (5.1)$$

Эту формулу можно записать в следующем виде:

$$p_{ош}(n) = C_n^1 \cdot p_{ош}^1 - C_n^2 \cdot p_{ош}^2 + C_n^3 \cdot p_{ош}^3 - \ldots C_n^i \cdot p_{ош}^i,$$

где $C_n^i = \dfrac{n!}{i!(n-i)!}$ – число сочетаний из n по i.

Использование избыточных кодов позволяет исправлять или обнаруживать в зависимости от кодового расстояния ошибки той или иной кратности. При независимых ошибках вероятность появления кратных ошибок определяется по формуле Бернулли:

$$p_{ош}(i, n) = C_n^i \cdot p_{ош}^i \left(1 - p_{ош}\right)^{n-i},$$

где $i = 1,2,3,...$ – кратность ошибок.

5.2.1. Эффективность кода в режиме исправления ошибок

Для кодов, исправляющих ошибки кратности до t, вероятность исправления ошибки определяется выражением:

$$p_{исп} = \sum_{i=1}^{t} p_{ош}(i, n), \text{ или } p_{исп} = \sum_{i=1}^{t} C_n^i \cdot p_{ош}^i \left(1 - p_{ош}\right)^{n-i}. \qquad (5.2)$$

Прием кодовых слов в режиме исправления ошибок в общем случае может сопровождаться следующими ситуациями:

кодовое слово принято без ошибок, вероятность этого события $p_{пр}(n)$;

кодовое слово принято с ошибкой, причем $p_{ош}(n)=1-p_{пр}(n)$;

кодовое слово принято с ошибкой, которая исправляется с вероятностью $p_{испр}$;

кодовое слово принято с ошибкой, которая не исправляется данным кодом, вероятность этого события $p_{ни}$.

Поскольку:

$$p_{ош}(n)=p_{испр}+p_{ни}.$$

то вероятность появления неисправляемых ошибок:

$$p_{ни}=p_{ош}(n)-p_{испр}.$$

Используя формулы (5.1) и (5.2), получим выражение для определения вероятности появления неисправляемых ошибок в канале без памяти:

$$\begin{aligned} p_{ни} &= 1-\left(1-p_{ош}\right)^n-\sum_{i=1}^{t} C_n^i\cdot p_{ош}^i\left(1-p_{ош}\right)^{n-i}\approx \\ &\approx \sum_{i=t+1}^{n} C_n^i\cdot p_{ош}^i\left(1-p_{ош}\right)^{n-i}. \end{aligned} \qquad (5.3)$$

Формула (5.3) позволяет вычислить вероятность ошибки в последовательности длиной n при передаче информации с помощью кода, исправляющие t – кратные ошибки в дискретном симметричном канале без памяти (ДСКБП).

Для каналов с памятью эта формула имеет вид:

$$p_{ни}\approx\left(\frac{n}{t+1}\right)^{1-\alpha}\cdot p_{ош}, \qquad (5.4)$$

где α – коэффициент группирования ошибок.

Для коротковолновых радиолинии $\alpha=0{,}3...0{,}4$; для радиорелейной линии $\alpha=0{,}3...0{,}5$; для проводных линий связи $\alpha=0{,}5...0{,}7$.

5.2.2. Эффективность кода в режиме обнаружения ошибок

В каналах с обратной связью помехоустойчивые коды используются в режиме обнаружения ошибок. В таких каналах осуществляется запрос повтор-

ной передачи кодовых комбинаций (или слов), принятых с ошибкой.

Для кодов, используемых в режиме обнаружения ошибок, характерны следующие ситуации:

кодовое слово принято без ошибок, с вероятностью $p_{пр}(n)$;

кодовое слово принято с ошибкой, которая обнаруживается, вероятность такого события $p_{оо}$;

кодовое слово принято с ошибкой, которая с вероятностью $p_{но}$ не обнаруживается, при этом:

$$p_{ош}(n) = p_{оо} + p_{но}.$$

Поскольку искаженные комбинации, которые обнаруживаются декодером, получателю не выдаются, то вероятность получения ошибочных комбинаций получателем оценивается только как $p_{но}$.

При использовании кода в режиме обнаружения вероятность такого события:

$$p_{но} = \sum_{i=s+1}^{n} W_i \cdot p_{ош}^{i} \left(1 - p_{ош}\right)^{n-i}, \qquad (5.5)$$

где W_i – весовая характеристика кода, которая указывает число кодовых комбинаций с весом i в коде, в которых не обнаруживается ошибка.

Для кодов Хэмминга данная характеристика определяется выражением:

$$W_i = \frac{2^k}{2^n} \cdot C_n^i$$

где $\frac{2^k}{2^n}$ – процентное содержание разрешенных слов с весом i.

Отсюда вероятность того, что ошибка не обнаруживается в зависимости от используемого канала определяется следующими соотношениями.

Для каналов без памяти:

$$p_{но} = 2^{k-n} \sum_{i=d}^{n} C_n^i \cdot p_{ош}^{i} \left(1 - p_{ош}\right)^{n-i}, \qquad (5.6)$$

Для каналов с памятью формула (5.5) примет вид:

$$p_{но} \approx 2^{k-n} \cdot \left(\frac{n}{d}\right)^{1-\alpha} \cdot p_{ош}. \qquad (5.7)$$

5.2.3. Эффективность помехоустойчивых кодов

Под эффективностью помехоустойчивого кода понимается отношение вероятности ошибки в элементе кодовой комбинации безизбыточного кода к вероятности ошибки на выходе декодера канала (в режиме исправления или обнаружения ошибок) для каналов с памятью и без памяти:

$$\eta = \frac{p_{ош}}{p_{ДКи(о)}}.$$

Известно (5.1), что вероятность ошибки в кодовой комбинации длиной n при безизбыточном кодировании:

$$p_{ош}(n) = 1 - \left(1 - p_{ош}\right)^n.$$

Вероятность ошибочного приема, на один элемент, на выходе декодера в режиме исправления ошибки, соответственно для каналов ДСКБП (5.3) и каналов с памятью (5.4) составляет:

$$p_{ДК(и)} \approx \frac{p_{ни}}{n}.$$

В режиме обнаружения ошибки для каналов ДСКБП (5.6) и каналов с памятью (5.7):

$$p_{ДК(о)} \approx \frac{p_{но}}{n}.$$

Таким образом, с использованием ранее полученных выражений для ДСКБП:

$$p_{ДК(и)} = \frac{1}{n} \cdot \sum_{i=t+1}^{n} C_n^i \cdot p_{ош}^i \left(1 - p_{ош}\right)^{n-i},$$

$$p_{ДК(о)} = \frac{2^{k-n}}{n} \cdot \sum_{i=d}^{n} C_n^i \cdot p_{ош}^i \left(1 - p_{ош}\right)^{n-i}.$$

Для канала с памятью:

$$p_{ДК(и)} \approx \frac{1}{n} \cdot \left(\frac{n}{t+1} \right)^{1-\alpha} \cdot p_{ош},$$

$$p_{ДК(о)} \approx \frac{2^{k-n}}{n} \cdot \left(\frac{n}{d} \right)^{1-\alpha} \cdot p_{ош}.$$

Следует отметить, что применение избыточного кода означает увеличение длины кодовой комбинации по сравнению с безызбыточным кодом. При этом увеличение n может производиться при сохранении прежней длительности передачи комбинации ($\tau_{кода} = k \cdot \tau_0 = const$) или при сохранении прежней длительности символа ($\tau_0 = const$).

В первом случае длительность символа будет уменьшаться. Следовательно, изменится отношение сигнал/шум на входе демодулятора приемника:

$$h_к^2 = \frac{k}{n} \cdot h^2.$$

Тогда вероятность ошибки на входе декодера:

$$p_{ош} = f\left[\frac{k}{n} \cdot h^2 \right].$$

Во втором случае $h_к^2 = h^2$, соответственно $p_{ош} = f\left[h^2\right]$.

Сравнивая величину $p_{ДК}$ с вероятностью ошибки символа для безызбыточного кода, можно установить, при каких условиях применение избыточного кода позволяет повысить помехоустойчивость приема.

5.3. Математические основы теории помехоустойчивого кодирования

Основой построения наиболее важных из известных кодов является их алгебраическая структура, которая облегчает изучение различных свойств кода, а также обеспечивает возможность практической реализации кодирующих и декодирующих устройств.

В данном разделе излагаются основные понятия алгебры, необходимые для изучения теории помехоустойчивых кодов. Основные теоремы приводятся

без доказательств. Заинтересованный читатель найдет их в [25, 26, 30, 33].

Важнейшими объектами изучения в алгебре являются алгебраические системы, т.е. множества, в которых определены одна или несколько операций, таких, например, как сложение или умножение. Под операцией в общем случае подразумевается однозначная функция двух переменных, которая может быть записана в виде $f(a,b)=a*b=c$, где $*$ – знак операции. Примерами алгебраических систем являются группы, кольца, поля и др. Их свойства рассматриваются в настоящем разделе.

5.3.1. Краткие сведения из теории чисел

Если a, b и c – целые числа и $a=bc$, то говорят, что a делится на b или что b является делителем a. Наибольшим общим делителем (НОД) двух чисел называется наибольшее целое положительное число, являющееся делителем обоих этих чисел. Говорят, что два числа взаимно просты, если их НОД равен 1. Для любой пары целых чисел a и b существует единственная пара чисел q (частное) и r (остаток), таких, что $a=qb+r$, где $0<r<|b|$.

Если два числа a и b дают при делении на число p один и тот же остаток, то говорят, что числа a и b сравнимы по модулю p. Сравнение записывается в виде $a=b(\mathrm{mod}\,p)$. Эквивалентным определением сравнимости двух чисел является делимость их разности на p.

Отметим основные свойства сравнений.

1. Если $a=c(\mathrm{mod}\,p)$ и $b=c(\mathrm{mod}\,p)$, то $a=b(\mathrm{mod}\,p)$.

2. Над сравнениями можно производить операции, аналогичные операциям над равенствами, т.е., если $a_1=b_1(\mathrm{mod}\,p)$ и $a_2=b_2(\mathrm{mod}\,p)$, то $(a_1+a_2)=(b_1+b_2)(\mathrm{mod}\,p)$, $(a_1-a_2)=(b_1-b_2)(\mathrm{mod}\,p)$ и $a_1d=b_1d(\mathrm{mod}\,p)$, где d – произвольное целое число.

3. Обе части сравнения и их модуль можно разделить на их общий делитель, т.е., если $a=a_1d$, $b=b_1d$ и $p=p_1d$, то из $a=b(\mathrm{mod}\,p)$ следует $a_1=b_1(\mathrm{mod}\,p_1)$.

4. Два числа, сравнимые по модулю p, сравнимы и по модулю l, если l –

любой делитель p.

Все числа, сравнимые по модулю p, образуют класс вычетов по модулю p. Любое число в классе называется вычетом по модулю p. Всем числам класса вычетов соответствует один и тот же остаток. Так как всего имеется p остатков: $0,1,2,...,p-1$, то существует p различных классов вычетов. Вычет, равный самому остатку, называется наименьшим неотрицательным вычетом. Выбрав из каждого класса вычетов по модулю p по одному вычету, получим полную систему вычетов по модулю p. Обычно в качестве полной системы вычетов используют наименьшие неотрицательные вычеты: $0,1,2,...,p-1$.

Пример 5.1. Пусть $p=4$, тогда числа $1,5,9,13,17,21$ и т.д. образуют класс вычетов по модулю 4. Наименьший вычет в этом классе равен 1, а полную систему вычетов по модулю 4 образуют числа $\{0,1,2,3\}$.

С учетом определенного выше понятия сравнения чисел по модулю p введены операции сложения и умножения чисел по модулю произвольного целого числа p.

При этом результат применения операции сложения (умножения) двух чисел по модулю p равен наименьшему вычету класса, к которому принадлежит число, получаемое в результате обычного сложения (умножения) чисел. Другими словами, результат применения операции сложения (умножения) чисел по модулю p равен остатку от деления числа, получаемого при обычном сложении (умножении) чисел, на модуль p.

Пример 5.2. При $p=5$ таблицы сложения и умножения чисел по модулю 5 выглядят следующим образом:

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

×	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

5.3.2. Группы

Определение группы

Группой G называется множество элементов, для которых определена некоторая операция $*$ и выполняются следующие аксиомы:

G.1. Операция может быть применена к любым двум элементам группы, в результате чего получается третий элемент группы, т.е., если $a \in G$ и $b \in G$, то $a * b \in G$.

G.2. Для любых трех элементов a, b и c из G $a*(b*c)=(a*b)*c$.

G.3. В G существует единичный элемент 1, т.е. такой, что $a*l=l*a=a$ для любого $a \in G$.

G.4. Для любого элемента $a \in G$ существует обратный элемент a^{-1} такой, что $a*a^{-1}=a^{-1}*a=l$.

Аксиома G.1 определяет замкнутость операции в группе. Обычно операции над элементами записывают в виде $a+b=c$ и называют сложением или в виде $a \cdot b=c$ и называют умножением, даже если они не являются обычными сложением и умножением. В соответствии с двумя записями операций различают аддитивную и мультипликативную группы.

Свойство операции, сформулированное в виде аксиомы G.2, называют ассоциативностью. Она означает, что порядок выполнения операций несущественен, и поэтому скобки не нужны.

Аксиома G.3 постулирует обязательное существование единичного элемента. Для аддитивной группы единичный элемент называют нулем, обозначают 0 и определяют из уравнения $0+a=a+0=a$. Для мультипликативной группы единичный элемент называют единицей и определяют из уравнения $1 \cdot a=a \cdot 1=a$.

Аксиома G.4 требует для каждого элемента группы существования обратного элемента. Если групповая операция – сложение, то элемент, обратный a, обозначается $(-a)$ и находится из уравнения $a+(-a)=(-a)+a=0$. Для мультипликативной группы обратный к a элемент обозначается a^{-1} и находится из уравнения $a \cdot a^{-1}=a^{-1} \cdot a=1$.

Группа называется коммутативной или абелевой, если кроме аксиом G.1 – G.5 выполняется следующая аксиома коммутативности.

G.5. Для двух произвольных элементов a и b из G справедливо $a*b=b*a$.

Примеры групп

Пример 5.3. Одна из простейших аддитивных групп состоит из двух элементов, одним из которых является единичный элемент 0. Второй элемент обозначим через a. В соответствии с G.4 должен существовать обратный элемент, такой, что $a+(-a)=0$. Значит, $(-a)=a$, и правило сложения записывается в виде: $H;\ 0+0=0;\ a+a=0$. При $H=(h_1,h_2,...,h_m)$ имеем правило сложения по модулю $p=2$.

Пример 5.4. Совокупность всех действительных чисел образует группу относительно операции обычного сложения. Единичным элементом группы (нулем) является число 0.

Пример 5.5. Совокупность всех действительных чисел без нуля образует мультипликативную группу. Единичным элементом при этом является 1, а обратным – число $1/a$.

Пример 5.6. Совокупность двоичных n-символьных комбинаций образует группу из 2^n элементов, если в качестве групповой операции используется посимвольное сложение по модулю 2. Так, если $a=(101110)$, $b=(111011)$, то $c=a+b=(010101)$. Единичным является элемент (000000), а обратный элемент равен самому элементу, т.к. $110010+110010=000000$.

Пример 5.7. Полная система вычетов по модулю 6 $(G=\{0,1,2,3,4,5\})$ является группой с операцией сложения по модулю 6. Единичным элементом этой группы является 0, а обратный элемент находится из равенства $a+(-a)=0(\mathrm{mod}\,6)$. Так, если $a=2$, то $(-a)=4$ и т.д.

Все рассмотренные в примерах группы являются абелевыми.

Теорема 5.1. Группа содержит один единичный элемент, и каждый элемент группы имеет единственный обратный элемент.

Легко видеть, что в примерах 1 – 5 утверждения теоремы выполняются.

Число элементов в группе называется порядком группы. Если порядок конечен, группа называется конечной, в противном случае – бесконечной группой. В примерах 1.3, 1.6 и 1.7 рассмотрены конечные группы 2-го, 2^n-го и 6-го порядков, а в примерах 1.4 и 1.5 – бесконечные группы.

Подгруппы

Подмножество элементов группы G называется подгруппой H, если оно удовлетворяет всем аксиомам группы. Для того чтобы определить, является ли H подгруппой G, надо проверить только замкнутость операции и наличие обратных элементов. Подгруппами группы, рассмотренной в примере 2, являются множества: целых чисел, чисел, делящихся на 3, и т. д.

Смежные классы

Пусть задана конечная группа $G=(g_1,g_2,...,g_n)$ содержащая подгруппу $H=(h_1,h_2,...,h_m)$. Табл. 5.1 составлена следующим образом. Первая строка состоит из элементов подгруппы: она начинается с единичного элемента 1, и каждый элемент подгруппы появляется в строке только один раз. Первым элементом второй строки может быть любой элемент группы, не вошедший в первую строку, а все остальные элементы получаются путем применения групповой операции $g_1 * h_i$. Аналогично образуются третья, четвертая и т.д. строки, каждая с неиспользованным прежде элементом группы в начале строки, до тех пор, пока каждый элемент группы не войдет в таблицу.

Таблица 5.1

Разложение группы на смежные классы

$h_1 = l$	h_2	h_3	⋯	h_n
$g_1 * h_1 = g_1$	$g_1 * h_2$	$g_1 * h_3$	⋯	$g_1 * h_n$
$g_2 * h_1 = g_2$	$g_2 * h_2$	$g_2 * h_3$	⋯	$g_2 * h_n$
⋯	⋯	⋯	⋯	⋯
$g_m * h_1 = g_m$	$g_m * h_2$	$g_m * h_3$	⋯	$g_m * h_n$

Полученная таблица задает разложение группы на смежные классы. Совокупность элементов в каждой строке называется левым смежным классом, а элемент в первом столбце строки называется образующим смежного класса.

Число смежных (т.е. неперекрывающихся) классов k в разложении группы по подгруппе называется индексом H в G.

Правые смежные классы получаются, если для нахождения элементов строк применить операцию $h_i * g_1$. Для коммутативной группы левый и правый смежные классы совпадают.

Отметим основные свойства смежных классов.

1. Смежные классы не имеют общих элементов. Если у двух смежных классов оказался общий элемент, то такие смежные классы совпадают.

2. Левый (правый) смежный класс содержит столько элементов, каков порядок группы H.

3. Порядок n конечной группы G есть произведение порядка m подгруппы H на ее индекс k в группе G (на число смежных классов).

Группу G можно рассматривать как объединение неперекрывающихся смежных классов.

Циклические группы

Пусть a – один из элементов конечной группы G порядка n. Обозначим элементы $a*a, a*a*a, ...$ через $a, a^2, a^3, ...$ (при использовании операции сложения элементы можно обозначать также $2a, 3a, ...$) и рассмотрим последовательность элементов $a, a^2, a^3, ...$ Так как группа G конечна, то существуют такие числа i и $j\,(j > i)$, что $a^i = a^j$ Но тогда $a^i = a^j = a^i a^{j-1}$ и $a^{j-i} = l$ (единичный элемент группы). Минимальное целое положительное число m такое, что $a^m = l$, называют порядком элемента a.

Очевидно, если m – порядок элемента a, то все m элементов $a, a^2, a^3, ..., a^{m-1}, a^m = l$ различны. Доказано, что множество элементов $H = (a, a^2, a^3, ..., a^{m-1}, a^m = l) = (l, a, a^2, a^3, ..., a^{m-1})$ является подгруппой группы G. Такая подгруппа называется циклической подгруппой, порожденной элементом a.

Если в группе G существует элемент a такой, что его порядок совпадает с порядком группы $\{m = n\}$, т.е. $G = \left(l, a, a^2, a^3, ..., a^{m-1}\right)$, то сама группа называется циклической. При этом элемент a называется порождающим элементом группы.

Теорема 5.2. Если a – порождающий элемент циклической группы порядка n, то a^k – порождающий элемент этой же группы, где k – число взаимно простое с n.

В примере 1.7 рассмотрена аддитивная циклическая группа 6-го порядка. Порождающими элементами этой группы являются 1 или 5. Циклическая группа с порождающим элементом a приведена в примере 1.3.

5.3.3. Кольца и поля

Определение кольца

Кольцом R называется множество элементов, на котором определены две операции – сложение и умножение, и в R выполняются следующие аксиомы:

R.1. Множество R является аддитивной абелевой группой.

R.2. Для любых двух элементов a и b из R определено их произведение: $a \cdot b = c \in R$ (замкнутость операции умножения).

R.3. Для любых трех элементов a, b и c из R выполняется ассоциативный закон, т.е. $a(bc) = (ab)c$ и $a + (b + c) = (a + b) + c$.

R.4. Для любых трех элементов a, b и c из R выполняется дистрибутивный закон, т.е. справедливы равенства: $a(b + c) = ab + ac$ и $(b + c)a = ba + ca$.

Заметим, что в кольце для операции умножения аксиомы G.3, G.4 и G.5 могут не выполняться. Если же операция умножения коммутативна в кольце, то такое кольцо называется коммутативным. Если в кольце существует единичный элемент относительно операции умножения (выполняется аксиома G.3), то это кольцо называется кольцом с единицей.

Пример 5.8. Все целые положительные и отрицательные числа и нуль образуют коммутативное кольцо с единицей относительно обычных операций сложения и умножения.

Пример 5.9. Легко убедиться, что полная система вычетов по модулю p также образует коммутативное кольцо с единицей относительно операций сложения и умножения по модулю p.

Определение поля

Полем F называют коммутативное кольцо с единицей, в котором каждый ненулевой элемент имеет мультипликативный обратный элемент (т.е. обратный по умножению).

Другими словами, полем называют множество, которое является аддитивной абелевой группой; ненулевые же элементы этого множества образуют мультипликативную абелевую группу, и выполняется закон дистрибутивности.

По аналогии с группами число элементов поля называется порядком поля. Поля, порядки которых конечны, называются конечными полями. Конечные поля имеют наибольшее значение в теории кодирования.

Отметим некоторые свойства полей, вытекающие из их определения.

1. Для любого элемента поля $a \cdot 0 = 0 \cdot a = 0$.

2. Для ненулевых элементов a и b поля $a \cdot b \neq 0$.

3. Для любых элементов a и b поля $a + b \neq 0$.

4. Если $a \cdot b = a \cdot c$ и $a \neq 0$, то $b = c$.

Пример 5.10 Множество всех действительных чисел образует поле. Существует также поле комплексных чисел, поле рациональных чисел, но не может быть поля целых чисел, поскольку обратные элементы по умножению, кроме единицы, не являлись бы целыми.

Пример 5.11. Множество чисел $(0, 1, 2, ..., p-1)$, где p – простое число, образует конечное поле, в котором сложение и умножение производятся по модулю p.

Пример 5.12. При $p = 2$ имеем простейшее двоичное поле, состоящее из двух элементов 0 и 1. Эти элементы являются соответственно единичными элементами относительно операций сложения и умножения по модулю 2, которые определяются правилами: $0 + 0 = 1 + 1 = 0$; $1 + 0 = 0 + 1 = 1$; $0 \cdot 0 = 0$; $0 \cdot 1 = 1 \cdot 0 = 0$; $1 \cdot 1 = 1$. Так как $(-1) = 1$, то операции сложения и вычитания в двоичном поле

совпадают, а так как $1^{-1}=1$, также совпадают операции умножения и деления. Это поле находит широкое применение в теории и технике помехоустойчивого кодирования. Более сложные конечные поля рассмотрены в 5.3.5.

Кольцо полиномов

Рассмотрим полином (многочлен) $f(x)=f_0+f_1x+...+f_mx^m$. Если коэффициенты f_i, $i=0,1,...,m$, при степенях x являются элементами поля F, то говорят, что полином $f(x)$ задан над полем F.

Степенью полинома называется наибольшая степень переменной x с ненулевым коэффициентом. Многочлен называется нормированным, если коэффициент при наивысшей степени x равен 1. Два полинома

$$f(x)=\sum_{i=0}^{m}f_ix^i \text{ и } g(x)=\sum_{i=0}^{n}g_ix^i \qquad (5.8)$$

называются равными, если они имеют одинаковую степень, т.е. $m=n$, и равные коэффициенты $f_i=g_i$, $i=0,1,...,m$. При этом считается, что $x^0=l$, где 1 – единичный элемент поля F. Полином, все коэффициенты которого равны нулю, называется нулевым. Степень нулевого полинома равна нулю.

В кольце полиномов операции сложения и умножения вводятся следующим образом. Для двух полиномов (5.8) их сумма

$$f(x)+g(x)=\sum_{(i)}(f_i+g_i)x^i$$

а произведение

$$f(x)g(x)=\sum_{(i)}\left(\sum_{j=0}^{i}f_ig_{i-j}\right)x^i$$

В частности, если $g(x)=\alpha$, $\alpha\in F$, то $g(x)f(x)=\sum_{(i)}\alpha f_ix^i$. Нетрудно проверить, что при введенных таким образом операциях сложения и умножения множество $R[x]$ полиномов является кольцом, которое называется кольцом полиномов над полем F.

Свойства делимости полиномов в кольце

Пусть $f(x)$ и $g(x)$ – два полинома степени m и n соответственно, причем $m>n$. Говорят, что $f(x)$ делится на $g(x)$, если в кольце $R(x)$ существует тре-

тий полином $Q(x)$ такой, что $f(x)=Q(x)g(x)$. Деление полиномов в кольце $R(x)$ не всегда возможно даже на ненулевой многочлен. Например, деление невозможно, если степень делимого меньше степени делителя.

Укажем основные свойства делимости полиномов в кольце.

1. Если $f_1(x)$ и $f_2(x)$ – полиномы из $R(x)$ и $f(x)$ делится на $g(x)$, а $g(x)$ делится на $f(x)$, то многочлены $f(x)$ и $g(x)$ отличаются друг от друга лишь множителем нулевой степени, т.е. $f(x)=\alpha g(x)$, где α – элемент поля.

2. Если каждый из полиномов $f_1(x)$ и $f_2(x)$ делится на $g(x)$, то их сумма $f_1(x)+f_2(x)$ и разность $f_1(x)-f_2(x)$ делятся на $g(x)$.

3. Если $f_1(x)$, $f_2(x)$ и $f_3(x)$ – полиномы из $R(x)$ и $f_1(x)$ делится на $f_2(x)$, а $f_2(x)$ делится на $f_3(x)$, то $f_1(x)$ делится на $f_3(x)$.

4. Ненулевые элементы поля F являются делителями любого полинома из $R(x)$.

5. Для любой пары полиномов $a(x)$ и $g(x)$ существует единственная пара многочленов $Q(x)$ (частное) и $r(x)$ (остаток) таких, что $a(x)=Q(x)g(x)+r(x)$ причем степень $r(x)$ меньше степени $g(x)$.

6. Полином $d(x)$ называется наибольшим общим делителем (НОД) полиномов $a(x)$ и $g(x)$, если $d(x)$ – полином наивысшей степени, который делит как $a(x)$, так и $g(x)$. НОД обозначается: $d(x)=НОД[a(x),f(x)]$ Два полинома называются взаимно простыми, если их НОД равен 1.

Полином, который делится только на себя и на элемент поля F, называется неприводимым над полем F.

Кольцо вычетов по модулю $g(x)$

При описании блочных кодов [25, 30, 33] широко используется понятие кольца вычетов по модулю некоторого полинома $g(x)$ с коэффициентами из поля F.

Для полиномов существуют понятия, аналогичные введенным в 5.8 для чисел, если заменить в этих понятиях слово «число» словом «полином». Так, ес-

ли при делении полиномов $a(x)$ и $f(x)$ из $R[x]$ на $g(x)$ получаются одинаковые остатки, то многочлены $a(x)$ и $f(x)$ сравнимы между собой по модулю многочлена $g(x)$ из $R[x]$ или $a(x) = f(x) (\mathrm{mod}\, g(x))$.

Все полиномы, сравнимые между собой по модулю $g(x)$, образуют класс вычетов по модулю $g(x)$, а каждый полином класса называется вычетом по модулю $g(x)$. Каждый класс характеризуется своим представителем, в качестве которого обычно выбирают полином, степень которого меньше степени $g(x)$. Количество классов вычетов по модулю $g(x)$ равно числу многочленов, степени которых меньше степени $g(x)$.

Совокупность классов вычетов по модулю $g(x)$ образует кольцо вычетов по модулю $g(x)$. В качестве операций сложения и умножения в этом кольце используются сложение и умножение по модулю $g(x)$.

Пример 5.13. Рассмотрим кольцо классов вычетов по модулю полинома $g(x) = x^2 + x + 1$ над двоичным полем. Полиномы вида $a(x) = Q(x)g(x) + r(x)$, где $r(x)$ – произвольный полином, степень которого меньше 2, при фиксированном $r(x)$ образуют класс вычетов по модулю $x^2 + x + 1$. Так как всего имеется 4 разных полинома $r(x)$ степени меньше 2, то возможны 4 следующие класса вычетов:

$$r(x) = 0 \quad \leftrightarrow \quad a(x) = Q(x)\left(x^2 + x + 1\right)$$

$$r(x) = 1 \quad \leftrightarrow \quad a(x) = Q(x)\left(x^2 + x + 1\right) + 1$$

$$r(x) = x \quad \leftrightarrow \quad a(x) = Q(x)\left(x^2 + x + 1\right) + x$$

$$r(x) = x + 1 \quad \leftrightarrow \quad a(x) = Q(x)\left(x^2 + x + 1\right) + x + 1$$

Здесь $Q(x)$ – произвольный полином. В качестве представителей классов обычно выбирают вычеты наименьшей степени, которые совпадают с полиномами $r(x)$ и образуют кольцо классов вычетов по модулю полинома $x^2 + x + 1$, т.е. множество $(0,1,x,x+1)$.

5.3.4. Векторное пространство

Определение вектора

Вектором называется упорядоченное множество из n элементов поля, обозначаемое как $[a_1, a_2, ..., a_n]$. Величины $a_i \in F$ называются компонентами (координатами) вектора. Число компонентов вектора n называется длиной вектора. Векторы считаются равными, если равны их соответствующие компоненты. Число ненулевых компонентов вектора называют весом вектора [33].

Сложение двух векторов длины n определяется следующим образом:

$$[a_1, a_2, ..., a_n] + [b_1, b_2, ..., b_n] = [a_1 + b_1, a_2 + b_2, ..., a_n + b_n].$$

Умножение элемента поля на вектор производится покомпонентно:

$$\alpha[b_1, b_2, ..., b_n] = [\alpha b_1, \alpha b_2, ..., \alpha b_n],$$

причем сложение и умножение компонентов векторов происходит по правилам сложения и умножения в поле F.

Для векторов введено понятие нормы [25, 33], которая для вектора A определяется как $\|A\| = \sum_{i=1}^{n} a_i^2$, где символ $\sum$ означает суммирование в поле действительных чисел. Если компоненты вектора принадлежат двоичному полю, то норма вектора совпадает с числом его ненулевых компонентов, т.е. с его весом.

Вектор $v = \alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k$, где α_i – элементы поля, называют линейной комбинацией векторов $u_1, u_2, ..., u_k$. Векторы $u_1, u_2, ..., u_k$ называются линейно зависимыми, если в F существуют такие элементы $\alpha_1, \alpha_2, ..., \alpha_k$, по крайней мере один из которых не равен нулю, такие что $\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_k u_k = 0$ и линейно независимыми в противном случае. Если векторы линейно зависимы, то любой из них может быть выражен через линейную комбинацию остальных.

Определение векторного пространства

Множество V называется векторным пространством, если для него выполняются следующие аксиомы:

V. 1. Множество V является аддитивной абелевой группой.

V.2. Для любого вектора $v \in V$ и любого скаляра – элемента α поля F

определено произведение αv, являющееся вектором. Это произведение определено так, что $lv = v$, где l – единичный элемент поля F.

V.3. Выполняются законы дистрибутивности

$$\alpha(v_1 + v_2) = \alpha v_1 + \alpha v_2 \text{ и } (\alpha + \beta)v = \alpha v + \beta v,$$

где α, β – скаляры, а v_1 и v_2 – векторы.

V.4. Выполняется закон ассоциативности

$$(\alpha\beta)v = \alpha(\beta v),$$

где α, β – скаляры, а v – вектор.

Свойства векторного пространства

1. Максимальное число линейно независимых векторов в V называется размерностью пространства V над полем F.

2. Совокупность n любых линейно независимых векторов называется базисом и-мерного пространства, если каждый из векторов пространства может быть представлен в виде линейной комбинации этих векторов. Векторы совокупности называются базисными.

3. Подмножество W векторного пространства V такое, что любая линейная комбинация векторов этого подмножества снова принадлежит W, называется подпространством пространства V. Легко проверить, что все векторы подпространства удовлетворяют аксиомам V.1 – V.4. Очевидно, что размерность подпространства не превышает размерности пространства, т.к. во всем пространстве содержится не более n линейно независимых векторов. Каждое подпространство можно рассматривать как самостоятельное пространство. Следовательно, каждое подпространство имеет свой базис.

4. Скалярным произведением двух векторов одинаковой длины n: $v = [a_1, a_2, ..., a_k]$ и $u = [b_1, b_2, ..., b_k]$ называется скаляр, определяемый как

$$(vu) = (a_1b_1 + a_2b_2 + ... + a_nb_n).$$

Можно показать, что $(vu) = (uv)$ и $(w(u + v)) = (wu) + (wv)$.

Если скалярное произведение двух векторов равно нулю, то говорят,

что эти векторы ортогональны. Два пространства называются взаимно ортогональными, если каждый вектор одного пространства ортогонален любому вектору другого пространства.

Множество всех векторов пространства V, ортогональных подпространству V_1, образуют подпространство V_2 пространства V. Подпространство V_2 часто называют нулевым пространством для V_1.

Можно показать, что если V_1 – подпространство размерности k n-мерного векторного пространства V, то размерность нулевого пространства равна $n-k$.

5. Для векторного пространства определено понятие расстояния между двумя векторами, которое совпадает с нормой разности этих векторов

$$d(A,B)=\|A-B\|=\sum_{i=1}^{n}(a_i-b_i)^2 ,$$

где суммирование производится в поле действительных чисел.

5.3.5. Конечные поля

Определение конечного поля

Ранее в 1.3.2 дано определение поля F как коммутативного кольца с единицей, в котором каждый ненулевой элемент имеет мультипликативный обратный элемент. В теории помехоустойчивых кодов весьма важное значение имеют поля, образованные конечным множеством элементов – так называемые конечные поля Галуа (Galois Field), обозначаемые GF. В связи с этим дадим их развернутое определение.

Конечным полем GF называется конечное множество элементов, замкнутое по отношению к двум заданным в нем операциям комбинирования элементов. Под замкнутостью понимается тот факт, что результаты операций не выходят за пределы конечного множества введенных элементов. Для конечных полей выполняются следующие аксиомы.

GF.1. Из введенных операций над элементами поля одна называется сложением и обозначается как $a+b$, а другая - умножением и обозначается как

ab.

GF.2. 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 $a \cdot a^{-1}=1$. Наличие обратных элементов позволяет наряду с операциями сложения и умножения выполнять также вычитание и деление: $a-b=a+(-b)$, $a/b = a \cdot b^{-1}$. Поэтому иногда просто говорят, что в поле определены все четыре арифметические операции (кроме деления на 0).

GF.3. Поле всегда содержит мультипликативную единицу 1 и аддитивную единицу 0, такие что $a+0=a$, и $a \cdot 1 = a$ для любого элемента поля.

GF.4. Для введенных операций выполняются обычные правила ассоциативности $a+(b+c)=(a+b)+c$, $a(bc)=(ab)c$, коммутативности $a+b=b+a$, $ab=ba$ и дистрибутивности $a(b+c)=ab+ac$.

GF.5. Результатом сложения или умножения двух элементов поля является третий элемент из того же конечного множества.

Аксиомы GF.1 – GF.5 являются общими для полей как с конечным, так и с бесконечным числом элементов. Специфику же конечного поля определяет аксиома GF.5, где ключевыми являются слова «из того же конечного множества».

Требование конечности множества определяет ряд ограничений как на количество элементов поля GF, так и на понятия «сложение» и «умножение».

Конечные поля существуют не при любом числе элементов, а только в том случае, если их количество – простое число p или его степень p^m, где m – целое. В первом случае поле $GF(p)$ называется простым, а во втором – расширением $GF(p^m)$ простого поля.

Очевидно, операции комбинирования элементов конечного поля не могут быть обычными сложением и умножением. Выполнение аксиомы GF.5 для простого конечного поля обеспечивается совершением арифметических операций по модулю числа p, которое носит название характеристики конечного поля. Можно убедиться, что в кольце вычетов по модулю p (см. 5.3.1) каждый нену-

левой элемент имеет обратный элемент тогда и только тогда, когда p – простое число [25, 26, 33]. Следовательно, кольцо вычетов по модулю простого числа p является простым полем $GF(p)$. Элементами этого поля являются целые числа $0, 1, 2, ..., p-1$. Операции сложения и умножения в таком поле производятся по модулю p. Пример простейшего двоичного поля $GF(2)$ приведен в 5.3.3.

Элементами β расширенного поля $GF(p^m)$ могут быть, например, все многочлены степени $m-1$ или меньше, коэффициенты которых лежат в простом поле $GF(p)$. Число p^m называется порядком расширенного поля и определяет количество различных многочленов.

Правила сложения и умножения полиномов – элементов расширенного конечного поля получаются из обычных правил сложения и умножения полиномов с последующим приведением результата по модулю некоторого специального многочлена $p(x)$ степени m. Такое приведение эквивалентно делению многочлена результата на $p(x)$ и использованию только остатка.

Очевидно, любые результаты вычислений в поле после приведения по модулю $p(x)$ должны оставаться обратимыми – только в этом случае наша система образует поле. Для этого используемый полином $p(x)$ должен быть неприводимым в поле $GF(p)$, т.е. его нельзя разложить на множители, используя только многочлены с коэффициентами из $GF(p)$. Это означает также, что $p(x)$ не имеет корней в поле $GF(p)$. Аналогом неприводимого полинома является простое число в поле вещественных чисел.

К сожалению, регулярных методов поиска неприводимых полиномов не существует, они обычно определяются перебором. К настоящему времени имеются подробные таблицы неприводимых полиномов [30, 33].

Особым свойством конечных полей является связь между собой всех ненулевых элементов β и возможность выражения каждого из них через один элемент α, называемый примитивным, как некоторую целую степень этого элемента. Множество $p^m - 1$ ненулевых элементов расширения $GF(p)$ образует циклическую мультипликативную группу (см. 5.3.2), т.е. элементы находятся

между собой в соотношении $1, \alpha, \alpha^2, ..., \alpha^{p^m-1}, \alpha^{p^m} = 1$. Примитивных элементов в $GF(p^m)$ может быть несколько.

Построение конечного поля

Построим конечное поле $GF(2)$ и его расширение $GF(2^4)$. Пусть элементами $GF(2)$ являются 0 и 1, а элементами $GF(2^4)$ – 16 всевозможных полиномов степени 3 и менее с коэффициентами из $GF(2)$:

$$0, 1, x, x+1, x^2, x^2+1, x^2+x, x^2+x+1, ..., x^3+x^2+x+1.$$

Теперь необходимо определить операции над элементами таким образом, чтобы их результаты не давали новых элементов, кроме уже введенных.

В поле $GF(2)$ обычные операции умножения (на 0 и 1) и деления (на 1) не выводят результат за пределы множества 0; 1. Однако при сложении и вычитании элементов это требование может уже не выполняться: $1+1=2$; $-1+(-1)=-2$ и т. д. Свойства конечного поля будут, очевидно, соблюдаться, если в качестве операции сложения использовать суммирование по модулю 2 $(\mathrm{mod}\,2)$:

$$0+0=0;\ 0+1=1;\ 1+0=1;\ 1+1=0. \qquad (5.9)$$

причем операции сложения и вычитания в поле $GF(2)$ совпадают. Этим мы будем пользоваться в дальнейшем, заменяя, например, полином вида x^n-1 на x^n-1 в тех случаях, когда полиномы заданы над полем характеристики 2. Если, однако, характеристика поля $p \neq 2$, такая замена неправомерна, и полиномы каждого вида нужно рассматривать самостоятельно.

В поле $GF(2^4)$ операцией, которая может вывести результат за пределы поля, является умножение многочленов. Обычное перемножение может дать полином степени больше 3, не принадлежащий множеству элементов $GF(2^4)$. Действительно, используя представление полиномов через векторы их коэффициентов [26], а также учитывая (5.9), получим

$$(1101)(1001) \leftrightarrow (x^3+x^2+1)(x^3+1) = x^6+x^5+x^3+x^3+x^2+1 =$$
$$x^6+x^5+(1+1)x^3+x^2+1 = x^6+x^5+x^2+1 \leftrightarrow 1100101.$$

Поэтому введем дополнительное условие, чтобы x удовлетворял некоторому уравнению степени $m=4$, например, $p(x)=x^4+x+1=0$ или $x^4=x+1$. Тогда

$x^5 = x^2 + x$; $x^6 = x^3 + x^2$; $x^7 = x^4 + x^3 = x^3 + x + 1$ и т.д., а $x^6 + x^5 + x^2 + 1 = x^3 + x^2 + x^2 + x + x^2 + 1 = x^3 + x^2 + x + 1 \leftrightarrow 1111$, т.е. не выходит за пределы поля $GF(2^4)$.

Нетрудно видеть, что результат проделанного преобразования полинома $x^6 + x^5 + x^2 + 1$ эквивалентен вычислению остатка от его деления на полином $p(x) = x^4 + x + 1$, т.е. приведению произведения многочленов по модулю $p(x)$:

$x^6 + x^5 + x^2 + 1$	$\leftrightarrow$	1100101	10011	$\leftrightarrow$	$x^4 + x + 1$
		10011	110	$\leftrightarrow$	$x^2 + x$ – частное
		10100			
		10011			
		01111			
		00000			
		1111	$\leftrightarrow$		$x^3 + x^2 + x + 1$ – остаток.

Или $x^6 + x^5 + x^2 + 1 = (x^2 + x)p(x) + x^3 + x^2 + x + 1 = (x^3 + x^2 + x + 1)(\mathrm{mod}\, p(x))$. Делением на полиномы первой степени x и $x+1$, и второй степени x^2, $x^2 + 1$ и $x^2 + x + 1$ можно убедиться, что рассматриваемый многочлен $p(x) = x^4 + x + 1$ неприводим над $GF(2)$, а следовательно, не имеет в нем корней. В противном случае $p(x)$ раскладывался бы на сомножители $x + 0 = x$ и $x+1$, ибо корнями в поле $GF(2)$ могут быть только 0 или 1.

Для нахождения примитивных элементов поля, как и неприводимых полиномов, приходится прибегать к таблицам. В нашем примере поля $GF(2^4)$, задаваемого многочленом $p(x) = x^4 + x + 1$, примитивным элементом является $\alpha = x$. Последовательно применяя равенства $\alpha^{i+1} = x\alpha^i$ и $p(x) = 0$ (или, что эквивалентно, $x^4 = x + 1$), получим упорядоченное по степеням примитивного элемента α множество элементов β, составляющее конечное поле $GF(2^4)$.

В табл. 5.2 даны различные представления элементов β поля $GF(2^4)$, заданного полиномом $p_1(x) = x^4 + x + 1$, а также полиномом $p_2(x) = x^4 + x^3 + 1$.

Представление элементов поля по степеням примитивного элемента α удобно, в частности, при умножении элементов друг на друга. Для этого доста-

точно сложить их степени по модулю $p^m - 1$ (применительно к табл. 5.2 – по модулю 15). Например,

$$\beta_{10}\beta_{13} = \left(x^2 + x + 1\right)\left(x^3 + x^2 + 1\right) = \alpha^{10}\alpha^{13} \leftrightarrow (10 + 13) \bmod 15 = 8 \leftrightarrow \alpha^8 = x^2 + 1.$$

Прямые вычисления дают то же, но более трудоемко:

$$\beta_{10}\beta_{13} = \left(x^2 + x + 1\right)\left(x^3 + x^2 + 1\right) = x^5 + x^4 + x^3 + x^4 + x^3 + x^2 + x^2 + x + 1 =$$
$$= x^5 + x + 1 = x^2 + x + x + 1 = x^2 + 1.$$

Таблица 5.2

Различные представления элементов поля $GF\left(2^4\right)$

Ненулевые элементы поля $GF\left(2^4\right)$	$p_1(x) = x^4 + x + 1$			$p_2(x) = x^4 + x^3 + 1$	
	Представление элементов поля через			Представление через	
	полином	вектор	степень α	вектор	степень γ
β_0	1	0001	$\alpha^0 = 1$	0001	1
β_1	x	0010	α^1	0111	γ^7
β_2	x^2	0100	α^2	1100	γ^{14}
β_3	x^3	1000	α^3	1111	γ^6
β_4	$x + 1$	0110	α^4	0110	γ^{13}
β_5	$x^2 + x$	0110	α^5	1011	γ^5
β_6	$x^3 + x^2$	1100	α^6	0011	γ^{12}
β_7	$x^3 + x + 1$	1011	α^7	1001	γ^4
β_8	$x^2 + 1$	0101	α^8	1101	γ^{11}
β_9	$x^3 + x$	1010	α^9	1000	γ^3
β_{10}	$x^2 + x + 1$	0111	α^{10}	1010	γ^{10}
β_{11}	$x^3 + x^2 + x$	1110	α^{11}	0100	γ^2
β_{12}	$x^3 + x^2 + x + 1$	1111	α^{12}	0101	γ^9
β_{13}	$x^3 + x^2 + 1$	1101	α^{13}	0010	γ^1
β_{14}	$x^3 + 1$	1001	α^{14}	1110	γ^8
			$\alpha^{15} = \alpha^0 = 1$		$\gamma^{15} = \gamma^0 = 1$

Ненулевые элементы $GF(2^4)$ расположены в порядке нарастания степени примитивного элемента и образуют циклическую группу порядка 15. При этом $\alpha^{15}=1$, $\alpha^{16}=\alpha$, $\alpha^{17}=\alpha^2$, … , $\alpha^{30}=1$ и т.д.

Нетрудно убедиться, что примитивным в поле $GF(2^4)$ является не только один элемент α, но и α^2, α^4, α^8 и ряд других (предлагается их отыскать самостоятельно), а α^3 и α^5 таковыми не являются.

Основные свойства конечных полей и полиномов

Связь между элементами конечного поля

Все ненулевые элементы β конечного поля $GF(2^m)$ являются степенями одного примитивного элемента:

$$\beta=\alpha^s,\quad s=0,1,2,...,p^m-2;\quad \alpha^{p^m-1}=1.$$

Порядок элемента поля

Порядком β элемента $\beta^{p^m}=\beta$ конечного поля называется наименьшее значение β, для которого $\beta^q=1$. Пусть $\beta=\alpha^i$. Поскольку ненулевые элементы β образуют циклическую группу, порядок элемента α^i может быть определен из равенства

$$q=\frac{p^m-1}{НОД\left[p^m-1,i\right]},$$

где $НОД$ – наибольший общий делитель. Порядки элементов x^q-1 лежат в пределах от 1 (элемент $\varphi(x)$) до n (примитивные элементы), но p^m-1 всегда кратно порядку элемента.

Возведение многочлена над полем $GF(p)$ в степень p

Если $\varphi(x)$ – произвольный многочлен, коэффициенты которого лежат в $GF(p)$, то $\varphi^p(x)=\varphi(x^p)$. Справедливость этого утверждения вытекает из того, что все по парные или многократные произведения в $\varphi^p(x)$ появляются с коэффициентами, которые делятся на p, и значит, равны 0 в $GF(p)$.

Так для многочлена над полем характеристики $p=2$ справедливо $\varphi^2(x)=\varphi(x^2)$, в чем можно убедиться на примере:

$$\varphi^2(x)=\left(x^2+x+1\right)=x^4+x^2+1+(1+1)\left(x^3+x^2+x\right)=x^4+x^2+1=\varphi\left(x^2\right),$$

Корни полиномов

Ключевым при построении кодов и их декодировании является вопрос о корнях полиномов, соответствующих кодовым комбинациям. Напомним, что из теории полиномов над полем вещественных чисел (не конечных!) известно, что полином степени m всегда имеет m корней, только не все они обязательно лежат в поле вещественных чисел (на вещественной оси). Часть корней может находиться в поле комплексных чисел как некотором расширении поля вещественных чисел.

Известная аналогия этому имеется и в конечных полях. Любой многочлен степени m, в том числе и неприводимый над полем $GF(p)$ (не имеющий корней среди элементов этого поля), всегда имеет m корней в расширении $GF\left(p^m\right)$, и этими корнями является часть элементов поля $GF\left(p^m\right)$. Как элементы конечного поля, корни находятся между собой в определенном соотношении. Если $\varphi(x)$ – неприводимый полином с коэффициентами из $GF(p)$ и β_1 – его корень, то $\beta_1^p,\beta_1^{p^2},\beta_1^{p^3},...$ также являются его корнями. В поле $GF\left(p^m\right)$ корнями неприводимого полинома степени m будут $\beta_1,\beta_2=\beta_1^2,\beta_3=\beta_1^4,...,\beta_m=\beta_1^{2^{m-1}}$.

Полиномы x^n-1

Для дальнейшего обсуждения процедур кодирования и декодирования полезно иметь в виду следующие свойства многочлена вида x^n-1. Для любого элемента β как циклической группы справедливо равенство $\beta^{p^m}=\beta$. Это означает, что любой из элементов β является корнем уравнения $x^{p^m}=x$ или, что то же самое, корнем полинома $x^{p^m}-x$ или $x\left(x^{p^m-1}-1\right)$. Нулевой элемент $\beta=0$ – корень полинома x, а каждый из ненулевых элементов поля $GF\left(p^m\right)$ – один из корней полинома $x^{p^m-1}-1$. Таким образом,

$$x^{p^m}-x=\prod_{i=1}^{p^m}(x-\beta_i). \qquad (5.10)$$

Пусть q – порядок элемента поля β, т.е. $\beta^q=1$. Следовательно, β – ко-

рень полинома $x^q - 1$. Если β является также и корнем неприводимого многочлена $\varphi(x)$, то $x^q - 1$ делится без остатка на $\varphi(x)$.

В более общем случае минимальное значение n, для которого произвольный многочлен $\varphi(x)$ без кратных корней делит $x^n - 1$, совпадает с наименьшим общим кратным (НОК) порядков корней $\varphi(x)$.

Многочлен $x^n - 1$ делится на $x^m - 1$ только в том случае, если n делится на m. Действительно, если корни $x^m - 1$ являются также корнями $x^n - 1$, то n должно делиться на m.

Циклотомические классы

Каждый из корней β_i полинома $\varphi(x)$ в поле $GF(p^m)$ есть степень примитивного элемента α. Показатели степеней, соответствующие корням

$$\beta_1 = \alpha^s,\ \beta_2 = \alpha^{sp},\ \beta_3 = \alpha^{sp^2},\ \beta_4 = \alpha^{sp^3},\ldots,$$

образуют циклотомический класс чисел $\{s, sp, sp^2, sp^3, \ldots\}$ по модулю $p^m - 1$, а весь набор показателей степеней примитивного элемента в поле $GF(p^m)$ распадается на не перекрывающиеся циклотомические классы K_s. Индекс s равен наименьшему из чисел в классе и называется представителем класса по модулю $p^m - 1$.

С другой стороны, как отмечалось в 5.3.5, каждый из $p^m - 1$ ненулевых элементов β поля $GF(p^m)$ является одним из корней полинома $x^{p^m-1} - 1$, который, в свою очередь, раскладывается на произведение неприводимых полиномов $\varphi_i(x)$ меньшей степени. Каждый из циклотомических классов содержит набор показателей степеней примитивного элемента, соответствующих корням одного из полиномов $\varphi_i(x)$.

Убедимся в этом на примере полинома $x^{15} - 1$ над полем $GF(2)$. Поскольку сложение и вычитание по модулю 2 здесь неразличимы, то записи $x^{15} - 1$ и $x^{15} + 1$ эквивалентны. Разложение $x^{15} + 1$ на неприводимые полиномы $\varphi_i(x)$ выглядит следующим образом [30]:

$$x^{15} + 1 = (x+1)(x^4 + x + 1)(x^4 + x^3 + x^2 + x + 1)(x^2 + x + 1)(x^4 + x^2 + 1).$$

В табл. 5.3 приведено распределение элементов поля β, представленных

степенями примитивного элемента α, по циклотомическим классам K_s с указанием соответствующих им неприводимых полиномов $\varphi_i(x)$.

Класс K_0 содержит один элемент, K_5 – два элемента, а классы K_1, K_3 и K_7 – по четыре элемента. Это значит, что неприводимый над полем $GF(2)$ полином, имеющий в качестве корня элемент α^0 поля $GF(2^4)$, должен быть полиномом первой степени, т.е. $\varphi_0(x)=x+1$. Корни α^5 и α^{10} принадлежат неприводимому полиному 2-й степени, который определяется по известному правилу:

$$\varphi_5(x)=(x-\textit{корень}1)(x-\textit{корень}2)=(x+\alpha^5)(x+\alpha^{10})=$$
$$=x^2+x\alpha^5+x\alpha^{10}+\alpha^{15}=x^2+x(x^2+x)+x(x^2+x+1)+1=x^2+x+1.$$

Остальные ненулевые элементы поля $GF(2^4)$ являются корнями неприводимых полиномов $\varphi_1(x)$, $\varphi_3(x)$, $\varphi_7(x)$ четвертой степени, вычисляемых аналогичным образом.

Имея в виду связь между корнями одного полинома, часто об его корнях говорят в единственном числе: «неприводимый полином имеет корень...», понимая под корнем один элемент поля, соответствующий, как правило, младшему из чисел циклотомического ряда, называемому его представителем.

Таблица 5.3

Распределение элементов поля $GF(2^4)$ по циклотомическим классам

Корни $\varphi_i(x)$				Циклотомические классы K_s	Полиномы $\varphi_i(x)$
β_1	$\beta_2=\beta_1^2$	$\beta_3=\beta_1^4$	$\beta_4=\beta_1^8$		
$\alpha^0=1$	$\alpha^0=1$	$\alpha^0=1$	$\alpha^0=1$	$K_0=\{0\}$	$x+1$
α	α^2	α^4	α^8	$K_1=\{1,2,3,4\}$	x^4+x+1
α^3	α^6	α^{12}	$\alpha^{24}=\alpha^9$	$K_3=\{3,6,12,9\}$	$x^4+x^3+x^2+x+1$
α^5	α^{10}	$\alpha^{20}=\alpha^5$	$\alpha^{40}=\alpha^{10}$	$K_5=\{5,10\}$	x^2+x+1
α^7	α^{14}	$\alpha^{28}=\alpha^{13}$	$\alpha^{56}=\alpha^{11}$	$K_7=\{7,14,13,11\}$	x^4+x^3+1

Минимальные многочлены

Рассмотренное распределение элементов конечного поля по циклотомическим классам позволяет лучше понять следующее важное в теории кодирования понятие. Минимальным многочленом или минимальной функцией элемента β поля $GF(p^m)$ называется многочлен $M(x)$ с коэффициентами из $GF(p^m)$ наименьшей степени, для которого β является корнем, т.е. $M(\beta)=0$. Обсудим его основные свойства.

Прежде всего, очевидно, что минимальный многочлен должен быть неприводимым, иначе он раскладывался бы на полиномы меньшей степени.

Любой другой полином, имеющий тот же корень β, что и минимальный, делится на $M(x)$. На $M(x)$ делится и полином $x^{p^m}-1$, т.к. корнями последнего в соответствии с (5.10) будут все ненулевые элементы поля $GF(p^m)$. Степень минимального многочлена определяется количеством компонентов циклотомического класса, которому соответствует его корень (табл. 5.3). Действительно, минимальный многочлен, показатели корней которого принадлежат циклотомическому классу K_s, может быть записан в виде

$$M^{(s)}(x)=(x-\beta_1)(x-\beta_2)\ldots=\prod_i(x-\beta_i)=\prod_{j\in K_s}(x-\alpha^j). \qquad (5.11)$$

Для $s=5$ (см. 5.3.5):

$$M^{(5)}(x)=(x+\alpha^3)(x+\alpha^6)(x+\alpha^{12})(x+\alpha^9)=x^4+x^3+x^2+x+1.$$

Аналогично для $s=3$:

$$M^{(3)}(x)=(x+\alpha^3)(x+\alpha^6)(x+\alpha^{12})(x+\alpha^9)=x^4+x^3+x^2+x+1.$$

С учетом (5.10) справедливо равенство

$$x^{p^m-1}-1=\prod_s M^{(s)}(x),$$

где s пробегает все множество классов по модулю p^m-1, т.е. многочлен $x^{p^m-1}-1$ раскладывается на произведение минимальных многочленов элементов, показатели которых принадлежат каждому из циклотомических классов по модулю p^m-1.

Минимальные многочлены элементов β и β^p равны. В частности, в поле

$GF(2^m)$ равны минимальные многочлены элементов β и β^2. В этом можно убедиться, обратив внимание на тот факт, что элементы β и β^2 всегда соответствуют одному циклотомическому классу (табл. 5.3), а следовательно, принадлежат набору корней одного неприводимого полинома. Более того, между собой равны минимальные многочлены всех элементов, соответствующих одному циклотомическому классу, т.к. любые два соседние из таких элементов находятся в соотношении β и β^2.

И еще об одном свойстве минимального многочлена, имеющем отношение к нахождению примитивных элементов поля. Минимальный многочлен, корнем которого является примитивный элемент поля, называется примитивным многочленом. Его степень всегда равна m. Для практических приложений важно иметь в виду следующее. В тех случаях, когда неприводимый многочлен $p(x)$, задающий операции в поле, является также и примитивным многочленом, примитивным элементом поля будет элемент $\alpha = x$.

Таблицы неприводимых многочленов [33] обычно содержат сведения о том, какие из многочленов являются примитивными, что позволяет избежать возможных затруднений в определении примитивных элементов поля. В табл. 5.4 приведены примитивные многочлены над $GF(2)$ для m от 1 до 20 [3, 30].

Помимо представленных в таблице примитивными являются также полиномы, векторы коэффициентов которых написаны в обратном порядке. Такие полиномы называются двойственными, или взаимными исходным.

Таблица 5.4

Примитивные многочлены до степени $m = 20$

$x+1$	x^6+x+1	$x^{11}+x^2+1$	$x^{16}+x^{12}+x^3+x+1$
x^2+x+1	x^7+x^3+1	$x^{12}+x^6+x^4+x+1$	$x^{17}+x^3+1$
x^3+x+1	$x^8+x^4+x^3+x^2+1$	$x^{13}+x^4+x^3+x+1$	$x^{18}+x^7+1$
x^4+x+1	x^9+x^4+1	$x^{14}+x^9+x^5+x+1$	$x^{19}+x^5+x^2+x+1$
x^5+x^2+1	$x^{10}+x^3+1$	$x^{15}+x+1$	$x^{20}+x^3+1$

Это пары x^3+x^2+1 и x^3+x+1; x^4+x^3+x+1 и x^4+x+1 и т.д. Использовавшийся ранее при построении $GF(2^4)$ полином $p(x)=x^4+x+1$ примитивен, на основании чего в качестве примитивного элемента поля был взят $\alpha=x$.

Изоморфизм конечных полей

Расширение конечного поля $GF(p^m)$ может быть задано разными полиномами одинаковых степеней m. В каком соотношении находятся эти поля? Прежде всего, очевидно, ненулевыми элементами любого поля порядка p^m является тот же полный набор всевозможных многочленов степени $m-1$ и ниже, отличающийся для разных полиномов $p(x)$ лишь порядком следования элементов p по степеням примитивного элемента.

В теории конечных полей доказывается, что все поля $GF(p^m)$ одного порядка p^m изоморфны («подобны по форме»), т.е. между $GF_1(p^m)$ и $GF_2(p^m)$ существует взаимнооднозначное отображение f друг на друга, сохраняющее операции сложения и умножения. Это означает, что для любых двух элементов β_i и β_j из $GF_1(p^m)$ справедливы соотношения

$$f(\beta_i+\beta_j)=f(\beta_i)+f(\beta_j), \tag{5.12}$$

$$f(\beta_i\beta_j)=f(\beta_i)f(\beta_j). \tag{5.13}$$

Нетрудно убедиться, что между полями, построенными на основе неприводимых полиномов $p_1(x)=x^4+x+1$ и $p_2(x)=x^4+x^3+1$ (табл. 5.2), существует взаимнооднозначное отображение: $\alpha=f(\gamma)=\gamma^7=x^4+x^3+1$. Простой подстановкой можно убедиться, что при таком отображении сохраняются операции сложения и умножения (5.12) и (5.13). Например, для сложения

$$f(\beta_4+\beta_7)=f(\alpha^4+\alpha^7)=f(\alpha^3)=\gamma^{21}=\gamma^6=\alpha^3=\beta_3,$$

$$f(\beta_4)+f(\beta_7)=f(\alpha^4)+f(\alpha^7)=\gamma^{28}+\gamma^{49}=\gamma^{13}+\gamma^4=\gamma^6=\beta_3.$$

Аналогично для умножения

$$f(\beta_4\beta_7)=f(\alpha^4\alpha^7)=f(\alpha^{11})=\gamma^2=\alpha^{11}=\beta_{11},$$

$$f(\beta_4)(\beta_7)=f(\alpha^4)f(\alpha^7)=\gamma^{28}\gamma^{49}=\gamma^{13}\gamma^4=\gamma^{17}=\gamma^2=\beta_{11}.$$

5.4. Линейные блочные коды

5.4.1. Система передачи дискретных сообщений

При передаче информации по каналам связи возможны ошибки вследствие помех и искажений сигналов. Для обнаружения и исправления возникающих ошибок используются помехоустойчивые коды. Упрощенная схема системы передачи информации при помехоустойчивом кодировании показана на рис. 5.3.

Кодер служит для преобразования поступающей от источника сообщений последовательности из k информационных символов в последовательность из n символов кодовых комбинаций (или кодовых слов). Совокупность кодовых слов образует код.

Множество символов, из которых составляется кодовое слово, называется алфавитом кода, а число различных символов в алфавите – основанием кода. В дальнейшем вследствие их простоты и наибольшего распространения рассматриваются главным образом двоичные коды, алфавит которых содержит два символа: 0 и 1.

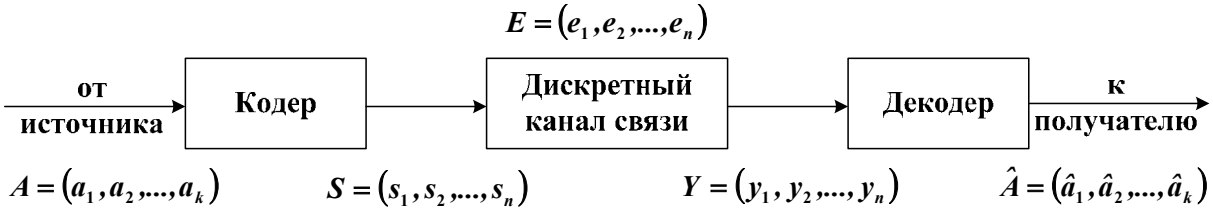


Рис. 5.3. Система передачи дискретных сообщений

Правило, по которому информационной последовательности сопоставляется кодовое слово, называется правилом кодирования. Если при кодировании каждый раз формируется блок A из k информационных символов, превращаемый затем в n-символьную кодовую комбинацию S, то код называется блочным. При другом способе кодирования информационная последовательность на блоки не разбивается, и код называется непрерывным.

С математической точки зрения кодер осуществляет отображение множества из 2^k элементов (двоичных информационных последовательностей) в

множество, состоящее из 2^n элементов (двоичных последовательностей длины n). Для практики интересны такие отображения, в результате которых получаются коды, обладающие способностью исправлять часть ошибок и допускающие простую техническую реализацию кодирующих и декодирующих устройств.

Дискретный канал связи – это совокупность технических средств вместе со средой распространения радиосигналов, включенных между кодером и декодером для передачи сигналов, принимающих конечное число разных видов. Для описания реальных каналов предложено много математических моделей, с разной степенью детализации отражающих реальные процессы. Ограничимся рассмотрением простейшей модели двоичного канала, входные и выходные сигналы которого могут принимать значения 0 и 1.

Наиболее распространено предположение о действии в канале аддитивной помехи. Пусть $S=(s_1,s_2,...,s_n)$ и $Y=(y_1,y_2,...,y_n)$ соответственно входная и выходная последовательности двоичных символов. Помехой или вектором ошибки называется последовательность из n символов $E=(e_1,e_2,...,e_n)$, которую надо поразрядно сложить с переданной последовательностью, чтобы получить принятую:

$$Y=S+E\,. \qquad (5.14)$$

Таким образом, компонента вектора ошибки $e_i=0$ указывает на то, что 2-й символ принят правильно ($y_i=s_i$), а компонента $e_i=1$ указывает на ошибку при приеме ($y_i\neq s_i$).Поэтому важной характеристикой вектора ошибки является число q ненулевых компонентов, которое называется весом или кратностью ошибки. Кратность ошибки – дискретная случайная величина, принимающая целочисленные значения от 0 до n.

Классификация двоичных каналов ведется по виду распределения случайного вектора E. Основные результаты теории кодирования получены в предположении, что вероятность ошибки в одном символе не зависит ни от его номера в последовательности, ни от его значения. Такой канал называется

стационарным и симметричным. В этом канале передаваемые символы искажаются с одинаковой вероятностью p, т.е. $p(e_i = 1) = p$, $i = 1,2,...,n$.

Для симметричного стационарного канала распределение вероятностей векторов ошибки кратности q является биноминальным:

$$p_n(q) = C_n^q \cdot p^q (1-p)^{n-q},$$

где C_n^q – число сочетаний из n элементов по q.

Вероятность искажения конкретных q символов (или вероятность появления одной конфигурации E_i вектора ошибки веса q) определяется по формуле

$$p(E_i) = p^q (1-p)^{n-q},$$

которая показывает, что при $p < 0{,}5$ вероятность $\beta_2 = \alpha^j$ является убывающей функцией q, т.е. в симметричном стационарном канале более вероятны ошибки меньшей кратности. Этот важный факт используется при построении помехоустойчивых кодов, т.к. позволяет обосновать тактику обнаружения и исправления в первую очередь ошибок малой кратности. Конечно, для других моделей канала такая тактика может и не быть оптимальной.

Декодирующее устройство (декодер) предназначено оценить по принятой последовательности $Y = (y_1, y_2, ..., y_n)$ значения информационных символов $\hat{A} = (\hat{a}_1, \hat{a}_2, ..., \hat{a}_k)$. Из-за действия помех возможны неправильные решения. Процедура декодирования включает решение двух задач: оценивание переданного кодового слова и формирование оценок информационных символов.

Вторая задача решается относительно просто. При наиболее часто используемых систематических кодах, кодовые слова которых содержат информационные символы на известных позициях, все сводится к простому их стробированию. Очевидно также, что расположение информационных символов внутри кодового слова не имеет существенного значения. Удобно считать, что они занимают первые k позиций кодового слова.

Наибольшую трудность представляет первая задача декодирования. При

равновероятных информационных последовательностях ее оптимальное решение дает метод максимального правдоподобия. Функция правдоподобия как вероятность получения данного вектора Y при передаче кодовых слов s_i, $i = 1,2,...,2^k$ на основании (5.14) определяется вероятностями появления векторов ошибок:

$$p\left(Y/S_i\right) = p\left(E_i\right) = p^{q_i}(1-p)^{n-q_i},$$

где q_i – вес вектора $E_i = Y + S_i$

Очевидно, вероятность $p\left(Y/S_i\right)$ максимальна при минимальном q_i. На основании принципа максимального правдоподобия оценкой $\hat{S}$ является кодовое слово, искажение которого для превращения его в принятое слово Y имеет минимальный вес, т. е. в симметричном канале является наиболее вероятным (НВ):

$$\hat{S} = Y + E_{HB}.$$

Если несколько векторов ошибок E_i имеют равные минимальные веса, то наивероятнейшая ошибка E_{HB} определяется случайным выбором среди них.

В качестве расстояния между двумя кодовыми комбинациями принимают так называемое расстояние Хэмминга, которое численно равно количеству символов, в которых одна комбинация отлична от другой, т.е. весу (числу ненулевых компонентов) разностного вектора. Расстояние Хэмминга между принятой последовательностью Y и всеми возможными кодовыми словами 5, есть функция весов векторов ошибок E_i:

$$d\left(S_i, Y_i\right) = \sum_{j=1}^{n}\left(s_{i,j} + y_i\right) = \sum_{j=1}^{n} e_{i,j} = q_i.$$

Поэтому декодирование по минимуму расстояния, когда в качестве оценки берется слово, ближайшее к принятой последовательности, является декодированием по максимуму правдоподобия.

Таким образом, оптимальная процедура декодирования для симметрично-

го канала может быть описана следующей последовательностью операций. По принятому вектору Y определяется вектор ошибки с минимальным весом $E_{НВ}$, который затем вычитается (в двоичном канале - складывается по модулю 2) из Y:

$$Y \to E_{НВ} \to S \to \hat{C} = Y + E_{НВ}.$$

Наиболее трудоемкой операцией в этой схеме является определение наивероятнейшего вектора ошибки, сложность которой существенно возрастает при увеличении длины кодовых комбинаций. Правила кодирования, которые нацелены на упрощение процедур декодирования, предполагают придание всем кодовым словам технически легко проверяемых признаков.

Широко распространены линейные коды, называемые так потому, что их кодовые слова образуют линейное подпространство над конечным полем. Для двоичных кодов естественно использовать поле характеристики $p = 2$. Принадлежность принятой комбинации Y известному подпространству является тем признаком, по которому выносится решение об отсутствии ошибок ($E_{НВ} = 0$).

Так как по данному коду все пространство последовательностей длины n разбивается на смежные классы (см. 5.3.2), то для каждого смежного класса можно заранее определить вектор ошибки минимального веса, называемый лидером смежного класса. Тогда задача декодера состоит в определении номера смежного класса, которому принадлежит Y, и формировании лидера этого класса.

5.4.2. Параметры линейного кода

Помехоустойчивое кодирование сообщений дискретного источника информации [25, 33] заключается в том, что поступающие k-символьные информационные комбинации $A = (a_1, a_2, ..., a_k)$ дополняются $n - k$ избыточными символами до n-символьных кодовых комбинаций $S = (s_1, s_2, ..., s_n)$. В процессе передачи последних по каналу связи под действием помех отдельные символы кодовой комбинации искажаются и трансформируются на приемной стороне в другие

символы из используемого для передачи алфавита.

Наиболее употребимы двоичные линейные коды. Такой код определяется как множество из 2^k кодовых n-символьных комбинаций, образующих линейное подпространство размерности k.

Линейные коды обозначаются (n,k,d_0). Здесь n – длина кода, число символов в кодовых словах или размерность пространства кодовых комбинаций; k – число информационных символов или размерность кода; $n-k$ – количество проверочных или избыточных символов. Числа n и k определяют относительную скорость передачи информации кодом, равную k/n двоичных единиц на 1 символ кодовой комбинации.

Третий параметр линейного кода – кодовое расстояние d_0 характеризует корректирующую способность помехоустойчивого кода и вводится как минимальное из расстояний Хэмминга (см. 5.4.1) при попарном сравнении кодовых слов. С кодовым расстоянием связаны кратности обнаруживаемых $q_{об}$ и исправляемых $q_{ис}$ ошибок, произошедших в пределах одной кодовой комбинации:

$$d_0 \geq q_{об} + 1 \text{ или } d_0 \geq 2q_{ис} + 1.$$

Число $q_{об}$ указывает, что код способен обнаруживать все конфигурации вектора ошибки, вес которых $q \leq q_{об}$. Число $q_{ис}$ указывает, что код способен исправлять все конфигурации вектора ошибки, вес которых $q \leq q_{ис}$.

При совмещении процедур обнаружения и исправления ошибок, причем $q_{об} > q_{ис}$ соотношение между d_0, $q_{об}$ и $q_{ис}$ имеет вид:

$$d_0 \geq q_{об} + q_{ис} + 1.$$

При фиксированных n и k большей помехоустойчивостью обладают коды с большим кодовым расстоянием. Линейные коды достаточно хорошо изучены и сведены в таблицы [30, 33].

5.4.3. Полиномиальные циклические коды

Весьма плодотворным оказалось представление n-символьных комбинаций линейных кодов в виде полиномов степени $n-1$. Для этого перенумеруем символы в кодовых комбинациях так, чтобы каждому кодовому слову $S=(s_{n-1},s_{n-2},...,s_0)$ можно было сопоставить многочлен $S=s_{n-1}x^{n-1}+s_{n-2}x^{n-2}+...+s_0$, и символы слова являлись коэффициентами многочлена. Например, кодовому слову 100101 соответствует многочлен x^5+x+1. Такой код, получивший название полиномиального, можно определить как множество всех многочленов степени $n-1$, содержащих в качестве множителя некоторый многочлен $g(x)$, называемый порождающим многочленом кода. Иными словами, полиномиальный код (n,k,d_0) есть множество всех многочленов степени $n-1$ или меньше, делящихся на $g(x)$, т.е. $S(x)=g(x)A(x)$. Степень полинома $g(x)$ равна $n-k$, а количество символов в соответствующем ему кодовом слове – $(n-k+1)$.

