

ББК 32.84
Г 97
УДК 621.396.6.001.63

Гуткин Л. С.

Г 97 Проектирование радиосистем и радиоустройств: Учеб.
пособие для вузов. — М.: Радио и связь, 1986. — 288 с.: ил.

Налагаются вопросы проектирования, являющиеся общими для информационных радиотехнических систем различных классов и назначений (передачи информации, радиолокация, радиоуправления и др.). Рассматриваются их основные показатели качества (точность, пропускная способность, помехоустойчивость, электромагнитная и экологическая совместимость, надежность, габаритные размеры, масса и стоимость). Анализируются особенности проектирования с учетом совокупности этих показателей. Указываются пути повышения эффективности эвристических решений при проектировании, основанные на системном подходе.
Для студентов вузов радиотехнических специальностей; полезна инженерам, занимающимся проектированием радиоэлектронных систем.

Г 2402020000-097
046(01)-86 93-86

ББК 32.84

Рецензенты: кафедра радиоуправления Казанского авиационного института; д-р техн. наук, проф. В. А. Вейцель

Учебное пособие

Лев Соломонович Гуткин

ПРОЕКТИРОВАНИЕ РАДИОСИСТЕМ И РАДИОУСТРОЙСТВ

Заведующая редакцией Г. И. Козырева
Редакторы В. Ю. Ярошевский, Т. М. Любимова
Художественный редактор Н. С. Шейн
Переплет художника Ю. А. Архангельского
Технический редактор А. Н. Золотарева
Корректор Н. В. Козлова
ИБ № 285

Сдано в набор 10.10.85 Подписано в печать 01.04.86
Т-08727 Формат 60×90/16 Бумага кн.-журнальная № 2 Гарнитура литературная
Печать высокая Усл. печ. л. 18,0 Усл. кр.-отт. 18,0 Уч.-изд. л. 20,05 Тираж 20 000 экз.
Изд. № 19942 Зак. № 112 Цена 1 р.
Издательство «Радио и связь», 101000 Москва, Почтамт, а/я 653

Московская типография № 5 ВГО «Союзучетиздат»
101000 Москва, ул. Кирова, д. 40

© Издательство «Радио и связь», 1986

ПРЕДИСЛОВИЕ

Почти все профилирующие дисциплины, изучаемые на радиофакультетах, в той или иной мере направлены на подготовку студентов в области проектирования радиосистем и радиоустройств.

Однако имеется ряд весьма важных общих вопросов проектирования радиосистем и радиоустройств, которые не освещаются с достаточной полнотой ни в одном из курсов и в соответствующих им учебных пособиях. К ним относятся следующие вопросы, которым и посвящена данная книга:

важнейшие показатели качества радиосистем и радиоустройств (и общие пути их улучшения) — показатели точности, пропускной способности, помехоустойчивости, электромагнитной совместимости, экологической совместимости, надежности, массы, габаритных размеров и стоимости;

общая характеристика задач и этапов проектирования; особенности проектирования систем и устройств с учетом совокупности (вектора) показателей качества. Пути сведения этой векторной задачи к скалярной путем применения безусловного и условных критериев предпочтения;

значение и содержание системного подхода к проектированию систем и входящих в них устройств.

В заключительной главе приводятся примеры оптимизации радиоустройств и радиосистем.

Все изложенное в данной книге относится к информационным системам, работающим в диапазоне не только радиоволн, но и в других диапазонах электромагнитных волн, а также к связным, акустоэлектронным и комбинированным системам.

Содержание книги базируется на материалах, имеющихся в опубликованной литературе, и материалах преддипломного курса, читаемого автором для студентов радиофакультета.

Данное пособие рассчитано в первую очередь на студентов старших курсов и дипломников, уже знакомых с остальными дисциплинами своей специальности. Оно может быть также полезно студентам при выполнении ими курсовых проектов и учебно-исследовательских работ и при углубленном изучении читаемых на этих курсах инженерных дисциплин.

Радиоэлектронные системы (РС) делятся на информационные и энергетические. Информационные РС предназначены для передачи, извлечения и преобразования информации (сообщений), т. е. для информационных целей. Радиосистемы, предназначенные для других целей, называются неинформационными или энергетическими. В данной книге рассматриваются лишь информационные РС, поэтому термин «информационные» в дальнейшем опускается.

По своему информационному назначению радиосистемы делятся на следующие основные классы [1]: системы передачи информации (СПИ); системы извлечения информации (СИИ); системы разрушения информации; комбинированные системы; системы радиоуправления.

Системы передачи информации предназначены для передачи информации (сообщений) из одних пунктов пространства в другие. К ним относятся, в частности, системы радиовещания и телевизионного вещания, системы связи, радиотелеметрические системы и системы передачи команд.

Системы извлечения информации предназначены для извлечения информации об объектах или процессах с помощью радиосредств. К ним относятся системы локации, астрономии, разведки природных ресурсов, измерения параметров радиотехнических сигналов и цепей и др.

Системы радиопротиводействия создают помехи системам передачи или извлечения информации для разрушения циркулирующей в них информации.

Комбинированной называется система, в состав которой входит две или более радиосистемы различных классов (например, СПИ и СИИ).

Системой радиоуправления называется система управления, в которой радиосредства выполняют основную функцию. Главной областью применения этих систем является управление полетом самолетов, ракет, космических и других летательных аппаратов. Радиосредства в таких системах применяются для извлечения и передачи информации. Система радиоуправления является обычно системой автоматического (или автоматизированного) управления с обратной связью, и радиосредства извлечения и передачи информации можно рассматривать как звенья этой замкнутой системы, входящие в нее наряду с другими звеньями (звеном «автопилот», кинематическими звеньями и др.).

Во всех упомянутых системах, за исключением систем радиопротиводействия, радиосредства предназначены для правильного воспроизведения информации (сообщений). Задачей же систем радиопротиводействия является разрушение информации; поэтому подход к проектированию таких систем имеет существенные особенности. В дальнейшем системы радиопротиводействия рассматриваться не будут и, следовательно, основными классами ра-

диосистем будут считаться радиоэлектронные системы передачи и извлечения информации, работающие автономно или в составе систем управления.

Основными из многочисленных требований, предъявляемых к таким системам, являются:

- точность (достоверность) воспроизведения сообщений;
- помехозащищенность;
- пропускная способность;
- электромагнитная совместимость (ЭМС);
- экологическая совместимость;
- надежность; малые масса и габаритные размеры;
- возможно меньшая стоимость.

Все эти требования характеризуются количественно соответствующими показателями (показателями точности, помехозащищенности, надежности и т. д.) $k_1, k_2, \dots, k_m$, называемыми показателями качества. Упорядоченная совокупность

$$K = \langle k_1, \dots, k_m \rangle \quad (1.1)$$

показателей качества образует вектор K качества системы. Задача проектирования системы состоит в выборе такого варианта S ее построения, который удовлетворяет всей совокупности исходных данных, включающей условия работы системы, ограничения на ее структуру и значения параметров и требования к вектору качества K . Как будет показано далее (§ 8.1), показатели качества $k_1, \dots, k_m$ целесообразно выбирать такими, чтобы каждый из них удовлетворял условию: чем меньше данный показатель k_i , тем лучше система при прочих равных условиях, т. е. при неизменных значениях всех остальных $(m-1)$ показателей. При этом один из наиболее распространенных критериев качества сводится к следующему: обеспечить минимальное значение одного из показателей (например, стоимости системы k_1) при условии, что все остальные показатели ($k_2, \dots, k_m$) не превосходят допустимых значений.

Выбор состава вектора K является одной из самых сложных и ответственных задач проектирования. Это объясняется тем, что степень удовлетворения системой даже какого-то одного требования (например, помехозащищенности) невозможно полно и достаточно охарактеризовать каким-то одним числом или даже совокупностью конечного набора чисел. Основные пути преодоления этой трудности рассматриваются в следующих главах. Общим критерием, которым при этом приходится руководствоваться (как и при всяком переходе от бесконечно сложной реальной ситуации к ее формализованному описанию), является отыскание компромисса между полнотой и точностью описания, с одной стороны, и простотой использования этого описания для проектирования, с другой. Например, для сравнительно полной и точной оценки качества системы нужно характеризовать совокупностью нескольких сотен или даже тысяч показателей $k_1, \dots, k_m$. Но выбрать оптимальный вариант построения системы по столь боль-

шому числу показателей практически невозможно даже при использовании наиболее совершенных ЭВМ. Поэтому практически на начальных этапах проектирования обычно учитывают лишь 10...20 наиболее важных показателей и увеличивают число учитываемых показателей до нескольких десятков лишь на последующих этапах.

Обоснование исходных данных производится исходя из назначения проектируемой системы и называется *внешним проектированием*. В отличие от этого выбор структуры и значений параметров системы, удовлетворяющих всей совокупности исходных данных, называется *внутренним проектированием* или *синтезом* системы.

Методы проектирования существенно зависят от назначения проектируемой системы, в первую очередь от того, какой системой она является: передачи или извлечения информации, гражданского или военного назначения. Однако даже для систем одного и того же класса (например, радиотехнических систем передачи информации гражданского назначения) исходные данные могут весьма существенно различаться — по требуемой дальности действия (в 10^5 раз и более), диапазону используемых электромагнитных волн (в 10^8 раз и более), пропускной способности (в 10^9 и более), массе (в 10^3 и более), стоимости (в 10^9 и более), массовости выпуска (в 10^7 и более), сроку службы (в 10^7 и более) и другим требованиям.

Такой огромный диапазон исходных данных делает невозможным создание единой теории проектирования, в одинаковой мере справедливой и эффективной для радиосистем различных классов и назначений. Для каждого класса или даже подкласса радиосистем должна разрабатываться и действительно разрабатывается соответствующая теория проектирования. Например, в последние годы успешно разрабатываются вопросы проектирования спутниковых систем связи, телеметрических, радиолокационных и др.

Однако имеется целый ряд важных вопросов проектирования, которые являются в значительной мере общими для информационных систем самых различных классов и назначений. Действительно, для всех радиосистем в той или иной мере существенны перечисленные выше восемь основных требований и улучшение соответствующих им показателей. Поэтому весьма важно установить, какими числовыми показателями следует характеризовать степень удовлетворения каждого такого требования (например, требования точности воспроизведения сообщений), от каких факторов зависит данный показатель качества и каковы основные пути его улучшения. Общей является также проблема сведения задачи векторного синтеза к скалярному. В самом деле, качество системы характеризуется не одним показателем качества, а совокупностью (вектором) показателей (1.1). Некоторые из этих показателей могут быть лучше у одного варианта (S') построения системы, а остальные — у другого (S''). В таких случаях вы-

бор лучшего (оптимального) варианта является далеко не тривиальной задачей.

Общими для различных классов радиосистем являются и такие важные вопросы проектирования, как необходимость системного подхода, повышения эффективности математических, экспериментальных и эвристических процедур и др. В данной книге рассматриваются только общие вопросы проектирования. Однако даже ограничение наиболее общими вопросами не позволяет изложить их в небольшой книге достаточно полно. Наконец, следует учесть, что теория проектирования находится в начальной стадии развития. По всем этим причинам изложенный в данной книге материал можно рассматривать лишь как введение в теорию проектирования радиосистем различных классов.

Почти все отмеченные выше и многие другие общие вопросы, рассматриваемые в книге, весьма существенны при проектировании не только радиосистем различных классов, но и их частей — радиопередающих, радиоприемных, антенно-фидерных, электронновычислительных и других устройств. Поэтому изложенный далее материал относится в значительной мере не только к радиосистемам, но и к радиоустройствам.

Для облегчения понимания последующего материала приведем основные понятия и сокращения, принятые в данной книге.

Основные понятия. *Множество* — совокупность элементов некоторой природы (людей, систем, точек, понятий и др.).

Система — совокупность взаимосвязанных элементов (техническая система, биологическая, общественная, система обозначений, мер, взглядов и др.). К основным видам связи между элементами системы принадлежат функциональные (обусловленные общностью назначения элементов) и динамические (проявляющиеся во взаимодействии процессов и явлений в различных элементах) связи.

Радиосистема (радиоэлектронная система) — техническая система, в которой радиосредства (радиоаппаратура) выполняют основные функции.

Радиосредства (радиоаппаратура) — радиоэлектронные средства (радиоэлектронная аппаратура) — любые технические средства (аппаратура), относящиеся к радиоэлектронике.

Подсистема — часть системы. Подсистемой может являться система более низкого (по сложности) иерархического уровня.

Радиоустройство (радиоэлектронное устройство) — радиоэлектронная часть системы.

Класс систем — совокупность (множество) систем, имеющих как общие признаки (отличающие данный класс от другого класса), так и различные признаки (которыми различаются системы одного и того же класса). Например, к классу радиосистем передачи информации относятся любые радиосистемы, предназначенные для передачи информации.

Реальная система — реально существующая, т. е. уже изготовленная система.

Модель системы — образ (описание) системы, находящейся в стадии проектирования, т. е. еще не существующей в законченном виде. Иначе говоря, модель системы — это то, что понимается под системой в процессе ее проектирования. Чем ближе этап проектирования к конечному, тем полнее модель системы и тем в большей мере она соответствует будущей реальной системе.

Радиокомплекс — сложная часть радиосистемы, содержащая радиосистемы более низкого иерархического уровня и радиоустройства. Например, спутниковая система связи, обеспечивающая связь между наземными пунктами, состоит из наземного радиокомплекса и бортового радиокомплекса (расположенного на искусственных спутниках Земли).

Информация — совокупность сведений о процессах и явлениях в некотором объекте (субъекте), имеющих значение (или могущих иметь значение в будущем) для процессов и явлений в том же самом или в другом объекте (субъекте).

Сообщения — единство сведений (информации) и их первичных физических носителей (электрического тока, акустического колебания, печатного текста и др.). Например, сообщениями являются печатный текст, электрический ток, содержащие некоторые сведения (информацию).

Сигнал первичный — сообщение, в котором носителем является электрический ток (напряжение).

Радиосигнал — напряжение (ток, электромагнитное или акустическое поле) высокой частоты, модулированное переносимым этим сигналом сообщением. Здесь под высокой понимается частота f_0 , существенно (в несколько раз или более) превышающая высшую частоту F_n спектра переносимого сообщения; F_n — высшая частота, за пределами которой (в сторону превышения) энергия спектра сообщения пренебрежимо мала.

Исходный показатель качества — любая числовая характеристика k_{in} системы (устройства), монотонно связанная с ее качеством: чем больше (чем меньше) k_{in} , тем лучше система (устройство) при прочих равных условиях.

Стандартный показатель качества — нормированный (безразмерный) показатель качества $k_t = k_{in}/k_{in0}$, удовлетворяющий следующим условиям:

- а) $k_t > 0$;
- б) чем меньше k_t , тем лучше система при прочих равных условиях;
- в) у идеальной (по данному показателю) системы $k_t \rightarrow 0$. Здесь k_{in0} — некоторое опорное (эталонное) значение показателя k_{in} .

Вектор качества $K = \langle k_1, \dots, k_i, \dots, k_m \rangle$ — упорядоченная совокупность стандартных показателей качества, учитываемых на данном этапе проектирования.

Глава 1

ОБЩАЯ ХАРАКТЕРИСТИКА ЗАДАЧ И МЕТОДОВ ПРОЕКТИРОВАНИЯ¹

1.1. ОСНОВНЫЕ ВИДЫ ЗАДАЧ ПРОЕКТИРОВАНИЯ

По степени новизны проектируемой системы различают следующие типичные задачи проектирования:

частичная модернизация существующей системы — изменение ее параметров (а иногда и структуры), вызывающее сравнительно небольшое улучшение одного или нескольких показателей качества;

существенная модернизация — изменение параметров и структуры системы, вызывающее значительное улучшение одного или нескольких основных показателей качества;

создание новой системы, основанной на новых принципах действия для резкого (например, на порядок или более) улучшения одного или нескольких основных показателей качества при решении тех же или новых задач.

По степени сложности задачи проектирования можно разделить на простые, средней сложности, сложные и весьма сложные. К простым будем относить такие задачи, для решения которых достаточен коллектив инженеров не более 10 человек. Решение задачи средней сложности требует нескольких десятков или сотен инженеров, а для решения сложных и весьма сложных задач необходимы несколько тысяч или десятков тысяч инженеров. Конечно, такое деление условно, но оно показывает, сколь велик диапазон сложности задач проектирования. Например, задача создания системы «Аполлон» для полета астронавтов на Луну была весьма сложной: в ее решении участвовали десятки тысяч инженеров различных профилей, сосредоточенные в 300 фирмах. При этом большая часть из них занималась разработкой системы радиоуправления кораблем «Аполлон».

Следует отметить, что в последние годы процент задач проектирования новых систем существенно возрос.

Очевидно, что организация процесса проектирования и применяемые методы сильно зависят от степени новизны и сложности решаемой задачи. В дальнейшем, там где это не оговаривается, предполагается, что речь идет о существенной модернизации системы или о создании новой системы и задача проектирования имеет среднюю или большую сложность.

¹ Эта глава написана в основном по книге [1] автора.

1.2. ОСНОВНЫЕ ЭТАПЫ ПРОЕКТИРОВАНИЯ

Различают следующие основные этапы проектирования: предварительное, эскизное, техническое.

Результатом первого этапа являются технические предложения (аван-проект) (сюда также относятся теоретические и экспериментальные НИР), а результатами второго и третьего — эскизный и технический проекты соответственно. Каждый этап включает как теоретические, так и экспериментальные исследования и завершается испытаниями лабораторных макетов (первый этап), экспериментальных (второй) и опытных (третий) образцов.

С точки зрения решаемых задач процесс проектирования можно разбить на следующие этапы:

- выбор и формулировка цели проектирования;
- обоснование исходных данных;
- определение принципов построения системы;
- аппаратурный синтез;
- конструирование;
- разработка технологии изготовления;
- разработка испытательной аппаратуры.

Под выбором и формулировкой цели понимают словесную (качественную) формулировку задачи проектирования: какая система и для решения каких задач должна быть разработана? Для этого должны быть предварительно проанализированы потребности общества (или какой-либо его части), возможности, предоставляемые современным уровнем развития науки и техники вообще и радиоэлектроники в частности, и учтены результаты прогнозирования дальнейшего развития радиоэлектронных и близких к ним (по структуре или по решаемым задачам) систем. Инициатива в постановке задачи проектирования может принадлежать как заказчикам, так и разработчикам.

На втором этапе проектирования дается количественная формулировка задачи проектирования, т. е. обоснование исходных данных — условий работы системы; ограничений, накладываемых на структуру системы и значения ее параметров; вида учитываемых показателей качества и предъявляемых к ним требований.

При обосновании исходных данных необходимо учесть назначение системы и все основные виды ее взаимодействия с окружающей средой, в частности с другими подсистемами (если данная система является подсистемой более крупной системы) и с другими системами, работающими в том же или соседнем диапазоне частот. Иногда обоснование исходных данных называют внешним проектированием, а все последующие этапы — внутренним. Обоснование исходных данных обычно производится совместными усилиями разработчиков, заказчиков и потребителей проектируемой системы и включает необходимые патентные исследования.

На третьем этапе проектирования (первом этапе внутреннего проектирования) определяют принципы построения системы. В результате должно быть обосновано деление проектируемой систе-

мы на подсистемы (устройства), сформулированы исходные данные для всех подсистем и определены (по меньшей мере в первом приближении) принципы их действия. На этом этапе выбирают рабочий диапазон волн и принципы извлечения, передачи и обработки информации, включая используемые физические явления, вид сигнала — переносчика информации, принципы пространственных и временных преобразований сигналов, объединения и разделения каналов и станций, модуляции и демодуляции, кодирования и декодирования (при проектировании некоторых систем может потребоваться решение лишь части перечисленных выше задач). Третий этап проектирования позволяет сравнительно точно определить значения лишь информационных показателей качества (точности, разрешающей способности, помехоустойчивости, пропускной способности и др.), а экономические и конструктивно-технологические показатели (стоимость, масса, объем, надежность и др.) оценивают весьма приближенно.

Задачей четвертого этапа (аппаратурного синтеза) является аппаратурная реализация выбранных на предыдущем этапе принципов построения системы. При этом выбирают элементную базу и оптимизируют структуру и параметры подсистем с более точным учетом не только информационных, но и всех остальных показателей качества, включая конструктивно-технологические и экономические. Одними из задач этого этапа являются также обоснование методики проведения испытаний, выбор состава измерительных средств и разработка нестандартной испытательно-измерительной аппаратуры.

На пятом этапе (конструирование) оптимизируют систему по ее конструктивно-технологическим и экономическим показателям, окончательно определяют эти показатели и составляют всю техническую документацию, необходимую для изготовления опытных образцов, а затем и для серийного или массового производства.

Шестой этап проектирования тесно связан с четвертым и пятым и включает в себя разработку технологии изготовления опытных, а затем и промышленных образцов. При этом создание технологии не может быть завершено, пока не будет решен полностью вопрос о конструкции системы, а определение окончательной конструкции и составление всей технической документации не могут быть выполнены до окончания разработки технологии. Поэтому пятый и шестой этапы проектирования проводят и завершают практически в одно и то же время и часто объединяют единым названием — этап конструирования и технологии. На этом этапе, как правило, необходимо тесное взаимодействие разработчиков системы с заводами-изготовителями образцов.

В меньшей степени, но также достаточно тесно связаны (в прямом и обратном направлениях) и все остальные (даже несмежные) этапы проектирования. В частности, в результате завершения этапов конструирования и технологии может потребоваться коррекция не только принципов построения системы, но и исходных данных и даже изменение цели проектирования (на-

пример, если окажется, что полученные экономические и конструктивно-технологические показатели системы не обеспечивают выполнения поставленных перед нею задач). Поэтому процесс проектирования является не только многоэтапным, но и многократно корректируемым по мере его выполнения.

Необходимо также подчеркнуть, что на всех этапах проектирования системы следует учитывать особенности ее производства, эксплуатации и утилизации. В противном случае разработанная система может оказаться непригодной или по меньшей мере неоптимальной.

Первые три этапа проектирования часто называют системотехническими, так как на этих этапах части системы (подсистемы) подробно не рассматриваются и главное внимание уделяется функционированию и характеристикам системы в целом. Этап аппаратного синтеза иногда называют схемотехническим, так как на этом этапе обосновывается принципиальная схема системы (на основе детального рассмотрения подсистем). Следовательно, процесс проектирования можно разбить на три больших этапа: системотехнический, схемотехнический и конструкторско-технологический. Соответственно инженеров, осуществляющих проектирование, делят на системотехников, схемотехников и конструкторов-технологов. Однако, хотя задачи, решаемые конструкторами и технологами, весьма тесно связаны, они имеют и существенные различия: конструктор должен сообщить технологу, что будет представлять собой разрабатываемая система, а технолог должен определить, как эту систему можно изготовить. Поэтому часто более справедливым оказывается деление инженеров, осуществляющих разработку радиосистем, на четыре группы: системотехников, схемотехников, конструкторов и технологов.

Следует отметить, что при проектировании интегральных микросхем и СВЧ блоков взаимодействие между инженерами-схемотехниками, конструкторами и технологами оказывается настолько тесным, что четко разграничить этапы аппаратного синтеза, конструирования и разработки технологии изготовления таких устройств невозможно.

1.3. ОРГАНИЗАЦИЯ ПРОЕКТИРОВАНИЯ

Организация проектирования в большой мере зависит от степени сложности системы. Однако обычно коллектив инженеров формируется для разработки не одной, а целого класса систем, близких по структуре или виду решаемых задач, с различными сроками окончания работы (для обеспечения непрерывной занятости всего коллектива). Проектирование сложной радиосистемы в целом поручается главному НИИ во главе с генеральным конструктором системы. Главной НИИ поручает разработку подсистем, образующих проектируемую систему, коллективам отраслевых НИИ во главе с главными конструкторами (ГК) подсистем. Кроме НИИ в проектировании системы принимают учас-

тие смежные организации (опытные и серийные заводы, учреждения Академии наук и вузы). При этом осуществляется тесное взаимодействие разработчиков аппаратуры с заказчиками. Так как в проектировании участвует много людей, использующих различные материальные средства (печатные и рукописные материалы, вычислительные машины, экспериментальные и заводские стенды, оборудование, приборы, материалы, детали и узлы и др.), управление этим процессом является весьма сложным и должно осуществляться на базе АСУ.

В каждом НИИ (головном или отраслевом) в разработке радиосистем и их частей принимают участие инженеры следующих основных специальностей:

- 1) инженеры-системотехники общего профиля;
- 2) радиоинженеры-системотехники;
- 3) радиоинженеры-схемотехники;
- 4) инженеры-конструкторы;
- 5) инженеры-технологи;
- 6) инженеры-математики (или математики-прикладники);
- 7) инженеры-экономисты и плановики.

Инженеры этих специальностей концентрируются в соответствующих отделах НИИ: теоретическом, комплексном, отраслевых, конструкторском, технологическом, математического обеспечения (включающем или дополняющем вычислительный центр) и планово-экономическом. Процент инженеров той или иной специальности в общем числе инженеров-разработчиков зависит от вида разрабатываемой системы, степени ее новизны и сложности. Однако в среднем наиболее крупными и примерно одинаковыми в процентном отношении являются специальности 3, 4, 5 и 6. Затем в порядке убывания их процентного веса идут специальности 2, 7 и 1. При этом радиоинженеры обычно составляют 30—40% от общего числа инженеров-разработчиков.

В процессе проектирования осуществляется тесное взаимодействие инженеров всех (или почти всех) указанных специальностей.

1.4. ОБЩАЯ ХАРАКТЕРИСТИКА МЕТОДОВ ПРОЕКТИРОВАНИЯ

В процессе проектирования используются математические, экспериментальные и эвристические методы.

При применении математических методов совокупность $D = (Y, O_s, K, \text{кр. } K)$ исходных данных формулируется математически, т. е. составляется математическое описание условий работы системы Y , ограничений O_s , налагаемых на структуру системы и значения ее параметров, состава s , K вектора качества K и критерия качества $\text{кр. } K$ системы. Затем определяют математическим путем целевые функции, т. е. зависимости частных показателей качества $k_1, \dots, k_m$ от структуры системы S и значений ее параметров при заданных условиях Y . Для полученного таким образом описания отыскивают математическими методами анализа

и синтеза алгоритмы работы и параметры системы, удовлетворяющие выбранному критерию качества. При решении указанной задачи широко используют цифровые и аналоговые ЭВМ для расчета, математического моделирования и автоматизированного проектирования¹.

К расчету относят вычисления по заранее полученным формулам — при фиксированных значениях входящих в них параметров (обычный расчет) и при вариациях параметров для нахождения экстремума функции одной или нескольких переменных (линейное и нелинейное программирование). К расчету могут относиться также некоторые задачи решения уравнений.

При математическом моделировании исходные уравнения, описывающие поведение системы и все приложенные к ней воздействия, приводят к виду, приемлемому для цифровой или аналоговой ЭВМ, и вводят в нее. В эту же ЭВМ вводят все ограничения и критерии качества. На выходе ЭВМ выдаются (в реальном, ускоренном или замедленном масштабе времени) результаты анализа или синтеза. Однако на современном уровне науки и техники решать на ЭВМ задачи синтеза структуры удается только для сравнительно простых радиосистем или лишь для некоторых частей систем.

Автоматизированное проектирование имеет много общего с математическим моделированием: оно проводится также на базе ЭВМ и предназначено для определения оптимальной структуры системы и ее параметров. Наиболее существенные отличия автоматизированного проектирования от математического моделирования на современном этапе состоят в следующем:

основными задачами математического моделирования являются анализ системы и оптимизация ее параметров, а задачей автоматизированного проектирования — синтез структуры, включая определение оптимальных значений параметров;

при автоматизированном проектировании ЭВМ выдает результаты в виде готовой технической документации (совокупности чертежей, таблиц, программ для ЭВМ и других необходимых документов), при математическом моделировании результаты, выдаваемые ЭВМ, не имеют такой степени законченности.

В настоящее время автоматизированное проектирование радиосистем применяется в основном при конструировании и разработке технологии изготовления, в меньшей степени — на схемотехническом уровне (главным образом для разработки интегральных микросхем, цифровых и СВЧ устройств) и в еще меньшей — на системотехническом уровне. Математическое моделирование применяют в основном на системотехническом уровне. Отсюда следует, что в настоящее время математическое моделирование и автоматизированное проектирование дополняют друг друга. Однако по мере развития науки и техники различия между ними будут умень-

шаться: автоматизированное проектирование будет еще больше внедряться и на уровне системотехники, а математическое моделирование по мере расширения возможностей ЭВМ сможет обеспечивать и синтез структуры, и изготовление технической документации. Уже в настоящее время математическое моделирование часто рассматривают как составную часть процесса автоматизированного проектирования.

Математические методы, включая расчеты на ЭВМ, математическое моделирование и автоматизированное проектирование, являются весьма мощным инструментом проектирования. Однако они предполагают наличие вполне определенного математического описания, которое, во-первых, требует экспериментальной проверки (ибо основным критерием истины является практика) и, во-вторых, существует далеко не на всех этапах проектирования. Для выбора и обоснования математического описания, а также для решения ряда дополнительных важных задач проектирования требуется эвристическая деятельность коллектора разработчиков, т. е. творческая деятельность, не поддающаяся математической формализации. Таким образом, математические исследования должны дополняться экспериментальными исследованиями и эвристической деятельностью разработчиков.

Различают следующие основные виды экспериментальных исследований: полунатурное моделирование, лабораторные исследования, полевые испытания, летные испытания, пробные пуски, испытания в эксплуатации.

Для ракетной и космической аппаратуры применяют все указанные виды исследований и испытаний, для аппаратуры другого назначения некоторые виды испытаний могут отсутствовать.

Полунатурное моделирование отличается от математического лишь тем, что часть звеньев включают в состав модели в виде натуральных макетов, а не моделируют на ЭВМ. Под лабораторными исследованиями понимают исследования натуральных макетов, проводимые в лабораториях, причем обычно реальные источники сигналов и внешних помех заменяют имитаторами, построенными на основе математических моделей этих сигналов и помех (генераторами сигналов, шума и т. д.) Сказанное означает, что как полунатурное моделирование, так и лабораторные исследования являются не чисто экспериментальными, а экспериментально-теоретическими.

При полевых испытаниях аппаратуру испытывают не в закрытых лабораторных помещениях, а в полевых условиях. При этом оказывается возможным заменить все или часть имитаторов сигналов и внешних помех реальными (натурными) источниками. Еще большее приближение к реальным условиям (при испытаниях авиационной или ракетной и космической аппаратуры) дают летные испытания, при которых бортовая часть системы размещается на самолете (обычно специально оборудованном). Однако при испытаниях ракетной и космической аппаратуры даже такие испытания не являются еще чисто экспериментальными, так

¹ Система автоматизированного проектирования может включать подсистему автоматического проектирования.

как при этом часть звеньев системы (кинематические, аэродинамические, а иногда и некоторые другие) имитируют (моделируют), реализуя на ЭВМ. При испытаниях ракет и космических аппаратов полностью экспериментальными можно считать лишь испытания, проводимые при пробных пусках и в процессе эксплуатации.

В процессе проектирования приходится решать три основных вида задач:

1. Синтез оптимальных алгоритмов преобразования информации (пространственно-временной обработки сообщений, сигналов и псмех).

2. Выбор оптимальных значений параметров системы (длины волны, занимаемой полосы частот, мощности излучения и т. п.).

3. Сравнение различных вариантов построения системы с целью выбора наилучшего варианта.

Для краткости задачу 1 часто называют *синтезом структуры* (алгоритмов), задачу 2 — *оптимизацией параметров* (параметрической оптимизацией), а задачу 3 — *дискретным выбором* системы (варианта ее построения). Обычно решение задач 1 и 2 предшествует дискретному выбору системы, так как перед сравнением вариантов системы нужно предварительно найти для каждого из этих вариантов оптимальные структуру и значения параметров. Но, с другой стороны, проводить синтез структуры и оптимизацию параметров для весьма большого числа вариантов практически невозможно — необходимо сначала хотя бы на основе имеющихся априорных данных исключить из рассмотрения часть заведомо неподходящих вариантов. Поэтому последовательность решения задач 1, 2 и 3 в процессе проектирования может изменяться.

Эвристический подход в процессе проектирования необходим при решении следующих задач:

- выбор и формулировка цели проектирования;
- выбор физических принципов действия системы;
- обоснование математических моделей системы, полезных и мешающих воздействий;
- выбор методов математического и экспериментального исследования;
- выбор элементной базы (при отсутствии жестких ограничений);

трактовка результатов исследования и принятие окончательных решений.

Эвристическая деятельность опирается на имеющийся опыт в разработке подобных систем или решений подобных задач и на результаты теоретических и экспериментальных исследований, проводимых в процессе проектирования.

Математические методы анализа и синтеза радиосистем и радиоустройств весьма разнообразны, рассматриваются практически во всех инженерных курсах, изучаемых студентами вузов, и

излагаются в учебных пособиях по этим курсам. Краткая общая характеристика этих методов дана в [1]. В данной книге рассматриваются следующие вопросы: методы определения и оптимизации показателей пропускной способности, ЭМС, ЭКС, надежности и стоимости; учет неопределенности исходных данных при определении и оптимизации показателей точности и помехоустойчивости и некоторые другие; системный подход к проектированию как важнейшая основа повышения эффективности эвристических решений.

1.5. СИСТЕМНЫЙ ПОДХОД КАК ОСНОВА ПРОЕКТИРОВАНИЯ

Оптимальное решение задачи проектирования возможно лишь на основе всестороннего, целостного рассмотрения проектируемой системы и ее развития (изменения) в процессе взаимодействия с окружающей средой. Лишь такой подход, называемый *системным*, способен привести к подлинно творческим, новаторским решениям, вплоть до крупных изобретений и научных открытий.

В основе системного подхода лежат следующие основные положения (принципы):

1. Учет всех этапов «жизненного цикла» разрабатываемой системы: проектирования, производства, эксплуатации и утилизации. Забвение этого принципа часто приводит к грубым просчетам. Например, проекты некоторых систем, в основу которых были положены весьма прогрессивные принципы их действия, остались нереализованными только потому, что оказались недостаточно технологичными: слишком трудоемкими и непригодными с точки зрения их производства.

2. Учет истории и особенно перспектив развития систем данного и близких классов. Историю важно учитывать потому, что системы, в прошлом призванные непригодными или устаревшими, могут в новых условиях, на новом уровне развития науки и техники стать (при соответствующей коррекции) очень хорошими и перспективными. Например, полупроводниковые приборы, появившиеся впервые более 50 лет назад, были вскоре признаны малопригодными и вытеснены электронными лампами. Но спустя примерно 30 лет они произвели подлинную революцию в радиоэлектронике и в значительной мере вытеснили электронные лампы.

Учет при проектировании прогноза развития систем данного и других взаимодействующих или конкурирующих классов необходим потому, что в противном случае разрабатываемая система может оказаться морально устаревшей вскоре после завершения разработки или даже до ее окончания.

3. Всестороннее рассмотрение взаимодействия системы с внешней средой. Основными видами такого взаимодействия являются: взаимодействие с природой и обществом в целом (учет экологических, экономических, социальных, политических, военных и других факторов);

обмен полезной информацией — получение (извлечение) и выдача ее;

обмен энергией и веществом;

обмен радиопомехи (т. е. помехами от электромагнитного излучения);

внешние воздействия на систему (изменение температуры, влажности, давления, механические нагрузки, радиация и т. д.);

взаимодействие с другими системами (входящими в общую систему более высокого иерархического уровня) в процессе решения общей задачи.

4. Учет основных видов взаимодействия внутри системы (между ее частями): функционального, конструктивного, динамического, информационного, энергетического и др.

5. Учет взаимодействия между элементной базой и системотехникой. Развитие элементной базы вызывает развитие системотехники (появление новых принципов построения систем и улучшение их показателей качества); в свою очередь, развитие системотехники предъявляет новые требования к элементной базе и стимулирует ее развитие.

Следует подчеркнуть, что все революционные этапы в развитии радиосистем были вызваны революционными изменениями элементной базы: появлением пролетных СВЧ радиоламп, полупроводниковых приборов, интегральных микросхем, микропроцессоров, оптоэлектронных приборов и т. д. Поэтому темпы развития радиосистем зависят от того, насколько полно используются в системотехнике новейшие достижения в области элементной базы и в какой мере возможности элементной базы удовлетворяют требованиям, выдвигаемым системотехникой. Важно также отметить, что все революционные изменения в элементной базе были обусловлены использованием процессов, происходящих в тех или иных физических телах на микроуровне, т. е. на уровне элементарных частиц. Иначе говоря, успешное развитие радиосистем в значительной мере определяется умелым использованием разработчиками явлений, происходящих на микроуровне. Поэтому разработчики, рассматривающие проектируемую систему только как совокупность «черных ящиков», могут лишиться возможности создать радикально лучшую систему.

6. Учет возможности изменения исходных данных и даже решаемой задачи в процессе проектирования, производства и эксплуатации системы. Отсюда вытекает необходимость:

вариации исходных данных (включая критерий качества) в процессе проектирования для оценки степени их критичности и получения более надежных результатов проектирования;

обеспечения возможно большей универсальности применения проектируемой системы, чтобы при изъятии или добавлении некоторых блоков система была пригодной для решения новых (других) задач.

7. Выделение главных показателей качества, подлежащих улучшению в первую очередь. Стремление улучшить возможно

большее число показателей качества (особенно на ранних этапах проектирования) может привести к потере лучшего решения, не говоря уже об излишнем увеличении длительности проектирования. В большинстве практических задач к главным относят показатели точности, помехоустойчивости, надежности, пропускной способности, массу (объем) и стоимость.

8. Сочетание принципов композиции, декомпозиции и иерархичности. Радиосистема может содержать десятки тысяч или даже миллионы элементов. Очевидно, невозможно одновременно оптимизировать все элементы даже с помощью самых совершенных ЭВМ. Нужно объединить эти элементы в блоки, устройства, подсистемы и рассматривать каждый блок (устройство, подсистему) как единое целое, т. е. производить *композицию элементов*. Однако обычно не удается сколь-нибудь сложную систему рассматривать как один блок, приходится разбивать ее на ряд подсистем, т. е. производить *декомпозицию*. Декомпозиция особенно необходима, если разработку отдельных подсистем приходится поручать различным организациям. В результате композиции и декомпозиции систему разбивают на ряд иерархических уровней, каждый из которых может содержать ряд частей (подсистем, устройств, блоков и т. д.).

Сочетание композиции, декомпозиции и иерархичности позволяет существенно упростить не только проектирование системы, но и ее производство, эксплуатацию и утилизацию. Однако при декомпозиции системы на подсистемы (устройства, блоки) необходимо формулировать исходные данные для проектирования каждой подсистемы так, чтобы они достаточно полно и правильно учитывали все основные виды взаимодействия между подсистемами в процессе выполнения системой поставленной перед ней задачи. Особое внимание при этом следует уделять обоснованию критерия качества каждой подсистемы. В противном случае подсистема, оптимальная в смысле ее критерия качества, может оказаться далеко не оптимальной или вообще неприемлемой с точки зрения критерия качества системы в целом.

9. Вскрытие основных технических противоречий, препятствующих улучшению качества системы и ускорению процесса ее разработки, а также отыскание приемов их преодоления (содержащие этого положения подробно раскрыты в [1]).

10. Правильное сочетание различных методов проектирования, в первую очередь математических, эвристических и экспериментальных, а в рамках математических методов — аналитических и проводимых с помощью ЭВМ.

11. Обеспечение должного взаимодействия в процессе проектирования специалистов различных уровней и профилей.

В [1] описывается ряд приемов и рекомендаций, облегчающих реализацию системного подхода.

Литература: *основная* [1, 25, 46], *дополнительная* [4, 58].

Глава 2 ТОЧНОСТЬ ВОСПРОИЗВЕДЕНИЯ СООБЩЕНИЙ

2.1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВОСПРОИЗВОДИМЫХ СООБЩЕНИЙ

Воспроизводимое сообщение x может иметь следующие основные виды:

1. x — дискретная величина, т. е. в фиксированный момент времени x может принимать лишь одно из значений $x_1, x_2, \dots, x_j, x_m$, где m — конечное число. Например, в случае СПИ дискретными являются цифровые сообщения — при передаче цифр в двоичной системе $m=2$, а при применении десятичной системы $m=10$. В радиолокации дискретным сообщением может быть, например, число целей в группе или число типов целей (число «образов»).

2. x — непрерывная величина, постоянная в течение времени воспроизведения T , но могущая иметь любое значение в некоторых пределах

$$x_{\min} \leq x \leq x_{\max}.$$

В системах СПИ такой характер часто имеет телеметрическая информация (данные о температуре, давлении, влажности и других непрерывных параметрах), а в радиолокации — данные о координатах и параметрах движения целей (если время T сравнительно невелико).

3. $x=x(t)$ — непрерывная функция времени (речь, телевизионный сигнал и т. д.).

4. $x=f(X, Y, Z, t)$ — функция времени t и пространственных координат X, Y, Z (например, информация о совокупности движущихся целей или первичная телевизионная информация).

5. $x=\langle x_1(t), \dots, x_N(t) \rangle$ — векторная функция, т. е. упорядоченная совокупность функций времени (или времени и пространственных координат).

В дальнейшем для краткости рассматриваются лишь случаи 1, 2 и 3, когда сообщение является дискретной или непрерывной величиной или непрерывной функцией времени. В месте приема (воспроизведения) сообщение x неизвестно и часто его можно считать случайной величиной (в случаях 1 и 2) или случайной функцией времени (случай 3) с некоторым (априорным) распределением $P(x)$.