Например, $g(x)=x^3+x+1\leftrightarrow 1011$ порождает множество полиномов вида

$$S(x)=g(x)A(x)=(x^3+x+1)(a_3x^3+a_2x^2+a_1x+a_0)=a_3x^6+a_2x^5+$$
$$+(a_1+a_3)x^4+(a_0+a_2+a_3)x^3+(a_1+a_2)x^2+(a_0+a_1)x+a_0,$$

т.е. код (7, 4, 3).

Естественно в качестве $A(x)$ взять полином, соответствующий информационной комбинации $A=(a_1,a_2,...,a_k)$, т.е. принять a_i в качестве информационных символов. Однако недостаток такого способа кодирования заключается в том, что получаемый код оказывается несистематическим, т.е. не имеет четкого разделения разрядов на информационные и избыточные.

На практике используется другой способ получения полиномов $S(x)$ по информационному полиному $A(x)$ и порождающему полиному $g(x)$ [25, 33]:

$$S(x)=A(x)x^{n-k}+R(x), \qquad (5.15)$$

где $R(x)$ – остаток от деления полинома $A(x)x^{n-k}$ на $g(x)$. В этом случае $S(x)$ также делится на $g(x)$ без остатка, а следовательно, S является кодовым словом систематического кода (n,k,d_0). Полином $R(x)$ соответствует комбинации $B=(b_1,b_2,...,b_{n-k})$ символов, называемых проверочными.

Связь между структурой $g(x)$ и кодовым расстоянием d_0 порождаемого им кода достаточно проста. Для того чтобы задаваемый полиномом $g(x)$ код имел кодовое расстояние d_0, количество отличных от нуля коэффициентов $g(x)$ должно быть не менее d_0. Действительно, с одной стороны, нулевая кодовая комбинация, соответствующая $S(x) \equiv 0$, принадлежит любому линейному коду, в том числе и задаваемому полиномом $g(x)$, так как при $A(x) = 0$ остаток $R(x)$ от деления $A(x)x^{n-k}$ на $g(x)$ также равен нулю. С другой стороны, кодовая комбинация вида $0\ 0...g(x)...0$, т.е. содержащая комбинацию, соответствующую $g(x)$ в $(n-k+1)$ смежных разрядах и нули в остальных, делится на $g(x)$ и тоже принадлежит порождаемому им коду. Эти комбинации должны отличаться не менее чем в d_0 разрядах, а значит, $g(x)$ должен иметь не менее d_0 единиц.

Весьма важный в практическом отношении подкласс полиномиальных кодов составляют циклические коды, обладающие тем свойством, что циклическая перестановка символов в кодовом слове дает другое слово, но того же кода. Порождающий полином $g(x)$ циклического кода должен быть делителем многочлена $x^n + 1$. Убедимся в этом.

Пусть кодовое слово S_2 получено из S_1 путем циклического сдвига символов влево, а самый левый символ переведен на освободившееся место на правом конце. Если S_1 имеет в качестве левого символа 0, то циклический сдвиг эквивалентен только умножению $S_1(x)$ на x: $S_2(x) = xS_1(x)$. Очевидно, если $S_1(x)$ делится на $g(x)$, то и $S_2(x)$ также делится на $g(x)$.

Если левый символ S_1 равен 1, то циклический сдвиг влево эквивалентен умножению $S_1(x)$ на x, вычитанию из произведения члена x^n и добавлению единицы в младший разряд: $S_2(x) = xS_1(x) - x^n + 1 = xS_1(x) - (x^n - 1)$.

Ясно, что для делимости $S_2(x)$ на $g(x)$ необходимо, чтобы многочлен $x^n + 1$ также делился на $g(x)$ (сложение и вычитание здесь эквивалентны).

В табл. 5.5 приведено разложение полинома $x^n + 1$ над $GF(2)$ на неприводимые многочлены для значений $n \le 31$. Более обширные сведения за-

интересованный читатель сможет найти в [30]. Таблица позволяет выбирать порождающие полиномы $g(x)$ в зависимости от числа n символов в коде и k информационных символов, поскольку степень $g(x)$ равна $n-k$.

Любой делитель полинома x^n+1 или любое произведение делителей может быть взято в качестве порождающего полинома хода. Например, при $n=15$ $g(x)=x^4+x+1$ и $g(x)=x^4+x^3+1$ порождают код (15,11,3) – код Хэмминга с кодовым расстоянием $d_0=3$; код (15,7,5) порождается полиномом $g(x)=(x^4+x+1)(x^4+x^3+x^2+x+1)$; код (15,5,7) – полиномом $g(x)=(x^2+x+1)(x^4+x+1)(x^4+x^3+x^2+x+1)$; код (15,4,8) – полиномом $g(x)=(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+x^2+x+1)$ и т.д.

Для $n=3,5,11,13,19,29$ разложение имеет вид: $x^n+1=(x+1)(x^{n-1}+x^{n-2}+\ldots+x+1)$, что определяет лишь два типа кодов этих длин.

1. Коды $(n,n-1,2)$ с простой проверкой на четность, кодовым расстоянием $d_0=2$ и порождающим полиномом $g(x)=x+1$, позволяющим лишь обнаруживать ошибки нечетной кратности.

2. Коды $(n,1,n)$ нечетной длины с повторением, $d_0=n$, $g(x)=x^{n-1}+x^{n-2}+\ldots+x+1$.

В таблице отсутствуют разложения для четных n. Но поскольку $(x^m+1)(x^m+1)=x^{2m}+x^m+x^m+1=(x^{2m}+1)$, то при четном n полином $(x^n+1)=(x^{n/2}+1)(x^{n/2}+1)$ и может быть сведен к произведению полиномов нечетных степеней. Например, $(x^6+1)=(x^3+1)(x^3+1)=(x+1)(x^2+x+1)(x+1)(x^2+x+1)$.

Неприводимые полиномы, являющиеся примитивными, подчеркнуты в табл. 5.5.

Полином $h(x)$, удовлетворяющий равенству $(x^n+1)=g(x)h(x)$, называется проверочным полиномом и тоже может быть использован для формирования кода.

Таблица 5.5

Разложение полинома x^n+1 на делители

n	Делители полинома x^n+1
7	$(x+1)(x^3+x+1)(x^3+x^2+1)$
9	$(x+1)(x^2+x+1)(x^6+x^3+1)$
15	$(x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$
17	$(x+1)(x^8+x^5+x^4+x^3+x+1)(x^8+x^7+x^6+x^4+x^2+x+1)$
21	$(x+1)(x^2+x+1)(x^3+x+1)(x^3+x^2+1)(x^6+x^4+x^2+x+1)(x^6+x^5+x^4+x^2+1)$
23	$(x+1)(x^{11}+x^{10}+x^6+x^5+x^4+x^2+1)(x^{11}+x^9+x^7+x^6+x^5+x+1)$
25	$(x+1)(x^4+x^3+x^2+x+1)(x^{20}+x^{15}+x^{10}+x^5+1)$
27	$(x+1)(x^2+x+1)(x^6+x^3+1)(x^{18}+x^9+1)$
31	$(x+1)(x^5+x^3+1)(x^5+x^2+1)(x^5+x^4+x^3+x^2+1)(x^5+x^4+x^3+x+1)\times$ $\times(x^5+x^4+x^2+x+1)(x^5+x^3+x^2+x+1)$

Еще один способ задания кода основан на использовании порождающей или проверочной матриц. Порождающая матрица G имеет k строк и n столбцов, содержит k базисных линейно независимых кодовых комбинаций. Наиболее удобна для пользования каноническая форма порождающей матрицы. Ее строки в своей информационной части образуют квадратную $k\times k$ единичную матрицу.

$$G=\begin{vmatrix}S_{\delta 1}\\S_{\delta 2}\\\dots\\S_{\delta k}\end{vmatrix}=\begin{vmatrix}1&0&\dots&0&b_{11}&b_{12}&\dots&b_{1n-k}\\0&1&\dots&0&b_{21}&b_{22}&\dots&b_{2n-k}\\\dots&\dots&\dots&\dots&\dots&\dots&\dots&\dots\\0&0&\dots&1&b_{k1}&b_{k2}&\dots&b_{kn-k}\end{vmatrix}. \qquad (5.16)$$

Любая другая комбинация кода может быть получена как взвешенная (с весами a_i от источника сообщений) сумма строк порождающей матрицы:

$$S=a_1S_{\delta 1}+a_2S_{\delta 2}+\dots+a_kS_{\delta k}.$$

соответствует произведению матриц

$$S=AG.$$

Так, для двоичного кода (7, 4, 3) порождающая матрица имеет вид:

$$G_{7\times4} = \begin{vmatrix} 1000101 \\ 0100111 \\ 0010110 \\ 0001011 \end{vmatrix}.$$

Очевидна связь между порождающим полиномом $g(x)$ и строками матрицы G. Избыточные символы b_i первой строки в соответствии с (5.15) есть остаток от деления n-символьной комбинации 100 ... 0 на $g(x)$, второй –остаток от деления комбинаций 010 ... 0 и т.д.

Проверочная матрица H имеет $n-k$ строк и n столбцов и связана с порождающей матрицей уравнением

$$HG^T = 0,$$

где T – символ транспонирования. Эквивалентное уравнение имеет вид:

$$GH^T = 0.$$

Для двоичного кода (7, 4, 3)

$$H = \begin{vmatrix} 1110100 \\ 0111010 \\ 1101001 \end{vmatrix}. \qquad (5.17)$$

Кодирование с помощью проверочной матрицы производится на основе уравнений:

$$SH^T = 0 \text{ или } b_j = \sum_{i=1}^{k} a_i h_{ij},\ j = 1,2,...,n-k,$$

где h_{ij} – элемент i-й строки и j-го столбца матрицы H.

Обнаружение и исправление на приемной стороне системы передачи информации части ошибок, произошедших в кодовых комбинациях в процессе передачи, составляет сущность декодирования. При сравнительно небольшой длине кодовых слов распространены так называемые неалгебраические методы декодирования [26]. Они основаны на анализе остатков от деления принятых комбинаций на порождающий полином $g(x)$ или реализации системы оценок каждого из принятых символов, построенной на основе порождающей или проверочной матриц.

Вопросы организации кодеров и декодеров в зависимости от структуры

полиномов $g(x)$ и $h(x)$, а также матриц G и H достаточно подробно рассмотрены в монографиях [3, 26]. Однако при больших длинах используемых кодов подобные методы декодирования реализуются слишком громоздко. Кроме того, мажоритарные или пороговые декодеры, основывающиеся на анализе системы оценок каждого символа принятой комбинации, имеют тот недостаток, что необходимую систему оценок удается составить не для всех кодов.

Дальнейшим развитием теории и практики помехоустойчивого приема сигналов являются алгебраические методы декодирования, использующие понятие корней полиномов, соответствующих кодовым комбинациям. Алгебраические процедуры эффективны, в частности, при декодировании кодов БЧХ.

5.4.4. Циклические коды и корни полиномов

Поскольку многочлен $S(x)$ каждого кодового слова делится на порождающий полином $g(x)$, то корни $g(x)$, при подстановке в $g(x)$ обращающие его в 0, являются также и корнями многочлена $S(x)$. Число таких корней равно степени порождающего полинома $n-k$.

Следовательно, многочлен $S(x)$ с коэффициентами из поля $GF(p)$ будет кодовым словом в том и только в том случае, если элементы $\beta_1, \beta_2, ..., \beta_{n-k}$ из расширения $GF(p^m)$ являются его корнями.

Установим связь элементов проверочной матрицы H с корнями порождающего полинома $g(x)$, а также обсудим, как находить сами корни и, наоборот, по заданным корням определять порождающий полином.

Заметим прежде, что до сих пор использовалась запись полиномов с расположенной слева старшей степенью переменной x. Такая форма записи была удобна для выполнения простейших алгебраических действий над полиномами: сложения, умножения, деления полиномов. Однако при предстоящем рассмотрении алгебраических процедур декодирования предпочтительна обратная запись многочленов, начиная с нулевой степени переменной.

Условие $S(\beta_j)=0$ в развернутом виде означает

$$s_0\beta_j^0 + s_1\beta_j^1 + \ldots + s_{n-2}\beta_j^{n-2} + s_{n-1}\beta_j^{n-1} = \sum_{i=1}^{n-1} s_j\beta_j^i = 0,\ j = \overline{1, n-k}\ . \qquad (5.18)$$

Такая запись эквивалентна матричной

$$SH^T = 0, \qquad (5.19)$$

Где

$$H = \begin{vmatrix} \beta_1^0 & \beta_1 & \beta_1^2 & \ldots & \beta_1^{n-1} \\ \beta_2^0 & \beta_2 & \beta_2^2 & \ldots & \beta_2^{n-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \beta_{n-k}^0 & \beta_{n-k} & \beta_{n-k}^2 & \ldots & \beta_{n-k}^{n-1} \end{vmatrix}.$$

Выражение (5.19) является определением проверочной матрицы H [3]. Строки H содержат степени корней порождающего полинома $g(x)$, а следовательно, и корней кодовых многочленов. Каждый элемент H есть m-символьный столбец p-ичного представления корня – элемента поля $GF(p^m)$.

В общем случае матрица H может содержать ряд строк, функционально связанных между собой и, следовательно, выражаемых друг через друга. Таковыми являются строки, соответствующие множеству $\beta, \beta^p, \beta^{p^2}, \beta^{p^3}, \ldots$ корней одного циклотомического класса показателей (см. 5.3.5). Функциональная зависимость подобных строк вытекает из свойства многочленов над конечным полем $GF(p)$: $S(\beta^p) = S^p(\beta)$.

Из каждого такого множества корней $\beta, \beta^p, \beta^{p^2}, \beta^{p^3}, \ldots$ следует выбрать по одному (любому), и соответствующие им строки оставить в матрице, а остальные удалить как функционально зависимые от удерживаемых.

Дальнейшее обсуждение проведем для многочленов над полем $GF(2)$. Рассмотрим несколько частных случаев.

1. Порождающий полином $g(x)$ неприводим и примитивен, а длина кода $n = 2^m - 1$. Коды таких длин называют примитивными независимо от вида порождающего полинома.

Очевидно, в этом случае $g(x)$ может быть использован в качестве полинома, задающего поле $GF(2^m)$. Корнем его является примитивный элемент поля α, а показатели всех корней принадлежат одному циклотомическому классу.

Проверочная матрица такого кода имеет вид

$$H = \left[\alpha^0 \alpha^1 \alpha^2 ... \alpha^{n-1}\right],$$

а сам код носит название кода Хэмминга.

В качестве примера выберем код (7, 4, 3), порождающий полином которого $g(x) = x^3 + x + 1$ примитивен, а проверочная матрица

$$H = \left[1 \alpha \alpha^2 ... \alpha^6\right] = \begin{vmatrix} 0010111 \\ 0101110 \\ 1001011 \end{vmatrix},$$

с точностью до порядка следования столбцов совпадает с ранее полученной проверочной матрицей (5.17) этого же кода. Смена порядка на обратный объясняется изменением в (5.18) порядка записи степеней переменной x.

2. Порождающий полином $g(x)$ неприводим, но не является примитивным, длина кода $n = 2^m - 1$. В этом случае элемент α поля $GF(2^m)$ не обязательно является его корнем, и при отыскании корня возникают дополнительные трудности. Таков, например, полином $g(x) = x^6 + x^3 + 1$ для кода с $n = 9$ (табл. 5.5). Следует прибегнуть к таблицам неприводимых полиномов.

В табл. 5.6 представлены полиномы степеней до 8, заимствованные из [33], где можно найти сведения о полиномах до 34-й степени.

Неприводимые многочлены даны в восьмеричном представлении. Каждую цифру соответствующего многочлену числа следует перевести в трехразрядное двоичное число и рассматривать его разряды как коэффициенты при степенях x многочлена. Например, среди полиномов степени 5 есть восьмеричное число 67. Это означает: $67 = 110111 \leftrightarrow x^5 + x^4 + x^2 + x + 1$.

Примитивные многочлены подчеркнуты. Первым среди многочленов данной степени помещен примитивный, и с его помощью строится поле $GF(2^m)$. Перед восьмеричным представлением стоит десятичное число, отделенное точкой и означающее показатель степени примитивного элемента, являющегося корнем следующего за ним полинома, причем большие степени записываются слева. Например, для нахождения корня полинома 6-й степени 127 нужно задать с помощью примитивного полинома 103 поле $GF(2^6)$ и в

Таблица 5.6

Неприводимые полиномы над полем $GF(2)$

Степень m полинома	Восьмеричное представление полиномов				
2	1.7				
3	1.13				
4	1.23	3.37	5.007		
5	1.45	3.75	5.67		
6	1.103	3.127	5.147	7.111	9.115
	11.155	21.007			
7	1.211	3.217	5.235	7.365	9.277
	11.325	13.203	19.303	21.345	
8	1.435	3.567	5.763	7.551	9.765
	11.747	13.453	15.727	17.023	19.545
	21.613	23.543	25.433	27.477	37.537
	43.703	45.471	51.037	85.007	

качестве корня β полинома 127 взять элемент α^3 этого поля. Каждый из многочленов табл. 5.6 является минимальным многочленом указанного перед ним корня.

Многочлен, двойственный неприводимому, также неприводим, а двойственный примитивному – примитивен. Двойственные многочлены не помещены в табл. 5.6, однако могут быть получены из представленных в ней переменой порядка следования коэффициентов при степенях x на обратный. Так, полином, двойственный 67, имеет вид: $111011 \leftrightarrow x^5 + x^4 + x^3 + x + 1$.

Показатель i степени примитивного элемента поля, соответствующий корню двойственного многочлена, определяется как $i = 2^m - 1 - j$, где j-степень примитивного элемента – корня приведенного в табл. 5.6 полинома. Поскольку $2^m - 1$ и j нечетны, то i всегда четно. Например, корнем полинома,

двойственного полиному 23, является элемент $\beta = \alpha^{14}$, так как $i = 15 - 1 = 14$, двойственного полиному 37 – элемент $\beta = \alpha^{12}$, и т.д.

Минимальный многочлен элемента α^j включен в таблицу, даже если степень многочлена меньше m (т.е. показатель j принадлежит циклотомическому классу, содержащему меньшее m число компонентов, например, классу K_5 (табл. 5.3)). Таким минимальным многочленом корня α^5 в поле $GF(2^4)$ является 007 второй степени. В табл. 5.6 подобные многочлены начинаются с нуля.

3. Порождающий полином $g(x)$ представляет собой произведение нескольких неприводимых многочленов, каждый из которых имеет корни в $GF(2^m)$, и по-прежнему $n = 2^m - 1$. Например, $g(x) = (x^2 + x + 1)(x^4 + x + 1) = x^6 + x^5 + x^4 + x^3 + 1$ (табл. 5.5) порождает код (15, 9, 3).

Очевидно, в этом случае среди $n - k$ корней полинома $g(x)$ будут корни каждого из сомножителей, отыскиваемые изложенным выше способом.

4. Общий случай: порождающий полином $g(x)$ либо неприводим, либо является произведением неприводимых многочленов; длина кода $n \neq 2^m - 1$.

Корни многочлена $x^n + 1$ для таких значений n являются элементами некоторого поля $GF(2^l)$, причем показатели этих элементов принадлежат одному циклотомическому классу K_5 по модулю $(2^l - 1)$. Полином $x^{2^l - 1} + 1$ имеет корнями все элементы поля $GF(2^l)$, в том числе и корни полинома $x^n + 1$. Следовательно, $x^{2^l - 1} + 1$ делится на $x^n + 1$, а $2^l - 1$ делится на n (см. 5.3.5), т.е. $2^l - 1 = nr$, где r – целое число, причем одному значению n может соответствовать множество пар чисел (l, r). Например, при $n = 9$ справедливы равенства $2^6 - 1 = 9 \cdot 7$; $2^{12} - 1 = 9 \cdot 455$ и т.д.

Поскольку корни $x^n + 1$ лежат в поле $GF(2^l)$, то корни неприводимых полиномов – делителей двучлена $x^n + 1$ также принадлежат этому полю, и их следовало бы обозначать показателями j степени примитивного элемента поля $GF(2^l)$, как это делалось в табл. 5.6. По так как одному значению и может отвечать несколько полей $GF(2^l)$, в перечнях кодов [33] корни помечаются

отношением j/l .

Табл. 5.7 содержит заимствованные из [30] двоичные циклические коды (n,k,d_0) нечетной длины до $n=23$ с указанием нормированных показателей j/l корней. Здесь же приведены значения конструктивных расстояний δ, необходимые для изложения материала следующих разделов.

Таблица 5.7

Нормированные показатели корней

n	k	d_0	δ	R	n	k	d_0	δ	R
7	4	3	3	1	15	4	8	8	0,1,3,5
7	3	4	4	0,1	15	3	5	5	1,3,7
9	3	3	3	1	15	2	10	10	0,1,3,7
9	2	6	6	0,1	17	9	5	4	1
15	11	3	3	1	17	8	6	6	0,1
15	10	4	4	0,1	21	16	3	3	7,3
15	9	3	3	1,5	21	15	4	3	0,7,3
15	9	4	3	3,5	21	14	4	4	0,1
15	8	4	4	0,1,5	21	13	4	3	7,9,3
15	8	4	3	0,3,5	21	12	4	3	0,7,9,3
15	7	3	3	1,7	21	11	6	6	0,3,1
15	7	5	5	1,3	21	10	5	5	7,3,1
15	6	6	6	0,1,3	21	9	5	6	9,3,1
15	6	6	6	0,1,7	21	8	6	6	0,9,3,1
15	5	3	3	1,5,7	21	6	7	7	3,5,1
15	5	7	7	1,3,5	23	12	7	5	1
15	4	6	6	0,1,5,7	23	11	8	6	0,1

Например, для кода (9, 3, 3) величина $j/l=1$. Это означает, что корнем порождающего полинома данного кода в поле $GF(2^6)$ является элемент α^7, в поле $GF(2^{12})$ – элемент α^{455} и т.д.

В заключение этого раздела обсудим задачу, обратную рассмотренной: по заданным корням требуется построить порождающий полином и код. Такая постановка характерна при формировании БЧХ-кодов.

Пусть $\beta_1, \beta_2, ..., \beta_r$ – элементы поля $GF(2^m)$, являющиеся заданными корнями, а $M_1(x), M_2(x), ..., M_r(x)$ – соответствующие им минимальные многочлены. Каждый из корней является некоторой степенью примитивного элемента поля. Если все показатели степени принадлежат разным циклотомическим классам по модулю $(2^m - 1)$, то в соответствии с (5.9), (5.10) и (5.11) порождающий полином

$$g(x) = \prod_{i=1}^{r} M_i \tag{5.20}$$

В общем случае, когда некоторые из заданных корней могут принадлежать одному циклотомическому классу, т.е. находиться между собой в соотношении $\beta, \beta^2, \beta^4, ...$, порождающий полином (5.20)

$$g(x) = HOK[M_1(x), M_2(x), ..., M_r(x)]. \tag{5.21}$$

5.4.5. Спектральное описание циклических кодов

Рассмотрим еще один подход к описанию полиномиальных кодов, который основан на использовании дискретного преобразования Фурье (ДПФ) кодовых последовательностей, заданных над конечным полем $GF(p)$. Данный подход, подробно изложенный в [30], позволяет в ряде случаев упростить процедуры кодирования и декодирования.

Пусть $V = (v_0, v_1, ..., v_{n-1})$ – последовательность из n элементов конечного поля $GF(p)$, причем n делит $p^m - 1$ для некоторого m, и пусть α – примитивный элемент порядка n в расширении поля $GF(p^m)$. Дискретным преобразованием Фурье вектора V над конечным полем $GF(p)$ называется последовательность $F^{(v)} = (f_0^{(v)}, f_1^{(v)}, ..., f_{n-1}^{(v)})$ элементов поля $GF(p^m)$ задаваемая равенством

$$f_j^{(v)} = \sum_{i=0}^{n-1} v_i \alpha^{ij}, \quad j = 0,1,...,n-1. \tag{5.22}$$

В матричной форме ДПФ может быть записано следующим образом

$$F^{(v)} = (v_0 v_1 ... v_{n-1}) \begin{vmatrix} 1 & 1 & 1 & ... & 1 \\ 1 & \alpha & \alpha^2 & ... & \alpha^{n-1} \\ ... & ... & ... & ... & ... \\ 1 & \alpha^{n-1} & \alpha^{2(n-1)} & ... & \alpha^{(n-1)^2} \end{vmatrix}.$$

Такое определение аналогично определению ДПФ в поле комплексных чисел, где α заменяется на корень n-й степени из единицы, равный $\exp\left(-j2\pi/n\right)$ В связи с такой аналогией оказывается удобным называть индекс i «дискретным временем», а последовательность $v_0, v_1, ..., v_{n-1}$ – временной последовательностью (функцией). Тогда индекс j можно назвать «частотой», а последовательность $f_0^{(v)}, f_1^{(v)}, ..., f_{n-1}^{(v)}$ – частотным спектром или просто спектром.

Если векторы V и $F^{(v)}$ связаны равенством (5.22), то существует обратное преобразование Фурье

$$v_i = \frac{1}{n} \sum_{j=0}^{n-1} f_j^{(v)} \alpha^{-ij} . \qquad (5.23)$$

Равенства (5.22) и (5.23) часто называют парой преобразований Фурье. Укажем на два наиболее важных свойства ДПФ.

1. Пусть U, V и W – временные последовательности, причем $w_i = u_i v_i$, $i = 0,1,...,n-1$. Тогда

$$f_j^{(v)} = \sum_{k=0}^{n-1} f_k^{(u)} f_{n-k}^{(v)} .$$

Справедливо и обратное утверждение. Если

$$f_j^{(w)} = f_j^{(u)} f_j^{(v)},\ i = 0,1,...,n-1, \text{ то } w_i = \sum_{l=0}^{n-1} u_l v_{i-l} .$$

Эти утверждения носят название теорем о свертке в частотной и временной областях.

2. Если вектор V во временной области и его преобразование Фурье $F^{(v)}$ заданы в виде полиномов

$$V(x) = \sum_{i=0}^{n-1} v_i x^i \text{ и } F^{(v)}(z) = \sum_{i=0}^{n-1} f_i^{(v)} z^i .$$

то элемент α^j поля $GF(p^m)$ является корнем полинома тогда и только тогда, когда частотный компонент $f_j^{(v)}$ равен нулю; элемент α^{-j} является корнем $F^{(v)}(z)$ тогда и только тогда, когда i-я компонента v_i равна нулю.

На основе спектрального подхода можно дать еще одно равнозначное определение циклическому коду как множеству таких слов над конечным полем $GF(p)$, у которых все спектральные компоненты, принадлежащие заданному множеству частот, называемых проверочными, равны нулю.

5.4.6. Простейшие блочные линейные коды

Коды длины n и размерности k могут иметь разные значения кодового расстояния d_0 и обладать поэтому разной помехоустойчивостью. Оптимальными называются коды, обеспечивающие при заданных n и k максимальную вероятность правильного приема кодового слова. Они, как правило, имеют наибольшее кодовое расстояние. Числа n и k определяют скорость кода, равную k/n двоичных единиц на 1 символ.

При заданной величине кодового расстояния d_0 существует нижняя граница необходимого количества избыточных символов в кодовых комбинациях. Выражение для нижней границы $(n-k)_{\min} = \log_2\left(\sum_{i=0}^{q_{ИС}^{(d_0-1)/2}} C_n^i\right)$ получается в результате приравнивания числа $2^{(n-k)}-1$ различных ненулевых синдромов количеству $\left(\sum_{i=0}^{q_{ИС}^{(d_0-1)/2}} C_n^i\right)$ исправляемых кодом ошибок, начиная с их нулевой кратности.

Коды, для которых достигается нижняя граница, называются совершенными или плотноупакованными. Такие коды исправляют все ошибки кратности до $q_{ИС}$ включительно и ни одной ошибки более высоких кратностей.

Рассмотрим некоторые простейшие оптимальные коды.

Код с простой проверкой на четность обозначается $(n, n-1, 2)$ и содержит слова с одним проверочным символом $S = (a_1, a_2, ..., a_{n-1}, b)$. Проверочный символ

есть сумма по модулю 2 информационных:

$$b = a_1 + a_2 + ... + a_{n-1}.$$

Видно, что $b = 1$, если число единиц в информационной последовательности нечетное, и $b = 0$, если число единиц – четное. Таким образом, наличие проверочного символа позволяет всем кодовым словам придать общий признак: четность числа единиц в слове.

Порождающая матрица кода имеет $n-1$ строки n столбцов:

$$G = \begin{vmatrix} 1 & 0 & ... & 0 & 1 \\ 0 & 1 & ... & 0 & 1 \\ ... & ... & ... & ... & ... \\ 0 & 0 & ... & 1 & 1 \end{vmatrix}.$$

Любые строки матрицы содержат 2 единицы и отличаются значениями символов на двух позициях, поэтому кодовое расстояние равно 2. Следовательно, код может лишь обнаруживать однократные ошибки $(q_{об} = 2)$.

Проверочная матрица кода содержит одну строку

$$H = \begin{vmatrix} 1 & 1 & ... & 1 \end{vmatrix}.$$

и указывает, что для проверки основного признака кодовых слов надо сложить по модулю 2 все принятые символы.

Декодирование кода основано на проверке четности числа единиц в принятой последовательности Y. Для этого вычисляется синдром, содержащий один компонент

$$C = YH^T = c_1 + c_2 + ... + c_{n-1} + c,$$

или с учетом $Y = S + E$ и $SH^T = 0$

$$C = e_1 + e_2 + ... + e_n.$$

Значение $C = 0$ соответствует четному числу единиц в Y. В этом случае принимается решение об отсутствии ошибок, т.е. полагается $S = S$. Ясно, что при действии ошибки четной кратности $C = 0$, и данное решение будет неправильным.

Если $C = 1$, то фиксируется наличие ошибки. Очевидно, значение $C = 1$ даст любая ошибка нечетной кратности, т.е. код обнаруживает часть ошибок

большей кратности, чем $q_{об}=1$. Несомненным достоинством этого кода является высокая скорость, характеризуемая величиной $(n-1)/n$.

Код нечетной длины с повторением обозначается $(n,1,n)$, имеет кодовые слова $[ab_1b_2...b_{n-1}]$ с одним информационным символом и $n-1$ проверочными, которые повторяют информационный:

$$b_1=b_2=...=b_{n-1}=a.$$

Порождающая матрица состоит из одной строки $G=[11...1]$, так что код с повторением имеет всего два слова: одно содержит только нули, второе – только единицы. Понятно, что кодовое расстояние равно n, отсюда

$$q_{об}=n-1,\ q_{ИС}=(n-1)/2.$$

Проверочная матрица кода:

$$H=\begin{vmatrix}1 & 1 & 0 & 0 & ... & 0\\ 1 & 0 & 1 & 0 & ... & 0\\ ... & ... & ... & ... & ... & ...\\ 1 & 0 & 0 & 0 & ... & 1\end{vmatrix}.$$

имеет $n-1$ строк и n столбцов и указывает, что сумма первого и любого другого символов кодового слова должна равняться 0. Число избыточных символов в коде достигает нижней границы, следовательно, коды нечетной длины с повторением относятся к совершенным.

Проверочная матрица кода $(n,n-1,2)$ с простой проверкой на четность совпадает с порождающей матрицей кода с повторением $(n,1,n)$, а порождающая матрица первого кода подобна проверочной матрице второго. Такие коды называются дуальными.

Коды Хэмминга имеют кодовое расстояние $d_0=3$, исправляют все однократные ошибки или обнаруживают двукратные, т.е. $q_{об}=2$, $q_{ИС}=1$. Зависимости проверочных символов от информационных выбраны так, что каждой однократной ошибке соответствует свое ненулевое значение синдрома. Поэтому для кодов Хэмминга число ненулевых синдромов равно числу символов в кодовых комбинациях (числу однократных ошибок):

$$n = 2^{n-k} - 1. \qquad (5.24)$$

Следовательно, для кодов Хэмминга достигается нижняя граница числа $(n-k) = \log_2(n+1)$ избыточных символов, а сами коды являются совершенными. Примеры кодов: (3, 1,3), (7,4, 3), (15, И, 3), (31,26, 3), (63, 57, 3),....

Единственный код Голея (23, 12, 7) завершает ряд совершенных кодов.

Расширенный код Хэмминга образуется из совершенного путем добавления общей проверки на четность, т. е. проверочного символа, равного сумме всех символов кода Хэмминга. Код имеет кодовое расстояние $d_0 = 4$, что позволяет исправить все однократные и одновременно обнаружить все двукратные ошибки. Такой режим целесообразен, в частности, в системах передачи информации с обратной связью.

При добавлении проверочного символа длина кода становится четной, а соотношение (5.24) преобразуется к виду $n = 2^{n-k-1}$. Расширенные коды Хэмминга образуют ряд: (4, 1,4), (8,4,4), (16, 11,4), (32, 26,4), (64,57,4),....

Коды этого вида относятся к квазисовершенным, т.е. исправляющим все ошибки кратности по $q_{ИС}$ включительно и часть ошибок кратности $q_{ИС}+1$.

5.5. Коды Боуза-Чоудхури-Хоквингема

5.5.1. Методы задания кодов БЧХ

Коды Боуза-Чоудхури-Хоквингема (БЧХ) составляют один из больших классов линейных кодов, исправляющих ошибки. Причем метод построения этих кодов задан явно.

Код БЧХ длины n, исправляющий $q_{ИС}$-кратные ошибки, это циклический блочный код над полем $GF(p)$, корнями порождающего многочлена которого являются $\beta^v, \beta^{v+1}, ..., \beta^{v+2q_{ИС}-1}$, где β – элемент конечного поля $GF(p^m)$; v – целое число.

В соответствии с этим определением и выражением (5.21) порождающий многочлен кода БЧХ может быть представлен наименьшим общим кратным

$$g(x) = HOK\left[M_v(x), M_{v+1}(x), ..., M_{v+2q_{ИС}-1}(x)\right],$$

где $M_j(x)$ – минимальные многочлены элементов β^j.

Доказано [30, 33], что наличие $2q_{ИС}$ корней полинома $g(x)$, указанных в определении кода, гарантирует исправление всех ошибок кратности, меньшей или равной $q_{ИС}$.

Основное внимание обратим на коды БЧХ, имеющие длину $n = p^m - 1$. Такие коды называются примитивными кодами БЧХ.

Часто выбирают $v = 1$ (случай кодов БЧХ в узком смысле), что, как правило, приводит к порождающему полиному наименьшей степени, а значит, и к наименьшему числу избыточных символов в кодовом слове. Кроме того, целесообразно выбрать $\beta = \alpha$ (α – примитивный элемент поля $GF(p^m)$, поскольку при этом получается наибольшая длина кодового слова. Список порождающих многочленов кодов БЧХ различных длин (вплоть до $n = 256$) имеется, например, в [26].

Построенные таким образом коды БЧХ, исправляющие как минимум $q_{ИС}$-кратные ошибки, характеризуются конструктивным расстоянием кода $\delta = 2q_{ИС} + 1$. Истинное минимальное расстояние d_0 кода БЧХ может оказаться больше, чем δ. Это означает, что ряд кодов БЧХ может исправлять ошибки кратности большей, чем та, которую задают при построении этого кода.

Найдем проверочную матрицу двоичного циклического кода БЧХ, исправляющего $q_{ИС}$-кратные ошибки. Учитывая свойство равенства минимальных многочленов с номерами j и $2j$ (см. 5.3.5), степень порождающего многочлена $g(x)$ может быть снижена. Действительно, если, например, $v = 1$, порождающий многочлен примет вид

$$g(x) = НОК\left[M_1(x), M_3(x), ..., M_{2q_{ИС}-1}(x)\right]. \qquad (5.25)$$

При этом проверочная матрица

$$H = \begin{vmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{(n-1)} \\ 1 & \alpha^3 & \alpha^6 & \dots & \alpha^{(n-1)3} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \alpha^{(2q_{ИС}-1)} & \alpha^{2(2q_{ИС}-1)} & \dots & \alpha^{(n-1)(2q_{ИС}-1)} \end{vmatrix}. \qquad (5.26)$$

Сравнивая эту матрицу с матрицей (5.19) кода Хэмминга, видим, что код Хэмминга представляет собой частный случай примитивного кода БЧХ, исправляющего однократные ошибки $(q_{ИС}+1)$.

Представляет интерес воспользоваться возможностью описания кодов БЧХ в спектральной области (см. 5.4.5). Как следует из свойств дискретного преобразования Фурье, спектры слов циклического кода БЧХ должны содержать нулевые компоненты с номерами $j = v, v+1, ..., v+2q_{ИС}-1$.

Таким образом, циклический код БЧХ, исправляющий $q_{ИС}$-кратные ошибки, можно определить как множество всех слов над полем $GF(p)$, для которых $2q_{ИС}$ последовательных компонентов спектра равна нулю. Указанное свойство кодов БЧХ используется при их декодировании.

К особенностям кодов БЧХ можно отнести тот факт, что с ростом длины n кода при фиксированном значении скорости кода k/n отношение δ/n стремится к нулю. В результате, несмотря на наличие у кодов БЧХ отмеченных положительных свойств, при больших длинах $(n>1000)$ приходится отдавать предпочтение другим кодам [3, 30].

5.5.2. Принципы декодирования кодов БЧХ

Коды БЧХ относятся к классу циклических, что позволяет применять для их декодирования любые методы, разработанные для циклических кодов. Однако для кодов БЧХ получены специальные эффективные методы декодирования, называемые *алгебраическими.*

Для уяснения принципа декодирования этими методами обратимся к двоичным кодам БЧХ $(p-2)$, исправляющим двукратные ошибки $(q_{ИС}=2)$.

В соответствии с (5.25) и (5.26) эти коды можно задать при помощи проверочной матрицы

$$H = \begin{vmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \dots & \alpha^{(n-2)} & \alpha^{(n-1)} \\ 1 & \alpha^3 & \alpha^6 & \alpha^9 & \dots & \alpha^{3(n-2)} & \alpha^{3(n-1)} \end{vmatrix}. \qquad (5.27)$$

и порождающего полинома $g(x) = НОК[M_1(x), M_3(x)]$ имеющего своими корнями

элементы конечного поля $GF(p^m)$.

Последовательность принятых символов $Y=(y_0y_1...y_{n-1})$ описывается многочленом $Y(x)$ степени $n-1$, который можно представить через порождающий многочлен $g(x)$ следующим образом:

$$Y(x)=Z(x)g(x)+E(x).$$

где $E(x)$ – полином степени $n-1$, соответствующий вектору ошибки (см. 5.4.1) $E=[e_0e_1...e_{n-1}]$.

Векторы исправляемых ошибок кратности $q \le q_{ИС}=2$ представляются полиномом

$$E(x)\begin{cases}0, & \text{если ошибки нет } (q=0);\\ e_ix^i, & \text{если ошибка в i - м символе } (q=1);\\ e_ix^i+e_jx^j, & \text{если ошибки в i - м и j - м символах } (q=2).\end{cases}.$$

В двоичном случае ненулевые компоненты вектора ошибки e_i и e_j равны 1, поэтому задача декодирования кода БЧХ, исправляющего двукратные ошибки, сводится к определению номеров i и j позиций ошибочных символов в принятом слове Y. При этом оценка $\hat{S}$ переданного кодового слова S может быть получена в виде $\hat{S}=Y\oplus\hat{E}$, где $\hat{E}$ – оценка вектора ошибки, а $\oplus$ – знак суммирования по модулю 2.

Как будет видно из дальнейшего изложения, более удобно пользоваться не значениями i и j, а однозначно связанными с этими номерами некоторыми элементами β_1 и β_2 конечного поля $GF(p^m)$, причем $\beta_1=\alpha^i$, $\beta_2=\alpha^j$. Поскольку порядок примитивного элемента α равен длине кодового слова $n=2^m-1$, то элементы β_1 и β_2 действительно однозначно сопоставляются номерам i и j. Значения β_1 и β_2 принято называть локаторами, т.е. указателями искаженных позиций последовательности символов Y.

Процедура декодирования двоичного кода БЧХ начинается с вычисления синдрома

$$C=YH^T, \qquad (5.28)$$

являющегося $(n-k)$-символьной комбинацией: $C=(c_0c_1...c_{n-k-1})$. В соответствии с

(5.14) и (5.19)

$$C = SH^T + EH^T = EH^T,$$

т.е. синдром не зависит от принятой комбинации S, а определяется только вектором ошибки E.

Подставив в (5.28) матрицу (5.27) кода БЧХ, получим

$$C = (c_1, c_3) = \left(\sum_{l=0}^{n-1} e_l \alpha^l, \sum_{l=0}^{n-1} e_l \alpha^{3l} \right).$$

Поскольку при максимальной заданной кратности ошибок $q_{ИС} = 2$ вектор E имеет лишь два ненулевых компонента на позициях с номерами i и j, то

$$\begin{cases} \alpha^i + \alpha^j = c_1, \\ \alpha^{3i} + \alpha^{3j} = c_3. \end{cases}.$$

Используя введенные выше локаторы β_1 и β_2 получим следующую систему уравнений:

$$\begin{cases} \beta_1 + \beta_2 = c_1, \\ \beta_1^3 + \beta_2^3 = c_3. \end{cases}. \qquad (5.29)$$

Теперь можно указать один из путей решения задачи декодирования кода БЧХ, исправляющего двукратные ошибки. Действительно, по принятой комбинации символов Y в соответствии с (5.28) можно вычислить значение синдрома. Тогда система (5.29) будет содержать два линейно-независимых уравнения с двумя неизвестными β_1 и β_2, а значит, может быть решена (в конечном поле $GF(p^m)$ относительно этих неизвестных.

Для отыскания локаторов β_1 и β_2 оказывается удобным преобразовать систему (5.29) с тем, чтобы свести задачу к поиску корней некоторого многочлена: $(x - \beta_1)(x - \beta_2) = x^2 + (\beta_1 + \beta_2)x + \beta_1\beta_2$, где учтено, что вычитание в $GF(2)$ равносильно сложению.

Преобразуем второе уравнение системы (5.28):

$$c_3 = \beta_1^3 + \beta_2^3 = (\beta_1 + \beta_2)(\beta_1^2 + \beta_1\beta_2 + \beta_2^2) = c_1(\beta_1\beta_2 + c_1^2).$$

Теперь квадратное уравнение представимо в виде $x^2 + c_1 x + \left(c_1^2 + {c_3}/{c_1} \right) = 0$.

Решая его, можно найти значения локаторов β_1 и β_2, однако по причинам, которые будут изложены в 5.5.3, отыскиваются корни уравнения

$$(1-z\beta_1)(1-z\beta_2)=1+c_1z+\left(c_1^2+{}^{c_3}\!/_{c_1}\right)z^2=0.$$

Многочлен в левой части уравнения называется многочленом локаторов ошибок. Корни этого многочлена z_1 и z_2 являются обратными к локаторам, т.е. $z_1=\beta_1^{-1}$ и $z_2=\beta_2^{-1}$.

Обычно многочлен локаторов ошибок обозначается $\sigma(z)$ и имеет вид

$$\sigma(z)=\sum_{l=0}^{q}\sigma_l z^l=\prod_{l=1}^{q}(1-\beta_l z),$$

где q – вес вектора ошибки E.

В рассматриваемом случае коэффициенты этого многочлена равны соответственно $\sigma_0=1$; $\sigma_1=c_1$; $\sigma_2=c_1^2+{}^{c_3}\!/_{c_1}$.

Наконец, для решения задачи декодирования двоичного кода БЧХ необходимо отыскать корни z_1 и z_2 многочлена локаторов ошибок с известными коэффициентами σ_0, σ_1 и σ_2. Очевидно, это можно сделать, вычисляя его значения при z, равном всем ненулевым элементам $GF(2^m)$.

Следует отметить, что стратегия декодирования, изложенная на примере двоичного кода БЧХ, исправляющего только ошибки до кратности 2, с успехом может быть обобщена и использована для случая кодов БЧХ, исправляющих произвольную кратность ошибок.

Декодирование недвоичных кодов БЧХ $(p>2)$ представляет собой более сложную задачу. Это обусловлено тем, что при декодировании таких кодов необходимо не только определить номера символов, в которых произошла ошибка, но и вычислить p-ичные значения компонентов вектора ошибки E. Особенности выполнения этих операций применительно к p-ичным кодам Рида-Соломона приведены в следующем разделе.

5.5.3. Методы реализации этапов декодирования кодов БЧХ

Изложенные в 5.5.2 основные идеи декодирования кодов БЧХ позволяют представить процесс декодирования в виде совокупности следующих трех основных этапов.

Этап 1. Вычисление синдрома C по принятой комбинации символов Y.

Этап 2. Нахождение коэффициентов многочлена локаторов ошибок $\sigma(z)$.

Этап 3. Вычисление корней многочлена $\sigma(z)$, определение взаимных с ним величин – локаторов ошибок и исправление ошибочных символов в принятой комбинации Y (для недвоичных кодов БЧХ требуется найти, помимо локаторов, оценки компонентов e_i вектора ошибки).

Каждый из перечисленных этапов процесса декодирования кодов БЧХ может быть реализован разными способами, существенно отличающимися друг от друга по сложности реализации. Рассмотрим основные из этих способов и проиллюстрируем их примером.

Этап 1. Отыскание синдрома C кода БЧХ по принятой последовательности символов $Y=(y_0 y_1 ... y_{n-1})$ может осуществляться путем деления многочлена $Y(x)$ на минимальные многочлены $M_j(x)$, задающие порождающий многочлен $g(x)$ двоичного кода БЧХ (5.25). Действительно:

$$Y(x)=Q_1(x)M_1(x)+R_1(x),$$

$$Y(x)=Q_2(x)M_2(x)+R_2(x),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$Y(x)=Q_{2q_{ИС}-1}(x)M_{2q_{ИС}-1}(x)+R_{2q_{ИС}-1}(x),$$

где $R_j(x)$ – остаток от деления многочлена $Y(x)$ на минимальный многочлен $M_j(x)$.

Подставив в j-е уравнение корень α^j минимального многочлена $M_j(x)$, имеем

$$Y(\alpha^j)=Q_j(\alpha^j)M_j(\alpha^j)+R_j(\alpha^j)=R_j(\alpha^j), \qquad (5.30)$$

поскольку $M_j(\alpha^j)=0$. Учитывая (5.27), получим, что j-й компонент синдро-

ма $c_j = Y(\alpha^j) = R_j(\alpha^j)$.

Таким образом, первый этап декодирования кодов БЧХ может быть сведен к вычислению остатков $R_j(x)$ от деления многочлена $Y(x)$ на минимальные многочлены $M_j(x)$ элементов $\alpha^j (j = 1,3,...,2q_{ИС} - 1)$ и подстановке этих элементов в многочлены остатка $R_j(x)$.

Сопоставляя выражение (5.30) с (5.23), описывающим прямое преобразование Фурье, можно сделать вывод о том, что компонент c_j синдрома есть j-й коэффициент ДПФ принятой последовательности символов Y. Учтем также, что j-й коэффициент ДПФ этой последовательности есть сумма j-х коэффициентов ДПФ кодового слова S и вектора ошибки E, поэтому

$$c_j = Y(\alpha^j) = S_j(\alpha^j).$$

Поскольку по определению кода БЧХ $S(\alpha^j) = 0$ для $j = 1,2,...,2q_{ИС}$, то компонент синдрома $c_j = E(\alpha^j)$, $j = 1,2,...,2q_{ИС}$.

Таким образом, блок компонентов синдрома как бы образует в частотной области окно, через которое можно наблюдать $2q_{ИС}$ из n частотных составляющих спектра вектора ошибки. Заметим, что для двоичных кодов БЧХ из этих $2q_{ИС}$ компонентов c_j компоненты с нечетными номерами $(j = 1,3,...,2q_{ИС} - 1)$ вычисляются в соответствии с (5.30), а с четными – по формуле с $c_{2j} = c_j^2$. Справедливость этой формулы следует из правила возведения многочлена над $GF(p)$ в степень p (см. 5.3.5).

Этап 2. Вычисления коэффициентов полинома локаторов ошибок кодов БЧХ, ориентированных на исправление $q_{ИС} > 2$ ошибок, оказываются наиболее трудоемкими. Разработаны различные алгоритмы решения этой задачи, отличающиеся скоростью реализации.

При больших значениях кратности ошибок, особенно интересных на практике, более эффективным оказывается метод, предложенный Берлекэмпом [3, 26]. Суть метода состоит в сведении задачи отыскания коэффициентов о,-многочлена локаторов ошибок к задаче нахождения коэффициентов авторегрес-

сионного фильтра (регистра сдвига минимальной длины с линейной обратной связью), генерирующего известную последовательность компонентов синдрома. Следует указать также на возможность решения ключевого уравнения при помощи алгоритма Евклида [3], являющегося рекуррентным и по сложности реализации близким к алгоритму Берлекэмпа.

Этап 3. На данном этапе необходимо отыскать корни z_i многочлена локаторов ошибок $\sigma(z)$, коэффициенты которого найдены на предыдущем этапе. Эти корни являются обратными локаторам β_j, т.е. $\beta_j = z_j^{-1}$ (см. 5.4). По вычисленным таким образом локаторам β_j для двоичных кодов БЧХ остается лишь исправить (проинвертировать) ошибочные символы, номер i которых соответствует локатору β_j. При декодировании недвоичных кодов БЧХ необходимо найти оценку компонентов $\hat{e}_i$ вектора ошибки и лишь затем произвести исправление, формируя оценки символов переданного кодового слова $\hat{x}_i = y_i + \hat{e}_i$.

Поиск элементов z_j поля $GF(p^m)$, удовлетворяющих условию $\sigma(z_i) = 0$, состоит в переборе всех элементов поля. Именно этот способ реализован в процедуре Ченя [30], широко используемой для решения задач третьего этапа декодирования кодов БЧХ. Сущность процедуры состоит в том, что при известных коэффициентах $\sigma_1, \sigma_2, ..., \sigma_r$ декодер вычисляет значения $\sigma(\alpha) = \sum_k \sigma_k \alpha^k$, $\sigma(\alpha^2) = \sum_k \sigma_k \alpha^{2k}$ и т.д. (здесь α – примитивный элемент поля $GF(p^m)$).

Если для некоторого $k = i$ значение $\sigma(\alpha^i) = 0$, то это означает, что α^i является корнем многочлена локаторов ошибки, а элемент $\beta_j = \alpha^{-i}$ является локатором, указывающим на ошибочную позицию в кодовой последовательности, которую следует исправить.

5.6. Коды Рида-Соломона

5.6.1. Основные определения

Коды Рида-Соломона (РС) определены над полем $GF(p^m)$. В дальнейшем

рассматриваются коды над конечными полями характеристики $p=2$, как наиболее распространенные. Поэтому символы кодовых слов являются элементами поля $GF(2^m)$. Коды РС, таким образом, это не двоичные, а 2^m- ичные коды БЧХ длины $N=2^m-1$. Они являются циклическими кодами с порождающим многочленом

$$g(x)=(x-\alpha^{v})(x-\alpha^{v+1})..(x-\alpha^{v+\delta-2}). \qquad (5.31)$$

где α – примитивный элемент поля $GF(2^m)$; v – целое; δ – конструктивное расстояние кода.

Выбор длины кода $N=2^m-1$ гарантирует, что многочлен (5.31) является делителем x^N-1, и следовательно, всегда будет выполнено требование к порождающему многочлену циклического кода. Действительно, все ненулевые элементы поля $\beta=\alpha^i$, $i=0,1,2,...,N-1$ являются корнями многочленах x^N-1, в чем легко убедиться подстановкой $x=\beta$: $\beta^N-1=(\alpha^N)^i-1=1-1=0$, т.к. $\alpha^N=1$. Поэтому многочлен x^N-1 представим в виде произведения линейных множителей $(x-\alpha^i)$ и делится на любой многочлен (5.31).

Слово кода РС имеет вид: $A_1A_2...A_KB_1B_2...B_{N-K}$, где A_i и B_j – соответственно информационные и избыточные символы, которые могут быть представлены m-разрядными двоичными векторами. Число информационных символов $K=(N-cm.g(x))$, где $cm.g(x)$ – степень порождающего полинома $g(x)$. Коды РС являются систематическими кодами с максимальным кодовым расстоянием $D_0=N-K+1$, равным конструктивному δ.

Обозначение символов слов, длины кода, размерности, кодового расстояния прописными буквами A_i, B_j, N, K и D_0 введено для того, чтобы отличить их от аналогичных параметров двоичных кодов.

Коды РС являются линейными кодами, поэтому, кроме задания с помощью порождающего многочлена (5.31), они могут быть заданы также проверочным многочленом $h(x)=(x^N-1)/g(x)$, порождающей G и проверочной H матрицами.

Коды РС существуют для всех $K = 1,2,...,N$. При $K = 1$ код содержит нулевое слово и $2^m - 1$ ненулевых слов, полученных умножением одного базисного вектора порождающей матрицы на все $2^m - 1$ ненулевых элементов конечного поля. Если в (5.31) положить $v = 1$, то для $K = 1$ базисный вектор равен единичному. Тогда все ненулевые кодовые слова составлены из повторяющихся N раз ненулевых элементов поля. Очевидно, кодовое расстояние равно N. Это пример кода с повторением над произвольным конечным полем (см. 5.4.4).

При $K = N$ код РС становится безызбыточным, его словами являются все $\left(2^m\right)^N$ последовательностей, символами в которых служат элементы поля $GF\left(2^m\right)$. Минимальное расстояние этого кода равно 1.

При $K = N - 1$ в кодовом слове имеется один проверочный символ, который равен взвешенной сумме информационных символов. Весами служат элементы поля $GF\left(2^m\right)$. Так как последовательности из K информационных символов имеют минимальное расстояние, равное 1, то добавление проверочного символа, выбранного специальным образом, может увеличить кодовое расстояние до 2. Напомним, что для двоичных кодов один проверочный символ может быть связан с информационными единственным способом: он равен сумме всех информационных, что при $K = N - 1$ приводит к кодам с простой проверкой на четность. Таким образом, переход к недвоичным кодам (расширению двоичного поля, над которым задается код) предоставляет большие возможности для конструирования новых кодов.

Отметим и другие особенности кодов РС по сравнению с двоичными кодами. Так, расстояние между векторами над полем $GF\left(2^m\right)$ по-прежнему определяется как расстояние Хэмминга, т. е. равно числу пар несовпадающих компонентов, хотя символы кодовых слов как m-разрядные векторы могут отличаться друг от друга в $1,2,3,...,m$ разрядах. Такое определение расстояния не учитывает, насколько сильно различаются символы слов, но зато позволяет упростить анализ кодов.

Коэффициентами порождающего $g(x)$ и проверочного $h(x)$ полиномов

служат элементы поля $GF(2^m)$, а не 0 и 1, как у соответствующих многочленов двоичных кодов. Получение кодовых слов с помощью порождающей матрицы означает не простое суммирование строк матрицы, а, как уже указывалось, суммирование базисных векторов, умноженных на различные элементы поля $GF(2^m)$.

Проверочная матрица кода РС есть проверочная матрица кода БЧХ длины N, являющаяся частью матрицы ДПФ последовательностей с компонентами из поля $GF(2^m)$.

Особенности декодирования кодов РС по сравнению с двоичными кодами БЧХ связаны с тем, что вектор ошибки имеет компоненты ε_i из поля $GF(2^m)$. Поэтому при декодировании кодов РС недостаточно указать номера искаженных символов, надо еще определить, насколько искажены символы, т.е. найти значения ε_i вектора ошибки. Последнее требование усложняет процедуру декодирования кодов РС по сравнению с декодированием двоичных кодов БЧХ.

Пример 5.14. Пусть имеется код РС, исправляющий одиночные, двойные и тройные ошибки. Корнями порождающего многочлена являются элементы $\alpha,\alpha^2,\alpha^3,\alpha^4,\alpha^5,\alpha^6$, где α – примитивный элемент поля $GF(2^4)$, образованного с помощью неприводимого полинома $p(x)=x^4+x+1$ (см. 5.3.5). Спектральные коэффициенты ДПФ кодовых слов в точках α^i, $i=1,2,...,6$, равны 0. Параметры кода: $N=15$, $K=9$, $\delta=D_0=N-K+1=7$, $q_{ИС}=3$.

Так как синдром не зависит от передаваемого слова S, то для рассмотрения процедуры декодирования достаточно задать вектор ошибки E. Допустим, что многочлен, описывающий вектор ошибки, $E(x)=\alpha^7x^3+\alpha^{11}x^{10}$, т.е. произошла двукратная ошибка, исказившая в слове $S=(s_0s_1s_2...s_{14})$ символы s_3 и s_{10}, причем принятые символы $y_3=s_3+\varepsilon_3=s_3+\alpha^7$, $y_{10}=s_{10}+\varepsilon_{10}=s_{10}+\alpha^{11}$ остальные – $y_i=s_i$.

Декодер по реализации Y должен найти оценку вектора ошибки $\hat{E}$, а затем оценку переданного слова $\hat{S}=Y+\hat{E}$. Если оценка $\hat{E}=E$, то $\hat{S}=S+E+\hat{E}=S$, и ошибка исправляется. Если $\hat{E}\neq E$, ошибка не исправляется.

Процедура нахождения E зависит от описания вектора ошибки E. Для указания номеров искаженных символов воспользуемся многочленом локаторов ошибок $\sigma(z)$ (см. 5.5.2). Для двоичных кодов многочлен $\sigma(z)$ полностью определяет вектор ошибки E. В случае кодов над полем $GF(2^m)$, каковыми являются коды РС, необходимо еще указать значения ошибок ε_i, что делается с помощью так называемого многочлена значений ошибок $\Omega(z)$ [3, 26]. Для нахождения ε_i требуется знание обоих многочленов $\sigma(z)$ и $\Omega(z)$, так как

$$\varepsilon_i = \Omega\left(\alpha^{-i}\right) \Big/ \sigma'\left(\alpha^{-i}\right). \qquad (5.32)$$

где i – номер искаженного символа; $\sigma'\left(\alpha^{-i}\right)$ – формальная производная многочлена $\sigma(z)$ в точке α^i, которая для кодов, заданных над полями характеристики 2, содержит только четные степени. Коэффициенты при нечетных степенях производной являются четными числами и обращаются в ноль.

Итак, произвольный вектор ошибки над полем $GF(2^m)$ может быть задан с помощью двух многочленов $\sigma(z)$ и $\Omega(z)$. Отметим, что эти многочлены определены с точностью до постоянного множителя β – элемента поля $GF(2^m)$. Многочлен $\beta\sigma(z)$ имеет множество корней, совпадающее с множеством корней многочлена $\sigma(z)$, и следовательно, оба определяют одни и те же номера искаженных символов.

Главная задача декодера состоит в оценке многочленов $\sigma(z)$ и $\Omega(z)$. Если эти оценки правильно описывают реальную ошибку, т.е. множества корней о $\hat{\sigma}(z)$ и $\sigma(z)$ совпадают, то такая ошибка исправляется. В противном случае исправления не произойдет.

Многочлены локаторов $\sigma(z)$, значений ошибок $\Omega(z)$ и синдрома $C(z)$ связаны ключевым уравнением [26, 30]:

$$\sigma(z)C(z) \equiv \Omega(z) \bmod\left(z^{2q_{\text{ис}}}\right). \qquad (5.33)$$

Трудность решения этого уравнения состоит том, что в соответствии с принципом максимального правдоподобия должны быть найдены многочлены $\hat{\sigma}(z)$ и $\hat{\Omega}(z)$ минимальных степеней, причем степень $\hat{\sigma}(z)$ должна быть больше

степени $\Omega(z)$, но менее степени $q_{\text{ИС}}$.

Разработаны различные методы решения ключевого уравнения, некоторые из которых уже упоминались в третьем разделе.

Заданная в примере конфигурация двукратной ошибки в соответствии с (5.29) описывается многочленом локаторов, имеющим корни, обратные элементам α^3 и α^{10}:

$$\sigma(z)=\beta\left(1-\alpha^3 z\right)\left(1-\alpha^{10}z\right)=\beta\left(1+\left(\alpha^3+\alpha^{10}\right)z+\alpha^3\alpha^{10}z^2\right)=\beta\left(1+\alpha^{12}z+\alpha^{13}z^2\right).$$

Напомним, что вычисления производятся в поле $GF\left(2^4\right)$, пример которого дан в табл. 5.2. Сложение удобно выполнять, используя полиномиальное или векторное представление элементов поля, а умножение – степенное представление. Поэтому $\alpha^3+\alpha^{10}=1000+0111=1111=\alpha^{12}$.

Этапы декодирования данного кода при сделанных предположениях о характере искажении описываются следующими операциями.

Этап 1. По принятой реализации символов Y вычисляются 6 компонентов синдрома $C_1,C_2,...,C_6$, являющихся коэффициентами ДПФ последовательности Y. Для заданного искажения $E(x)=\alpha^7x^3+\alpha^{11}x^{10}$ компоненты синдрома определяются по формуле

$$C_j=Y\left(\alpha^j\right)=E\left(\alpha^j\right)=\alpha^7\alpha^{3j}+\alpha^{11}\alpha^{10j},\ j=1,2,...,6.$$

Используя табл. 5.2 задания элементов поля $GF\left(2^4\right)$, находим

$$C_j=E(\alpha)=\alpha^7\alpha^3+\alpha^{11}\alpha^{10}=\alpha^{10}+\alpha^{21}=\alpha^{10}+\alpha^6=0111+1100=1011=\alpha^7.$$

Читателю предлагается самостоятельно произвести вычисления остальных коэффициентов ДПФ и убедиться, что $C_2=\alpha^{12}$, $C_3=\alpha^6$, $C_4=\alpha^{12}$, $C_5=\alpha^{14}$, $C_6=\alpha^{14}$.

Результатом выполнения первого этапа является многочлен синдрома

$$C(z)=\alpha^{14}z^5+\alpha^{14}z^4+\alpha^{12}z^3+\alpha^6z^2+\alpha^{12}z+\alpha^7.$$

Этап 2. Декодер решает ключевое уравнение (5.33) при $q_{\text{ИС}}=3$, оценивает многочлены $\sigma(z)$ и $\Omega(z)$ с помощью алгоритма Евклида нахождения НОД. Не вдаваясь в детали этого алгоритма, с которыми можно познакомиться в [26], приведем результат решения (5.33):

$$\sigma(z) = \alpha^9 z^2 + \alpha^8 z + \alpha^{11}, \quad \Omega(z) = \alpha^2 z + \alpha^3.$$

Анализ показывает, что $\sigma(z) = \alpha^{-4}\left(\alpha^{13} z^2 + \alpha^{12} z + 1\right) = \alpha^{-4}\sigma(z)$, т.е. в данном случае многочлен локаторов определен правильно.

Этап 3. Исправление ошибок производится при сложении символов принятой комбинации Y с компонентами ε_i вектора ошибки на позициях, номера которых обратны корням многочлена $\sigma(z)$. Корни $\sigma(z)$ определяются путем непосредственной подстановки в многочлен всех ненулевых элементов поля $\beta = \alpha^{-i}$, $i = 0,1,2,...,2^m - 1$. Если $\sigma\left(\alpha^{-i}\right) = 0$, то предполагается, что искажен символ ε_i.

Проделав такую подстановку, найдем корни многочлена $\sigma(z)$: $z = \alpha^{-3}$ и $z = \alpha^{-10}$. Действительно, для элемента α^{-3}: $\sigma\left(\alpha^{-3}\right) = \alpha^9\alpha^{-6} + \alpha^8\alpha^{-3} + \alpha^{11} = 1000 + + 0110 + 1110 = 0000$.

Для оценки вектора ошибки $E(x) = \varepsilon_3 x^3 + \varepsilon_{10} x^{10}$, т.е. для определения ε_3 и ε_{10}, найдем производную многочлена локаторов $\sigma'(z) = 2\alpha^9 z + \alpha^8 = \alpha^8$ и с помощью формулы (5.32) Получим:

$$\varepsilon_3 = \Omega\left(\alpha^{-3}\right) \Big/ \alpha^8 = \left(\alpha^2\alpha^{-3} + \alpha^3\right) \Big/ \alpha^8 = \alpha^{-9} + \alpha^{-5} = \alpha^6 + \alpha^{10} = \alpha^7,$$

$$\varepsilon_{10} = \Omega\left(\alpha^{-10}\right) \Big/ \alpha^8 = \left(\alpha^2\alpha^{-10} + \alpha^3\right) \Big/ \alpha^8 = \alpha^{11}.$$

Оценка многочлена ошибки $E(x) = \varepsilon_3 x^3 + \varepsilon_{10} x^{10}$ совпадает с $E(x)$, по предположению исказившим кодовое слово. Поэтому $S = S$, и двукратная ошибка исправлена.

В заключение рассмотрим декодирование данного кода, если кратность ошибки больше трех, например, $q_{\text{ИС}} = 4$. Такая ошибка описывается многочленом локаторов 4-й степени. Декодер, конечно, «не знает» о кратности произошедших искажений и выполняет алгоритм декодирования, предназначенный для исправления ошибок кратности не более 3.

На первом этапе определяется синдром, содержащий по-прежнему 6 компонентов. Синдром ненулевой, так как кратность ошибки меньше кодового

расстояния $D_0 = 7$, и ошибка кратности 4 обнаруживается. В результате решения ключевого уравнения (если это окажется возможным) будет найден многочлен локаторов, степень которого не превышает 3. Ясно, что множества корней многочленов разных степеней не совпадают (многочлен локаторов не имеет кратных корней). Поэтому корни $\sigma(z)$, найденные на третьем этапе декодирования, будут неправильно указывать номера искаженных символов. Ошибка кратности 4 обнаруживается, но не исправляется.

Коды РС имеют важное теоретическое и практическое значение, так как при заданных N и K имеют максимальное кодовое расстояние, используются для обнаружения и исправления пакетов ошибок и построения высокоэффективных каскадных кодов.

5.6.2. Обнаружение и исправление пакетов ошибок

Основные результаты теории помехоустойчивого кодирования получены в предположении, что передача информации производится по симметричному каналу, для которого постулируется независимость компонентов вектора ошибок как друг от друга, так и от передаваемых символов. Эта удобная модель значительно упрощает теорию кодов, но зато не всегда удовлетворительно отражает процессы передачи информации.

В реальных каналах наблюдаются всплески искажений, причинами которых могут быть, например, коммутационные помехи, быстрые замирания радиосигнала. Чтобы приблизить модель симметричного канала к действительности, вводится понятие пакета ошибок.

Пакетом ошибок длины l называется вектор ошибок E, все ненулевые компоненты которого расположены на отрезке из l подряд следующих позиций, причем в начале и конце отрезка расположены ненулевые компоненты. При $l = 1$ имеем пакет длиной в один символ, т.е. однократную ошибку.

Такая модель пакета в упрощенном виде, без привлечения характеристик коррелированных случайных процессов описывает зависимость отдельных искажений друг от друга. Различают однократные и многократные пакеты в за-

висимости от того, сколько пакетов искажают кодовое слово.

Для обнаружения и исправления пакетов ошибок разработано много способов [33]. Так, для борьбы с пакетами ошибок используется перемежение символов. Оно заключается в том, что символы одного кодового слова на передающей стороне разносятся друг от друга (распределяются по времени) на расстояние, превышающее длину пакета. В образовавшемся промежутке передаются символы других кодовых слов, так что друг за другом следуют символы не одного, а разных кодовых слов.

Пакет ошибок, действуя на $q_{\text{ИС}}$ подряд следующих символов, тем не менее, искажает только один символ каждого кодового слова. В результате перемежения модель канала с пакетами ошибок трансформируется в модель канала с независимыми ошибками. На приемной стороне символы каждого кодового слова выделяются из принятой последовательности, собираются вместе и декодируются по правилам, рассчитанным на исправление независимых ошибок. Подробно вопросы построения перемежителей рассмотрены в [26]. Здесь же поясним использование кодов РС для обнаружения и исправления пакетов ошибок.

На рис. 5.4 схематично показаны последовательность из $k = Km$ двоичных информационных символов a_i и случайно расположенный, одиночный пакет ошибок длины l.

Рис. 5.4. Символы кода РС и пакет ошибок

Применение кодов РС, заданных над полем $GF\left(2^m\right)$, основано на разбие-

нии двоичной последовательности на K блоков, содержащих по m двоичных символов. Эти блоки рассматриваются как двоичное представление элементов поля $GF(2^m)$ и, следовательно, кодированию подвергается последовательность $A_1, A_2, ..., A_K$ информационных символов. Видно, что такое укрупнение уменьшает длину пакета искажений, действующих на символы A_i. Если в двоичной последовательности пакет может исказить l символов, то для символов кода РС максимальная кратность ошибок снижается из-за случайного расположения пакета приблизительно до $l/m + 1$.

Таким образом, задача обнаружения и исправления пакетов ошибок сводится к использованию кода РС с кодовым расстоянием, позволяющим исправлять ошибки кратности лишь до $l/m + 1$. Процедура исправления в этом случае ничем не отличается от исправления независимых ошибок. Для кодов РС также возможно перемежение символов. В отличие от двоичных кодов, когда перемежение соответствует перестановке одиночных двоичных символов, перемежение символов кода РС означает перестановку блоков из m двоичных символов.

5.7. Коды Рида-Маллера

5.7.1. Задание и декодирование кодов Рида-Маллера

Коды Рида-Маллера относятся к линейным двоичным кодам, имеющим большие кодовые расстояния и исправляющим благодаря этому много ошибок. Они пригодны для каналов с малым отношением сигнал/помеха. Этот класс кодов интересен и потому, что с ним связаны многие другие сигналы, применяемые в радиотехнических системах: ортогональные и биортогональные сигналы, симплексные коды, m-последовательности и коды Хэмминга.

Будем рассматривать простейшие коды Рида-Маллера, слова которых являются линейными комбинациями некоторых двоичных функций обладающих полезными для практики свойствами. Сразу укажем, что эти функции выбраны такими, что их отображение в поле действительных чисел дает систему ортого-

нальных функций. Это свойство используется при декодировании.

Данное ограничение означает, что в базис кода не входят произведения двоичных функций, т.е. рассматривается код Рида-Маллера 1-го порядка. Некоторые сведения о кодах Рида-Маллера более высоких порядков имеются в [30].

Кодовое слово длины n обычно рассматривается как булева функция (или ее инверсия), заданная в 2^m точках, т.е. на наборах из m двоичных элементов. Можно многими способами нумеровать позиции кодового слова m-разрядными двоичными векторами. Ясно, что, как в случае кодов Хэмминга, такая перестановка не влияет на помехоустойчивость получаемых кодов. Будем нумеровать позиции кодового слова числами в двоичной системе счисления $(v_1 v_2 ... v_m)$, где $v_i = 0;1$ для $i = 1,2,...,m$. Ввиду линейности кодов Рида-Маллера каждый символ кодового слова S_i представим линейной комбинацией

$$S_i = a_1 v_1 + a_2 v_2 + ... + a_m v_m ,$$

или ее инверсией

$$1 + S_i = a_0 1 + a_1 v_1 + a_2 v_2 + ... + a_m v_m ,$$

где $a_0, a_1, a_2, ..., a_m$ – известные информационные символы.

В соответствии с определением порождающей матрицы (5.16) и правилом покомпонентного сложения векторов элементы $(1, v_1, v_2, ..., v_m)$ являются столбцами матрицы G. Для $m = 3$ порождающая матрица размера $m + 1 = 4$ на $n = 8$ имеет вид:

$$G = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{vmatrix}.$$

Столбцы матрицы G без верхней строки представляют собой последовательность чисел, записанных в двоичной системе счисления (младшие разряды внизу). Таким образом, столбцы матрицы можно рассматривать как последовательность состояний двоичного суммирующего счетчика:

$$G = \begin{vmatrix} 1 \\ S_{\delta 1} \\ \dots \\ S_{\delta m} \end{vmatrix}, \qquad (5.34)$$

где 1 – последовательность из единиц; $S_{\delta 1}$ – последовательность состояний последнего (старшего) разряда счетчика; $S_{\delta m}$ – последовательность состояний первого (младшего) разряда. Отметим, что перестановка столбцов и строк порождающей матрицы приводит к эквивалентным кодам.

Кодовое слово есть линейная комбинация базисных векторов (строк матрицы G):

$$S = (s_1 s_2 \dots s_n) = a_0 1 + a_1 S_{\delta 1} + a_2 S_{\delta 2} + \dots + a_m S_{\delta m}. \qquad (5.35)$$

Вид матрицы (5.34) указывает простой способ формирования базисных векторов и получения кодового слова. Схема кодирующего устройства для $m = 3$ (рис. 5.5) содержит трехразрядный двоичный счетчик, вырабатывающий функции $S_{\delta 1}, S_{\delta 2}, S_{\delta 3}$, и комбинационную схему, реализующую булеву функцию (5.35). Естественно, длительность информационных символов, подаваемых в этот кодер, предполагается равной длительности кодового слова, т.е. в данной случае 8 длительностям символов канала.