Если x — дискретная случайная величина, то $P(x)$ — распределение вероятностей значений $x_1, \dots, x_m$, удовлетворяющее условию нормировки

$$\sum_{j=1}^m P(x_j) = 1. \quad (2.1)$$

Если x — непрерывная случайная величина, то $P(x)$ — плотность вероятности и условие нормировки имеет вид

$$\int_{x_{\min}}^{x_{\max}} P(x) dx = 1. \quad (2.2)$$

Наконец, если $x=x(t)$ — случайная функция времени, спектр которой можно считать ограниченным сверху частотой f_a , то

$$P(x) = P(x_1, \dots, x_j, \dots, x_n) \quad (2.3)$$

— n -мерная плотность вероятности, а $x_j=x(j\Delta t)$ — значение функции $x(t)$ в момент отсчета $t_j=j\Delta t$, где

$$\Delta t = 1/2 f_a, n = T/2 \Delta t = 2 f_a T \gg 1. \quad (2.4)$$

При этом условие нормировки имеет вид

$$\int_{(X)} \dots \int P(x_1, \dots, x_n) dx_1 \dots dx_n = 1,$$

где (X) — область всех возможных значений переменных, по которым производится интегрирование. Распределение $P(x)$ в месте приема может быть известно полностью, частично или вообще неизвестно. В последнем случае обычно полагают это распределение равномерным, т. е. принимают, что

$$P(x) = \text{const}. \quad (2.5)$$

Здесь и далее (2.5) является сокращенной формой записи следующего соотношения:

$$P(x) = \begin{cases} \text{const} & \text{при } x \in X, \\ 0 & \text{вне области } X. \end{cases}$$

Такое допущение может существенно упростить задачу синтеза. Кроме того, в ряде задач равномерное распределение является наименее благоприятным, т. е. соответствует наихудшему возможному случаю.

2.2. ПОКАЗАТЕЛИ ТОЧНОСТИ ВОСПРОИЗВЕДЕНИЯ ДИСКРЕТНЫХ СООБЩЕНИЙ

Пусть сообщение является дискретной случайной величиной и значения этой величины в различных сеансах воспроизведения статистически независимы. Тогда полной характеристикой степени достоверности воспроизведения сообщений является матрица условных вероятностей ошибок вида

$$\|P_{ij}\| = \begin{vmatrix} P_{11} & \dots & P_{1j} & \dots & P_{1m} \\ P_{21} & \dots & P_{2j} & \dots & P_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ P_{m1} & P_{m2} & P_{m3} & \dots & P_{m,m-1} \end{vmatrix}, j \neq i. \quad (2.6)$$

Здесь $P_{ij}=P(x_i|x_j)$, $j \neq i$, — вероятность воспроизведения сообщения x_i при условии, что действительное значение сообщения рав-

но x_j . В матрице (2.6) элементы P_{jj} отсутствуют, так как они соответствуют вероятностям правильного воспроизведения, а не ошибок. Очевидно, число элементов этой матрицы

$$M = m^2 - m = m(m-1). \quad (2.7)$$

При этом достоверность воспроизведения сообщения характеризуется совокупностью M условных вероятностей P_{ij} ошибок ($i = \overline{1, m}, j = \overline{1, m}, j \neq i$), т. е. M показателями качества. Из (2.7) следует, что при $m \geq 3$ оказывается $M \geq 6$ и с ростом m число M быстро возрастает.

Учитывая, что радиосистема должна характеризоваться показателями не только точности, но и помехозащищенности, пропускной способности, надежности и т. д., оценивать точность большим числом показателей обычно неприемлемо. Поэтому при $m > 2$ (а иногда и при $m = 2$) стремятся заменить совокупность (матрицу) M показателей точности (достоверности) некоторым единым результирующим (обобщенным) показателем. Обычно в качестве такого результирующего показателя выбирают взвешенную сумму частных показателей (условных вероятностей P_{ij} ошибок) вида

$$R = \sum_{i=1}^m \sum_{j=1}^m c_{ij} P(x_j) P(x_i|x_j), \quad (2.8)$$

где

$$c_{ij} \begin{cases} = 0 & \text{при } i = j, \\ > 0 & \text{при } i \neq j. \end{cases}$$

Выражение (2.8) можно получить путем следующих рассуждений. Введем некоторую функцию

$$I(x_i, x_j) = c_{ij}, \quad (2.8')$$

назначаящую потери (числа), соответствующие каждой комбинации (x_i, x_j) результата воспроизведения x_i и истинного значения x_j сообщения. Эта функция называется функцией потерь. Очевидно, она должна удовлетворять условиям

$$I(x_i, x_j) \begin{cases} = 0 & \text{при } i = j, \\ > 0 & \text{при } i \neq j. \end{cases} \quad (2.8'')$$

Поскольку аргументы x_i и x_j случайны, потеря $I(x_i, x_j)$ также случайна. Поэтому в качестве меры точности выбирается математическое ожидание потери

$$R = M[I(x_i, x_j)] = \sum_{i=1}^m \sum_{j=1}^m P(x_i, x_j) I(x_i, x_j).$$

С учетом (2.8') это выражение принимает вид, совпадающий с (2.8). Величину R называют *средним риском* (средней потерей).

Из (2.8) следует, что в качестве весов при вероятностях ошибок $P(x_i|x_j)$ выбирают произведения априорных вероятностей

$P(x_j)$ на весовые коэффициенты c_{ij} , учитывающие относительную опасность ошибок того или иного вида. Если все виды ошибок одинаково опасны, то

$$c_{ij} = \begin{cases} 0 & \text{при } i = j, \\ 1 & \text{при } i \neq j \end{cases} \quad (2.9)$$

и формула (2.8) принимает вид

$$R = \sum_{i=1}^m \sum_{j=1}^m P(x_i) P(x_i|x_j), \quad i \neq j. \quad (2.10)$$

Например, при $m = 2$

$$R = P(x_1) P(x_2|x_1) + P(x_2) P(x_1|x_2). \quad (2.11)$$

Из (2.10) и (2.11) нетрудно убедиться, что выражение (2.10) есть не что иное, как *полная вероятность $P_{\text{ош}}$ ошибки* воспроизведения. Следовательно, если ошибки всех видов одинаково опасны, т. е. справедливо (2.9), то средний риск R совпадает с полной вероятностью ошибки:

$$R = P_{\text{ош}} = \sum_{i=1}^m \sum_{j=1}^m P(x_j) P(x_i|x_j), \quad i \neq j. \quad (2.12)$$

Если относительная опасность ошибок различных видов неизвестна, то обычно полагают все ошибки одинаково опасными и, следовательно, характеризуют точность полной вероятностью ошибки.

Выражение (2.12) можно преобразовать к следующему виду, часто более удобному для вычислений:

$$R = P_{\text{ош}} = 1 - \sum_{j=1}^m P(x_j) P(x_j|x_j) = 1 - P_{\text{прав}}, \quad (2.13)$$

где

$$P_{\text{прав}} = \sum_{j=1}^m P(x_j) P(x_j|x_j) \quad (2.14)$$

— полная вероятность правильного воспроизведения сообщений. Из соотношений (2.12) — (2.14) следует, что для вычисления полной вероятности $P_{\text{ош}}$ в общем случае нужно знать априорное распределение $P(x_j)$ сообщений. Исключение составляет случай, когда условные вероятности правильного воспроизведения $P(x_j|x_j)$ оказываются одинаковыми для всех значений сообщения, т. е. когда

$$P(x_j|x_j) = P_{\text{прав.ус}} \quad (2.15)$$

не зависит от номера j . Тогда из (2.13) получаем

$$P_{\text{ош}} = 1 - P_{\text{прав.ус}} \quad (2.16)$$

т. е. вероятность ошибки не зависит от априорного распределения $P(x_j)$.

Если известно, что все значения сообщения равновероятны, то

$$P(x_j) = 1/m, \quad (2.17)$$

и формулы (2.13), (2.14) принимают вид

$$R = P_{\text{ош}} = 1 - P_{\text{прав.ср}}, \quad (2.18)$$

где

$$P_{\text{прав.ср}} = \frac{1}{m} \sum_{j=1}^m P(x_j|x_j) \quad (2.19)$$

— среднее значение условной вероятности $P(x_j|x_j)$ правильного воспроизведения сообщений (при $j=1, m$).

Если распределение $P(x_j)$ неизвестно, то, как уже отмечалось, обычно полагают все значения сообщения равновероятными, т. е. вычисляют полную вероятность ошибки по формулам (2.18), (2.19).

Таким образом, в большинстве случаев удастся, по крайней мере в качестве первого приближения, характеризовать точность (достоверность) воспроизведения дискретных сообщений одним показателем — средним риском R или его частным значением — полной вероятностью $P_{\text{ош}}$ ошибки или полной вероятностью правильного воспроизведения сообщения ($P_{\text{прав.}}$, $P_{\text{прав.ср}}$ или $P_{\text{прав.ус.}}$).

2.3. ПОКАЗАТЕЛИ ТОЧНОСТИ ВОСПРОИЗВЕДЕНИЯ НЕПРЕРЫВНЫХ СООБЩЕНИЙ, ПОСТОЯННЫХ В ПРОЦЕССЕ ИХ ВОСПРОИЗВЕДЕНИЯ

Рассмотрим теперь воспроизведение сообщения, являющегося непрерывной случайной величиной, не изменяющейся в течение времени его воспроизведения. Пусть x^* — результат воспроизведения (в месте приема) сообщения x . Тогда ошибка воспроизведения

$$\Delta x = x^* - x. \quad (2.20)$$

Будем полагать, что значения сообщения x в различных сеансах (опытах) его воспроизведения статистически независимы. Тогда априорное распределение значений сообщения полностью характеризуется одномерной плотностью вероятности $P(x)$, удовлетворяющей условию нормировки (2.2).

Ошибка Δx также является случайной величиной. Для данного образца системы она может принимать случайные значения при переходе от одного сеанса воспроизведения сообщения (длительностью T) к другому, т. е. по ансамблю сеансов. Для данного сеанса (интервала воспроизведения T) она может быть случайной при переходе от одного образца (экземпляра) системы к другому, т. е. по ансамблю образцов системы. Короче говоря, случайность ошибки может проявляться как в пределах ансамбля

сеансов (по сеансам), так и в пределах ансамбля образцов (по образцам).

В системах однократного действия (например, в системах самонаведения) ансамбль сеансов состоит всего из одного сеанса; поэтому имеет смысл рассматривать случайность лишь по образцам. В уникальной системе (например, в уникальном радиотелескопе), наоборот, ансамбль образцов содержит всего один образец и, следовательно, можно рассматривать лишь ансамбль сеансов. Однако многие радиосистемы не уникальны и предназначены для многократного действия. Поэтому в них случайность ошибки Δx вызывается как факторами, проявляющимися при переходе от одного образца к другому (например, разбросом параметров системы при переходе от одного образца к другому), так и по ансамблю сеансов (например, из-за действия помех и нестабильностей параметров).

Некоторые источники ошибок обладают свойством эргодичности — статистические характеристики, определенные по ансамблю образцов, совпадают со статистическими характеристиками, определяемыми для данного образца по ансамблю сеансов (например, таким свойством обладают ошибки, вызванные стационарным действием внутреннего шума приемных устройств).

Из сказанного следует, что в некоторых случаях достаточно учитывать случайность ошибки Δx только по сеансам (или только по образцам); однако в общем случае необходимо учитывать оба вида источников случайности.

С учетом сделанных замечаний рассмотрим основные показатели точности воспроизведения непрерывных сообщений. Как и в случае дискретных сообщений, точность можно характеризовать средним риском R , определяемым соотношением

$$R = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, x^*) P(x) P(x^*|x) dx dx^*. \quad (2.21)$$

Здесь $I(x, x^*)$ — функция потерь (риска), удовлетворяющая условию [аналогичному (2.8'')]

$$I(x, x^*) \begin{cases} = 0 & \text{при } x^* = x, \\ > 0 & \text{при } x^* \neq x, \end{cases} \quad (2.22)$$

$P(x^*|x)$ — плотность вероятности воспроизведения сообщения x^* при условии, что действительное значение сообщения равно x .

Сравнивая выражения (2.21), (2.22) с соответствующим выражением (2.8) для дискретных сообщений, нетрудно убедиться, что они в значительной мере аналогичны: при переходе от дискретных сообщений к непрерывным двойная сумма (по i и j) заменяется двойным интегралом (по x и x^*), функция потерь c_{ij} , дискретно зависящая от номеров i и j , превращается в непрерывную функцию потерь $I(x, x^*)$, а вероятности $P(x_i)$ и $P(x_i|x_j)$ переходят в соответствующие плотности вероятности $P(x)$ и $P\{x^*|x\}$.

Из (2.21) следует, что средний риск R является математическим ожиданием потери:

$$R = M [I(x, x^*)]. \quad (2.23)$$

Функция потерь назначает потери, соответствующие различным комбинациям истинного значения сообщения x и результата его воспроизведения x^* , т. е. учитывает опасность появления тех или иных отклонений x^* от x . Наиболее часто в качестве функции потерь выбирают квадратичную функцию

$$I(x, x^*) = (x - x^*)^2 = (\Delta x)^2. \quad (2.24)$$

При этом выражения (2.21) и (2.23) принимают вид

$$R = M [(\Delta x)^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x^*)^2 P(x) P(x^*|x) dx dx^*, \quad (2.25)$$

т. е. средний риск совпадает со средним квадратом ошибки воспроизведения сообщений.

Удобство функции потерь вида (2.24) состоит в том, что она, с одной стороны, обычно удовлетворительно соответствует требованиям, предъявляемым к радиосистеме, а с другой — в большинстве случаев оказывается удобной при математическом анализе и синтезе системы.

Из (2.21) и (2.25) следует, что для вычисления среднего риска R (или его частного вида — среднего квадрата ошибки) необходимо знать априорное распределение $P(x)$ сообщений. Если $P(x)$ неизвестно, то вместо среднего риска R точность приходится характеризовать условным риском R_x , определяемым для каждого данного значения сообщения x :

$$R_x = \int_{-\infty}^{\infty} I(x, x^*) P(x^*|x) dx^*, \quad (2.26)$$

При квадратичной функции потерь (2.24) это выражение принимает вид

$$R_x = M_x [(\Delta x)^2] = \int_{-\infty}^{\infty} (x - x^*)^2 P(x^*|x) dx^*. \quad (2.26')$$

Очевидно, условный риск R_x является в общем случае некоторой функцией от x :

$$R_x = f(x). \quad (2.27)$$

Поэтому оказывается, что точность системы характеризуется не числом (средним риском, средним квадратом ошибки), а некоторой функцией, что весьма неудобно при проектировании. Для преодоления этой трудности применяют различные методы. Наиболее распространенными являются минимаксный метод¹ и замена неизвестного распределения равномерным. При минимаксном мето-

де точность воспроизведения сообщений оценивается максимальным (по всем возможным значениям x) значением $R_{\max}$ условного риска, т. е. величиной

$$R_{\max} = \max_{(x)} R_x. \quad (2.28)$$

При этом наилучшей по точности считается такая система (вариант ее построения), у которой величина $R_{\max}$ минимальна.

Другой метод состоит в том, что неизвестное априорное распределение полагается равномерным:

$$P(x) = \begin{cases} h & \text{при } x = [x_{\min}, x_{\max}], \\ 0 & \text{вне этих пределов,} \end{cases} \quad (2.29)$$

где константа h определяется из условия нормировки (2.2):

$$h = 1/(x_{\max} - x_{\min}). \quad (2.30)$$

При этом выражения (2.21) и (2.26) принимают следующий вид:

$$R = h \int_{(x)} \int_{(x^*)} I(x, x^*) P(x^*|x) dx dx^*, \quad (2.31)$$

$$M [(\Delta x)^2] = h \int_{(x)} \int_{(x^*)} (x - x^*)^2 P(x^*|x) dx dx^*. \quad (2.32)$$

Наряду со средним риском R (или средним квадратом ошибки $M[(\Delta x)^2]$) важными статистическими характеристиками ошибки Δx являются ее математическое ожидание

$$M[\Delta x] = \int_{(x)} \int_{(x^*)} (x - x^*) P(x) P(x^*|x) dx dx^* \quad (2.33)$$

и дисперсия

$$D_{\Delta x} = M[(\Delta x - M[\Delta x])^2] = M[(\Delta x)^2] - (M[\Delta x])^2. \quad (2.34)$$

Если математическое ожидание ошибки равно нулю, то ошибка Δx называется несмещенной. В этом случае, как следует из (2.34),

$$D_{\Delta x} = M[(\Delta x)^2]. \quad (2.35)$$

Если априорное распределение $P(x)$ неизвестно, то приходится вместо безусловных значений математического ожидания и дисперсии вычислять их условные значения, определяемые для каждого фиксированного значения сообщения x , т. е. вместо (2.33) и (2.34) определять условное математическое ожидание ошибки

$$M_x[\Delta x] = \int_{(x^*)} (x - x^*) P(x^*|x) dx^* \quad (2.36)$$

и условную дисперсию ошибки

$$D_{x, \Delta x} = M_x[(\Delta x)^2] - (M_x[\Delta x])^2. \quad (2.37)$$

Однако при этом показатели точности $M_x[\Delta x]$ и $D_{x, \Delta x}$ оказываются функциями от x , что делает неудобным их практическое использование. Для преодоления этой трудности применяют те же

¹ Минимаксный метод, вообще говоря, можно применять и при известном $P(x)$, если вести расчет на худший, а не на средний случай.

методы, что и при определении среднего риска R или среднего квадрата ошибки $M[(\Delta x)^2]$, например полагают распределение $P(x)$ равномерным или применяют минимаксный критерий.

Из изложенного следует, что точность воспроизведения непрерывных сообщений часто приходится характеризовать не одним, а двумя показателями — математическим ожиданием ошибки и ее дисперсией (или средним риском). Это объясняется тем, что ошибка Δx в общем случае имеет две составляющие существенно различного характера:

$$\Delta x = \Delta x_c + \Delta x_{\phi}, \quad (2.38)$$

где Δx_c — систематическая, а Δx_{ϕ} — флуктуационная составляющие.

Систематической составляющей ошибки (или, короче, систематической ошибкой) называется такая составляющая Δx , которая при переходе от одного сеанса к другому не изменяется или изменяется по неслучайному (детерминированному) закону. Если эта составляющая не изменяется и при переходе от одного образца к другому, то она является безусловно (строго) систематической. Если же систематическая ошибка при переходе от одного образца к другому изменяется случайным образом, то она является условно (нестрого) систематической. Это означает, что безусловная систематическая ошибка Δx_c есть не что иное, как безусловное математическое ожидание ошибки:

$$\Delta x_c = M[\Delta x], \quad (2.39)$$

где безусловно математическое ожидание $M[\Delta x]$ в общем случае находится путем усреднения как по большому (теоретически бесконечно большому) числу сеансов, так и по большому (бесконечно большому) числу однотипных образцов. Эта ошибка строго детерминирована. В отличие от этого условная систематическая ошибка $\Delta x_{c,yc}$ является условным математическим ожиданием ошибки:

$$\Delta x_{c,yc} = M_{yc}[\Delta x]. \quad (2.40)$$

Здесь условное математическое ожидание $M_{yc}[\Delta x]$ определяется при условии, что образец системы фиксирован и усреднение производится лишь по большому (бесконечно большому) числу сеансов. Очевидно, для данного образца величина $\Delta x_{c,yc}$ неслучайна, но при переходе от одного образца к другому изменяется случайным образом.

В литературе обычно не делают различия между условной и безусловной систематическими ошибками, что в общем случае недопустимо. Одни авторы понимают под систематической только безусловную систематическую ошибку, другие — только условную; третьи называют систематической ошибкой математическое ожидание полной ошибки, не уточняя, о каком математическом ожидании (условном или безусловном) идет речь; четвертые, говоря о систематической ошибке, вообще не дают ее строгого ма-

тематического определения. Выше отмечалось, что усреднение по образцам можно не производить лишь в случаях, когда источник обладает свойством эргодичности или проектируемая система является уникальной. Поэтому в общем случае необходимо четко представлять, о каком усреднении идет речь — только по сеансам или также и по образцам. От этого будет зависеть как значение показателя точности, так и пути его улучшения. Так как со статистическим усреднением приходится иметь дело при вычислении не только математического ожидания ошибки, но и ее дисперсии и среднего риска, то во всех этих случаях также нужно четко представлять о каком усреднении идет речь — безусловном или условном, т. е. на каком множестве определена усредняемая случайная величина.

С учетом этих замечаний вернемся к представлению ошибки в виде суммы (2.38) систематической и флуктуационной составляющих. Так как мы условились под Δx_c понимать безусловную систематическую ошибку, т. е. строго детерминированную составляющую, то Δx_{ϕ} — строго случайная (флуктуационная) составляющая ошибки, безусловное математическое ожидание которой равно нулю:

$$M[\Delta x_{\phi}] = 0. \quad (2.41)$$

Из (2.38) следует, что

$$\Delta x_{\phi} = \Delta x - \Delta x_c = \Delta x - M[\Delta x], \quad (2.42)$$

т. е. флуктуационная ошибка есть не что иное, как отклонение полной ошибки от ее математического ожидания.

Из (2.34) и (2.42) следует, что

$$D_{\Delta x} = M[(\Delta x_{\phi})^2], \quad (2.43)$$

т. е. дисперсия (безусловная) ошибки равняется среднему квадрату флуктуационной составляющей ошибки. Если вместо (2.38) представить полную ошибку в виде

$$\Delta x = \Delta x_{c,yc} + \Delta x_{\phi,yc}, \quad (2.44)$$

где $\Delta x_{c,yc}$ — условная систематическая ошибка, определяемая соотношением (2.40), то

$$\Delta x_{\phi,yc} = \Delta x - \Delta x_{c,yc} \quad (2.45)$$

есть условная флуктуационная составляющая ошибки.

В общем случае $\Delta x_{c,yc} \neq \Delta x_c$ и соответственно $\Delta x_{\phi,yc} \neq \Delta x_{\phi}$. Очевидно, $M[\Delta x_{\phi,yc}] = 0$ и составляющая $\Delta x_{\phi,yc}$ является отклонением полной ошибки от ее условного математического ожидания. Из изложенного можно сделать следующие основные выводы.

1. Точность воспроизведения сообщений, являющихся непрерывными случайными величинами, приходится характеризовать (оценивать) обычно не одним, а несколькими показателями.

2. В качестве таких показателей обычно выбирают статистические характеристики ошибки, определяемые путем статистического усреднения: средний риск, средний квадрат ошибки, среднее

значение (математическое ожидание) ошибки, дисперсия ошибки (средний квадрат флуктуационной составляющей ошибки).

3. Значения показателей качества, определяемых статистическим усреднением, в общем случае существенно зависят от того, по какому ансамблю производится это усреднение: только по ансамблю сеансов или по ансамблю образцов (условное усреднение) или как по ансамблю сеансов, так и по ансамблю образцов (безусловное усреднение). От вида усреднения зависит, в частности, и систематическая ошибка (математическое ожидание ошибки). Поэтому в общем случае необходимо различать безусловные и условные показатели качества, безусловную и условную систематические ошибки.

4. Для определения всех приведенных показателей качества необходимо знать закон распределения вероятностей ошибки Δx или располагать достаточно большим экспериментальным статистическим материалом, позволяющим производить статистическое усреднение. Если ни тех, ни других данных нет, приходится характеризовать точность значительно грубее — максимальным значением $\Delta x_{\max}$ ошибки, определяемым путем грубых оценок сверху возможного значения ошибки Δx .

Если источники ошибки нестационарны, то введенные выше показатели точности (математическое ожидание ошибки, дисперсия и др.) оказываются зависящими от начала отсчета времени. При этом, если необходимо характеризовать точность не функцией времени, а числом, приходится находить максимальное или среднее по времени (за достаточно большое время) значение соответствующего показателя точности.

2.4. ПОКАЗАТЕЛИ ТОЧНОСТИ ВОСПРОИЗВЕДЕНИЯ СООБЩЕНИЙ, ИЗМЕНЯЮЩИХСЯ ВО ВРЕМЕНИ

В этом случае

$$x = x(t), x^* = x^*(t) \text{ и } \Delta x = \Delta x(t)$$

— некоторые случайные функции времени и возможны самые различные подходы к выбору меры точности воспроизведения сообщения $x(t)$. Наиболее распространенными являются следующие меры точности:

1. Математическое ожидание ошибки $M[\Delta x(t)]$.
2. Средний квадрат ошибки $M[\Delta x^2(t)]$.
3. Дисперсия ошибки

$$D_{\Delta x} = M[(\Delta x(t) - M[\Delta x(t)])^2] = M[(\Delta x(t))^2] - (M[\Delta x(t)])^2. \quad (2.46)$$

4. Математическое ожидание интегральной квадратичной оценки

$$M \left[\int_0^T [\Delta x(t)]^2 dt \right].$$

5. Средний риск R .

Для образования среднего риска R функции $x(t)$ и $x^*(t)$ удобно полагать имеющими спектр, ограниченный высшей частотой F_a . Тогда в соответствии с теоремой Котельникова можно полагать, что каждая реализация процесса $x(t)$ характеризуется совокупностью (вектором) своих отсчетов

$$x = \langle x_1, \dots, x_i, \dots, x_n \rangle,$$

где

$$x_i = x(i \Delta t), i = \overline{1, n}; \quad \Delta t = 1/2 F_a. \quad (2.47a)$$

Аналогично

$$x^* = \langle x_1^*, \dots, x_j^*, \dots, x_n^* \rangle,$$

где

$$x_j^* = x^*(j \Delta t), j = \overline{1, n}, \quad \Delta t = 1/2 F_a. \quad (2.47b)$$

Тогда по аналогии с (2.21) и (2.22) вводится функция потерь $I(x, x^*)$ и находится ее среднее значение

$$R = M[I(x, x^*)] = \int \int I(x, x^*) P(x) P(x^*|x) dx dx^*. \quad (2.48)$$

В развернутой (т. е. в векторной) записи

$$R = \int \dots \int I(\langle x_1, \dots, x_n \rangle, \langle x_1^*, \dots, x_n^* \rangle) P(x_1, \dots, x_n) P(x_1^*, \dots, x_n^* | x_1, \dots, x_n) dx_1, \dots, dx_n dx_1^*, \dots, dx_n^*. \quad (2.48')$$

т. е. является $2n$ -мерным интегралом, где $n \gg 1$. Поэтому его удается вычислить обычно лишь при гауссовской или марковской аппроксимации случайных процессов $x(t)$ и $x^*(t)$ или при учете лишь независимых выборочных значений этих процессов.

Функция потерь $I(x, x^*)$ назначает потери, соответствующие различным комбинациям векторов x и x^* . Часто удобно ее выбирать по аналогии с квадратичной функцией потерь (2.24) в виде квадратичной формы

$$I(x, x^*) = (x^* - x)' B (x^* - x), \quad (2.49)$$

где $B = \|b_{ij}\|$ — симметричная матрица размера $n \times n$, назначаемая потери (веса), соответствующие различным комбинациям ошибок $\Delta x_i = x_i^* - x_i$ и $\Delta x_j = x_j^* - x_j$. Штрих означает транспонирование.

В развернутом виде выражению (2.49) соответствует

$$\begin{aligned} I(\langle x, \dots, x_n \rangle, \langle x_1^*, \dots, x_n^* \rangle) &= \sum_{i=1}^n \sum_{j=1}^n b_{ij} (x_i^* - x_i) (x_j^* - x_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^n \Delta x_i \Delta x_j. \end{aligned} \quad (2.50)$$

В частном, но важном случае, когда

$$B = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

— диагональная единичная матрица, выражение (2.50) принимает вид

$$I = \sum_{i=1}^n (\Delta x_i)^2,$$

который является дискретным аналогом интеграла $\int_0^T (\Delta x(t))^2 dt$.

При этом средний риск совпадает (при $n \rightarrow \infty$) с математическим ожиданием интегральной квадратичной оценки.

Приведенные меры точности основаны на проведении статистического усреднения. Как и для сообщений, постоянных за время сеанса, это усреднение в общем случае должно производиться как по сеансам, так и по образцам. Для нестационарных процессов результаты усреднения по сеансам оказываются зависящими от начала отсчета сеансов (времени).

Наиболее просто оценивать точность, если процессы $x(t)$ и $x^*(t)$ можно считать стационарными и эргодическими. Тогда достаточно производить усреднение лишь по времени и результаты усреднения не зависят от начала отсчета времени. Иногда допустимо предельное упрощение, сводящееся к замене векторов x и x^* дискретными скалярными величинами x_i и x_j соответственно, т. е. к воспроизведению случайных величин (а не процессов), рассмотренному в § 2.2.

Действительно, пусть на интервале воспроизведения $[0, T]$ $x(t)$ имеет вид последовательности дискретных случайных величин $x_1, \dots, x_j, \dots, x_l$ (рис. 2.1), каждая из которых существует на интервале Δt_k и может принимать на этом интервале одно из m значений (такая ситуация имеет место, например, при передаче двоичной цифровой информации. В этом случае $m=2$ и на каждом интервале Δt_k передается один двоичный символ — нуль или единица).

Если значения сообщения $x(t)$ и ошибки $\Delta x(t)$ в различных интервалах Δt_k , $k=1, \dots, l$, статистически независимы и имеют одинаковые одномерные распределения, а длительность этих интервалов одинакова ($\Delta t_1 = \dots = \Delta t_l = \Delta t$), то очевидно, что задача воспроизведения сообщения $x(t)$ сводится к независимому воспроизведению сообщения на каждом элементарном интервале (к так называемому поэлементному приему), т. е. к воспроизведению дискретной случайной величины x_k . При этом точность воспроизведения сообщения может характеризоваться показателями, рассмотренными в § 2.1, например полной вероятностью $P_{\text{ош}}$ ошибки вос-

произведения элементарного символа x_k . Иногда такой упрощенный (поэлементный) подход применяют в качестве первого приближения и в тех случаях, когда не все из упомянутых выше предположений полностью справедливы (например, если нет уверенности в том, что ошибки воспроизведения сообщения в различных элементарных интервалах статистически независимы).

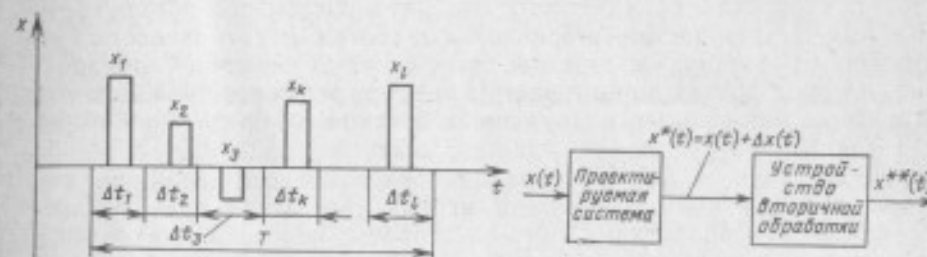


Рис. 2.1

Рис. 2.2

Рассмотрим теперь другой крайний случай, в котором знания показателя (2.46) может оказаться совершенно недостаточно для характеристики точности воспроизведения сообщений, являющихся функциями времени. Этот случай имеет место, если получатель сообщений обладает существенной избирательностью (линейной или нелинейной) по отношению к характеру ошибок воспроизведения сообщений. Например, если получателем информации является человек, то при получении акустической или визуальной информации он по-разному реагирует на составляющие ошибки (помех), имеющие одинаковое среднеквадратическое значение, но расположенные в различных участках частотного диапазона. Для учета этого обстоятельства при оценке точности вводят так называемые психометрические (при воспроизведении акустической информации) и визометрические (при воспроизведении визуальной информации) коэффициенты (см. например, [6]). Если воспроизводимая информация предназначена не для человека, а для какого-либо автоматического устройства, то это устройство также может обладать существенной избирательностью.

При решении некоторых технических задач оказывается целесообразным разбивать процесс воспроизведения сообщений на несколько последовательных этапов, которые называются соответственно первичной, вторичной, третичной и т. д. обработкой информации. Например, при измерении с Земли параметров эллиптической орбиты космического аппарата (ее большой полуоси, эксцентриситета, угла наклона и др.) окончательные данные о значениях этих параметров вырабатываются в центральном командно-измерительном пункте на основании данных, поступающих от нескольких измерительных пунктов, расположенных по земной поверхности на значительных расстояниях друг от друга. При этом в каждом вынесенном пункте производится первичная

обработка информации, сводящаяся, например, к измерению дальности, радиальной скорости и угловых координат космического аппарата относительно этого пункта. Данные измерений, содержащие некоторые ошибки, поступают в центральный пункт, в котором производится вторичная обработка — определение параметров орбиты на основе поступивших неточных данных о координатах и скоростях космического аппарата. Очевидно, погрешность на выходе устройства вторичной обработки будет зависеть не только от средних квадратов погрешностей первичной обработки, но и от других характеристик этих погрешностей — частотных спектров, одномерных и двумерных законов распределения и т.д.

Таким образом, если в данной системе осуществляется не окончательная, а лишь первичная обработка сообщений, то для этой системы роль получателя играет устройство последующей (вторичной) обработки. С другой стороны, человека или другого конечного получателя информации, обладающего линейной или нелинейной избирательностью к ошибкам воспроизведения сообщений, также можно рассматривать как устройство вторичной обработки информации. Поэтому в общем случае следует полагать, что проектируемая система выдает информацию некоторому устройству вторичной обработки (рис. 2.2), выходом которого является итоговый (конечный) результат $x^{**}(t)$ воспроизведения сообщений. Для нахождения этого результата может потребоваться знание не только простейших характеристик ошибки $\Delta x(t)$, но и дополнительных характеристик — ее энергетического спектра $g_{\Delta x}(f)$, функции корреляции, одномерного распределения $P(\Delta x)$ и др.

Пусть например, устройством вторичной обработки является линейный стационарный фильтр с такой передаточной функцией $K_{\Phi}(p)$, что для сообщения $x(t)$ можно полагать $K_{\Phi}(p)=1$. Тогда, как следует из рис. 2.2,

$$x^{**}(t) = K_{\Phi}(p) [x(t) + \Delta x(t)] \approx x(t) + K_{\Phi}(p) \Delta x(t).$$

При этом ошибка воспроизведения сообщения после вторичной обработки

$$\Delta x^{**}(t) = x^{**}(t) - x(t) = K_{\Phi}(p) \Delta x(t). \quad (2.51)$$

Пусть ошибка $\Delta x(t)$ может считаться стационарным случайным процессом с нулевым математическим ожиданием и энергетическим спектром $g_{\Delta x}(f)$, т. е.

$$M[(\Delta x)^2] = D_{\Delta x} = \int_0^{\infty} g_{\Delta x}(f) df. \quad (2.52)$$

Тогда средний квадрат (и дисперсия) ошибки Δx^{**} (в установившемся режиме) в соответствии с (2.51)

$$M[(\Delta x^{**})^2] = D_{\Delta x^{**}} = \int_0^{\infty} |K_{\Phi}(j2\pi f)|^2 g_{\Delta x}(f) df. \quad (2.53)$$

Следовательно, в рассматриваемом случае для нахождения среднего квадрата (дисперсии) итоговой (конечной) ошибки воспроизведения сообщений необходимо знать энергетический спектр $g_{\Delta x}(f)$ ошибки $\Delta x(t)$ проектируемой системы.

Если ширина спектра $g_{\Delta x}(f)$ больше полосы пропускания фильтра (рис. 2.3), то вместо (2.53) можно приближенно полагать

$$M[(\Delta x^{**})^2] = D_{\Delta x^{**}} = g_{\Delta x}(0) \Delta F_{\Phi}, \quad (2.54)$$

где $\Delta F_{\Phi} = \int_0^{\infty} |K_{\Phi}(j2\pi f)|^2 df$ — эквивалентная шумовая полоса пропускания фильтра.

Из (2.54) следует, что в этом случае средний квадрат (дисперсия) итоговой ошибки определяется не средним квадратом (дисперсией) ошибки $\Delta x(t)$ проектируемой системы, а спектральной плотностью этой ошибки на нулевой частоте $g_{\Delta x}(0)$. Из (2.52) и рис. (2.3) следует, что

$$M[(\Delta x)^2] = D_{\Delta x} = g_{\Delta x}(0) \Delta f_{\Phi}, \quad (2.55)$$

где

$$\Delta f_{\Phi} = \int_0^{\infty} \frac{g_{\Delta x}(f)}{g_{\Delta x}(0)} df \quad (2.56)$$

— эквивалентная ширина спектра ошибки $\Delta x(t)$. Поэтому соотношение (2.54) можно представить также в виде

$$M[(\Delta x^{**})^2] = D_{\Delta x^{**}} = M[(\Delta x)^2] \Delta F_{\Phi} / \Delta f_{\Phi}, \quad (2.57)$$

т. е. в результате вторичной обработки средний квадрат (дисперсия) ошибки уменьшается в $\Delta f_{\Phi} / \Delta F_{\Phi}$ раз. Поэтому мерой точности проектируемой системы (первичной обработки) является не средний квадрат (дисперсия) ошибки $\Delta x(t)$, а его отношение к эквивалентной ширине спектра ошибки, т. е. величина

$$K_{\Delta x}(0) = M[(\Delta x)^2] / \Delta f_{\Phi}. \quad (2.58)$$

Пусть, например, сравниваются два варианта построения проектируемой системы, причем второму варианту соответствует в 3 раза большее значение среднего квадрата (дисперсии) ошибки и в 6 раз большее значение эквивалентной ширины спектра этой ошибки. Тогда, как следует из (2.57) [или (2.58)], второй вариант обеспечит получение большей точности итогового воспроизведения, чем первый, и, следовательно, должен считаться по точности превосходящим первый вариант, несмотря на большее значение среднего квадрата (дисперсии) ошибки.

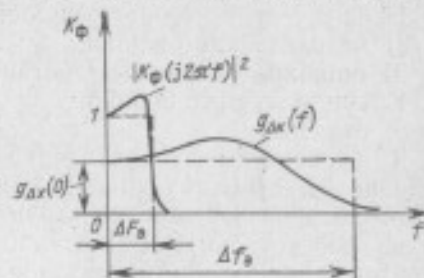


Рис. 2.3

Этот пример показывает, что в некоторых случаях оценка точности воспроизведения сообщения $x(t)$ лишь средним квадратом (или дисперсией) ошибки может оказаться не только недостаточной, но и совершенно неверной.

Отсюда следует и более общий вывод: при наличии взаимодействия проектируемой системы с другими системами или устройствами необходимо тщательно проанализировать характер этого взаимодействия с тем, чтобы установить, какие именно по-

казатели качества данной системы достаточно правильно отражают характер этого взаимодействия.

В заключение отметим, что при воспроизведении непрерывных сообщений (постоянных в течение сеанса воспроизведения или изменяющихся во времени) часто удобно оценивать точность не абсолютным, а относительным средним квадратом ошибки, т. е. величиной

$$\delta_s^2 = M[(\Delta x)^2]/M[x^2] \quad (2.59)$$

или

$$\delta_s^2 = M[(\Delta x)^2]/x_{\text{макс}}^2 \quad (2.60)$$

Здесь $M[x^2]$ и $x_{\text{макс}}$ — средний квадрат сообщения и квадрат его максимального значения соответственно.

Из (2.59) и (2.60) следует, что

$$\delta_s = P_{\Delta x} \delta, \quad (2.61)$$

где

$$P_{\Delta x} = x_{\text{макс}}/\sqrt{M[x^2]} \quad (2.62)$$

— пик-фактор сообщения, т. е. отношение максимального значения сообщения к среднеквадратическому значению. Для сообщения с нулевым математическим ожиданием и нормальным законом распределения можно полагать $P_{\Delta x} = 3 \dots 4$, а для синусоидального сообщения $P_{\Delta x} = \sqrt{2} = 1.41$. Если характер сообщения неизвестен, то в первом приближении полагают $P_{\Delta x} = 2 \dots 3$.

2.5. ИСТОЧНИКИ ОШИБОК

По характеру источников ошибки можно разбить на следующие основные группы:

- 1) физические неопределенности;
- 2) методические ошибки;
- 3) ошибки, вызванные действием помех;
- 4) аппаратные ошибки;
- 5) отказы.

К физическим неопределенностям можно отнести те неопределенности, существующие в природе, влияние которых на точность воспроизведения конструктор в принципе не в состоянии полностью устранить. Важнейшими из них являются тепловые и квантовые шумы, неточность знания физических законов и мировых постоянных, входящих в эти законы (скорости света c , заряда электрона e , постоянной Больцмана k , постоянной Планка h , гравитационной постоянной f и др.) Ошибки, вызванные действием шумов, обычно относят к третьей группе. Неточность знания физических законов имеет двойкий смысл — объективный и субъективный. Объективная неточность физических законов обусловлена тем обстоятельством, что все явления в природе взаимосвязаны и находятся в состоянии непрерывного изменения.

Каждый же закон выделяет из всего бесконечного множества взаимосвязанных явлений лишь главные и не учитывает влияния явлений и процессов, являющихся для данной пространственно-временной области второстепенными. Субъективная неточность состоит в том, что даже для данной ограниченной пространственно-временной области человечество может не знать с достаточной точностью законы природы и должно в процессе своего познания природы корректировать эти законы. Например, еще сравнительно недавно не учитывались в должной мере релятивистские и квантовые эффекты, требующие корректировки законов классической физики. В частности, при определении доплеровского сдвига частоты не учитывался релятивистский эффект, проявляющийся при скоростях движения объектов, сравнимых со скоростью света; при определении шумовой температуры приемника не учитывались квантовые шумы, существенно влияющие в диапазоне дециметровых и более коротких электромагнитных волн, и т. д.

В соотношения, описывающие основные законы природы, входят физические постоянные, неизменные для всех доступных человечеству в настоящее время явлений и называемые поэтому мировыми постоянными (c , e , k , h , f и др.). Все они известны с некоторой погрешностью (которая по мере развития науки и техники непрерывно уменьшается). В ряде случаев это может приводить к существенным погрешностям воспроизведения сообщений. Например, при измерении дальности до объекта активной импульсной РЛС дальность определяется по формуле

$$r = c \Delta t / 2, \quad (2.63)$$

где Δt — измеряемый временной сдвиг между зондирующим и отраженным импульсами.