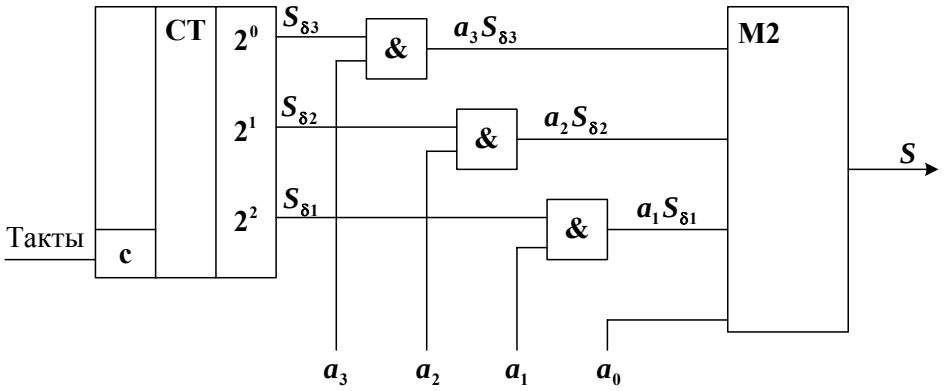


Рис. 5.5. Кодер кода Рида-Маллера длины n=8

Двоичный вектор $S = (s_1 s_2 \dots s_n)$ с компонентами $s_ш = 0;1$ может быть отображен в вектор $W = (w_1, w_2, \dots, w_n)$ с действительными компонентами $w_i = \pm 1$. Для этого надо «0» в двоичном векторе заменить на (+1), а «1» – на (–1). Такое отобра-

жение можно определить формулами:

$$W = (-1)^S,\ w_i = (-1)^{S_i}. \qquad (5.36)$$

В табл. 5.8 приведены все 16 кодируемых информационных последовательностей и соответствующие им кодовые слова. Обратим внимание, что кодовые слова правой половины таблицы являются инверсией слов левой половины.

Тогда операции суммирования двоичных последовательностей будет соответствовать покомпонентное умножение последовательностей с элементами ± 1. Можно говорить о том, что аддитивная группа двоичных векторов отображается в мультипликативную группу действительных векторов (см. 5.3.2). Конечно, отображение (5.36) не требует для технического воплощения дополнительного оборудования, так как делается разработчиком аппаратуры

Таблица 5.8

Кодовые слова кода Рида-Маллера

Информационные символы	Кодовое слово	Информационные символы	Кодовое слово
0000	00000000	1000	11111111
0001	01010101	1001	10101010
0010	00110011	1010	11001100
0011	01100110	1011	10011001
0100	00001111	1100	11110000
0101	01011010	1101	10100101
0110	00111100	1110	11000011
0111	01101001	1111	10010110

мысленно и связано с разной трактовкой одних и тех же уровней напряжений (высоких и низких) в технических устройствах.

Если применить отображение (5.36) к строкам матрицы (5.34), по определению получим известные функции Радемахера [10]:

$$R_1 = (-1)^{S_{\delta 1}},\ R_2 = (-1)^{S_{\delta 2}},\ldots, R_m = (-1)^{S_{\delta m}}.$$

Известно, что всевозможные произведения функций Радемахера образуют полную ортогональную систему функций, которые называются функциями Уолша, и им соответствуют слова кода Рида-Маллера. Ортогональность функций означает, что для двух произвольных функций Уолша W_i и W_j, $i \neq j$ скалярное произведение

$$(W_i, W_j) = \sum_{k=1}^{n} w_{ik} w_{jk} = 0.$$

т.е. число позиций, на которых символы последовательностей совпадают, равно числу позиций, на которых последовательности отличаются, и равно поэтому $n/2$. Из определения расстояния Хэмминга (см. 5.4.1) заключаем, что кодовое расстояние кода Рида-Маллера равно $n/2$. Отметим, что для пар слов, являющихся инверсией друг друга, расстояние равно n.

Таким образом, коды Рида-Маллера имеют длину $n = 2^m$, $m+1$ информационный символ, кодовое расстояние $d_0 = n/2 = 2^{m-1}$. Отображение двоичных кодовых слов в область действительных чисел ± 1 дает множество функций Уолша, включающее 2^m функций Уолша W_i, $i = 1,2,...,n$, и 2^m противоположных функций $(-W_i)$, $i = 1,2,...,n$. Множество сигналов, составленных из функций Уолша и противоположных им, называется системой биортогональных сигналов. Если в систему не включать противоположные сигналы, то получим систему ортогональных сигналов, которые используются в качестве адресов абонентов в системах множественного доступа с кодовым разделением каналов. Применение ортогональных сигналов в качестве канальных позволяет разделять их в таких системах без взаимных помех [20].

Для кодов Рида-Маллера разработаны достаточно эффективные алгоритмы порогового (мажоритарного) декодирования, изложенные в [30]. Здесь рассмотрим декодирование кодов Рида-Маллера по принципу максимума правдоподобия. Для симметричного канала это совпадает с декодированием по минимуму расстояния между векторами, при котором в качестве оценки переданно-

го вектора S берется слово, ближайшее к принятому вектору Y.

Имея в виду преобразование (5.36), рассмотрим коэффициент корреляции F_j между принятым вектором Y и функцией Уолша W_j. При $a_0 = 0$

$$F_{j0} = (Y, W_j) = \sum_{i=1}^{n} y_i w_{ij},$$

где y_i и w_{ij} принимают значения ± 1.

Поскольку при совпадении знаков y_i и w_{ij} их произведение равно 1, а при несовпадении – (–1), то

$$F_{j0} = n_c - n_{\text{нс}} = n_c - 2n_{\text{нс}} = n - 2d(Y, W_j),$$

где n_c и $n_{\text{нс}}$ – соответственно числа совпадающих и несовпадающих символов в Y и W_j, а $n = n_c + n_{\text{нс}}$. При $a_0 = 1$, очевидно, получим

$$F_{j1} = 2d(Y, W_j) - n.$$

Таким образом, оптимальный алгоритм декодирования предполагает следующие этапы:

1. Вычисление 2^m коэффициентов корреляции F_j между Y и функциями Уолша W_j, $j = 1,2,...,2^m$.

2. Поиск максимального по абсолютной величине коэффициента $|F_j|_{\max}$.

3. Принятие решения по правилу: $c_0 = 0$, если $F_j > 0$, и $c_0 = 1$, если $F_j < 0$.

Следовательно, данная процедура представляет собой многоканальный корреляционный прием. Ее сложность пропорциональна числу n^2 операций сложения и вычитания. Разработаны быстрые алгоритмы декодирования кодов Рида-Маллера, по своей сути аналогичные алгоритмам быстрого преобразования Фурье [30].

5.7.2. Симплексные коды и m-последовательности

Симплексным кодом называется линейный код, порождающая матрица которого равна проверочной матрице кода Хэмминга, т.е. симплексный код дуален к коду Хэмминга [30]. Свое название код получил потому, что его слова,

рассматриваемые как многомерные векторы, образуют правильную многомерную фигуру, называемую симплексом.

Так как существует много эквивалентных друг другу кодов Хэмминга, отличающихся перестановкой столбцов проверочной матрицы, то имеется и много эквивалентных симплексных кодов. Среди них есть и код с порождающей матрицей, столбцами которой являются целые числа, записанные в двоичной системе счисления.

Подобная матрица уже рассматривалась в 5.7.1 как часть порождающей матрицы кода Рида-Маллера. Действительно, если в матрице (5.34) исключить первую строку и первый столбец, то получим все $2^m - 1$ ненулевых различных столбцов, т.е. порождающую матрицу симплексного кода. Исключение первой строки, состоящей из единиц, можно трактовать как формирование слов кода Рида-Маллера при $a_0 = 0$, а исключение первого столбца означает исключение первого (нулевого) символа из слов Рида-Маллера. Таким образом, симплексный код получается из кода Рида-Маллера путем отбрасывания при $a_0 = 0$ первого символа, равного 0 для всех 2^m слов (см. левую часть табл. 5.8).

Так как при $a_0 = 0$ множество слов кода Рида-Маллера соответствует ортогональным функциям Уолша, отличающимся друг от друга в 2^{m-1} позициях, а исключение одинакового для всех слов символа не изменяет расстояния, то отсюда основное свойство симплексного кода: все его слова находятся друг от друга на одинаковом расстоянии, равном 2^{m-1}. Итак, симплексный код имеет длину $n = 2^m - 1$, число информационных символов $k = m$ (на единицу меньше, чем у кода Рида-Маллера), кодовое расстояние $\mathrm{d}_0 = 2^{m-1} = n/2$.

В технических приложениях большое значение имеет циклический симплексный код, столбцы порождающей матрицы которого являются степенями примитивного элемента поля Галуа $GF(2^m)$:

$$G = \left[1, \alpha, \alpha^2, ..., \alpha^{n-1}\right]. \qquad (5.37)$$

Порождающая матрица вида (5.37) указывает простой способ формирования базисных векторов кода путем последовательного умножения элементов

поля $GF(2^m)$ на α. Устройства умножения на α представляют собой сдвигающий регистр, охваченный линейной обратной связью [10].

Отметим, что помехоустойчивость симплексного кода всегда выше, чем ортогонального, так как при равных кодовых расстояниях слово симплексного кода на 1 элемент короче и, следовательно, энергия, приходящаяся на один символ, больше. Правда, при увеличении m выигрыш уменьшается, стремясь к 0.

На первый взгляд каждая комбинация симплексного кода похожа на случайную последовательность единиц и нулей, из-за чего слова этого кода часто называют псевдослучайными последовательностями (ПСП), m-последовательностями (отмечая связь с параметром кода m), линейными рекуррентными последовательностями, последовательностями максимальной длины. Все эти названия отражают специфические свойства слов симплексного кода.

ПСП широко используются в радиолокации и радионавигации, в асинхронных системах передачи информации, что обусловлено хорошими корреляционными свойствами этих сигналов.

Перечислим основные структурные свойства m-последовательностей. Рассмотрим бесконечную m-последовательность, получаемую, например, в результате бесконечного умножения элементов поля на α или с выхода генератора m-последовательности.

1. Последовательность $S_0, S_1, S_2, ..., S_l, ...$ удовлетворяет рекуррентному соотношению для всех $l \geq m$:

$$S_l = S_{l-1}h_{m-1} + S_{l-2}h_{m-2} + ... + S_{l-m}h_0, \qquad (5.38)$$

где $h_i = 0;1$ – известные двоичные числа (коэффициенты минимального многочлена примитивного элемента поля).

Это свойство следует из того, что генератором m-последовательности является регистр с линейной обратной связью. Символы $S_{l-1}, S_{l-2}, ..., S_{l-m}$ записаны в m его ячейках, а S_l есть выход линейной цепи обратной связи.

2. Период m-последовательности равен $2^m - 1$ и максимален по сравне-

нию с периодами других последовательностей, получаемых на m-разрядном сдвигающем регистре, но с другим законом функционирования цепи обратной связи.

Это объясняется тем, что степени примитивного элемента α пробегают все $2^m - 1$ ненулевых элементов поля, т.е. все различные m-разрядные векторы.

3. В m-последовательности число единиц равно 2^{m-1}, а число нулей $2^{m-1} - 1$, что очевидно, так как слово симплексного кода может быть получено путем исключения первого нулевого символа из слов кода Рида-Маллера, у которых число единиц и нулей одинаково и равно 2^{m-1}. Приблизительно равное соотношение символов 0 и 1 имеет место для отрезка из m символов, взятого в любом месте бесконечной последовательности. Для наблюдателя, не знающего закона формирования последовательности (5.5), появление 0 и 1 происходит примерно с равными вероятностями. Если же рекуррентное соотношение (5.5) известно, то по любому отрезку из m символов можно восстановить все оставшиеся символы последовательности. Этим и объясняется название «псевдослучайная» последовательность.

4. Сумма m-последовательности и ее циклического сдвига есть m-последовательность, удовлетворяющая тому же рекуррентному соотношению, но с другим сдвигом.

Это следует из того, что циклический сдвиг ПСП принадлежит симплексному коду, а сумма кодовых слов линейного кода есть слово данного кода (см. 5.4.1).

Таким образом, симплексный циклический код содержит один нулевой вектор, а остальные $2^m - 1$ являются циклическими сдвигами одной m-последовательности, заданной (5.38). Эти сдвиги, называемые также фазами ПСП, можно пронумеровать различными m-разрядными векторами и сопоставить их с состояниями сдвигающего регистра. Отсюда понятен способ получения ПСП с заданной фазой: в сдвигающий регистр генератора записывается номер фазы, и за и тактов работы генератора формируется требуемая ПСП.

Перейдем теперь к описанию корреляционных свойств сигналов, соответ-

ствующих m-последовательностям. Для оценки этих свойств в зависимости от применения сигналов используют периодическую автокорреляционную функцию (ПАКФ), периодическую взаимно-корреляционную функцию (ПВКФ) или непериодические автокорреляционные (АКФ) и взаимно-корреляционные функции (ВКФ).

Так как корреляционная функция определена для последовательностей с действительными компонентами, то далее предполагается, что двоичные символы m-последовательности заменены на +1 и –1 с помощью отображения (5.36). По определению, нормированная ПАКФ бесконечной последовательности $S_0, S_1, S_2, ..., S_l, ...$ с периодом n

$$\rho(\tau) = \rho(\tau + n) = \frac{1}{n}\sum_{j=0}^{n-1}(-1)^{S_j + S_{j+\tau}},$$

где τ – число тактов сдвига.

После несложных преобразований можно получить

$$\rho(\tau) = \frac{(n_c - n_{нс})}{n},$$

где n_c и $n_{нс}$ – соответственно числа совпадений и несовпадений символов последовательности S и ее циклического сдвига, причем $n_c + n_{нс} = n$. Тогда

$$\rho(\tau) = \frac{(n_c - 2d_0)}{n} = \frac{\left(2^m - 2\cdot 2^{m-1}\right)}{n} = \frac{-1}{n},$$

т.е. боковые лепестки ПАКФ m-последовательности постоянны и равны $-1/n$.

Вид ПВКФ m-последовательностей зависит от многих факторов, в том числе и от соотношения между периодами ПСП: периоды равны, взаимно просты или произвольны. Для частных случаев получены или точные, или верхние границы значений взаимно-корреляционных функций. Например, нормированная ПАКФ m-последовательностей с взаимно простыми периодами имеет при любых сдвигах одинаковые значения, равные $p_c - p_{нс}$, где p_c и $-p_{нс}$ – соответственно вероятности совпадения и несовпадения символов двух ПСП.

Корреляционные свойства одиночных (непериодических) ПСП также за-

висят от многих факторов. Многочисленные исследования показывают, что можно подобрать несколько ПСП или отрезков одной ПСП с удовлетворительными АКФ и ВКФ. Относительно боковых лепестков АКФ одиночных ПСП известно, что их максимальный уровень не превышает $1/\sqrt{n}$.

Отметим, что m-последовательности являются базовыми для получения других ансамблей сигналов. Так, последовательности Голда, обладающие достаточно хорошими ВКФ, формируются путем сложения двух разных m-последовательностей одинаковых периодов.

Существование разных ПСП одного периода объясняется наличием в конечном поле Галуа нескольких примитивных элементов.

5.7.3. Связь между блочными кодами

Обобщим вопросы кодирования, изложенные в предыдущих разделах, и поясним связи между некоторыми из рассмотренных кодов, схематично показанные на рис. 5.6. В прямоугольниках указаны длины n, размерности k, кодовые расстояния d_0, обозначены порождающие G и проверочные H матрицы анализируемых кодов. Стрелками указаны операции, с помощью которых можно от одного кода перейти к другому. Эти операции можно рассматривать как способы построения новых кодов из заданных. Перечислим некоторые из таких способов.

Расширение кода означает добавление к кодовому слову проверочного символа, равного сумме всех символов преобразуемого кодового слова. Новый проверочный символ есть общая проверка на четкость. В результате все слова имеют четное число единиц, и кодовое расстояние увеличивается на единицу, если оно были нечетным числом. При четном кодовом расстоянии расширение не увеличивает кодового расстояния, так как для векторов с минимальным четным весом общая проверка на четность равна 0, и вес расширенного вектора не изменяется.

Проверочная матрица расширенного кода Хэмминга образуется путем

введения в матрицу H исходного кода нулевого столбца и строки из единиц:

$$H_{рас} = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & & & & \\ 0 & & [H] & & \\ 0 & & & & \end{vmatrix}.$$

Выкалывание символов является операцией, обратной расширению.

Удлинение кода означает добавление в базис кода вектора $[1\ 1\ 1 \dots 1]$ и выполнение операции расширения. Порождающая матрица удлиненного кода $G_{удл}$ может быть записана в виде:

$$G_{удл} = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & & & & \\ 0 & & [G] & & \\ 0 & & & & \end{vmatrix}.$$

Увеличение числа векторов в базисе означает увеличение размерности кода на 1, а добавление единичного вектора – включение в код дополнительно всех инверсий кодовых слов, что, конечно, не может увеличить минимального расстояния. Понятно, что если кодовое расстояние исходного кода было четным, то в результате удлинения оно не изменится.

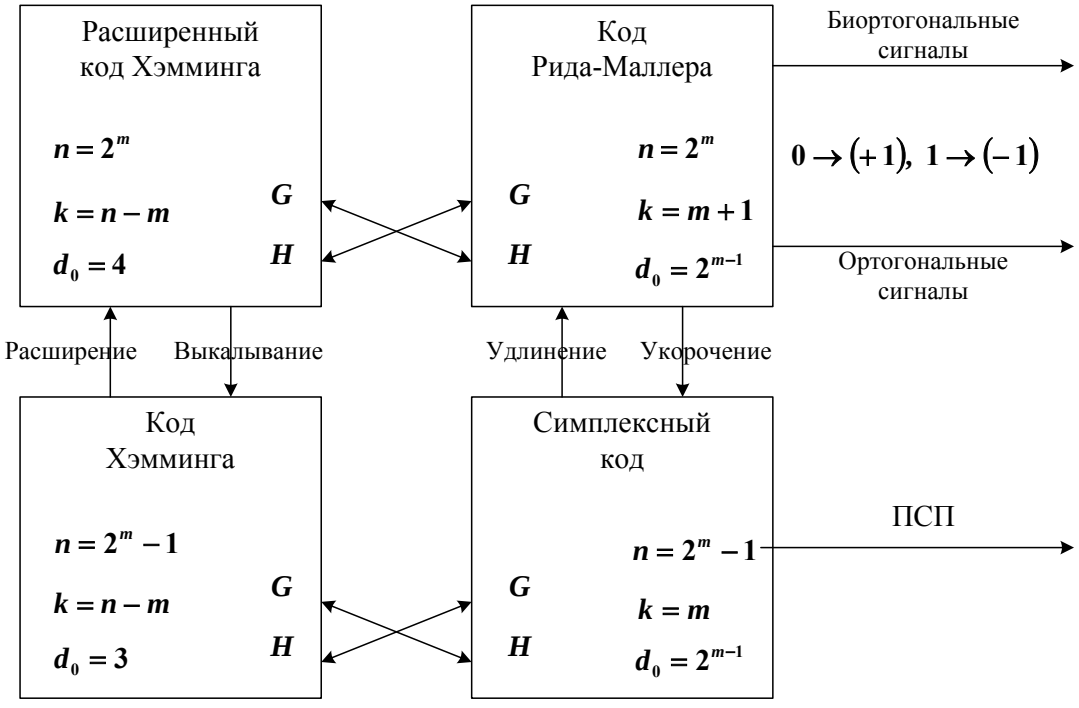


Рис. 5.6. Операции над кодами

Укорочение кода – операция обратная удлинению. Она состоит в выборе из кода тех слов, у которых первый символ равен 0, и последующему исключения этого равного для всех слов символа.

Из рис.5.6 видно, что расширенный код Хэмминга и код Рида-Маллера, а также код Хэмминга и симплексный код дуальны, так как проверочные матрицы H первых кодов в каждой паре являются порождающими G для вторых и наоборот.

При замене двоичных символов 0 на (+1), а 1 на (–1) совокупность слов Рида-Маллера преобразуется в множество биортогональных сигналов, включающее все функции Уолша и противоположные им. Если при замене двоичных символов ограничиться подпространством кода Рида-Маллера при $a_0 = 0$, то получим множества ортогональных сигналов. Применение указанной замены двоичных символов на (+1) и (–1) в ненулевых словах симплексного кода дает m-последовательности разных сдвигов.

5.8. Сверточные коды

5.8.1. Основные параметры

Сверточные коды относятся к непрерывным рекуррентным кодам. Они называются непрерывными, так как последовательность информационных символов при кодировании не разбивается на блоки. Теоретически проверочные символы могут зависеть от неограниченно удаленных информационных. Это позволяет считать сверточные коды обобщением блочных.

Рекуррентными эти коды называются потому, что соотношения, связывающие проверочные символы с информационными, справедливы для любого участка информационной последовательности. В сверточных кодах так же, как и в блочных, выделяют классы систематических и несистематических кодов. Напомним, что в словах систематического кода известны позиции с информационными и проверочными символами.

Термин «сверточные коды» объясняется тем, что кодовое слово можно рассматривать как свертку отклика линейной системы (кодера) и входной ин-

формационной последовательности. Поэтому сверточные коды являются линейными, для которых сумма любых кодовых последовательностей также является кодовой последовательностью.

Структура слова систематического сверточного кода схематично изображена на рис. 5.7.

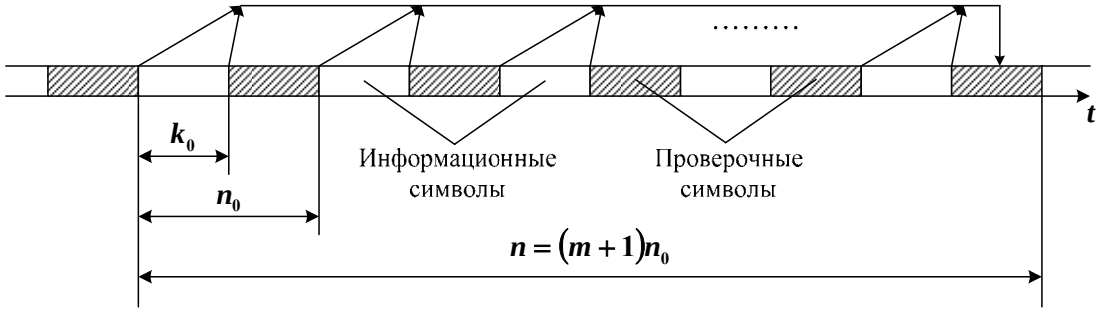


Рис. 5.7. Структура слова сверточного кода

Не заштрихованные участки отмечают позиции, на которых расположены информационные символы (кадры), а заштрихованные – позиции с проверочными символами. Таким образом, кодовое слово состоит из элементарных блоков длиной n_0 с k_0 информационными символами, что обеспечивает скорость k_0/n_0. При анализе сверточных кодов удобно считать, что информационные символы, стоящие на одинаковых позициях в элементарных блоках, принадлежат одной информационной последовательности. Таким образом, при использовании кода k_0/n_0 производится кодирование k_0 информационных последовательностей, которые формируются, например, путем периодического подключения источника информации на k_0 входных шин кодера.

Стрелками на рис. 5.7 условно показаны связи между проверочными и информационными символами. Эти связи имеют место для любого участка кодовой последовательности. Поскольку технически реализация кодеров возможна при ограниченном объеме запоминающих устройств, то влияние информационного символа на проверочные распространяется также на конечное число позиций.

На практике используются различные определения длин сверточных кодов. Следуя [3], назовем длиной кодового ограничения величину $\upsilon = mk_0$, где m – число кадров, хранящихся в кодере. Число позиций между информационным и максимально удаленным зависимым от него проверочным символом называется кодовой длиной блока: $n = (m+1)n_0$.

Вследствие того, что проверочный символ зависит от информационных из $m+1$ кадров, для построения хороших сверточных кодов не требуется увеличивать длину элементарного блока. Это определяет основное отличие сверточных кодов от блочных и их достоинства. Так, кодеры и декодеры сверточных кодов, исправляющих однократные ошибки и пакеты ошибок, имеют меньшую сложность реализации.

5.8.2. Способы задания сверточного кода

Для получения слова систематического кода надо по известным информационным символам найти проверочные и расположить их на заданных позициях элементарного блока. Для простоты положим $k_0 = 1$, т. е. будем кодировать одну последовательность информационных символов $a_0, a_1, a_2, ..., a_{i-1}, a_i, ...$, которая представляется многочленом

$$A(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{i-1} x^{i-1} + a_i x^i + ...$$

Требуется найти последовательности проверочных символов, которые также можно описать многочленами:

$$B_1(x) = b_{10} + b_{11} x + b_{12} x^2 + ... + b_{1i} x^i + ...,$$

$$B_2(x) = b_{20} + b_{21} x + b_{22} x^2 + ... + b_{2i} x^i + ...,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$B_{n_0-1}(x) = b_{n_0-10} + b_{n_0-11} x + b_{n_0-12} x^2 + ... + b_{n_0-1i} x^i + ...$$

Укажем некоторые способы задания сверточных кодов, которые во многом напоминают способы задания блочных.

1. Проверочные символы определяются по известным информационным с помощью $n_0 - 1$ рекуррентных соотношений:

$$b_{1i} = \sum_{j \le i} \alpha_j a_j \,,$$

$$b_{2i} = \sum_{j \le i} \beta_j a_j \,,$$

$$\cdots\cdots\cdots\cdots$$

$$b_{n_0-1i} = \sum_{j \le i} \gamma_j a_j \,.$$

где α, β, ... ,γ – известные двоичные коэффициенты; суммирование проводится по модулю 2.

2. Последовательность проверочных символов $B_j(x)$ находится с помощью порождающего многочлена

$$g_j(x) = g_o + g_1 x + g_2 x^2 + \ldots + g_m x^m$$

по формуле

$$B_j(x) = A(x) g_j(x), \qquad (5.39)$$

Отсюда следует, что проверочный символ b_{ji} является сверткой информационной последовательности и коэффициентов порождающего многочлена

$$b_{ji} = g_o a_i + g_1 a_{i-1} + g_2 a_{i-2} \ldots + g_m a_{i-m} \,. \qquad (5.40)$$

Как уже указывалось, соотношение (5.40) дало название этому классу кодов. Для задания всех проверочных последовательностей требуется $n_0 - 1$ порождающий многочлен. Максимальная степень многочлена определяет число кадров, хранимых в кодере, т.е. длину кодового ограничения.

Если кодируется k_0 информационных последовательностей $A_1(x), A_2(x), \ldots, A_k(x)$, то систематический сверточный код со скоростью k_0 / n_0 задается с помощью $k_0(n_0 - k_0)$ порождающих многочленов, а несистематический – с помощью $k_0 n_0$ многочленов. Порождающие многочлены хороших сверточных кодов приведены, например, в работе [3].

3. Сверточный код задается с помощью графа и кодовой решетки. Кодовое слово изображается последовательным соединением ребер графа, структура которого похожа на ветвящееся дерево. Поэтому сверточные коды относятся к классу древовидных.

4. Для сверточного кода могут быть построены порождающая и проверочная матрицы. По сравнению с теми же матрицами блочного кода эти матрицы (как кодируемая и принятая последовательности) полубесконечны.

Пример 5.15. Пусть требуется применить простейший систематический сверточный код с параметрами: $k_0 = 1$, $n_0 = 2$, кодовым ограничением $m = 1$, скоростью $k_0/n_0 = 1/2$ бит на один символ. Кодовое слово является полубесконечной последовательностью символов $a_0b_0, a_1b_1, a_2b_2, ..., a_{i-1}b_{i-1}, a_ib_i, ...$, состоящей из элементарных блоков a_ib_i. Рассмотрим способы задания этого кода.

1. Проверочный символ определяется через информационные линейным рекуррентным соотношением

$$b_i = a_i + a_{i-1}. \qquad (5.41)$$

Видно, что значение проверочного символа зависит от значений информационных символов, входящих в текущий и предыдущий кадры. Предыдущий информационный символ должен быть запомнен в кодере, для чего необходим один двоичный элемент памяти.

2. Порождающий многочлен этого кода $g(x) = g_o + g_1x = 1 + x$. С помощью заданного многочлена $g(x)$ найдем проверочные символы для следующей информационной последовательности $a_0 = 0$, $a_1 = 1$, $a_2 = 0$, $a_3 = 1$, $a_4 = 1, ...$, которая описывается многочленом $A(x) = x + x^3 + x^4 + ...$.

Используя (5.39), получим $B(x) = A(x) + xA(x) = x + x^3 + x^4 + x^2 + x^4 + x^5 + ...$
$... = x + x^2 + x^3 + ...$, что соответствует $b_0 = 0$, $b_1 = 1$, $b_2 = 1$, $b_3 = 1$, $b_4 = 0$.

3. Графические способы задания поясним по схеме кодера, для построения которого следует пользоваться заданием сверточного кода с помощью соотношения (5.41) или порождающего многочлена $g(x) = 1 + x$.

Схема кодирующего устройства приведена на рис. 5.8. Кодер представляет собой линейную систему (со сложением по модулю 2), отклик которой на входную единицу равен (11), что соответствует коэффициентам многочлена $g(x)$. При подаче на вход информационной последовательности выходная по-

следовательность образуется в соответствии с (5.41). Слово систематического сверточного кода формируется с помощью электронного ключа S, который поочередно подключает шины информационных и проверочных символов к выходу. Такт работы ключа в два раза меньше такта поступления и сдвига информационных символов.

Схема на рис. 5.8 позволяет пояснить задание сверточного кода с помощью решетки. *Решеткой* называется граф, узлы которого находятся в полубесконечной прямоугольной координатной сетке и связаны ребрами. Число узлов в каждом столбце конечно, а конфигурация ребер, соединяющих узлы каждого столбца с узлами следующего столбца, одинакова для всех столбцов. Каждый столбец отображает набор возможных состояний кодера. Поэтому ребра показывают изменение состояния кодера при подаче на вход новою информационного символа. Маркировка ребер соответствует последовательности кодовых символов, передаваемых в канал связи, т.е. комбинации элементарного блока из n_0 символов.

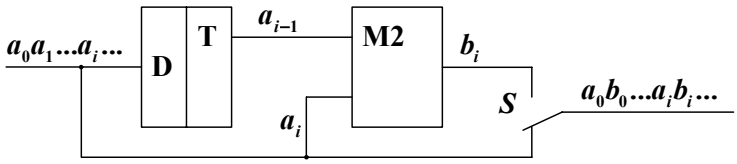


Рис. 5.8. Кодер сверточного кода 1/2

Для рассматриваемого примера кодовая решетка изображена на рис. 5.9 в предположении, что верхнее ребро, исходящее из каждого узла, соответствует поступающему информационному символу $a_i = 0$, а нижнее $a_i = 1$. Поэтому на первых двух фрагментах решетки верхние ребра помечены наборами $0b_0$ и $0b_1$, а нижние – $1b_0$ и $1b_1$. Для получения маркированной решетки требуется определить согласно (5.41) значения проверочных символов. Результаты вычислений отражены на правых фрагментах решетки.

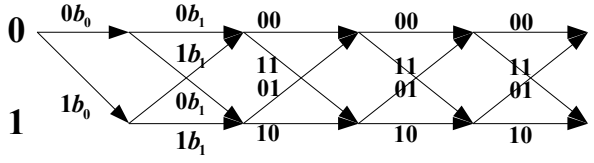


Рис. 5.9. Кодовая решетка

Решетка описывает код в том смысле, что каждой последовательности информационных символов соответствует свой путь по решетке, а маркировка ребер, составляющих этот путь, дает кодовое слово. Так, ранее рассмотренной

последовательности 01011 соответствует путь и кодовое слово 00 11 01 11 10, в котором последовательность проверочных символов совпадает с получаемой согласно (5.40).

Кодовая решетка удобна для наглядного пояснения принципов декодирования сверточных кодов. Напомним, что в двоичном симметричном канале оптимальной оценкой переданного кодового слова является слово, ближайшее к принятой комбинации, т.е. путь по кодовой решетке, отстоящий от этой последовательности на минимальное расстояние Хэмминга. Поиск такого пути и составляет сущность алгоритма декодирования Витерби, который рассматривается в следующем параграфе.

Пример 5.16. Пусть дан несистематический сверточный код с параметрами $k_0 = 1$, $n_0 = 2$, скоростью $1/2$, кодовым ограничением $m = 2$. Он задается многочленами $G_1(x) = x^2 + x + 1$ и $G_2(x) = x^2 + 1$.

Схема кодера приведена на рис.5.10. В отличие от примера 5.14 кодовое слово состоит из элементарных блоков, не содержащих информационные символы. Первые символы каждого блока s_{1i} являются сверткой информационной последовательности и многочлена $G_1(x)$, а вторые s_{2i} - сверткой той же последовательности и многочлена $G_2(x)$. Заметим, что если один из многочленов приравнять 1, то получим систематический сверточный код.

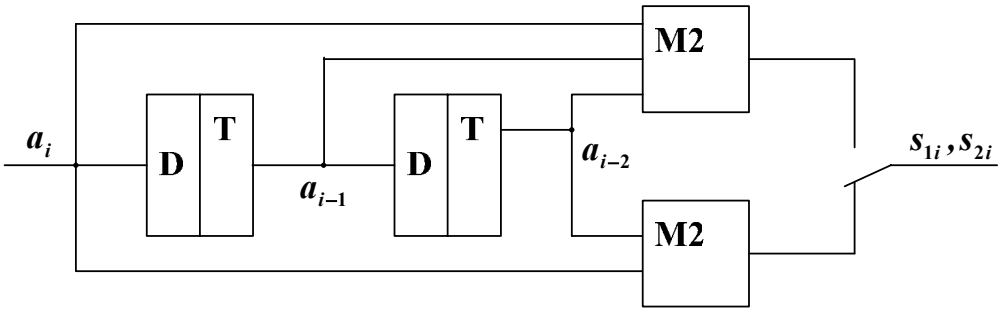


Рис. 5.10. Схема кодера

Кодовая решетка, построенная по правилам примера 5.14, показана на рис. 5.11.

Кодер имеет 2-разрядный сдвигающий регистр с четырьмя состояниями, следовательно, столбцы решетки содержат по четыре узла, помеченных содер-

жимым регистра. Поэтому левый символ в обозначении узла равен последнему информационному символу, поступившему в регистр. Порядок обозначения узлов выбран так, что при $a_i = 0$ регистр переходит в следующее состояние по верхнему ребру, а при $a_i = 1$ – по нижнему. Маркировка ребер совпадает с комбинацией элементарного блока, посылаемого в канал. По-прежнему информационной последовательности соответствует путь на кодовой решетке и кодовое слово. Если входные символы 0 1 1 0 , то по решетке находим кодовое слово 00 11 01 01.

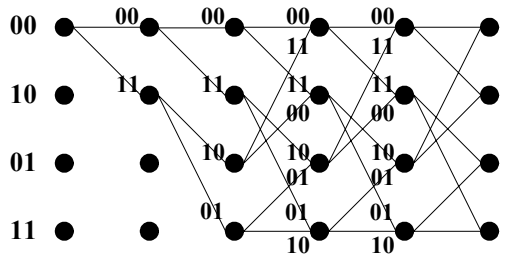


Рис. 5.11. Решетка кода

5.8.3. Алгоритм декодирования Витерби

Для сверточных кодов разработаны алгоритмы синдромного, последовательного декодирования и декодирования по максимуму правдоподобия (алгоритм Витерби). На практике широко используется последний метод, что объясняется простотой реализации при небольших длинах кодового ограничения и получаемым выигрышем от кодирования.

Алгоритм Витерби является рекуррентной процедурой, направленной на поиск пути по кодовой решетке, ближайшего к принимаемой последовательности. Как уже указывалось, декодирование по минимуму расстояния является оптимальным в канале с независимыми ошибками. Основные операции алгоритма поясним при декодировании кода примера 5.15.

Пусть для простоты передается нулевое кодовое слово, а в канале произошла трехкратная ошибка, так что принятая последовательность имеет вид 10 10 00 00 10 00 ... 00 ... Результаты поиска ближайшего пути после приема 14 элементарных блоков показаны на рис. 5.12. Промежуточные этапы работы декодера при сделанных предположениях подробно рассмотрены в [3].

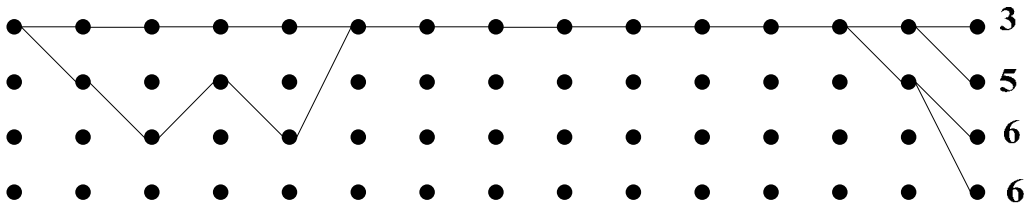


Рис. 5.12. Пример работы алгоритма Витерби

На правой части рисунка видны четыре пути, ведущие в каждый узел решетки. Рядом проставлены расстояния (меры расходимости) этих путей от принятой последовательности на отрезке из 14 блоков. Мера верхнего пути значительно меньше мер нижних. Поэтому можно предположить, что верхний путь наиболее вероятен. Однако декодер Витерби, не зная следующих фрагментов принимаемой последовательности, вынужден запомнить все четыре пути на время приема L элементарных блоков. Число L называется шириной окна декодирования. Понятно, что для уменьшения ошибки декодирования следует выбирать L достаточно большим, в несколько раз превышающим длину блока, что, естественно, усложняет декодер. В данном случае $L = 15$.

Отметим, что тактика выбора и последующего анализа только одного пути с наименьшим расстоянием составляет сущность более экономного последовательного декодирования [3, 26].

На средней части рис. 5.12 видно, что все пути имеют общий отрезок и, следовательно, прием новых блоков не может повлиять на конфигурацию этого участка наиболее правдоподобного пути. Поэтому декодер уже может принимать решение о значении информационных символов, соответствующих этим элементарным блокам. Поскольку рассматриваемый отрезок составлен из верхних ребер кодовой решетки, то согласно правилу ее построения оценки информационных символов равны 0.

Левая часть рисунка демонстрирует возможную ситуацию неисправляемой ошибки. Существует два пути с одинаковыми мерами расходимости. Декодер может разрешить эту неопределенность двумя способами. Отметить этот участок как недостоверный или принять одно из двух решений: информацион-

ная последовательность равна 00000... или 10100.... Очевидно, расширение окна декодирования не позволяет исправить такую ошибку. Ее исправление возможно при использовании кода с большей корректирующей способностью.

Поступление из канала нового элементарного блока вызывает сдвиг картинки в окне декодирования влево. В результате левое ребро пути исчезает, а справа появляется новый столбец решетки, к узлам которого должны быть продолжены сохраненные пути от узлов предыдущего столбца. Для этого выполняются следующие операции.

1. Для каждого узла нового столбца вычисляются расстояния между принятым блоком и маркировкой ребер, ведущих в данный узел.

2. Полученные меры расходимости ребер суммируются с расстоянием путей, которые они продолжают.

3. Из двух возможных путей оставляется путь с меньшим расстоянием, а другой отбрасывается, так как следующие поступающие блоки не могут изменить соотношения расстояний этих путей. В случае равенства расстояний или случайно выбирается один путь, или сохраняются оба.

В результате этих операций к каждому узлу нового столбца вновь ведет один путь. Например, пусть новый блок из канала равен 00. Рассмотрим продолжение пути к нижнему узлу решетки, в который можно попасть из состояния кодера 10 по ребру 01 или из состояния 11 по ребру 10 (рис. 5.11). В обоих случаях расстояние этих ребер от принятого блока 00 равно 1. Однако суммарное расстояние пути, продолженного из состояния 10, равно 6, а пути из состояния 11 равно 7. Поэтому второй путь будет отброшен вместе с ребром 01, которое входило в нижний узел на предыдущем шаге декодирования (рис. 5.12). Оценка информационного символа производится по левому ребру пути, находящемуся в окне декодирования. Согласно правилу построения кодовой решетки принимается, что информационный символ равен 0, если ребро верхнее, и равен 1, если ребро нижнее.

5.9. Помехоустойчивость систем передачи дискретных сообщений

5.9.1. Две процедуры приема сигналов

При использовании помехоустойчивых кодов передаваемые сообщения $X(t)$ предварительно дискретизируются по уровню и по времени. В интервале кодового слова $T = n\tau$ передается одно из дискретных сообщений $X_i, i = \overline{1, M_x}$. Для передачи используется, соответственно, M_x разных сигналов $S_i(t), i = \overline{1, M_x}$.

Принимаемый сигнал $y(t)$ во многих случаях можно представить в виде суммы

$$y(t) = S_i(t) + n(t), \quad 0 \le t \le T. \tag{5.42}$$

полезного сигнала и аддитивного гауссовского белого шума $n(t)$ со спектральной плотностью N_0. Задача приема заключается в том, чтобы по наблюдениям $Y_T = (y(t), 0 \le t \le T)$ входного процесса $y(t)$ на интервале $(0, T)$ определить, какой из сигналов $S_i(t)$ был передан.

Выбор наилучшего (оптимального) приемника требует введения критерия, на основе которого можно сравнивать качество различных методов. В рассматриваемом случае передачи дискретных сообщений таким критерием качества может служить вероятность ошибки, т.е. вероятность того, что после анализа реализации $\mathrm{Y_T}$ входного процесса, содержащей сигнал $S_i(t)$, будет принято неверное решение о передаче сигнала $S_j(t), j \ne i$.

Оптимальный приемник обеспечивает минимальную вероятность ошибки или, что то же самое, максимальную вероятность правильного решения $p(S_i / Y_T)$, вычисленную для данной реализации (максимальную апостериорную вероятность сигнала $S_i(t)$).

Введенный критерий качества позволяет указать наилучший способ обработки реализации входного процесса. Действительно, необходимо вычислить все апостериорные вероятности $p(S_i / Y_T)$ и сравнить их между собой. Затем следует выбрать сигнал S_j, апостериорная вероятность $p(S_i / Y_T)$ которого мак-

симальна, а вероятность ошибки $1-p(S_i/Y_T)$ минимальна. Для вычисления $p(S_i/Y_T)$ воспользуемся формулой Байеса

$$p(S_i/Y_T)=\frac{1}{w(Y_T)}p(S_i)w(Y_T/S_i), \qquad (5.43)$$

где $p(S_i)$ – априорная вероятность использования сигнала S_i для передачи сообщения; $w(Y_T)$ и $w(Y_T/S_i)$ – безусловная и условная плотности распределения вероятностей Y_T, которые при известной (принятой) реализации входного процесса Y_T превращаются в конкретные числа. Величина $w(Y_T)$ не зависит от S_i и может не учитываться при сравнении апостериорных вероятностей. Напротив, функция $w(Y_T/S_i)$ играет основную роль в задаче построения оптимального приемника. Эта функция после подстановки отрезка реализации Y_T входного процесса зависит лишь от переданного сигнала S_i и называется функцией правдоподобия.

Для того чтобы подчеркнуть, что $w(Y_T/S_i)$ после приема сигнала и подстановки Y_T не является плотностью распределения, введем ее новое обозначение:

$$L(S_i)=w(Y_T/S_i).$$

Таким образом, оптимальный прием может быть осуществлен на основе вычисления $L(S_i)$ и выбора максимального значения произведения $p(S_i)L(S_i)$ При равных вероятностях $p(S_i)=1/M_x$, $i=\overline{1,M_x}$, оптимальный прием базируется на вычислении лишь функции правдоподобия и определении ее максимума. Такой приемник называется приемником максимального правдоподобия.

Из статистической радиотехники [20] известно, что для модели (5.42)

$$L(S_i)=C\exp\left(-\frac{1}{N_0}\int_0^T(y(t)-S_i(t))^2dt\right).$$

Максимум функции правдоподобия достигается при таком сигнале S_i, для которого минимально значение интеграла

$$d_i^2(Y_T,S_i)=\int_0^T(y(t)-S_i(t))^2dt\,. \qquad (5.44)$$

Полученный результат имеет ясную геометрическую трактовку. Действительно, $d_i^2(Y_T, S_i)$ является расстоянием между функциями $Y_T = (y(t), t \in (0,T))$ и $S_i(t), t \in (0,T)$ в гильбертовом пространстве со среднеквадратической метрикой. С этой точки зрения оптимальное решение по наблюдениям Y_T заключается в выборе такого из возможных переданных сигналов $S_i(t), i = \overline{1, M_x}$, который находится ближе других к Y_T. Рис 5.13 иллюстрирует оптимальный выбор S_2.

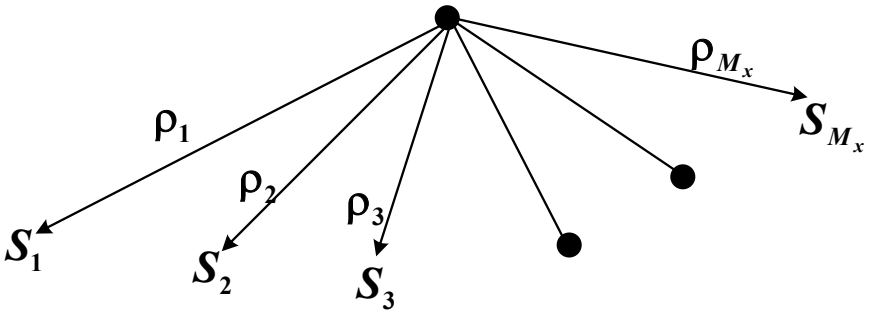


Рис. 5.13. Геометрическая интерпретация приема сигналов

Преобразуем теперь (5.44) к следующему виду:

$$d_i^2(Y_T, S_i) = \int_0^T y^2(t)dt - 2\int_0^T y(t)S_i(t)dt + \int_0^T S_i^2(t)dt .$$

Первый интеграл не зависит от i, а при равных энергиях сигналов $\int_0^T S_i^2(t)dt$ минимум d_i^2 достигается при максимальном значении корреляционного интеграла:

$$\lambda_i = \int_0^T y(t)S_i(t)dt, i = \overline{1, M_x} . \qquad (5.45)$$

Структурная схема приемника (рис. 5.14), использующего вычисления по формуле (5.45), включает набор корреляторов и блок выбора наибольшего из чисел $\lambda_i, i = \overline{1, M_x}$.

На выходе приемника вырабатывается наилучшая с точки зрения минимума вероятности ошибки оценка $\hat{\epsilon}_j$ переданного сообщения по принятой реализации Y_t. Рассмотренную процедуру называют приемом сигналов в целом.

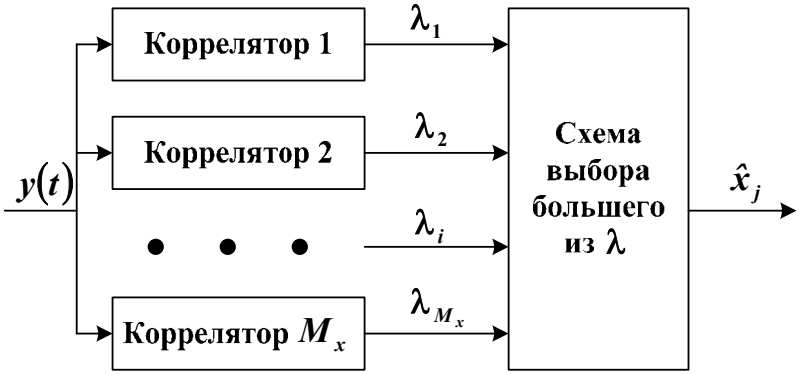


Рис. 5.14. Оптимальный приемник

При переходе к кодам значительной длины число используемых в системе сигналов и соответствующее число корреляторов растет очень быстро $\left(M_x = 2^k\right)$, и сложность всего устройства может оказаться неприемлемо большой.

Существенного упрощения можно достичь, используя посимвольный прием сигналов, то есть разбивая процедуру приема на две части:

оптимальный прием каждого символа и принятие решения о его значении;

декодирование полученных кодовых слов. В этом случае структура приемника наиболее проста (рис. 5.15).

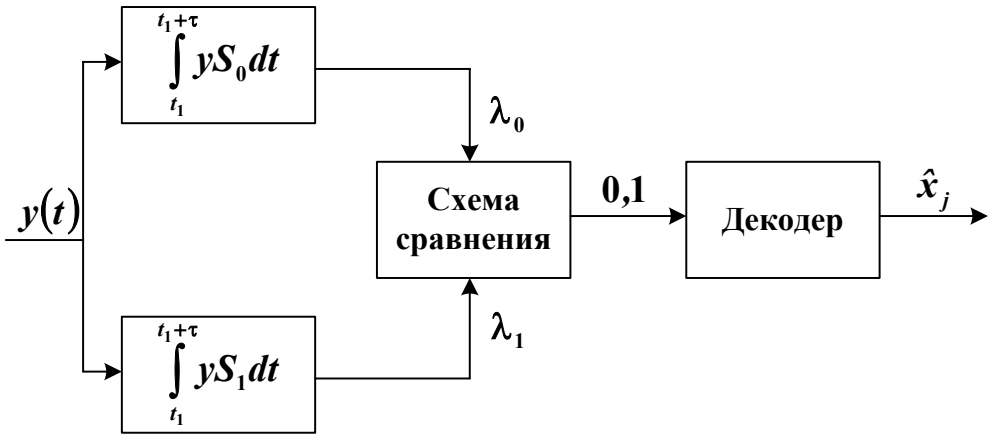


Рис. 5.15. Посимвольный прием

Для бинарного канала связи приемник содержит два коррелятора, схему сравнения и декодер. На выходе схемы сравнения появляется сигнал «0», если $\lambda_0 > \lambda_1$, или сигнал «1», если $\lambda_0 \le \lambda_1$. Возможно и дальнейшее упрощение схемы за счет объединения двух корреляторов в один, используя в качестве опорного

сигнала разность $S_0(t)-S_1(t)$.

Платой за существенное снижение сложности алгоритма при переходе от оптимального к посимвольному приему является увеличение вероятности ошибки. Поэтому нашей очередной задачей будет анализ характеристик рассмотренных алгоритмов и их сравнение при различных видах кодов.

5.9.2. Помехоустойчивость систем передачи информации при оптимальной процедуре приема

Пусть по каналу связи передается сигнал $S_i(t)$. Очевидно, прием будет правильным, когда выходной сигнал λ_i коррелятора (рис. 5.14) окажется наибольшим. Поэтому вероятность $p_{ПРi}$ правильного приема сигнала $S_i(t)$ может быть найдена как вероятность совместного выполнения системы неравенств $\lambda_i > \lambda_j; i,j=\overline{1,M_x}, i\neq j$. При фиксированном значении λ_i легко находится условная вероятность выполнения этой системы неравенств:

$$p_{ПРi}(\lambda_i)=\int_{-\infty}^{\lambda_i}\ldots\int_{-\infty}^{\lambda_i} w(\lambda_1,\ldots,\lambda_{M_x}/\lambda_i)d\lambda_1\ldots d\lambda_{i-1}d\lambda_{i+1}\ldots d\lambda_{M_x},$$

где $w(\lambda_1,\ldots,\lambda_{M_x}/\lambda_i)$ – условная плотность распределения $\lambda_1\ldots\lambda_{i-1}\lambda_{i+1}\ldots\lambda_{M_x}$ при фиксированном λ_i.

Безусловная вероятность правильного приема сигналов в системе находится с помощью усреднения $p_{ПРi}(\lambda_i)$ по λ_i и i:

$$p_{ПР}=\sum_{i=1}^{M_x} p(S_i)\int_{-\infty}^{+\infty}\int_{-\infty}^{\lambda_i}\ldots\int_{-\infty}^{\lambda_i} w(\lambda_1,\ldots,\lambda_{M_x})d\lambda_1 d\lambda_2\ldots d\lambda_{M_x}. \qquad (5.46)$$

В общем случае интегрирование (5.46) является сложной математической задачей. Поэтому ограничимся рассмотрением частного случая ортогональных сигналов, а также близких к ним симплексных сигналов, для которых величины λ_i и $\lambda_j, j\neq i$ независимы. В этом случае совместная плотность распределения может быть записана в виде произведения:

$$w(\lambda_1,\ldots,\lambda_{M_x})=\prod_{j=1}^{M_x} w(\lambda_j).$$

Вследствие свойства эквидистантности ортогональных кодов вероятности $p_{ПРi}$ равны между собой, следовательно,

$$p_{ПРi}(\lambda_i) = \int_{-\infty}^{+\infty} w(\lambda_i) \left(\int_{-\infty}^{+\infty} w(\lambda) d\lambda \right)^{M_x - 1} d\lambda_i ,$$

где $w(\lambda_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-(\lambda_i - U_c)^2 \Big/ 2\sigma^2 \right]$; $w(\lambda) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\lambda^2 \Big/ 2\sigma^2 \right]$; U_c – сигнальная составляющая на выходе i-го коррелятора; σ^2 – дисперсия шумовой составляющей на выходах корреляторов.

Переходя к новым переменным $t_1 = \lambda / \sigma$ и $t_1 = (\lambda_i - U_c) / \sigma$, преобразуем выражение для вероятности ошибки к виду:

$$p_{ПР} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2} \right) \Phi^{M_x - 1} \left(t + \frac{U_c}{\sigma} \right) dt .$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{t_1^2}{2} \right) dt_1$ – интеграл вероятности. Величина $\frac{U_c}{\sigma}$ (отношение сигнал/шум) может быть найдена следующим образом. При действии белого шума выходной сигнал коррелятора совпадает с выходным сигналом согласованного фильтра. Следовательно, можно воспользоваться известной формулой для отношения сигнал/шум на выходе согласованного фильтра: $U_c / \sigma = \sqrt{2E / N_0}$, где E – энергия сигнала. В теории связи обычно вводится параметр $E_0 = E / k$, характеризующий энергию сигнала на один бит передаваемой информации при числе информационных символов k. Окончательно $U_c / \sigma = \sqrt{2h_0 k}$, где $h_0 = E_0 / N_0$ – параметр, принятый в системах связи.

Выражение для вероятности правильного приема сигнала

$$p_{ПР} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2} \right) \Phi^{M_x - 1} \left(t + \sqrt{2h_0 k} \right) dt . \qquad (5.47)$$

часто называется интегралом В.А. Котельникова. Можно показать, что для симплексных сигналов справедливо следующее выражение:

$$p_{ПР} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2}\right) \Phi^{M_x - 1}\left(t + \sqrt{2h_0 k(1-\rho)}\right) dt . \qquad (5.48)$$

где ρ – коэффициент корреляции сигналов $S_i(t)$ и $S_j(t), j \neq i$.

Сравнение помехоустойчивости систем передачи информации, использующих разные коды, по величине $p_{ПР}$ не всегда удобно, так как коды могут иметь разное число k информационных символов. С изменением k меняется как $p_{ПР}$, так и количество передаваемой информации. Поэтому вероятность правильного приема приводится к одному биту передаваемой информации, для чего вводят новую характеристику $Q_Э$, равную эквивалентной вероятности искажения одного бита информации. Реальный канал связи заменяется эквивалентным каналом без избыточности, но так, чтобы вероятности $p_{ПР}$ были одинаковы в обоих каналах. В системе без избыточности $p_{ПР} = (1 - Q_Э)^k$ поэтому

$$Q_Э = \sqrt[k]{p_{ПР}} \approx \frac{1 - p_{ПР}}{k} . \qquad (5.49)$$

Для анализа помехоустойчивости при разных значениях параметра h_0, имеющего смысл приведенного отношения сигнал/шум, по формулам (5.47), (5.48) можно вычислить вероятности $p_{ПР}$ и найти $Q_Э$ с помощью (5.49).

На рис. 5.16 сплошными кривыми представлены результаты таких расчетов для случаев $k = n$ (безызбыточное кодирование), а также симплексных кодов (7,3), (15,4), (1023,10). Как следует из рисунка, с ростом числа k информационных разрядов вероятность $Q_Э$ ошибки монотонно падает.

Представляет интерес предел $Q_Э$ при $k \to \infty$, что соответствует переходу ко все более сложным кодам. Для его определения заметим, что функция $\Phi(z)^{M_x - 1}$ при увеличении $k = \log_2 M_x$ приближается по форме к единичному скачку в некоторой точке z_0, значение которой определяется из уравнения $\Phi(z)^{M_x - 1} = 0{,}5$ и равно $z_0 \approx \sqrt{2k \ln 2}$. При $k \to \infty$ формула (5.47) приводится к виду:

$$p_{ПР} = \lim_{k \to \infty} \frac{1}{\sqrt{2\pi}} \int_{z_0 - \sqrt{2h_0 k}}^{+\infty} \exp\left(-\frac{t^2}{2}\right) dt = \begin{vmatrix} 1, h_0 > \ln 2, \\ 0, h_0 \leq \ln 2. \end{vmatrix}$$

Итак, при возрастании числа k информационных разрядов в коде величина

$Q_Э$ монотонно падает, если $h_0 > \ln 2$. При достаточно большом k система приобретает пороговый эффект (рис. 5.16, кривая $k \to \infty$); при $h_0 > h_{пор} = \ln 2$, $Q_Э \to 0$; при $h_0 \le h_{пор} = \ln 2$, $Q_Э \to 0{,}5$.

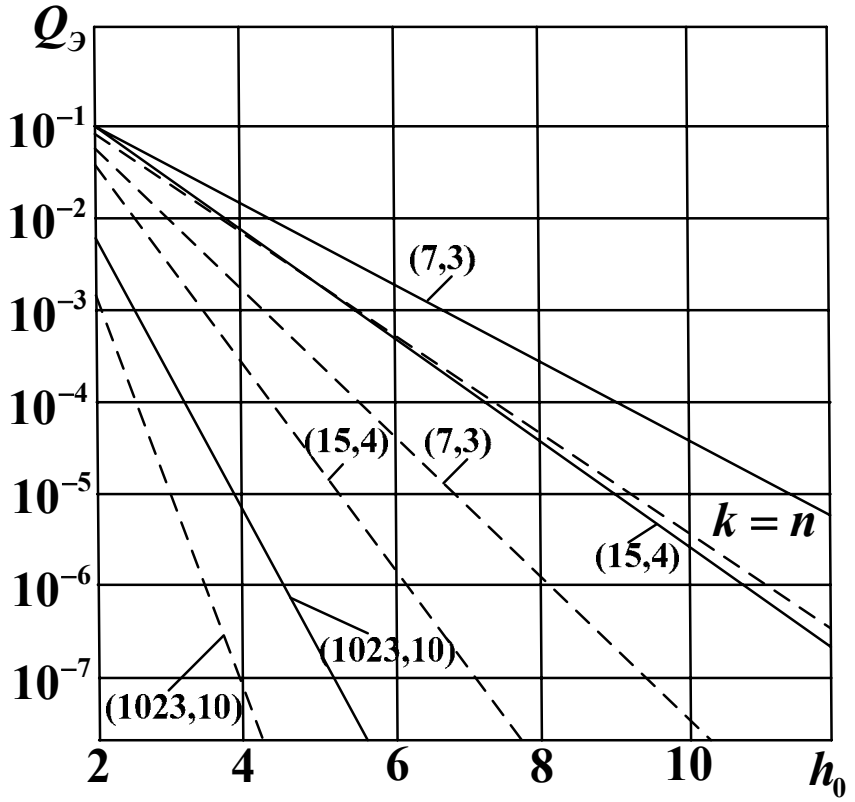


Рис. 5.16. Зависимости вероятности ошибочного декодирования при оптимальном приеме (пунктир) и посимвольном приеме (сплошная линия)

В диапазоне $h_0 > \ln 2$ возможна передача сообщений со сколь угодно малой вероятностью ошибки. Достигается это только ценою увеличения блока кодируемых информационных символов и соответствующего возрастания времени задержки при кодировании и декодировании, а также сложности оборудования на обеих сторонах системы связи.

5.9.3. Помехоустойчивость систем передачи информации при посимвольном приеме сигналов

Вероятность правильного приема одного символа в этом случае находится

с помощью формулы (5.48) при $M_x = 2$. При использовании для передачи сигналов фазовой манипуляции (ФМн), характеризуемой коэффициентом корреляции между сигналами $\rho = -1$,

$$p_{ПР1ФМ} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t+\sqrt{4E_1/N_0}} \exp\left(-\frac{t_i^2}{2}\right) dt_i dt, \qquad (5.50)$$

где E_1 – энергия одного символа. После замены $V = \frac{t_i + t}{\sqrt{2}}$, $V_1 = \frac{t_i - t}{\sqrt{2}}$ выражения для вероятности правильного ($p_{ПР1ФМ}$) и ошибочного ($p_{ФМ}$) приема символа при ФМн преобразуются к следующему виду:

$$p_{ПР1ФМ} = \Phi\left(\sqrt{\frac{2E_1}{N_0}}\right); \; p_{ФМ} = 1 - \Phi\left(\sqrt{\frac{2E_1}{N_0}}\right). \qquad (5.51)$$

Аналогично при использовании частотной (ЧМн) или амплитудной (АМн) манипуляций:

$$p_{ПР1ЧМ} = \Phi\left(\sqrt{\frac{E_1}{N_0}}\right); \; p_{ЧМ} = 1 - \Phi\left(\sqrt{\frac{E_1}{N_0}}\right). \qquad (5.52)$$

$$p_{ПР1АМ} = \Phi\left(\sqrt{\frac{E_1}{2N_0}}\right); \; p_{АМ} = 1 - \Phi\left(\sqrt{\frac{E_1}{2N_0}}\right). \qquad (5.53)$$

По результатам оценки каждого символа декодер выносит решение о всем принятом сигнале. Эффективность работы системы передачи информации целесообразно оценить раздельно для следующих двух вариантов организации системы передачи информации.

Случай заданного канала связи

Пусть параметры двоичного симметричного канала связи (длительность символа τ и вероятность его искажения p) являются неизменными, а при смене кода (изменении k/n) меняется скорость передачи. Следовательно, такой источник сообщений должен быть управляемым [10] в отношении своей производительности.

Вероятность правильного принятия решения декодером по всей комбинации символов равна сумме вероятностей появления ошибок, исправляемых де-

кодером, включая ошибку нулевой кратности. Для совершенных кодов

$$p_{ПР} = \sum_{i=0}^{q_{испр}} C_n^i p^i (1-p)^{n-i} ;$$

для любых других

$$p_{ПР} = \sum_{i=0}^{q_{испр}} C_n^i p^i (1-p)^{n-i} + \Delta p , \qquad (5.54)$$

где p – вероятность искажения одного символа; Δp – вероятность исправления ошибок кратности выше $q_{испр} = \left[\frac{d_0 - 1}{2} \right]$; $[\bullet]$ – означает целую часть числа; d_0 – кодовое расстояние. Процедуры исправления ошибок кратности выше $q_{испр}$ заметно усложняют декодер и поэтому реализуются не всегда.

Расчет эффективности системы передачи информации сводится к вычислению $p_{ПР}$, $Q_Э$ по формулам (5.49), (5.54) для заданных k, n и d_0. На рис. 5.17 приведены зависимости $Q_Э$ от вероятности p искажения одного символа для разных кодов.

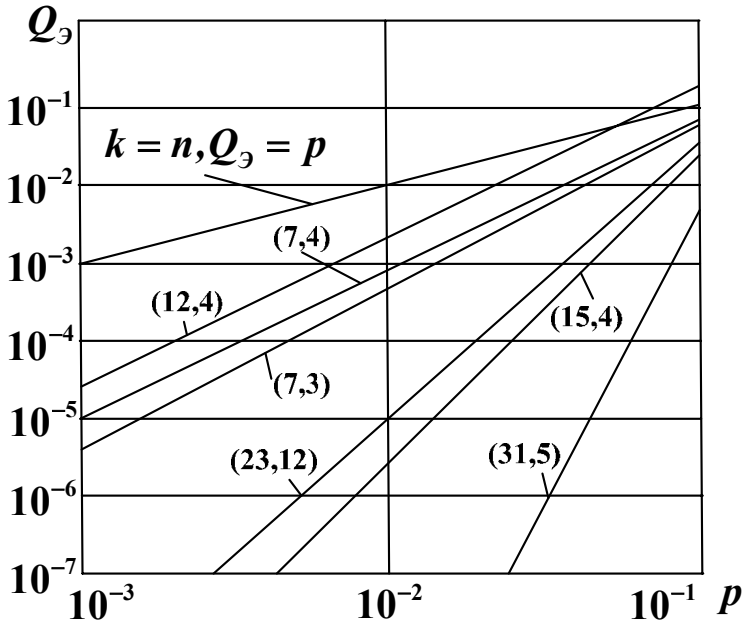


Рис. 5.17 Зависимости вероятности ошибочного декодирования при заданном канале связи

Все коды в этом случае дают заметный выигрыш по сравнению с безызбы-

точной передачей. Коды (7,3) и (7,4) близки по эффективности. Остальные коды располагаются в порядке возрастания кодового расстояния и числа избыточных символов.

Модель заданного канала связи хорошо соответствует случаю организации обмена информацией между блоками одного устройства, например, вычислительного комплекса. Из приведенных кривых следует, в частности, что применение кода (15.7) выгоднее трехкратного дублирования семи информационных символов. В этом случае выше и скорость передачи (15 тактов вместо 21), и помехоустойчивость.

Случай заданной производительности источника

Пусть источник сообщений характеризуется производительностью $H'(x)$.

Применяя для передачи тот или иной код (n,k), необходимо обеспечивать скорость передачи информации по каналу связи, соответствующую производительности $H'(x)=\frac{k}{n\tau}$, где τ – длительность одного символа.

Разные коды имеет разное отношение k/n. Следовательно, со сменой кода в общем случае должна меняться длительность символа, а значит, и полоса пропускания канала связи, и энергия на один символ $E_1=\frac{E}{n}=\frac{E_0 k}{n}$, что следует учесть в формулах (5.51), (5.52), (5.53). Так, при ФМ вероятность ошибочного опознания символа

$$p_{ФМ}=1-\Phi\left(\sqrt{\frac{2E_0k}{N_0n}}\right)=1-\Phi\left(\sqrt{2h_0\frac{k}{n}}\right). \qquad (5.55)$$

Задаваясь значениями h_0 для параметров конкретного кода k, n и d_0, по формуле (5.55) вычисляют вероятности $p_{ФМ}$ ($p_{ЧМ}$ или $p_{АМ}$) искажения одного символа в блоке посимвольного приема, а затем по (5.54) и (5.49) находят $p_{ПР}$ и $Q_Э$.

На рис. 5.16 пунктиром нанесены кривые для симплексных кодов (7,3), (15,4), (31,5), (1023,10), совершенных кодов Хэмминга (3,1) и (7,4) и Голея (23,12), БЧХ-кода (15,7) при использовании ФМн и посимвольной процедуре

приема.

Сопоставление кривых посимвольного приема с соответствующими (сплошными) кривыми оптимальной процедуры, а также безызбыточной передачи $(k = n)$ показывает следующее.

1. Применение помехоустойчивых кодов в системах с постоянной производительностью источника менее эффективно, чем при заданном канале связи (рис. 5.17). Это объясняется тем обстоятельством, что в данном случае переход к коду с большим кодовым расстоянием, а следовательно, большим числом избыточных символов, сопровождается уменьшением энергии E_1 каждого символа.

2. Посимвольный прием симплексных кодов по сравнению с оптимальной процедурой дает проигрыш 2...4 дБ в отношении сигнал/шум или 2...3 порядка в величине $Q_Э$.

3. При малом числе информационных разрядов ряд кодов дает большую вероятность ошибки, чем в отсутствии избыточности $(k = n)$. Таковы коды (3,1), (7,3); код (15,4) приносит выигрыш лишь при $h_0 > 10$.

4. Совершенные коды при посимвольном приеме дают лучшие результаты, чем симплексные. Так, код (7,4) лучше кодов (7,3) и (15,4).

Следует отметить, что при безызбыточном кодировании кодовое расстояние $d_0 = 1$, а следовательно, посимвольный прием принципиально равноценен оптимальной процедуре.

Более подробно с вопросами помехоустойчивости систем передачи информации, использующих кодирование, можно познакомиться по работам [25].

Контрольные вопросы

1. Какая задача кодирования решается кодером канала?

2. Поясните физический смысл помехоустойчивого кода?

3. На рис. 2.21 представлена частично заполненная диаграмма сверточного кода $1/2$. Сформируйте полную диаграмму состояний.

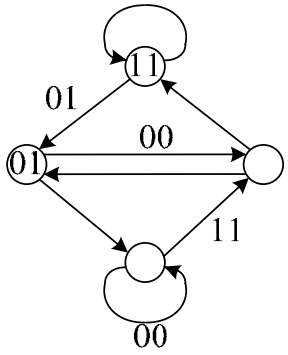


Рис. 5.21. Диаграмма состояний регистра кодера

4. Закодируйте сверточным кодом $\frac{1}{2}$, двоичную информационную последовательность: 01110011101. Покажите путь на решеточной диаграмме.

5. Запишите полиномы связи для сверточного кодера представленного на рис. 2.22.

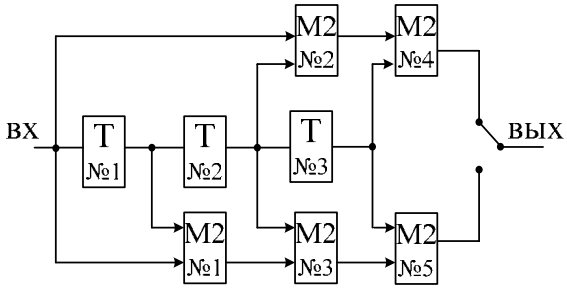


Рис. 5.22. Кодирующее устройство

6. Декодируйте сверточным кодом $\frac{1}{2}$, по алгоритму Витерби, информационную последовательность: 11110010111111011110101. Правильно ли принята эта последовательность?

ГЛАВА 6. СИГНАЛЫ С ИМПУЛЬСНОЙ МОДУЛЯЦИЕЙ

6.1. Методы импульсной модуляции

До сих пор рассматривались модулированные (манипулированные) колебания, где в качестве несущей применялось гармоническое колебание. В системах связи широко используется несущая в виде периодической последовательности видеоимпульсов. Модуляция с такой несущей называется импульсной модуляцией. Применение импульсных методов позволяет осуществить многоканальную радиосвязь с временным разделением каналов.

6.1.1. Импульсные методы передачи непрерывных сигналов

При импульсной модуляции несущее колебание имеет характер периодической последовательности импульсов (рис. 6.1,б) и может быть записано следующим выражением [32]:

$$s_{И}(t) = s_0 \sum_{k=-\infty}^{\infty} s_1(t - t_0 - kT_i), \qquad (6.1)$$

где $s_1(t)$ – функция, описывающая форму одиночного импульса, это последовательность, в которой $s_1(t)$ чаще всего прямоугольный однополярный импульс, характеризующийся параметрами:

амплитудой импульса s_0;

длительностью импульса τ_0;

начальной фазой t_0;

частотой повторения (тактовая частота) $F_i = \frac{1}{T_i}$, обычно $T_i >> \tau_0$.

В зависимости от параметра импульса, который подвергается модуляции различают четыре основных вида модуляции: амплитудно-импульсная (АИМ), длительно-импульсная модуляция (ДИМ) (широтно-импульсная (ШИМ)), частотно-импульсная (ЧИМ) и фазоимпульсная (ФИМ).

Рассмотрим пример формирования импульсно модулированных сигналов, если в качестве модулирующего сигнала взять гармоническое колебание $s_M(t)$

(рис. 6.1,а), а в качестве несущего периодическую последовательность импульсов вида (6.1) (рис. 6.1,б).

При АИМ (рис. 6.1,в) пропорционально модулирующему колебанию $s_M(t)$ изменяется амплитуда импульсов, а прочие параметры остаются неизменными.

При ДИМ (ШИМ) (рис. 6.1,г) пропорционально модулирующему колебанию $s_M(t)$ изменяется длительность (ширина) импульсов. Возможны два метода: ОДИМ (односторонняя ДИМ) и ДДИМ (двухсторонняя ДИМ). При ОДИМ перемещается один из фронтов, а второй при этом фиксированный. При ДДИМ изменяются два фронта, а положение середины не меняется.

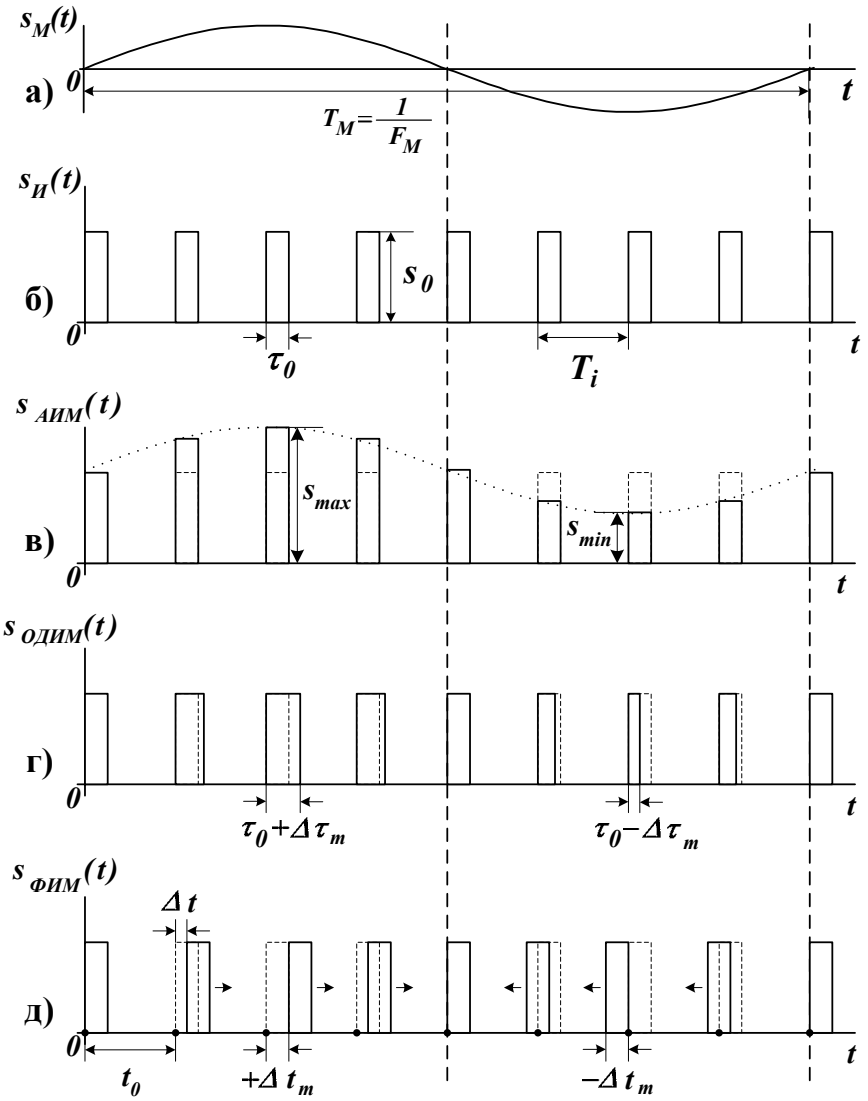


Рис. 6.1. Временные диаграммы

При ФИМ (рис. 6.1,д) пропорционально модулирующему колебанию $s_M(t)$ изменяется начальная фаза, т.е. положение импульса на временной оси относительно тактовых точек.