Из (2.63) следует, что погрешность Δc знания скорости света вызывает (при идеально точном измерении временного сдвига Δt) погрешность измерения расстояния

$$\Delta r = r \Delta c / c. \quad (2.64)$$

При большом значении расстояния r (например, при измерении расстояния до планеты Венера) эта погрешность может быть для ряда задач существенной, несмотря на то, что уже в настоящее время отношение $\Delta c/c$ весьма мало.

Неточность знания человечеством физических законов и входящих в них констант сказывается обычно лишь в некоторых задачах, требующих исключительно высокой точности воспроизведения сообщений. В отличие от этого все остальные (рассматриваемые далее) источники ошибок оказываются весьма существенными в подавляющем большинстве практических задач.

К методическим относят ошибки, вызванные погрешностями теории (методов), использованной при проектировании системы (за исключением уже отмеченных погрешностей основных физических законов) и выбранных алгоритмов преобразования инфор-

мации. Методические ошибки могут быть осознанными или неосознанными. На ошибки первого вида разработчики идут сознательно ради упрощения анализа и синтеза системы или улучшения каких-либо ее свойств. Типичным примером такой ошибки является ошибка, возникающая в результате преобразования непрерывных сообщений в цифровые вследствие их дискретизации во времени и квантования по уровню, называемая шумом квантования. В этом случае истинные (аналоговые) сообщения заменяются их приближенными (цифровыми) аппроксимациями сознательно для получения возможности использования преимуществ цифровой обработки информации по сравнению с аналоговой, в первую очередь ради ослабления действия других источников ошибок — разброса и нестабильности характеристик аппаратуры, нелинейности тракта и др. Другим примером осознанной методической ошибки является часто делаемое допущение о том, что эллиптическая орбита космического аппарата, имеющая малый эксцентриситет e , является круговой, т. е. что $e=0$. Это допущение позволяет существенно упростить как проектирование радиосистемы контроля или управления космическим аппаратом, так и саму эту радиосистему.

Поскольку происхождение осознанных методических ошибок разработчикам системы известно еще на ранних этапах проектирования, они могут принять меры к тому, чтобы значения этих ошибок были допустимыми (например, выбирая достаточно малые интервалы дискретизации по времени и квантования по уровню при преобразовании аналоговых сообщений в цифровые).

В отличие от этого неосознанными называются такие методические ошибки, о существовании которых разработчики могут не знать в процессе всего проектирования системы или, во всяком случае, на его ранних стадиях. В какой-то мере такие ошибки неизбежны, так как все явления и процессы в природе взаимосвязаны, а разработчики системы могут учесть лишь ничтожно малый процент этих явлений и процессов и вследствие этого в ряде случаев пропустить некоторые существенные факторы.

Типичным примером такой ошибки явился существенный просчет, допущенный американскими инженерами при проектировании долговременной орбитальной станции «Скайлэб». При расчетах они предполагали, что эта многотонная станция будет существовать настолько долго, что до того, как она начнет терять высоту, будут разработаны эффективные средства ее пополнения горючим и ремонта, а в случае необходимости и безопасной ликвидации на орбите с помощью запускаемых с Земли космических кораблей многократного действия. Однако вопреки расчетам оказалось, что станция начала терять высоту на несколько лет раньше расчетного срока, когда корабли многократного действия разрабатывать еще не удалось. Это привело к тому, что тяжелые куски многотонной станции, не сгоревшие при ее входе в атмосферу, с большой скоростью врезались в различные точки земной поверхности и лишь случайно не привели к тяжелым последствиям. Как

выяснилось, просчет разработчиков состоял в том, что они не учли весьма медленного, но оказавшегося существенным увеличения плотности атмосферы на высоте орбиты «Скайлэб» (из-за нагрева атмосферы Земли промышленными отходами), которое привело к дополнительному (по сравнению с расчетами) торможению станции.

Очевидно, для уменьшения влияния неосознанных методических ошибок нужно, во-первых, весьма внимательно подходить к выявлению состава возможных источников существенных ошибок и, во-вторых, обращать особое внимание на сроки и объем экспериментальных исследований и проверок.

Весьма важным, а во многих случаях основным источником ошибок являются всевозможные мешающие воздействия, возникающие как во внешней по отношению к радиосистеме среде, так и внутри самой системы. Часть этих мешающих воздействий называют помехами, а остальные считают источниками аппаратных ошибок и отказов. Такое деление имеет условный характер и поэтому не может быть однозначным.

В большинстве случаев к помехам относят, в первую очередь, следующие виды мешающих воздействий:

- внутренний шум радиоприемных устройств;
- флуктуации сигнала, возникающие в процессе его распространения от передатчика к приемнику;
- мешающие отражения сигнала (от смежных объектов, Земли, облаков, местных предметов и др.);
- атмосферные и космические помехи;
- индустриальные помехи;
- помехи от радиостанций (непреднамеренные);
- преднамеренные (организованные) помехи;
- внутрисистемные помехи, т. е. помехи одним частям системы, создаваемые возмущениями, возникающими в других частях (кроме внутреннего шума).

Основными источниками аппаратных ошибок являются: линейные и нелинейные искажения сообщений, вызываемые неидеальностью амплитудных, амплитудно- и фазочастотных характеристик радиоканалов;

- разброс параметров аппаратуры при ее изготовлении;
- нестабильность напряжений источников питания;
- изменения температуры, влажности и давления;
- механические возмущения — вибрации, тряска, удары, переход к невесомости и др.

Отказами обычно называют длительные аномальные ошибки любого происхождения. Отказы могут вызываться самыми разнообразными причинами, в первую очередь обрывами и короткими замыканиями в цепях системы (так называемые внезапные отказы), старением и (или) истощением ресурсов источников питания, топлива и др. (постепенные отказы), а также длительным действием мощных помех.

Из изложенного вытекает, что источники существенных ошибок весьма разнообразны как по своему происхождению, так и по характеру действия. Поэтому общая теория точности распадается на четыре раздела: аппаратную точность, помехозащищенность, электромагнитную совместимость и надежность. Границы между этими разделами в некоторой степени условны и обычно таковы: в теории аппаратной точности рассматривают физические неопределенности, методические и аппаратные ошибки; в теории надежности — отказы; в теории ЭМС — помехи от радиостанций (непреднамеренные), а также промышленные и внутрисистемные помехи; ошибки, вызываемые остальными видами помех, являются предметом теории помехозащищенности.

От вида учитываемых источников ошибок зависит и содержание понятия «точность воспроизведения сообщений» или, короче, просто точность системы. Если при определении погрешности учитываются все существенные источники ошибки, рассматриваемые в теориях аппаратной точности, надежности, помехозащищенности и ЭМС, то такая точность называется обобщенной или точностью в широком смысле этого слова. Если учитываются лишь методические и аппаратные, а также ошибки, вызванные физическими неопределенностями, то точность называют аппаратной или просто точностью (в узком смысле этого слова). При рассмотрении действия только отказов термин «точность» заменяют на «надежность». Наконец, при учете ошибок, вызванных лишь действием помех, применяют (в зависимости от вида помех) термины «помехозащищенность» и «электромагнитная совместимость».

При этом обобщенная точность системы характеризуется совокупностью показателей аппаратной точности, надежности, помехозащищенности и электромагнитной совместимости. Однако, как будет показано далее, такое независимое (отдельное) рассмотрение действия различных групп источников ошибок не всегда оправдано. Поэтому в последние годы появляется все больше теоретических работ, посвященных исследованию обобщенной точности с одновременным учетом действия всех или по меньшей мере нескольких групп источников существенных ошибок, например с одновременным учетом действия помех, аппаратных ошибок и отказов.

2.6. СУММИРОВАНИЕ ОШИБОК, ВЫЗВАННЫХ РАЗЛИЧНЫМИ ИСТОЧНИКАМИ

Вычисление результирующей ошибки, вызванной совокупным действием всех существенных источников, является одной из самых сложных, если не самой сложной проблемой, которую приходится решать в процессе проектирования радиосистемы. Это объясняется не только большим числом различных источни-

ков, но и существенно различным характером их действия, а также необходимостью в ряде случаев учета их взаимодействия.

Различный характер действия отдельных источников (или групп источников) приводит к необходимости описывать их действие различными показателями точности. Например, аномальные ошибки правильнее оценивать не средним квадратом, а вероятностью возникновения такой ошибки за заданное время (например, за время одного сеанса); мультипликативную погрешность, проявляющуюся в изменении масштаба воспроизведения сообщения x , целесообразнее оценивать не разностью $\Delta x = x^* - x$, а отношением x^*/x и т. д.

Это обстоятельство делает желательным отдельный (независимый) анализ действия каждого источника (или группы источников) ошибок. Однако при таком рассмотрении не учитывается эффект взаимодействия отдельных источников, пренебрежение которым в ряде случаев совершенно недопустимо. Взаимодействие источников ошибок может быть разнообразным. Например, если в системе осуществляется автоматическое слежение за отражаемым целью сигналом по угловым координатам и по частоте, то при наличии независимых аддитивных помех появление существенной динамической ошибки по угловым координатам приводит к уменьшению мощности принимаемого антенной сигнала. Это вызывает увеличение флуктуационной ошибки как в канале слежения за целью, так и в частотно-следящем канале. Уменьшение точности слежения по частоте приводит, в свою очередь, к дополнительному уменьшению мощности сигнала на выходе приемника и, значит, к дополнительному увеличению погрешности слежения за целью как по частоте, так и по угловым координатам. Следовательно, в рассматриваемом случае взаимодействуют различные источники ошибок (динамической и флуктуационной) как в каждом канале системы, так и между каналами. В любой реальной радиосистеме, даже средней сложности, имеется много путей существенного взаимодействия различных источников ошибок.

Таким образом, с одной стороны, для определения результирующей (обобщенной) точности радиосистемы желательно, а иногда и необходимо учитывать взаимодействие между всеми основными источниками ошибок; с другой же стороны, как уже отмечалось, теоретический учет такого взаимодействия в ряде случаев весьма сложен. Экспериментальный учет также может оказаться слишком сложным, так как требует проведения огромного количества статистических испытаний в весьма разнообразных условиях, некоторые из которых даже в принципе невозможно реализовать (особенно при испытаниях военных систем).

Все эти обстоятельства и привели к тому, что теория обобщенной точности радиосистем находится пока в начальном состоянии и существенное развитие получили лишь отдельные теории аппаратной точности, надежности, помехозащищенности и ЭМС.

Рассмотрим приближенные методы оценки результирующей ошибки, вызванной совокупным действием всех основных ее источников. Пусть воспроизводимое сообщение x и ошибка его воспроизведения $\Delta x_i (i=\overline{1, N})$ являются стационарными случайными величинами, а вероятность аномальных ошибок пренебрежимо мала. Тогда суммарную ошибку системы можно представить в виде

$$\Delta x_p = \sum_{i=1}^N \Delta x_i, \quad (2.65)$$

где Δx_i — ошибка, создаваемая i -м источником; N — общее число источников.

В первом приближении пренебрегают взаимодействием всех источников, т. е. считают, что величины Δx_i статистически независимы и каждую из них можно вычислить в предположении, что остальные источники ошибок отсутствуют. Обычно такое допущение справедливо тем в большей мере, чем выше точность действия системы и всех ее подсистем (устройств), т. е. чем меньше относительные погрешности $\delta_i = \sqrt{M[(\Delta x_i)^2]}/x_{i \text{ макс}}$ при $i=\overline{1, N}$. Из (2.65) следует, что

$$M[\Delta x_p] = \sum_{i=1}^N M[\Delta x_i], \quad (2.66)$$

т. е. результирующая систематическая ошибка равна сумме систематических ошибок, вызываемых отдельными источниками. Так как мы приняли величины Δx_i статистически независимыми, то дисперсия результирующей ошибки равна сумме дисперсий отдельных составляющих:

$$D_{\Delta x_p} = \sum_{i=1}^N D_{\Delta x_i}, \quad (2.67)$$

Поэтому средний квадрат результирующей ошибки

$$M[(\Delta x_p)^2] = \left(\sum_{i=1}^N M[\Delta x_i] \right)^2 + \sum_{i=1}^N D_{\Delta x_i}. \quad (2.68)$$

Поскольку составляющие Δx_i приняты статистически независимыми, из (2.65) следует, что закон распределения результирующей ошибки Δx_p близок к смещенному нормальному (при любых законах распределения составляющих $\Delta x_i, i=\overline{1, N}$), если выполняются следующие дополнительные условия нормализации.

1. Число слагаемых достаточно велико, т. е. $N \gg 1$. В первом приближении достаточно, если

$$N \geq 10. \quad (2.69)$$

2. Дисперсия каждого слагаемого много меньше дисперсии суммы, т. е. $D_{\Delta x_i} \ll D_{\Delta x_p}, i=\overline{1, N}$. В первом приближении достаточно, если

$$D_{\Delta x_i} \leq 0,1 D_{\Delta x_p}, i=\overline{1, N}. \quad (2.70)$$

Из (2.67) следует, что если слагаемых всего десять ($N=10$), то условие (2.70) может выполняться лишь, если все составляющие ошибки имеют примерно одинаковые дисперсии. Если же $N \gg 10$, то закон распределения может быть близким к нормальному и при существенно различных частных дисперсиях $D_{\Delta x_i} (i=\overline{1, N})$.

Рассмотрим теперь случай воспроизведения дискретных сообщений. Предположим, что точность (достоверность) воспроизведения таких сообщений характеризуется полной вероятностью $P_{\text{ош}}$ ошибки. Пусть при действии только i -го источника значение этой вероятности равно $P_{\text{ош } i}$, т. е. вероятность правильного воспроизведения

$$P_{\text{пр } i} = 1 - P_{\text{ош } i}.$$

Найдем вероятность правильного воспроизведения сообщения $P_{\text{пр}}$ с учетом действия всех N источников. Будем полагать для упрощения, что ошибки, вызываемые различными источниками, статистически независимы. Тогда

$$P_{\text{пр}} = \prod_{i=1}^N P_{\text{пр } i}, \text{ т. е. } 1 - P_{\text{ош}} = \prod_{i=1}^N (1 - P_{\text{ош } i}),$$

где $P_{\text{ош}} = 1 - P_{\text{пр}}$ — полная вероятность ошибки с учетом действия всех N источников. Поэтому

$$P_{\text{ош}} = 1 - \prod_{i=1}^N (1 - P_{\text{ош } i}). \quad (2.71)$$

Если $N P_{\text{ош } i} \leq 0,1$, т. е.

$$P_{\text{ош } i} \leq 0,1/N \quad (2.72)$$

(что обычно имеет место), то из (2.71) приближенно получается

$$P_{\text{ош}} \approx \sum_{i=1}^N P_{\text{ош } i}. \quad (2.73)$$

Следовательно, при справедливости сделанных допущений полная вероятность ошибки, вызванной совокупным действием всех N источников, равна сумме полных вероятностей ошибок, вызванных в отдельности каждым из этих источников.

Из (2.66), (2.67) и (2.73) следует, что как при непрерывных, так и при дискретных сообщениях приближенные соотношения, характеризующие вклад каждого источника ошибки в результирующую ошибку (а при дискретных сообщениях — в результирующую вероятность ошибки), получаются весьма простыми. Значительно сложнее решить вопрос об оптимальном распределении требований, предъявляемых к различным источникам ошибок. Рассмотрим этот вопрос на примере соотношения (2.73) для дискретных сообщений.

Пусть максимально допустимое значение полной вероятности ошибки, вызванной совокупным действием всех N источников, равно $P_{\text{ош, макс}}$. Требуется обосновать допустимое значение $P_{\text{ош } i}$ вероятности ошибки для каждого из N источников. Если распределить усилия (требования) между всеми источниками равномерно, т. е. принять

$$P_{\text{ош } 1} = P_{\text{ош } 2} = \dots = P_{\text{ош } N}, \quad (2.74)$$

то из (2.73) получается, что

$$P_{\text{ош } i} = P_{\text{ош, макс}}/N. \quad (2.75)$$

Однако равномерное распределение усилий (2.74) в общем случае не является оптимальным. Действительно, в общем случае ослабить мешающее действие какого-то одного, например l -го, источника может быть значительно сложнее, чем уменьшить до такого же значения действие другого, например k -го, источника: для обеспечения равенства $P_{\text{ош } l} = P_{\text{ош, макс}}/N$ может потребоваться значительно больше усилий (затрат времени, труда, материалов, деталей и др.), чем для обеспечения $P_{\text{ош } k} = P_{\text{ош, макс}}/N$. В такой ситуации может оказаться более правильным предъявить к источнику k более жесткие требования, чем к источнику l , т. е. принять $P_{\text{ош } k} < P_{\text{ош } l}$. Результирующее усилие (например, стоимость), необходимое для обеспечения данных вероятностей ошибок $P_{\text{ош } 1}, \dots, P_{\text{ош } N}$, является некоторой функцией

$$C = f(P_{\text{ош } 1}, \dots, P_{\text{ош } N}) \quad (2.76)$$

этих вероятностей. Поэтому оптимальными будут такие значения вероятностей $P_{\text{ош } 1}, \dots, P_{\text{ош } N}$, которые обеспечат минимум величины C при одновременном выполнении условия

$$P_{\text{ош } 1} + \dots + P_{\text{ош } N} \leq P_{\text{ош}}.$$

Следовательно, оптимальное распределение усилий, т. е. оптимальное допустимое значение вероятностей ошибок, должно находиться путем решения задачи

$$C = f(P_{\text{ош } 1}, \dots, P_{\text{ош } N}) = \min_{(S)} \quad (2.77)$$

при $P_{\text{ош } 1} + \dots + P_{\text{ош } N} \leq P_{\text{ош}}$.

Однако при решении этой задачи необходимо располагать зависимостью (2.76), которую в большинстве практических случаев найти не удастся. При этом вместо строгого решения задачи об оптимальном распределении усилий (требований) приходится применять приближенные методы. Например, если очевидно без детальных расчетов, что для некоторых из N источников можно сделать значение вероятности $P_{\text{ош } i}$ пренебрежимо малым по сравнению с $P_{\text{ош, макс}}$ без существенного усложнения (или удорожания) системы, то на это следует пойти и исключить такие источники из дальнейшего рассмотрения. При этом для дальнейшего рассмотрения остаются лишь N' источников (где $N' < N$), таких, для которых целесообразность уменьшения допустимой вероятности

ошибки ниже значения $P_{\text{ош, макс}}/N'$ не очевидна. Между этими N' источниками усилия можно распределить в первом приближении равномерно, т. е. положить для каждого (j -го) из этих источников

$$P_{\text{ош, макс } j} = P_{\text{ош, макс}}/N'. \quad (2.78)$$

Если в ходе дальнейшего проектирования выяснится, что для некоторых из N' источников выполнить требование (2.78) значительно сложнее (дороже), чем для других источников, следует перераспределить усилия, т. е. увеличить значение $P_{\text{ош, макс } j}$ для одних источников и соответственно уменьшить для других (с тем, чтобы суммарная вероятность ошибки осталась равной $P_{\text{ош, макс}}$).

2.7. ОБЩАЯ ХАРАКТЕРИСТИКА ПУТЕЙ УМЕНЬШЕНИЯ ОШИБОК

Поскольку источники ошибок весьма разнообразны как по происхождению, так и по характеру вызываемых ими погрешностей, существует много различных путей ослабления влияния тех или иных источников, рассматриваемых в теориях аппаратурной точности, надежности, помехозащищенности и ЭМС.

Все методы повышения точности можно разбить на три основные группы:

направленные на устранение самих источников мешающих воздействий или подавление этих воздействий в месте их возникновения;

направленные на ослабление мешающего действия различных источников в тех частях системы, в которых это действие оказывается наиболее сильным;

комбинированного действия, позволяющие, во-первых, устранить некоторые источники ошибок и, во-вторых, ослабить действие ряда других источников.

К первой группе методов можно отнести, в частности:

подавление источников промышленных помех внутри этих источников (например, применением экранировки);

уменьшение шумовой температуры радиоприемных устройств; ослабление взаимных помех между радиостанциями разнесением их рабочих частот;

повышение скрытности действия системы, затрудняющее организацию преднамеренных помех;

переход в диапазон сантиметровых или более коротких волн, в котором интенсивность многих видов помех (атмосферных, промышленных и от радиостанций) резко уменьшается;

ослабление влияния нестабильности температуры внутри всего отсека с аппаратурой путем стабилизации температуры внутри этого отсека;

устранение влияния внутреннего нагрева аппаратуры значительным уменьшением потребляемой энергии (например, заменой электронных приборов с раскаленными катодами соответствующими

щими приборами с холодными катодами; повышением коэффициента полезного действия мощных блоков системы);

уменьшение числа механических контактов применением интегральной технологии изготовления аппаратуры;

замена механических блоков управления электронными;

стабилизация питающих напряжений.