При ЧИМ – частота следования импульсов.

При любом виде импульсной модуляции передача непрерывного модулирующего сигнала осуществляется отдельными импульсами. Чем чаще следуют импульсы в несущем колебании, тем меньше интервал T_i, а значит, тем точнее отображается сам модулирующий сигнал.

Частота повторения импульсов находится исходя из необходимой точности восстановления непрерывного колебания $s_M(t)$ при его демодуляции. Минимальное значение частоты повторения определяется теоремой Котельникова $F_{\min} = \frac{1}{\Delta t} = 2F_B$ где F_B – максимальная частота в спектре $s_M(t)$. Так, для передачи телефонного сигнала с $F_B = 3400$ Гц значение $\Delta t \le \frac{1}{2 \cdot 3400} \le 147$ мксек. Обычно частоту следования импульсов берут с некоторым запасом $F_B = 8$ кГц, что соответствует $\Delta t = 125$ мксек. Длительность самих импульсов в современных системах связи исчисляется долями микросекунды.

Рассмотрим, как формируется импульсно-модулированное колебание.

В импульсных системах связи, используют высокочастотное колебание, при этом модуляция осуществляется в два этапа сначала модулирующее колебание $s_M(t)$ управляет информационным параметром периодической последовательности видеоимпульсов, выполняющей роль промежуточного переносчика. В результате образуется модулированная последовательность видеоимпульсов (рис. 6.1,в,г,д). На втором этапе полученная модулированная последовательность видеоимпульсов используется для манипуляции гармонического высокочастотного несущего колебания. Тем самым осуществляется перенос модулированных видеоимпульсов на частоту несущего колебания Ω_i. В этом случае получается двойная модуляция. Чаще всего применяется вторичная амплитудная манипуляция. Колебания с двойной модуляцией сокращенно обозначаются пятью буквами, например ФИМ–АМ (фазоимпульсная модуляция и вторичная амплитудная манипуляция), АИМ–АМ.

При приеме выделение модулирующего сигнала $s_M(t)$ осуществляется также в два этапа. Сначала радиоимпульсы подвергаются демодуляции, т.е. преобразуются в короткие видеоимпульсы соответствующей высоты, а затем из импульсной последовательности выделяется модулирующее колебание. Второй этап выполняется с помощью «идеального» фильтра низких частот, пропускающего все частоты от нуля до F_B. Как известно, при подаче на вход такого

фильтра короткого импульса выходное напряжение будет пропорционально амплитуде импульса и описывается функцией вида $\frac{\sin x}{x}$. В результате на выходе фильтра получается исходное колебание $s_M(t)$ в соответствии с рядом Котельникова (п. 1.5).

Заметим, что спектр реальных колебаний не имеет резкого ограничения по частоте, а идеальный ФНЧ нереализуем, поэтому восстановление $s_M(t)$ всегда осуществляется приближенно.

6.1.2. Спектральные характеристики импульсных методов модуляции

Представление о спектральном составе импульсно-модулированных колебаний можно получить, рассмотрев спектр при АИМ.

Спектр модулирующего колебания представлен одной составляющей на частоте F_M (рис. 6.2,а). Спектр несущего колебания определяется периодической последовательностью импульсов (рис. 6.2,б).

Амплитудно-частотный спектр АИМ сигнала показан на рис. 6.2. Обратим внимание, что спектр содержит постоянную составляющую, составляющую на частоте модулирующего сигнала F_M и составляющие на частотах F_i, $i = 1,2,...$, при этом около каждой составляющей на частотах F_i, $i = 1,2,...$, находятся боковые частоты, отстоящие на частоту модулирующего сигнала $F_i \pm F_M$.

Наличие в спектре составляющей с частотой модулирующего сигнала F_M позволяет выделять ее с помощью ФНЧ. Если последовательность видеоимпульсов модулируется не простым гармоническим колебанием, а сигналом тональной частоты (речевой сигнал) с полосой $F_{\min} - F_{\max}$, то в спектре АИМ сигнала вместо частот F_M будут присутствовать спектральные составляющие в полосе $F_{\min} - F_{\max}$ (рис. 6.3). Из-за сравнительно низкой помехоустойчивости АИМ обычно используется несамостоятельно, а в качестве промежуточной процедуры при формировании сигналов.

Амплитудно-частотный спектр ОДИМ сигнала показан на рис. 6.2,г. Состав спектра аналогичен рассмотренному случаю АИМ, но имеет более сложную структуру. Однако значения амплитуд высших спектральных составляющих быстро убывают и при демодуляции также можно использовать ФНЧ. При этом возможно ограничение импульсов по амплитуде; это делает систему более помехоустойчивой.

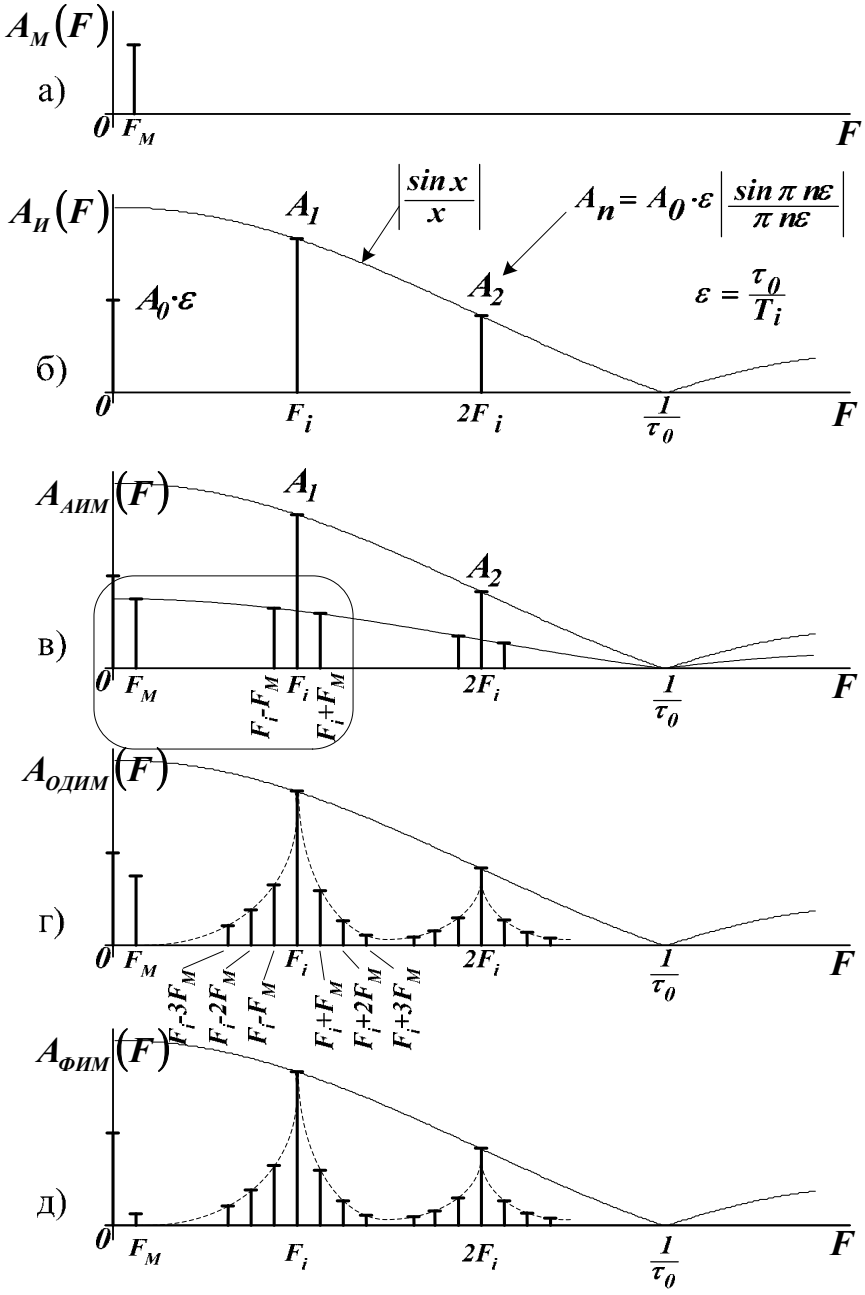


Рис. 6.2. Спектры сигналов импульсной модуляции

Амплитудно-частотный спектр ФИМ сигнала показан на рис. 6.2,д. По своей структуре он близок к спектру ДИМ, однако спектральная составляющая на частоте модулирующего сигнала F_M меньше, чем при ДИМ и АИМ в 50 и более раз. Это объясняется тем, что, информация заложена в положении импульсов, а их сдвиги при модуляции невелики. Следовательно среднее значение частоты модулирующего сигнала принятой ФИМ последовательности также мало. В этом случае применять ФНЧ нецелесообразно. Для де-

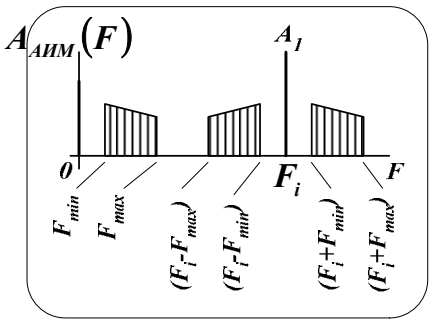


Рис. 6.3. Участок спектра сигнала АИМ при сложном информационном сигнале

модуляции ФИМ сигналы предварительно преобразуют в АИМ или ДИМ, и после этого применяют стандартные ФНЧ.

6.2. Помехоустойчивость непрерывных каналов связи с импульсной модуляцией

6.2.1. Помехоустойчивость систем передачи с импульсными методами модуляции

В реальных системах форму канальных импульсов стремятся сделать близкой к колоколообразной (рис. 6.4,а). У таких импульсов энергия сосредоточена в более узкой полосе частот, чем у прямоугольных.

Важным параметром импульсов является крутизна фронта (см. рис. 6.4,а) $S_Ф = \frac{ds(t)}{dt} = tg\varphi$, которая при максимальном значении улучшает выигрыш системы модуляции. На огибающей импульса можно найти такую точку d (точку касания касательной и огибающей) в которой крутизна фронта максимальна ($S_Ф = \max$).

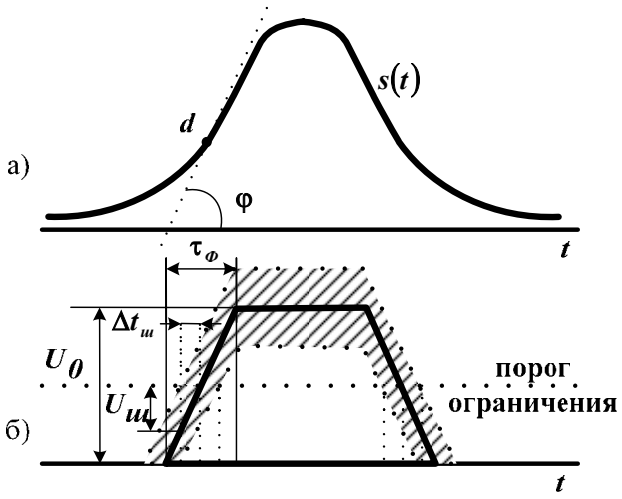


Рис. 6.4. Формы канального импульса

Для удобства анализа колоколообразную форму импульса заменим трапецеидальной с амплитудой U_0 и длительностью фронта $\tau_Ф$ (рис. 6.4,б – сплошная линия). Крутизна фронта такого импульса $S_Ф = \frac{U_0}{\tau_Ф}$. Заштрихованная область условно представляет собой множество состояний информационного импульса, обусловленное воздействием помехи $n(t)$ на входе устройства временного разделения и демодуляции. В результате помехи возникает паразитная модуляция канального импульса по амплитуде, длительности и фазе, что вызывает неодинаковый характер искажений демодулированного сигнала при АИМ, ДИМ (ШИМ), ФИМ.

Помехоустойчивость систем связи оценивается через выигрыш, который можно представить в виде выражения:

$$g = \frac{h_{вых}}{h_{вх}}, \qquad (6.2)$$

где $h_{вх}$ – отношение сигнал/шум (ОСШ) на входе приемника, а $h_{вых}$ – ОСШ на выходе канала тональной частоты.

Определим вначале выигрыш g для систем связи с АИМ.

Отношение сигнал/шум на входе приемника определяется выражением:

$$h_{вх} = \frac{U_0}{U_{ш\ вх}}.$$

Отношение сигнал/шум на выходе канала тональной частоты:

$$h_{вых} = \frac{U_m}{U_{швых}} = h_{вх} m_{AM}, \qquad (6.3)$$

т.к. максимальное значение полезного сигнала на выходе демодулирующего ФНЧ:

$$U_m = U_0 m_{AM} \varepsilon K_{ФНЧ},$$

где $m_{AM} = \frac{U_{\max} - U_{\min}}{U_{\max} + U_{\min}}$ – индекс амплитудной модуляции; $\varepsilon = \frac{\tau_0}{T_i}$ – коэффициент заполнения; $K_{ФНЧ}$ – коэффициент передачи ФНЧ. Обозначим $\varepsilon K_{ФНЧ} = \nu$, где ν – коэффициент пропорциональности, тогда $U_m = U_0 \nu m_{AM}$.

Эффективное значение напряжения помехи на выходе индивидуального канала:

$$U_{швых} = \nu \cdot U_{ш\ вх},$$

тогда:

$$h_{вых} = \frac{U_0 \nu m_{AM}}{\nu U_{ш\ вх}} = \frac{U_0 m_{AM}}{U_{ш\ вх}} = h_{вх} m_{AM},$$

откуда величина выигрыша для АИМ:

$$g_{АИМ} = m_{AM}.$$

Таким образом, при АИМ отношение сигнал/шум на выходе демодулятора пропорционально глубине амплитудной модуляции. Максимальная помехоустой-

чивость достигается при $(m_{AM}=1)$, однако в реальных системах связи всегда $m_{AM}<1$. Поэтому отношение сигнал/шум на выходе демодулятора АИМ всегда ниже, чем на входе.

Оценим теперь помехоустойчивость систем связи с ФИМ (ДИМ). В этом случае сигнал на выходе демодулятора пропорционален девиации длительности $\Delta\tau_m$ (ДИМ) и положения Δt_m (ФИМ) канальных импульсов, соответственно (рис. 6.1,г,д).

При ФИМ, как и при ДИМ, можно исключить значительную часть искажений, а также аддитивных шумов, путем двухстороннего ограничения импульсов по амплитуде и их последующего восстановления по длительности. Предположим, что ограничитель обладает порогом ограничения $\frac{U_0}{2}$ (рис. 6.4,б), а время переходных процессов пренебрежительно мало. Тогда на выходе ограничителя будут формироваться модулированные по соответствующему параметру импульсы.

Отношение сигнал/шум на входе демодулятора, можно оценить на основании подобия треугольников (рис. 6.4,б) через отношение:

$$h_{вхФИМ(ДИМ)}=\frac{U_0}{U_{ш\ вх}}=\frac{\tau_Ф}{\Delta t_ш}. \qquad (6.4)$$

Отношение сигнал/шум на выходе канала тональной частоты будет определяться выражением (6.3).

Максимальное значение полезного сигнала на выходе демодулятора пропорционально девиации канальных импульсов Δt_m при ФИМ $U_m=\nu\cdot\Delta t_m$ и $\Delta\tau_m$ при ДИМ $U_m=\nu\cdot\Delta\tau_m$ (рис. 6.5):

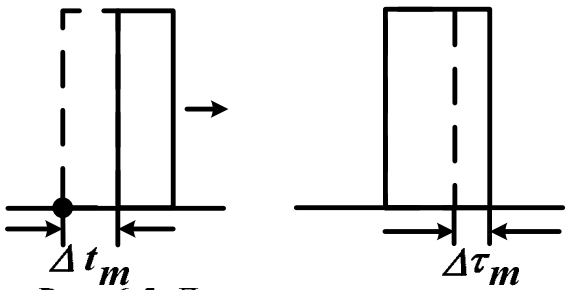


Рис. 6.5. Девиации положения и длительности канальных импульсов при ФИМ и ДИМ

Эффективное напряжение шума на выходе демодулятора:

$$U_{швых}=\nu\cdot\Delta t_ш,$$

тогда:

$$h_{выхФИМ} = \frac{U_m}{U_{швых}} = \frac{v\Delta t_m}{v\Delta t_ш} = \frac{\Delta t_m}{\Delta t_ш}, \qquad (6.5)$$

$$h_{выхДИМ} = \frac{\Delta \tau_m}{\Delta t_ш}. \qquad (6.6)$$

Подставляя (6.4) и (6.5) в (6.2), получим выигрыш при ФИМ:

$$g_{ФИМ} = \frac{\Delta t_m}{\tau_Ф}.$$

Аналогично подставляя (6.4) и (6.6) в (6.2), получим выигрыш при ДИМ:

$$g_{ДИМ} = \frac{\Delta \tau_m}{\tau_Ф}.$$

Очевидно, выигрыш при ФИМ (ДИМ) прямо пропорционален девиации положения (длительности) канальных импульсов Δt_m ($\Delta \tau_m$) и обратно пропорционален длительности фронта. Таким образом, порог ограничения реальных ФИМ (ДИМ) видеоимпульсов необходимо выбирать на таком уровне, при котором крутизна фронта канальных импульсов максимальна (например, в точке d на рис. 6.4,а)

Так как на практике всегда выполняются неравенства $\tau_Ф < \Delta t_m$ при ФИМ и $\tau_Ф < \Delta \tau_m$ при ДИМ, то $g_{ФИМ,ДИМ} > 1$, т.е. отношение сигнал/шум на выходе демодуляторов ФИМ (ДИМ) выше, чем на их входах.

Сравним между собой системы с ДИМ и ФИМ. Для получения требуемой девиации среднюю длительность канальных импульсов при ДИМ приходиться выбирать большей, чем при ФИМ. Это приводит к тому что $P_{срШИМ} > P_{срФИМ}$. Следовательно в системах с ФИМ при сохранении средней мощности передатчика имеется возможность увеличить амплитуды канальных радиоимпульсов и тем самым повысить ОСШ на входе приемника. Кроме того, в системах с ДИМ для безискаженной передачи самых коротких импульсов приходиться выбирать полосу пропускания приемопередающих трактов более широкой чем это необходимо для импульсов средней длительности.

В то же время в системах с ФИМ длительность всех канальных импульсов одинакова и неизменна, что позволяет лучше согласовывать полосу про-

пускания приемопередающих трактов со спектрами сигналов и тем самым достичь более высокой помехоустойчивости. Кроме того, при ФИМ есть возможность дополнительно снизить влияние помех за счет регенерации канальных импульсов не только по амплитуде (как при ШИМ), но и по длительности, что невозможно в системе с ШИМ.

Поэтому в современных системах связи с ВРК преимущественно применяется ФИМ, а АИМ и ШИМ используются как промежуточные процедуры при формировании и обработки импульсно-модулированных сигналов.

6.2.2. Порог помехоустойчивости системы передачи с импульсными методами модуляции

Всем системам связи, в которых используются нелинейные методы модуляции, присущ так называемый пороговый эффект [6].

В системах связи с ФИМ-АМ (ЧМ) момент прихода информационного импульса определяется как момент перехода его огибающей через некоторый пороговый уровень (рис.6.6,а), который обычно выбирается равным половине амплитуды импульсного сигнала.

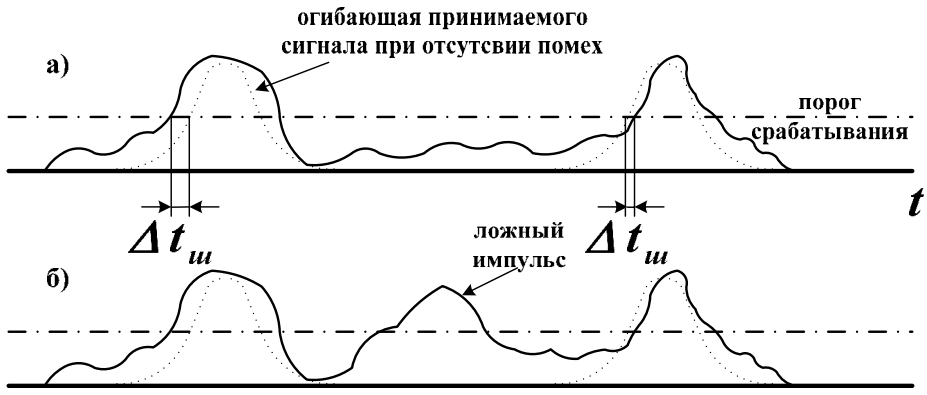


Рис. 6.6. Ограничение сигналов: при слабой помехе («а»); при сильной помехе («б»)

При наличии помех происходит искажение огибающей принимаемого импульса смещается момент пересечения порога. Чем круче фронт импульса, тем меньше смещение, вызываемое помехой.

При большом уровне помех отдельные их выбросы превышают пороговый уровень и вызывают ложные срабатывания оконечных устройств в промежутках между информационными импульсами (рис. 6.6,б).

Если значения помехи с большой вероятностью достигают и превышают

порог, то прием становится практически невозможным. В этом случае считают, что система достигла порога помехоустойчивости, при котором слабый сигнал подавляется сильной помехой.

Величина средней пороговой мощности на входе приемника с ФИМ-АМ выражается: $P_{ср.пор} > P_{и.пор} \cdot \varepsilon$,

где $P_{и.пор}$ – импульсная пороговая мощность.

Анализ показывает [6], что импульсная пороговая мощность определяется выражением: $P_{и.пор} = 1{,}55 \cdot 10^{-19} \cdot \dfrac{n_ш}{\tau_0}$,

где $n_ш$ – коэффициент шума приемника; τ_0 – длительность канального импульса на уровне 0,5.

Тогда средняя пороговая мощность на входе приемника с ФИМ-АМ определяется выражением [6]:

$$P_{ср.пор} = 1{,}55 \cdot 10^{-19} \cdot \frac{n_ш}{T_i} = 1{,}55 \cdot 10^{-19} \cdot n_ш \cdot F_i .$$

Средняя пороговая мощность на входе приемника многоканальной системы связи с ФИМ-АМ определяется выражением [6]:

$$P_{ср.пор} = 1{,}55 \cdot 10^{-19} \cdot n_ш \cdot F_i \cdot N_k ,$$

где N_k – число каналов.

В системах связи с ФИМ-ЧМ пороговая мощность определяется выражением [6]:

$$P_{ср.пор} = n_ш \cdot k \cdot T \cdot \Delta f_{пр} \cdot \left(4 + 4{,}76 \cdot \lg \frac{\Delta f_{прм}}{2\Delta F_{вф}} \right)$$

где $\Delta f_{прм}$ – полоса пропускания приемника;

$\Delta F_{вф}$ – полоса пропускания видеофильтра $\left(\Delta F_{вф} = \dfrac{0.5}{\tau_0} \right)$;

$k \cdot T = 4 \cdot 10^{-21} \left[\text{Вт}/\text{Гц} \right]$ – произведение постоянной Больцмана на абсолютную температуру.

Средней пороговой мощностью считают такую, при которой отношение

сигнал/шум на выходе канала ТЧ отклоняется от линейного закона в сторону уменьшения на 1-2 дБ. На рис. 6.7 представлена зависимость выходного ОСШ от мощности сигнала на входе приемника для систем связи с ФИМ-АМ и ФИМ-ЧМ. Из рисунка видно, что при уменьшении мощности сигнала на входе приемника $(P_{пр})$ ниже $P_{ср.пор1}$ ОСШ $\left(\frac{P_c}{P_ш}\right)_{выхТЧ}$ резко падает. Поэтому при проектировании систем связи с ФИМ-АМ (ЧМ) основная задача состоит в том, чтобы определить тот минимальный уровень мощности на входе приемника, при котором шумовая защищенность канала ТЧ (или вероятность ошибки при приеме дискретной информации) соответствует заданному значению.

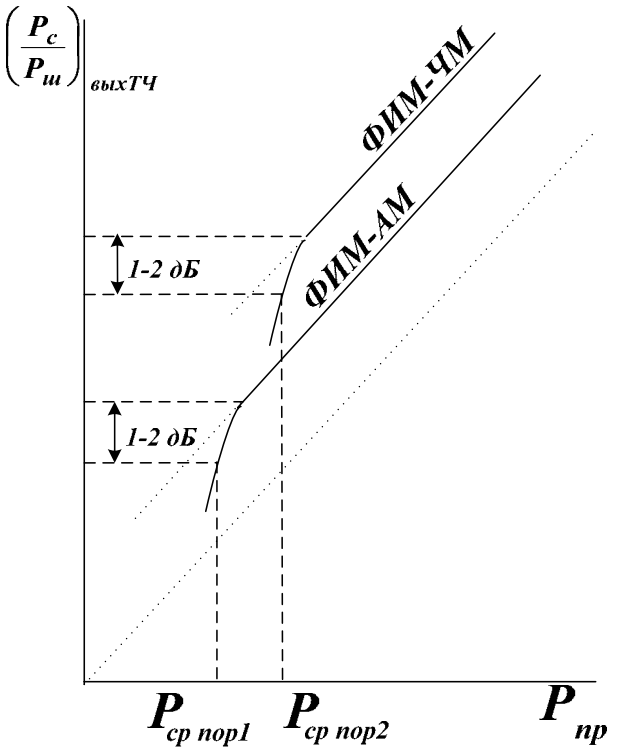


Рис. 6.7. Определение пороговой мощности

6.3. Цифровые методы передачи непрерывных сообщений

Для передачи непрерывных сигналов можно использовать дискретный канал, преобразуя непрерывный сигнал в цифровой с помощью АЦП, а на приемной стороне цифровой сигнал в непрерывный с помощью ЦАП.

Применение цифровых систем передачи (ЦСП) дает возможность объединения различных видов связи на единой цифровой основе, а также широко использовать современную элементную базу, обеспечивая стабильность характеристик, надежность, и хорошие массо-габаритные показатели.

6.3.1. Передача сигналов с импульсно-кодовой модуляцией

Принцип АЦП на основе импульсно-кодовой модуляции (ИКМ) включает дискретизация во времени, квантование по уровню (амплитуде) и кодирование

[5, 6, 18, 20, 21].

Процесс формирования ИКМ сигнала поясним с помощью упрощенной структурной схемы (рис.6.8) и временных диаграмм (рис. 6.9).

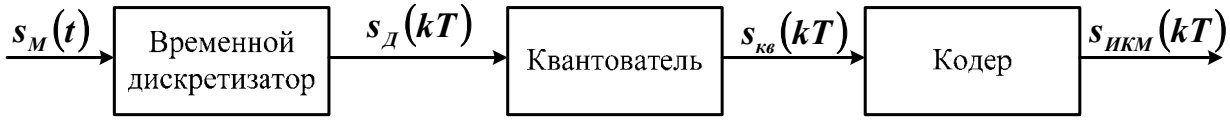


Рис. 6.8. Обобщенная структурная схема процесса формирования ИКМ сигнала

Дискретизация заключается в том, что непрерывный сигнал $s_M(t)$ (рис. 6.9,а) заменяется отсчетами $s_Д(kT)$ (на рис. 6.9,б), следующими через одинаковые интервалы времени $\Delta t = \frac{1}{2F_{\max}}$. Например, для речевого сигнала, где $F_{\max} = 3{,}4$ кГц, принят стандартный интервал $t = 125$ мкс, $2F_{\max} \approx 8$ кГц .

Процесс дискретизации эквивалентен импульсной модуляции. Для примера на рис. 6.9,б приведен случай АИМ.

Различают равномерное и неравномерное квантование. При квантовании устанавливается количество уровней (L) разрешенных для передачи.

Процесс квантования состоит в следующем текущие значения сигнала соответствующее моменту отсчета $s_Д(kT)$ заменяется ближайшим дискретным значением $s_{кв}(kT)$ (уровнем), такая операция подобна округлению и приводит к ошибке:

$$\varepsilon_{кв}(t) = s_Д(kT) - s_{кв}(kT),$$

где $\varepsilon_{кв}(t)$ – шум квантования, величина которого обычно считается случайной, равномерно распределенной в пределах $-0{,}5\delta \le \varepsilon_{кв} \le 0{,}5\delta$. Дисперсия шума квантования $\sigma^2_{КВ} = \frac{\delta^2}{12}$.

Разницу между двумя соседними уровнями s_k и s_{k-1} называют шагом квантования:

$$\delta_k = s_k - s_{k-1} = \left(s_{\max} - s_{\min}\right) \Big/ (L-1).$$

При равномерном квантовании шаг квантования δ имеет постоянную ве-

личину. В системе ИКМ с равномерным квантованием как большие, так и малые сигналы кодируются с одним и тем же шагом квантования. Если выбор шага квантования был ориентирован на малые сигналы, то для больших сигналов создается избыточное качество воспроизведения. Кроме того, вероятность появления больших сигналов мала. По этим причинам можно считать, что выбранная разрядность кода n не всегда используется эффективно.

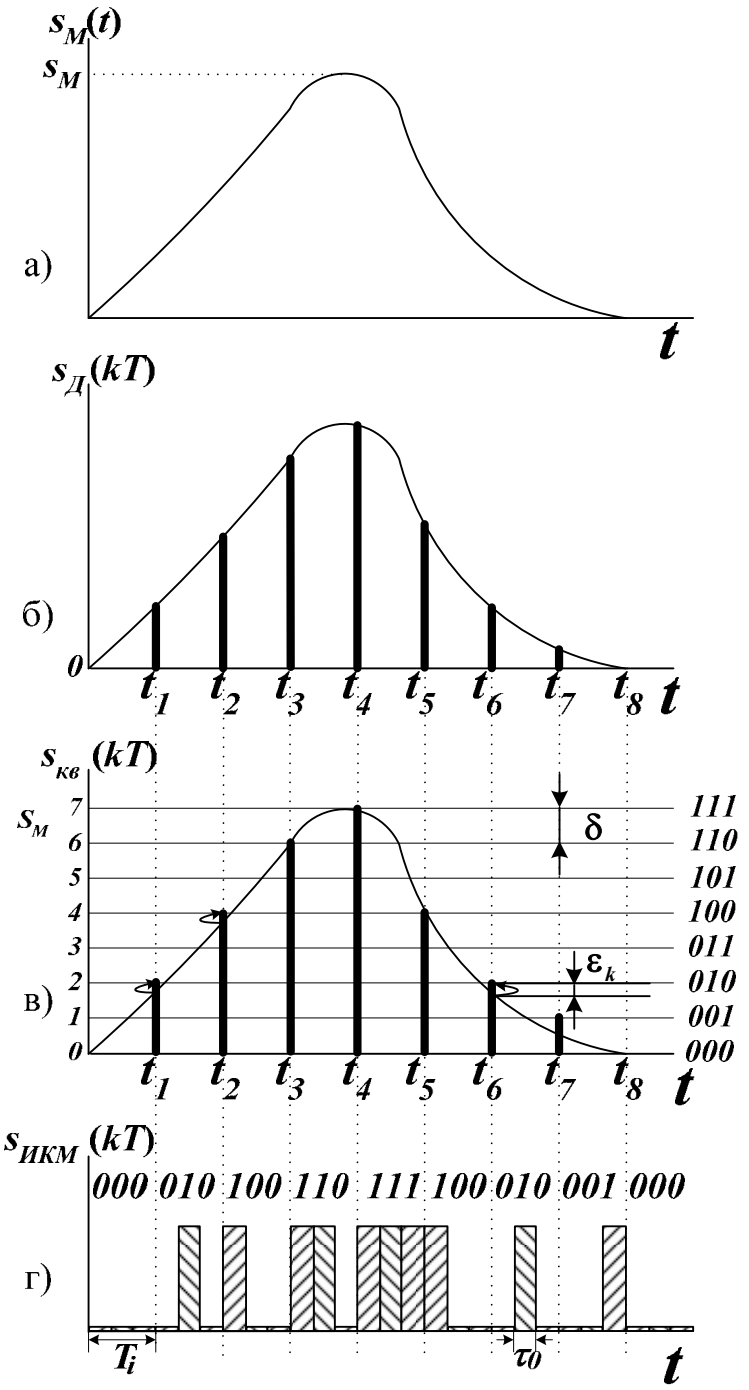


Рис. 6.9. Формирование ИКМ сигнала

Можно реализовать более высокую точность передачи, если применить неравномерное квантование, предполагающее для больших значений входных сигналов увеличение шага квантования. Выбор характеристики квантователя позволяет добиться одинакового качества восстановления сигналов как малой, так и большой величины.

Трудности реализации неравномерного квантования устраняются предварительным нелинейным преобразованием – компрессией («сжатием») аналогового сигнала.

Компрессированные отсчеты сигнала затем подвергаются равномерному квантованию. Для компенсации нелинейного искажения отсчетов на приемной

стороне осуществляют их обратное преобразование – экспандирование («растяжение»). Совместный процесс компрессирования и экспандирования называется компандированием сигнала.

Таким образом, компандирование обеспечивает передачу с меньшими шумами квантования сигналов, обладающих малой средней мощностью (с большим пикфактором), например речевых.

При кодировании происходит преобразование квантованных значений $s_{кв}(kT)$ в n разрядные кодовые комбинации. Например, при количестве уровней $L = 8 = 2^3$, в десятичной системе счисления этим уровням соответствуют номера от 0 до 7 (рис. 6.9,в). В двоичной системе счисления им соответствуют трехразрядные кодовые комбинации, в данном случае от 000 до 111 (рис. 6.20,в). Полученная импульсная последовательность представлена на рис. 6.9,г.

Повышение разрядности, во-первых, связано с определенными трудностями технической реализации быстродействующих многоразрядных кодеков и, во-вторых, требует значительного увеличения пропускной способности систем связи, что не всегда возможно. Преодоление указанных трудностей возможно, например, за счет применения неравномерного квантования.

6.3.2. Передача сигналов с дельта модуляцией

Дельта-модуляция (ДМ) – особый вид импульсной модуляции, при которой так же, как и при ИКМ, аналоговый сигнал $s_M(t)$ представляется в виде дискретных отсчетов времени, квантованных по амплитуде. ДМ основана на существовании зависимости между отсчетами в речевом сигнале. При ДМ используется только один разряд для квантования разности соседних отсчетов. В этот разряд записывается полярность разности [5, 18, 20, 21, 32]. Важным элементом схемы (рис.6.10) при ДМ является компара-

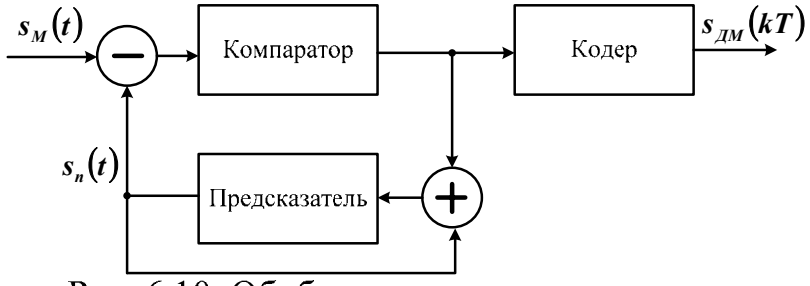


Рис. 6.10. Обобщенная структурная схема процесса формирования ДМ сигнала

тор, который разность $s_M(t)-s_n(t)$ входного сигнала $s_M(t)$ и предсказанного $s_n(t)$, имеющего ступенчатый вид, квантует на два уровня. На выходе компаратора появляется значение $+1$ если входной сигнал больше предсказанного (разность положительна); и -1 если он меньше предсказанного сигнала (разность отрицательна).

Закон возрастания (уменьшения) величины шага δ для предсказания $s_{пр}(kT)=s_{пр}((k-1)T)\pm\delta$ выбирается исходя из статистических характеристик передаваемых сообщений. В частности, величина δ может возрастать по линейному закону, по закону геометрической прогрессии, по экспоненциальному закону и другим законам, обеспечивающим требуемую точность передачи информации. При использовании постоянного шага (рис.6.11,б), необходимо иметь высокую тактовую частоту с целью предотвращения перегрузок по крутизне. Наиболее широко применяются методы адаптивной ДМ, один из которых иллюстрируется на рис. 6.11,в. При таком методе шаг предсказания δ меняется автоматически в соответствии с законом изменения крутизны (производной) сигнала $s_M(t)$. На рис. 6.11,в показано, что участку сигнала с большой крутизной соответствуют большие шаги квантования, что позволяет устранить

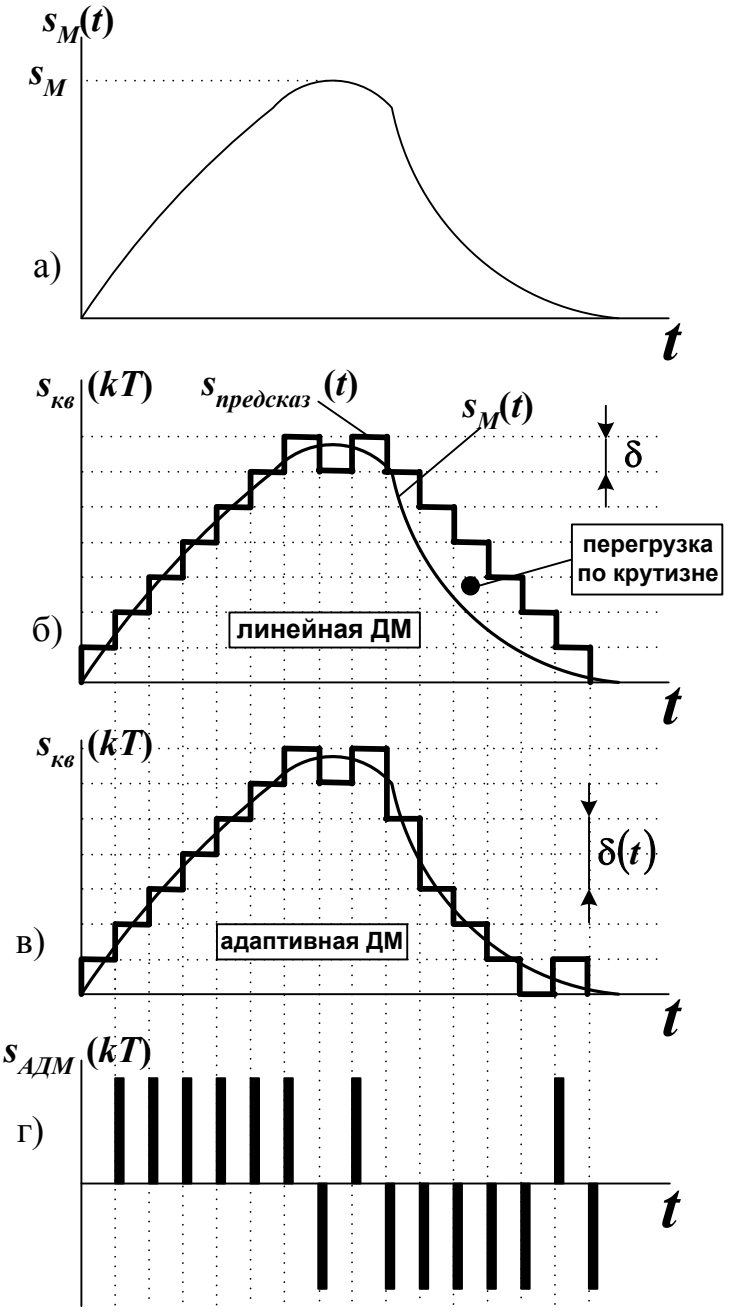


Рис. 6.11. Формирование ДМ сигнала

искажения.

При создании цифровых систем связи применение ДМ является перспективным благодаря следующим особенностям:

устройства кодирования и декодирования ДМ сигналов характеризуются более простыми, чем при ИКМ, схемными решениями, что важно с точки зрения их надежности и стоимости;

ДМ сигналы по сравнению с ИКМ имеют большую устойчивость к сбою символов в каналах связи, поскольку «вес» каждого символа ограничен лишь значениями δ. Вследствие этого пороговые свойства систем связи с ДМ несколько лучше, чем при ИКМ;

в системах связи с ДМ предъявляются менее жесткие требования к работе системы синхронизации.

6.3.3. Квантование сигналов в системах с ИКМ и ДМ

Качество квантования в системах ИКМ может характеризоваться отношением сигнал/шум на выходе цифро-аналогового преобразователя:

$$h_{вых}^2 = \frac{P_{свых}}{P_{швых}}.$$

Как уже отмечалось, особенностью цифровых методов передачи непрерывных сообщений является то, что из-за погрешностей квантования цифровой сигнал отличается от оригинала даже при полном отсутствии помех и искажений в канале. Эти отличия учитываются с помощью шума квантования, дисперсия которого определяется выражением [6, 20, 21, 32, 39]:

$$\sigma_{КВ}^2 = \frac{\delta^2}{12}.$$

Расчеты показывают, что для получения высокого качества передаваемых речевых сообщений необходимо выбирать $L = 1000 \div 2000$. При этом потребуется $10 - 11$ – разрядное кодирование ($L = 2^n = 2^{11} = 2048$).

Помимо шума квантования, на выходе приемника с ИКМ может возникать дополнительный шум из-за ошибочного приема информационных симво-

лов кодовых комбинаций, что происходит вследствие воздействия тепловых шумов приемника и других источников помех.

Цифровым системам с ДМ также присущи искажения, которые можно подразделить на следующие виды:

искажения, обусловленные методом преобразования; к ним относятся шумы квантования и шумы перегрузки;

искажения, обусловленные несовершенством технической реализации кодеков, например шум в молчащем канале из-за разбаланса приращений аппроксимирующего напряжения в интеграторе, либо из-за конечной чувствительности компаратора;

искажения, возникающие в канале связи из-за воздействия переходных помех, тепловых шумов и т. д.

При увеличении δ растет дисперсия шума квантования:

$$\sigma_{КВ}^2 = k_{кв} \frac{\Delta F_{ФНЧ} \delta^2}{F_i}, \qquad (6.7)$$

где $k_{кв}$ – коэффициент, зависящий от вида входного сигнала. Например, при гармоническом модулирующем сигнале равен $k_{кв} = 1/3$.

Определим отношение сигнал/шум $\frac{P_c}{\sigma_{КВ}^2}$ на выходе декодера для случая, когда модулирующий сигнал является гармоническим $U_M(t) = U_m \cos 2\pi F_M t$.

При максимальной мощности сигнала [6]:

$$P_c = \frac{\delta^2 F_i^2}{8\pi^2 F_M{}^2}. \qquad (6.8)$$

из (6.7) и (6.8) получим:

$$\left(\frac{P_c}{\sigma_{КВ}^2}\right)_{вых} = \frac{1}{8\pi^2 k_{кв}} \frac{F_i^3}{F_M^2 \Delta F_{ФНЧ}}. \qquad (6.9)$$

Из выражения (6.9) следует, что отношение $\frac{P_c}{\sigma_{КВ}^2}$ на выходе канала связи очень критично к величине частоты следования информационных символов F_i. Чем больше частота F_i, тем более высокой становится защищенность каналов.

Кроме того, это отношение существенно зависит от частоты модулирующего сигнала F_M. Чем выше F_M, тем меньше отношение сигнал/шум. Следовательно, более высокие частоты спектра сигнала воспроизводятся хуже, чем низкие, что является существенным недостатком систем связи с линейной ДМ.

Контрольные вопросы

1. Как выбирается период следования импульсов несущей при импульсных видах модуляции?

2. В чем сходство и различие понятий: модуляция, манипуляция, импульсная модуляция?

3. Когда возникает пороговый эффект при использовании импульсных методов модуляции?

4. На каких операциях основано применение цифровых методов модуляции?

5. Что такое шум квантования и как оценивается его мощность?

ГЛАВА 7. МЕТОДЫ ПРИЕМА СИГНАЛОВ В СЛОЖНЫХ УСЛОВИЯХ

7.1. Прием сигналов в каналах с замираниями

7.1.1. Сущность замираний и их классификация

Ранее были рассмотрены методы оптимального приема сигналов в каналах с постоянными параметрами. В частности, модуль коэффициента передачи канала μ и время задержки τ предполагались неизменными, хотя и не всегда известными. Однако в ряде случаев, например в линиях радио, радиорелейной и тропосферной связи, эти величины являются переменными. Основной особенностью названных линий является многолучевое распространение радиоволн (рис. 7.1). Принятый сигнал является суммой сигналов, пришедших по M различным траекториям:

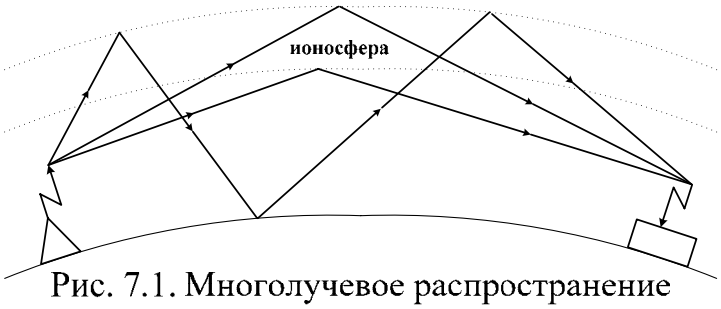


Рис. 7.1. Многолучевое распространение сигналов

$$U(t)=\sum_{k=1}^{M}\mu_k S(t-\tau_k).$$

Коэффициент передачи $\mu_k=\mu_k(t)$ и время задержки $\tau_k=\tau_k(t)$ из-за изменчивости свойств среды распространения оказываются переменными. Поэтому параметры сигнала $U(t)$также изменяются во времени.

Многолучевость сигнала в сочетании с подвижностью пространственных неоднородностей приводит к появлению так называемых «быстрых» замираний, а также к рассеянию сигнала во времени, характеризуемом временем многолучевости $\tau_{МЛ}$. Величина $\tau_{МЛ}$ непрерывно изменяется и зависит от протяженности сеанса связи и ширины диаграммы направленности антенн. Для малоканальных тропосферных линий максимальное время многолучевости лежит в пределах 0,07...0,4 мкс. Флюктуации времени многолучевости приводят к взаимному влиянию соседних импульсов в дискретных системах связи, а также к появлению переходных шумов в аналоговых многоканальных системах. Кроме того, изменяется время прихода сигнала (время группового замедления сигна-

лов), что существенно затрудняет реализацию синхронных методов приема.

Изменение во времени коэффициентов передачи $\mu_k(t)$ называется замираниями. В зависимости от поражаемого частотного диапазона замирания делятся на селективные и общие (рис. 7.2). В зависимости от времени существования они разделяются на быстрые (с квазипериодом $0{,}1 \div 10$ с) и медленные (с квазипериодом от десятков минут до нескольких часов). Такое деление условно, поскольку существуют замирания и с промежуточными квазипериодами.

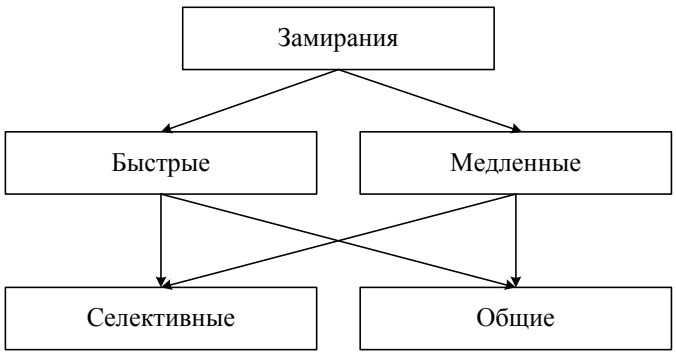


Рис. 7.2. Классификация замираний

Быстрыми принято считать замирания на коротких отрезках времени, со случайным квазипериодом (в пределах от десятых долей до единиц секунд, но не более 10 мин) с глубиной до 30 дБ, подчиняющихся чаще всего релеевскому закону [6]. При приеме сигналов в разнесенных по пространству или частоте антеннах быстрые замирания, как правило, независимы или слабо зависимы.

Изменение отражающих свойств, например, за счет изменения метеоусловий в объеме рассеяния, приводит к общим «медленным» замираниям. Глубина медленных замираний составляет $20 \div 30$ дБ, квазипериод – единицы минут ÷ часы, а распределение обычно подчиняется логарифмически нормальному закону. Для медленных замираний характерна коррелированность по всем ветвям разнесения.

Медленные замирания вызываются изменением условий рефракции и медленным изменением параметров тропосферных неоднородностей. Наблюдаются они обычно при прохождении теплых и холодных фронтов воздуха и образовании на трассе интенсивных инверсионных слоев, приводящих к значительному возрастанию требуемого уровня сигнала.

Наличие медленных замираний приводит к необходимости создания в системах связи «запаса» энергетического потенциала для компенсации их

влияния.

Плотность вероятностей величины μ выражается обобщенным законом Рэлея [6, 22, 42]:

$$w(\mu)=\frac{\mu}{\sigma^2}\exp\left(-\frac{\mu_0^2+\mu^2}{2\sigma^2}\right)I_0\left(\frac{\mu_0\mu}{\sigma^2}\right),\quad \mu\geq 0, \tag{7.1}$$

где μ_0 – регулярная составляющая коэффициента передачи; σ^2 – параметр, характеризующий флюктуирующую составляющую; $I_0(*)$ – модифицированная функция Бесселя нулевого порядка. Нередко регулярная составляющая μ_0 оказывается равной нулю и выражение (7.1) переходит в обычное распределение Рэлея:

$$w(\mu)=\frac{\mu}{\sigma^2}\exp\left(-\frac{\mu^2}{2\sigma^2}\right),\quad \mu\geq 0.$$

В этом случае замирания называют рэлеевскими. Чем больше относительная величина регулярной составляющей μ_0, тем меньше глубина замираний. Самыми глубокими в обычных условиях являются рэлеевские замирания.

Найдем вероятность ошибок при различных видах манипуляции с учетом замираний. Если при некотором фиксированном значении μ условную вероятность ошибки обозначить p_μ, то безусловная вероятность ошибок при медленных общих замираниях

$$p=\int_0^\infty p_\mu w(\mu)d\mu.$$

Условная вероятность ошибок p_μ вычисляется для канала без замираний.

Вычисление интеграла приводит к результату [6, 20]:

$$p=\frac{1+b^2}{2+2b^2+\gamma^2h_{cp}^2}\exp\left(-\frac{b^2\gamma^2h_{cp}^2}{2+2b^2+\gamma^2h_{cp}^2}\right),$$

где $b^2=\frac{\mu_0^2}{2\sigma^2}$ – отношение мощностей регулярной и флуктуирующей составляющих сигнала; $h_{cp}^2=\frac{\mu_{cp}^2}{\mu^2}h^2$; $\mu_{cp}^2=\mu_0^2+2\sigma^2$ – среднее значение квадрата коэффициента передачи.

Вероятность ошибки уменьшается с увеличением среднего отношения энергии сигнала к спектральной плотности мощности помехи $h_{ср}^2$, а также с увеличением параметра γ. Так при ОФМн ($\gamma = \sqrt{2}$) она меньше, чем при ЧМн ($\gamma = 1$) или АМн ($\gamma = 1/\sqrt{2}$). Существенно влияет на вероятность ошибки параметр b. Наибольшая вероятность ошибки имеет место при рэлеевских замираниях, когда $b = 0$ и

$$p = \frac{1}{2 + \gamma^2 h_{ср}^2}. \qquad (7.2)$$

В частности, для ЧМн ($\gamma = 1$)

$$p = \frac{1}{2 + h_{ср}^2},$$

а для ОФМн ($\gamma = \sqrt{2}$)

$$p = \frac{1}{2 + 2h_{ср}^2}.$$

7.1.2. Принципы разнесенного приема сигналов

Для повышения верности приема при замираниях переданное сообщение передается не по одному, а по двум или нескольким каналам связи. С этой целью могут использоваться различные средние частоты (разнесение по частоте) или передача в разные отрезки времени (разнесение по времени). Но наиболее широкое применение получил в радиосвязи метод приема сигналов на разнесенные антенны, находящиеся друг от друга на расстоянии нескольких длин волн (пространственно разнесенный прием), или принимающие различные поляризационные составляющие электромагнитного поля (поляризационно разнесенный прием).

Повышение эффективности при разнесенном приеме достигается в том случае, если замирания в различных ветвях разнесения не коррелированны или слабо коррелированны друг с другом. Поэтому в то время, когда в одних ветвях уровень сигнала оказывается очень низким, в других ветвях он может быть высоким и по ним легко восстановить переданное сообщение. Если замирания в

ветвях слабо коррелированны, то вероятность одновременного падения уровней в нескольких ветвях может быть достаточно мала.

Пространственно–разнесенный прием, когда производится одновременный прием сигналов одного передатчика несколькими приемниками на разнесенные в пространстве антенны. Такой способ является наиболее распространенным. Параметр разнесения обычно задают в виде нормированного расстояния $r_н = r/\lambda$, где r – проекция расстояния между антеннами на направление прихода радиоволн; λ – длина волны. Очевидно, величина $r_н$ существенно зависит от расположения антенн относительно направления трассы связи.

Частотно-разнесенный прием, когда сигналы, передаются одновременно на нескольких частотах одним или несколькими передатчиками. При частотно-разнесенном приеме величина разноса рабочих частот определяется интервалом корреляции замираний по спектру и в декаметровом диапазоне волн обычно составляет $0{,}5 \div 2$ кГц. Частотно-разнесенный прием применяется не только для борьбы с замираниями сигналов, но и является эффективным методом повышения устойчивости КВ связи при воздействии сосредоточенных по спектру помех.

Основной недостаток частотного разнесения состоит в расширении полосы частот, занимаемой системой связи, что приводит к увеличению взаимных помех, т.е. к ухудшению условий электромагнитной совместимости средств радиосвязи.

Временной разнесенный прием осуществляется с помощью многократно передаваемых на одной и той же частоте сигналов через некоторые интервалы времени. Временное разнесение сигналов накладывает ограничения на скорость передачи информации, так как интервал повторения сигнала должен превосходить среднюю длительность замираний в канале связи. Несмотря на это, принципы временного разнесения широко используются в системах с обратной связью по решению, т.е. с автоматическим запросом ошибок и повторением информации.

Эффективность того или иного метода разнесенного приема во многом

определяется способом обработки разнесенных сигналов на приемной стороне. Наибольшее распространение на практике получили способы линейного сложения и автоматического выбора ветви разнесения.

7.2. Методы борьбы с замираниями сигналов

7.2.1. Методы борьбы с замираниями в аналоговых системах связи

В аналоговых системах возможен ряд вариантов разнесенного приема, отличающихся способами объединения ветвей и формирования результирующего колебания $U_P(t)$.

В системах связи применяются линейные методы додетекторного объединения ветвей. При этом результирующее колебание на выходе схемы объединения представляет собой линейную комбинацию M входных колебаний отдельных ветвей разнесения:

$$U_P(t) = \sum_{i=1}^{M} k_i \cdot U_i(t), \qquad (7.3)$$

где $U_i(t) = \mu_i S(t) + n_i(t)$ – принятый сигнал в i-й ветви; $S(t)$ – переданный сигнал; μ_i – коэффициент передачи, зависящий от условий распространения сигнала; $n_i(t)$ – помеха в i-й ветви; k_i –весовые коэффициенты, величина которых зависит от конкретного метода объединения ветвей.

Среди линейных методов объединения ветвей большое распространение в технике пространственного разнесенного приема получили простое и оптимальное линейное сложение сигналов. Кроме того, часто применяется автоматический выбор наилучшей ветви разнесения. Различают автовыбор по наибольшему значению сигнала и автовыбор по наибольшему превышению сигнала над помехой.

Система автовыбора с переключением приемников

В этой системе путем переключения приемников из M ветвей выбирается та, в которой сигнал имеет наибольшее значение. Для такой системы коэффициенты в выражении (7.3) выбираются следующим образом:

$$k_i = \begin{cases} 1 \text{ при } (i = j) \\ 0 \text{ при } (i \neq j) \end{cases},$$

где j – индекс лучшего в каком либо смысле сигнала.

Структурная схема приемного устройства с системой автовыбора для сдвоенного приема, в которой приемники переключаются по низкой частоте (после детектора) приведена на рис. 7.3. Возможна аналогичная система с переключением приемников до детектора (на промежуточной частоте).

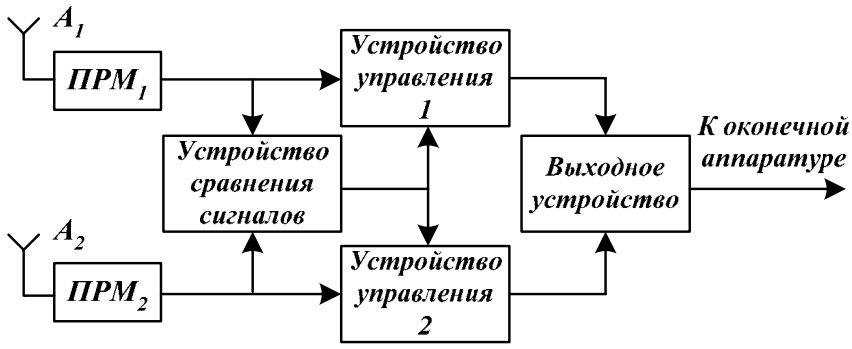


Рис. 7.3. Схема сдвоенного пространственно разнесенного приема сигналов с автовыбором

В соответствии с принципами работы системы автовыбора огибающая сигнала на выходе схемы объединения представляет собой случайную величину $U_{ср}(t) = \max U_i(t)$, где номер ветви i может меняться случайно от одного интервала анализа состояния ветвей к другому (рис. 7.4).

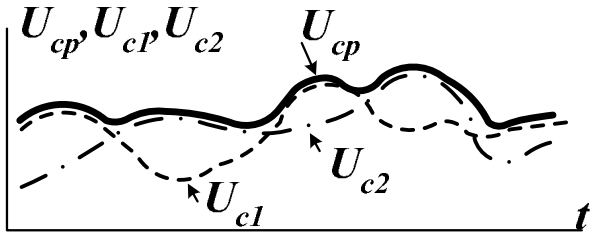


Рис. 7.4. Формирование результирующего сигнала при разнесенном приеме

Оценку различных систем разнесенного приема производят по вероятности ошибки. Если замирания в отдельных ветвях неселективные и определяются законом Рэлея, то вероятность ошибки в рассматриваемой системе автовыбора может быть рассчитана по формуле [6, 42]:

$$p_M = \frac{1}{2} \cdot \frac{M!}{\left(1 + h_{ср}^2/2\right)\left(2 + h_{ср}^2/2\right)\ldots\left(M + h_{ср}^2/2\right)}, \qquad (7.4)$$

где $h_{ср}^2 = \left(P_c/P_ш\right)_{ср} = \left(P_c/P_ш\right)_i$ – среднее отношение мощности сигнала и помехи (шума) в i-й ветви на входе приемника.

Если среднее значение отношения сигнал/шум в ветвях достаточно вели-

ко ($h_{ср}^2 >> 1$), то формулу (7.4) можно упростить:

$$p_M \approx \frac{M!}{2} \cdot \left(\frac{2}{h_{ср}^2} \right)^M . \qquad (7.5)$$

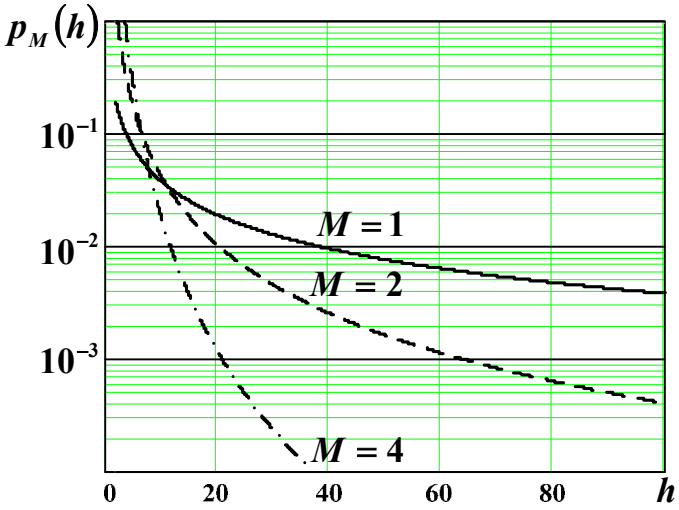


Рис. 7.5. Вероятность ошибки в системе автовыбора при различном числе ветвей разнесения (M)

На основе этого выражения построены графики (рис. 7.5) из которых следует, что применение разнесенного приема позволяет существенно уменьшить вероятность ошибки по сравнению с одиночным приемом флуктуирующих сигналов.

Система автовыбора с переключением антенн

В системе автовыбора с переключением антенн при M - кратном разнесении используется всего лишь один приемник. Структурная схема такой системы при сдвоенном приеме приведена на рис. 7.6. Она состоит из двух разнесенных антенн, одного приемника и переключающего устройства (блока автовыбора). Блок автовыбора подключает антенну, сигнал которой больше установленного порога U_0. Выбранный сигнал, например U_1 от антенны A_1, принимается до тех пор, пока $U_1 > U_0$. Как только U_1 станет меньше U_0, блок автовыбора переключает приемник на антенну A_2 и производится прием с этой антенны, пока $U_2 > U_0$. Если U_1 и U_2 одновременно будут

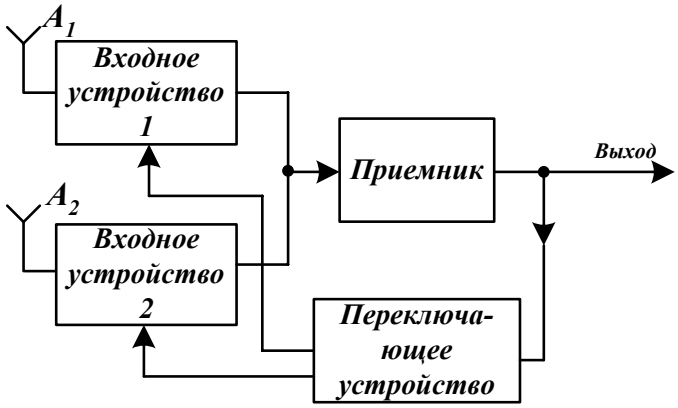


Рис. 7.6. Структурная схема системы автовыбора с переключением антенн при сдвоенном приеме

меньше U_0, то система будет находиться в состоянии поиска (периодического переключения антенн), пока хотя бы на одной из антенн сигнал не станет больше порога.

Вероятность ошибки при разнесенном приеме ЧМн сигналов с M антеннами, можно определить по формуле [6, 22, 42]:

$$p_M = \frac{1}{h_{ср}^2 + 2}\left[\left(1 - \exp\left\{-\frac{h_{прг}^2}{h_{ср}^2}\right\}\right)^{M-1} + \left(1 - \left(1 - \exp\left\{-\frac{h_{прг}^2}{h_{ср}^2}\right\}\right)^{M-1}\right)\exp\left\{-\frac{h_{ср}^2}{h_{прг}^2}\right\}\right],$$

где $h_{прг}^2 = \left(P_c \big/ P_ш\right)_{прг}$ – пороговое значение отношения мощностей сигнала и помехи.

На рис. 7.7 приведены зависимости вероятности ошибок для различного числа ветвей разнесения. Анализ этих зависимостей показывает, что увеличение кратности разнесения (от 2 до 5) приводит к значительному снижению вероятности ошибок. Наименьшая вероятность ошибки при сдвоенном приеме обеспечивается при некотором значении $h_{прг}^2 = \left(h_{прг}^2\right)_{opt}$, и определяется следующим выражением [42]:

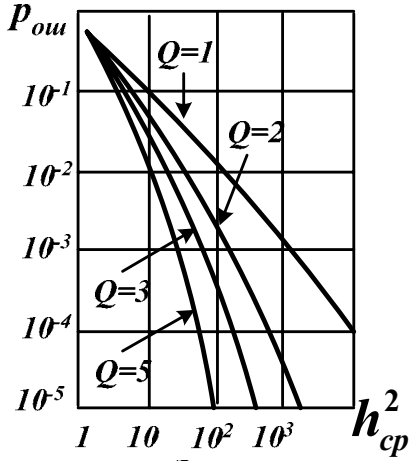


Рис. 7.7. Зависимости минимальной вероятности ошибок для различного числа ветвей разнесения

$$p_2 = \frac{1}{h_{ср}^2 + 2}\left[1 - \frac{h_{ср}^2}{h_{ср}^2 + 2}\left(\frac{2}{h_{ср}^2 + 2}\right)^{\frac{2}{h_{ср}^2}}\right]. \qquad (7.6)$$

Это выражение позволяет упростить расчеты для наиболее часто используемых систем сдвоенного приема.

Система линейного сложения сигналов

В случае линейного сложения сигналов с одинаковыми средними уровнями (средней мощностью) в каждой ветви результирующий сигнал представляет собой сумму M случайных независимых функций, имеющих обычно рэлеевские распределения замираний.

Вероятность ошибки в схеме линейного сложения для сдвоенного приема

сигналов ЧМн при независимых рэлеевских замираниях при $h_{ср}^2 >> 1$ может быть найдена по формуле [42]:

$$p_2 \approx \frac{3{,}24}{h_{ср}^2}. \qquad (7.7)$$

Сравнивая записанное выражение с (7.2), убеждаемся, что энергетический проигрыш при переходе от оптимального приема к линейному не превышает 0,2 дБ.

На рис. 7.8 приведена упрощенная структурная схема приемного устройства с системой линейного сложения сигналов на промежуточной частоте, принимаемых двумя антеннами [22]. В состав схемы входят два антенных устройства с высокочастотными блоками (ВЧ), смесители (СМ), гетеродины (Гет), усилители промежуточной частоты (УПЧ), устройство сравнения и автоматической подстройки фазы (ФАПЧ) и устройство автоматической регулировки усиления (АРУ).

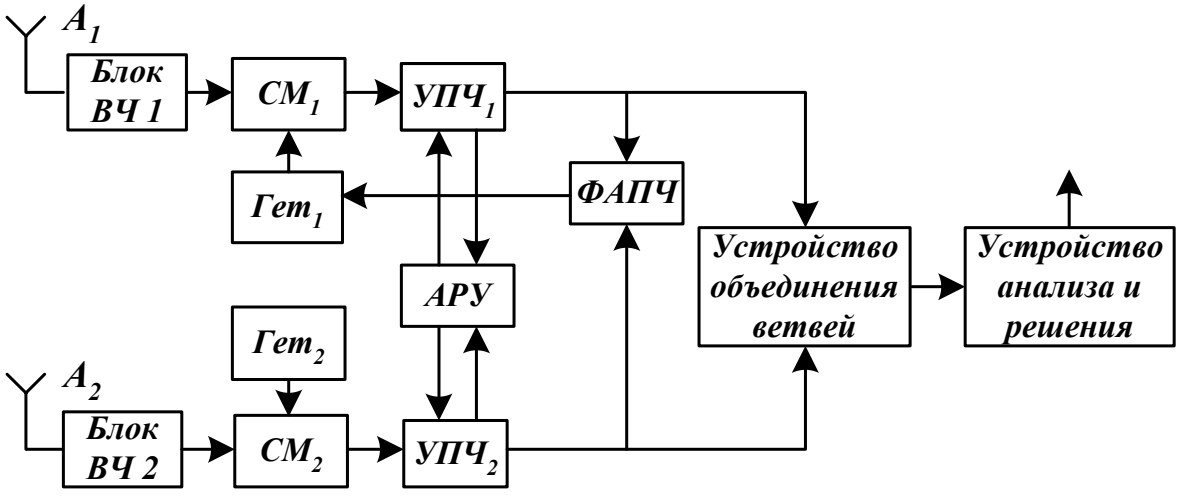


Рис. 7.8. Упрощенная структурная схема приемного устройства с системой линейного сложения сигналов на промежуточной частоте

Линейность сложения в схеме обеспечивается общим для двух приемников устройством АРУ. При такой схеме коэффициенты усиления обоих приемников близки и определяются наибольшим из складываемых сигналов. Система предусматривает когерентность складываемых сигналов, которая обеспечивается автоматической подстройкой фазы сигнала одного приемника к фазе сигнала другого приемника в ФАПЧ. Практически это достигается автоматической

подстройкой фазы одного из гетеродинов приемника.

Система оптимального линейного сложения сигналов

Принципиальным недостатком линейного сложения сигналов с одинаковыми весовыми коэффициентами является то, что ветви с плохим отношением сигнал/шум вносят заметный вклад в шумовую составляющую результирующего колебания и незначительный в сигнальную составляющую. Если выбирать весовые коэффициенты k_i при линейном объединении ветвей так, чтобы они учитывали фактическое состояние каждой ветви, определяемое величиной h_i^2, то при определенном правиле такого выбора можно добиться максимального отношения сигнал/шум на выходе устройства объединения. Так как величина h_i^2 в каждой ветви обычно медленно меняется во времени, то возможно обеспечить такое изменение весовых коэффициентов k_i для всех ветвей, при которых величина отношение сигнал/шум на выходе схемы объединения достигает своего максимального значения.

Вероятность ошибки при оптимальном линейном сложении ЧМн сигналов в условиях рэлеевских замираний можно рассчитать по формуле [6, 42]:

$$p_M = \frac{1}{2}\left(1+\frac{h_{cp}^2}{2}\right)^{-M}.$$

При большом отношении сигнал/шум в ветвях ($h_{cp}^2 >> 1$) и $M = 2$:

$$p_2 \approx \frac{2}{h_{cp}^4}. \qquad (7.8)$$

Сравнивая соотношения (7.8), (7.7), (7.6), (7.5), характеризующие вероятность ошибки при применении соответствующей системы разнесенного приема, можно сделать вывод, что схема оптимального линейного сложения сигналов обеспечивает наибольшую помехоустойчивость.

Для более полной характеристики методов разнесенного приема необходимо располагать законами распределения огибающей сигнала на выходе схемы объединения. Указанные кривые законов распределения для систем разнесенного приема при автовыборе и оптимальном линейном сложении ($M = 2$ и $M = 3$) приведены на рис. 7.9 [42].

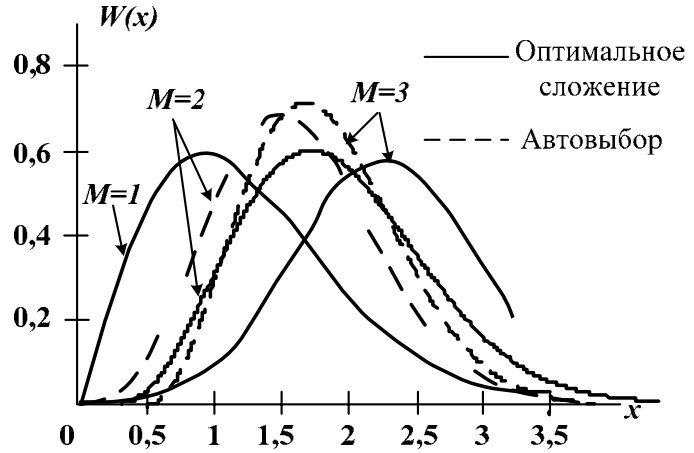


Рис. 7.9. Кривые законов распределения для систем разнесенного приема при автовыборе и оптимальном линейном сложении

На этом же рисунке представлен закон Рэлея, соответствующий одиночному приему флуктуирующего сигнала ($M = 1$). Из анализа рисунка следует, что при увеличении числа ветвей разнесения кривые распределения для оптимального линейного сложения смещаются вправо, приближаясь к нормальному распределению.

7.2.2. Методы борьбы с замираниями в цифровых системах связи

Основным методом повышения помехоустойчивости цифровых линий связи в условиях быстрых замираний также является разнесенный прием. Вместе с тем в цифровых системах передачи данных могут быть эффективно использованы широкополосные и составные сигналы, решающая обратная связь в сочетании с помехоустойчивым кодированием, а также методы адаптивного приема. Адаптация может производиться либо на приеме, либо на передаче. В последнем случае для получения информации о состоянии прямого тракта и соответствующего изменения параметров передаваемого сигнала используется канал обратной связи.

Наряду с известными методами разнесения (пространственным, частотным, угловым) особое внимание в цифровых системах уделяется комбинированным методам, например, частотно-временному, при котором предполагается использование последовательных или параллельных многочастотных сигналов (МЧС). Последовательные и параллельные МЧС при фазовой манипуляции показаны на рис. 7.10, где $T = \tau_1 + \tau_2 + \tau_3 + \tau_4$ – длительность информационной посылки; τ_i $i = 1,2,3,4$ – длительность передачи информационной посылки на частоте; f_k $k = 1,2,3,4$ – частоты МЧС. Очевидно, применение последовательного или параллельного МЧС предполагает использование соответственно одного или не-

скольких передающих устройств.

Частотное разнесение может осуществляться одним или несколькими передатчиками, но для передачи M сигналов одним передатчиком требуется его значительная пиковая мощность и высокая линейность амплитудной характеристики. Поэтому, как правило, параллельный МЧС реализуется несколькими передатчиками.

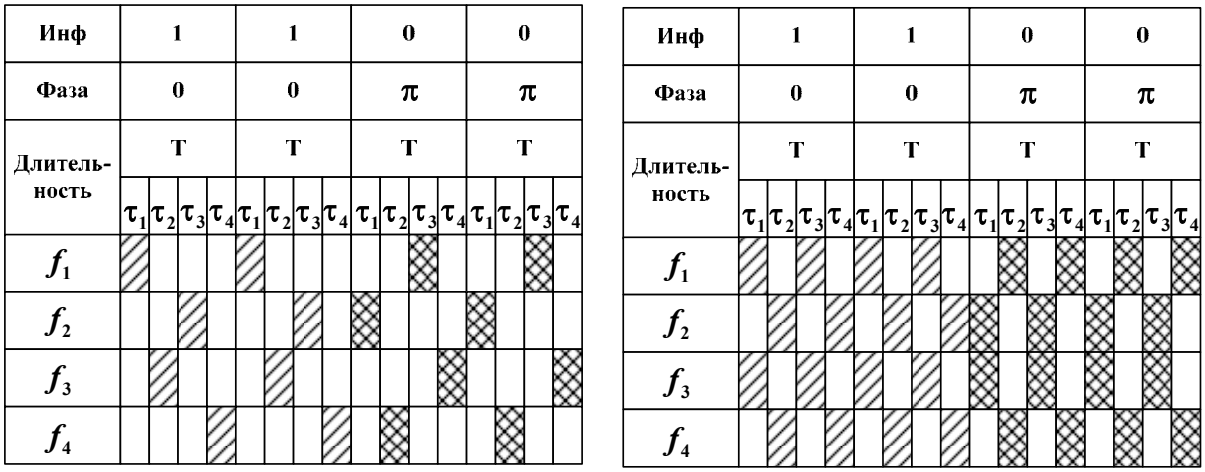


Рис. 7.10. Последовательный и параллельный многочастотные сигналы

Адаптивный прием сигналов основан на возможности измерения характеристик канала передачи сигналов и использовании полученных данных для соответствующей регулировки параметров передаваемого сигнала. Условием создания замкнутых адаптивных систем является наличие канала обратной связи. Упрощенная структурная схема адаптивной линии связи приведена на рис. 7.11. Она может обеспечивать функционирование линии как с одной, оптимальной в текущий момент времени частотой (f_i), так и с группой частот $(f_1, f_2, ..., f_M)$. Достоинство такой схемы – уменьшение габаритов и массы оборудования. Недостатками схемы являются наличие обратного канала, возможность реализации только автовыбора, а также переходные процессы вследствие переключений.

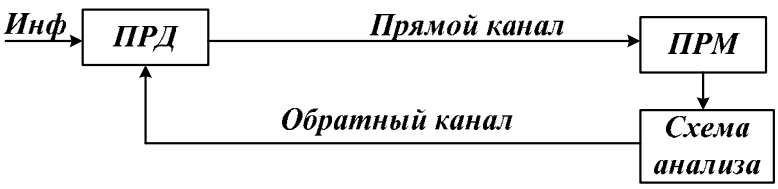


Рис. 7.11. Упрощенная структурная схема адаптивной линии связи

Обратная связь может быть использована для получения сведений об ис-

кажениях сигнала в прямом канале до демодуляции (обратная связь по сигналу). При этом на передающую сторону поступает информация об уровне сигнала или его искажениях. Обратная связь может быть также по сообщению. В этом случае она связывает выход демодулятора или декодирующего устройства с соответствующими блоками на передаче. При этом контролируются не только сигналы, но и решения, принимаемые приемным устройством. При передаче дискретной информации можно использовать оба типа обратной связи, тогда как для аналоговой информации используется только первый.

В реальных многолучевых каналах цифровой связи ошибки имеют тенденции к группированию. Основной причиной этого является падение значения сигнала ниже допустимого, при котором появляются длинные серии ошибок. Для борьбы с этим явлением трудно использовать какой-либо исправляющий код, поскольку требуемая избыточность кода будет очень велика.

Для борьбы с сериями ошибок можно использовать временное разнесение сигналов, основанное на передаче одной и той же информации в моменты времени, отстоящие друг от друга на время Δt, превышающее длительность замираний.

Возможным путем реализации помехоустойчивой передачи информации в этих условиях является создание системы с декорреляцией ошибок. При этом сообщение кодируется обычным способом, но соседние символы кодовой комбинации передаются по радиоканалу не в реальном масштабе времени, а через промежутки времени, близкие к Δt. В свободные промежутки времени передаются символы других кодовых комбинаций.

7.3. Методы борьбы с межсимвольной интерференцией

Особенностью радиосвязи на большие расстояния часто является передача информации в условиях общих замираний и межсимвольной интерференции.

Межсимвольная интерференция (МСИ) это искажения сигнала за счет откликов на более ранние символы, которые могут проявлять себя как помехи. МСИ зависит от вида АЧХ и ФЧХ фильтров в тракте передаче, структуры и па-

раметров кодовой последовательности.

7.3.1. Причины возникновения и сущность межсимвольной интерференции

Спецификой многих линий дальней радиосвязи (тропосферных, спутниковых и др.) является многолучевый характер распространения радиосигнала (рис. 7.1). Сигнал в точке приема представляет собой сумму большого числа элементарных сигналов с разными амплитудами и случайным временем запаздывания. Отдельные лучи могут запаздывать друг относительно друга на значительную величину, что и вызывает МСИ. В зависимости от степени искажения формы импульса различают большие (рис. 7.12) и малые (рис. 7.13) межсимвольные помехи.

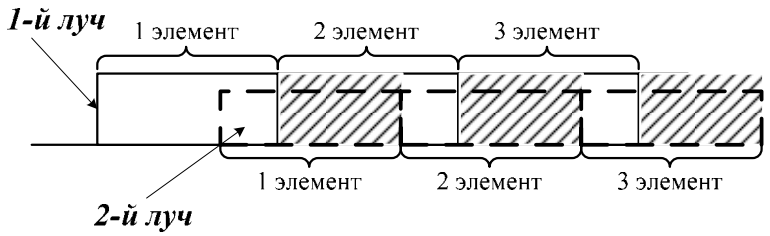


Рис. 7.12. Большие межсимвольные помехи

Степень искажения формы импульса при наложении сигналов зависит от разности времен распространения радиоволн по различным путям. Обычно разность времени распространения по максимальному и минимальному путям называют временем многолучевости ($\Delta\tau_{мл}$). Для расстояний связи $R \cong 150км$ величина $\Delta\tau_{мл}$ лежит в пределах 0,2—0,5 мкс. Если длительность импульса (T) меньше времени многолучевости то возникают большие межсимвольные помехи. Если же длительность импульса намного превышает время многолучевости, то межсимвольные помехи мало влияют на прием, т.к. в данном случае лишь небольшая часть элемента оказывается пораженной помехой.

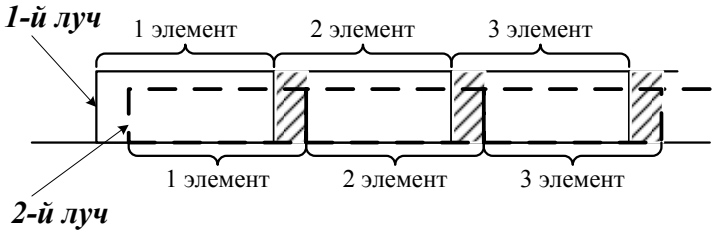


Рис. 7.13. Малые межсимвольные помехи

7.3.2. Обработка сигналов в каналах с межсимвольной интерференцией

Помехоустойчивость цифровых радиосистем в низкоскоростном режиме работы характеризуется малым влиянием помех межсимвольной интерференции (МСИ), поскольку в этом случае длительность импульса (T) много больше времени многолучевости ($\Delta\tau_{мл}$) (рис. 7.14). В общем случае для борьбы с МСИ применяются следующие методы.

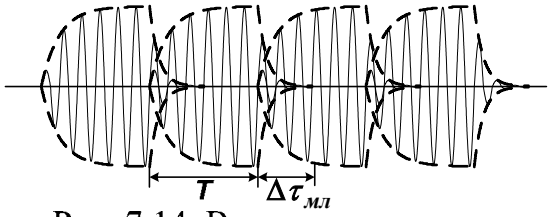


Рис. 7.14. Вид радиосигнала на входе приемника в канале с МСИ

Прием со стробированием импульсов

Сигнал в приемнике подвергается стробированию, т. е. из посылки длительностью T вырезается та ее часть, где проявление переходных процессов от предыдущего символа минимально (рис. 7.15).

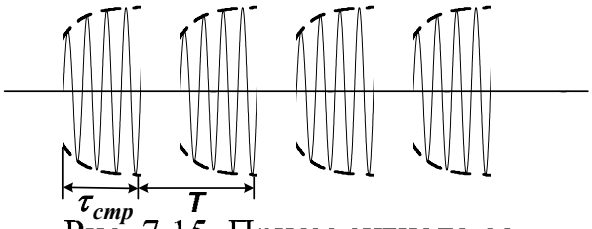


Рис. 7.15. Прием сигнала со стробированием импульсов

Применение сигналов с пассивной паузой

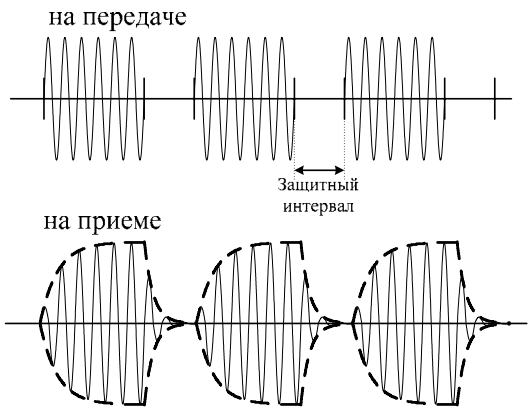


Рис. 7.16. Передача сигнала с пассивной паузой

При передаче сигнала между информационными импульсами вводится защитный интервал длительностью больше, чем $\Delta\tau_{МЛ}$. В этом случае влияние лучей смежных информационных символов исключается. Форма сигнала, соответствующая рассмотренному случаю, представлена на рис. 7.16.

Применение многопозиционных сигналов

Использование многопозиционной манипуляции с основанием кода $m > 2$ позволяет уменьшить техническую скорость передачи B по сравнению с ин-

формационной скоростью V в $\log_2 m$ раз. Например, при четырехпозиционной манипуляции $\log_2 4 = 2$ $B/V = 0{,}5$, т.е. длительность элементарных посылок, излучаемых передатчиком, увеличивается в два раза.

Оптимальный метод приема требует разработки алгоритмов разделения лучей, их фазирования, взвешивания по максимуму отношения сигнал/шум (масштабирования) и квазикогерентного весового сложения.

7.3.3. Помехоустойчивость в каналах с межсимвольной интерференцией

Помехоустойчивость в условиях МСИ оценивается вероятностью ошибки [42]:

$$p_{ош,мл} = \frac{\Delta\tau_{мл}^2}{3\lambda_E T^2}\left[1 + \ln\left(\frac{3\lambda_E T^2}{4\pi\Delta\tau_{мл}^2} + 1\right)\right],$$

где λ_E – параметр учитывает метод манипуляции: для когерентного приема ОФМн $\lambda_E = 2{,}7 \div 3$; для некогерентного – $\lambda_E = 2$. Приведенное выражение относится к случаю однократной передачи (одной ветви приема).

Вероятность ошибки с использованием основных методов устранения межсимвольных помех находится по следующим формулам.

При введении защитного интервала на передаче или приеме:

$$P_{МСИ}^{ЗИ} = 0.65 \cdot \exp\left(-0.443 \cdot \left(\lambda_E \cdot h \cdot \sqrt{D_{МЛ}} + 0.75\right)^2\right),$$

Для случая применения m - позиционного кодирования:

$$P_{МСИ}^{m} = \left[1 - F\left(\sqrt{2h^2}\sin\frac{\pi}{m}\right)\right],$$

Если используется обратная связь по решению, то:

$$P_{МСИ}^{ОСР} = \left[1 - F\left(\sqrt{2h^2}\right)\right] + \left[1 - F\left(\sqrt{4h^2\left(1 - R\sqrt{\beta}\right)}\right)\right],$$

где R и β – параметры описывающие импульсную реакцию канала.

На основе расчетов построены графики (рис. 7.17), позволяющие сравнить помехоустойчивость СПИ в условиях МСИ. Как видно из рисунка, ис-

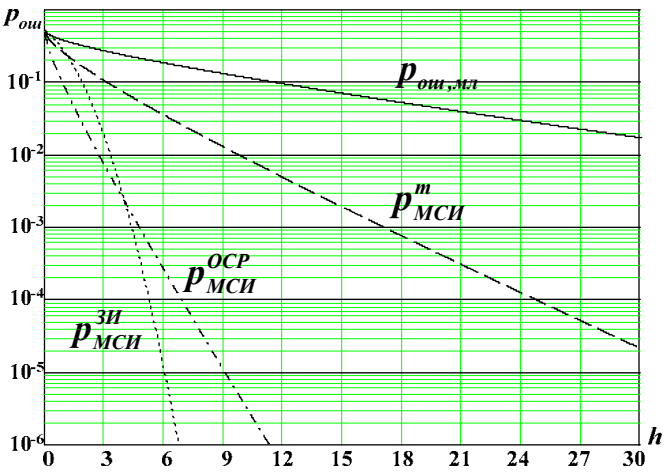


Рис. 7.17. Вероятность ошибки приема дискретных сигналов в условиях МСИ

пользование любого метода устранения межсимвольных помех обеспечивает выигрыш по помехоустойчивости. При малых отношениях сигнал/шум наибольший выигрыш (до 10 дБ) обеспечивают системы с обратной связью по решению, а при больших отношениях – введение защитного интервала.

7.4. Прием дискретных сообщений в каналах с сосредоточенными по спектру и импульсными помехами

В реальных каналах связи наряду с флуктуационными гаусовскими помехами типа белого шума действуют сосредоточенные по времени (импульсные) помехи и сосредоточенные по спектру помехи.

7.4.1. Общая характеристика сосредоточенных по спектру и импульсных помех

Во многих случаях помеха состоит из отдельных импульсов, длительность которых τ_n существенно меньше длительности элемента сигнала T [6, 8], а спектр помехи значительно шире спектра сигнала. Такие помехи называются импульсными (рис. 7.18).

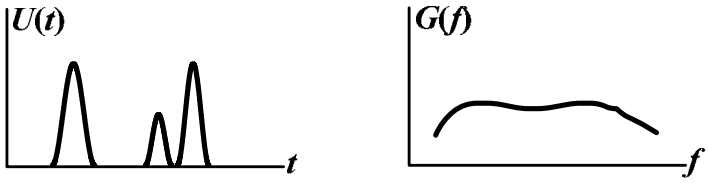


Рис.7.18. Импульсная помеха во временной и частотной областях

К сосредоточенным по времени (импульсным) помехам относятся помехи в виде одиночных коротких импульсов различной интенсивности и длительности, следующих через случайные, достаточно большие промежутки времени. Причинами импульсных помех являются грозовые разряды, радиостанции, работающие в импульсном режиме, линии электропередачи и другие энергоустановки, системы энергообеспечения транспорта и др.

Кроме импульсных помех, могут существовать протяженные по времени помехи, спектр которых занимает такую же полосу частот, как и сигнал, или даже более узкую. Эти помехи называют сосредоточенными по спектру (рис. 7.19).

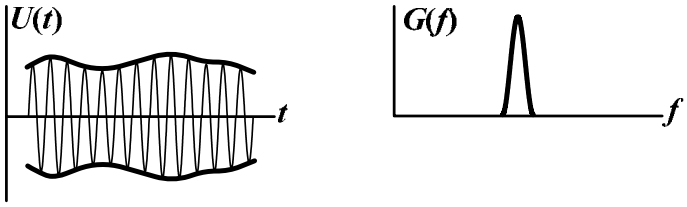


Рис.7.19. Узкополосная помеха во временной и частотной областях

К сосредоточенным по спектру помехам относятся помехи посторонних радиостанций, генераторов высокой частоты различного назначения (медицинские, промышленные, бытовые и др.), переходные помехи от соседних каналов многоканальных систем. Обычно это гармонические или модулированные колебания с шириной спектра меньшей или соизмеримой с шириной спектра полезного сигнала. В диапазоне декаметровых волн такие колебания являются основным видом помех.

Воздействие сосредоточенной по спектру помехи

Поскольку далеко не всякий элемент сигнала принимается в присутствии сосредоточенной помехи, то вероятность ошибки p можно выразить произведением:

$$p = p_n \cdot p\left(o/n\right). \qquad (7.9)$$

где p_n – вероятность того, что на вход решающей схемы поступила сосредоточенная помеха, а $p\left(o/n\right)$ – условная вероятность того, что произойдет ошибка символа при воздействии сосредоточенной помехи, которая зависит от мощности сигнала, мощности сосредоточенной помехи, вида сигнала, частоты сигнала, частоты помехи и т.д. Для различных систем связи определены аналитические зависимости условной вероятности ошибок от названных факторов, эти зависимости можно найти в специальной литературе [6, 8].

Теоретические и экспериментальные исследования показывают, что в любых системах связи существует некоторое отношение $P_n/P_c = k$, называемое

порогом или коэффициентом помехоустойчивости, такое, что при $P_n/P_c < k$ условная вероятность ошибки $p(o/n) \cong 0$. Если же отношение P_n/P_c выше k, то условная вероятность ошибки может быть велика. Экспериментальные исследования реальных приемников ЧМ показывают, что при $P_n/P_c < 0{,}85 - 0{,}95$ ошибки не возникают, а при $P_n/P_c > 1$ вероятность ошибки практически равна $1/2$. Коэффициент k различен для разных систем связи. Так, для когерентной системы с фазовой манипуляцией $k = 1$, для системы с амплитудной манипуляцией $k = 1/2$. Коэффициент k может быть и значительно выше единицы, если используются широкополосные сигналы, занимающие полосу частот $F >> \frac{1}{T}$.

Воздействие импульсной помехи

Для вероятности ошибки, вызываемой импульсной помехой, также справедливо выражение (7.9), где под p_n следует теперь понимать вероятность того, что за время существования элемента сигнала на вход решающей схемы поступил импульс помехи, а под $p(o/n) = 0{,}5$ – условную вероятность ошибочного приема символа, при условии прихода импульса помехи. Воздействие импульсной помехи на прием дискретных сигналов тоже носит пороговый характер. Если интенсивность импульсной помехи (на входе решающей схемы) меньше некоторой величины, то она не вызывает ошибок, т. е. $p(o/n) = 0$. При увеличении интенсивности сверх этой величины условная вероятность ошибок быстро возрастает.

7.4.2. Борьба с сосредоточенными и импульсными помехами

Борьба с сосредоточенными и импульсными помехами может быть направлена либо на понижение вероятности p_n попадания помехи с уровнем более порогового на вход решающей схемы, либо на уменьшение уровней вероятности ошибки $p(o/n)$. В свою очередь, вероятность попадания помехи на вход

решающей схемы можно уменьшить, воздействуя на источники помех либо на структуру приемного устройства. Поэтому все действия по борьбе с помехами подразделяются на три группы [6, 8, 20, 21]:

борьба с помехами в месте их возникновения;

защита от попадания помех на вход решающей схемы;

повышение помехоустойчивости системы связи путем выбора соответствующих форм сигналов.

Борьба с помехами в месте их возникновения

Все источники помех с точки зрения борьбы с ними можно подразделить на контролируемые, находящиеся в пределах рассматриваемой системы, и неконтролируемые, находящиеся вне системы, следовательно, не поддающиеся непосредственному воздействию или регулированию.

Для уменьшения уровней сосредоточенных и импульсных помех во всех развитых странах мира разработаны законодательные акты, регламентирующие допустимый уровень и частотный диапазон электромагнитных излучений. С этой же целью существуют международные органы, разрабатывающие допустимые нормы и контролирующие их соблюдение отдельными странами.

Следует иметь в виду, что выделенный нашей стране международными соглашениями диапазон частот является национальным достоянием и эффективное его использование представляет собой не только техническую, но и государственную задачу.

Для уменьшения уровней сосредоточенных и импульсных помех до международных или государственных норм и, следовательно, для уменьшения вероятности попадания p_n помехи предусматриваются:

уменьшение уровня и ширины спектра побочных излучений передающих устройств при строгой регламентации допустимой ширины полезной части спектра сигнала, а также ограничение излучаемой мощности;

экранировка излучающих блоков аппаратуры связи, постановка схем искрогашения на различных энергетических устройствах промышленной, научной, медицинской или бытовой аппаратуры;

целесообразное размещение электрических систем, в частности, средств связи на местности, при одновременной регламентации работы системы по времени;

оптимальное распределение и назначение частот всем видам радиотехнических систем, обеспечивающее минимально возможные взаимные помехи.

Защита от попадания сосредоточенных по спектру помех на вход решающей схемы

Защита от узкополосных помех в радиосвязи является одной из наиболее важных задач, решаемых при разработке радиоприемных и антенных устройств. Способность приемного устройства пропустить на вход решающей схемы сигнал и задержать или существенно ослабить сосредоточенные помехи называется избирательностью. Она обеспечивается благодаря отличиям сигнала от помех по направлению прихода (пространственная избирательность), по спектру (частотная избирательность), по времени существования (временная избирательность), по начальной фазе (фазовая избирательность) и по форме (избирательность по форме).

Наибольшее значение в радиосвязи имеют пространственная и частотная избирательность. Пространственную избирательность обеспечивают узконаправленные приемные антенны.

Частотная избирательность основана на том, что каскады приемника до решающей схемы обладают частотной характеристикой, пропускающей только ту часть спектра, где расположена основная мощность сигнала, и сильно подавляющей остальные участки спектра, в которых может находиться помеха. Если спектр сигнала шире спектра помехи, то иногда можно избавиться от помехи, лежащей в той же полосе частот, что и сигнал. Для этого сумму сигнала и помехи пропускают через режекторный (заграждающий) фильтр, настраиваемый так, чтобы «вырезать» ту полосу частот, где сосредоточена помеха, сохранив достаточную часть спектра сигнала, чтобы по ней можно было восстановить переданное сообщение.

Методы компенсации импульсных помех, несмотря на все их многообра-

зие, основаны на широкополосности спектра помехи, что позволяет построить дополнительный компенсационный тракт (рис.7.20), расстроенный относительно частоты сигнала.

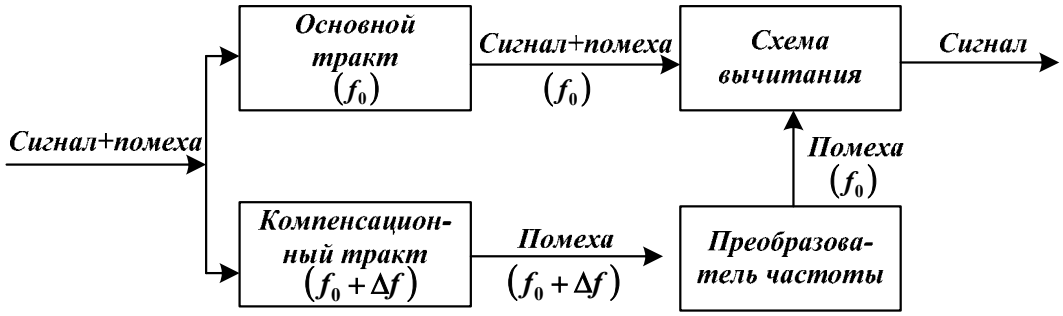


Рис.7.20. Обобщенная структурная схема компенсатора импульсных помех

Сигнал проходит только через основной тракт, тогда как импульсная помеха создает напряжение на выходах обоих трактов. С помощью преобразователей частоты и фазовращателей помеха в компенсационном тракте преобразуется так, чтобы она совпадала с помехой в основном тракте, что позволяет произвести компенсацию в схеме вычитания. Несмотря на кажущуюся простоту этой схемы, в действительности трудно добиться хорошей компенсации помехи, так как для этого необходима высокая стабильность амплитудно-частотных и фазо-частотных характеристик обоих трактов. Другой недостаток схемы заключается в том, что наличие компенсационного тракта приводит к ухудшению помехоустойчивости относительно флуктуационных и сосредоточенных по спектру помех.

Наиболее широко применяются методы защиты от импульсных помех, основанные на амплитудном ограничении. Поясним сущность этих методов, полагая вначале, что импульсная помеха состоит из идеальных дельта функций. Предположим, что на входе приемника включен двусторонний амплитудный ограничитель, характеристика которого показана на рис.7.21,а.

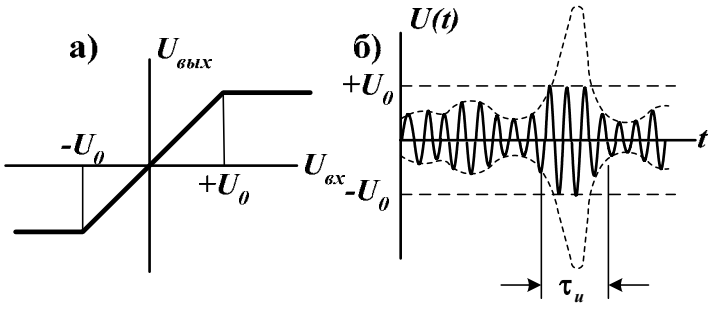


Рис.7.21. Характеристика ограничителя («а») и сигнал на его входе и выходе («б»)

При подаче на его вход напряжения $U_{вх}$, лежащего в пределах $-U_0 < U_{вх} < U_0$, на выходе получится напряжение $U_{вых} = U_{вх}$ [6, 42].

Но если входное напряжение превысит U_0, то выходное напряжение окажется ограниченным и не будет по абсолютной величине превосходить U_0 (рис.7.21,б). Если уровень ограничения U_0 выбран выше максимального напряжения, создаваемого сигналом, флуктуационными и сосредоточенными помехами, то в отсутствие импульсной помехи тракт приемника будет линейным. При появлении импульсной помехи она окажется ограниченной по уровню U_0. Поскольку ее длительность мала, то мала и равна ее «площадь» и спектральная плотность энергии. Такая ограниченная импульсная помеха вызовет незначительную реакцию в фильтрах решающей схемы и, следовательно, не будет создавать ошибок.

Подавление сосредоточенных по спектру помех требует строгой линейности тракта приемника вплоть до фильтра, выделяющего спектр сигнала. Ограничитель же представляет собой принципиально нелинейное устройство, и если на его вход поступят сосредоточенные по спектру помехи, то возникающие комбинационные частоты с большой вероятностью могут попасть в полосу частот, занимаемую сигналом.

Таким образом, требования к схеме приемника для защиты от импульсных и узкополосных помех оказываются взаимно противоречивыми. Для подавления импульсных помех следует вводить нелинейность (ограничитель) в той части тракта, которая предшествует фильтру, определяющему частотную избирательность. Подавление же узкополосных помех, наоборот, связано с требованием линейности этой части тракта.

Более «терпимыми» к сосредоточенным помехам являются методы защиты от импульсных помех, основанные на запирании приемника на время действия импульсной помехи. Такая схема работает в линейном режиме, пока нет импульсных помех. При возникновении импульса срабатывает устройство мгновенной автоматической регулировки усиления (МАРУ), снижающей уси-

ление приемника практически до нуля, т.е. до его полного запирания (рис. 7.22).

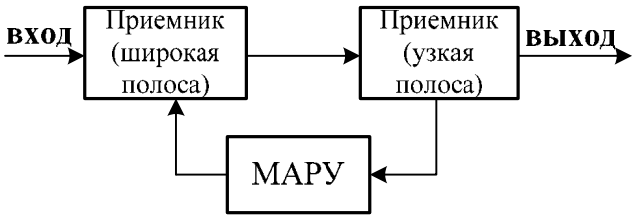


Рис.7.22. Схема с запиранием приемника на время действия импульсной помехи

Чтобы время запирания приемника было достаточно малым и не охватывало значительную часть элемента сигнала, необходимо устройство МАРУ располагать в широкополосной части тракта, где длительность импульсов помехи существенно меньше T.

Одновременная защита от импульсных и сосредоточенных помех

Для подавления импульсных помех с сохранением удовлетворительной избирательности относительно узкополосных помех часто применяется способ, получивший название ШОУ (широкая полоса – ограничитель – узкая полоса). Его сущность заключается в том, что для подавления импульсной помехи используется амплитудный ограничитель, который включается между двумя фильтрами (рис. 7.23). Первый из этих фильтров, называемый широкополосным, обеспечивает отсеивание сосредоточенных помех, расположенных на оси частот достаточно далеко от спектра сигнала, но имеет полосу пропускания $F_ш$ более широкую, чем полоса частот F_y, занимаемая сигналом.

Рис.7.23. Схема с использованием амплитудного ограничителя между двумя фильтрами

Сосредоточенные помехи попадают в полосу пропускания $F_ш$ с большей вероятностью, чем в полосу частот F_y. Однако вероятность того, что в полосе пропускания $F_ш$ будет много мощных сосредоточенных помех, которые на выходе ограничителя создадут комбинационные частоты, попадающие в спектр сигнала, невелика, так как $F_ш$ не намного больше F_y. Все же помехи, которые прошли через широкополосный фильтр и ограничитель, не следует пропускать

на решающую схему, поскольку они могут вызвать ошибки. Для этого служит узкополосный фильтр, с полосой пропускания F_y, завершающий функции частотной избирательности, в частности, узкополосный фильтр может быть согласован с сигналом [6, 42].

Заметим, что мощная помеха, прошедшая через широкополосный фильтр, может при прохождении через ограничитель «подавить» сигнал, т. е. сильно уменьшить его мощность. Несмотря на то, что последующий узкополосный фильтр и отсеет эту помеху, мощность сигнала может оказаться недостаточной для нормальной работы решающей схемы. Поэтому в схеме ШОУ всегда предусматривается большой запас усиления после узкополосного фильтра.

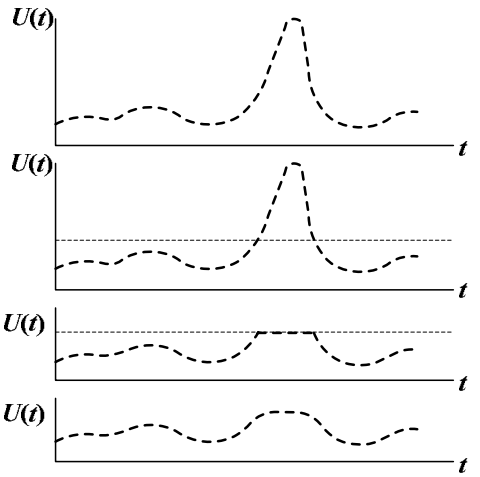


Рис. 7.24. Изменение огибающей импульса в схеме ШОУ

Для пояснения физического процесса подавления импульсной помехи в схеме ШОУ на рис. 7.24 показаны изменения огибающей импульса. На выходе широкополосного фильтра импульс имеет большую амплитуду, но относительно малую длительность. Ограничитель «срезает» амплитуду импульса и делает ее равной амплитуде сигнала, но не изменяет его длительности. В узкополосном фильтре длительность импульса увеличивается примерно в $\frac{F_ш}{F_y}$ раз и во столько же раз уменьшается его амплитуда. Благодаря этому амплитуда импульсной помехи на выходе схемы ШОУ оказывается приблизительно в $\frac{F_ш}{F_y}$ раз меньше амплитуды сигнала и не вызывает ошибок.

Более эффективными путями одновременной защиты от сосредоточенных и импульсных помех, являются комбинированные методы разнесенного приема, например, разнесенный прием, одновременно по времени и по частоте. Из ветвей частотного разнесения выбирается та, в которой меньше интенсивность сосредоточенных помех, а из ветвей разнесения по времени – такая в которой

отсутствует импульсная помеха. Если число ветвей достаточно велико, то с большой вероятностью найдется одна ветвь, не пораженная помехой.

7.5. Компенсация помех и искажений в канале

Анализ различных показателей эффективности функционирования СЭС позволяет сделать вывод, что защитить передаваемую информацию от помех можно либо повышением энергетики линий связи (за счет увеличения мощности передатчика, применением специальных антенн и др.), либо усложнением сигналов и алгоритмов их обработки.

Первый путь во многих случаях неприемлем ввиду ограничения на излучаемую мощность по требованиям электромагнитной совместимости, технико-экономическим, эксплуатационным и другим соображениям. Развитие элементной базы позволяет считать второй путь более перспективным.

Помехозащищенность систем связи можно повысить за счет применения широкополосных сигналов.

Широкополосными называют сигналы, у которых произведение ширины спектра на длительность намного больше единицы ($F \cdot T >> 1$). Такое произведение называют базой сигнала, $B = F \cdot T$ [18].

База сигнала может быть расширена B- кратным повторением сигнала во времени, либо в отведенной полосе частот (рис.7.25).

Существует два основных способа расширения базы сигналов:

скачкообразное изменение несущей частоты, при котором каждый символ сообщения передают с помощью набора дискретных частот (fast frequency hopping) (рис.7.25,а);

прямое расширение спектра частот, за счет уменьшения длительности единичного элемента (direct sequence spreading) (рис.7.25,б).

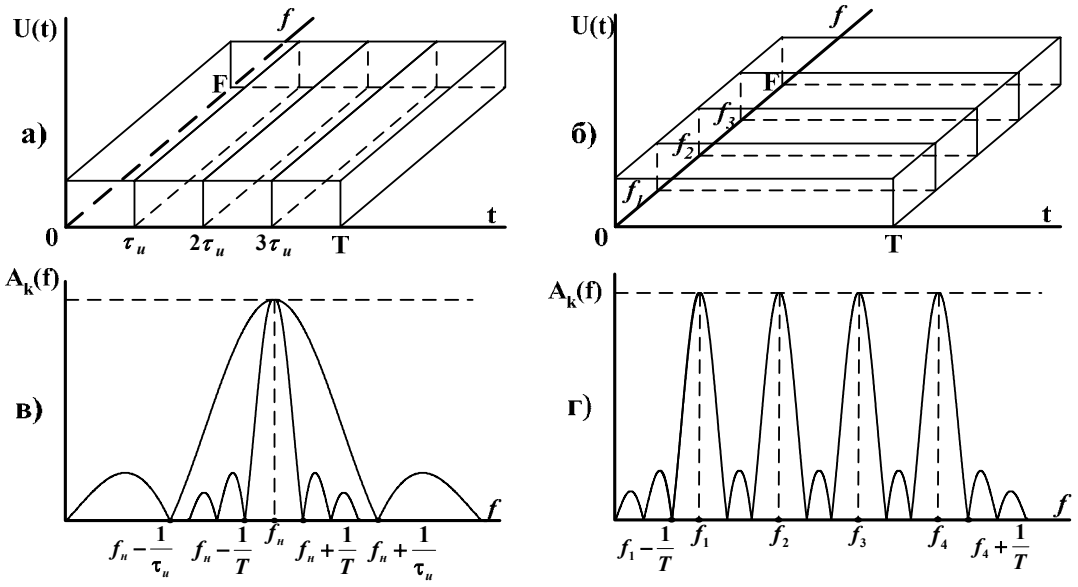


Рис. 7.25. Сигналы с В-кратным повторением по времени («а»), по частоте («б») и их спектры («в», «г»)

В последние годы сложные сигналы стали находить применение не только в службах ведомственной связи, но и в коммерческих системах телекоммуникаций. Это объясняется постоянным ростом потребности в услугах радиосвязи, а поскольку выделенный частотный ресурс жестко ограничен, приходится использовать его более эффективно. На применении сложных сигналов базируется технология, позволяющая использовать уже занятые частотные диапазоны при соблюдении условий полной электромагнитной совместимости. В ближайшем будущем стратегия развития систем радиосвязи будет заключаться в улучшении показателей спектральной эффективности, возрастании надежности оборудования, повышении качества и снижении стоимости услуг связи. Поэтому сложные сигналы, на применении которых основана CDMA-технология, будут широко использоваться в перспективных системах связи.

Основная идея технологии сложных сигналов базируется на преобразовании узкополосных сигналов с шириной спектра $\Delta F = 1/T$ в широкополосные сигналы с шириной спектра $\Delta f = 1/\tau$ при постоянстве энергии сигналов E.

Сигналы с большой базой обеспечивают ряд преимуществ:

высокую помехозащищенность систем связи:

эффективную борьбу с искажениями сигналов в канале связи:

одновременную работу многих абонентов в общей полосе частот за счет кодового разделения каналов:

совместимость передачи информации с измерением параметров движения объектов;

более эффективное использование спектра частот на ограниченной территории.

Общая тенденция цифровизации систем связи повышает значимость сложных сигналов. Так, одной из особенностей построения цифровых систем передачи является возможность использования аппаратуры помехозащищенной радиосвязи (АПР) входящей в состав высокочастотного тракта. АПР предназначена для формирования широкополосного сигнала на передачу и обработки его на приеме. Могут использоваться методы фазовой манипуляции широкополосными сигналами (ФМ-ШПС) (иначе называемый ФМ-ПСС – фазовая манипуляция псевдослучайными сигналами) или псевдослучайное переключение рабочих частот (ППРЧ) в сочетании с режекцией помех по частоте.

В перспективе могут найти применение более сложные виды сигналов, например, манипуляция с минимальным частотным сдвигом, с когерентным приемом. Этот вид модуляции позволяет снизить уровень внеполосных излучений, а также повысить помехоустойчивость.

Использование сложных сигналов и специальных методов их обработки в цифровых системах передачи позволяет повысить кратность разнесения при приеме, избавляясь от необходимости иметь на линии большой энергетический запас на замирания. На практике нашли применение параллельные и последовательные многочастотные сигналы.

7.5.1. Принцип работы радиолинии с ФМ ПСС (ФМ ШПС)

Широкое применение в технике связи из класса сложных сигналов нашли фазоманипулированные псевдослучайные сигналы (ФМ ПСС). Структурная схема СЭС с ФМ ПСС представлена на рис. 7.26.

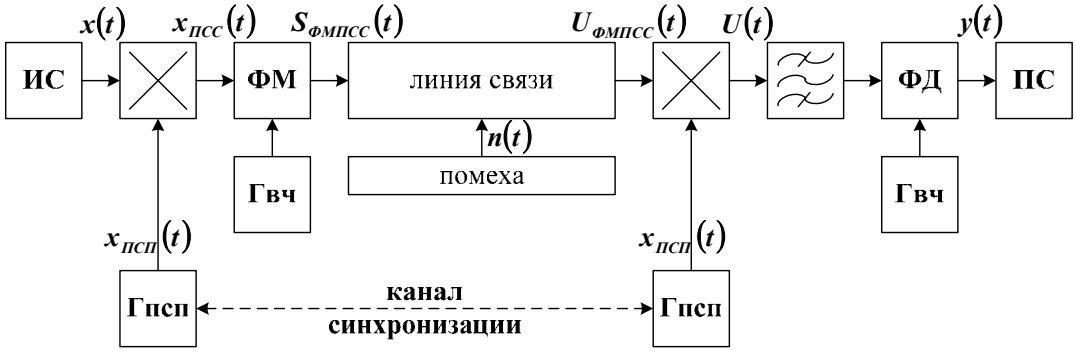


Рис.7.26 Структурная схема СЭС с ФМ ПСС

В состав схемы входят: ИС – источник дискретных сообщений; $\Gamma_{псп}$ – генератор псевдослучайной последовательности (ПСП); $\Gamma_{вч}$ – генератор ВЧ колебаний; ⊠ – умножитель; ФМ – фазовый модулятор; ФД – фазовый демодулятор; ПС – получатель дискретных сообщений.

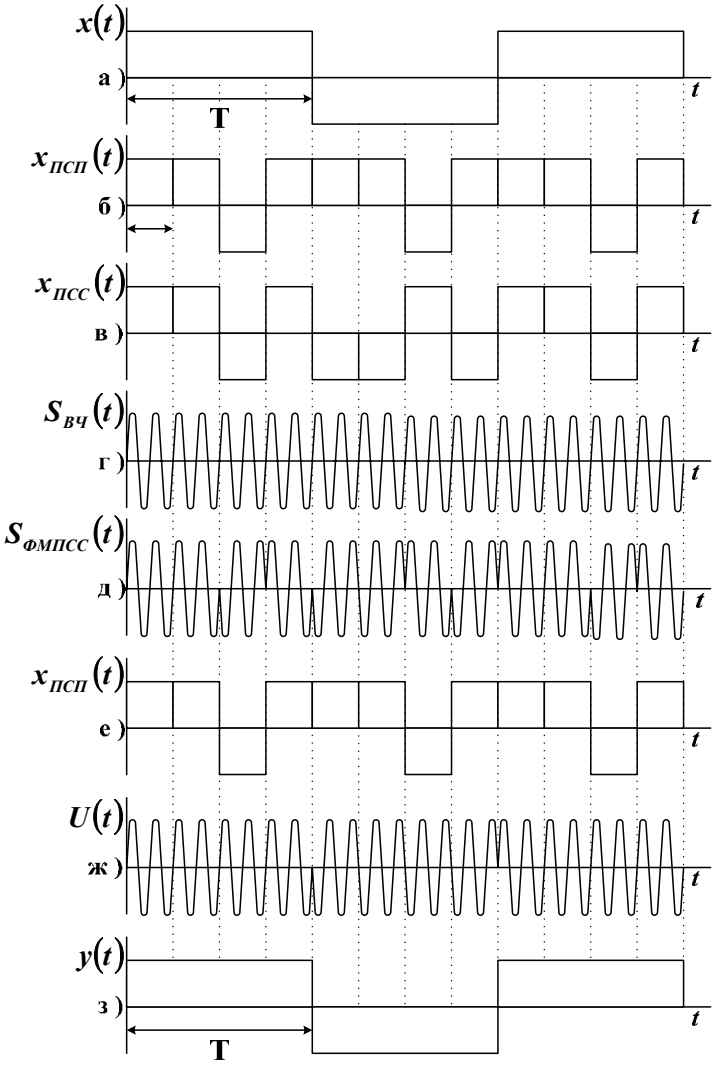


Рис.7.27. Формирование и обработка ФМ ПСС

Вид сигналов при формировании и обработке ФМ ПСС представлен на рис. 7.27. Источник сообщений выдает одно из двух сообщений: $+1$ или -1 с равной вероятностью и длительностью T (рис. 7.27,а). Сообщения следуют со скоростью V_u.

Генератор ПСП формирует последовательность импульсов длительностью τ_u, которая следует со скоростью V_m (рис. 7.27,б).

На входы фазового модулятора поступают соответственно последовательность $x_{ПСС}(t)$ и напряжение опорного (высокочастотного) колебания $S_{ВЧ}(t)$ (рис. 7.27,г). С выхода ФМ фазоманипулированный псевдослучайный сигнал

$S_{ФМПСС}(t)$ (рис. 7.27,д) поступает в линию связи.

На приемной стороне генератор ПСП синхронизирован с передающим генератором ПСП и выдает на вход умножителя точно такую же псевдослучайную последовательность (рис. 7.27,е). На выходе умножителя будет наблюдаться фазоманипулированная последовательность с изменениями фазы колебания не чаще, чем через длительность T (в соответствии с длительностью первичного сигнала) (рис. 7.27,ж). Пройдя через полосовой фильтр ($F = 1/T$), радиосигнал поступает на вход фазового демодулятора, с выхода которого получателю будет поступать исходное информационное сообщение со скоростью $V_и$ (рис. 7.27,з). Нетрудно показать, что база ФМ ПСС $B = F \cdot T = T/\tau_и = V_T/V_и$, где T – длительность информационного символа; $\tau_и$ – длительность импульса ПСП.

7.5.2. Помехоустойчивость радиолинии с ФМ ПСС

Помехоустойчивость радиолинии передачи дискретных сообщений характеризуется вероятностью ошибки на выходе демодулятора. Оценим, как будет изменяться вероятность ошибочного приема при применении широкополосного сигнала с базой B.

Для определенности положим, что мощность преднамеренной помехи значительно больше мощности внутренних шумов приемника и что энергия помехи равномерно распределена во всей полосе частот радиосигнала Δf.

Отношение сигнал/помеха на входе приемника

$$h^2_{вх} = \frac{E_{c0}}{E_{n0}} = \frac{E_{c0}}{G(f) \cdot \Delta f},$$

где E_{c0} – энергия единичного импульса ПСП; $E_{n0} = G(f) \cdot \Delta f$ – энергия помехи в полосе Δf; $G(f)$ – спектральная плотность мощности помехи.

При когерентном способе обработки ФМ сигнала $P_{ош} = \left[1 - F\left(\sqrt{2h^2}\right)\right]$, где $h^2 = \frac{E_c}{E_n}$ – отношение сигнал/помеха (ОСП) на входе демодулятора.

Энергия сигнала на выходе полосового фильтра, согласованного с шири-

ной спектра первичного сигнала длительностью T, $E_c = B \cdot E_{c0}$, а энергия помехи $E_n = B \cdot G(f) \cdot \Delta F = B \cdot G(f) \cdot \frac{\Delta f}{B} = E_{n0}$. Отсюда $h^2 = E_c / E_n = B \cdot h_{вх}^2$, т.е. ОСП на входе демодулятора в B раз больше, чем на входе приемника. Таким образом, для ФМ ПСС при когерентном способе обработки

$$p_{ФМПСС}^{кг} = \left[1 - F\left(h_{вх}\sqrt{2B}\right)\right].$$

При некогерентном способе для ОФМ ПСС:

$$p_{ОФМПСС}^{нкг} = \frac{1}{2} \cdot e^{-B \cdot h_{вх}^2}.$$

Следовательно, если использовать широкополосный сигнал с базой намного большей единицы это позволит получить выигрыш в отношении сигнал/шум на входе демодулятора в B раз.

7.5.3. Принципы работы радиолиний с ППРЧ

Системы связи, использующие для передачи информации несколько частот, выбираемых по определенному закону, называют системами связи с псевдослучайным переключением рабочих частот (ППРЧ). Структурная схема системы связи с ППРЧ представлена на рис. 7.28.

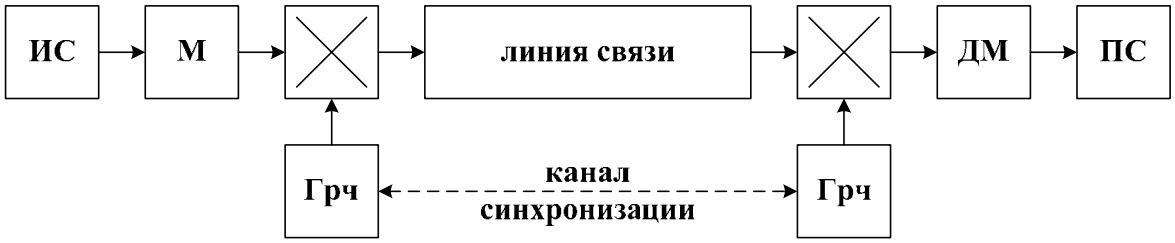


Рис. 7.28. Структурная схема СЭС с ППРЧ

Поясним принцип работы схемы. На один вход умножителя подается сигнал на промежуточной частоте с выхода модулятора, на второй вход – частота с выхода генератора рабочих частот (Грч). Генератор рабочих частот вырабатывает ряд частот, которые по определенному закону подаются на вход умножителя. Частота сигнала в радиолинии будет изменяться в соответствии с этим законом.

На приемной стороне осуществляется обратная операция за счет синхронной работы генераторов рабочих частот приемника и передатчика.

В системах связи с ППРЧ передавать информацию можно с помощью модуляции любого вида, хотя реализация ФМ затруднена из-за сложности обеспечения фазовой синхронизации радиосигнала на различных частотах.

Различают системы связи с быстрым и медленным ППРЧ [18]. Если время работы радиолинии на одной частоте ($T_{пер}$) соизмеримо или менее длительности информационного символа $T_{пер} \le T_и$ (рис. 7.29,а), то ППРЧ называют быстрым (рис. 7.29,б). Если $T_{пер} > T_и$, то такое ППРЧ называют медленным (рис. 7.29,в).

По порядку использования рабочих частот различают системы связи с последовательным ППРЧ, если в каждый момент времени передача ведется на одной частоте (рис. 7.29,б) и с параллельным ППРЧ, если передача ведется одновременно на нескольких частотах (рис. 7.29,в).

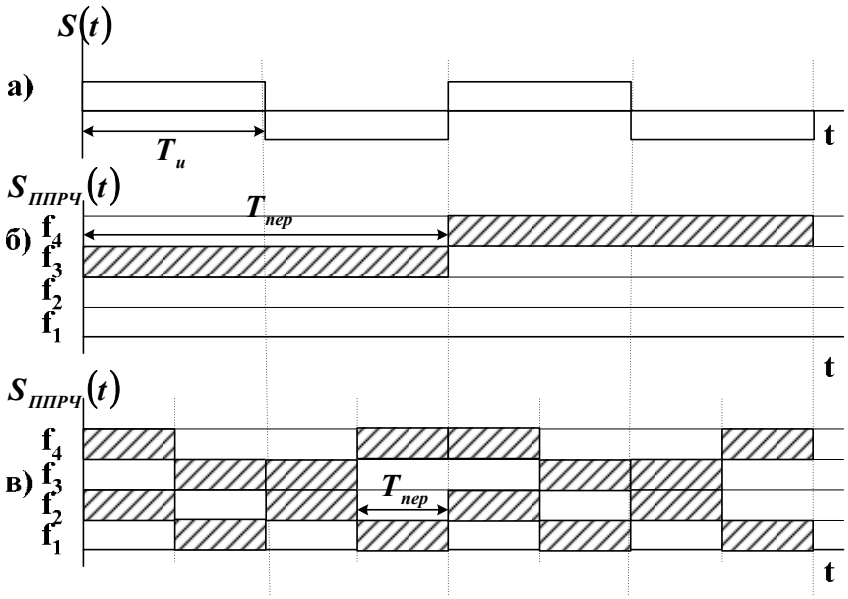


Рис. 7.29. Сигналы: исходный («а») последовательное медленное ППРЧ («б»); параллельное быстрое ППРЧ («в»)

Для сигналов с ППРЧ вводится понятие базы, характеризующей расширение спектра $B = \Delta f_{прч} / \Delta F_c$, где $\Delta f_{прч}$ – ширина используемого для передачи диапазона частот; ΔF_c – ширина спектра передаваемого сигнала.

В системах связи могут применяться составные сигналы, использующие и ФМ ПСП и ППРЧ. В этом случае база сигнала будет определяться выражением $B = M \cdot L$, где M – база ФМ ПСП; L – база сигнала ППРЧ.

7.5.4. Помехоустойчивость радиолиний с ППРЧ

Рассмотрим, как изменяется вероятность ошибки при использовании различных видов ППРЧ. Если применяется последовательное ППРЧ в полосе час-

тот $\Delta f_{прч}$ с шириной спектра сигнала ΔF_c, то при оптимальном приеме противник вынужден распределять энергию помехи равномерно во всей полосе $\Delta f_{прч}$. В этом случае спектральная плотность мощности помехи в полосе ΔF_c уменьшается в B раз, причем база $B = \dfrac{\Delta f_{прч}}{\Delta F_c}$. Вероятность ошибки определяется так же, как и для ФМ ПСС. Для ППРЧ с ЧМн сигналом $p^{к2}_{ППРЧ} = \dfrac{1}{2} \cdot \exp\left(- B \cdot \dfrac{h^2}{2} \right)$.

Если применяется параллельное ППРЧ, передача ведется одновременно на N частотах. Решение в приемном устройстве принимается по мажоритарному принципу, т.е. по большинству принятых решений на различных частотах.

Принимая допущение, что ошибки приема на различных частотах не зависят друг от друга, результирующая вероятность ошибки определяется по формуле Бернулли:

$$p_{ППРЧ} = \sum_{i=\mathrm{int}\left(N/2\right)}^{N} C_N^i \cdot p_{ош}^i \cdot \left(1 - p_{ош}\right)^{N-i}, \qquad (7.10)$$

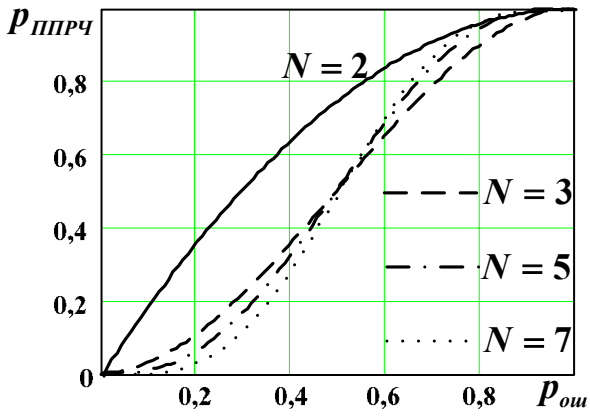


Рис. 7.30. Зависимости вероятности ошибок в системах с параллельным ППРЧ

где $i = \mathrm{int}\left(N/2\right)$ – ближайшее к $N/2$ целое число большее или равное $N/2$.

На рис. 7.30 показаны зависимости в соответствии с (7.10) при различном числе частот N. Анализ кривых показывает, что наибольший выигрыш наблюдается при нечетном числе частот, т.е. при $N = 3,5,7,\ldots$.

В заключении необходимо отметить, что при одном и том же значении базы выигрыш ППРЧ и ФМ ПСС одинаков, однако обработка сигналов с ППРЧ при ширине их спектра выше десятков МГц технически реализуется проще.

Контрольные вопросы

1. Какие замирания оказывают наибольшее влияние на помехоустойчи-

вость приема сигналов?

2. Какие способы борьбы с замираниями сигналов применяются в цифровых системах?

3. Какие методы применяются для борьбы с медленными замираниями, а какие для борьбы с быстрыми?

4. Какие существуют способы разнесения сигналов?

5. Чем характеризуется величина межсимвольной интерференции?

6. Как определяют помехоустойчивость приема сигналов в каналах с межсимвольной интерференцией?

7. Определить помехоустойчивость оптимального некогерентного приема ОФМн сигналов, если отношение сигнал/шум равно $\boldsymbol{h^2 = 2{,}57}$ и выработать предложения на обеспечению требуемой помехоустойчивости если $p^*_{ош} = 2 \cdot 10^{-5}$, при использовании сложных сигналов.

ГЛАВА 8. МНОГОКАНАЛЬНАЯ СВЯЗЬ И РАСПРЕДЕЛЕНИЕ ИНФОРМАЦИИ

8.1. Методы распределения ресурса общего канала

На практике существует потребность передачи больших объемов информации многих пользователей при ограниченных возможностях, когда уже както сформировались телефонные и телеграфные сети, определены линии и каналы связи, распределен ресурс рабочих радиоволн между странами.

В связи с этим остро стоит задача организации наиболее эффективного доступа нескольких пользователей к единому ресурсу (частотно-временному и энергетически-пространственному).

8.1.1. Классификация систем передачи информации, использующих единый ресурс

Любой сигнал занимает определенную полосу частот, существует некоторое время, обладает ограниченной энергией и распространяется в определенной области пространства. В соответствии с этим выделяют четыре вида ресурса канала: частотный, временной, энергетический и пространственный.

Проблема эффективного использования ресурса общего канала особенно обострилась из-за необходимости организации оперативного обмена данными и обеспечения связи с объектами в информационных системах различного назначения в условиях неравномерности и непредсказуемости запросов потребителей во времени. При решении проблемы распределения ресурса общего канала применяются методы мультиплексирования и множественного доступа (multiple access). Понятия «мультиплексирование» и «множественного доступа» сходны тем, что они предполагают распределение ресурса между пользователями. В то же время между ними есть и существенные различия. Так при мультиплексировании ресурс канала связи распределяется через общее оконечное оборудование, формирующие групповой сигнал $S_{\Sigma}(t)$. При множественном дос-

тупе, $S_\Sigma(t)$ образуется в результате сложения сигналов пользователей непосредственно в канале (рис. 8.1, где ИС – источник сообщения, ПРД - передатчик, ПРМ - приемник, ПС – получатель сообщения). Множественный доступ характерен для спутниковых каналов, радиоканалов, каналов мобильной связи [5, 6, 20, 39].

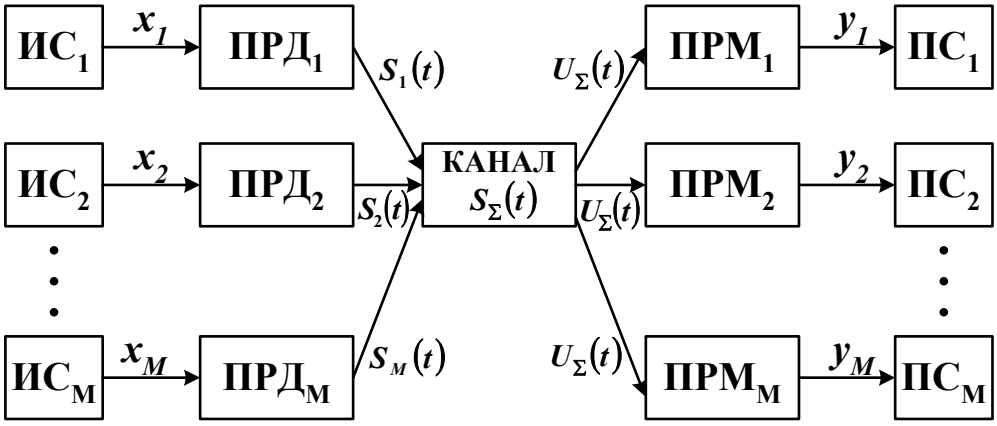


Рис. 8.1. Система передачи с множественным доступом

Принято считать, что мультиплексирование основано на общем аппаратурном обеспечении, в то время как множественный доступ (МД) использует определенные процедуры (протоколы), реализуемые с помощью программного обеспечения, хранящегося в памяти каждого терминала.

На рис. 8.2 представлены методы мультиплексирования.

Рис. 8.2. Методы мультиплексирования

В большинстве случаев для осуществления операции уплотнения канала источнику сообщений выделяется специальный сигнал, называемый каналь-

ным. Промодулированные сообщениями канальные сигналы объединяются, в результате чего образуется групповой сигнал $S_{гр}(t)$. Если операция объединения линейна, то $S_{гр}(t) = S_{\Sigma}(t)$ будет линейным групповым сигналом. Он, как правило, образуется линейным суммированием промодулированных канальных сигналов.

В системах так называемого комбинационного уплотнения групповой сигнал формируется посредством определенной логической (нелинейной) обработки, в результате которой каждый элемент сформированного сигнала отображает информацию (комбинацию символов) от всех ИС. Классическим примером такой системы является система двукратного частотного телеграфирования. Для передачи четырех комбинаций символов двух каналов используется четыре частоты: $f_1 \to 00, \quad f_2 \to 01, \quad f_3 \to 10, \quad f_4 \to 11$.

Устройство разделения линейного группового сигнала $S_{\Sigma}(t)$ представляет собой набор линейных избирательных цепей, каждая из которых выделяет только свой канальный сигнал и в идеальном случае совсем не реагирует на другие канальные сигналы. Для осуществления подобного идеального разделения необходимо и достаточно, чтобы промодулированные канальные сигналы составляли ансамбль линейно независимых сигналов. В качестве таких сигналов обычно используют ансамбли ортогональных сигналов.

В классе линейного уплотнения по виду отличительного признака канального сигнала различают временное разделение каналов (ВРК), частотное (ЧРК) и разделение каналов по форме сигналов, называемое кодовым разделением каналов (КРК). Вместо термина «разделение» применяют и термин «уплотнение». При ЧРК полоса частот общего канала Δf разделяется на несколько более узких полос Δf_i, каждая из которых образует канал ИС. При ВРК вся полоса Δf предоставляется поочередно через определенные интервалы времени различным источникам для передачи сообщений. При КРК нет деления общего канала между ИС ни по частоте, ни по времени. Канальные сигналы различных ИС, перекрываясь по времени и частоте, остаются ортогональными за счет раз-

личия формы, что и обеспечивает их разделение.

Возможны варианты комбинирования указанных методов. Так, в мобильной связи в качестве метода МД широко используются комбинации ЧРК и ВРК, ВРК и КРК. В первой комбинации каждый частотный канал предоставляется нескольким пользователям на определенные промежутки времени. При второй комбинации в полосе частот Δf формируют каналы с временным разделением, которые предоставляются нескольким пользователям на принципах КРК.

При организации многоканальной передачи информации, применяемые для уплотнения канальные сигналы могут быть заранее определенным образом распределены между источниками сообщений. Такое уплотнение называется уплотнением с закрепленными каналами. Соответствующая ему многоканальная система передачи также будет называться системой с закрепленными каналами. Возможна и такая организация многоканальной передачи информации, когда канальные сигналы не распределяются заранее между источниками, а выделяются каждому источнику по мере необходимости. Такое уплотнение называется уплотнением с незакрепленными каналами. Очевидно, для правильного разделения каналов в системах с незакрепленными каналами необходимо каким-либо образом передать на приемную сторону адресную информацию.

Основные понятия и определения, введенные для многоканальных систем, применимы и для систем МД. К настоящему времени изучено и предложено большое число разнообразных методов МД. Они различаются способом распределения коллективного ресурса канала (фиксированный или динамический), природой процессов принятия решения (централизованные или распределенные), а также степенью адаптации режима доступа к изменяющимся условиям.

Множественный доступ характерен для спутниковых каналов (в этом случае применяют термин «многостанционный доступ»), радиоканалов (пакетная радиосвязь), каналов мобильной связи, а также для многоточечных телефонных линий, локальных сетей.

Все существующие методы МД можно сгруппировать и выбрать в качестве основания классификации способ управления распределением ресурса об-

щего канала (рис.8.3).

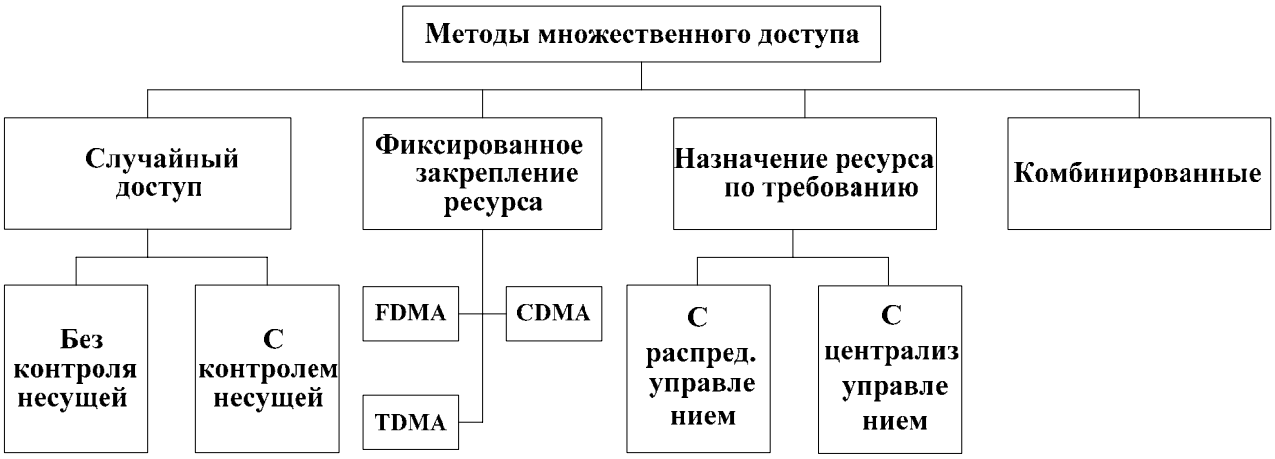


Рис. 8.3. Методы множественного доступа

Протоколы случайного доступа. При случайном МД весь ресурс канала связи представляется как один канал, доступ в который происходит случайно, в результате чего возможно столкновение пакетов передаваемой информации. Корреспондентам предлагается совершить определенную последовательность действий с целью разрешения конфликта. Каждый пользователь при необходимости может передавать данные в канал, не выполняя явного согласования с другими пользователями. Наличие обратной связи позволяет взаимодействующим корреспондентам контролировать прохождение передаваемой информации.

Возможны два варианта реализации стратегии случайного доступа: без контроля несущей и с контролем несущей.

Случайный доступ без контроля несущей состоит в том, что при необходимости передать данные, терминал пользователя сразу начинает передачу пакетов. Поскольку пакеты передаются без синхронизации между собой, возможно их наложение, что вызывает взаимные помехи. При возникновении такого конфликта, подтвержденного сигналом обратной связи, терминалы повторяют передачу искаженных пакетов. Во избежании повторения конфликтов промежутки времени до начала повторной передачи на каждом терминале выбираются случайно.

Случайный доступ с контролем несущей предполагает возможность кон-

тролировать наличие передачи информации другими корреспондентами. В случае отсутствия передачи данных незанятые временные промежутки имеются для передачи своей информации. В случае столкновения пользователи задерживают передачу пакетов на интервал времени Δt. В настоящее время существуют две разновидности протокола: настойчивый и ненастойчивый. Различие заключается в том, что в первом случае пользователи подвижных объектов, обнаруживая столкновения, начинают передачу сразу, а при втором через определенный интервал времени.

Протоколы фиксированного закрепления ресурса канала обеспечивают статическое распределение ресурса канала между пользователями. Наиболее типичными представителями протоколов данного типа являются многостанционный доступ с частотным разделением (FDMA), многостанционный доступ с временным разделением (TDMA), многостанционный доступ с кодовым разделением (CDMA).

Фиксированное закрепление ресурса канала не может обеспечить динамически изменяющиеся требования пользователей сети, т.е. имеет жесткое управление.

Методы назначения ресурса по требованию позволяют избавиться от недостатков, присущих вышеперечисленным методам, но предполагают подробную и четкую информацию о требованиях пользователей сети.

По природе процессов принятия решения методы назначения ресурса по требованию подразделяют на централизованные и распределенные.

Централизованные методы назначения ресурса по требованию, характеризуются наличием запросов на передачу со стороны терминалов источника сообщения. Принятие решения о предоставлении ресурса осуществляется центральной станцией.

Соответствующие протоколы отличаются наличием жестко закрепленных за каждым подвижным объектом каналов резервирования и наличием центральной станции управления. Протоколы характеризуются высоким значением коэффициента использования пропускной способности базовой станции, одна-

ко критичны к нарушениям функционирования системы управления.

По способу резервирования, определяющему действия центральной станции пользователей сети, существует два метода назначения ресурса по требованию с централизованным управлением.

Распределенные методы назначения ресурса по требованию отличаются тем, что все пользователи производят одни и те же операции, не прибегая к помощи центральной станции, и используют дополнительную служебную информацию, которой обмениваются друг с другом. Все алгоритмы с распределенным управлением требуют обмена управляющей информацией между пользователями. Протоколы характеризуются жестким закреплением каналов резервирования за подвижным объектом. При этом на каждом объекте имеется таблица закрепления запросных каналов, следовательно, любой подвижный объект в любой момент времени имеет информацию о состоянии всей сети.

Комбинированные методы представляют собой комбинации предыдущих методов распределения ресурса, и реализуют стратегии, в которых выбор метода является адаптивным для различных пользователей с целью получения характеристик используемого ресурса канала, близких к оптимальным. В качестве критерия оптимальности, как правило, принимается коэффициент использования пропускной способности канала. На основе протоколов данного типа осуществляется подстройка параметров под конкретную обстановку в сети.

Таким образом, каждый из рассмотренных способов распределения ресурса обладает достоинствами и недостатками. На практике целесообразно иметь всю совокупность методов и осуществлять адаптивный переход от одного метода к другому при определенных изменениях рабочих условий.

8.1.2. Постановка задачи объединения и разделения сигналов

Передача сообщений с малой вероятностью ошибок возможна в случае, когда пропускная способность канала связи C превышает производительность источника H' [6, 20, 39]:

$$C \geq H', \tag{8.1}$$

где H' производительность источника, определяемая выражением:

$H'(x)=V_И \cdot H(x)$; $V_И$ – скорость передачи символов; $H(x)$ – энтропия источника, т.е. среднее количество информации, приходящееся на один символ.

Производительность источников сообщений, как правило, значительно меньше пропускной способности существующих каналов связи. Это позволило повысить эффективность использования канала путем передачи по нему сообщений нескольких источников. Такое использование канала называют уплотнением.

Очевидно, при мультиплексировании (объединении и разделении) канала должно выполняться условие:

$$H'_M = \sum_{i=1}^{M} H'_i \le C, \qquad (8.2)$$

где M – число независимых источников.

Системы связи, в которых используют мультиплексирование, часто называют многоканальными (рис. 8.4.).

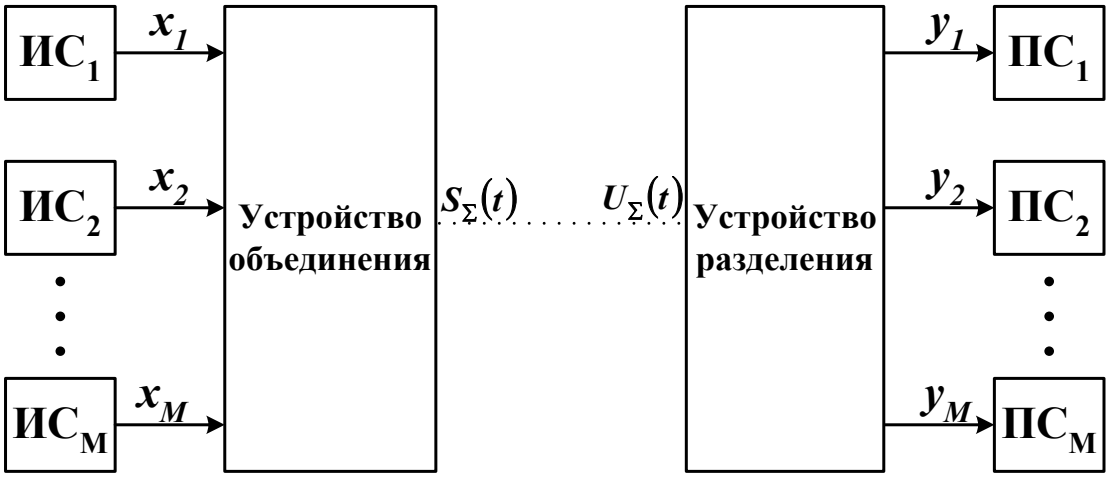


Рис. 8.4. Многоканальная система передачи сообщений

Пусть M источников посылают сообщения $x_1, x_2, ..., x_M$. Задача объединения сигналов подразумевает преобразование совокупности $\{x_i\}$ в групповой сигнал $S_\Sigma(t)$. Групповой сигнал проходит по уплотняемому каналу и на устройство разделения поступает в смеси с аддитивной помехой $n(t)$:

$$U_\Sigma(t) = S_\Sigma(t) + n(t). \qquad (8.3)$$

Задача устройства разделения состоит в разделении группового сигнала

$U_\Sigma(t)$ и преобразовании его в совокупность сообщений $y_1, y_2, ..., y_M$.

Для этого необходимо осуществить выбор системы функций $\{S_i\}$ таким образом, чтобы обеспечить восстановление x_i из y_i по принятому групповому сигналу $U_\Sigma(t)$.

Чтобы исключить возможные влияния индивидуальных сигналов друг на друга, обычно они выбираются взаимно ортогональными, т.е. для любой пары сигналов $S_i, S_j, \mathrm{i} \neq \mathrm{j}$ должно выполняться одно из условий [5, 6, 20]

$$\int_{-T/2}^{T/2} S_i(\omega, \Theta, t) \cdot S_j(\omega, \Theta, t) dt = 0,$$

где T – длительность элемента сигнала; ω – частота сигнала; Θ - пространственный угол наблюдения сигнала.

Пути решения задачи объединения и разделения сигналов

Устройство разделения должно определить, какой символ сообщения передавался каким источником. Для того, чтобы групповой сигнал мог переносить информацию о сообщениях всех M источников, необходимо, чтобы число различимых реализаций группового сигнала m_Σ на каждом отрезке времени было не меньше числа всех возможных состояний совокупности M источников на этом отрезке. Для дискретных источников с одинаковым объемом алфавита $m = 2$, необходимое число реализаций группового сигнала составляет $m_\Sigma \geq m^M = 2^M$. Системы уплотнения, основанные на данном принципе, относятся к системам комбинационного типа.

Комбинационные системы уплотнения применяются для многоканальной системы передачи дискретных сообщений. В этих системах ни групповой сигнал ни его отдельные параметры не могут считаться суммой индивидуальных сигналов. Групповой сигнал определяется совокупностью сочетаний символов в индивидуальных каналах. Структурная схема системы разделения группового сигнала комбинационного типа показана на рис. 8.5.