Вторая группа методов повышения точности весьма разнообразна. Перечислим лишь некоторые из них:

применение пространственно-временной обработки в радиопередающих и особенно радиоприемных устройствах для ослабления действия различных помех;

применение принципов компенсации, инвариантности и комплексирования для ослабления действия различных источников аппаратных ошибок и помех;

резервирование аппаратуры и встроенный контроль для повышения надежности действия системы;

применение принципа обратной связи — охват отдельных блоков или цепей системы обратной связью (отрицательной или положительной, избирательной или неизбирательной) для повышения точности воспроизведения сообщений (уменьшения влияния внутренних нестабильностей, уменьшения линейных и нелинейных искажений, подавления некоторых видов помех и др.);

термостатирование наиболее критичных к изменениям температуры блоков системы;

амортизация частей системы, наиболее критичных к механическим воздействиям;

экранировка отдельных блоков от мешающего воздействия электрических и магнитных полей;

применение калибровки для уменьшения систематических ошибок;

стабилизация частоты передатчиков и гетеродинов приемников;

предварительная тренировка вновь изготовленной аппаратуры.

Основным методом комбинированного действия является переход от аналоговой обработки сообщений и сигналов к двоичной цифровой обработке. При этом устраняется весьма важный источник искажений сообщений — нелинейные искажения. Резко уменьшается влияние нестабильностей характеристик тракта передачи сообщений. Благодаря использованию больших интегральных схем, содержащих в небольшом объеме огромное число цифровых элементов, появляется возможность реализовать более помехоустойчивые, в частности, адаптивные, алгоритмы пространственно-временной обработки информации, в большей мере применять резервирование аппаратуры, заменять механические блоки электронными и проводить ряд других мероприятий, повышающих обобщенную точность воспроизведения сообщений.

Но следует иметь в виду, что при переходе к цифровой обработке имеет место появление или усиление действия некоторых источников ошибок. Во-первых, как уже упоминалось, появляется новый источник — шум квантования. Во-вторых, расширяется (на

порядок или более) требуемая полоса пропускания системы, что может привести к усилению действия некоторых видов помех. Кроме того, для перехода к цифровой обработке весьма высокочастотных радиосигналов пропускная способность современных электронно-вычислительных устройств может оказаться недостаточной.

Однако в большинстве случаев переход на цифровую обработку позволяет существенно повысить обобщенную точность воспроизведения сообщений.

В зависимости от характера источников ошибок все упомянутые и многие другие пути повышения обобщенной точности рассматриваются соответственно в теориях аппаратурной точности, надежности, помехозащищенности и ЭМС, которым посвящены последующие главы. Здесь же ограничимся лишь методами, основанными на использовании принципов компенсации, инвариантности и комплексирования. Эти три принципа, тесно связанные между собой, характерны тем, что позволяют уменьшить мешающее воздействие весьма широкого класса возмущений (например, столь разнородных, как температурные нестабильности и организованные помехи) и тем самым повысить не только аппаратурную точность, но и помехозащищенность, ЭМС и надежность. Это объясняется тем, что для реализации указанных принципов требуется лишь небольшая априорная информация о мешающих воздействиях.

2.8. ПРИНЦИП КОМПЕНСАЦИИ

Принцип компенсации состоит в таком построении системы, при котором мешающее действие возмущений в одних каналах системы компенсируется действием тех же возмущений в других каналах. Пусть на систему, дающую на своих выходах эффекты $u_1(t), \dots, u_j(t), \dots, u_N(t)$, действуют мешающие возмущения $f_1(t), \dots, f_i(t), \dots, f_M(t)$ (рис. 2.4). Скомпенсировать мешающее действие возмущения $f_i(t)$ на выходной эффект $u_j(t)$ означает сделать этот эффект независимым или, как говорят, инвариантным относительно $f_i(t)$. Желательно, чтобы эта инвариантность (независимость) сохранялась не только в отсутствие других мешающих возмущений, но и при их наличии, а также чтобы была обеспечена инвариантность каждого из выходов $u_1(t), \dots, u_N(t)$ относительно всех мешающих возмущений $f_1(t), \dots, f_M(t)$. Однако столь полной и совершенной инвариантности обычно не удается

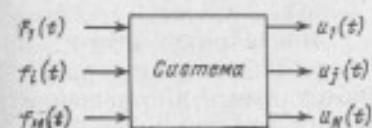


Рис. 2.4



Рис. 2.5

добиться не только путем применения принципов компенсации, но и всем комплексом мер повышения точности. Поэтому наиболее часто задача компенсации ставится более узко: скомпенсировать действие лишь некоторых из M воздействий на некоторые из N выходов системы.

В простейшем случае требуется полностью или частично скомпенсировать действие на один из выходов $u(t)$ лишь одного из возмущений $f(t)$ (рис. 2.5). При этом вначале обычно полагают остальные $(M-1)$ возмущений отсутствующими и лишь на втором этапе выясняют, насколько они могут ухудшить эффект компенсации. Следовательно, вначале задача ставится следующим образом: построить систему так, чтобы возмущение $f(t)$, проходящее по одним каналам, компенсировалось тем же возмущением, проходящим по другим каналам. Отсюда следует, что для обеспечения компенсации система должна иметь не менее двух каналов, по которым воздействие $f(t)$ проходит на выход $u(t)$.

Пусть имеются два таких канала и их можно считать по отношению к воздействию $f(t)$ линейными стационарными цепями с нулевыми начальными условиями. Тогда каналы полностью определяются своими передаточными функциями $K_1(p)$ и $K_2(p)$, система, приведенная на рис. 2.5, принимает вид, изображенный на рис. 2.6, и

$$u(t) = [K_1(p) - K_2(p)] f(t).$$

Поэтому условие полной (идеальной) компенсации в этом случае таково:

$$K_1(p) = K_2(p), \text{ т.е. } K_1(j\omega) = K_2(j\omega). \quad (2.79)$$

Это означает, что должно иметь место равенство амплитудно-фазочастотных характеристик каналов во всем диапазоне частот от 0 до ∞ , что выполнить невозможно. Однако практически условие (2.79) достаточно выполнить лишь приближенно и притом в том диапазоне частот $F_{\text{н}} \dots F_{\text{в}}$, за пределами которого энергия мешающего воздействия $f(t)$ пренебрежимо мала. При этом компенсация будет хотя и не полной, но обеспечит значительное ослабление действия мешающего возмущения.

Образование в системе двух или более каналов, обеспечивающих компенсацию, может производиться либо только ради компенсации мешающих возмущений, либо также и для решения других задач, необходимых для извлечения информации. Для краткости будем говорить, что в первом случае система имеет искусственную многоканальность, а во втором случае — естественную.

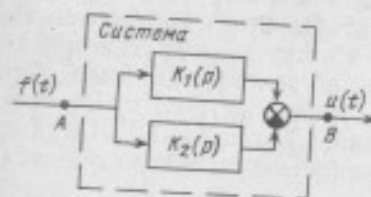


Рис. 2.6

Рассмотрим, например, двухканальную систему пеленгации с мгновенным фазовым сравнением сигналов (рис. 2.7). В этой системе информация об измеряемом угловом отклонении θ от плоскости xOy содержится в разности фаз сигналов $u_1(t)$ и $u_2(t)$, принятых антеннами A_1 и A_2 , фазовые центры которых O_1 и O_2 разнесены вдоль оси x на некоторое расстояние d . При ненаправленных антеннах и малых значениях угла θ

$$u_1(t) = U(t) \cos \varphi_1(t), \quad u_2(t) = U(t) \cos \varphi_2(t)$$

$$\text{и } \Delta \varphi(t) = \varphi_1(t) - \varphi_2(t) = (2\pi d/\lambda) \theta(t).$$

В этом случае сигнал, принятый каждой антенной в отдельности, не позволяет измерить угол θ . Следовательно, в данном случае двухканальность системы обусловлена не необходимостью компенсации каких-либо мешающих воздействий, а самим принципом извлечения информации (о размере угла θ), т. е. имеет не искусственный, а естественный характер.

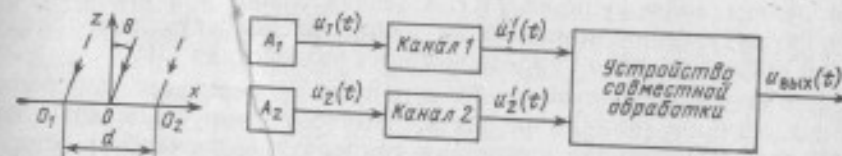


Рис. 2.7

Однако наличие в системе двух каналов позволяет при идентичных характеристиках этих каналов скомпенсировать ряд мешающих воздействий, в частности флуктуации фазы пеленгуемого радиосигнала и влияние неустойчивости температур, давления и влажности на фазовые характеристики каналов 1 и 2.

Примером искусственной компенсации является система компенсации помех, проникающих по боковым лепесткам диаграммы направленности основной антенны A_1 (рис. 2.8). В этом случае в отсутствие помех по боковым лепесткам

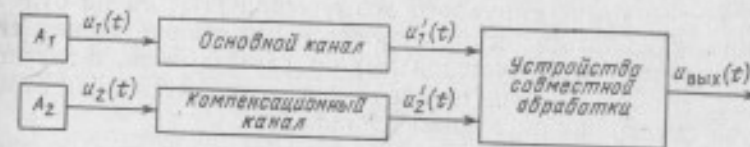


Рис. 2.8

компенсационный канал не нужен — полезные сообщения воспроизводятся основным каналом. Если же по какому-то боковому лепестку диаграммы направленности основной антенны (например, по первому) в основной канал проникает мощная внешняя помеха с известного направления, то действие помехи можно скомпенсировать приемом этой же помехи компенсационной антенной A_2 . При этом в некоторой степени может скомпенсироваться и полезный сигнал, проходящий по главному лепестку основной антенны A_1 ; однако при правильном выборе параметров компенсационного канала ослабление полезного сигнала может быть сделано допустимым.

Одним из первых в истории радиотехники применением искусственной компенсации была компенсация влияния неустойчивости температуры на резонансную частоту колебательного контура из L и C (при этом емкость C контура выполняется в виде параллельного соединения двух конденсаторов, имеющих температурные коэффициенты противоположных знаков).

В настоящее время в радиоэлектронике компенсационные методы (искусственные и естественные) находят весьма широкое применение для ослабления влияния самых разнообразных по своей физической природе и характеру действия источников ошибок. Основное преимущество принципа компенсации состоит в том, что для его применения часто достаточно весьма малой априорной информации об источниках ошибок. Например, если компенсация может быть осуществлена по схеме рис. 2.6, то при выполнении условия (2.79) она полностью устраняет влияние воздействия $f(t)$ при любом виде функции $f(t)$. Следовательно, в данном случае для осуществления компенсации нужно располагать лишь следующей информацией о характеристиках воздействия $f(t)$.

1. Знать физическую природу и пути прохождения мешающего воздействия (например, знать, что компенсируемым воздействием является влияние изменения температуры на частоту гетеродина приемника) для того, чтобы установить, что именно следует в данном случае понимать под каналами 1 и 2 системы.

2. Иметь уверенность, что интенсивность воздействия $f(t)$ не может быть столь велика, чтобы нарушить линейность каналов 1 и 2.

3. Иметь грубое представление о порядке низшей и высшей частот $F_{\text{н}}$ и $F_{\text{в}}$ спектра воздействия $f(t)$ с тем, чтобы обеспечить возможно более точное выполнение равенства (2.79) именно в диапазоне частот $F_{\text{н}} \dots F_{\text{в}}$.

Поскольку результат компенсации в малой степени зависит от характеристик компенсируемого возмущения $f(t)$, он не будет существенно ухудшаться, если в процессе функционирования системы характеристики возмущения $f(t)$ будут весьма значительно изменяться. Это обстоятельство является дополнительным важным преимуществом принципа компенсации.

Однако компенсационным методам присущи и весьма существенные недостатки, в частности следующие:

1. Не всегда возможно скомпенсировать мешающее воздействие, не ухудшив при этом условий прохождения на выход полез- ных сообщений.

2. Вследствие неидеальности и неустойчивости характеристик каналов невозможно обеспечить полную компенсацию — в большинстве случаев удается ослабить интенсивность мешающих воздействий не более чем на 10—15 дБ.

3. Введение каналов для компенсации некоторых воздействий может увеличить мешающее действие других источников ошибок или даже привести к появлению дополнительных источников оши-

бок. Поэтому при использовании принципа компенсации необходимо следить за тем, чтобы получаемые при этом преимущества не сопровождались существенными отрицательными последствиями.

2.9. ПРИНЦИП ИНВАРИАНТНОСТИ

Принцип инвариантности, как и принцип компенсации, состоит в том, чтобы обеспечить инвариантность (независимость) одного или нескольких из выходов $u_1(t), \dots, u_N(t)$ системы (рис. 2.4) относительно одного или нескольких возмущений $f_1(t), \dots, f_M(t)$. В этом смысле принципы инвариантности и компенсации совпадают. Некоторые различия между ними состоят лишь в следующем.

1. При использовании принципа компенсации заранее постулируется, что инвариантность должна быть достигнута путем такого построения системы, при котором действие возмущений в одних каналах системы компенсируется действием тех же возмущений в других каналах системы. При применении же принципа инвариантности такого постулирования не производится, т. е. заранее не оговаривается, что инвариантность должна быть достигнута применением компенсации. Отсюда на первый взгляд кажется, что принцип инвариантности носит более общий характер. Однако в теории инвариантности доказывается, что полную инвариантность можно обеспечить лишь при наличии в системе не менее двух каналов, и это достигается путем компенсации (так называемый принцип двухканальности Петрова). Иначе говоря, доказывается, что полная инвариантность может быть достигнута лишь при использовании компенсации.

2. Можно говорить о компенсации влияния мешающих воздействий, но нельзя говорить о компенсации полезных воздействий. В то же время можно говорить об обеспечении инвариантности (независимости) некоторого вспомогательного (промежуточного) выходного эффекта системы от полезного воздействия. Например, можно требовать инвариантности (независимости) сигнала рассогласования (сигнала ошибки) следящей системы относительно управляющего воздействия. В этом смысле принцип инвариантности действительно имеет несколько более общий характер, чем принцип компенсации.

Теория инвариантности начала интенсивно развиваться начиная с основополагающей статьи Г. В. Шипанова и последовавших за ней фундаментальных работ академиков Б. Н. Петрова, Н. Н. Лузина, В. С. Кулебакина и других ученых. К настоящему времени наиболее существенные для практики результаты этой теории получены для линейных стационарных замкнутых следящих систем при нулевых начальных условиях, в которых требуется обеспечить инвариантность одного из выходов системы $u(t)$ относительно одного из возмущений $f(t)$ [9, 10]. Так как поведе-

ние такой системы полностью характеризуется операторными передаточными функциями, то

$$u(t) = K(p) f(t), \quad (2.80)$$

где $K(p)$ — передаточная функция системы от входа $f(t)$ к выходу $u(t)$. Поэтому условие инвариантности имеет следующий вид: обеспечить

$$u(t) \equiv 0 \quad (2.81)$$

при любом виде воздействия $f(t)$ [требуется лишь, чтобы при любом виде $f(t)$ система оставалась линейной]. Из (2.80) и (2.81) следует, что условие инвариантности может быть представлено в виде

$$K(p) = 0. \quad (2.82)$$

Б. Н. Петров доказал, что необходимым (но не достаточным) условием выполнимости тождества (2.82) является возможность представить передаточную функцию системы в виде разности двух передаточных функций:

$$K(p) = K_1(p) - K_2(p) = 0. \quad (2.82')$$

Это означает, что в системе должно быть не менее двух каналов, по которым воздействие $f(t)$ проходит на рассматриваемый выход системы (принцип двухканальности Петрова). Кроме этого необходимого условия должны выполняться также соответствующие условия физической реализуемости и устойчивости системы.

Таким образом, для обеспечения полной (строгой) инвариантности требуется, чтобы условие (2.82) или (2.82') было физически реализуемым (т. е. чтобы для его выполнения не требовалось включения в состав системы физически нереализуемых звеньев) и не противоречило условиям устойчивости системы как при его точном выполнении, так и при некотором практически неизбежном нарушении [вследствие неидеальной идентичности и стабильности передаточных функций $K_1(p)$ и $K_2(p)$]. Кроме того, поскольку инвариантность обеспечивается применением компенсации, необходимо следить за отсутствием упоминавшихся выше возможных отрицательных последствий компенсации, например за тем, чтобы компенсация воздействия $f(t)$, проходящего на данный выход $u(t)$, не приводила к увеличению влияния других воздействий на данный и остальные выходные эффекты системы, а также к влиянию воздействия $f(t)$ на остальные $(N-1)$ выходных эффектов.

Практически процедура отыскания принципа построения инвариантной системы (в упомянутом выше классе линейных следящих систем) обычно сводится к следующему. Составляется структурная схема системы с указанием вида всех передаточных функций и мест приложения управляющего и мешающих воздействий (рис. 2.9). Поскольку система линейна, можно рассматривать действие различных возмущений независимо. Сначала рассматри-

вается возможность обеспечения инвариантности одного из выходов $u_j(t)$ относительно одного из воздействий $f_i(t)$ (при остальных воздействиях тождественно равных нулю). Для этого находится передаточная функция $K(p)$ от $f_i(t)$ к $u_j(t)$. Если (что обычно имеет место) условие инвариантности (2.82) не выпол-

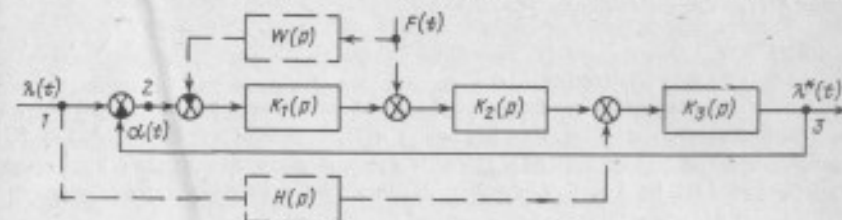


Рис. 2.9

няется, то полагают, что $K(p) = K'(p)$, и пытаются создать дополнительный (компенсационный) канал с передаточной функцией $[-K''(p)]$ от $f_i(t)$ к $u_j(t)$, удовлетворяющей условию инвариантности

$$K''(p) = K'(p). \quad (2.83)$$

Если в принципе такой дополнительный канал создать возможно, то проверяют физическую реализуемость требуемых для создания этого канала дополнительных (корректирующих) звеньев и устойчивость системы как при идеальном выполнении условия (2.83), так и при его небольших нарушениях [вследствие отклонений $K''(p)$ от $K'(p)$]. Если все эти требования выполняются, то это значит, что требуемая инвариантность выхода $u_j(t)$ относительно возмущения $f_i(t)$ может быть обеспечена. После этого следует проверить отсутствие упомянутых отрицательных последствий компенсации с учетом всех остальных $(M-1)$ возмущений и всех выходов.

Во многих случаях обеспечить даже в принципе полную (строгую) инвариантность выходного эффекта $u_j(t)$ относительно входного воздействия не удастся [например, вследствие того, что требуемые для выполнения условия (2.83) корректирующие звенья оказываются физически нереализуемыми или нарушаются условия устойчивости системы]. Однако при этом часто все же удается обеспечить ослабление мешающего эффекта на выходе $u_j(t)$ (от воздействия $f_i(t)$) хотя и не до нуля, но до некоторой достаточно малой величины ε . При этом обеспечивается, как говорят, инвариантность системы с точностью до ε или, короче, ε -инвариантность (почти — инвариантность) системы. Практически, т. е. с учетом не только принципиальных, но и технических возможностей, всегда удастся реализовать лишь ε -инвариантность.

Для иллюстрации приведенных положений рассмотрим применение принципа инвариантности к следящей системе, структурная схема которой изображена

на рис. 2.9 сплошными линиями. На систему действуют управляющее воздействие $\lambda(t)$ и помеха $F(t)$. Ошибка слежения $\alpha(t) = \lambda(t) - \lambda^*(t)$. Поскольку система линейна, можно рассматривать эффекты воздействий $\lambda(t)$ и $F(t)$ независимо. Рассмотрим сначала действие лишь $\lambda(t)$, т. е. положим $F(t) = 0$. Выясним возможность обеспечения инвариантности (независимости) ошибки $\alpha(t)$ от воздействия $\lambda(t)$, т. е. выполнения условия

$$\alpha(t) = 0 \quad (2.84)$$

при любом законе изменения $\lambda(t)$.