В согласованных фильтрах (СФ) данной схемы осуществляется сравнение принятой кодовой комбинации с эталонными образцами, на которые настроены фильтры. На вход устройства разделения поступит групповой сигнал только с того согласованного фильтра, эталонный сигнал которого совпал с принятым сигналом.

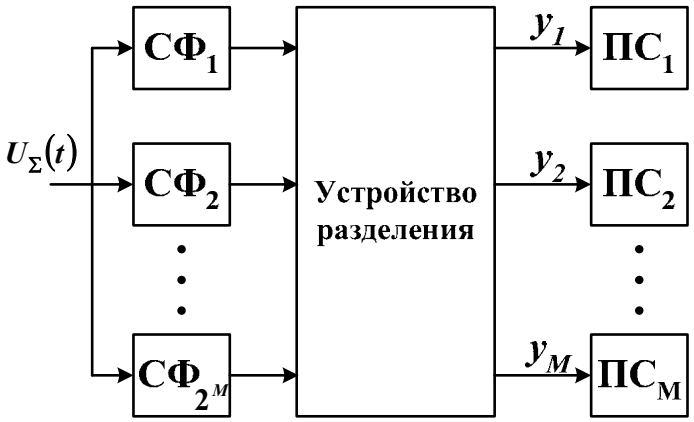


Рис. 8.5. Структурная схема системы разделения группового сигнала комбинационного типа

Очевидно, с увеличением кратности уплотнения сложность схемы быстро возрастает. Например, при $M = 10$, число ветвей равно $m = 2^{10} = 1024$. Поэтому подобные системы обычно используют в системах уплотнения при малом числе каналов $M \le 4...5$.

Наиболее широко распространены системы, в которых групповой сигнал образуется путем простого сложения индивидуальных сигналов, каждый из которых переносит информацию только об одном из сообщений. Такие системы уплотнения называют раздельными; их применяют для передачи как дискретных, так и непрерывных сообщений. Они нашли преимущественное применение на практике ввиду простоты образования каналообразующей аппаратуры. Действительно, решающая схема может быть построена как объединение M индивидуальных решающих схем для отдельных сообщений (рис. 8.6).

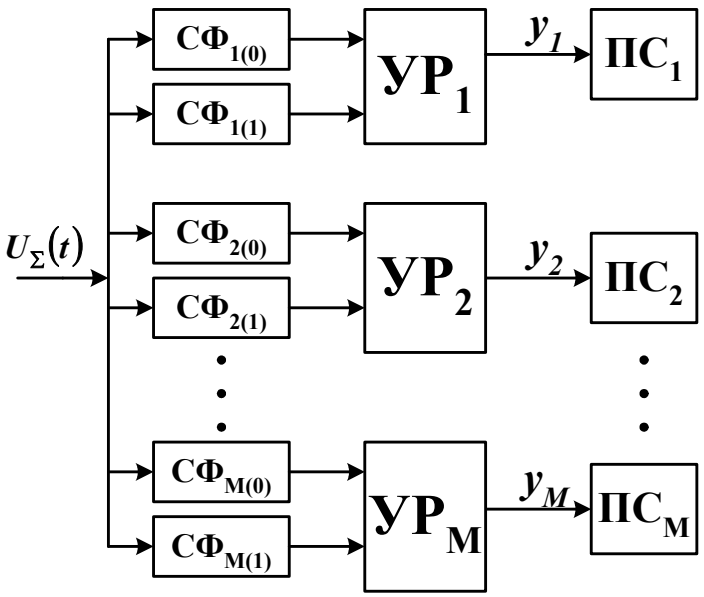


Рис. 8.6. Структурная схема системы разделения группового сигнала раздельного типа

В этом случае число груп-

повых реализаций m равно сумме реализаций каждого индивидуального сигнала. Для двоичных сигналов число ветвей в общей решающей схеме $m = 2 \cdot M$,. Поэтому даже при кратностях уплотнения, измеряемых сотнями и тысячами, решающая схема остается технически осуществимой.

8.1.3. Энергетическая и спектральная цена уплотнения

Для сопоставления различных систем уплотнения по их помехоустойчивости и эффективности использования полосы частот, вводятся понятия энергетической и спектральной цены уплотнения.

Любая система уплотнения требует увеличения мощности группового сигнала, чтобы обеспечить заданную верность приема индивидуальных сообщений. Энергетической ценой уплотнения называют отношение [6, 39]

$$\eta(M) = \frac{P_\Sigma}{P_1},$$

где P_Σ – мощность группового сигнала; P_1 – мощность индивидуального сигнала. Энергетическая цена уплотнения показывает, во сколько раз требуется увеличить среднюю мощность группового сигнала, чтобы при данной системе уплотнения и заданной верности передавать M сообщений вместо одного.

В раздельной системе уплотнения с ортогональными индивидуальными сигналами влияние индивидуальных сигналов друг на друга отсутствует, поэтому $\eta(M) = \frac{P_\Sigma}{P_1} = M$. При других методах уплотнения энергетическая цена может оказаться большей или меньшей числа передаваемых сообщений M .

Для количественной оценки эффективности использования полосы частот уплотненного канала вводится понятие спектральной цены уплотнения [6, 39]:

$$\beta(M) = \frac{\Delta f_\Sigma}{\Delta f_1},$$

где Δf_Σ – ширина спектра группового сигнала; Δf_1 – ширина спектра индивидуального сигнала. В большинстве систем уплотнения спектр группового сигнала занимает более широкую полосу частот, чем при передаче одного сообщения. Сравнение полос должно производиться при одинаковых условиях, например одинаковом основании кода, методе модуляции, длительности элемента груп-

пового сигнала.

8.2. Частотное разделение каналов

Практика построения современных телекоммуникационных систем и сетей показывает, что наиболее дорогостоящими звеньями трактов передачи являются линии связи (кабельные, волоконно-оптические, радиосвязи, радиорелейные и др.). Поскольку экономически нецелесообразно использовать дорогостоящую линию связи для передачи информации единственной пары абонентов (от источника к получателю сообщений и обратно при дуплексной связи), то возникает задача построения многоканальных систем передачи, обеспечивающих передачу большого числа сообщений различных источников информации по общей линии связи. Многоканальные системы так же, как и одноканальные, могут быть аналоговыми и цифровыми.

8.2.1. Принцип частотного объединения и разделения каналов

При частотном разделении каналов для передачи данных различных источников сообщений используются определенные поддиапазоны частот. Функциональная схема простейшей системы многоканальной связи с частотным разделением каналов (ЧРК) представлена на рис. 8.7. [5, 6, 21, 39]

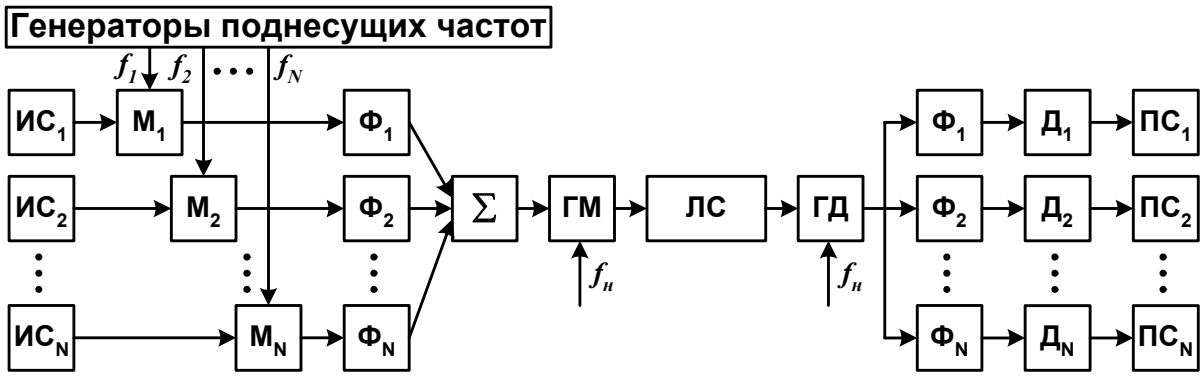


Рис. 8.7. Функциональная схема СЭС с ЧРК

Основные этапы образования спектра $A_\Sigma(f)$ группового сигнала показаны на рис.8.8. Пусть в СЭС осуществляется одновременная работа N корреспондентов. В соответствии с передаваемыми сообщениями первичные сигналы от источников сообщений, имеющие энергетические спектры $A_1(F), A_2(F) \ldots A_N(F)$,

модулируют поднесущие частоты f_k каждого канала.

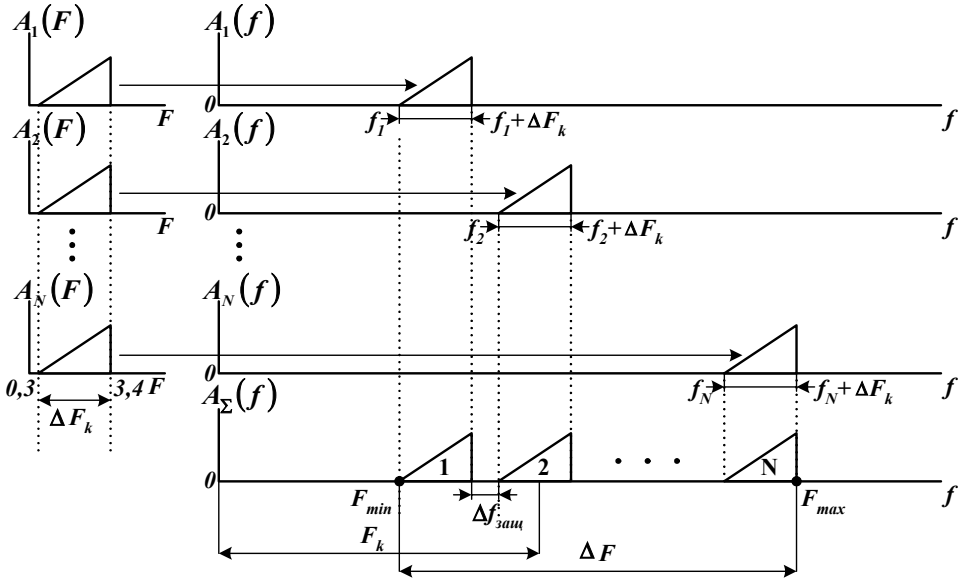


Рис. 8.8. Формирование группового спектра при ЧРК

Наиболее распространенный при ЧРК вариант - однополосная амплитудная модуляция. Полученные на выходе полосовых фильтров $\Phi_1, \Phi_2, ..., \Phi_N$ канальные сигналы суммируются, и их совокупность $A_\Sigma(f)$ поступает на групповой модулятор. Здесь спектр $A_\Sigma(f)$ с помощью колебания несущей частоты $f_н$ переносится в область частот, отведенную для передачи данной группы каналов, т. е. групповой сигнал преобразуется в так называемый линейный сигнал, передаваемый по линии связи – кабелю, радио, радиорелейной, спутниковой линии связи. При этом может использоваться любой вид модуляции. На приемном конце осуществляется вся совокупность обратных преобразований. Групповым демодулятором линейный сигнал преобразуется в групповой, из которого с помощью фильтров выделяются канальные сигналы. С помощью детекторов канальные сигналы преобразуются в ПЭС поступающие к получателям.

8.2.2. Групповой сигнал, его структура и характеристики

В многоканальных радиосистемах передачи с ЧРК обычно используют аппаратуру объединения и разделения каналов, применяемую в проводных системах. Это обеспечивает простоту сопряжения тех и других систем и отражает общую тенденцию к унификации оборудования на сетях связи. Для унифика-

ции аналоговых многоканальных систем за основной или стандартный канал принимают канал тональной частоты (канал ТЧ), обеспечивающий передачу сообщений с полосой частот 300...3400 Гц, соответствующей основному спектру телефонного сигнала. Многоканальные аналоговые системы формируются путем объединения каналов ТЧ в группы, обычно кратные 12 каналам. Для снижения переходных помех вводятся защитные частотные интервалы $\Delta f_{защ} = 0{,}9 кГц$. Таким образом на один канал ТЧ выделяется полоса частот 4кГц.

В частотной области реальный групповой сигнал (рис. 8.8) характеризуется следующими параметрами:

шириной полосы частот группового сигнала ΔF;

нижней $F_{\min}$ и верхней $F_{\max}$ граничными частотами;

числом каналов N;

эффективно передаваемыми полосами частот каналов ΔF_k;

значениями средних частот каналов на оси частот F_k;

значениями поднесущих частот f_k;

защитными полосами частот между каналами $\Delta f_{защ}$.

Построение многоканальных систем подчиняется иерархическому принципу. В табл. 8.1 представлены основные характеристики иерархии частотного объединения каналов.

Таблица 8.1

Основные характеристики иерархии частотного объединения каналов

Группа каналов	Полоса частот, кГц	Число каналов и групп
Канал ТЧ	0,3…3,4	–
Первичная группа (ПГ)	60…108	12 каналов ТЧ
Вторичная группа (ВГ)	312…552	5 ПГ = 60 каналов ТЧ
Третичная группа (ТГ)	812…2044	5 ВГ = 300 каналов ТЧ
Четверичная группа (ЧГ)	8516…12388	3 ТГ = 900 каналов ТЧ

Многоканальный (групповой) сигнал имеет сложную структуру, которая зависит от общего количества каналов, числа работающих в данный момент ка-

налов, затуханий абонентских линий, индивидуальных особенностей абонентов. Кроме того, часть каналов ТЧ используется не для передачи речевых сигналов, а для вторичного уплотнения (тональный телеграф), передачи бинарной информации и т.п.; периодически по каналам посылаются сигналы вызова. Поэтому величины средней и пиковой мощности группового сигнала и его пик–фактора зависят от числа каналов и непостоянны во времени, что во многом определяет качество функционирования группового тракта. Так, например, если пиковые напряжения группового сигнала выходят за пределы линейных участков амплитудных характеристик групповых усилителей, модуляционной характеристики передатчика, демодуляционной характеристики приемника, то в этих элементах тракта возникает режим перегрузки, вызывающий искажения сигналов и переходные помехи.

При проектировании и разработке многоканальных систем передачи информации возникает необходимость количественной оценки параметров групповых сообщений на различных ступенях преобразования, в частности, сигналов на входе линейного тракта. Эти параметры определяются соответствующими частотными, информационными и энергетическими характеристиками. Первые две группы характеристик и связанные с ними параметры могут находиться в системах с ЧРК на основе принципа «пропорционального роста». Так, например, сопоставляя сообщения, получающиеся в результате объединения 12 и 60 каналов ТЧ, можно утверждать, что сообщение вторичной группы по сравнению с сообщением первичной группы занимает в 5 раз более широкую полосу, и соответственно его максимальная информационная нагрузка в 5 раз выше, чем у сообщения первичной группы. Это является следствием того, что спектры сообщений в соседних каналах не перекрываются, а источники сообщений считаются однородными по своим параметрам.

Упомянутый принцип «пропорционального роста» нельзя распространить на энергетические характеристики, такие как мгновенная мощность группового сообщения, его пик - фактор, динамический диапазон и др. Это связано со следующими особенностями многоканальных систем с ЧРК:

мгновенные значения групповых сообщений являются продуктом «взаимодействия» (сложения) мгновенных значений сообщений всех объединяемых каналов;

расчет общей мощности группового сообщения обычно производится в предположении, что не менее 95 % каналов используются только для телефонной связи. Это означает, что в расчетах следует учитывать случайный характер канальных сигналов, изменяющиеся в широких пределах;

мгновенная мощность группового сообщения определяется не общим числом N объединяемых каналов, а числом так называемых «активных» каналов. Если один из абонентов молчит (например, слушает другого абонента) или имеет место пауза между словами или фразами, то соответствующий канал в данный момент времени к числу активных не относится; канал считается активным лишь в те интервалы времени, когда по нему передается сообщение.

Из рассмотренных особенностей следует, что при оценке энергетических показателей групповых сообщений следует руководствоваться лишь усредненными характеристиками, найденными с учетом соответствующих статистических закономерностей.

8.3. Временное разделение каналов

В настоящее время передача информации в радиорелейных линиях связи осуществляется как методом частотного разделения каналов (ЧРК), так и методом временного разделения каналов (ВРК). Радиорелейные линии с ЧРК-ЧМ обладают сравнительно высокими технико-экономическими показателями, однако они имеют существенный недостаток: трудность осуществления передачи части каналов для группы абонентов, находящихся на промежуточных станциях линии. Каждое ответвление связано с разуплотнением каналов на промежуточных станциях, выделением части каналов для группы абонентов и повторим уплотнением каналов для передачи остальных по линии связи.

Широкое применение нашли радиорелейные линии с ВРК. Их основное достоинство состоит в простоте выделения групп каналов, что весьма важно

при создании подвижных радиорелейных станций.

Временное разделение основано на возможности передачи вместо непрерывных сигналов последовательных импульсов (отсчетов). Поскольку при импульсной передаче период следования импульсов обычно намного больше их длительности (импульсы имеют большую скважность), между импульсами одного сигнала остается промежуток, на котором можно разместить импульсы от других сигналов. В настоящее время уже реализованы многоканальные системы с временным разделением 12, 15, 30, 120, 480 речевых сигналов.

Радиорелейные линии с ВРК предполагают использование как аналоговых: амплитудно - импульсная модуляция (АИМ), широтно - импульсная модуляция (ШИМ), фазо - импульсная модуляция (ФИМ), так и цифровых: импульсно кодовая модуляция (ИКМ), дельта – модуляция (ДМ) методов импульсной модуляции.

8.3.1. Принцип временного разделения каналов

Многоканальные системы с ВРК широко используются для передачи аналоговой и дискретной информации.

Принцип временного объединения каналов удобно пояснить с помощью синхронно вращающихся распределителей на передающей и приемной стороне (рис. 8.9).

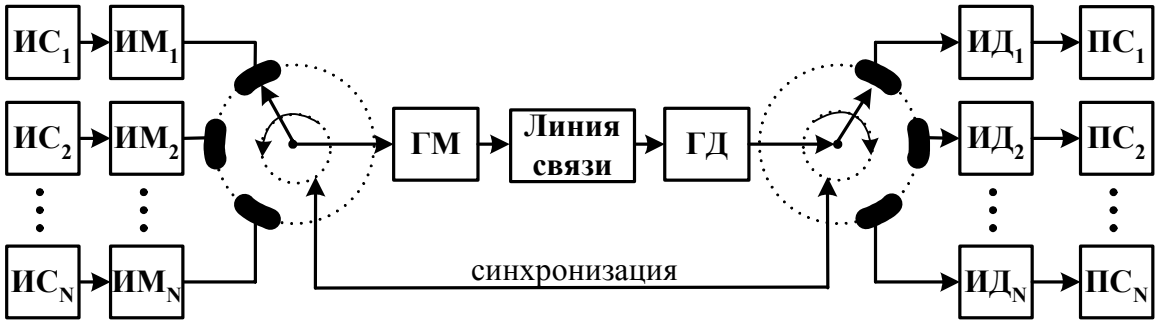


Рис. 8.9. Упрощенная блок-схема СЭС с ВРК

Основные этапы образования группового сигнала $S_{\Sigma}(t)$ показаны на рис.8.10.

Информация от N источников аналоговых сигналов поступает на входы соответствующих индивидуальных импульсных модуляторов АИМ (ШИМ, ФИМ). Формируемые отсчеты сигналов $S_1(t)$ на выходе первого импульсного модулятора ($ИМ_1$) (рис. 8.10,в), $S_2(t)$ на выходе второго импульсного модулятора ($ИМ_2$) (рис. 8.10,г) берутся через одинаковый интервал $\Delta t = \frac{1}{2F_{\max}}$, но с таким сдвигом Δ во времени, чтобы они не перекрывались.

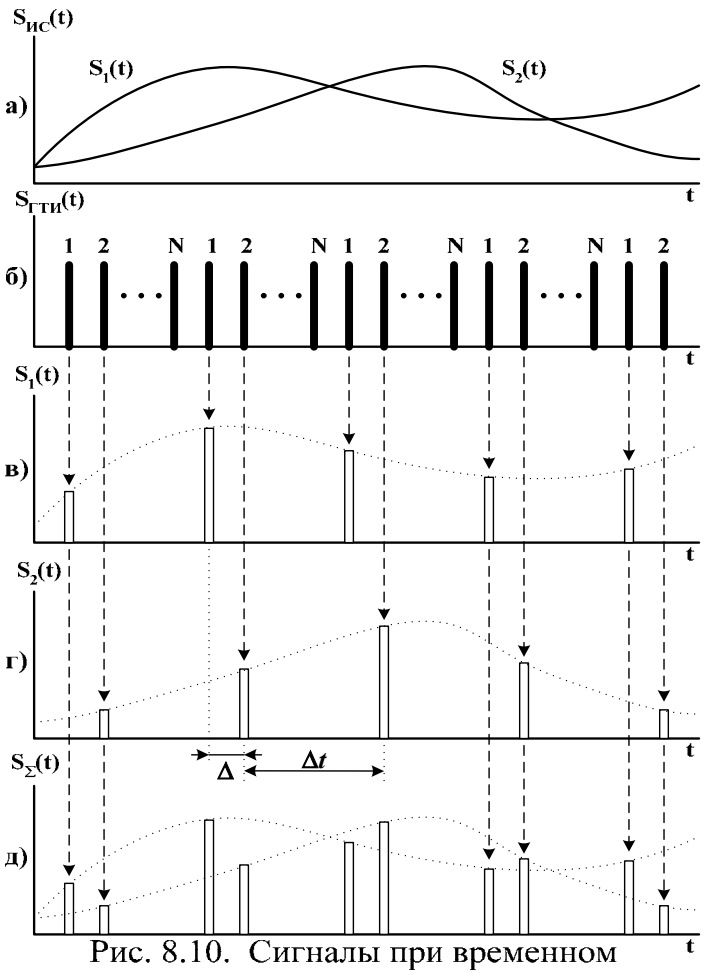


Рис. 8.10. Сигналы при временном разделении каналов

Затем передающий распределитель считывает импульсы от всех источников, формируя сигнал $S_\Sigma(t)$ (рис. 8.10,д), спектр которого с помощью группового модулятора (ГМ) переносится в область частот, отведенных для данной линии связи. Групповой сигнал $S_\Sigma(t)$, передаваемый по линии связи, несет информацию как от первого, так и о второго источника одновременно. На приемной стороне с выхода группового демодулятора (ГД) импульсы группового сигнала $S_\Sigma(t)$ поступают на вращающиеся контакты приемного распределителя для формирования канальных последовательностей $S_1(t)$, $S_2(t)$ и т.д. из которых на выходе импульсных детекторов формируются непрерывные сигналы поступающие к получателям сообщений [5, 6, 20, 21, 39].

Следует подчеркнуть, что рис. 8.9 служит лишь для иллюстрации идеи временного уплотнения и не отражает современных технических методов коммутации. В действительности аппаратура временного уплотнения обходится без механических распределителей, которые заменены электронными распре-

делителями, выполняющими те же функции (рис. 8.11).

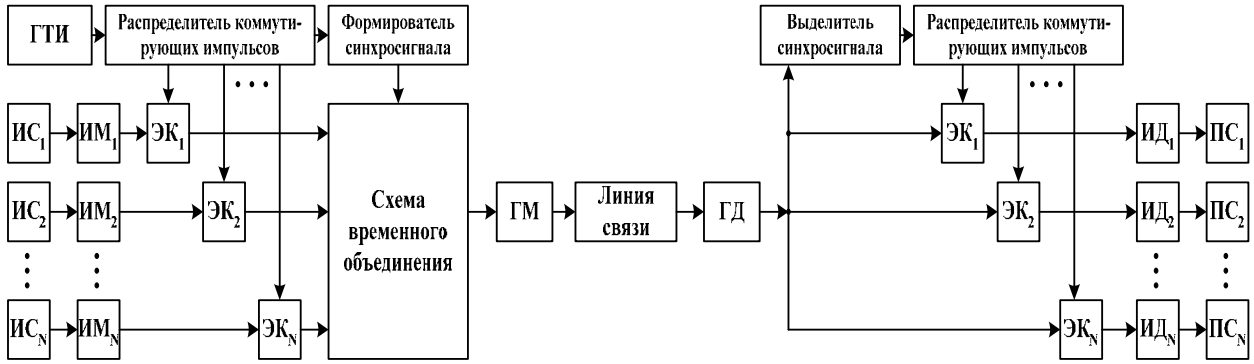


Рис.8.11. Схема многоканальной связи с ВРК.

Выходы всех импульсных модуляторов подключены к «своим» электронным ключам, работой которых управляет распределитель коммутирующих импульсов. В свою очередь, распределитель запускается от генератора тактовых импульсов.

Временное разделение сигналов осуществляется устройством, упрощенная структурная схема которого представлена на рис. 8.11. Принятый групповой радиосигнал в групповом демодуляторе преобразуется в групповую импульсную видеопоследовательность и поступает одновременно на входы выделителя синхросигнала и канальные электронные коммутаторы.

Процесс временного разделения производится в два этапа. На первом – этапе вхождения системы в синхронизм происходят поиск, обнаружение и выделение сигналов синхронизации, после чего запускается распределитель канальных коммутирующих импульсов. Распределитель формирует на своих выходах импульсы требуемой длительности и такой очередности, при которой в каждый канальный интервал открывается лишь один электронный коммутатор соответствующего канала.

На втором этапе производится демодуляция каждого канального импульса, после чего сигналы принимаемых каналов подаются к получателям аналоговой информации.

При временном разделении каналов важнейшую роль играет система синхронизации, алгоритм работы которой каждый раз выбирается индивидуально для принятого способа импульсной модуляции, способа временного объ-

единения каналов, структуры сигналов синхронизации и т.д.

8.3.2. Характеристики группового сигнала систем с ВРК

Коммутирующие импульсы на выходах распределителя появляются в определенной последовательности (рис. 8.12): каждому каналу отведен временной интервал T_k, в течение которого передается информация только данного канала. Следующий канал будет подключен лишь после того, как отключится предыдущий, что достигается с помощью электронных ключей, управляемых коммутирующими импульсами [5, 20, 21].

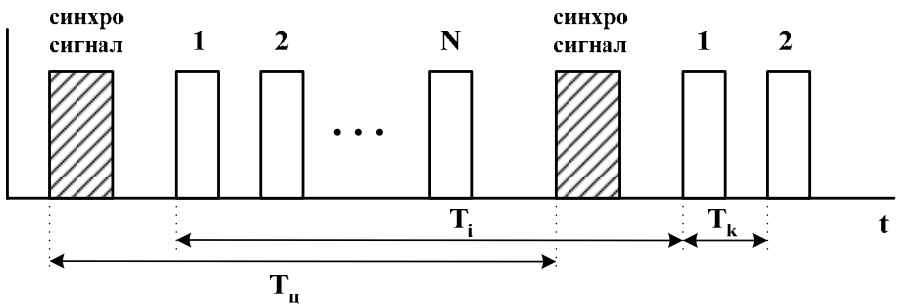


Рис. 8.12. Параметры группового сигнала

Каждый импульсный модулятор будет подключаться через электронный ключ к схеме временного объединения только в канальные интервалы времени длительностью $T_k = \frac{T_i}{N}$, выделенные в групповом цикле для передачи информации по данному каналу. Период повторения этих интервалов для T_i одного и того же канала называют периодом дискретизации по времени или тактовым интервалом. Очевидно, коммутация каналов должна происходить синхронно и синфазно на передающей и приемной сторонах линии связи.

В настоящее время применяются два способа коммутации каналов в схеме временного объединения. Первый способ заключается в том, что каждый канал подключается периодически и переход от одного канала к другому происходит в строго заданной очередности. Системы с таким способом опроса называют синхронными. При втором способе опрос производится непериодически и не в заранее заданной очередности, а произвольным образом. В этом случае система является асинхронной.

В рассматриваемой синхронной системе для безошибочного опознавания и разделения информационных (канальных) импульсов на приемной стороне в

групповой многоканальный сигнал вводятся специальные импульсы, называемые импульсами синхронизации. Синхросигнал формируется специальным устройством и вводится в групповой видеосигнал на строго определенную временную позицию (рис. 8.12), благодаря чему на приемной стороне каждый информационный импульс попадает на вход только своего тракта обработки. Обычно синхросигнал – отличается от информационных видеоимпульсов каким-либо параметром (амплитудой, длительностью, фазой и т.д.), что необходимо для его надежного обнаружения и выделения.

Группу видеоимпульсов (информационных и синхронизирующих), полученных на выходе схемы временного объединения в результате однократного опроса всех источников, сигналов, называют циклом, а соответствующий этому временной интервал – длительностью цикла $T_Ц$ (рис. 8.12). Заштрихованный на рис. 8.12 импульс называют сигналом цикловой (групповой) синхронизации. Поскольку видеоимпульсы различных каналов следуют друг за другом в строгой очередности, в их временном положении заключена информация о номере канала. В дальнейшем групповой видеосигнал подается на вход группового модулятора, где преобразуется в линейный радиосигнал, пригодный для передачи по линии связи.

При временном объединении предполагается, что каждый элемент индивидуального сигнала локализован во времени; вне интервала передачи данного сигнала он должен быть равен нулю. Но сигналы конечной длительности имеют бесконечно протяженный спектр. Реальные каналы связи ограничивают спектр последовательности импульсов, что приводит к растяжению импульсов во времени и создает возможность попадания их в соседние временные интервалы, предназначенные для других каналов (рис. 8.13). В результате возникают переходные помехи.

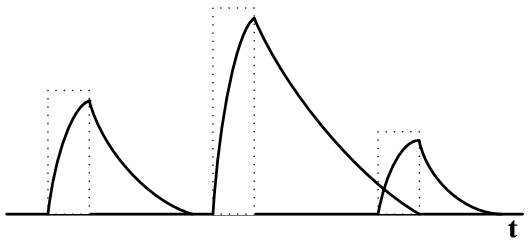


Рис. 8.13. Взаимное перекрытие импульсов

Для снижения переходных помех требуется либо расширять полосу пропускания группового тракта, либо уменьшать N, что приводит к неполному использова-

нию пропускной способности канала. Применяют также защитные интервалы времени между индивидуальными сигналами (рис. 8.14). Это позволяет снизить влияние переходных помех до допустимого уровня, но соответственно увеличивает спектральную цену уплотнения.

Необходимо отметить, что причиной переходных помех может быть и многолучевое распространение, в результате которого сигнал одного источника накладывается на сигнал последующего источника. С этим приходится считаться главным образом в коротковолновых радиоканалах. Для защиты от таких переходных помех целесообразно также применять защитные интервалы, длительность которых должна быть не меньше максимального времени запаздывания между лучами.

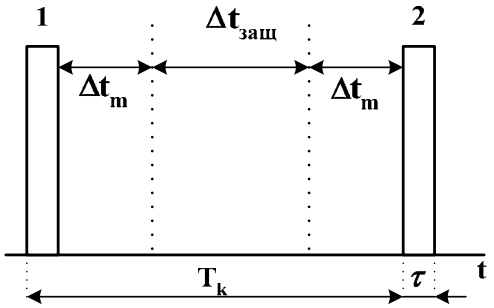


Рис. 8.14. Защитный интервал

8.4. Разделение сигналов по форме

Для разделения сигналов могут использоваться не только частота (ЧРК) и время (ВРК), но и форма сигналов. Разделение каналов по форме пока не нашло такого широкого использования, как частотное и временное. Его настоящее применение и перспективы в наибольшей степени связаны с множественным доступом в мобильных и спутниковых системах. В мобильной связи кодовое разделение рассматривается как один из основных видов обеспечения множественного доступа в плане реализации концепции развития систем мобильной связи IMT-2000.

Технология разделения каналов по форме предполагает возможность одновременной работы группы разнообразных радиосредств (мобильные терминалы, отдельные радиостанции, земные станции спутниковой связи и т. д.) в общей полосе частот ΔF. Сигналы радиосредств $S_i(t)$ образуют суммарный (групповой) сигнал $S_\Sigma(t)=\sum_{i=1}^{N} S_i(t)$, который поступает на приемные устройства пользователей. Взаимная ортогональность сигналов $\{S_i(t)\}$ обеспечивает корре-

ляционному приемнику выделение необходимого сигнала $S_i(t)$ из $S_\Sigma(t)$.

Асинхронно-адресные системы связи

В ряде случаев осуществить точную синхронизацию затруднительно. С этим приходится сталкиваться, например, при организации оперативной связи между подвижными объектами (автомобилями, самолетами) или при организации оперативной связи с использованием искусственных спутников Земли в качестве ретрансляторов. В этих случаях могут быть использованы системы асинхронной многоканальной связи, когда сигналы всех абонентов передаются в общей полосе частот, а каналы не синхронизированы между собой во времени. В системах со свободным доступом каждому каналу (абоненту) присваивается определенная форма сигнала, которая и является отличительным признаком, "адресом" данного абонента, отсюда и название асинхронно адресные системы связи (ААСС).

Адрес абонента может кодироваться в виде псевдослучайных (шумоподобных) сигналов или в виде последовательности нескольких радиоимпульсов с одинаковым или различным частотным заполнением. Если радоимпульсы имеют различное частотное заполнение, то говорят, что адрес кодируется в виде частотно-временной матрицы (ЧВМ). Адреса различаются как интервалами времени между радиоимпульсами, так и частотами их заполнения.

Рассмотрим принцип работы ААСС на основе обобщенной структурной схемы (рис. 8.15).

Передаваемые сообщения, полученные от источников $ИС_1,...,ИС_N$, подвергаются импульсной модуляции. В одних системах используется ФИМ, в других - некоторые разновидности дельта-модуляции. Затем каждый импульс, полученный в результате первичной импульсной модуляции, преобразуется в адресную последовательность из n импульсов, разделенных паузами [6, 21].

Формирование адресных последовательностей осуществляется с помощью линии задержки (ЛЗ), имеющую l отводов, как показано на рис. 8.15.

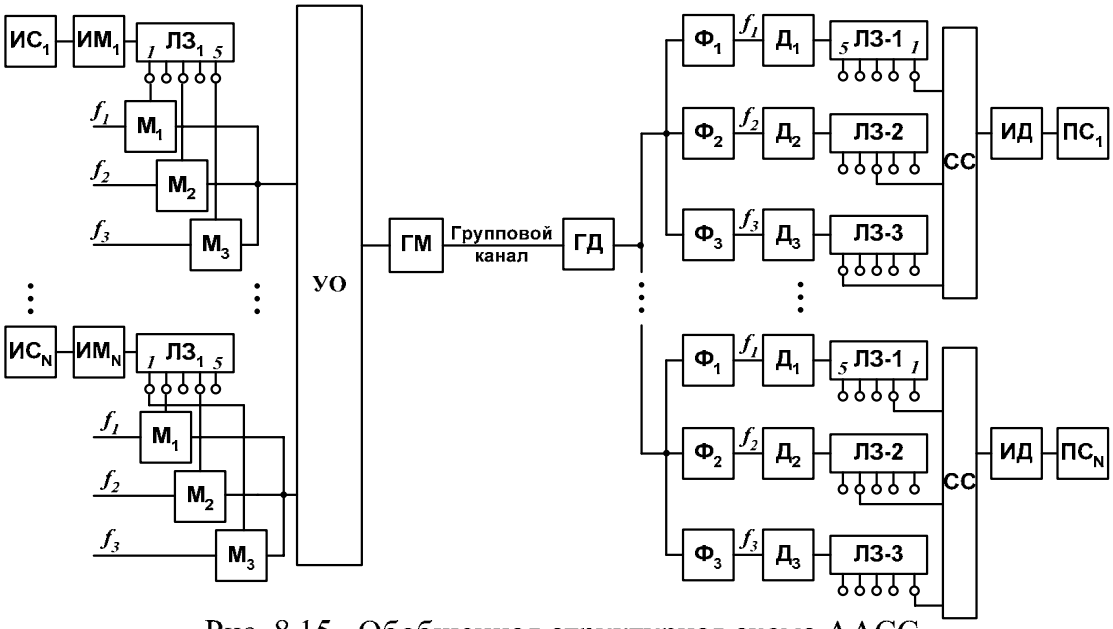


Рис. 8.15. Обобщенная структурная схема ААСС

Для формирования адреса используется только n отводов из l, причем для другого адреса применяется другое сочетание n отводов. Эти $\boldsymbol{n}$ импульсов различаются частотой своего заполнения (всего таких частот в системе уплотнения m) и могут занимать l различных положений во времени. Для примера, на рис. 8.16 представлен вариант построения таких адресных последовательностей для системы с $n = m = 3$ и $l = 5$.

Таким образом, импульс, полученный в результате первичной импульсной модуляции сообщением, разделяется в линии задержки на n импульсов. Каждый из этих n импульсов может занимать одно из l положений во времени и передается на своей частоте.

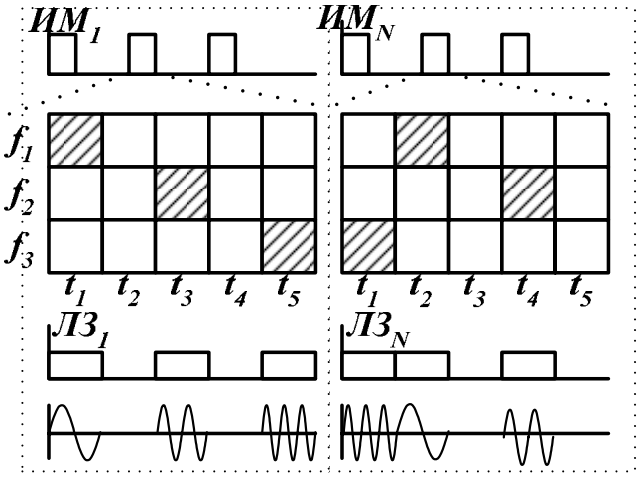


Рис. 8.16. Адресные последовательности для ААСС

Варьируя положения импульсов во времени относительно первого импульса, а также частоты заполнения импульсов, можно получить большое число адресных кодовых комбинаций (большую кратность уплотнения).

Каждый индивидуальный приемник представляет собой нелинейное устройство, содержащее линии задержки и схему совпадения (СС),

и реагирует только на определенную последовательность радиоимпульсов (рис. 8.17). Приемник имеет n полосовых фильтров $\Phi_1,\Phi_2,...,\Phi_n$, настроенных на соответствующие частоты. Выходные импульсы каждого фильтра детектируются и поступают на линии задержки, спроектированные в соответствии с присвоенным данному приемнику адресом так, чтобы все n импульсов на выходах совпали по времени. На нелинейной схеме совпадений (СС) появляется импульс только при том условии, что задержанные входные импульсы во всех ветвях совпали. Если же с выходов линий задержек на вход схемы совпадения хотя бы один из импульсов поступает неодновременно с остальными, то сигнал на выходе СС не появится. Благодаря этому приемник реагирует лишь на присвоенную ему адресную кодовую комбинацию.

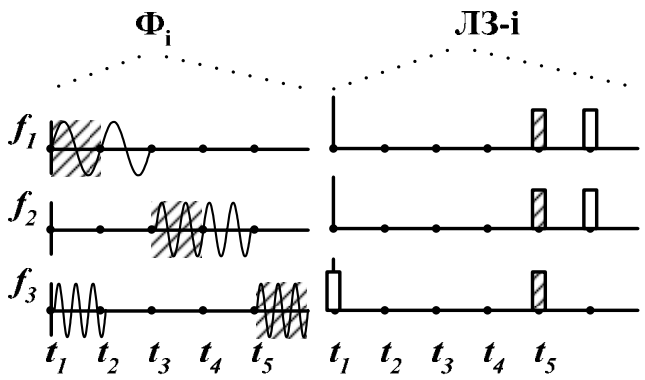


Рис. 8.17. Процесс разделения сообщений

Описанный процесс разделения сообщений (т.е. выделения только присвоенной приемнику адресной кодовой комбинации) поясняет рис. 8.17. На вход приемника поступает групповой сигнал, содержащий, в частности, два сообщения (заштрихованные и незаштрихованные радиоимпульсы). Приемное устройство реагирует лишь на присвоенную ему адресную частотно-временную комбинацию, отображенную заштрихованными импульсами, т.е. выделяет сообщение. Импульсы с выхода схемы совпадения преобразуются в принятое сообщение в импульсном демодуляторе (ИД) в соответствии с примененной импульсной модуляцией.

Для того чтобы установить связь с определенным абонентом, достаточно выбрать соответствующие n положений индивидуальной линии задержки на передатчике согласно адресной кодовой комбинации. Никаких частотных перестроек в этих системах не требуется, что очень удешевляет аппаратуру и обеспечивает ее надежность.

Контрольные вопросы

1. В каких случаях используются многоканальные системы связи?

2. Какие существуют принципы объединения и разделения каналов?

3. Поясните принцип работы систем с ЧРК.

4. Каковы причины переходных искажений в системах с ЧРК?

5. Поясните принцип работы системы с ВРК.

6. Как должна выбираться длительность цикла $\boldsymbol{T_Ц}$ в системах с ВРК?

7. В чем состоит сущность разделения сигналов по форме?

8. Какие способы объединения и разделения каналов нашли наибольшее распространение в системах связи?

9. В чем состоит основное различие между многоканальными системами и системами множественного доступа?

10. Какие существуют протоколы построения систем множественного доступа?

ГЛАВА 9. ЭФФЕКТИВНОСТЬ СИСТЕМ СВЯЗИ

9.1. Оценка эффективности систем связи

В предыдущих разделах система электрической связи (СЭС) рассматривалась как совокупность отдельно функционирующих элементов. Вместе с тем СЭС представляет собой сложный комплекс, характеризующийся иерархичностью структуры, наличием прямых, обратных и перекрестных связей между элементами. Следовательно, необходимо рассматривать работу СЭС в целом, для чего нужно определить алгоритмы ее функционирования с учетом взаимодействия и свойств элементов. Для решения таких задач воспользуемся системным подходом (системным анализом).

9.1.1. Подходы к оценке эффективности

Системный подход представляет собой совокупность общих принципов и рекомендаций, определяющих научную и практическую деятельность исследователя при анализе и синтезе сложных объектов.

Принцип системного подхода базируется на представлении объекта как сложной системы с учетом ее специфических связей и свойств. Система определяется как целостное образование, состоящее из связанных между собой элементов. Поэтому система обладает собственными свойствами, не вытекающими непосредственно из свойств ее элементов.

Свойства системы прежде всего определяются ее целевым назначением (целями функционирования), которое трактуется как совокупность задач, решаемых данной системой. Для получения желаемого результата необходимо совершить определенную совокупность операций, направленных на достижение поставленной задачи. Эти операции реализуются за счет использования некоторых ресурсов системы. В СЭС такими операциями являются кодирование, модуляция, усиление сигнала, демодуляция, декодирование, селекция сигналов и т.п., а ресурсами системы являются мощность сигнала и полоса частот канала. Таким образом СЭС имеет все признаки сложной системы.

Весьма важен анализ взаимодействия СЭС со средой. Среда в СЭС – это

не только линия связи (среда распространения сигнала), используемая для передачи сигналов от передатчика к приемнику, но и другие системы естественного и искусственного происхождения, оказывающие определенные воздействия на систему связи. Обычно это мешающие воздействия (помехи и искажения), затрудняющие качественную передачу информации по каналу связи. Необходимость борьбы с вредными воздействиями помех существенно усложняет СЭС.

Для исследования СЭС создается ее модель, в которой отображены наиболее существенные свойства и признаки. Математическая модель СЭС представляет собой совокупность математических соотношений, отображающих структуру системы, алгоритмы ее функционирования, статистические характеристики канала, сигнала и помех, технические и экономические показатели системы. Стохастический характер помех и непредсказуемость сообщений и сигналов обусловливают широкое использование вероятностных моделей.

При выборе комплексного показателя технико-экономической эффективности системы исходят из того, что он должен иметь прямую связь с ее целевым назначением, объективно характеризовать основные свойства, быть чувствительным к изменению определяющих параметров системы и наряду с этим должен быть достаточно простым, чтобы им можно было пользоваться. Проблема заключается в том, что не все цели системы можно адекватно отразить в количественной форме. Например, трудно численно оценить степень удовлетворения потребности людей в общении с помощью средств связи. Тем не менее, решение вопросов выбора наиболее целесообразных вариантов СЭС в конечном итоге сводится к решению задач оптимизации этих систем по выбранным критериям качества.

9.1.2. Критерии эффективности

Обобщенной характеристикой эффективности систем связи является коэффициент использования канала по пропускной способности (информационная эффективность) который характеризует реальную скорость передачи ин-

формации R по отношению к пропускной способности C канала связи [5, 20, 21, 32]:

$$\eta = \frac{R}{C}. \tag{9.1}$$

Информационная эффективность η всегда меньше единицы; чем ближе η к единице, тем совершеннее система.

Для оценки эффективности систем связи вводятся также коэффициент использования канала по мощности (энергетическая эффективность)

$$\beta = \frac{R}{P_c / N_0} \tag{9.2}$$

и коэффициент использования канала по полосе частот (частотная эффективность)

$$\gamma = \frac{R}{\Delta F}. \tag{9.3}$$

В этих формулах P_c – мощность сигнала; N_0 – спектральная плотность шума; ΔF – ширина полосы частот, занимаемой сигналом.

Предельные возможности системы передачи информации можно оценить с помощью выражения для пропускной способности гауссовского непрерывного канала связи с полосой частот ΔF :

$$C = \Delta F \log_2\left(1 + \frac{P_c}{P_ш}\right). \tag{9.4}$$

Здесь $P_c = E_b B$ – средняя мощность сигнала: E_b – энергия, затрачиваемая на передачу одного бита информации; $B = 1/T_b$ – скорость передачи информации источника; T_b – время передачи источником одного бита информации; $P_ш = N_0 \Delta F$ – средняя мощность шума в полосе частот.

В реальных СЭС скорость передачи информации B [Бит/с], меньше пропускной способности непрерывного канала: $B \le C$. Можно показать, что после элементарных преобразований это неравенство приводится к виду [5, 21, 32]:

$$\beta \le \frac{\gamma}{2^{\gamma} - 1}, \tag{9.5}$$

$$\text{где} \quad \beta = \frac{1}{h_э^2} = \frac{N_0}{E_b}. \tag{9.6}$$

Тогда информационная эффективность для гауссовского непрерывного канала может быть найдена по формуле [5, 20, 21, 32]:

$$\eta = \frac{\gamma}{\log_2\left(\gamma/\beta + 1\right)}. \tag{9.7}$$

Согласно теореме Шеннона, при соответствующих способах передачи и приема величина η может быть сколь угодно близкой к единице. При $\eta = 1$ получаем предельную зависимость между β и γ:

$$\beta = \frac{\gamma}{2^\gamma - 1}. \tag{9.8}$$

Наглядно данная зависимость представляется в виде кривой на $\beta\gamma$ плоскости (рис. 9.1). Эта зависимость, часто называется границей (пределом) Шеннона: она отражает наилучший обмен между β и γ в непрерывном канале. Анализ соотношения (9.6) и предела Шеннона показывает, что повышение частотной эффективности (т.е. снижение затрат полосы $\frac{1}{\gamma}$) требует увеличения энергетических затрат (снижения энергетической эффективности). Для непрерывного канала частотная эффективность изменяется в пределах от 0 до ∞, в то время как энергетическая эффективность ограничена сверху [20, 21]:

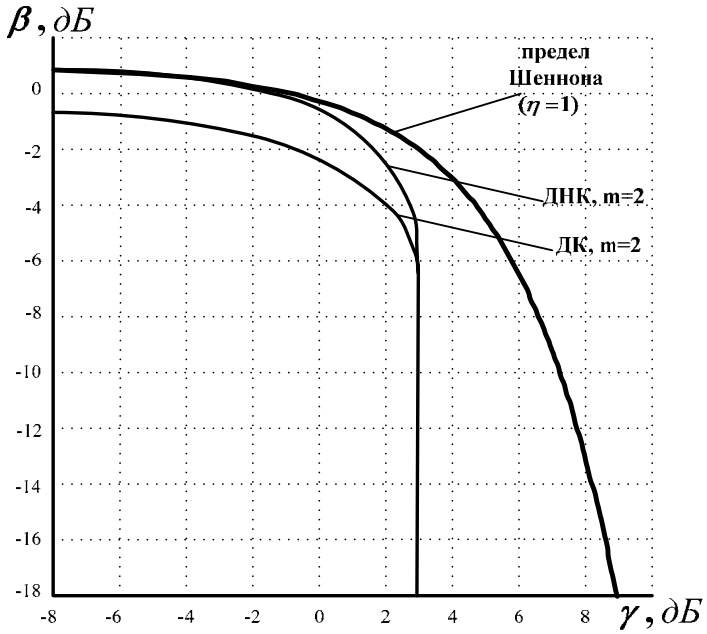


Рис. 9.1. Связь частотной (γ) и энергетической (β) эффективности

$$\beta_{\max} = \lim_{\gamma \to 0} \beta = \lim_{\gamma \to 0} \frac{\gamma}{2^\gamma - 1} = \frac{1}{\ln 2} \approx 1{,}443.$$

Аналогичные предельные зависимости $\beta = f(\gamma)$ можно получить и для

других моделей канала, если в (9.2) и (9.3) вместо скорости R подставить выражение для пропускной способности соответствующего канала. Предельные зависимости $\beta\gamma$ - номограммы позволяют определить системы, удовлетворяющие заданным требованиям по энергетической и частотной эффективности, и установить, насколько эти показателя близки к предельным.

9.1.3. Эффективность аналоговых и цифровых систем

В системах передачи дискретных сообщений сигнал формируется с помощью кодирования и модуляции. При этом кодирование осуществляется обычно в два этапа: кодирование источника с целью сокращения его избыточности и кодирование канала с целью уменьшения вероятности ошибки за счет введения избыточности кода. При этом выражение (9.1) для информационной эффективности системы передачи дискретных сообщений можно представить в виде произведения [5, 20]:

$$\eta = \frac{R}{C} = \eta_{ки} \cdot \eta_{кк} \cdot \eta_{м} . \qquad (9.9)$$

где $\eta_{ки}$ – эффективность кодера источника; $\eta_{кк}$ – эффективность кодера канала; $\eta_{м}$ – эффективность модема, зависящая от вида модуляции и способа обработки сигнала в канале.

Средняя скорость передачи информации в системе при использовании многопозиционных сигналов длительностью T равна $R = \frac{\log_2 m}{T} R_{кк}$ (бит/с), где $R_{кк} = k/n$ – скорость помехоустойчивого кода. Тогда энергетическая эффективность [5, 20, 21, 32]

$$\beta = \frac{R}{P_c / N_0} = \frac{R_{кк} \log_2 m}{E_0 / N_0} = \frac{1}{E_b / N_0}, \qquad (9.10)$$

частотная эффективность может быть найдена по формуле

$$\gamma = \frac{R}{\Delta F} = \frac{\log_2 m}{T \Delta F}, \qquad (9.11)$$

где $E_0 = P_c T = E_b R_{кк} \log_2 m$ – энергия сигнала; $E_b = \frac{E_0}{R_{кк} \log_2 m}$ – энергия, затрачи-

ваемая на передачу одного бита информации.

Значения $h_э^2 = E_b / N_0$ можно определить по известным формулам или графикам, рассчитанным для вероятности ошибки $p_{ош} = f\left(h^2 = h_э^2 R_{кк} \log_2 m\right)$.

На рис. 9.1 приведены предельные кривые $\beta = f(\gamma)$ для симметричных двоичных дискретных и дискретно-непрерывных каналов. При этом выходом ДНК считается согласованный фильтр в оптимальной схеме приема дискретных сообщений при примитивном кодировании ($R_{кк} = 1$).

Для двухпозиционных систем $m = 2$ предельное значение полосы пропускания канала равно частоте манипуляции. В этом случае частотная эффективность (предел Найквиста) будет иметь наибольшее значение, равное $\gamma_{\max} = 2$. Было показано, что в двоичных симметричных каналах с различными видами модуляции максимум энергетической эффективности наступает при $p_{ош} = 0{,}5$, однако удельная скорость передачи при этом стремится к нулю. Предельные значения показателей эффективности достигаются при $\boldsymbol{R = C}$ и при малой вероятности ошибки. Для определения β и γ могут использоваться приближенные формулы:

$$\gamma \approx \frac{\log_2 m}{n}; \quad \beta \approx \frac{1}{E_b / N_0}; \tag{9.12}$$

где n – размерность сигнала, в m-позиционной системе. В табл.9.1 приведены значения m и формулы для приближенных расчетов γ некоторых ансамблей сигналов.

Таблица 9.1

Формулы для приближенных расчетов частотной эффективности некоторых ансамблей сигналов

Ансамбль сигналов	Ортогональный	Биортогональный	Симплексный
m	n	$2n$	$n+1$
γ	$\frac{\log_2 m}{m}$	$\frac{2\log_2 m}{m}$	$\frac{\log_2 m}{m-1}$

В реальных системах вероятность ошибки всегда имеет ненулевое значение и $\eta < 1$. В этих случаях при заданном значении $p_{ош} = const$ можно определить отдельно β и γ и построить кривые $\beta = f(\gamma)$.

В координатах β и γ каждому варианту реальной системы будет соответствовать точка на плоскости (рис. 9.2) [5, 20, 21, 32]. Все эти точки располагаются ниже предельной кривой Шеннона и ниже предельной кривой соответствующего канала. Ход этих кривых зависит от вида модуляции, метода кодирования и способа обработки сигналов. Около графиков на рис. 9.2 указано число позиций дискретного сигнала m. Кривые рассчитаны на основании формул оценки помехоустойчивости различных методов модуляции (раздел 3) для оптимального приема сигналов при вероятности ошибки на бит $p_b = 10^{-5}$. При этом занимаемая полоса частот для ЧМн $\Delta F = \dfrac{m}{T \log_2 m}$, а для ФМн (АМн) $\Delta F = \dfrac{1}{T \log_2 m}$.

Анализ рис. 9.2 показывает, что в системах с ЧМн при увеличении числа позиций m энергетическая эффективность β увеличивается, а частотная эффективность γ уменьшается. В системах с ФМн и ОФМн, наоборот, с увеличением m коэффициент β уменьшается, а γ – увеличивается. Таким образом, условия обмена β на γ за счет изменения числа позиций сигналов в системах связи с ЧМн и ФМн различны.

Представленные на рис. 9.2 результаты позволяют определить системы, удовлетворяющие заданным требованиям по энергетической и частотной эффективности, и установить, насколько эти показатели близки к предельным.

После выбора системы по показателям β и γ, информационная эффективность вычисляется с использованием формулы (9.7).

Например, для сигналов АМн-2 показатель информационной эффективности составляет $\eta \approx 0{,}228$, а для ЧМн-2 $\eta \approx 0{,}145$; для ФМн-2 $\eta \approx 0{,}25$, а для ФМн-4 $\eta \approx 0{,}47$.

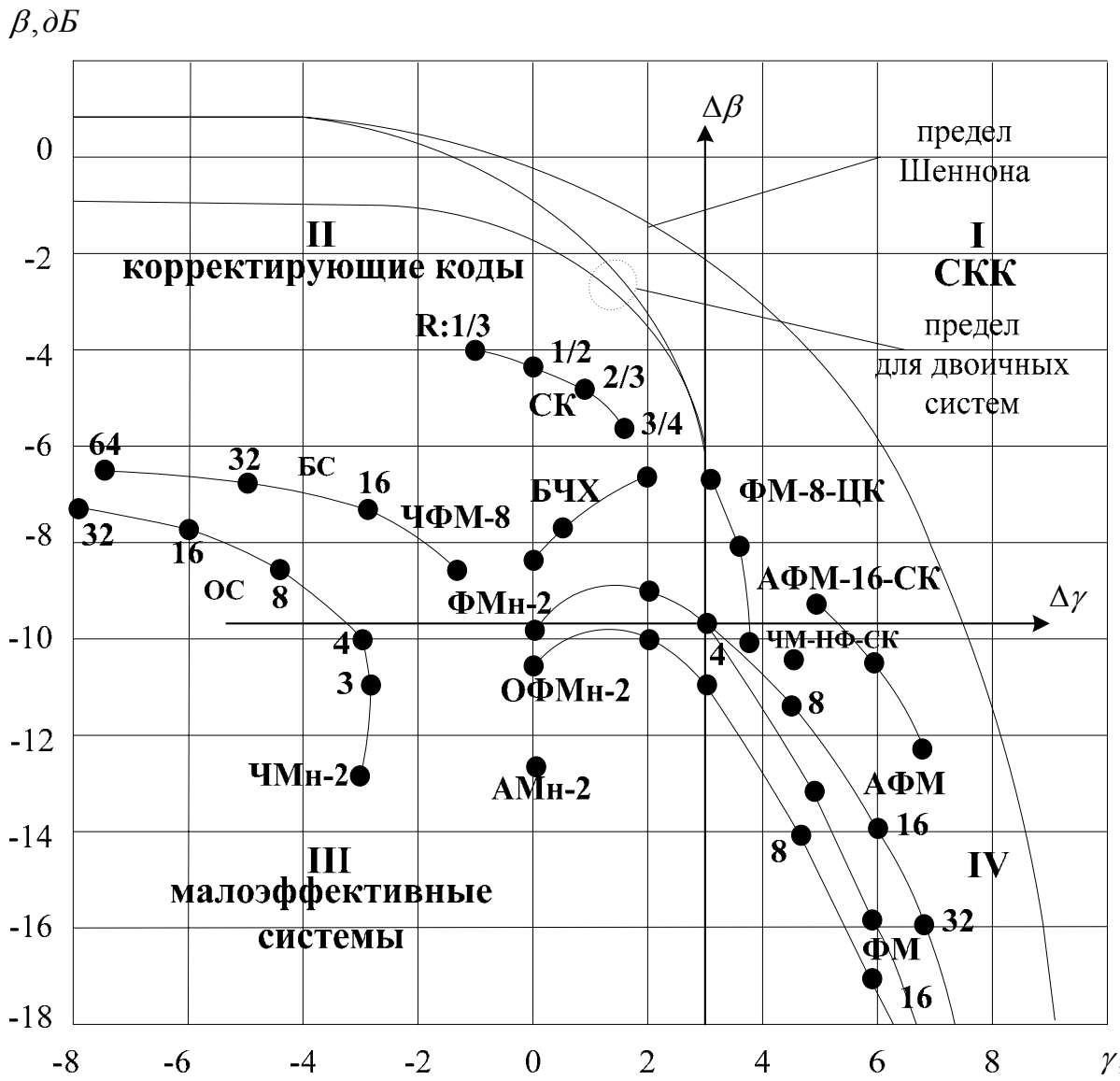


Рис. 9.2. Кривые энергетической и частотной эффективности цифровых систем связи

Анализ предельных кривых показывает, что эффективность дискретных систем передачи можно существенно повысить, если вместо двоичных применять многопозиционные сигналы ($m > 2$).

Эффективность передачи непрерывных сообщений в значительной степени зависит от вида модуляции. Для сравнительного анализа различных видов модуляции обычно используют выигрыш по отношению сигнал/шум ($\rho_{вых}$) и коэффициент использования пропускной способности каналов связи (ν)[21]:

$$\eta = \frac{\log \rho_{вых}}{\nu \cdot \log\left(\rho_{вых} / g + 1\right)} \qquad (9.13)$$

В табл. 9.2 приведены данные сравнительного анализа эффективности различных видов модуляции, полученные при $\rho_{вых} = 40$ дБ и пик-факторе $\text{П} = 3$ для гауссовского канала при оптимальной обработке сигналов [5, 20, 21, 32].

Таблица 9.2

Значения выигрыша и информационной эффективности некоторых систем передачи непрерывных сообщений

Система модуляции	$\nu = \frac{F}{F_c}$	$g = \frac{\rho_{вых}}{\rho_{вх}}$	$g' = \frac{g}{\nu}$	$\eta = \frac{R}{C}$
АМ	2	0,2	0,1	0,42
БМ	2	2	1	0,50
ОМ	1	1	1	1
ФМ	20	222	11,1	0,12
ЧМ	20	666	33,3	0,17
ФИМ-АМ	20	666	33,3	0,17
ИКМ-АМ	20	250	12,5	0,23
ИКМ-ЧМ	20	500	25	0,32
ИКМ-ФМ	20	1000	50	0,48
ИС	20	6310	315	1

Анализ показывает, что наибольшая информационная эффективность достигается при однополосной модуляции, однако значение обобщенного выигрыша для этого вида модуляции ($g' = 1$) свидетельствует о том, что в системе отсутствует выигрыш по помехоустойчивости. Одноканальные системы ЧМ и ФИМ примерно равноценны. В этих системах, а также в цифровых системах с ИКМ, высокая помехоустойчивость может быть достигнута с помощью увеличения ширины спектра сигнала, т.е. за счет частотной избыточности. При больших индексах ФМ и ЧМ приближаются по помехоустойчивости к идеальной системе (выигрыш составляет десятки и сотни раз), но информационная эффективность таких систем мала (0,12 – 0,17) из-за большой частотной избыточности. Основными способами повышения эффективности передачи непре-

рывных сообщений являются устранение избыточности, статистическое уплотнение и применение цифровых видов модуляции.

Аналоговые системы ОМ, АМ и узкополосная ЧМ обеспечивают высокую частотную эффективность при сравнительно низкой энергетической эффективности. Применение этих систем целесообразно в каналах с хорошей энергетикой (при больших значениях $\rho_{вх}$) или в тех случаях, когда требуемое значение $\rho_{вых}$ мало. Цифровые системы обеспечивают высокую β-эффективность при достаточно хорошей γ-эффективности. В каналах с ограниченной энергетикой (при малых значениях $\rho_{вх}$) преимущества цифровых систем особенно заметны. При высоком качестве передачи, когда требуемые значения $\rho_{вых}$ велики, широкополосная ЧМ и цифровые системы обеспечивают примерно одинаковую эффективность.

В многоканальных системах эффективность связи снижается за счет несовершенства системы разделения сигналов.

Расчеты показывают, что наиболее эффективным является метод временного разделения каналов; менее эффективен метод частотного разделения. При временном разделении пропускная способность не зависит от числа каналов, т.к. в каждый момент времени передается только один сигнал. При ЧРК пропускная способность канала с ограниченной средней мощностью сигнала также не зависит от числа каналов. При разделении по форме между n парциальными каналами делится только мощность, полоса частот и время передачи используются одновременно всеми сигналами. В этом случае информационная эффективность уменьшается с увеличением n, причем амплитудное ограничение сигнала слабо влияет на эту зависимость.

Показатели частотной, энергетической и информационной эффективности для систем с множественным доступом определяются на основании суммарной скорости передачи СЭС, зависящей от методов формирования и обработки информационных сигналов в парциальных каналах и методов доступа.

9.2. Выбор сигналов и помехоустойчивых кодов

Эффективность систем передачи дискретных сообщений можно существенно повысить путем применения многопозиционных сигналов и корректирующих кодов.

9.2.1. Многопозиционные сигналы

Ансамбль сигналов $s_i(t)$, где $i \in \overline{0, m-1}$, на отрезке $[0, T]$ можно представить в виде [5, 21, 32]:

$$s_i(t) = \sum_{j=0}^{n} C_{ij} \varphi_j(t). \qquad (9.14)$$

Здесь n-число отсчетов на интервале T, а $\{\varphi_j(t)\}$ – система базисных ортонормированных функций:

$$\int_{-T/2}^{T/2} \varphi_i(t)\varphi_j(t) = \begin{cases} 1, \text{ при } i = j \\ 0, \text{ при } i \neq j \end{cases}. \qquad (9.15)$$

Геометрически каждому сигналу ансамбля $s_i(t)$ соответствует точка (или вектор) в n-мерном пространстве с координатами $(C_{i1}, C_{i2}, ..., C_{in})$. В соответствии с формулой (1.8), энергия сигнала

$$E = \int_0^T s_i^2(t) dt, \qquad (9.16)$$

а расстояние (1.42) между сигналами

$$d(s_i, s_j) = \sqrt{[s_i(t) - s_j(t)]^2} = \sqrt{E_i + E_j - 2B_{ij}E_{ij}}, \qquad (9.17)$$

где

$$B_{ij} = \frac{1}{E_{ij}} \int_0^T s_i(t) s_j(t) dt, \qquad (9.18)$$

коэффициент взаимной корреляции рассматриваемых сигналов.

На рис. 9.2 приведены $\beta\gamma$-диаграммы для некоторых ансамблей многопозиционных сигналов. Центральное место на рис. 9.2 занимают кривые для систем с сигналами ФМн-4, которые относятся к классу многопозиционных при $m = 4$. В цифровых сетях система ФМн-4 является наиболее распространенной и принята в качестве стандарта, поэтому при сравнительной оценке эффективно-

сти систем она принята за эталон. Если начало координат перенести в точку, соответствующую ФМн-4, то в новой системе координат по вертикальной оси будет отсчитываться энергетический выигрыш $\Delta\beta$ рассматриваемых систем по сравнению с ФМн-4, а по горизонтальной оси – выигрыш $\Delta\gamma$ по удельной скорости. В этой системе координат все возможные системы связи можно условно разделить на четыре группы, соответствующие четырем квадрантам на плоскости.

Малоэффективные системы (III квадрант), имеющие относительно ФМн-4 проигрыш по β и γ, например, АМн-2, ЧМн-2. Системы с высокой энергетической эффективностью (II квадрант), обеспечивающие выигрыш по β и проигрыш по γ (системы с корректирующими кодами). Системы с высокой частотной эффективностью (IV квадрант), обеспечивающие выигрыш по γ и проигрыш по β (системы с многопозиционными ФМн и АФМ сигналами). Высокоэффективные системы (I квадрант), позволяющие получить одновременно выигрыш по обоим показателям β и γ на основе применения сложных сигнально-кодовых конструкций).

Можно выделить также два класса многопозиционных сигналов. К первому отнесем так называемые «плотные» сигналы, когда с ростом объема ансамбля m при фиксированной размерности n расстояние между сигнальными точками уменьшается, а удельная скорость γ возрастает при соответствующем снижении энергетической эффективности β. Примерами таких сигналов служат многопозиционные ФМн и АФМ.

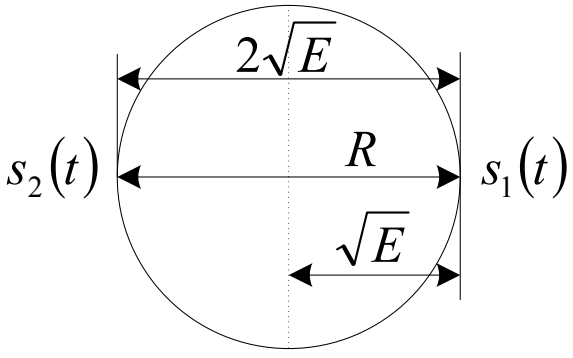


Рис. 9.3. Двухпозиционный сигнал (ФМн-2)

Примером сигналов, у которых сигнальные векторы располагаются на прямой, являются двоичные противоположные сигналы ФМн-2 (рис. 9.3). Им соответствует два симметрично распо-

ложенных относительно начала координат вектора длиной $\sqrt{E}$. Расстояние между сигналами $d_{12} = 2\sqrt{E}$, а коэффициент корреляции $R_{12} = -1$.

К этому же классу относятся и широко используемые сигналы с фазовой манипуляцией и числом позиций $m = 4$ (рис. 9.4). Сигналы ФМн-4 имеют одинаковые энергии, а сигнальные точки располагаются на одинаковом расстоянии от начала координат. На амплитудно-фазовой плоскости они образуют квад-

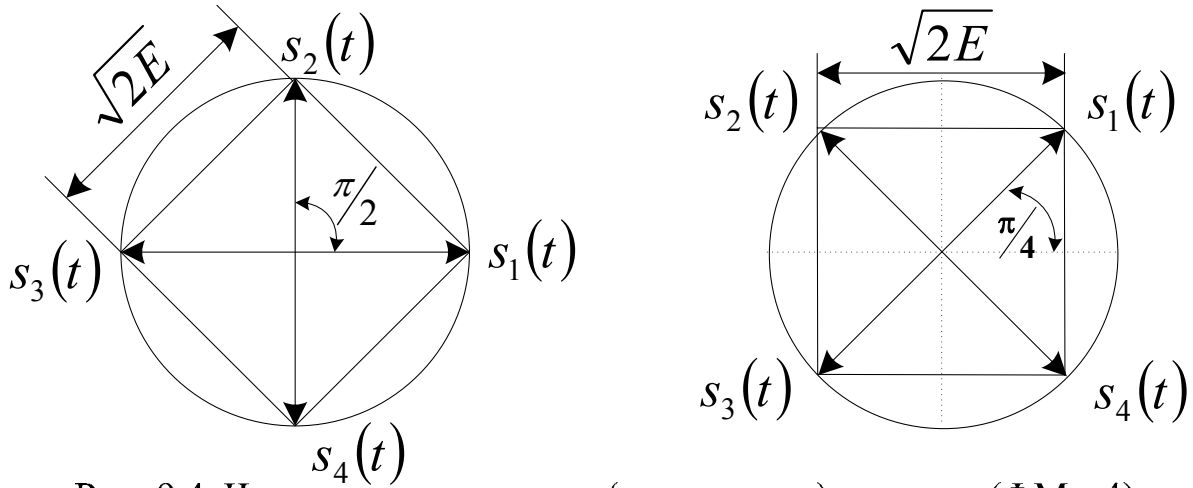


Рис. 9.4. Четырехпозиционные (двукратные) сигналы (ФМн-4)

ратную сеть. Сигналы этого ансамбля отличаются только начальными фазами. Расстояние между ближайшими сигнальными точками равно $\sqrt{2E}$, а между противоположными сигналами $2\sqrt{E}$.

Многопозиционные сигналы с ФМн-8 образуют круговую сеть с равномерным распределением точек по окружности (рис. 9.5).

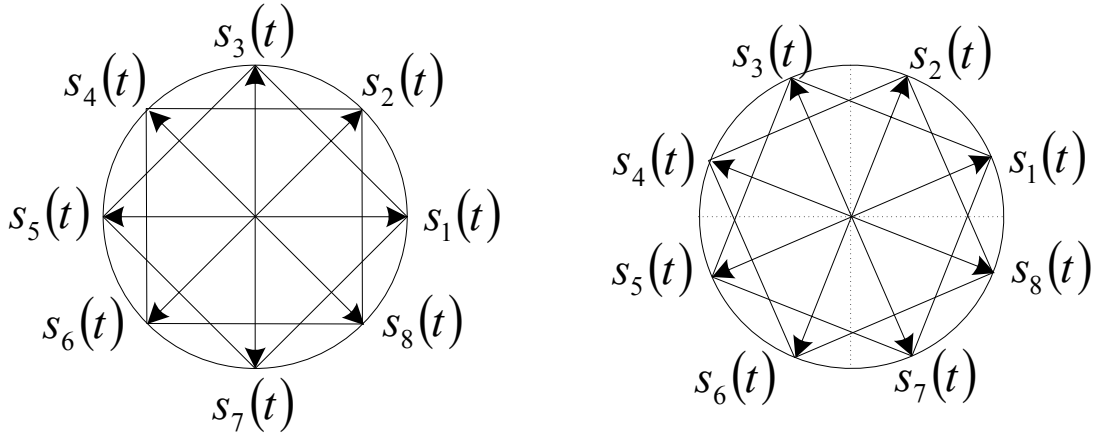


Рис. 9.5. Восьмипозиционные (трехкратные) сигналы (ФМн-8)

Для сигналов АФМ–4 три сигнала равномерно распределены по окружности, а четвертый расположен в центре окружности (рис. 9.6,а).

На рис. 9.6,б показано также расположение сигнальных точек в восьми позиционной системе с амплитудно-фазовой модуляцией (АФМ-8).

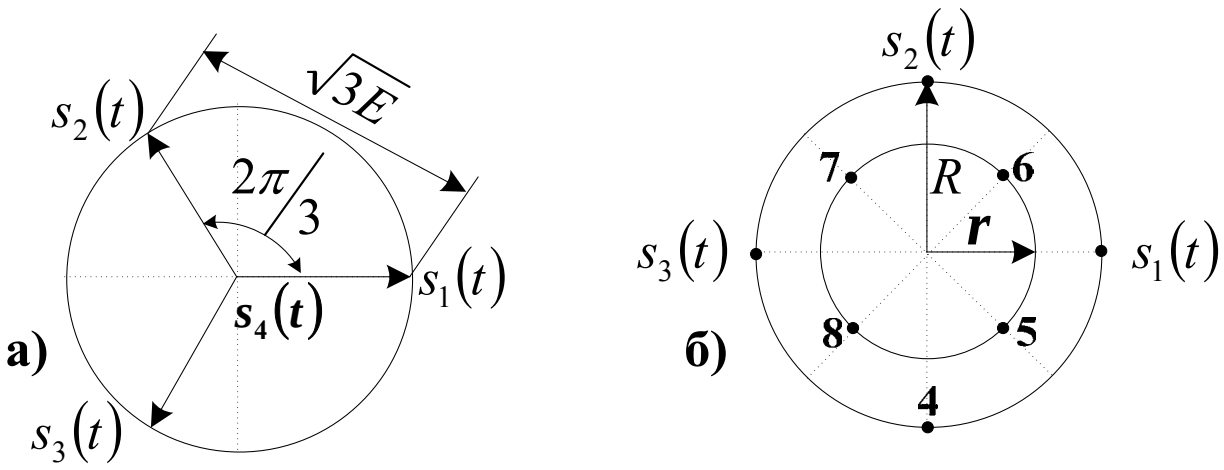


Рис. 9.6. Сигналы амплитудно-фазовой модуляции:
а) АФМ-4, б) АФМ-8

Ко второму классу отнесем ортогональные, биортогональные и симплексные сигналы. Это примеры «разнесенных» сигналов, когда с увеличением m увеличивается расстояние между сигнальными точками и соответственно увеличивается энергетическая эффективность за счет снижения частотной эффективности.