Очевидно, выполнение этого условия обеспечивает равенство нулю динамической составляющей ошибки слежения $\alpha(t)$. В соответствии с (2.82) и (2.83) для выполнения условий (2.84) должны существовать два канала прохождения воздействия $\lambda(t)$ от точки 1 к точке 2 (рис. 2.9) с передаточными функциями $K'(p)$ и $-K''(p)$, где

$$K''(p) = K'(p). \quad (2.85)$$

Из рис. 2.9 следует, что в системе существует лишь один канал с передаточной функцией $K'(p) = 1$. Поэтому необходимо создать второй канал с передаточной функцией $-K''(p)$, удовлетворяющей условию

$$K''(p) = K'(p) = 1. \quad (2.86)$$

Из рис. 2.9 следует, что такой канал может быть создан, если удастся создать дополнительную корректирующую цепь, показанную на рис. 2.9 штриховой линией и имеющую передаточную функцию $H(p)$, удовлетворяющую условию

$$H(p)K_3(p) = 1. \quad (2.87)$$

Действительно, при этом воздействие $\lambda(t)$ может пройти от точки 1 к точке 2 через звенья $H(p)$ и $K_3(p)$ и далее по цепи основной обратной связи. При этом образуется второй канал с передаточной функцией¹

$$K''(p) = H(p)K_3(p)$$

и при выполнении условия (2.82) оказывается

$$K''(p) = 1,$$

т. е. выполняется условие инвариантности (2.86). Таким образом, выполнение условия (2.87) обеспечивает инвариантность ошибки относительно воздействия $\lambda(t)$, т. е. нулевую динамическую ошибку при любом виде воздействия $\lambda(t)$.

Однако точное выполнение условия (2.87) физически невозможно. Действительно, условие (2.87) можно представить в виде $H(j\omega) \cdot K_3(j\omega) = 1$ для всех $\omega = 0 \dots \infty$. Но у любого физически реализуемого блока с учетом всегда существующих паразитных параметров модуль коэффициента передачи при $\omega \rightarrow \infty$ стремится к нулю. Поэтому при $\omega \rightarrow \infty$

$$K_3(j\omega) \rightarrow 0, H(j\omega) \rightarrow 0$$

и, следовательно, $H(j\omega)K_3(j\omega) \rightarrow 0$, что противоречит условию (2.87). Поэтому условие (2.87) можно выполнить лишь приближенно в ограниченном диапазоне частот и, следовательно, обеспечить лишь ϵ -инвариантность. Кроме того, в радиотехнических следящих системах физически невозможно реализовать

¹ Знак минус перед $H(p)K_3(p)$ здесь не ставится, так как выше было принято, что передаточная функция от точки 1 к точке 2 (рис. 2.9) равна $-K''(p)$, а не $K''(p)$.

требуемый в соответствии со схемой рис. 2.9 «отвод» от управляющего воздействия $\lambda(t)$. Это объясняется тем, что в таких следящих системах слежение осуществляется не за самим входным сигналом реальной физической (технической) системы, а лишь за каким-либо параметром этого сигнала. Например, в системах фазовой автоподстройки (ФАП) происходит слежение за фазой $\varphi(t)$ входного сигнала, т. е. $\lambda(t) = \varphi(t)$. На входе же реальной системы ФАП действует не фаза $\varphi(t)$, а смесь сигнала $U \cos \varphi(t)$ с помехами. Поэтому в радиотехнических следящих системах для создания второго канала от $\lambda(t)$ к $\alpha(t)$ необходимо дополнительное измерительное устройство для измерения параметра $\varphi(t)$.

Рассмотрим теперь возможность обеспечения инвариантности выхода $\lambda^*(t)$ относительно помехи $F(t)$ [при $\lambda(t) = 0$]. В схеме на рис. 2.9 существует всего один канал, по которому помеха $F(t)$ проходит на выход системы, т. е. в точку 3. Он имеет передаточную функцию¹

$$K'(p) = K_2(p)K_3(p). \quad (2.88)$$

Для образования второго канала с передаточной функцией $-K''(p) = -K'(p)$ можно применить корректирующую цепь с передаточной функцией $W(p)$, показанную на рис. 2.9 штриховой линией. Действительно, при этом становится возможным прохождение помехи $F(t)$ к точке 3 последовательно через звенья $W(p)$, $K_1(p)$, $K_2(p)$ и $K_3(p)$. В этом случае, очевидно,

$$K''(p) = W(p)K_1(p)K_2(p)K_3(p) \quad (2.89)$$

и условия инвариантности (2.82') и (2.83) принимают вид $K''(p) = K'(p)$, т. е.

$$K_2(p)K_3(p) = W(p)K_1(p)K_2(p)K_3(p).$$

Отсюда следует

$$W(p)K_1(p) = 1 \quad (2.90)$$

или, что эквивалентно,

$$W(j\omega) = 1/K_1(j\omega) \quad (2.91)$$

для всех $\omega = 0 \dots \infty$.

Очевидно, что с учетом паразитных параметров можно выполнить условия (2.91) лишь приближенно и в ограниченном диапазоне частот, т. е. обеспечить лишь ϵ -инвариантность системы относительно помехи $F(t)$.

Кроме того, выше предполагалось, что можно организовать «подачу» помехи $F(t)$ на вход блока $K_1(p)$ [через звено с передаточной функцией $W(p)$]. Однако схема на рис. 2.9 в общем случае является лишь эквивалентной схемой реальной системы, в которой реальная помеха приведена (пересчитана) ко входу звена с передаточной функцией $K_2(p)$, т. е. $F(t)$ — это некоторое приведенное значение помехи, не обязательно совпадающее с самой реальной помехой. Например, в радиотехнической системе слежения по угловым координатам [при отсутствии блока $K_1(p)$] воздействие $F(t)$ может соответствовать реальной помехе, действующей на вход антенной системы пеленгатора, но приведен-

¹ В рассматриваемом случае, т. е. при нахождении условия компенсации $-K''(p) = -K'(p)$, передаточные функции от источника $F(t)$ к точке 3 можно определять без учета обратной связи, выходящей из точки 3, так как по условию компенсации эффект в этой точке, создаваемый источником $F(t)$, должен быть равен нулю и, следовательно, указанная обратная связь не может на него повлиять.

ной к угловой ошибке $\alpha(t)$, т. е. в этом случае реальная помеха $E_n(t)$ и ее эквивалент $F(t)$ не совпадают даже по размерности: $E_n(t)$ — напряженность поля помехи, например, в вольтах на метр, а $F(t)$ — ее эквивалент в угловых единицах, например в радианах или в градусах. Поэтому в ряде случаев для реализации корректирующей цепи $W(p)$ может потребоваться дополнительный измеритель напряжения (напряженности поля) помехи.

Из приведенных примеров следует, что во многих случаях (особенно в радиотехнических системах) для реализации даже ε -инвариантности может потребоваться применение совокупности нескольких измерителей полезных и (или) мешающих возмущений и совместная обработка получаемых от них данных, т. е. так называемое комплексирование измерителей. Иначе говоря, в большинстве случаев требуется сочетание принципов инвариантности и комплексирования.

2.10. ПРИНЦИП КОМПЛЕКСИРОВАНИЯ

Под комплексированием понимается совместная обработка результатов воспроизведения различными устройствами (системами) одного и того же сообщения или нескольких различных, но статистически связанных сообщений. Целью комплексирования является получение точности воспроизведения сообщений более высокой, чем у отдельного устройства (измерителя).

Принципиальная возможность повышения точности обусловлена при этом двумя основными факторами. Во-первых, комплексированная система оказывается по отношению к каждому сообщению не одноканальной, а многоканальной, что позволяет использовать рассмотренные выше принципы компенсации и инвариантности. Во-вторых, воспроизводимые сообщения в отдельных устройствах полностью или частично статистически связаны, а многие из аддитивных мешающих воздействий, возникающих в отдельных устройствах, обычно статистически независимы; поэтому при объединении выходов отдельных устройств суммирование полезных эффектов оказывается близким к арифметическому, а мешающих — к квадратичному. Вследствие этого относительная интенсивность мешающих эффектов на выходе уменьшается.

Комплексирование может в принципе применяться в самых разнообразных информационных системах и устройствах. Однако наибольшее применение оно получило в измерительных системах и системах управления, так как именно в таких системах часто имеет место измерение одних и тех же или статистически связанных параметров устройствами различной физической природы. Например, часто применяется комплексирование радиолокационного измерителя дальности r до объекта и измерителя его радиальной скорости V_r , являющейся первой производной этой дальности (или измерителей угловых координат и угловых скоростей направления на объект), причем измерение дальности и скорости

(или угловых координат и угловых скоростей) основывается на различных физических принципах. В системах управления летательными аппаратами часто применяется комплексирование радиотехнических, тепловых и инерциальных измерителей.

Однако при комплексировании измерителей приходится преодолевать ряд трудностей, которые могут существенно снизить эффективность комплексирования. К таким трудностям относятся, в первую очередь, необходимость согласования систем координат, в которых производится измерение различными измерителями, а иногда и начала отсчета времени. Кроме того, приходится согласовывать передаточные функции измерителей. Поскольку эти передаточные функции могут иметь совершенно различную физическую природу, изменение внешних условий (температуры, влажности, давления и др.) влияет на них по-разному. Поэтому согласование, проведенное в некоторых определенных условиях, в дальнейшем может существенно нарушаться.

Анализ принципов комплексирования в сжатой форме приведен в ряде книг ([10, 11] и др.).

Литература: основная [1, 10, 11, 25], дополнительная [8, 9, 24, 45, 46].

Глава 3

ПОМЕХОЗАЩИЩЕННОСТЬ

3.1. ОБЩАЯ ХАРАКТЕРИСТИКА ПОКАЗАТЕЛЕЙ ПОМЕХОЗАЩИЩЕННОСТИ И ПУТЕЙ ИХ УЛУЧШЕНИЯ

Помехозащищенность радиосистемы характеризует ее способность обеспечивать требуемую точность воспроизведения сообщений и пропускную способность (быстродействие) с учетом действия возможных помех. В общем случае помехозащищенность системы обеспечивается помехоустойчивостью и скрытностью ее действия. Скрытность действия затрудняет возможному противнику обнаружение факта функционирования данной радиосистемы и определение характеристик излучаемых ею радиосигналов с целью создания эффективных преднамеренных радиопомех или уничтожения системы наводимыми на нее снарядами. Помехоустойчивость обеспечивает нормальное функционирование системы в условиях действия определенной совокупности непреднамеренных и преднамеренных (организованных) помех. В большинстве радиосистем невоенного назначения учитывать возможные действия противника и соответственно обеспечивать скрытность действия системы не требуется. При этом помехозащищенность системы совпадает с ее помехоустойчивостью.

Скрытность действия системы может обеспечиваться активными и пассивными способами. Активные способы сводятся к радиомаскировке, т. е. излучению дополнительных ложных радиосигналов, настолько близких к основным рабочим радиосигналам, чтобы противник не мог отличить ложные сигналы от рабочих, но в то же время настолько отличающихся от рабочих сигналов, чтобы не создать существенных дополнительных помех маскируемой системе. Пассивные способы должны обеспечить скрытность действия системы даже в условиях отсутствия радиомаскировки.

Наивысшая степень скрытности, которую будем называть абсолютной скрытностью, обеспечивается, если противник не может обнаружить даже факта излучения радиосигнала. Такая скрытность может быть достигнута при применении сложных (шумоподобных) сигналов, если выполняется условие

$$N_{oc} \leq \varepsilon N_{osh} \quad (3.1)$$

где ε — достаточно малая величина (например, 10^{-2}); N_{osh} — спектральная плотность внутреннего шума радиоприемного устройства противника, приведенного к выходу антенной системы этого устройства; N_{oc} — спектральная плотность шумоподобного сигнала на выходе упомянутой антенной системы. Очевидно,

$$N_{oc} = P_c / \Delta f_c \quad (3.2)$$

где Δf_c — ширина спектра шумоподобного сигнала, а P_c — его средняя мощность на выходе антенной системы разведприемника.

Из (3.1) и (3.2) следует прежде всего, что для обеспечения абсолютной скрытности требуется уменьшать среднюю мощность P_c сигнала. Последнее может обеспечиваться рядом мер, включающих уменьшение излучаемой мощности рабочего радиосигнала (что требует повышения чувствительности радиоприемного устройства защищаемой системы и, следовательно, повышения его помехоустойчивости), выбор (если это возможно) такой рабочей волны, которая распространялась бы лишь в пределах прямой видимости, формирование (если это возможно) диаграммы направленности излучения радиопередающего устройства, дающей в направлении на противника минимум излучения.

Из (3.2) на первый взгляд кажется, что чем больше ширина спектра сигнала Δf_c при данной (неизменной) его средней мощности P_c , тем выше скрытность системы. Однако в действительности это не всегда так. Если значение P_c сравнительно невелико (например, вследствие большой удаленности разведприемника от подавляемой системы), то расширение полосы от некоторого значения Δf_{c1} , при котором условие (3.1) еще существенно не выполняется ($N_{oc}/N_{osh} \gg \varepsilon$), до значения Δf_{c2} (где $\Delta f_{c2} \gg \Delta f_{c1}$) может привести к тому, что неравенство (3.1) будет приближенно выполняться и скрытность приблизится к абсолютной, т. е. действительно повысится. Однако в другом случае, когда мощность P_c велика (например, вследствие близости разведприемника к передатчику подавляемой системы), условие (3.1) будет далеко от выполнения

даже при увеличении ширины спектра от Δf_{c1} до некоторого значения Δf_{c2} и факт наличия излучения передатчика вполне может быть обнаружен. В то же время увеличение ширины спектра излучаемого сигнала может ускорить обнаружение излучения разведприемником. Поэтому при большом значении мощности P_c увеличение ширины спектра Δf_c может привести не к улучшению, а к ухудшению скрытности. Таким образом, расширение ширины спектра сигнала Δf_c повышает скрытность излучения по отношению к весьма удаленным разведприемникам, но понижает по отношению к расположенным на малом расстоянии.

Обеспечение абсолютной скрытности возможно далеко не всегда, поэтому приходится применять дополнительные меры для повышения скрытности; к ним относятся, в частности, уменьшение времени излучения радиосигнала и изменение его параметров во времени по случайным (для противника) законам. Поскольку обеспечение скрытности действия требуется далеко не во всех задачах проектирования, в дальнейшем будет рассматриваться лишь проблема обеспечения помехоустойчивости системы.

Любая радиосистема подвержена действию разнообразных помех; поэтому обеспечение требуемой помехоустойчивости является необходимым и в значительной мере определяет принципы построения системы и значения большинства ее параметров. В качестве показателей помехоустойчивости системы могут использоваться все рассмотренные в гл. 2 показатели точности воспроизведения сообщений — средний риск, средний квадрат ошибки, ее дисперсия, математическое ожидание и др. При этом задача обеспечения требуемой помехоустойчивости может формулироваться, например, в следующем виде:

при заданных видах и характеристиках помех (и заданных требованиях к пропускной способности или быстродействию системы) обеспечить значения показателей точности, не выходящие за пределы допустимых, например обеспечить выполнение условия $P_{osh} \leq P_{osh, макс}$, где $P_{osh, макс}$ — максимально допустимое значение полной вероятности ошибки P_{osh} .

Помехоустойчивость системы тем выше, чем лучше (меньше) значения соответствующих показателей точности при заданных видах и параметрах помех или чем в большем диапазоне видов помех и значений их параметров обеспечивается получение заданных (допустимых) значений показателей точности. Основными путями повышения помехоустойчивости являются следующие.

1. Выбор в качестве рабочего такого диапазона волн, в котором действие помех минимально. Обычно это диапазон дециметровых или более коротких волн.

2. Улучшение энергетики радиоприемных, т. е. повышение отношения средней мощности (энергии) сигнала к мощности (энергии) аддитивных независимых помех на выходе антенных систем радиоприемных устройств. Такое улучшение обеспечивается, в первую очередь, уменьшением шумовой температуры радиоприемных устройств, увеличением средней мощности радиопередающих ус-

тройств и коэффициентов усиления передающих и приемных антенных систем (для увеличения энергии сигнала желательно также увеличивать время T , отводимое на воспроизведение сообщений; однако при этом уменьшаются быстродействие и пропускная способность системы).

3. Применение в радиосистеме помехоустойчивых принципов пространственно-временной обработки информации (ОПВ). Поскольку эти принципы в наибольшей мере определяют структуру и значения параметров проектируемой (информационной) радиосистемы, рассмотрим этот вопрос более подробно.

В общем случае выбор ОПВ включает выбор организации связи, типов и характеристик антенных систем, применяемых несущих колебаний сигналов, видов их модуляции сообщениями, принципов демодуляции, кодирования и декодирования сообщений, линейной и нелинейной фильтрации, способов компенсации, комплексирования и других видов обработки сигналов и сообщений. Под организацией связи здесь понимается организация в целом процесса извлечения и (или) передачи и обработки информации. Она включает в первую очередь выбор числа радиолиний, их состава (приемные или приемно-передающие радиолинии), рабочих диапазонов волн, размещения в пространстве радиопередающих и радиоприемных устройств, принципов взаимодействия отдельных радиолиний. Поэтому при окончательном выборе принципов ОПВ должна учитываться вся совокупность показателей качества проектируемой системы.

В общем случае ОПВ не распадается на независимые (отдельные) пространственную обработку (ОП) и временную обработку (ОВ) и должна рассматриваться как единая пространственно-временная обработка. В дальнейшем ради краткости будем рассматривать ее применительно лишь к одной радиолинии системы, а также полагать, что радиолиния содержит в общем случае как радиоприемное, так и радиопередающее устройство. При этом ОПВ радиолинии состоит из двух процедур — пространственно-временной обработки в радиопередающем¹ и в радиоприемном устройствах, которые будем обозначать соответственно ОПВ₁ и ОПВ₂. В общем случае обе эти процедуры также нельзя выбирать (оптимизировать) независимо друг от друга, т. е. необходимо синтезировать одновременно пару (ОПВ₁, ОПВ₂).

Результаты синтеза этой пары существенно зависят от числа и характеристик учитываемых при синтезе помех. Однако в общем случае для разных принципов построения системы существенными могут оказаться не одни и те же, а различные наборы видов помех и значений их параметров (особенно в случае систем военного назначения). Поэтому, строго говоря, необходимо одновременно (совместно) производить синтез тройки (виды и характеристики помех, ОПВ₁, ОПВ₂). Однако строгое математическое

решение такой задачи обычно невыполнимо и приходится применять метод последовательных приближений. При этом на первом этапе проектирования производится синтез пары (ОПВ₁, ОПВ₂) для выбранного в значительной мере эвристически набора помех (и их характеристик). На втором этапе уточняется, какие из учетных и дополнительных видов помех существенны для найденной пары (ОПВ₁, ОПВ₂), и для этого уточненного набора помех находится уточненная пара (ОПВ₁, ОПВ₂) пространственно-временной обработки. Затем, если потребуется, дополнительно последовательно уточняются виды и характеристики помех и принципы пространственно-временной обработки.

В большинстве случаев даже при полностью заданных характеристиках помех произвести достаточно строгим математическим методом совместный синтез всех частей пары (ОПВ₁, ОПВ₂) не удастся. Поэтому обычно задача синтеза пары (ОПВ₁, ОПВ₂) разбивается на ряд более частных и соответственно более простых подзадач, последовательное решение которых позволяет найти близкий к оптимальному принцип пространственно-временной обработки. Наиболее типичными из таких частных подзадач являются следующие.

Задача 1. Синтез характеристик передающего и приемного антенных устройств (их диаграмм направленности и поляризационных характеристик) при заданных принципах временной обработки.

Задача 2. Синтез временной обработки (пары «сигнал — фильтр») при заданных характеристиках антенных устройств.

Задача 3. Синтез ОПВ в приемном устройстве при заданной ОПВ в передающем устройстве.

Задача 4. Синтез временной обработки в приемном устройстве.

Задача 5. Синтез временной обработки в передающем устройстве при условии, что временная обработка в приемном устройстве выбирается согласованной (в заранее установленном смысле) с обработкой в передающем устройстве. Важнейшей подзадачей задачи 5 является синтез формы несущего колебания сигнала.

В последующих параграфах рассматриваются в основном лишь задачи 2, 3 и 4. При этом количественные результаты приводятся ради краткости, лишь для случаев, когда сигнал не содержит неизвестных (в месте приема) неинформационных параметров, т. е. известен точно, за исключением значений переносимых им сообщений. Очевидно, наличие у сигнала неизвестных неинформационных параметров может лишь ухудшить помехоустойчивость системы.

¹ Все или часть операций (обработки), проводимых в передатчике для образования излучаемого сигнала, называют также формированием сигнала.

3.2. СИНТЕЗ ОПТИМАЛЬНОЙ ПО ПОМЕХОУСТОЙЧИВОСТИ ВРЕМЕННОЙ ОБРАБОТКИ (ПАРЫ «СИГНАЛ — ФИЛЬТР»)

Математический синтез пары «сигнал — фильтр»¹ при произвольных характеристиках помех весьма сложен. Он резко упрощается, если единственной помехой является аддитивный независимый стационарный нормальный белый шум $u_m(t)$ с известной спектральной плотностью N_0 (например, внутренний шум радиоприемного устройства). В этом случае смесь сигнала и помехи на входе радиоприемного устройства на интервале $[0, T]$ воспроизведения сообщения x может быть представлена в виде

$$y(t) = u_c(t, x) + u_m(t), \quad (3.3)$$

где $u_c(t, x)$ — сигнал произвольной формы, известный в месте приема точно, за исключением неизвестного сообщения x с априорным распределением $P(x)$.