Если сигнальные точки выбрать на линиях, совпадающих с ортогональными областями на расстояниях $\sqrt{E}$ от начала координат, то получим систему ортогональных сигналов. Число сигналов в таком ансамбле $m = n$.

Двоичные ортогональные сигналы являются примером сигналов, у которых сигнальные точки располагаются в плоскости (рис. 9.7).

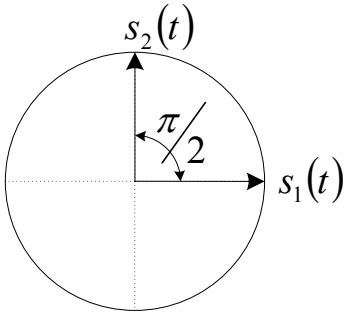


Рис. 9.7. Двоичные ортогональные сигналы (ЧМн-2)

Им соответствуют два ортогональных вектора на плоскости длиной $\sqrt{E}$. Расстояние между сигналами $d_{12} = \sqrt{2E}$, а коэффициент корреляции $R_{12} = 0$.

Если в качестве сигналов взять отрезки гармо-

нических колебаний разных частот $\omega_1, \omega_2, ..., \omega_m$, удовлетворяющих условию ортогональности, то получим сигналы многочастотной модуляции (МЧМ). Ортогональные сигналы образуют эквидистантную систему; расстояния между любыми двумя сигнальными точками одинаковы: $d_{ij} = \sqrt{2E}$. Перспективным вариантом ЧМн сигналов являются частотно-манипулированные сигналы с непрерывной фазой.

Биортогональные сигналы образуются путем добавления к каждому ортогональному сигналу противоположного. При этом общее число сигналов удваивается: $m = 2n$.

Симплексные сигналы отстоят друг от друга на одинаковом расстоянии. В n-мерном пространстве они образуют правильный симплекс с числом вершин $m = n + 1$. В двумерном пространстве сигнальные точки лежат в вершинах равностороннего треугольника (рис. 9.8). Расстояние между сигнальными точками симплексного ансамбля $d = \sqrt{2(n+1)\frac{E}{n}}$. При $n = 1$ симплексные сигналы совпадают с противоположными. Для ансамблей с большим объемом $(n >> 1)$ симплексные сигналы по своим свойствам и в частности по помехоустойчивости близки к ортогональным $d \approx \sqrt{2E}$.

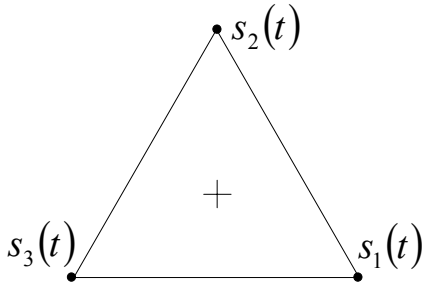


Рис. 9.8. Симплексные сигналы при ***n*=2**

Построение ансамблей многопозиционных сигналов можно осуществить и на основе двоичных последовательностей: для этого можно использовать элементарную матрицу Уолша–Адамара (A_0). Формирование матриц высшего порядка подчинено следующему правилу: матрица младшего порядка трижды повторяется в позитивной и один раз в негативной форме. При достижении размерности матрицы 4×4, она уже представляет собой ансамбль многопозиционных ортогональных сигналов с $m = 4$ [5, 21]:

$$A_0 = +1,\quad A_1 = \begin{vmatrix} +1 & +1 \\ +1 & -1 \end{vmatrix},\quad A_2 = \begin{vmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{vmatrix}. \qquad (9.19)$$

Каждая строка этой матрицы (последовательность двоичных символов) образует один сигнал; нетрудно убедиться в том, что строки (столбцы) этой матрицы взаимно ортогональны. Дополняя матрицу A_2 инверсиями строк, можно получить матрицу B, представляющую ансамбль $m = 8$ биортогональных сигналов:

$$B = \begin{vmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \\ -1 & -1 & -1 & -1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & +1 & -1 \end{vmatrix}. \qquad (9.20)$$

Ансамбли с большим числом сигналов строятся аналогично. В асинхронно-адресных системах широко используются ансамбли «почти ортогональных» сигналов, которые также формируются на основе двоичных последовательностей. Это известные рекуррентные псевдослучайные D и M-последовательности.

Приведенные на рис. 9.2 кривые позволяют количественно оценить обменный выигрыш (проигрыш) различных систем. Так, применение многопозиционных АФМ сигналов с $m = 16$ позволяет получить частотный выигрыш в 2 раза ($\Delta\gamma = 3$ дБ) в обмен на снижение энергетического выигрыша $\Delta\beta$ более чем на 4 дБ. Получить энергетический выигрыш в обмен на снижение удельной скорости можно с помощью ортогональных и биортогональных сигналов.

9.2.2. Корректирующие коды

Наряду с многопозиционными сигналами для повышения эффективности СЭС широко используются помехоустойчивые коды. Применение корректирующих кодов позволяет повысить верность передачи сообщений или при за-

данной верности повысить энергетическую эффективность системы. Это особенно важно для систем с малой энергетикой, например, систем спутниковой связи.

На практике используются как блочные, так и непрерывные коды. На рис. 9.2 приведены кривые эффективности для циклического кода Боуза-Чоудхури-Хоквингема (БЧХ) и для сверточного кода (СК) с декодированием по алгоритму Витерби.

Применение циклического кода позволяет получить энергетический выигрыш $\Delta\beta = 2...4$ дБ, а сверточного кода $\Delta\beta = 4...6$ дБ в обмен на снижение частотной эффективности примерно в 2 раза (3 дБ).

Энергетический выигрыш $\Delta\beta$ от применения помехоустойчивого кодирования тем больше, чем выше требуемая верность передачи. Для непрерывного канала с белым гауссовским шумом при требуемой вероятности ошибки 10^{-5} предельный энергетический выигрыш кодирования по сравнению с ФМн-2 без кодирования при оптимальном когерентном приеме составляет примерно 10 дБ.

Расчетные кривые на рис. 9.2 показывают, что применение циклического кода в канале с ФМн или сверточного кода в канале с АФМ позволяет повысить одновременно энергетическую, так и частотную эффективности. Построение таких высокоэффективных систем на основе сигнально-кодовых конструкций ведет к неизбежному увеличению сложности системы. Не пропускная способность, а сложность является ограничивающим фактором при построении высокоэффективных систем. Задача состоит в том, чтобы построить систему, удовлетворяющую высоким показателям эффективности, при допустимой сложности.

При современной элементной базе затраты на реализацию кодирующих и декодирующих устройств значительно сократились. В то же время стоимость энергетики канала практически не изменилась. Таким образом, «цена» выигрыша $\Delta\beta$ за счет кодирования может быть существенно меньше цены того же выигрыша, полученного за счет увеличения энергетики канала (мощности сигнала или размеров антенн).

Отметим, что выбор способов кодирования и модуляции зависит от характеристик канала. Улучшение этих характеристик, например, путем адаптации к помехам и оценивания искажений сигнала и их последующей компенсации, снижает потери в канале и создает лучшие условия для применения корректирующих кодов.

9.3. Оптимизация систем связи

Повышение таких важнейших показателей систем электрической связи, как скорость и верность передачи, связано со значительными частотными и энергетическими затратами. Сравнение между собой различных СЭС осуществляется по степени использования ими основных ресурсов канала связи (пропускной способности, мощности, занимаемой полосы частот), выражаемой через показатели информационной, энергетической и частотной эффективности. Создание СЭС, в которых достигаются близкие к предельным показатели эффективности, требует совместного согласования кодека и модема с учетом статистических свойств непрерывного канала.

9.3.1. Согласование методов модуляции и кодирования

Эффективный путь повышения удельной скорости передачи информации заключается в увеличении числа используемых сигналов m на интервале T. Однако увеличение m приводит к уменьшению расстояния между ближайшими сигналами ансамбля и снижению энергетической эффективности.

При высоких требованиях к верности передачи целесообразным становится применение помехоустойчивых кодов, которые позволяют повысить энергетическую эффективность за счет снижения частотной. Помехоустойчивое кодирование позволяет снизить необходимую величину мощности сигнала поскольку расстояние между кодовыми комбинациями увеличивается. Одновременное требование большой скорости и верности передачи в условиях ограниченного частотного и энергетического ресурса может быть выполнено при использовании многопозиционных сигналов и помехоустойчивых кодов.

При многопозиционной модуляции, когда по каналам связи передается блок из n кодовых символов, важно также правильно выбрать манипуляционный код, определяющий правило сопоставления с каждым передаваемым сигналом определенного блока кодовых символов. Общий принцип заключается в том, что большему расстоянию по Хэммингу между кодовыми блоками должно соответствовать большее расстояние по Евклиду между отображающими их сигналами.

Создание СЭС, в которых достигаются близкие к предельным показатели эффективности, требует совместного согласования кодека и модема с учетом статистических свойств непрерывного канала. Это означает, что кодирование и модуляцию необходимо рассматривать как единый процесс формирования сигнала, а демодуляцию и декодирование – как процесс оптимального приема сигнально-кодового блока в целом.

Согласование модуляции и кодирования сводится к поиску такого заполнения сигнального пространства, при котором обеспечивается высокая удельная скорость (сигналы расположены достаточно плотно) и одновременно высокая помехоустойчивость (сигналы достаточно далеко друг от друга).

Комбинирование различных ансамблей m-ичных сигналов, помехоустойчивых и манипуляционных кодов порождает множество конструкций. Однако только согласованные варианты обеспечивают повышение частотно-энергетической эффективности СЭС. Эти варианты называют сигнально-кодовыми конструкциями (СКК).

Рассмотрим обобщенную схему передачи дискретных сообщений, приведенную на рис.9.9.

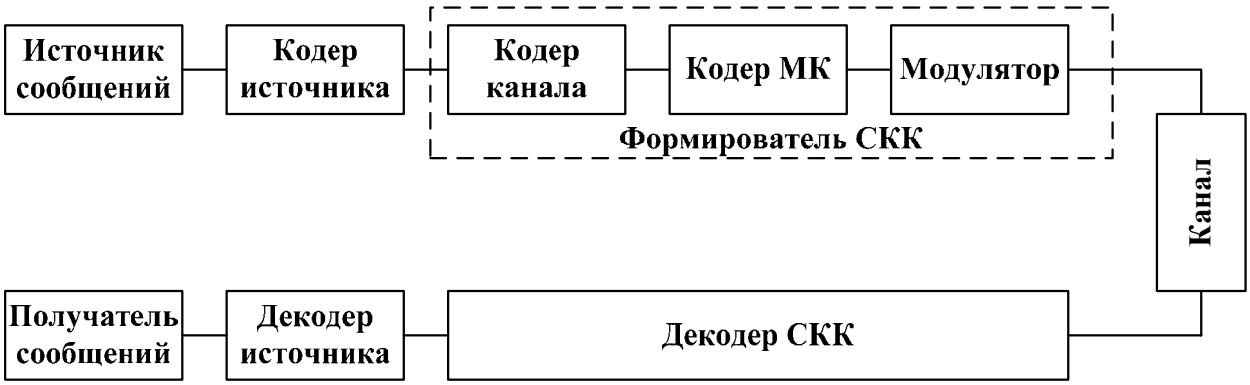


Рис. 9.9. Обобщенная схема передачи дискретных сообщений

Последовательность символов с выхода кодера канала разбивается на блоки по n символов. Отображение блоков в сигналы, формируемые модулятором, осуществляется по правилу манипуляционного кодирования, т.е. манипуляционный код определяет правило соответствия блоков кодовых символов m-ичным сигналам. Например, в случае двоичного канального кода каждому из $m = 2^n$ кодовых блоков ставится в соответствие один из 2^n сигналов.

Оптимальная процедура приема СКК заключается в обработке сигнально-кодового блока в целом. Поэтому демодулятор, декодеры канала и манипуляционный декодер рассматриваются как единое устройство – декодер СКК (рис.9.9).

Декодер СКК строится так, чтобы минимизировалась вероятность ошибки приема. Оптимальный декодер реализует принцип максимального правдоподобия. При белом гауссовском шуме выбирается кодовое слово, находящееся на минимальном евклидовом расстоянии от принятого.

Декодирование МК можно рассматривать как последний этап обработки сигнально–кодового блока оптимальным декодером СКК. При этом декодер канала работает в метрике Евклида с сигналами, а не с их двоичными представлениями по правилу манипуляционного кода. Схема поэлементного приема, наоборот, ориентирована на применение декодера канала в метрике Хемминга, т.е. обработку двоичных величин после декодера манипуляционного кода.

Следовательно, достижение наибольшей эффективности возможно при декодировании по алгоритму максимального правдоподобия сигнально-

кодового блока «в целом».

Необходимо отметить, что в принципе любой сигнальный ансамбль на выходе последовательно соединенных кодека и модема может быть отнесен к СКК. Введение понятия СКК отражает подход к модуляции и кодированию как процессу объединения сигналов и кодов в единую эффективную конструкцию.

9.3.2. Классификация сигнально – кодовых конструкций

В основе формирования СКК лежат операции отображения информационной последовательности в кодовую путем внесения избыточности и кодовой последовательности в канальную заданием манипуляционного кода. Помехоустойчивое кодирование, повышающее энергетическую эффективность СЭС, является одной из важнейших операций формирования СКК. Получаемый при этом энергетический выигрыш от кодирования зависит от степени увеличения минимального сигнального расстояния между разрешенными кодовыми блоками. В качестве сигнального для канала АБГШ используется расстояние Евклида. Асимптотический энергетический выигрыш определяется формулой[21]:

$$ЭВК[\text{дБ}] = 20\lg\left(\frac{d_{ef}}{d_e}\right), \qquad (9.21)$$

где d_{ef} – минимальное евклидово расстояние между разрешенными кодовыми блоками; d_e – минимальное евклидово расстояние между различными некодированными последовательностями канальных символов одинаковой мощности с кодированными символами.

Согласно (9.21), для получения больших величин энергетического выигрыша при построении СКК необходимо подбирать коды, максимизирующие минимальное евклидово расстояние между разрешенными кодовыми комбинациями.

В основу классификации СКК можно положить отличительные особенности по типам помехоустойчивого кода, по типам ансамблей сигналов и по способам согласования модуляции и кодирования.

По типу помехоустойчивых кодов все СКК могут быть поделены на два

больших класса: СКК на основе блочных кодов и СКК на основе непрерывных кодов. Кроме того, отдельный класс составляют СКК на основе каскадных кодов, в которых применяются одновременно блочные и непрерывные коды. Каждый из классов делится на группы по конкретным видам кода.

Среди блочных наиболее употребимыми являются коды Хэмминга, Голея, БЧХ, Рида–Соломона, Рида–Маллера и др.

Непрерывные коды на практике представлены сверточными кодами, которые обладают дополнительными свойствами линейности, и постоянства во времени.

При использовании сверточного кода практически удобным является случай, когда при объеме ансамбля сигналов $m = 2^{k+1}$ скорость сверточного кода выбирается равной $R_{кк} = k/(k+1)$. Тогда частотная эффективность у системы с кодированием и без него одна и та же. Поскольку каждый кодовый блок длиной $(k+1)$ переносится одним двумерным сигналом, то и СКК считается также двумерной. Декодирование СКК ведется обычно по алгоритму Витерби, реализующему принцип максимального правдоподобия. Одно из важнейших преимуществ СК заключается в простоте применения алгоритма Витерби для мягкого решения на выходе демодулятора.

Любая СКК вне зависимости от способа согласования модуляции и кодирования представляет собой каскадный код с ансамблем сигналов на внутренней ступени и одним или несколькими помехоустойчивыми кодами на внешней. При использовании нескольких помехоустойчивых кодов говорят о построении СКК на основе обобщенного каскадного кода.

По типу ансамблей сигналов СКК делятся на конструкции с одномерными, двумерными и многомерными сигналами.

Многомерные сигналы составляются из более простых (одномерных, двумерных) сигналов. При использовании в качестве составляющих двумерных сигналов число позиций M, соответствующих каждому n–мерному сигналу, определяется выражением $M = m^{n/2}$, где m – позиционность двумерного сигнала.

Каждый n–мерный сигнал в этом случае образуется последовательностью $n/2$ двумерных сигналов. Например, для получения многомерного сигнала с $n=6$ требуется последовательность из трех двумерных сигналов, например ФМн-4.

Способы согласования модуляции и кодирования условно можно разделить на две группы: согласование кодом Грея и согласование на основе разбиения ансамбля на вложенные подансамбли.

Сигнально-кодовые конструкции, принадлежащие первой группе, представляют собой результат согласования известных двоичных помехоустойчивых кодов с многопозиционным ансамблем сигналов путем использования кода Грея в качестве манипуляционного кода. Поскольку ошибки происходят за счет переходов в области соседних сигналов, то кодовые блоки, соответствующие соседним сигналам, должны различаться наименьшим числом двоичных символов.

Вторая группа включает в себя достаточно большое число типов СКК, различающихся модификациями методов согласования. Разбиение осуществляется таким образом, что подансамбли содержат равное количество сигналов, расстояния между соседними сигналами подансамблей одинаковы, а минимальные расстояния между сигналами подансамблей увеличиваются с каждым шагом разбиения. Широкое практическое применение получило согласование путем разбиения ансамбля на вложенные подансамбли, когда внешними кодами являются сверточные коды. В основе синтеза СКК со сверточными кодами лежит поиск кодов, максимизирующих евклидово расстояние, причем обычно эти коды не являются оптимальными в метрике Хэмминга. У решетчатой диаграммы, описывающей сверточные коды в метрике Евклида, переходы между состояниями промаркированы не двоичными блоками, а сигнальными точками.

Таким образом, достижение близких к предельным показателей частотно-энергетической эффективности цифровых систем связи предполагает согласование кодека и модема с учетом статистических свойств непрерывного канала. Одно из решений подобного согласования представляют сигнально-кодовые конструкции сверточного кодирования. Мягкое декодирование по алгоритму

Витерби обеспечивает энергетический выигрыш порядка 3...7 дБ без расширения занимаемом полосы частот.

9.4. Характеристики основных типов СКК

9.4.1. Согласование канала кодом Грея

Рассмотрим СКК, представляющие собой результат согласования известных двоичных помехоустойчивых кодов с многопозиционным ансамблем сигналов путем использования в качестве манипуляционного кода Грея (табл. 9.3).

Комбинации кода в табл. 9.3 получены по следующему правилу. Кодовая комбинация натурального кода складывается по модулю 2 с такой же комбинацией, сдвинутой на один разряд вправо, при этом младший разряд сдвинутой комбинации отбрасывается.

Таблица 9.3

Результат согласования двоичных помехоустойчивых кодов с кодом Грея

Десятичное число	Натуральный двоичный код	Код Грея
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100

Поскольку ошибки чаще происходят за счет переходов в области соседних сигналов, то кодовые блоки, соответствующие соседним сигналам, должны различаться наименьшим числом двоичных символов.

На рис. 9.10 приведены примеры кода Грея для ансамблей одномерных (АМ-4) и двумерных (ФМ-8, КАМ-16) сигналов [5, 21, 32].

Несмотря на достаточно высокие показатели энергетической эффективности при мягком решении в демодуляторе и декодировании алгоритмом Витерби, согласование кодом Грея не является оптимальным.

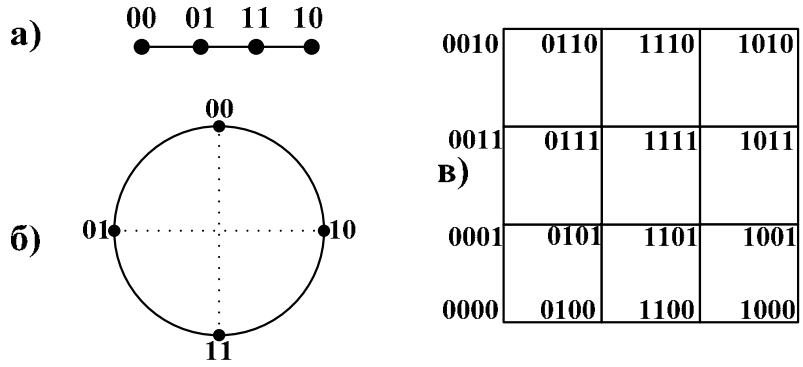


Рис. 9.10. Коды Грея ансамблей: одномерных: а) АМ-4; б) ФМ-4; двумерных: в) КАМ-16

Двоичные коды, оптимальные по критерию максимума хэммингова расстояния, будут оптимальны и по критерию максимума свободного евклидова расстояния, если при отображении двоичных подблоков в сигнальные точки ансамбля выполняется принцип: большему расстоянию Хэмминга $d_{h\max}$, соответствует большее расстояние по Евклиду $d_{e\max}$. Простейшие ансамбли сигналов АМн-2, ФМн-2, ФМн-4 этому условию для кода Грея удовлетворяют.

В табл. 9.4 показаны комбинации (подблоки) двоичного кода длиной $n=3$, а также расстояния d_h и d_e при использовании кода Грея для ФМн-8 (см. рис. 9.11).

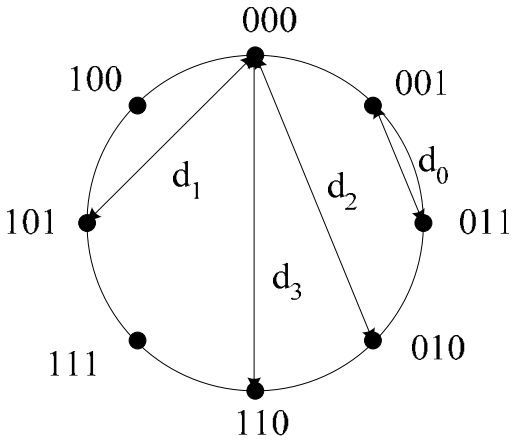


Рис. 9.11. Расстояние между сигнальными точками ФМн-8

Как следует из таблицы, сформулированный принцип соответствия большему расстоянию Хэмминга большего расстояния Евклида для всех вариантов отображения не выполняется. Например, для комбинации 111 большему расстоянию Хэмминга $d_h=3$ соответствует не самое большое расстояние Евклида $d_h=1{,}848$, и т.д.

Таблица 9.4

Соответствие расстояний Хэмминга и Евклида для сигналов ФМн-8

Кодовые комбинации	000	001	011	010	110	111	101	100
d_h	0	1	2	1	2	3	2	1
d_e	0	0,765	1,414	1,848	2,000	1,848	1,414	0,765

Таким образом поскольку манипуляционный код Грея для сложных сигналов не обеспечивает оптимального согласования кодека и модема, необходимо найти методы дальнейшего повышения свободного евклидова расстояния d_{ef} и, соответственно, энергетической эффективности β.

9.4.2. Согласование на основе разбиения ансамбля на вложенные подансамбли

В начале 80 х гг. Унгербоек (Ungerboeck G.) опубликовал статью, в которой, анализируя СКК на базе ансамбля ФМн-8 и сверточного кода со скоростью $R_{кк} = k/(k+1)$, сформулировал ряд правил построения СКК. Поэтому СКК построенные по этим правилам (Trellis-Coded Modulation – TCM), часто называют СКК Унгербоека.

По способу согласования модуляции и кодирования СКК Унгербоека относятся к конструкциям, полученным на основе разбиения ансамбля сигналов на вложенные подансамбли. Разбиение осуществляется таким образом, что подансамбли содержат равное количество сигналов, расстояния d_e между соседними сигналами подансамблей одинаковы, минимальные расстояния $d_{e\min}$ между сигналами подансамблей увеличиваются с каждым шагом разбиения; при этом левая ветвь разбиения кодируется символом «0», а правая «1». Считывание кодовой комбинации, соответствующей сигнальной точке на амплитудно-фазовой плоскости, осуществляется снизу вверх. Разбиение для ансамбля сигналов ФМн-8 представлено на рис. 9.12.

Как следует из рис. 9.12, исходный ансамбль разбивается на подансамбли при максимальном увеличении наименьших расстояний $d_{e\min}$ между сигналами

внутри подансамблей $d_0 < d_1 < d_2 < d_3$. Разбиение осуществляется поэтапно. В данном примере три этапа, заключающихся в разбиении каждого из подансамблей предыдущего этапа на 2 равноэлементных подансамбля.

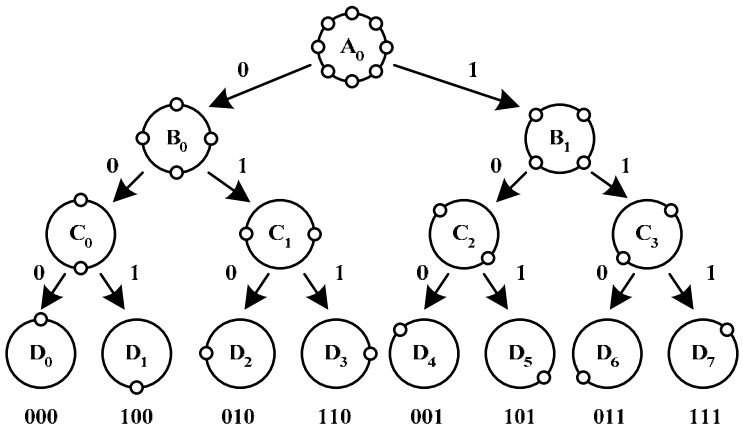


Рис. 9.12. Разбиение ансамбля сигналов ФМн-8

В общем случае количество этапов i полного разбиения ансамбля из m сигналов на вложенные подансамбли определяется выражением:

$$i_{\max} = \log_2 m, \qquad (9.22)$$

т. е. совпадает с кратностью ансамбля n.

В ансамбле из m сигналов кратности n каждой сигнальной точке соответствует блок двоичных символов $b = \left[b^{n-1}, b^{n-2}, ..., b^0\right]$. Соответствие между кодовым блоком b и сигнальной точкой на плоскости определяет манипуляционный код.

Достижение наибольшей помехоустойчивости непосредственно связано с увеличением евклидова расстояния между передаваемыми сигнальными последовательностями. Решетчатая диаграмма сверточного кода (5.6.3), ребра которой промаркированы сигнальными точками, полностью отображает весь набор разрешенных сигнальных последовательностей. Таким образом, величина свободного евклидова расстояния d_{ef} зависит от маркировки ребер решетчатой диаграммы сигнальными точками (канальными символами).

Унгербоек на примере ансамбля сигналов ФМн-8 (см. рис. 9.12) сформулировал четыре необходимых правила маркировки ребер сигнальными точками:

все сигнальные точки используемого ансамбля сигналов должны встречаться с одинаковой частотой и с определенной степенью регулярности и симметричности;

переходы из одного и того же состояния соответствуют сигналам из подансамблей B_0 или B_1;

переходы в одно и то же состояние соответствуют сигналам из подансамблей B_0 или B_1;

параллельные переходы между состояниями соответствуют сигналам из подансамблей C_0 или C_1, или C_2, или C_3.

Как показывает анализ, СКК Унгербоека имеют несколько более высокие частотно энергетические характеристики по сравнению с традиционными СКК, при той же сложности реализации. Это определило их бурное внедрение в технике связи. Но известные правила построения СКК Унгербоека, хотя и снижают размерность переборной задачи синтеза, но не обеспечивают гарантированное построение СКК с максимальными частотно-энергетическими характеристиками. В то же время, основной целью работ в области синтеза систем сигналов и СКК является поиск таких способов их формирования и обработки, которые при заданных ограничениях на сложность устройств формирования и приема, временные задержки, позволяли бы приблизиться к известной шенноновской границе.

При построении многомерных СКК возникает проблема выбора манипуляционного кода, поскольку известные методы его построения (правила построения кодов Грея и разбиения ансамбля на вложенные подансамбли Унгербоека) не всегда позволяют согласовать евклидовы и хеммингговы расстояния. Именно с этим связаны многие проблемы построения многомерных СКК.

Синтез многопозиционных ансамблей сигналов и СКК, построенных на их основе, является одним из направлений решения более общей задачи статистического согласования вероятностных характеристик передаваемого информационного сигнала и вероятностных характеристик канала. В рамках этих традиционных задач, такое согласование осуществляется на уровне канальных символов или их блоков (супербукв канала). При этом подходы к построению алфавита таких супербукв (ансамблей сигналов и СКК) могут существенно отличаться между собой, но направлены на решение этой общей проблемы.

Известно, что ансамбль сигналов, соответствующий полному двоичному коду длины n в пространстве соответствующей размерности n, построенный заменой «1» на «–1», а «0» на «+1», соответственно, обладает практически иде-

альным манипуляционным кодом. Минимальным хеммингговым расстояниям таких ансамблей соответствуют ребра n-мерного куба, которые характеризуются и минимальными евклидовыми расстояниями.

Кодовые комбинации и соответствующие им координаты сигнальных векторов приведены в табл. 9.5; графическое изображение ансамбля представлено на рис. 9.13.

При приеме сигналов такого ансамбля минимальная ошибка (ошибочный прием одной координаты сигнальной точки) приводит к неправильному приему одного бита информации. Ошибочный прием двух координат сигнальной точки приводит к искажению двух бит информации и так далее. Однако, если рассмотреть зависимость между хеммингговыми $d_h(i,j)$ и евклидовыми $d_e(i,j)$ расстояниями для такого ансамбля, то можно выявить следующую закономерность, связывающую эти две величины [5, 32]:

$$d_e(i,j) = 2r\sqrt{d_h(i,j)}, \qquad (9.23)$$

где r – радиус сферы.

Таблица 9.5

Взаимосвязь кодовых комбинаций манипуляционного кода и координат сигнальных векторов

Манипуляционный код	Координаты сигнальных векторов
000	+1,+1,+1
001	+1,+1,–1
010	+1,–1+1
011	+1,–1,–1
100	–1,+1,+1
101	–1,+1,–1
110	–1,–1,+1
111	–1,–1,–1

Таким образом, взаимосвязь между евклидовыми и хеммингговыми расстояниями в многомерном ансамбле сигналов нелинейная, хотя большему хеммингговому расстоянию будет соответствовать большее евклидово расстояние.

Если мощность и энергия сигналов являются постоянными величинами, не зависящими от номера, то ансамбли таких сигналов считают сигналами поверхностно-сферической упаковки.

В противном случае ансамбли сигналов рассматривают как объемные упаковки. Сохранение манипуляционного кода, принятого для простого трехмерного куба, в значительной мере сохраняет пропорциональность между евклидовыми и хеммингловыми расстояниями и поэтому будет наилучшим и для наиболее плотного ансамбля. Для других комбинаций манипуляционных кодов для сигнальных векторов изначально не будет соблюдаться взаимная пропорциональность между евклидовыми и хеммингловыми расстояниями.

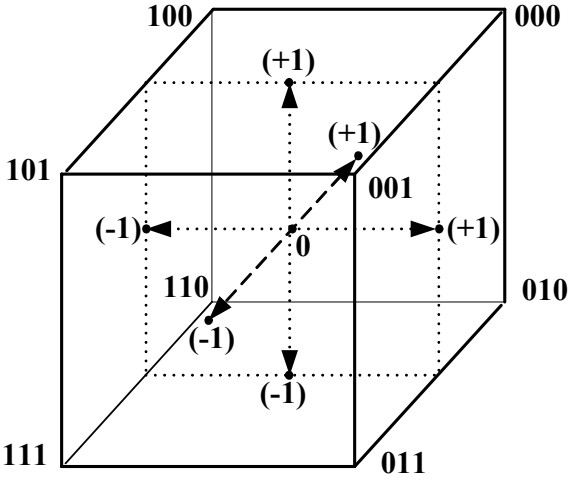


Рис. 9.13. Поверхностно-сферическая укладка обычного куба для ФМн-8

Таким образом, практически невозможно создать идеальный манипуляционный код и, следовательно, целесообразно строить манипуляционные коды, у которых хотя бы частично выполняется взаимосвязь между евклидовыми и хеммингловыми расстояниями.

9.5. Алгоритмы цифровой обработки сигналов

9.5.1. Дискретные сигналы и их спектры

Дискретизация непрерывного сигнала. С аналитической точки зрения процедуру получения дискретизированного (дискретного) сигнала $u_T(t)$ удобно рассматривать как непосредственное умножение непрерывного сигнала $u(t)$ на вспомогательную последовательность $y(t)$ дискретизирующих прямоугольных импульсов единичной амплитуды

$$u_T(t) = u(t) \cdot y(t). \qquad (9.24)$$

Длительность дискретизирующих импульсов $\tau_и$ должна быть много меньше интервала дискретизации Δt.

Принцип формирования дискретного сигнала показан на рис. 7.7, б…в, где изображены графики функций $u(t)$ $u_T(t)$ $y(t)$. При этом реальный дискретный сигнал $u_T(t)$ имеет вид импульсно-модулированного колебания, т. е. АИМ-сигнала.

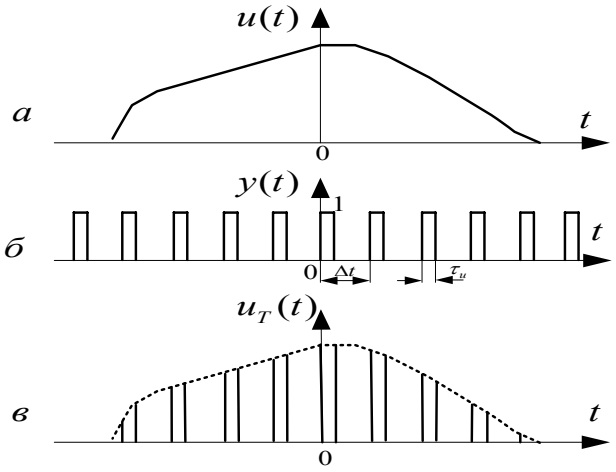


Рис.9.14. Дискретизация сигналов

Спектр дискретного сигнала. Чтобы дать оценку требованиям к длительности дискретизирующих импульсов, определим спектральный состав дискретного сигнала $u_T(t)$. Пусть некоторый непрерывный сигнал $u(t)$ имеет спектральную плотность $S(\omega)$. Представим последовательность дискретизирующих прямоугольных импульсов $y(t)$ рядом Фурье, в котором частота $\omega_1 = 2\pi/\Delta t$:

$$y(t) = \frac{\tau_и}{\Delta t}\left(1 + 2\sum_{n=1}^{\infty} S_n \cos n\omega_1 t\right), \qquad (9.25)$$

где

$$S_n = \frac{\sin(n\omega_1 \tau_и/2)}{n\omega_1 \tau_и/2}. \qquad (9.26)$$

Подставив формулу (9.25) в (9.24), получим

$$u_T(t) = \frac{\tau_и}{\Delta t}u(t) + 2\frac{\tau_и}{\Delta t}\sum_{n=1}^{\infty} S_n u(t)\cos n\omega_1 t. \qquad (9.27)$$

Проанализируем первое и второе слагаемые этого выражения отдельно. Первому слагаемому соответствует спектральная плотность $S(\omega)$ исходного сигнала $u(t)$. К произведению $u(t)\cos n\omega_1 t$ второго слагаемого применим прямое преобразование Фурье. Используя формулу Эйлера и проведя несложные математические выкладки, запишем

$$S(j\omega) = S(\omega) = \int_{-\infty}^{\infty} u(t)\cos n\omega_1 t \cdot e^{-j\omega t}dt = \frac{1}{2}\int_{-\infty}^{\infty} u(t)\cdot e^{-j(\omega - n\omega_1)t}dt + \frac{1}{2}\int_{-\infty}^{\infty} u(t)\cdot e^{-j(\omega + n\omega_1)t}dt$$

В этом выражении первый интеграл представляет собой спектральную плотность сигнала $u(t)$ на частотах $\omega - n\omega_1$ а второй - ту же спектральную

плотность, но на частотах $\omega + n\omega_1$. Поэтому

$$\int_{-\infty}^{\infty} u(t)\cos n\omega_1 t \cdot e^{-j\omega t} dt = \frac{1}{2}\left[S(\omega - n\omega_1) + S(\omega + n\omega_1)\right]. \qquad (9.28)$$

Следовательно, дискретному сигналу вида (9.27) соответствует спектральная плотность

$$S_T(\omega) = \frac{\tau_и}{\Delta t}\left[S(\omega) + \sum_{n=1}^{\infty} S_n S(\omega - n\omega_1) + \sum_{n=1}^{\infty} S_n S(\omega + n\omega_1)\right]. \qquad (9.29)$$

Поскольку при $n = 0$ коэффициент $S_n = 1$, запишем

$$S_T(\omega) = \frac{\tau_и}{\Delta t} \cdot \sum_{n=-\infty}^{\infty} S_n S(\omega - n\omega_1) = \frac{\tau_и}{\Delta t} \cdot \sum_{n=-\infty}^{\infty} \frac{\sin(n\omega_1 \tau_и/2)}{n\omega_1 \tau_и/2} S(\omega - n\omega_1). \qquad (9.30)$$

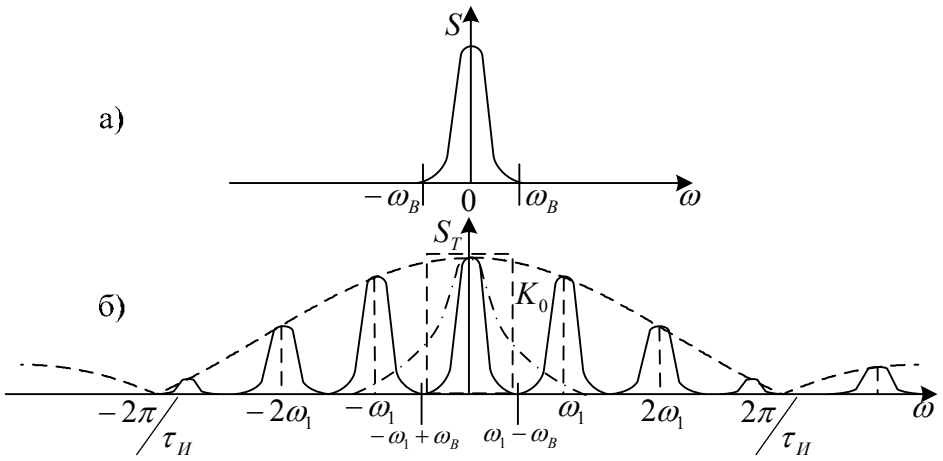


Рис.9.15. Спектры сигналов: а - непрерывного; б - дискретного

График спектра дискретного сигнала, полученного из непрерывного, показан на рис.9.15,б.

Полученные результаты позволяют сделать следующие выводы:

спектральная плотность $S_T(\omega)$ дискретного сигнала $u_T(t)$ представляет собой бесконечную последовательность спектральных плотностей $S(\omega)$ исходного непрерывного сигнала $u(t)$, сдвинутых друг относительно друга на частоту ω_1;

огибающая спектральной плотности $S_T(\omega)$ дискретного сигнала $u_T(t)$ с точностью до коэффициента $1/\Delta t$ повторяет огибающую спектральной плотности дискретизирующего прямоугольного импульса.

Чтобы восстановить непрерывный сигнал $u(t)$ из дискретного $u_T(t)$, достаточно выделить центральную часть спектра $S_T(\omega)$. На практике это осуществляют с помощью идеального ФНЧ, имеющим коэффициент передачи

$$K(\omega) = K_0, \qquad -\omega_в \le \omega \le \omega_в. \tag{9.31}$$

Вместе с тем известно, что идеальный ФНЧ физически нереализуем и может служить лишь теоретической моделью для пояснения принципа восстановления непрерывного сигнала на основе теоремы Котельникова. Реальный ФНЧ имеет частотную характеристику, которая либо охватывает несколько лепестков спектра (штрих - пунктирная линия на рис. 9.15,б), либо имеет конечную крутизну ската характеристики и не полностью охватывает центральный лепесток. В практических схемах интервал дискретизации, определяемый формулой $\Delta t/2F_в$, уменьшают в 2...5 раз. В этом случае отдельные составляющие спектра дискретного сигнала не перекрываются, как это и показано на рис. 9.15, б, и могут быть разделены фильтрами.

При уменьшении длительности дискретизирующего импульса $\tau_и$, амплитуды спектральных составляющих с ростом частоты убывают медленнее. В предельном случае, при $\tau_и \to 0$ спектр дискретного сигнала будет представлять собой бесконечную последовательность «копий» спектров исходного сигнала, имеющих равную амплитуду. Если одновременно с уменьшением длительности увеличивать амплитуду импульса так, чтобы его площадь оставалась неизменной и равной единице, то дискретизирующие сигналы преобразуются в последовательность дельта-функций:

$$y(t) = \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t).$$

В этом случае формула (9.24) запишется следующим образом:

$$u_T(t) = u(t) \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t) = u(k\Delta t) \sum_{k=-\infty}^{\infty} u(k\Delta t)\delta(t - k\Delta t). \tag{9.32}$$

Спектральная плотность дискретного сигнала в этом случае примет вид:

$$S_T(\omega) = \frac{1}{\Delta \mathrm{t}} \cdot \sum_{n=-\infty}^{\infty} S(\omega - n\omega_1). \tag{9.33}$$

Пример 9.1. Непрерывный сигнал, имеющий форму прямоугольного импульса напряжения с единичной амплитудой и длительностью $\tau_и$, дискретизирован 10 отсчетами. Определить спектр дискретного сигнала.

Решение. Для нахождения спектра воспользуемся формулой (9.33). В ней частота $\omega_1 = 2\pi/\Delta t = 20\pi/\tau_и$, а интервал дискретизации $\Delta t = \tau_и/10$. Тогда

$$S_T(\omega)=\frac{10}{\Delta t}\cdot\sum_{n=-\infty}^{\infty}\tau_{и}\frac{\sin(\frac{\omega\tau_{и}}{2}-\frac{20n\pi}{\tau_{и}}\cdot\frac{\tau_{и}}{2})}{\frac{\omega\tau_{и}}{2}-\frac{20n\pi}{\tau_{и}}\cdot\frac{\tau_{и}}{2}}=10\cdot\sum_{n=-\infty}^{\infty}\frac{\sin(\frac{\omega\tau_{и}}{2}-10n\pi)}{\frac{\omega\tau_{и}}{2}-10n\pi}.$$

Возможность представления дискретных сигналов $u_T(t)$ в форме (9.32) существенно упрощает их анализ. В частности, спектральную плотность $S_T(\omega)$ можно вычислить непосредственно по совокупности временных отсчетов $\{u(k\Delta t)\}$. Действительно, применив прямое преобразование Фурье $S(j\omega)=S(\omega)=\int_{-\infty}^{\infty}u(t)e^{-j\omega t}dt$ к соотношению (9.32) для отсчетов только с положительными номерами $k=0,1...,\infty$, со, получим с учетом фильтрующего свойства дельта-функции:

$$S_T(\omega)=\int_0^{\infty}\sum_{k=0}^{\infty}u(k\Delta t)\cdot e^{-j\omega t}\delta(t-k\Delta t)dt=\sum_{k=0}^{\infty}u(k\Delta t)e^{-j\omega k\Delta t}. \qquad (9.34)$$

При этом существенно сокращается время обработки реальных непрерывных сигналов.

9.5.2. Алгоритмы дискретного и быстрого преобразований Фурье

Как и при анализе аналоговых сигналов, дискретные сигналы можно представить во временной и частотной областях. В настоящее время обработку дискретных сигналов чаще всего проводят в частотной области, что диктуется значительными сокращениями объема цифровой аппаратуры и времени обработки.

Пусть дискретной обработке подвергается аналоговый импульсный сигнал $u(t)$ длительностью $\mathrm{T_и}$, имеющий спектральную плотность $S(\omega)$ (рис. 9.16, а, б). Теоретически можно предположить, что дискретизация сигнала производится периодической последовательностью дельта-функций

$$y(t)=\sum_{k=0}^{N-1}\delta(t-k\Delta t), \qquad (9.35)$$

где $N=T_{и}/\Delta t$ — требуемое число отсчетов, отвечающих теореме Котельникова.

Подставив в (9.32) пределы суммирования от 0 до $N-1$, и заменив здесь и далее для упрощения и уменьшения объема формул $u(k\Delta t)=u_k$, запишем выражение для дискретного сигнала (рис. 9.16, е)

$$u_T(t) = u(t)\sum_{k=0}^{N-1}\delta(t-k\Delta t) = \sum_{k=0}^{N-1}u_k\cdot\delta(t-k\Delta t)\,. \qquad (9.36)$$

На основании формулы (9.36) можно сделать вывод, что спектр данного дискретного сигнала имеет периодическую структуру с периодом по оси частот $\omega_1 = 2\pi/\Delta t$ (рис. . 9.16, г). Мысленно продолжим дискретный сигнал периодически с интервалом $T_и$ (рис. . 9.16, д). $|C_n|$

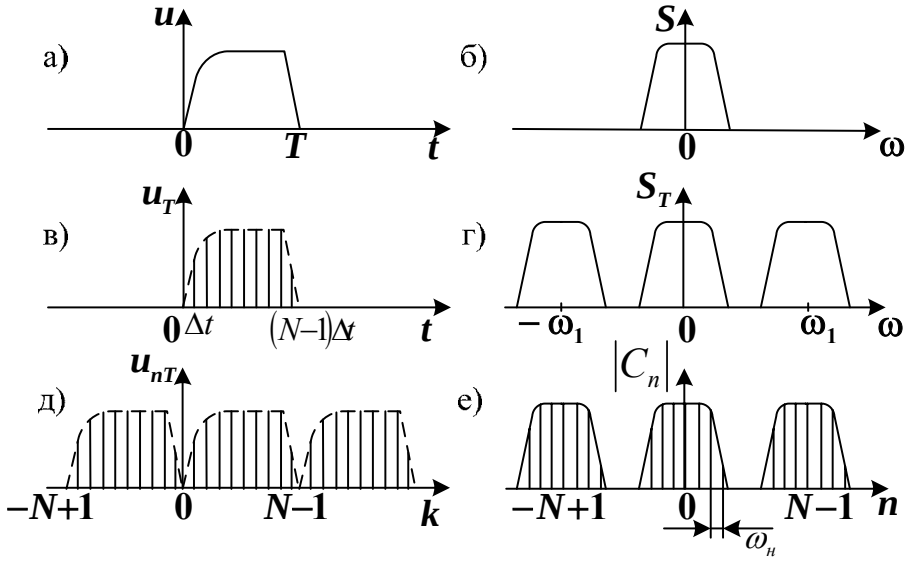


Рис. 9.16. Графики к выводу ДПФ:

а,б - аналоговый сигнал и его спектр; *в,г* - дискретный сигнал и его спектр;

д - периодическая последовательность дискретного сигнала; *е* - ДПФ сигнала

$$u_{nT}(t + nT_и) = u_T(t)\,,\ n = 0,\pm1,\pm2,\ldots.$$

По аналогии с представлением периодических непрерывных сигналов $u(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_1 t}$, где $C_n = \frac{1}{T}\int_{-T/2}^{T/2} u(t)\cdot e^{jn\omega_1 t}dt$ - комплексная амплитуда n -й гармоники. Дискретную функцию $u_{nT}(t)$ можно разложить в комплексный ряд Фурье:

$$u_{nT}(t) = \sum_{n=-\infty}^{\infty} C_n e^{jn\omega_н t}\,, \qquad (9.37)$$

где $\omega_н = 2\pi/T_u = 2\pi/(N\cdot\Delta t)$ - частота дискретизации сигнала.

Коэффициенты этого ряда

$$C_n = \frac{1}{T_и}\int_0^{T_и} u_T(t)\cdot e^{jn\omega_н t}dt = \frac{1}{T_и}\int_0^{T_и} u_T(t)\cdot e^{j2\pi nt/T_u}dt\,. \qquad (9.38)$$

Для определения коэффициентов проделаем следующее. Подставим фор-

мулу (9.36) в (9.38) и заменим параметр $T_u = N \cdot \Delta t$. Введем безразмерную переменную $y = t/\Delta t$ и запишем

$$C_n = \frac{1}{T_и}\int_0^{T_и}\sum_{k=0}^{N-1} u_k \delta(t - k\Delta t)\cdot e^{-j2\pi nt/T_u}\,dt = \frac{1}{N}\sum_{k=0}^{N-1} u_k \int_0^{N}\sum_{k=0}^{N-1}\delta(y-k)\cdot e^{-j2\pi ny/N}dy\,.$$

Используя фильтрующее свойство дельта – функции, находим

$$C_n = \frac{1}{N}\sum_{k=0}^{N-1} u_k \cdot e^{-j2\pi nk/N}\,. \qquad (9.39)$$

Это называется дискретным преобразованием Фурье (ДПФ). Дискретное преобразование Фурье по существу представляет собой алгоритм вычисления гармонических составляющих спектра C_n по заданным дискретным отсчетам u_k аналогового сигнала $u(t)$, что значительно сокращает время обработки. Характерный вид модулей коэффициентов C_n показан на рис. 9.16,е.

Следует отметить ряд свойств ДПФ, которые вытекают из определения (9.39).

1. Дискретное преобразование Фурье обладает свойством линейности: линейной комбинации дискретных сигналов соответствует линейная комбинация их ДПФ.

2. Коэффициент C_0 представляет собой среднее значение (постоянную составляющую) всех дискретных отсчетов сигнала

$$C_0 = \frac{1}{N}\sum_{k=0}^{N-1} u_k\,.$$

3. Число различных коэффициентов C_n равно числу отсчетов N за длительность сигнала $T_и$; при $n = N$ коэффициент $C_n = C_0$.

Пример 9.2. Определить коэффициенты ДПФ дискретизированного прямоугольного импульса единичной амплитуды, заданного четырьмя отсчетами $(N = 4)$.

Решение. Используя основную формулу (9.39), вычислим пять первых коэффициентов ДПФ: $C_0 = 4/4 = 1$;

$$C_1 = \frac{1}{4}\sum_{k=0}^{N-1}(1 + e^{-j\pi/2} + e^{-j\pi} + e^{-j3\pi/2}) = 0;\ C_2 = \frac{1}{4}\sum_{k=0}^{N-1}(1 + e^{-j\pi} + e^{-j2\pi} + e^{-j3\pi}) = 0$$

$$C_3 = \frac{1}{4}\sum_{k=0}^{N-1}(1 + e^{-j3\pi/2} + e^{-j3\pi} + e^{-j9\pi/2}) = 0;\ C_4 = \frac{1}{4}\sum_{k=0}^{N-1}(1 + e^{-j2\pi} + e^{-j4\pi} + e^{-j6\pi}) = 1\,.$$

При изучении теории ДПФ возникает очевидный вопрос: можно ли по

известным коэффициентам ДПФ вычислить отсчетные значения u_k непрерывного сигнала? По аналогии с периодическими сигналами представим заданную периодическую последовательность отсчетов комплексным рядом Фурье. Заменив в (7.25) $t = k\Delta t$, $\omega_н = 2\pi/(N \cdot \Delta t)$ и, учитывая, что суммируется конечное число членов ряда, запишем

$$u_k = \sum_{n=0}^{N-1} C_n e^{j2\pi nk/N} . \qquad (9.40)$$

Данное соотношение определяет алгоритм обратного дискретного преобразования Фурье (ОДПФ). Формулы (9.39) и (9.40) являются аналогами прямого и обратного преобразований Фурье для непрерывных сигналов.

Выражение (9.39) показывает, что для определения одного коэффициента ДПФ сигнальной последовательности из N отсчетов, необходимо выполнить около N операций умножения на комплексное число и столько же сложений, а для нахождения всех коэффициентов объем вычислений составит N^2. В частности, при $N = 2^{10} = 1024$ надо осуществить более миллиона (1024^2) умножений и сложений. Если длины обрабатываемых массивов превышают тысячу единиц, то дискретная спектральная обработка сигналов в реальном масштабе времени требует высокопроизводительных вычислительных комплексов.

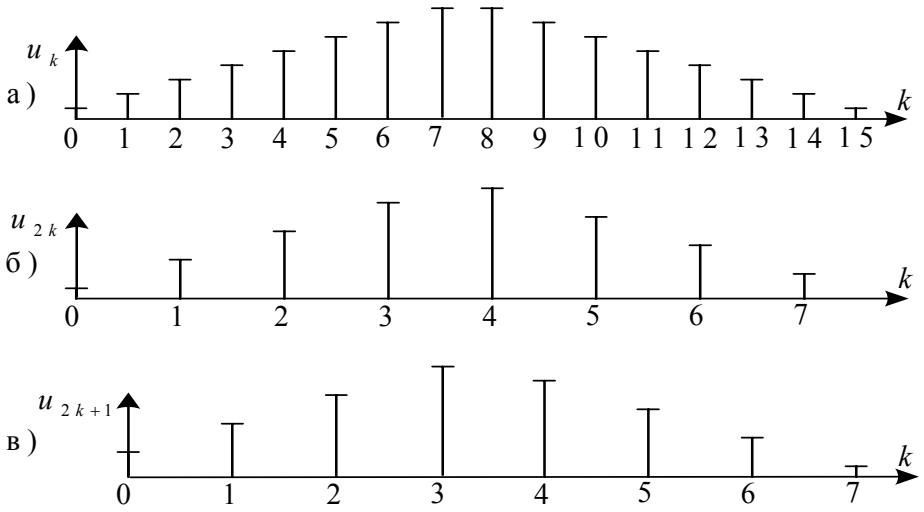


Рис.9.17. Разбиение последовательности u_k на две подпоследовательности: *а* - входная; *б* - с четными номерами; *в* - с нечетными номерами

Многократно сократить число операций позволяет быстрое преобразование Фурье (БПФ), обеспечивающее вычисление коэффициентов ДПФ за меньшее число операций. В основу БПФ положен принцип разбиения заданной по-

следовательности отсчетов дискретного сигнала на несколько промежуточных последовательностей. Для этого число отсчетов N разделяется на множители (например, $N = 8 = 2 \cdot 2 \cdot 2, N = 60 = 3 \cdot 4 \cdot 5$). Затем определяются спектры этих промежуточных последовательностей и через них находится спектр всего сигнала. В зависимости от состава, числа и порядка следования указанных множеств можно создать различные алгоритмы БПФ. В цифровой технике удобно обрабатывать сигнальные последовательности со значениями N, являющимся степенью числа два (4, 8, 16 и так далее). Это позволяет многократно делить входную последовательность отсчетов на подпоследовательности.

Пусть требуется вычислить ДПФ дискретного сигнала $\{u(k\Delta t)\} = \{u_k\}$, имеющего четное число отсчетов (рис. 9.17, а), причем $N = 2^r$; r - целое число.

Представим входную последовательность в виде двух подпоследовательностей с четными и нечетными номерами и половинным числом членов в каждой (рис. 9.17, б,в): $u_{\text{чт}} = u_{2k}$; $u_{\text{нч}} = u_{2k+1}; k = 0,1,2,..., N/2-1$.

Коэффициенты ДПФ для последовательностей с четными и нечетными номерами запишем отдельно:

$$\frac{1}{N}\sum_{k=0}^{N/2-1} u_{2k} \cdot e^{-j2\pi n 2k/N} = \frac{1}{N}\sum_{k=0}^{N/2-1} u_{2k} \cdot e^{\frac{-j2\pi nk}{N/2}} = C_{n\text{чт}}$$

$$\frac{1}{N}\sum_{k=0}^{N/2-1} u_{2k+1} \cdot e^{-j2\pi n(2k+1)/N} = \frac{1}{N} e^{-j2\pi n/N} \sum_{k=0}^{N/2-1} u_{2k+1} \cdot e^{\frac{-j2\pi nk}{N/2}} = e^{-j2\pi n/N} C_{n\text{нч}}. \qquad (9.41)$$

Коэффициенты C_n результирующего ДПФ входной последовательности можно выразить через параметры $C_{n\text{чт}}$ и $C_{n\text{нч}}$ двух вновь введенных подпоследовательностей. Анализ (9.41) показывает, что в диапазоне номеров отсчетов от 0 до $N/2-1$, ДПФ входной последовательности определяется соотношением:

$$C_n = C_{n\text{чт}} + e^{-j2\pi n/N} C_{n\text{нч}},\ n = 0,1,2,..., N/2-1. \qquad (9.42)$$

Так как ДПФ четной и нечетной последовательностей являются периодическими, с периодом $N/2$, то $C_{n\text{чт}} = C_{(n+N/2)\text{чт}}$; $C_{n\text{нч}} = C_{(n+N/2)\text{нч}}$.

Запишем экспоненциальный множитель в формуле (9.42) при $n \geq N/2$, т.е. для ДПФ $C_{(N/2+n)\text{нч}}$, в виде:

$$e^{-j\frac{2\pi(N/2+n)}{N}} = e^{-j\pi} \cdot e^{-j\frac{2\pi n}{N}} = -e^{-j\frac{2\pi n}{N}}$$

С учетом двух последних выражений находим коэффициенты ДПФ вход-

ной последовательности для отсчетов с номерами от $N/2$ до $N-1$:

$$C_{N/2+n} = C_{n\text{чт}} - e^{-j2\pi n/N} C_{n\text{нч}},\ n = 0,1,2,\ldots, N/2-1. \qquad (9.43)$$

Соотношения (9.42) и (9.43) полностью определяют алгоритмы вычисления коэффициентов с помощью БПФ. Отметим, что экспоненциальные фазовые множители $e^{-j2\pi n/N}$ в этих алгоритмах учитывают влияние сдвига нечетной подпоследовательности относительно четной.

Чтобы еще уменьшить число вычислений, четную и нечетную подпоследовательности также разбивают каждую на две промежуточные части. Разбиение продолжают вплоть до получения простейших двухэлементных последовательностей. Определив ДПФ данных простейших пар отсчетов, можно вычислить ДПФ четырехэлементных, восьмиэлементных и так далее подпоследовательностей. При объединении ДПФ четной и нечетной подпоследовательностей используют алгоритмы (9.42) и (9.43), подставляя в них соответствующие значения номеров N и n.

Нетрудно заметить, что вычисления по формулам (9.41) не потребуют операций умножения, в (9.41) имеются только сложение и вычитание комплексных чисел. Учитываться же должны лишь операции умножения в алгоритмах (9.42) и (9.43) для различных n при разбиениях массива отчетов на мелкие подпоследовательности. Число этих операций при первом разбиении составляло $N/2$. Такое же число $N/2$ операций требуется выполнить при каждом следующем разбиении. Таким образом, вдвое увеличивается число подпоследовательностей и вдвое сокращается наибольшее число n в формулах (7.30), (7.31).

Вычисление коэффициентов ДПФ последовательности из N отсчетов по алгоритмам БПФ требует примерно $N\log_2 N$ операций умножения. Алгоритмы БПФ сокращают число операций по сравнению с алгоритмами ДПФ в $N^2/(N\log_2 N) = N\log_2 N$ раз. Например, при количестве отсчетов $N = 2^{10}$, имеем $\log_2 N = 10$ и сокращение числа операций составляет $N\log_2 N \approx 100$. При очень больших массивах отсчетов входного сигнала выигрыш в скорости обработки может достигать нескольких тысяч.

Таким образом, в алгоритмах БПФ выполняются операции сложения и вычитания с умножением одного из компонентов на экспоненциальный множитель $e^{-j2\pi n/N}$. Эту базовую для БПФ операцию очень удобно представлять сиг-

нальным графом, называемым в цифровой технике «бабочкой».

БПФ по рассмотренному методу (его называют *методом прореживания* отсчетов во времени) осуществляют, как правило, в следующем порядке. Сначала для получения желательного при обработке сигнала порядка следования отсчетов $u(k), k = 0,1,2,..., N-1$, выполняется двоично-инверсная перестановка элементов исходной последовательности $u(l)$, $l = 0,1,2,..., N-1$. Для этого записывают порядковые номера элементов $u(l)$ в двоичном коде и инвертируют порядок следования разрядов. Новый порядок следования элементов $u(k)$ определяется номерами, полученными после инверсии разрядов.

Пример при *N=4*

$u(l)$			$u(k)$
0→	00→	00→	0→
1→	01→	10→	2→
2→	10→	01→	1→
3→	11→	11→	3→

Новый порядок следования элементов: $u(0), u(2), u(1), u(3)$. После этого поступают так. На первом этапе вычислений определяют двух точечные ДПФ "новой" последовательности $u(k)$, объединяя попарно элементы этой последовательности. На втором этапе из двух точечных ДПФ получают четырех точечные ДПФ, пользуясь основной базовой операцией данного метода (см. ниже). Затем четырех точечные ДПФ объединяют в восьми точечные и т.д.

Базовые операции $X = A + W_N^k B$ и $Y = A - W_N^k B$ показывают, как два входных числа *A* и *B* объединяются для получения двух выходных чисел X и Y. Для метода прореживания во времени базовая операция изображается «бабочкой», представленной на рис. 9.18. Надпись W_N^k у стрелки, идущей вверх, означает умножение W_N^k на величину *B*.

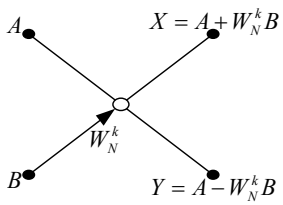


Рис. 9.18. Операция «бабочка», используемая
при реализации алгоритма БПФ

При вычислении двух точечного ДПФ $k = 0$ и выходные числа X и Y определяются без операции умножения $X = A + B$, $Y = A - B$.

Пример 9.3. Построим граф вычисления БДНФ с прореживанием во времени для *N*=4 (рис. 9.19).

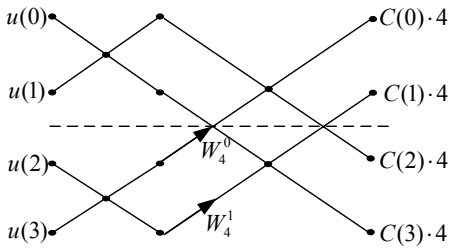


Рис. 9.19. Граф для вычисления БПФ при N=4

Учитывая, что $W_4^0 = 1$, $W_N^1 = e^{-j\frac{\pi}{2}} = -j$, получаем согласно приведенному графу

$$4 \cdot C(0) = u(0) + u(2) + u(1) + u(3)$$
$$4 \cdot C(1) = u(0) - u(2) - j \cdot u(1) - u(3)$$
$$4 \cdot C(2) = u(0) + u(2) - u(1) + u(3)$$
$$4 \cdot C(3) = u(0) - u(2) + j \cdot u(1) - u(3)$$

На рис. 9.20 показан граф вычисления БДПФ с прореживанием во времени для N=8.

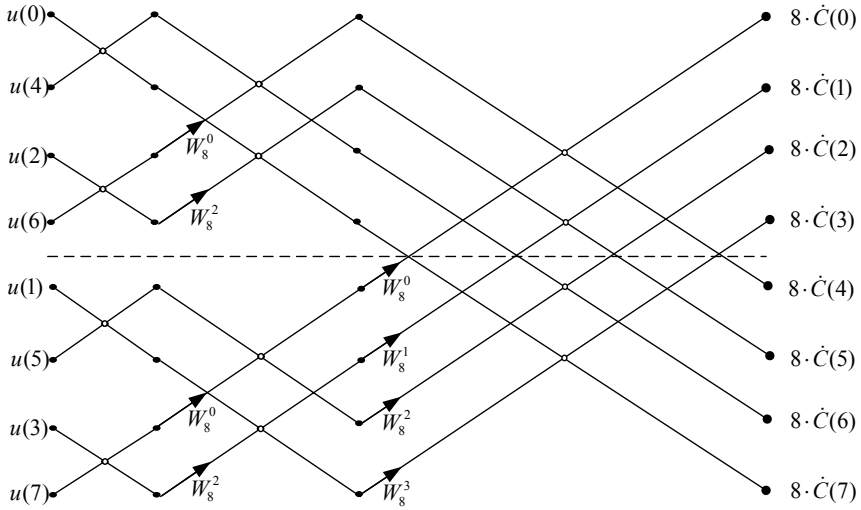


Рис. 9.20. Граф для вычисления БПФ при N=8

Контрольные вопросы

1. Что называют эффективностью СЭС и как она определяется количественно?

2. Как можно осуществить обмен эффективности между β и γ?

3. Докажите, что предельные показатели эффективности $\beta_{\max}$ и $\gamma_{\max}$ для двоичного канала с ортогональными сигналами на 3 дБ меньше, чем для канала с противоположными сигналами.

4. Сравните эффективность систем с ФМн, ЧМн и АМн двоичными сигналами.

5. Как изменится энергетическая эффективность системы передачи дискретных сообщений при уменьшении требуемой вероятности ошибки? Уменьшится она или увеличится? Покажите это на примере двоичного канала.

6. Какой общий вывод о связи между помехоустойчивостью и эффективностью можно сделать для систем с ФМн, ЧМн и АМн двоичными сигналами?

7. На какой основе могут быть построены системы, в которых достигается повышение как энергетической, так и частотной эффективности?

ГЛАВА 10. ТЕОРЕТИКО-ИНФОРМАЦИОННЫЕ ОСНОВЫ КРИПТОЗАЩИТЫ СООБЩЕНИЙ В ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ

10.1. Классификация криптографических систем

Криптология занимается развитием двух противоположных по задачам направлений: криптографии и криптоанализа.

Криптография занимается исследованием методов защиты информации и анализом их эффективности. 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 (лишение противника возможности извлечь информацию из канала связи) и проблему целостности (лишение противника возможности изменить сообщение так, чтобы изменился его смысл, или ввести ложную информацию в канал связи).

Проблемы конфиденциальности и целостности информации тесно связаны между собой, поэтому методы решения одной из них часто применимы для решения другой.

Криптография дает возможность преобразовать информацию таким образом, что ее прочтение (восстановление) возможно только при знании ключа. Под криптографическим ключом подразумевается некоторая секретная информация, известная законному собственнику (пользователю) информации и неизвестная нарушителю.

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. Обычно криптосистема представляет собой совокупность аппаратных и программных средств, которую можно изменить только при значительных затратах времени и средств, тогда как ключ является легко изменяемым объектом. Именно поэтому стойкость криптосистемы определяется только секретностью ключа. Другое почти общепринятое допущение в криптоанализе состоит в том, что криптоаналитик имеет в своем распоряжении шифротексты сообщений [1, 19, 31, 36].

10.1.1. Основные определения и понятия теории криптографической защиты

Криптографическим алгоритмом защиты информации называется последовательность действий, обеспечивающая преобразование защищаемой информации по правилу, заданному ключом.

Криптоалгоритмы могут выполнять шифрование (дешифрование) сообщений, формирование и проверку аутентификаторов сообщений, генерацию ключей, необходимых для выполнения других криптоалгоритмов, и т.п.

Криптоаналитик, т.е. сторона, использующая методы криптоанализа, стремится противодействовать алгоритмам защиты информации.

Нарушение (взлом) криптографического алгоритма – событие, при котором нарушитель, не знающий секретной ключевой информации криптоалгоритма, способен систематически срывать цель защиты информации. Например, криптографический алгоритм шифрования сообщений считается взломанным, если не знающий секретного ключа дешифрования нарушитель способен сис-

тематически восстанавливать сообщения из криптограмм, затрачивая на это некоторое допустимое для него время.

Криптографический протокол зашиты информации – совокупность используемых криптографических алгоритмов и правил их выполнения, определяющих порядок взаимодействия участников информационного обмена для достижения определенной цели защиты информации.

Криптографические протоколы играют важную роль в криптографии, так как ориентированы на обеспечение безопасного взаимодействия участников информационного обмена. Многочисленные примеры нарушения безопасности информации в современных защищенных информационных системах чаще всего относятся к различным нарушениям (взлому) криптопротоколов.

Нарушение (взлом) криптографического протокола - событие, при котором нарушитель, не обладающий возможностью взломать непосредственно сами используемые в протоколе криптоалгоритмы, используя слабости протокола, способен сорвать цель защиты информации.

Например, криптопротокол аутентификации пользователей информационной системы будет взломан независимо от использованных алгоритмов формирования и проверки, если пользователи будут выбирать себе аутентификаторы (пароли) из ограниченного множества легко угадываемых слов и имен. В ряде руководств по защите информации в ЭВМ отмечается, что всего несколько сотен паролей составляют более 90% всех используемых паролей пользователей. Поэтому нарушителю несложно перебрать наиболее «популярные» варианты паролей, чтобы с высокой вероятностью осуществить успешный несанкционированный доступ к ресурсам и услугам информационной системы.

Криптографическая система зашиты информации – совокупность используемых криптографических алгоритмов, протоколов и процедур формирования, распределения, передачи и использования криптографических ключей.

Для описания криптосистемы как математической модели удобно использовать также следующее определение: криптографическая система зашиты информации есть полное множество криптопреобразований, используемых отпра-

вителем и получателем сообщений. В контексте этого определения ключи можно рассматривать как выбранные из множества возможных для использованных криптоалгоритмов конкретных пар криптопреобразований, выполняемых участниками информационного обмена.

Криптографической стойкостью системы защиты информации называется ее способность противостоять атакам противоборствующей стороны. Стойкость криптографической системы является ее главной характеристикой и может рассматриваться и оцениваться с различных точек зрения.

10.1.2. Классификация криптографических систем защиты информации

В соответствии с выполняемыми задачами по защите информации можно выделить два основных класса криптографических систем [31, 36]:

криптосистемы, обеспечивающие секретность информации;

криптосистемы, обеспечивающие подлинность (аутентичность) информации.

Такое разделение обусловлено тем, что задача защиты секретности информации (сохранения ее в тайне) принципиально отличается от задачи защиты подлинности (аутентичности) информации, а поэтому должна решаться другими криптографическими методами.

Классификация криптосистем в соответствии с выполняемыми ими задачами по защите информации представлена на рис. 10.1.

Криптосистемы, обеспечивающие секретность информации, разделяются на системы шифрования и системы криптографического кодирования информации.

Системы шифрования информации исторически являются самыми первыми криптографическими системами. Например, в одном из первых трудов о военном искусстве, принадлежащим перу Энея Тактикуса, в главе «О секретных сообщениях», описывались принципы построения и использования в древней Спарте (IV век до нашей эры) средств шифрования информации. Спартан-

цы для передачи сообщений с театров военных действий использовали так называемую скиталу – механический шифратор в виде цилиндра. При шифровании сообщение записывалось побуквенно на узкую ленту, намотанную на скиталу, вдоль образующей этого цилиндра. После этого лента разматывалась и в промежутках дописывались произвольные буквы. Неизвестным для противоборствующей стороны ключом являлся диаметр скиталы. Интересно отметить, что первое дошедшее до нас имя криптоаналитика также связано со скиталой: Аристотель предложил перехваченную ленту с зашифрованным сообщением наматывать на конус, и то место, где появлялось осмысленная фраза, определяло неизвестный диаметр скиталы (ключ системы шифрования).

10.1. Классификация криптографических систем защиты информации

В общем случае шифрование сообщения (информации) есть обратимое

преобразование сообщения, не зависящее от самого сообщения, с целью скрытия его содержания. Зашифрованное сообщение называется шифрограммой. Преобразование сообщения в шифрограмму описывается функцией шифрования; преобразование шифрограммы в сообщение описывается функцией дешифрования.

Другим методом обеспечения секретности информации является криптографическое кодирование. Криптографическое кодирование информации есть в общем случае преобразование по ключу сообщений в кодограммы, зависящее от самих сообщений, с целью скрытия их содержания. Системами криптографического кодирования информации называются криптографические системы, в которых защита информации по ключу основана на использовании ее избыточности. Термин «криптографическое кодирование» используется, чтобы подчеркнуть отличие этого вида криптографического преобразования от других видов некриптографических преобразований информации, таких как помехоустойчивое кодирование и эффективное кодирование (главы 4 и 5).

Криптосистемы аутентификации информации предназначены для контроля ее подлинности, но в ряде случаев они способны эффективно обеспечить контроль целостности сообщений при различных деструктивных воздействиях.

Данный класс криптосистем может быть разделен в зависимости от решаемой задачи на системы аутентификации информации (сообщений) и системы аутентификации источников информации (корреспондентов, пользователей, сетей, систем и т. п.). Методы аутентификации информации различаются в зависимости от условий обеспечения подлинности информации.

Рассмотрим пример, когда требуется проверить подлинность информации, передаваемой от отправителя к ее получателю, безусловно доверяющих друг другу; пользователи друг друга не могут обманывать и только внешний нарушитель может искажать информацию. Криптосистемы аутентификации сообщений для таких условий используют формирование и проверку имитовставок сообщений. В соответствии с ГОСТ 28147-89 имитовставка это отрезок информации фиксированной длины, полученный по определенному правилу из

открытых данных и ключа, и добавленный к зашифрованным данным для обеспечения имитозащиты. Имитозащита сообщений – их преобразование для защиты от навязывания нарушителем ложных и ранее передававшихся сообщений. Получатель зашифрованного сообщения и его имитовставки, имея такой же секретный ключ, способен из расшифрованного сообщения заново сформировать имитовставку и при ее совпадении с полученной имитовставкой из канала связи убедиться в отсутствии искажений.

В случае, когда требуется проверить подлинность информации, передаваемой от отправителя к ее получателю, не доверяющих друг другу, криптосистемы аутентификации на основе имитовставок не эффективны.

Подлинность информации в условиях взаимного недоверия сторон может быть обеспечена с использованием так называемой цифровой подписи сообщения, формируемой отправителем и проверяемой получателем сообщений. Невозможность выполнения каких-либо действий отправителя за получателя и получателя за отправителя при использовании цифровой подписи сообщения обусловлена тем, что они для формирования и проверки цифровой подписи используют различную ключевую информацию. Большинство криптографических систем и протоколов аутентификации объектов построены на основе криптосистем цифровой подписи сообщений.

Криптосистемы, обеспечивающие доступность информации, в настоящее время не являются самостоятельным классом и строятся на основе принципов, заимствованных из криптосистем аутентификации информации и криптосистем обеспечения секретности информации.

Таким образом, краткое рассмотрение возможных методов защиты информации свидетельствует о том, что многие задачи защиты информации наиболее эффективно решаются криптографическими методами, а ряд задач вообще может быть решен только с использованием криптографических методов защиты информации.

10.1.3. Оценка стойкости криптосистем

При оценке стойкости произвольных криптографических систем защиты информации обычно придерживаются принципа Керкхоффа: стойкость криптосистемы должна быть обеспечена и тогда, когда нарушителю известно полное ее описание. Поэтому при анализе стойкости криптографической системы будем предполагать, что противоборствующей стороне известно детальное описание системы, статистические характеристики используемого языка сообщений, пространства возможных ключей и криптограмм; она может иметь некоторую информацию о контексте сообщения и т.п. Единственное, чего не не должен знать нарушитель – секретный криптографический ключ, используемый пользователями криптографической системы защиты информации [31].

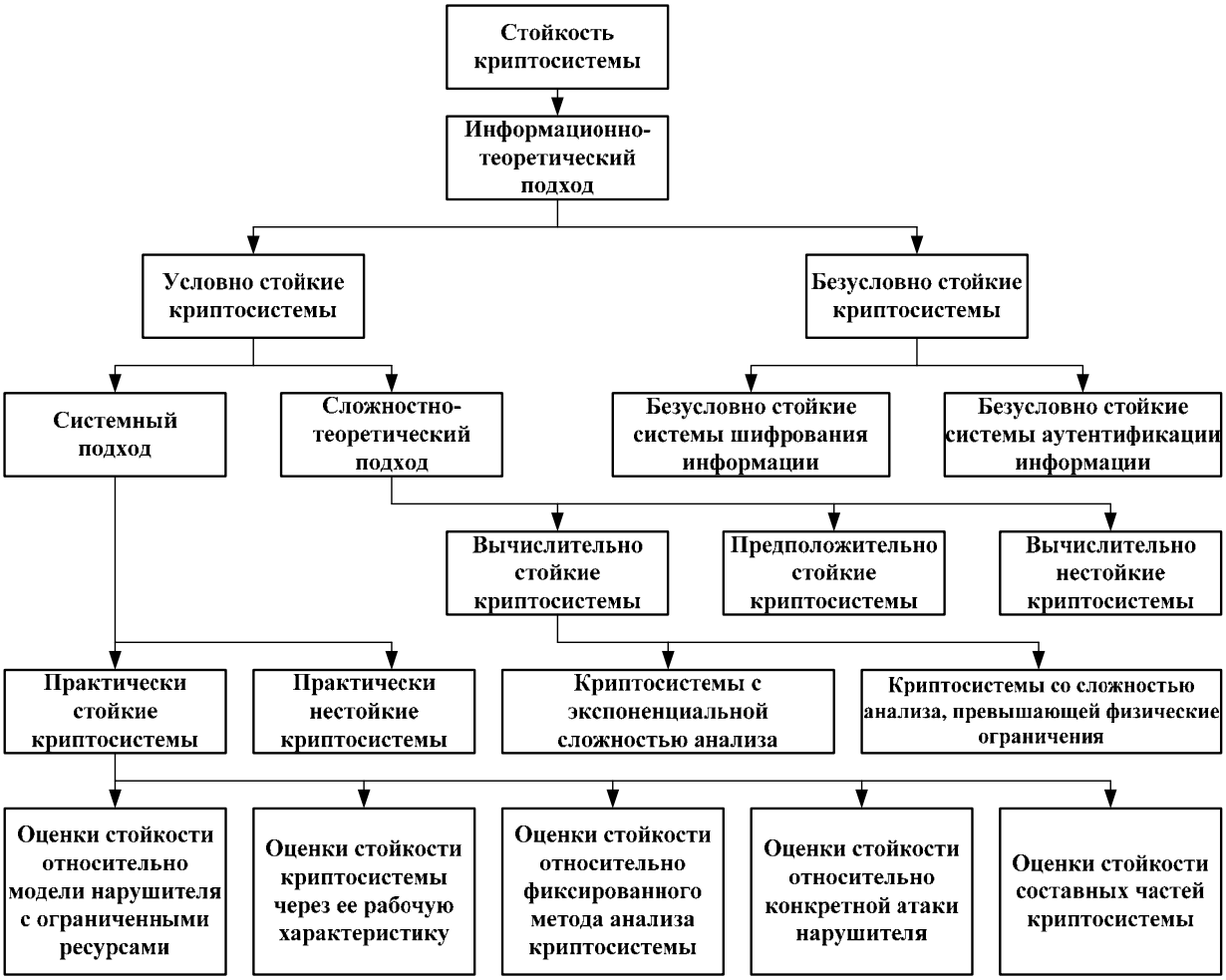
Для оценки стойкости криптографических систем защиты информации используются различные подходы, среди которых наибольший интерес представляют информационно-теоретический, сложностно-теоретический и системный подходы.

В соответствии с информационно-теоретическим подходом к оценке стойкости криптографических систем они могут быть разделены на, безусловно стойкие и на условно стойкие криптосистемы. Стойкость безусловно стойких криптографических систем не зависит ни от каких возможностей нарушителя и условий ее определения и не может быть уменьшена ни при каких обстоятельствах.

Стойкость условно стойких криптографических систем зависит от возможностей противоборствующей стороны и условий ее определения, и ее оценки могут меняться в зависимости от многих факторов.

Выяснение вопроса, является ли криптосистема безусловно или условно стойкой, составляет важную задачу информационно-теоретического подхода к оценке стойкости произвольных криптографических систем защиты информации. Если в рамках информационно-теоретического подхода криптосистема признана условно стойкой, то уточнить степень ее стойкости можно с использованием сложностно-теоретического и системного подходов. В научно-

технической литературе информационно-теоретический подход иногда относят к классу теоретических подходов к оценке стойкости криптосистем, а остальные – к классу практических подходов. На рис. 10.2 приведена классификационная схема оценок стойкости криптографических систем защиты информации.



10.2. Классификационная схема оценок стойкости криптографических систем защиты информации

Среди средств защиты информации от возможных атак нарушителя выделяются средства криптографической защиты информации. На них могут возлагаться следующие основные задачи:

обеспечение секретности (конфиденциальности) передаваемой, обрабатываемой и хранимой информации;

обеспечение целостности передаваемой, обрабатываемой и хранимой информации;

обеспечение подлинности сообщений и корреспондентов (пользователей), а также подлинности взаимодействующих сетей и систем;

установление авторства передаваемых и хранимых сообщений;

обеспечение доступности для законных корреспондентов (пользователей) информации, ресурсов и услуг;

обеспечение целостности самих средств криптографической защиты информации.

10.2. Функции, используемые в криптографических системах

Принципы построения криптографических систем защиты информации основаны на использовании математических функций специального вида, которые должны легко вычисляться законными пользователями, знающими «ключ», и очень сложно для всех не обладающих ключом.

10.2.1. Общее описание функций, используемых в криптографических системах

Рассмотрим пример произвольной функции $y = f(x)$, которую зададим графически (рис. 10.3). Пусть задано множество $X = \{a,b,c,d,e\}$ и множество $Y = \{1,2,3,4,5\}$. Напомним, что функция определяется двумя множествами X и Y, и правилом f, которое назначает каждому элементу из множества X один элемент из множества Y. Множество X называется областью определения функции, а множество Y областью ее значений.

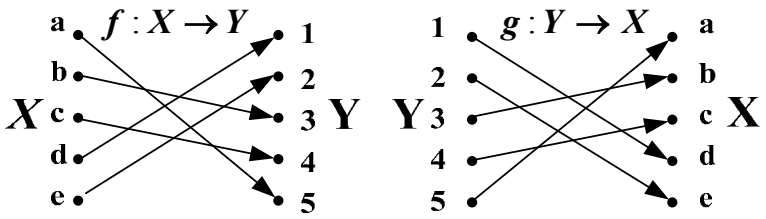


Рис. 10.3. Биективная функция f и обратная к ней $g = f^{-1}$

Элемент y из множества Y является образом элемента x, а элемент x является прообразом y. Отображение элементов из множества X в множество Y записывают так: $f : X \to Y$.

Множество всех элементов y, имеющих хотя бы один прообраз, называется образом функции f и обозначается $\mathrm{Im}(f)$.

Функция называется однозначной (отображением один в один), если каж-

дый элемент из множества Y является образом не более одного элемента из множества X. Функция f называется биекцией, если она является однозначной и $\mathrm{Im}(f) = Y$. Функция вида $g = f^{-1}$ называется обратной к f.

Среди биективных функций есть класс функций называемых инволюциями, которые наиболее часто используются для построения симметричных криптографических систем защиты информация.

Биективная функция называется инволюцией, если у функции совпадает область определения и область ее значений, т.е. $X = Y = S$, а также обратная функция с прямой $f = f^{-1}$. Пример инволюции для множества $S = \{1,2,3,4,5\}$ показан на рис. 10.4.

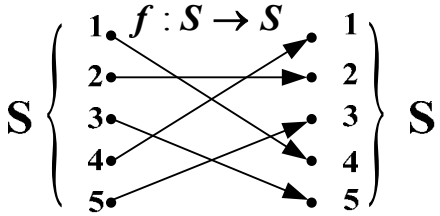


Рис. 10.4. Инволюция f для множества S={1,2,3,4,5}

Существование обратной функции является основой построения систем шифрования информации, с помощью которой можно однозначно дешифровать криптограммы в сообщения.

Последовательное применение сначала функции шифрования, а затем функции дешифрования к произвольному сообщению $x \in S$ однозначно восстанавливает данное сообщение: $f(f(x)) = x$.

10.2.2. Однонаправленные функции

Особую роль в криптографии играют однонаправленные функции, которые в общем случае не являются биективными.

Однонаправленной называется такая функция f, для которой легко определить значение функции $y = f(x)$, но практически невозможно отыскать для заданного y такое x, что $y = f(x)$.

Для построения криптографических систем защиты информации чаще используются однонаправленные функции, для которых обратное преобразование существует и однозначно, но вычислительно нереализуемо. Они называются вычислительно необратимыми функциями.

В качестве примера однонаправленной функции $y = f(x)$ рассмотрим широко известную функцию дискретного возведения в степень: $y = \alpha^x \pmod p$, где x – целое число от 1 до $p-1$ включительно, а вычисление производится по модулю p, где p – очень большое простое число; α – целое число $(1 < \alpha < p)$.

Напомним, что простым числом называется целое число, которое не делится ни на какие числа, кроме себя самого и единицы.

Пример 10.1. Для примера возьмем небольшое простое число $p = 7$; тогда для осуществления преобразований можно выбрать примитивный элемент $\alpha = 3$, так как $\alpha^1 \pmod 7 = 3$, $\alpha^2 \pmod 7 = 3^2 \pmod 7 = 9 \pmod 7 = 2$, $\alpha^3 \pmod 7 = 6$, $\alpha^4 \pmod 7 = 4$, $\alpha^5 \pmod 7 = 5$, $\alpha^6 \pmod 7 = 1$.

Функция $y = \alpha^x \pmod p$ вычисляется сравнительно просто, а обратная к ней функция $x = \log_y p$ является вычислительно сложной практически для всех $(1 < y < p)$ при условии, что не только p велико, но и $(p-1)$ имеет большой простой множитель (лучше всего, если это будет другое простое число, умноженное на 2). В связи с этим такую задачу называют задачей нахождения дискретного логарифма или задачей дискретного логарифмирования.

Задача дискретного логарифмирования состоит в том, что для известных целых α, p, y необходимо найти целое число x. Однако алгоритм вычисления дискретного логарифма за приемлемое время пока не найден. Поэтому модульная экспонента считается однонаправленной функцией.

По современным оценкам теории чисел при целых числах $\alpha \approx 2^{664}$ и $p \approx 2^{664}$ решение задачи дискретного логарифмирования потребует около 10^{26} операций, что 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.

Одним из первых применений однонаправленных функций было решение задачи обеспечения безопасности и использования пароля, по которому осуществляется доступ пользователя к ресурсам и услугам в автоматизированных системах.

Открытое значение y вместе с именем пользователя может быть помещено в список паролей доступа, хранящихся в ЭВМ. Законный пользователь для получения доступа в автоматизированную систему предъявляет свое число x. ЭВМ вычисляет по этому числу значение однонаправленной функции $y = \alpha^x (\mathrm{mod}\, p)$ и сравнивает с хранящимся значением y. При совпадении этих значений пользователь становится идентифицированным и получает требуемый доступ.

Кроме однонаправленных функций, не имеющих вычислительно простого обратного отображения даже для законных пользователей, знающих секретную ключевую информацию, в криптографии широко используются однонаправленные функции, для которых знание секретного ключа дает возможность законному пользователю вычислительно просто находить обратное отображение. Они получили название однонаправленных функций с потайным ходом, иногда их называют однонаправленными функциями с лазейкой.

Однонаправленные функции с потайным ходом

Быстрое развитие криптографии в последние два десятилетия во многом стало возможным благодаря открытию американскими учеными В. Диффи и М. Хэлманом однонаправленных функций с потайным ходом и их использованием для различных криптосистем защиты информации [1, 31].

Однонаправленная функция с потайным ходом есть однонаправленная функция f_z с дополнительным свойством, таким, что, зная информацию z потайного хода для каждого $y \in \mathrm{Im}(f)$ вычислительно просто определить $x \in X$, удовлетворяющее уравнению $y = f_z(x)$.

Для нарушителя, не знающего информации z потайного хода, нахождение отображения $y = f_z^{-1}(x)$ может быть сделано вычислительно нереализуемым.

Поэтому информация z может служить секретным ключом для пользователя функций с потайным ходом.

Однонаправленные функции с потайным ходом относятся к вычислительно необратимым функциям.

Функция вычислительно необратима, если при попытке формирования алгоритма нахождения обратного отображения к ней противник наталкивается на непреодолимую вычислительную проблему.

Оценивая стойкость криптосистем, построенных на основе известных однонаправленных функций с потайным ходом, отметим, что ни одна из них не является безусловно стойкой. Это объясняется тем, что нарушитель с теоретически бесконечными вычислительными ресурсами способен вычислять обратное отображение к таким функциям.

На основе однонаправленных функций с потайным ходом можно построить криптосистемы аутентификации информации в условиях взаимного недоверия корреспондентов, системы шифрования информации, в которых отправители сообщений могут пользоваться несекретными ключами шифрования, криптосистемы обмена секретной ключевой информации по открытым каналам связи, а также многие другие криптосистемы.

К настоящему времени предложено большое количество однонаправленных функций с потайным ходом, построенных на основе известных вычислительно сложных математических задач. Наиболее часто для построения однонаправленных функций с потайным ходом используется сложность решения следующих теоретико-числовых задач:

отыскание дискретного логарифма элемента в большом конечном поле или группе (криптосистема открытого распространения ключей Диффи-Хэллмана, криптосистема шифрования и криптосистема цифровой подписи сообщений Эль-Гамаля, криптосистема цифровой подписи сообщений Шнорра и другие криптосистемы) [1, 31, 36];

разложение больших чисел на простые множители (криптосистема шифрования и криптосистема цифровой подписи сообщений РША, криптосистема

цифровой подписи сообщений Рабина и другие криптосистемы) [1, 19];

задача об укладке целочисленного ранца (класс ранцевых систем шифрования информации Меркля-Хэллмана) [1, 36];

декодирование неизвестных получателю кодов Гоппы (класс систем шифрования информации Мак-Эллиса) [1].

Рассмотрим конкретные однонаправленные функции с потайным ходом, послужившие основой для широко используемых на практике криптографических систем защиты информация.

Однонаправленная функция РША с потайным ходом

В 1978 году была предложена первая однонаправленная функция с потайным ходом, положенная в основу широко используемой на практике несимметричной криптографической системы РША. Первые буквы фамилий ее авторов (Р. Ривеста, А. Шамира и Л. Адлемара) образовали общепринятое название предложенной ими функции и криптосистемы. Для описания однонаправленной функции РША с потайным ходом требуются некоторые сведения из элементарной теория чисел [1, 19, 31].

Однонаправленная функция РША с потайным ходом определяется как дискретное возведение значения x в степень ключа e: $f_z(x): y = x^e (\mathrm{mod}\, n)$, где $z\{p, g, d\}$ – информация потайного хода; p и g являются большими простыми числами; x – положительное целое число, не превосходящее $n = p \cdot g$; а значение e – положительное целое число, не превосходящее $\varphi(n)$ – функции Эйлера, для которого наибольший общий делитель $(e, \varphi(n)) = 1$.

Пусть $f_z(x)$ имеет обратную функцию вида $f_z^{-1}(y): x = y^d (\mathrm{mod}\, n)$, где значение d есть единственное положительное целое, меньшее $\varphi(n)$ и удовлетворяющее условию $d \cdot e = 1 (\mathrm{mod}\, \varphi(n))$.

Исследования однонаправленной функции РША с потайным ходом показали, что практически все попытки противостоящей стороны получить информацию о потайном ходе эквивалентны разложению $n = p \cdot g$ на множители. Поэтому в последние десятилетия интенсивно исследовались методы разложения

составного числа на множители. В математике такая задача называется задачей факторизацией составного числа и в течение столетий она привлекала внимание многих ученых. Известный наилучший алгоритм факторизации составного числа имеет субэкспоненциальную вычислительную сложность.

За последние годы в области разработки эффективных методов факторизации достигнуты существенные успехи, поэтому для обеспечения требуемой безопасности применения однонаправленной функции РША с потайным ходом должны использоваться числа p и g, размерностью многие сотни и даже тысячи бит.

Однонаправленная функция Эль-Гамаля с потайным ходом

Ранее была рассмотрена однонаправленная функция на основе вычисления дискретных логарифмов в алгебраической группе. Поле Галуа $GF(p)$, где p – простое число, является более сложной алгебраической структурой по сравнению с группой, над его элементами можно выполнять операции сложения и умножения, а в группе - только сложение или только умножение. Например, рассмотренная ранее однонаправленная функция Диффи и Хеллмана, послужившая основой криптосистемы открытого распространения ключей, использует операцию умножения над элементами группы [1, 31, 36].

Задача вычисления дискретных логарифмов в алгебраическом поле формулируется следующим образом. При заданных простом числе p, порождающем поле $GF(p)$, и примитивном элементе g $(0 < g < p)$ по элементу y поля отыскать элемент x $(0 \le x < p-1)$ такой, что выполняется тождество $y = g^x \pmod p$.

Число x является ключом формирования цифровой подписи сообщений отправителя и должно храниться отправителем сообщений в секрете, а значение y сообщается всем как открытый ключ проверки цифровой подписи сообщений отправителя.

На основе однонаправленной функции Эль-Гамаля с потайным ходом, как для функции РША, можно построить несимметричную систему шифрования информации.

Безопасность использования однонаправленной функции Эль-Гамаля с потайным ходом основана на вычислительной сложности задачи дискретного логарифмирования в алгебраическом поле большой размерности. Объем вычислений для ее решения выше, чем при факторизации составного числа.

Однонаправленная функция с потайным ходом на основе алгебраических уравнений по модулю 2

Значение однонаправленной функции $f(x)$ с потайным ходом зависит от аргументов M и K:

$$C = f(M, K) = (M_{d-1}, M_d),$$

где M – вектор сообщения, состоящий из двух частей M_0 и M_1 длиной по n бит каждая: $M = (M_0, M_1)$; K – вектор ключа, который определяется совокупностью подвекторов: $K = \{K_1, K_2, ..., K_{d-1}\}$.

Над вектором M циклически выполняются d раз операции вида:

$$M_i = M_{i-2} \oplus f(M_{i-1} K_{i-1}), \quad 2 \le i \le d,$$

где $f(\ )$ – некоторое фиксированное нелинейное преобразование, а знак $\oplus$ означает сложение по модулю два.

Рассмотренный принцип построения однонаправленной функции с потайным ходом используется при построении широкого класса блочных систем шифрования (класс блочных шифров Фейстеля) к которому принадлежат известный американский алгоритм шифрования данных DES и отечественный алгоритм шифрования согласно ГОСТ 28147–89 [1, 31, 36].

10.2.3. Криптографические хэш-функции

Понятие хэш-функций было определено в 1979 году в работах американского математика Р. Меркля, однако еще ранее в автоматизированных системах широко использовались некриптографические хэш-функции для оптимизации размещения и поиска данных [1, 31].

Хэш функции (хэширующие функции (ХФ)) произвольного вида принадлежат к классу однонаправленных функций без потайного хода.

ХФ $h(X)$ отображает сообщение X произвольной длины l в последовательность символов (хэш-код H) фиксированной длины m; для снижения скорости выбирают $m/l << 1$.

Если различные сообщения X отображаются в один и тот же хэш-код, то такая ХФ допускает коллизии (склеивания) сообщений.

Основными свойствами ХФ являются чувствительность к изменениям текста сообщения, отсутствие эффективных алгоритмов поиска коллизий, однонаправленность и простота вычисления $h(X)$.

Одной из важнейших характеристик ХФ является индекс хэширования (склеивания) $I(h,S)$ функции h, где S – словарь множества сообщений $\{X\}$. Если $I(h,S)=0$, то коллизий не происходит; каждое сообщение хэшируется в свой, отличный от других, хэш-код. ХФ, обеспечивающие $I(h,S)=0$, называются совершенными ХФ.

Если в процессе хэширования сообщений используется секретный ключ K, то такая функция $H=h(X,K)$ называется криптографической ХФ с секретным ключом (рис. 10.5,а). Криптографические ХФ, не использующие секретного ключа для хэширования сообщений $H=h(X)$, называются бесключевыми криптографическими ХФ (рис. 10.5,б). Бесключевые криптографические ХФ могут быть разделены на однонаправленные ХФ и устойчивые к коллизиям ХФ.

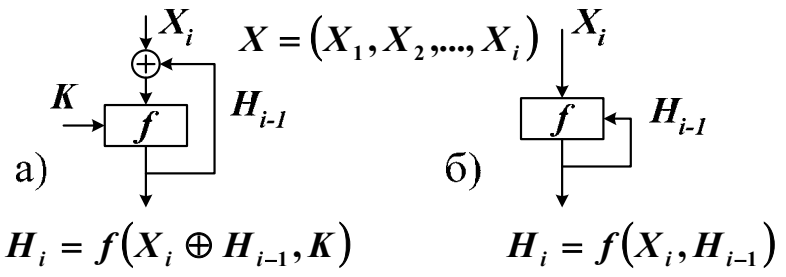


Рис. 10.5. Итеративное хэширование сообщения с ключом («а») и без ключа («б»)

Большинство известных бесключевых ХФ основано на разбиении произвольно длинных сообщений на блоки фиксированной длины и их последовательной обработке криптографической ХФ. Этот метод называется итеративным хэшированием. Хэшируемое сообщение X делится на i блоков от X_1 до X_i длиной по b бит. Если длина сообщения не кратна длине блока, то сообще-

ние должно дополняться до длины, кратной длине блока. Для инициализации процесса итеративного хэширования необходимо задать стартовый вектор хэширования H_0 длиной n бит. Хэширование сообщения осуществляется на основе функции f, которая образует выходное значение длиной n при задании блока исходного текста X_i и хэш-значения H_{i-1} предыдущего блока текста:

$$H_i = f(X_i, H_{i-1}),$$

где H_i – значение хэш-кода на i-й итерации хэширования.

Значение хэш-кода $h(X)$ всего сообщения $X = (X_1, X_2, ..., X_i)$ определяется как значение хэш-кода на последней итерации хэширования.

В настоящее время разработано много способов хэширования. В качестве примера рассмотрим однонаправленную ХФ вида:

$$h(X) = X^2 (\mathrm{mod}\, n). \qquad (10.1)$$

Процедура вычисления $h(X)$ является рекуррентной и применяется к сообщению X, разбитому на блоки $X = (X_1, X_2, ..., X_k)$:

$$H_i = \left[(H_{i-1} + X_i)^2\right] \mathrm{mod}\, n, \quad i = 1,2,...,k,$$

где H_0 – произвольное начальное число.

Пример 10.2. Пусть $n = 33$, а сообщение «ДВА» представлено номерами букв в русском алфавите, т.е. $X = (5,3,1)$. Выберем произвольно $H_0 = 4$. Тогда из (10.1) получим:

$$H_1 = \left[(4+5)^2\right] \mathrm{mod}\, 33 = 15, \quad H_2 = \left[(15+3)^2\right] \mathrm{mod}\, 33 = 27, \quad H_3 = \left[(27+1)^2\right] \mathrm{mod}\, 33 = 25.$$

Сообщение после хэширования имеет вид $X = (15,27,25)$ или «НЩЧ».

Криптографические ХФ в настоящее время широко используются для обеспечения безопасности информации (установление подлинности сообщений) и аутентификации пользователей криптографических систем и сетей.

В криптографических системах защиты информации ХФ используют для формирования дешифрующих последовательностей в шифрообразующих устройствах, для обеспечения секретности непрерывных и дискретных сообщений, а также для формирования случайных чисел в криптографических системах и во многих других приложениях.

10.3. Модели криптографических систем

10.3.1. Системы шифрования

Среди криптографических систем, обеспечивающих сохранение информации в тайне, наибольшее распространение получили системы шифрования информации. Рассмотрим обобщенную модель системы шифрования представленную на рис. 10.6.

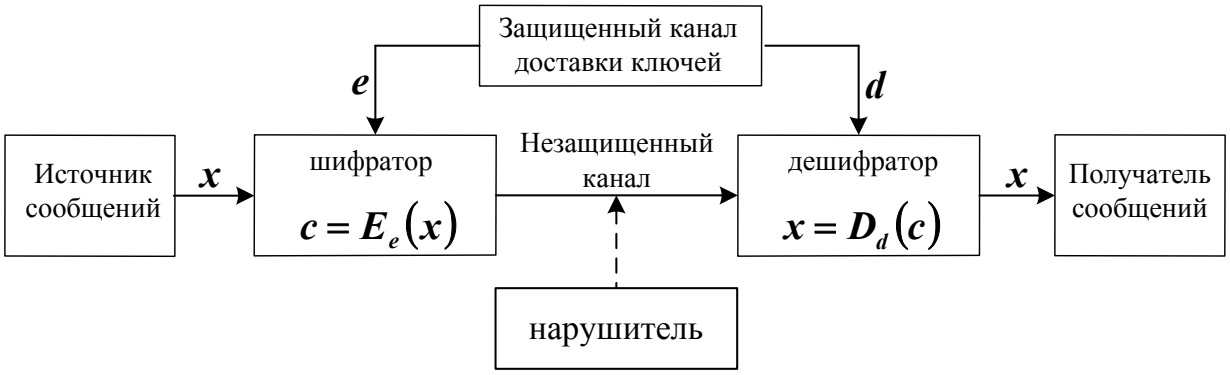


Рис. 10.6. Обобщенная модель системы шифрования информация

Источник сообщений генерирует сообщения x, которые необходимо сохранить в тайне от нарушителя при передаче по незащищенному каналу. В системе имеется защищенный от нарушителя источник ключевой информации, который вырабатывает некоторый ключ e, предназначенный для шифрования сообщений отправителем сообщений и ключ d, предназначенный для дешифрования криптограмм получателем. Ключи шифрования и дешифрования связаны друг с другом и позволяют восстановить сообщение из криптограммы. Сформированная ключевая информация передается по защищенному каналу ее доставки. Под защищенным будем понимать канал передачи информации, в котором нарушитель не способен на успешные атаки. Отправитель сообщений шифрует сообщение x по ключу e, используя шифрующее преобразование $c = E_e(x)$.

Образованная криптограмма c передается по незащищенному каналу передачи информации получателю. На приеме получатель способен из криптограммы однозначно восстановить сообщение x по ключу d, используя дешифрующее преобразование $x = D_d(c)$.

Для однозначного восстановления сообщения из криптограммы требует-

ся, чтобы дешифрующее преобразование D являлось обратным к шифрующему преобразованию E при использовании ключей d и e соответственно $D_d^{-1} = E_e$.

Системы шифрования информации разделяются на два больших класса: симметричные и несимметричные. Система шифрования информации называется симметричной, если для любой допустимой пары ключей (e,d) вычислительно просто определить один ключ, зная другой, т.е. из e можно вычислить d и, зная d, «легко» определить e. В таких системах оба ключа должны быть секретными. Во многих симметричных системах ключ шифрования совпадает с ключом дешифрования: $e = d = K$. Поэтому симметричные криптосистемы иногда называют одноключевыми системами или системами с секретным ключом.

Система шифрования информации называется несимметричной, если для любой допустимой пары ключей (e,d) вычислительно невозможно определить ключ дешифрования d, зная ключ шифрования e. В несимметричной системе шифрования ключ шифрования e может быть несекретным (открытым), известным для всех, включая нарушителя. Поэтому такие криптосистемы иногда называют системами с открытым ключом или двухключевыми системами. В таких системах должна обеспечиваться секретность ключа дешифрования d.

Несимметричные системы шифрования удобны для практического использования тем, что при доставке ключей отправителям сообщений не надо обеспечивать секретность ключевой информации шифрования сообщений.

Известно [36], что максимальная степень защищенности информации от чтения достигается, если произвольные передаваемые сообщения x и наблюдаемые нарушителем соответствующие им криптограммы c статистически независимы:

$$p\left(x_i \middle/ c_j\right) = p(x_i).$$

Для приближения характеристик реальных шифраторов к характеристикам идеального используют сжатие сообщений до шифрования и рандомизацию шифруемых сообщений. Идея рандомизации заключается в уменьшении избыточности шифруемых сообщений за счет специального кодирования, обес-

печивающего равную вероятность появления символов, но длина сообщений при этом увеличивается.

Основной характеристикой шифра является криптостойкость, которая обычно 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 отвечают шифры перестановок, шифры замены, шифры гаммирования и шифры, основанные на аналитических преобразованиях шифруемых данных [1, 19, 31, 36].

Шифрование перестановкой заключается в том, что символы исходного 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 исходного текста заменяются символами того же или другого алфавита в соответствии с заранее обусловленной схемой замены. Возможны моно- и многоалфавитные подстановки. В случае моноалфавитных подстановок каждый символ исходного текста преобразуется в символ шифрованного текста по одному и тому же закону. При многоалфавитной подстановке преобразование меняется от символа к символу. Для обеспечения высокой криптостойкости требуется использование сложных ключей.

Шифрование гаммированием заключается в том, что символы исходного текста складываются с символами некоторой псевдослучайной последовательности, именуемой гаммой шифра. Примером может служить, поразрядное сло-

жение сообщения x и гаммы z при формировании криптограммы $c = x \oplus z$. На приеме необходимо генерировать такую же псевдослучайную последовательность (z) тогда дешифрование будет осуществляться на основе следующего преобразования: $\tilde{x} = c \oplus z = x \oplus z \oplus z = x$. Стойкость шифрования определяется в основном длиной (периодом) неповторяющейся части гаммы шифра. Поскольку с помощью ЭВМ можно генерировать гамму шифра очень большой длины, то данный способ является одним из основных для шифрования информации в автоматизированных системах.

10.3.2. Традиционные симметричные криптосистемы

Симметричные системы шифрования информации подразделяются на блочные и поточные системы.

В блочной системе шифрования информации сообщение разбивается на информационные блоки фиксированной длины n бит и весь блок шифруется одновременно. Такие системы получили название блочных шифров; они представляют собой семейство обратимых преобразований блоков исходного текста. Фактически блочный шифр это система подстановки блоков. В настоящее время блочные шифры наиболее распространены на практике. Российский и американский стандарты шифрования относятся именно к этому классу.

К элементарным блочным шифрам относятся шифры подстановки и перестановки. В шифре подстановки каждый символ сообщения заменяется символом, определяемым функцией подстановки E_e. Вид функции подстановки задается ключом e. Каждый символ сообщения является информационным блоком фиксированной длины n бит. Символы сообщений принадлежат алфавиту A объемом 2^n: $A = \{0,1,...,2^n - 1\}$.

Каждый символ сообщения, записанный в верхней строке соответствия

$$\begin{matrix} \{0 & 1 & 2 & ... & 2^n - 1\}, \\ \{E_e(0) & E_e(1) & E_e(2) & ... & E_e(2^n - 1)\}, \end{matrix}$$

отображается с помощью функции подстановки E_e в соответствующий символ криптограммы того же алфавита, записанный под ним в нижней строке.

Если функция подстановки E_e фиксированная (ключ e является константой), то такой шифр называется простой заменой, или одноалфавитной подстановкой. Такой шифр в силу его простоты широко использовался в дипломатической и военной связи несколько веков тому назад.

Например, пусть в качестве алфавита сообщений и криптограмм использован алфавит русского языка. Зададим фиксированную функцию подстановки следующим образом:

(*А Б В Г Д Е Ж З И Й К Л М Н … Ю Я*),
(*П Г И Р Ш Д В Л Х Т Щ Я О Б … Н У*).

При шифровании очередной символ сообщения отыскивается в верхней строке и заменяется на соответствующий символ криптограммы, записанный снизу.

Для обеспечения однозначности операция дешифрования должна описываться подстановкой, обратной к подстановке операции шифрования.

В шифре перестановки сообщение делится на блоки фиксированной длины n и каждый символ переставляется в пределах блока в соответствии с функцией перестановки E_e. Вид функции перестановки определяется ключом e. Функция перестановки E_e каждый символ блока сообщения, записанный в верхней строке соответствия:

$$\begin{matrix}\{0 & 1 & 2 & \ldots & n-1\},\\ \{E_e(0) & E_e(1) & E_e(2) & \ldots & E_e(n-1)\}\end{matrix}$$

отображает в соответствующий символ блока криптограммы, записанный под ним в нижней строке

Пусть, например, требуется зашифровать сообщение: «ПЕРЕСТАНОВКА» шифром перестановки при $n=6$. Зададим фиксированную функцию перестановки E_e вида:

$$\begin{matrix}(0 & 1 & 2 & 3 & 4 & 5),\\ (4 & 2 & 5 & 0 & 1 & 3).\end{matrix}$$

Разобьем сообщение на два последовательных блока «ПЕРЕСТ» и «АНОВКА» длиною по 6 символов. С помощью функции перестановки полу-

чим блоки криптограммы «СРТПЕЕ» и «КОААНВ». Дешифрование выполняется, используя обратную перестановку.

Рассмотренные элементарные шифры подстановки и перестановки сами по себе не способны обеспечить требуемую в современных условиях стойкость криптографических систем, так как они уязвимы к атакам нарушителя на основе статистического анализа символов сообщений естественных языков, например таких, как телеграфные сообщения, написанные на русском языке. Однако, комбинируя элементарные шифры, можно построить широко используемые на практике стойкие композиционные шифры.

Пусть $\{X\}$, $\{Y\}$ и $\{Z\}$ есть конечные множества и пусть существуют функции $f:\{X\}\to\{Y\}$ и $g:\{Y\}\to\{Z\}$. Композицией функций f и g, обозначаемой $f*g$, называется функция, отображающая множество $\{X\}$ в множество $\{Z\}$ и определяемая как $f*g(x)=f(g(x))$ для всех x, принадлежащих множеству $\{X\}$.

На рис. 10.7 показан пример композиции функций f и g. Композиция функций легко расширяется на случай произвольного числа функций, позволяя из элементарных функций шифрования синтезировать композиционные шифры с высокой стойкостью.

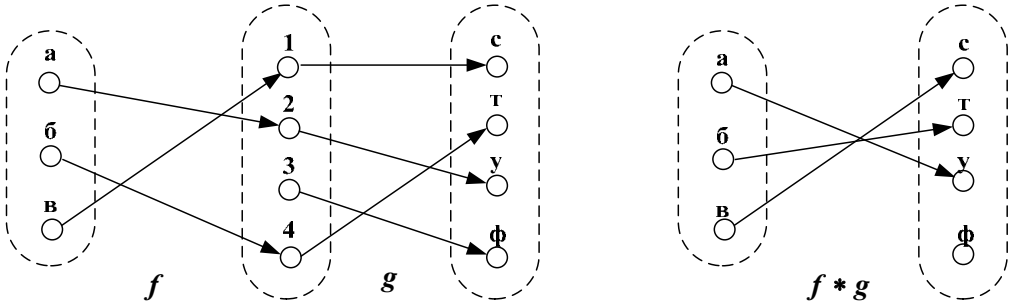


Рис. 10.7. Композиция функций f и g

В качестве элементарных шифрующих функций удобно использовать инволюции, т. е., функции, у которых прямая функция совпадает с обратной к ней функцией. Использование инволюций позволяет существенно упростить построение блочных шифраторов и дешифраторов, так как композиционная функция шифрования совпадает с соответствующей функцией дешифрования. Если исходные функции $E_{e_1}^1, E_{e_2}^2, ..., E_{e_t}^t$ являются инволюциями, то композиция

таких функций $E_e = E_{e_1}^1 * E_{e_2}^2 * ... * E_{e_t}^t$ имеет обратную функцию вида $E_e^{-1} = \left(E_{e_1}^1\right)^{-1} * \left(E_{e_2}^2\right)^{-1} * ... * \left(E_{e_t}^t\right)^{-1}$, где $e_1, e_2, ..., e_t$ – ключи шифрования; они же ключи дешифрования соответствующих исходных функций.

На практике удобно использовать подкласс композиционных шифров, называемый составными шифрами. В составных шифрах над блоком шифруемого сообщения многократно выполняются подстановки и перестановки. Последовательное использование подстановок и перестановок в составном шифре реализует два основных принципа построения симметричных блочных систем шифрования: рассеивание и перемешивание. Рассеивание заключается в распространении влияния всех символов блока открытого сообщения на все символы блока криптограммы, что позволяет скрыть статистические свойства шифруемой информации. Высокая степень рассеивания достигается многократными подстановками, зависящими не только от ключа, но и от промежуточных результатов шифрования. Одновременно рассеивание проявляется во влиянии каждого символа ключа на все символы блока криптограммы, что исключает восстановление взаимосвязи статистических свойств открытого и шифрованного текста. Перемешивание заключается в усложнении взаимосвязи статистических свойств сообщения и полученной из него криптограммы. Высокая степень перемешивания достигается при многократных перестановках в процессе криптопреобразований. Хорошее рассеивание и перемешивание обеспечивается использованием составного шифра, состоящего из последовательности элементарных шифрующих функций, каждая го которых вносит небольшой вклад в значительное суммарное рассеивание и перемешивание [1].

Современные составные шифры очень часто строятся как итеративные шифры, в которых над блоком сообщения многократно выполняется некоторый набор одних и тех же преобразований, называемых круговой функцией шифрования f. При выполнении каждой итерации круговой функции шифрования используется некоторая часть ключа шифрования, называемая подключом. Очередность выборки подключей из ключа шифрования называется расписани-

ем использования ключа шифрования. Итеративный процесс шифрования блока сообщения M можно записать в виде рекуррентной формулы:

$$C_0 = x,$$
$$C_i = f(C_{i-1}, K),\ i = 1,2,...,t,$$
$$C = C_t,$$

где C_i – значение блока криптограммы на i -й итерации шифрования; круговая функция шифрования f отображает предыдущее значение блока криптограммы C_{i-1} под управлением подключа K_i в очередное значение блока криптограммы C_i. Итоговое значение C определяется как значение блока криптограммы C_t на последней итерации шифрования. В качестве начального значения криптограммы C_0 используется значение блока сообщения x.

Итерационные блочные шифры вычислительно просто строить по схеме Фейстеля [1]. Для этого блоки сообщения и блоки криптограмм длиной по n бит разбиваются на левую L и правую R половины. Криптографическому преобразованию на каждой итерации шифрования подвергается только одна половина шифруемого сообщения; на следующей итерации половины меняются местами. Такой способ построения блочного шифра позволяет относительно просто при большом количестве итераций (циклов) обеспечить хорошее рассеивание и перемешивание. Схема одной итерации шифрования блочного шифра Фейстеля представлена на рис. 10.8. Алгоритм шифрования можно описать следующей рекуррентной процедурой. Вначале производится разбиение блоков на левую и правую половины: $(L_0, R_0) = x$. Затем над правым полублоком выполняются криптографические преобразования по круговой функции шифрования:

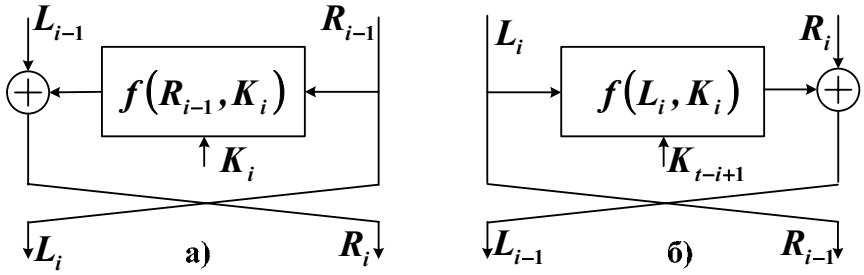


Рис. 10.8. Схемы шифрования («а») и дешифрования («б») одной итерации блочного шифра Фейстеля

$$R_i = L_{i-1} \oplus f(R_{i-1}, K_i),\ i = 1,2,...,t, \qquad (10.2)$$

полублоки меняются местами ($L_i = R_{i-1}$) и находится итоговое значение блока криптограммы $C = (L_t, R_t)$.

При дешифровании производится инициализация ($(L_t, R_t) = C$), а затем над левым полублоком выполняются криптографические преобразования по круговой функции шифрования f:

$$L_{i-1} = R_i \oplus f(L_i, K_i),\ i = 1,2,...,t; \qquad (10.3)$$

полублоки меняются местами ($R_{i-1} = L_i$) и формируется блок сообщения $x = (L_0, R_0)$.

В блочном шифре Фейстеля для дешифрования используется та же самая круговая функция f, что и для шифрования, но порядок использования подключей при дешифровании меняется на противоположный.

В реальных шифрах часто используется упрощенная схема Фейстеля. Для этого после t итераций шифрования левая и правая половины криптограмм меняются местами, что позволяет для дешифрования вместо выражения (10.3) использовать выражение (10.2) с учетом противоположного порядка применения подключей. Это существенно упрощает построение устройств шифрования; примером служит отечественный алгоритм криптографической защиты данных согласно ГОСТ 28147-89.

10.4. Алгоритмы криптографической защиты

Большинство практически используемых современных блочных шифров, таких как DES, ГОСТ РФ 28147-89, IDEA, построены, как шифры Фейстеля [1]. Их основные параметры представлены в табл. 10.1.

10.4.1. Алгоритм криптографической защиты информации согласно ГОСТ 28147-89

В Российской Федерации принят ГОСТ 28147-89 «Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования». Этот ГОСТ определяет алгоритм криптографической защиты

информации для сетей передачи данных, вычислительных систем, ЭВМ и устройств обработки информации. Он предназначен для шифрования как секретной, так и конфиденциальной информации и не накладывает ограничений на гриф секретности обрабатываемой информации.

Таблица 10.1

Параметры основных современных блочных шифров

Шифр	Длина блока n	Длина блока в битах	Число итераций
DES	64	56	16
LOKI	64	64	16
FEAL	64	128	$2^k, k \geq 5$
LUCIFER	128	128	16
PES	64	128	8
IDEA	64	128	8
ГОСТ 28147-89	64	256	32

Российский стандарт криптографического преобразования данных построен с учетом недостатков стандарта шифрования DES и обеспечивает существенно больший уровень безопасности информации [1, 36]. В отечественном стандарте длина оперативного ключа, вводимого пользователем, увеличена до 256 бит и введена возможность замены общего для сети обмена данных долговременного ключа длиной 512 бит. Рассматриваемый криптоалгоритм построен по модифицированной схеме блочных шифров Фейстеля и описывается 32 циклами шифрования каждого блока сообщения, что обеспечивает практически равновероятное влияние каждого бита шифруемого сообщения и ключа на все элементы криптограммы. В российском стандарте определен режим имитозащиты передаваемых сообщений, причем пользователь может сам устанавливать допустимую вероятность навязывания нарушителем ложных сообщений.

Криптографический алгоритм по ГОСТ 28147-89 обеспечивает защиту информации в следующих режимах:

шифрование/дешифрование данных в режиме простой замены (подста-

новки);

шифрование/дешифрование данных в режиме гаммирования;

шифрование/дешифрование данных в режиме гаммирования с обратной связью,

режим выработки имитовставки

10.4.2. Режимы работы алгоритма криптографического преобразования данных

Шифрование и дешифрование данных в режиме простой замены

Рассмотрим работу алгоритма в режиме простой замены [1, 36]. Схема алгоритма шифрования в данном режиме представлена на рис 10.9. В ее состав входят: ключевое запоминающее устройство (КЗУ) объемом 256 бит (для оперативного ключа), состоящее из восьми 32-разрядных накопителей $K_0, K_1, ..., K_7$; блок подстановок S (для долговременного ключа) объемом 512 бит состоящий из восьми узлов замены $S_1, S_2, ..., S_8$, с памятью 64 бита каждый. В каждом узле замены составляют ключевую таблицу, разбитую на 16 строк по 4 бита в каждой строке; регистр сдвига, осуществляющий циклический сдвиг вектора на 11 бит в сторону старших разрядов.

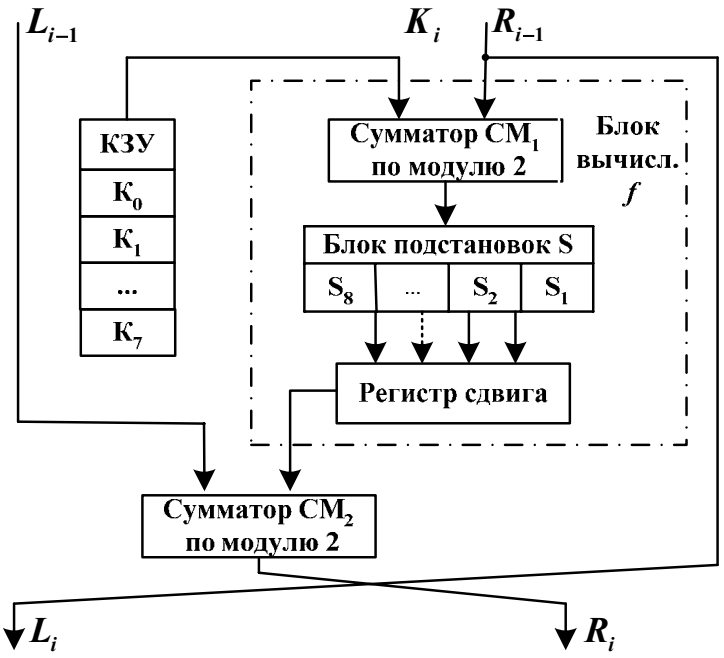


Рис. 10.9. Цикл режима простой замены

Сообщение, подлежащее шифрованию, разбивается на блоки x_k длиной 64 бита. Каждый блок разделяется на начальные левый L_0 и правый R_0 подблоки длиной по 32 бита. В соответствии с упрощенной схемой Фейстеля в течение 32 циклов выполняются фиксированные криптографические преобразования.

В первом цикле шифрования начальный правый подблок R_0 длиной 32 бита и оперативный ключ K_0, в соответствии с расписанием использования ключа из накопителя, складываются в сумматоре CM_1 по модулю 2.

На выходе сумматора полученный полублок длиной 32 бита разбивается на восемь четырех разрядных векторов, являющихся адресом для выбора одной из 16 строк ключевой таблицы, и поступающих на соответствующие узлы замены $S_1, S_2, ..., S_8$ блока подстановок S. С помощью каждого четырехразрядного вектора выбирается одна из 16 строк ключевой таблицы соответствующего узла замены. По четыре бита с восьми узлов замены считываются на выход блока подстановок S, объединяясь в 32-разрядный выходной вектор, который поступает на вход регистра сдвига. Сдвинутый на 11 бит вектор складывается в сумматоре CM_2 с начальным левым полублоком L_0. Результат сложения будет являться правым полублоком R_1. В качестве левого полублока L_1 используются значения начального правого полублока R_0 над которыми выполняется следующий цикл шифрования и т.д. После 32-го цикла шифрования полублоки R_{32} и L_{32} меняются местами и составляют блок криптограммы C.

Процесс шифрования в режиме простой замены блока сообщения x_k под управлением ключа K в блок криптограммы C записывается в виде: $C = E_K(x_k)$. Дешифрование в режиме простой замены выполняется аналогично шифрованию и отличается обратным порядком использования тех же ключей.

Шифрование и дешифрование данных в режиме гаммирования

Криптосхема алгоритма шифрования в режиме гаммирования представлена на рис. 10.10. В состав схемы, кроме рассмотренных ранее элементов, входят регистры хранения (N_1, N_2). Так же, как и в режиме простой замены, выполняется заполнение КЗУ и блока подстановок S.

На вход схемы поступают равновероятно и взаимонезависимо сформированные 64 бита синхропосылки S^0, которая разделяется на начальные левый SL_0 и правый SR_0 подблоки длиной по 32 бита. Подблоки синхропосылки зашифровываются в режиме простой замены в течение 32 циклов шифрования и

записываются в накопителях N_1 и N_2 как значения VR_0 и VL_0 соответственно. Затем эти значения складываются в сумматорах CM_1 и CM_2 по модулю 2 с 32-битовыми криптографическими константами c_1 и c_2.

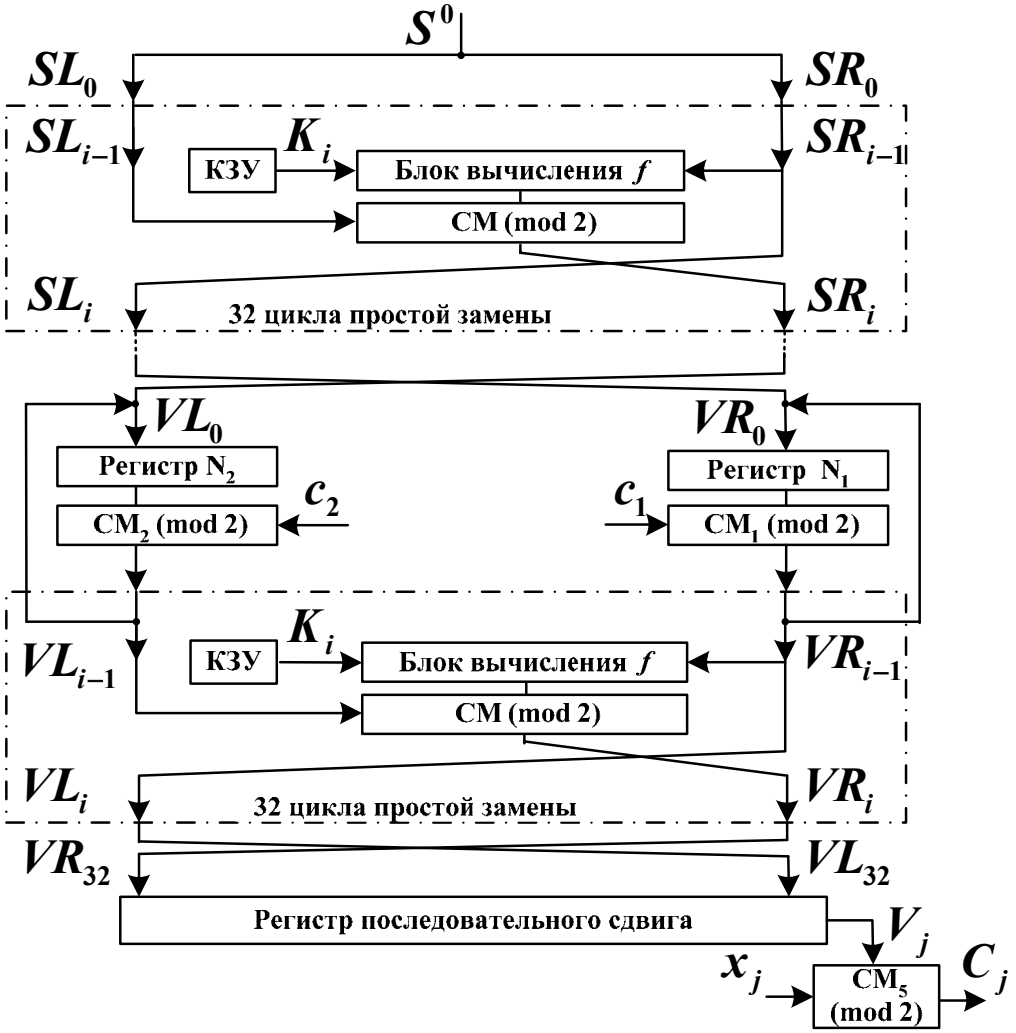


Рис. 10.10. Режим гаммирования

Новые значение VR_0 и VL_0, поступающие с выхода сумматоров CM_1 и CM_2 запоминаются в накопителях N_1 и N_2. В течение 32 циклов значения VR_0 и VL_0 зашифровываются в режиме простой замены, образуя первый 64-разрядный блок шифрующей гаммы V_1, состоящий из правого и левого полублоков.

Полублоки шифрующей гаммы записываются в регистр последовательного сдвига, из которого побитно считываются для шифрования битов первого 64-разрядного блока открытого сообщения x_1. Очередной бит блока сообщения шифруется в режиме гаммирования путем поразрядного сложения по модулю 2 в сумматоре CM_3 с очередным битом шифрующей гаммы. Сформированные та-

ким образом 64 бита составляют первый блок криптограммы C_1.

Для шифрования второго и последующих блоков открытого сообщения из регистров хранения считываются значения, складываются с новыми криптографическими константами и полученные значения запоминаются в этих же регистрах. Далее формирование и использование второго и последующих блоков шифрующей гаммы выполняется в соответствии с рассмотренным алгоритмом. По каналу связи последовательно передаются синхропосылка S^0 и сформированные блоки $C_1, C_2, \ldots$ криптограммы.

Дешифрование в режиме гаммирования выполняется аналогично шифрованию. Из полученной синхропосылки S^0 последовательно формируются блоки дешифрующей гаммы, которые используются для дешифрования принятых блоков криптограммы.

Для исключения снижения криптостойкости при повторном использовании одного и того же оперативного ключа необходимо в каждом сеансе связи использовать неповторяющуюся синхропосылку S^0.

Шифрование и дешифрование данных в режиме гаммирования с обратной связью

Криптосхема, реализующая алгоритм шифрования в режиме гаммирования с обратной связью, показана на рис. 10.11.

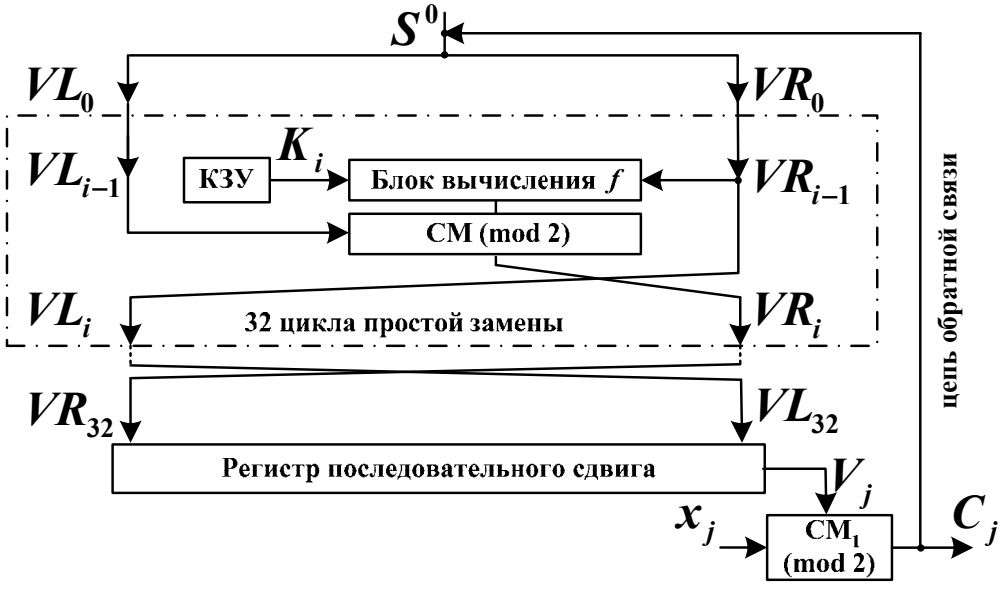


Рис. 10.11. Режим гаммирования с обратной связью

Оперативный ключ, состоящий из 256 бит, вводится в ключевое запоминающее устройство (КЗУ), долговременный ключ длиной 512 бит, аналогично режиму простой замены и гаммирования, записывается в блок подстановок S.

Как и в режиме гаммирования, генерируется синхропосылка, разделяется на начальные левый VL_0 и правый VR_0 подблоки и зашифровывается в режиме простой замены, образуя первый блок V_1 шифрующей гаммы, который записывается в регистр последовательного сдвига.

Первый 64-разрядный блок открытого сообщения x_1 шифруется поразрядным сложением по модулю 2 в сумматоре CM_1 с первым блоком шифрующей гаммы V_1.

Сформированные таким образом 64 бита составляют первый блок криптограммы C_1, который передается в канал связи и одновременно по цепи обратной связи подается как синхропосылка на вход схемы, где шифруется в режиме простой замены и формирует второй блок шифрующей гаммы V_2.

Аналогично шифруется второй блок открытого сообщения x_2, сформированный блок криптограммы C_2 по цепи обратной связи используется для формирования очередного блока шифрующей гаммы и т.д.

Криптосхема алгоритма расшифрования в режиме гаммирования с обратной связью аналогична криптосхеме алгоритма шифрования. Используется идентичное режиму шифрования заполнение КЗУ и блока подстановки. Из полученной синхропосылки S^0, по аналогии с процессом формирования первого блока шифрующей гаммы, формируется первый блок дешифрующей гаммы, который используется для побитного дешифрования первого принятого блока криптограммы. Затем первый блок криптограммы используется для формирования второго блока дешифрующей гаммы, предназначенного для дешифрования второго принятого блока криптограммы и т.д.

Данный алгоритм повышает имитостойкость шифрованной связи. Однако для выполнения современных требований к имитозащищенности ГОСТ предписывает специальный режим имитозащиты данных.

Режим выработки имитовставки

Для имитозащиты шифруемых данных, состоящих из 64-разрядных блоков $x_1, x_2, ..., x_N$ при $N > 2$, формируется имитовставка E_c. Криптосхема, реализующая алгоритм выработки имитовставки, представлена на рис 10.12.

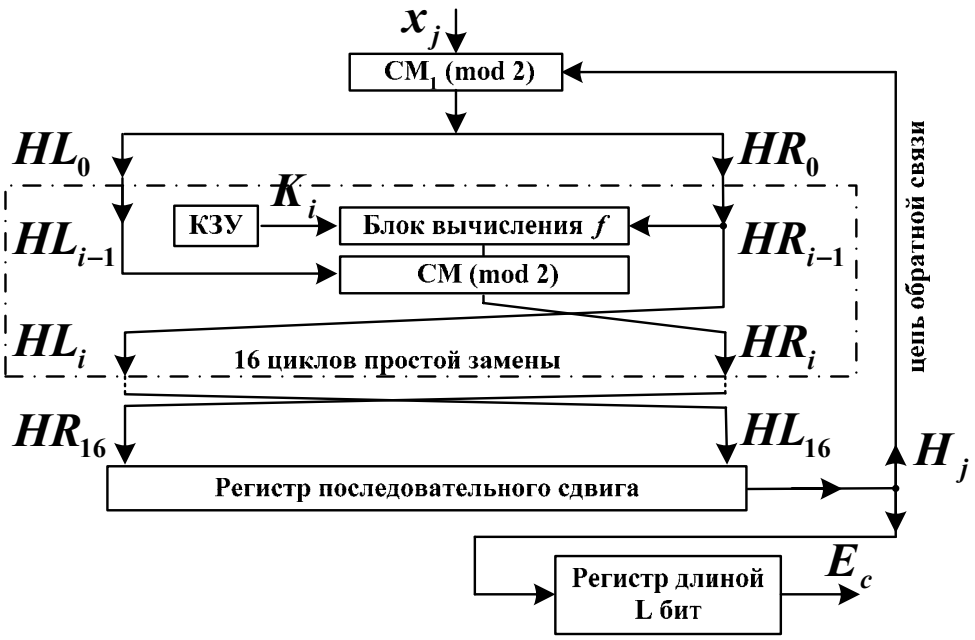


Рис. 10.12. Режим выработки имитовставки

В ней используются те же самые оперативный и долговременный ключи для имитозащиты данных, как и для шифрования.

Первый блок открытых данных x_1 разделяется на левый HL_0 и правый HR_0 полублоки, которые зашифровываются в режиме простой замены в течение 16 циклов шифрования, образуя первый блок хэш-кода открытых данных H_1.

Сформированный первый блок хэш-кода переписывается в регистр последовательного сдвига и побитно складывается по модулю 2 в сумматоре CM_1 со вторым блоком открытых данных x_2 с образованием полублоков HL_0 и HR_0.

Полублоки HL_0 и HR_0 используются для формирования второго блока хэш-кода H_2, который складывается с третьим блоком открытых данных x_3. Действия выполняются до тех пор, пока не будет обработан последний блок открытых данных x_N.

В качестве имитовставки E_c используются первые L бит правого полублока хэш-кода HR_{16}^N. Длина L имитовставки определяется требованиями по имитозащищенности связи.

Сформированная имитовставка E_c передается по каналу связи после зашифрованных данных или поступает в память ЭВМ. Стандарт определяет, что имитозащита данных реализуется одинаково для всех режимов их шифрования. Стандарт допускает в первых блоках данных размещение служебной информации (адреса сообщения, метки времени, синхропосылки и др.), которая может не шифроваться. При высоких требованиях к имитозащищенности метка времени обязательно должна присутствовать в передаваемой информации.

Криптосхема, реализующая алгоритм проверки имитовставки, аналогична криптосхеме ее выработки. Принятые блоки криптограммы расшифровываются, и заново формируется имитовставка, которая сравнивается с принятой. При их несовпадении принятые данные считаются ложными.

Данный режим должен использоваться для обеспечения имитозащищенности связи в условиях, когда исключена возможность ввода ложной информации со стороны законных пользователей шифрованной связи.

Рассмотренный криптографический алгоритм защиты информации по ГОСТ 28147-89 во всех описанных режимах обеспечивает высокую криптографическую стойкость и допускает программную или аппаратную реализацию.

Контрольные вопросы

1. В чем состоит принципиальное отличие криптографических и криптоаналитических методов?

2. Что называют криптографическим ключом?

3. Поясните сущность и сравните между собой криптографический алгоритм, криптографический протокол и криптографическую систему зашиты информации.

4. Что подразумевают под аутентификацией информации?

5. Перечислите криптографические функции?

6. В чем состоит принципиальное отличие симметричных и несимметричных систем шифрования информации?

7. Перечислите режимы защиты информации по ГОСТ 28147-89. В чем

состоит существенное отличие методов защиты информации по ГОСТ 28147-89 от методов согласно стандарта шифрования DES?

ЗАКЛЮЧЕНИЕ

В процессе развития теории электрической связи были решены вопросы, связанные с формированием и оптимальной обработкой сигналов в каналах связи с различными характеристиками, предложены эффективные методы помехоустойчивого кодирования. В то же время эти достижения еще далеко не полностью используются в современной технике связи. Существует известный разрыв между достижениями науки и их внедрением, определяемый необходимым временем на разработку и освоение в производстве новой аппаратуры.

Переход к цифровым методам передачи сообщений и цифровой обработке сигналов при широком использовании микропроцессорной техники обеспечивает возможности интеграции средств связи и вычислительных средств. На этой основе создаются интегральные цифровые сети, в которых достигается не только объединение по видам связи и услуг, но и общая программно-аппаратная реализация систем передачи, обработки, коммутации, управления и контроля. Интегральные сети, объединяющие в единый комплекс вычислительные и информационные системы на базе современных ЭВМ, образуют единую информационно-коммуникационную сеть. Информационно-коммуникационные сети являются технической основой информатизации отраслей, регионов, стран, мирового сообщества.

Проблемы информатизации предъявляют высокие требования как к вычислительной технике, так и к технике связи. Для техники связи – это прежде всего требования высоких скоростей (порядка гигабит и более в секунду), малых вероятностей ошибок (порядка $\mathbf{10^{-10}...10^{-11}}$), больших дальностей передачи (до **100** млн. км в системах космической связи), малого веса и энергопотребления аппаратуры. Однако, даже при использовании современных технологий, проектирование систем связи с высокими показателями эффективности ставит перед теорией электрической связи ряд новых нерешенных задач и проблем.

Настоящее пособие призвано дать основополагающие знания по современным технологиям преобразования и передачи сообщений и сигналов по ка-

налам связи. Эти знания, являясь базовыми, позволят читателям видеть и понимать состояние и тенденции развития современных информационных систем, ориентироваться во всем их многообразии, в том числе и по многочисленным публикациям. К рассмотренным в пособии технологиям относятся: кодирование, алгоритмы АЦП непрерывных сигналов, модуляция, сигнально-кодовые конструкции, распределение ресурса общего канала, расширение частотного спектра, сжатие и защита информации.

Ограниченный объем пособия не позволил включить в него вопросы оптимальной фильтрации непрерывных сообщений, многие разновидности помехоустойчивых кодов, справочный материал (таблицы некоторых функций). Все они войдут в учебник «Теория электрической связи» подготавливаемый к выпуску коллективом авторов.

СПИСОК ЛИТЕРАТУРЫ

1. Баричев С.Г., Гончаров В.В., Серов Р.Е. Основы современной криптографии. – М.: Горячая линия. Телеком, 2001. – 120с.

2. Березюк Н.Т., Андрущенко А.Г., Мощицкий С.С. и др. Кодирование информации (двоичные коды). / Под ред. Н.Т. Березюка. – Харьков: Вища школа, 1978. – 252 с.

3. Блейхут Р. Теория и практика кодов, контролирующих ошибки. – М.: Мир, 1986.

4. Брейсуэлл Р. Преобразование Хартли. / Пер. с английского А.И. Папкова. – М.: Мир, 1990. – 175с.

5. Борисов В.А., Калмыков В.В., Ковальчук Я.М. и др. Радиотехнические системы передачи информации. / Под ред. В.В. Калмыкова. – М.: Радио и связи, 1990. – 304с.

6. Бураченко Д.Л., Клюев Н.Н., Коржик В.И., Финк Л.М. и др. Общая теория связи. / Под ред. Л.М.Финка. – Л.: ВАС, 1970. – 412с.

7. Вальд А. Статистически решающие функции. Позиционные игры.. – М.: Наука, 1967. – 522с.

8. Варакин Л. Е. Теория систем сигналов. – М.: Сов. радио, 1978. – 304с.

9. Васильев К.К. Методы обработки сигналов: Учебное пособие. – Ульяновск: УлГТУ, 2001. – 80с.

10. Васильев К. К., Новосельцев Л. Я., Смирнов В. Н. Основы теории помехоустойчивых кодов: Учеб. пособие. – Ульяновск: УлГТУ, 2000. – 91с.

11. Винер Н.Я. Математика. – М.: Наука, 1967. – 300с.

12. Галлагер Р. Теория информации и надежная связь / Пер. с англ. под ред. М.С. Пинскера и Б.С. Цыбакова. – М.: Сов. радио, 1974. – 720с.

13. Глушков В.А., Нестеренко А.Г. Теория электрической связи. Часть 1. Дискретные сигналы. Учебное пособие. Ульяновск: УФВУС, 2003. – 96с.

14. Глушков В.А., Нестеренко А.Г., Попов Н.А. Теория электрической связи. Учебное пособие. Часть 2. Помехоустойчивость. – Ульяновск: УВВИУС,

2007. – 78с.

15. Глушков В.А., Нестеренко А.Г., Попов Н.А. Телекоммуникационные системы. Учебное пособие. Часть 1. Аналоговые и цифровые сигналы. – Ульяновск: УВВИУС, 2007. – 131с.

16. Глушков В.А., Нестеренко А.Г., Чикалев С.Б. Телекоммуникационные системы. Учебное пособие. Часть 2. Принципы построения систем связи. – Ульяновск: УВВИУС, 2007. – 118с.

17. Гоноровский И.С., Демин М.П. Радиотехнические цепи и сигналы. – М.: Радио и связь, 1994. – 480с.

18. Григорьев В.А., Григорьев С.В. Передача сообщений. / Под ред. В.А. Григорьева. – СПб.: ВУС, 2002. – 460с.

19. Жельников В. Криптография от папируса до компьютера. – М.: ABF, 1996. – 336с.

20. Зюко А.Г., Кловский Д.Д., Назаров М.В., Финк Л.М.. Теория передачи сигналов. – М.: Радио и связь, 1986. – 304с.

21. Зюко А.Г., Кловский Д.Д., Коржик В.И., Назаров М.В. Теория электрической связи. Учебник для вузов. / Под ред. Д.Д. Кловского. – М.: Радио и связь, 1999. – 432с.

22. Зюко А.Г., Фалько А.И., Панфилов И.П., Банкет В.Л., Иващенко П.В. Помехоустойчивость и эффективность систем передачи информации. / Под ред. А.Г. Зюко. – М.: Радио и связь, 1985. – 272с.

23. Игнатов В. А. Теория информации и передачи сигналов: Учебник для вузов. – М.: Радио и связь, 1991. – 280с.

24. Каганов В.И. Радиотехнические цепи и сигналы. Компьютеризированный курс: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2005. – 432с.

25. Кассами Т., Токура Н., Ивадари Е., Инагаки Я. Теория кодирования/ Пер. с япон. под ред. Б. С. Цыбакова и С. И. Гельфанда. – М.: Мир, 1978. – 576с.

26. Кларк Дж. мл., Кейн Дж. Кодирование с исправлением ошибок в системах цифровой связи / Пер. с англ. под ред. Б.С. Цыбакова. – М.: Радио и

связь, 1987. – 392с.

27. Колмогоров А.Н. Интерполирование и экстраполирование стационарных последовательностей. – М.: Изв. АН СССР. Сер. Матем., 1941, №5. С. 3–14.

28. Котельников В.А. Теория потенциальной помехоустойчивости. – М.: Госэнергоиздат, 1956. – 152с.

29. Левин Б.Р. Теоретические основы статистической радиотехники. – М.: Радио и связь, 1989. – 653с.

30. Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А. Теория кодов, исправляющих ошибки. – М.: Связь, 1979. – 744с.

31. Оков И.Н. Криптографические системы защиты информации. – СПб.: ВУС, 2001. – 236с.

32. Панфилов И.П., Дырда В.Е. Теория электрической связи. – М.: Радио и связь, 1991. – 344с.

33. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки / Пер. с англ. под ред. РЛ. Добрушина и С.И. Самойленко. – М.: Мир, 1976. – 596с.

34. Прокис Джон. Цифровая связь / Пер. с англ. под ред. Д.Д. Кловского. – М.: Радио и связь, 2000.

35. Ройтенберг Я.Н. Автоматическое управление. – М.: Наука, 1978. – 552с.

36. Романец Ю.В., Тимофеев П.А., Шаньгин В.Ф. Защита информации в компьютерных системах и сетях/ Под ред. В.Ф. Шаньгина. – М.: Радио и связь, 2001. – 376с.

37. Сифоров В.И. О влиянии помех на прием импульсных сигналов «Радиотехника», 1947, № 1.

38. Скляр Б. Цифровая связь. Теоретические основы и практическое применение. – М.: Изд. Дом «Вильямс», 2003. – 1104с.

39. Теплов Н.Л. Теория передачи сигналов по электрическим каналам связи. – М.: МО СССР, 1976. – 424с.

40. Тихонов В.И. Оптимальный прием сигналов. – М.: Радио и связь,

1983. – 320с.

41. Фано Р. Передача информации. Статистическая теория связи: Пер. с англ. под ред. Р. Л. Добрушина. – М.: Мир, 1965. – 438с.

42. Финк Л.М. Теория передачи дискретных сообщений. – М.: Советское радио, 1970. – 728с.

43. Харкевич А.А Избранные труды. Т.3. Теория информации. Опознание образов. – М.: Наука, 1972. – 524с.

44. Хартли Р. Передача информации. Сборник: Теория информации и ее применение. / Под ред. А.А. Харкевича. – М.: Физматгиз, 1959.

45. Хелстром К. Статистическая теория обнаружения сигналов. – М.: ИЛ, 1963.

46. Хинчин А.Я. Понятия энтропии в теории вероятностей. Успехи мат. наук, 1953, №3.

47. Хинчин А.Я. Об основных теоремах теории информации. Успехи мат. наук, 1956, №1.

48. Хинчин А. Работы по математической теории массового обслуживания. – М.: Физмат, 1963. – 236с.

49. Шеннон К. Работы по теории информации и кибернетике / Пер. с англ. под ред. Н.А. Железнова. – М.: ИЛ, 1963. – 829с.

50. Harry Nyquist. Certain factors affecting telegraph speed. – Bell System Technical Journal, 3, 1924. С.324–346.

ОГЛАВЛЕНИЕ