При этом синтез пары «сигнал — фильтр» может быть проведен в два этапа. На первом отыскивается оптимальный (в смысле выбранного критерия) приемник («фильтр») для сигнала, записанного в общем виде $u_c(t, x)$, т. е. без предварительной конкретизации его формы. На втором этапе определяется, как должна быть конкретизирована функция $u_c(t, x)$, чтобы найденный оптимальный приемник давал наилучшие [по всем возможным видам сигнала $u_c(t, x)$] результаты в смысле выбранного для приемника критерия оптимальности. Тем самым находится пара «сигнал — фильтр», наилучшая (по помехоустойчивости) среди всех возможных комбинаций способов временной обработки в передатчике и приемнике.

В соответствии с этой процедурой рассмотрим сначала синтез оптимального приемника. Апостериорное распределение сообщения x при известной реализации $y(t)$

$$P(x|y) = k P(x) P(y|x). \quad (3.4)$$

Здесь k — константа, не зависящая от x и поэтому не влияющая на результаты синтеза; $P(y|x)$ — плотность вероятности реализации $y(t)$ при данном x .

Без потери общности можно полагать, что процесс $y(t)$ имеет спектр, ограниченный некоторой сколь угодно высокой частотой f_n , так что реализацию $y(t)$ можно рассматривать как упорядоченную совокупность (вектор) выборочных значений, т. е. полагать

$$y(t) = \langle y_1, \dots, y_i, \dots, y_n \rangle, \quad (3.5)$$

где $y_i = y(i\Delta t)$; $n = T/\Delta t = 2f_n T \gg 1$.

¹ Здесь и далее термин «сигнал — фильтр» является сокращенным обозначением понятия «совокупность алгоритмов временной обработки в передающем и приемном устройствах», т. е. фильтр может включать в себя как линейные, так и нелинейные устройства.

Так как помеха $u_m(t)$ в данном случае имеет распределение

$$W_m(u_m) = c \exp \left[-\frac{1}{N_0} \int_0^T u_m^2(t) dt \right]$$

(где c — не существенная для дальнейшего константа) и является независимой по отношению к сообщению x , то из (3.3) следует, что

$$P(y|x) = W_m(y - u_c) = c \exp \left[-\frac{1}{N_0} \int_0^T [y - u_c(t, x)]^2 dt \right],$$

и (3.4) принимает вид

$$P(x|y) = c' P(x) \exp \left[-\frac{1}{N_0} \int_0^T [y - u_c(t, x)]^2 dt \right]. \quad (3.6)$$

Здесь c' — несущественная константа, поскольку решение о значении x , которое должно быть выдано на выходе приемного устройства, зависит лишь от формы распределения $P(x|y)$, но не от его масштаба по оси ординат.

Выражение (3.6) можно представить в виде

$$P(x|y) = c'' P(x) \exp \left[\frac{2}{N_0} \int_0^T y u_c(t, x) dt \right] \exp \left[-\frac{Q_x}{N_0} \right]. \quad (3.7)$$

Здесь

$$c'' = c' \exp \left[-\frac{1}{N_0} \int_0^T y^2(t) dt \right]$$

— константа [для данной реализации $y(t)$], не зависящая от x и поэтому не влияющая на результаты синтеза;

$$Q_x = \int_0^T u_c^2(t, x) dt$$

— удельная (т. е. выделяемая на единичном сопротивлении) энергия сигнала при данном значении сообщения x . Для широкого класса сигналов величина Q_x не зависит от x . Априорное распределение $P(x)$ сообщений, как отмечалось в гл. 2, часто можно также полагать не зависящим от x в пределах $[x_{\min}, x_{\max}]$ и равным нулю вне этих пределов. Поэтому, как следует из (3.7), вид функции $P(x|y)$ определяется главным образом корреляционным интегралом

$$J(x) = \int_0^T y u_c(t, x) dt. \quad (3.8)$$

Структура оптимального приемника определяется не только корреляционным интегралом (3.8), но и характером сообщения x (дискретная величина, непрерывная величина, функция времени), а также видом критерия оптимальности. Однако в ряде работ по теории оптимального приема (например, [5, 23, 40]) показано,

что при указанном виде помехи структура оптимального приемного устройства может быть представлена в виде, изображенном на рис. 3.1. Вычислитель $J(x)$ должен определять значения корреляционного интеграла (3.8) для всех возможных значений воспроизводимого сообщения x и выдавать результаты этого вычисления на решающее устройство. Решающее устройство на основе поступившей на его вход информации должно вынести решение x^* .

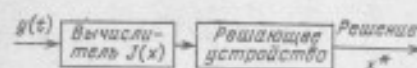


Рис. 3.1

В подавляющем большинстве случаев наиболее сложной частью оптимального приемника является вычислитель корреляционного интеграла, а не решающее устройство. В частности, если сообщение является дискретной случайной величиной (независимой в разных сеансах его воспроизведения), то, как показано в § 11.2, задача решающего устройства при любом критерии оптимальности сводится к простому линейному безынерционному преобразованию совокупности значений $J(x_1), \dots, J(x_m)$ корреляционного интеграла $J(x)$, выдаваемых его вычислителем.

Обработка поступающей на вход приемника смеси $y(t)$ по закону (3.8) называется согласованной [с формой сигнала $u_c(t, x)$], так как каждое значение интеграла $J(x)$ определяется для соответствующего ожидаемого сигнала $u_c(t, x)$, т. е. для соответствующего образца (копии) сигнала. В случае дискретного сообщения достаточно вычислить m значений интеграла (3.8) для m образцов сигнала $u_c(t, x_1), \dots, u_c(t, x_m)$. В частности, в задаче обнаружения сигнала $m=2$,

$$u_c(t, x_1) = u_c(t), u_c(t, x_2) = 0 \quad (3.9)$$

и, как следует из (3.8), достаточно вычислить одно значение корреляционного интеграла

$$J(x_1) = \int_0^T y u_c(t) dt, \quad (3.10)$$

соответствующее образцу (копии) обнаруживаемого сигнала $u_c(t)$. При этом задача решающего устройства (в схеме на рис. 3.1) сводится к сравнению значения $J(x_1)$ с некоторым порогом U_0 .

Таким образом, для рассмотренного вида помехи оптимальная обработка сигнала в приемном устройстве сводится к обработке, согласованной с ожидаемым сигналом $u_c(t, x)$ и состоящей в основном в вычислении корреляционного интеграла вида (3.8).

Поскольку при получении этого результата на форму сигнала $u_c(t, x)$ (при данном x) не накладывалось никаких ограничений (за исключением того, что эта форма должна быть точно известна в месте приема), задача отыскания оптимальной пары «сигнал — фильтр» сводится к указанному выше второму этапу синтеза — отысканию такой формы сигнала $u_c(t, x)$ [точнее, такого закона преобразования сообщения x в сигнал $u_c(t, x)$], которая при условии согласованной (в указанном выше смысле) с этим сигналом

обработки обеспечивает минимальное (наилучшее) значение показателя точности воспроизведения сообщения (например, минимальное значение среднего риска, или полной вероятности ошибки, или среднего квадрата ошибки воспроизведения сообщений).

Для дискретных сообщений получены достаточно полные результаты решения этой задачи, которые сводятся к следующему (см., например, [5]).

1. При обнаружении сигнала, т. е. в случае (3.9), качество обнаружения (полная вероятность ошибки $P_{\text{ош}}$ или средний риск R) не зависит от формы сигнала и определяется лишь значением его удельной энергии

$$Q = \int_0^T u_c^2(t) dt. \quad (3.11)$$

В частности, при $P(x_1) = P(x_2) = 0,5$ энергия сигнала, требуемая для обеспечения заданной полной вероятности ошибки $P_{\text{ош}}$ (при $P_{\text{ош}} \leq 10^{-3}$),

$$Q \approx 4 N_0 \ln(1/P_{\text{ош}}). \quad (3.12)$$

2. При воспроизведении сообщения, которое может иметь одно из m равновероятных значений, и использовании сигналов с одинаковыми энергиями

$$Q(x_1) = \dots = Q(x_m) = Q. \quad (3.13)$$

(т. е. при различении m равновероятных сигналов с равными энергиями) оптимальными являются так называемые симплексные сигналы. При этом энергия сигнала, обеспечивающая получение заданной полной вероятности ошибки $P_{\text{ош}}$ (при $P_{\text{ош}} \leq 0,1$)

$$Q \approx N_0(m-1)[2 \ln(1/P_{\text{ош}}) + \ln(m-1) - 2,8]/m. \quad (3.14)$$

Если вместо симплексных применить ортогональные сигналы, т. е. сигналы, удовлетворяющие условию ортогональности

$$\int_0^T u_c(t, x_i) u_c(t, x_j) dt = \begin{cases} 0 & \text{при } i \neq j, \\ Q & \text{при } i = j, \end{cases} \quad (3.15)$$

то требуемая для получения той же вероятности $P_{\text{ош}}$ энергия сигнала увеличивается по сравнению с (3.14) в

$$Q_{\text{орт}}/Q_{\text{опт}} = m/(m-1) \quad (3.16)$$

раз, т. е. достигает (при $P_{\text{ош}} \leq 0,1$) значения

$$Q = N_0[2 \ln(1/P_{\text{ош}}) + \ln(m-1) - 2,8]. \quad (3.17)$$

Из (3.16) следует, что при $m \geq 3$ ортогональные сигналы дают сравнительно небольшой проигрыш по сравнению с симплексными, который с ростом m быстро уменьшается. С другой стороны, система с ортогональными сигналами обычно проще в реализации. Поэтому симплексные сигналы находят применение главным образом лишь при $m=2$. В этом случае они сводятся к противоположным сигналам, т. е. удовлетворяют условию

$$u_c(t, x_2) = -u_c(t, x_1) \quad (3.18)$$

(к этому классу относятся, в частности, сигналы с манипуляцией фазы на 180°).

Таким образом, в случае воспроизведения дискретных сообщений на фоне единственной помехи в виде аддитивного независимого нормального белого шума удастся сравнительно просто найти оптимальную пару «сигнал — фильтр»: оптимальным фильтром является линейная согласованная в смысле (3.8) обработка, осуществляемая корреляторами или согласованными фильтрами. При этом оптимальными сигналами являются симплексные, если требуется, чтобы энергии всех возможных сигналов $u_c(t, x_1), \dots, u_c(t, x_m)$ не превосходили допустимого значения Q . При $m > 2$ близкими к оптимальным являются ортогональные сигналы. Если же требуется лишь обнаружить сигнал $[m=2, u_c(t, x_2)=0]$, помехоустойчивость определяется энергией Q сигнала и не зависит от его формы.

При изменении условий синтеза, в первую очередь при изменении вида помех, приведенные выше результаты синтеза оптимальной пары «сигнал — фильтр» могут оказаться неприменимыми. В частности, возникают следующие трудности и особенности.

1. При ряде видов помех и тем более для совокупности различных помех оптимальную пару «сигнал — фильтр» до сих пор не удалось найти.

2. Для коррелированных помех (в частности, пассивных) обработка, согласованная в смысле (3.8), не является оптимальной.

3. При обнаружении флуктуирующего сигнала качество обнаружения оказывается зависящим не только от энергии сигнала, но и от его формы. Так, для уменьшения влияния флуктуаций сигнала при его отражении от цели (в системах радиолокации) или при многолучевом распространении (в системах СПИ) может оказаться целесообразным применение сигнала, энергия которого существенно разнесена по спектру частот, например двух- или трех-частотного.

3.3. СИНТЕЗ ПОМЕХОУСТОЙЧИВОЙ ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ ОБРАБОТКИ В ПРИЕМНИКЕ

Пространственно-временная обработка в приемных устройствах в широком смысле этого понятия включает совместную обработку колебаний, принятых антенными элементами или подсистемами, расположенными в различных пунктах пространства, включая и пункты, разнесенные на большие расстояния в пределах земного пространства или даже за его пределами. В частности, обработка такого типа имеет место в системах управления воздушным движением, многопозиционных системах радиолокации, спутниковых системах связи и в ряде других сложных систем. Называемая *многопозиционной* или *разнесенной*, она может включать несколько иерархических ступеней — первичную обработку, вторичную и т. д. При этом в каждом из разнесенных пунктов производится первичная, а иногда и вторичная обработка, а по-

следующие ступени ее реализуются при совместной обработке информации, поступающей от различных пунктов. Выбор числа разнесенных пунктов и их расположения в пространстве относят обычно к организации процесса извлечения или передачи информации.

В узком смысле под пространственно-временной понимают совместную обработку колебаний, принятых различными антенными подсистемами, расположенными в пределах одного пункта (места) пространства, имеющего сравнительно небольшую протяженность (обычно несколько десятков метров или меньше). При этом каждая антенная подсистема может состоять как из элементарного вибратора с весьма слабонаправленным действием, так и из совокупности дискретных или дискретно-непрерывных элементов, например из антенной решетки или совокупности одного или более элементов, расположенных в фокусе параболического рефлектора. Антенная решетка, в свою очередь, может быть одномерной (линейной), двумерной (плоской) или трехмерной (объемной). Таким образом, приемная антенна в целом может содержать дискретные, непрерывные и дискретно-непрерывные подсистемы и иметь одномерный, двумерный или трехмерный характер.

Оптимальная (по помехоустойчивости) пространственно-временная обработка информации, извлекаемой из пространства такой приемной антенной, в общем случае не может быть разделена на две ступени — пространственную и временную. Однако в ряде важных случаев такое разделение оказывается возможным, по меньшей мере приближенно (т. е. с небольшим отступлением от оптимальности). Задача математического синтеза оптимальной пространственно-временной обработки (даже однопунктной) весьма сложна по ряду причин. Во-первых, из-за разнообразия и сложности действующих на антенну помех и неопределенности характеристик многих из них. Во-вторых, из-за разнообразия ограничений на структуру и параметры синтезируемой обработки. Дополнительные усложнения возникают в случае пространственной широкополосности принимаемых сигналов, т. е. при выполнении неравенства

$$\Delta f_c |t_1 - t_2| > 1. \quad (3.19)$$

Здесь Δf_c — ширина основной части спектра сигнала; $|t_1 - t_2|$ — разность времен распространения сигнала от «точки» его излучения до крайних точек раскрытия антенны.

Несмотря на указанные трудности, теория математического синтеза оптимальной приемной пространственно-временной обработки в последние годы успешно развивается и при некоторых разумных ограничениях и приближениях позволяет получить полезные для практики рекомендации. В наиболее доступном для инженеров и студентов виде эта теория изложена в [20, 40, 41] и изучается наиболее подробно в курсе радиолокации. Поэтому здесь мы остановимся весьма кратко лишь на основных положениях и некоторых выводах этой теории.

В общетеоретическом плане теория оптимальной (по помехоустойчивости) пространственно-временной обработки является логическим распространением теории оптимальной (по помехоустойчивости) временной обработки, основанной на статистических решениях, из временной области $t \in [T]$ на пространственно-временную $[t \in T, \mathbf{r} \in [V]]$. Здесь t — время, а $\mathbf{r}$ — радиус-вектор данной точки раскрыва антенны относительно его центра; $[T]$ — интервал времени, отводимый на прием сообщений, а $[V]$ — объем (плоскость, линейный размер) раскрыва антенны. При этом смесь сигнала и помех, принятая совокупностью элементов антенны, можно представить в виде

$$y(t, \mathbf{r}) = u_c(t; \mathbf{r}, \mathbf{x}, \mathbf{a}) + n(t, \mathbf{r}), \quad (3.20)$$

где $\mathbf{x}$ — вектор полезных (подлежащих воспроизведению) параметров сигнала, а $\mathbf{a}$ — вектор неинформационных параметров, не известных в месте приема; $\mathbf{n}$ — вектор напряжений, создаваемых аддитивными помехами. Отличие от задачи временной обработки состоит лишь в том, что векторы $\mathbf{y}$, $\mathbf{u}_c$ и $\mathbf{n}$ являются теперь функциями не только времени t , но и пространственных координат $\mathbf{r}$. Поэтому все основные понятия и методы теории статистических решений (априорные и апостериорные распределения, функции правдоподобия, средний риск, методы максимальной апостериорной вероятности, максимального правдоподобия, минимального среднего риска и др.) в принципе распространяются на пространственно-временную обработку, если $\mathbf{y}(\cdot)$, $\mathbf{u}_c(\cdot)$ и $\mathbf{n}(\cdot)$ рассматривать как функции не только времени t , но и пространственного вектора $\mathbf{r}$. Однако увеличение размерности аргумента этих функций, вызываемое переходом от t к $\langle t, \mathbf{r} \rangle$, в общем случае резко усложняет решение получаемых сравнительно просто исходных векторно-матричных уравнений. Получение такого решения может оказаться весьма сложным даже при применении самых мощных современных ЭВМ. Поэтому наиболее полезные для практики результаты получены лишь при ряде упрощающих предположений, которые обычно сводятся к следующему:

помехи $\mathbf{n}(t, \mathbf{r})$ считаются стационарными гауссовскими или марковскими процессами;

сигналы $\mathbf{u}_c(\cdot)$ — пространственно-узкополосные, т. е. в отличие от (3.19) выполняется условие $\Delta f_c |t_1 - t_2| \ll 1$. Это означает, что запаздыванием комплексной огибающей сигнала $U_c(t) \exp[j\psi_c(t)]$ в одних точках раскрыва антенны относительно других его точек можно пренебречь;

антенная система является одномерной решеткой ($\mathbf{r} = z$ — скалярная величина).

Особенно простыми и наглядными результаты синтеза пространственно-временной обработки получаются в следующих случаях.

1. Требуется обнаружить наличие пространственно-узкополосного сигнала в удаленной неподвижной относительно антенны точке пространства с известными координатами. Единственной поме-

хой является внутренний шум приемного устройства. В этом случае пространственно-временная обработка распадается на пространственную (линейную) и временную. Задачей оптимальной пространственной обработки является максимальное усиление сигнала. Оптимальная временная обработка синтезируется независимо от пространственной, т. е. в предположении заданной антенны.

Такой же результат получается, если к внутреннему шуму добавляется внешний тепловой шум, равномерно распределенный в окружающем антенну пространстве, а антенна состоит из идентичных приемных элементов.

2. Отличается от предыдущего лишь тем, что внешняя помеха содержит конечное число статистически независимых внешних источников, действующих из известных направлений пространства, не совпадающих с направлением на источник полезного сигнала. Каждый источник помехи создает гауссовский шум, настолько широкополосный, что в полосе спектра сигнала его можно считать белым, т. е. некоррелированным. Однако напряжение, наведенное таким источником в антенне, пространственно (т. е. по элементам антенны) коррелировано. В этом случае оптимальная пространственно-временная обработка также распадается на оптимальные пространственную (линейную) и временную. Оптимальная пространственная обработка при этом осуществляет такое весовое суммирование (комплексное) выходных напряжений элементов антенны, которое обеспечивает оптимальный компромисс между накоплением (пространственным) составляющих, порожденных полезным сигналом, и взаимной компенсацией составляющих, порожденных внешними помехами, т. е. компромисс между усилением антенны в направлении на полезный сигнал и ослаблением («провалами») в направлениях на источники помех. При этом чем больше интенсивность внешних помех по сравнению с интенсивностью внутренних шумов приемной системы, тем более глубокими оказываются оптимальные провалы в направлении на составляющие источники помех. Оптимальная временная обработка при этом синтезируется независимо от пространственной, т. е. в предположении, что антенная система полностью задана.

3. Требуется определить (измерить, оценить) координаты (в одной плоскости) неподвижной (относительно антенны) точечной цели, создающей в антенне пространственно-узкополосный сигнал. Помеха состоит из внутреннего шума приемного устройства и внешних равномерно распределенных тепловых шумов. Приемной антенной является линейная решетка, содержащая идентичные элементы. В этом случае оптимальная пространственно-временная обработка также распадается на пространственную (линейную) и последующую временную.

Из приведенных примеров видно, что в ряде простых, но достаточно важных случаев оптимальная пространственно-временная обработка распадается на две процедуры: оптимальную линейную пространственную и последующую оптимальную временную. При этом пространственная обработка обеспечивает оптимальное взве-

шенное суммирование комплексных выходных напряжений элементов антенны, а оптимальная временная обработка может синтезироваться в предположении, что алгоритм пространственной обработки задан.

3.4. СИНТЕЗ ПОМЕХОУСТОЙЧИВОЙ ВРЕМЕННОЙ ОБРАБОТКИ В ПРИЕМНИКЕ

При заданных характеристиках антенной системы исходными для синтеза являются статистические характеристики смеси $y(t)$ сигнала и помех на выходе антенной системы. В общем случае антенная система может иметь не один, а m выходов и смесь сигнала и помех имеет вид вектора

$$y(t) = \langle y_1(t), \dots, y_l(t), \dots, y_m(t) \rangle, \quad (3.21)$$

где $y_l(t)$ — смесь сигнала и помех на l -м выходе антенной системы. Если антенная система имеет лишь один выход, смесь $y(t)$ может быть представлена в виде

$$y(t) = u_c(t, x, \alpha_1, \dots, \alpha_r) + u_{ш}(t), \quad (3.22)$$

где x — воспроизводимое сообщение; $\alpha_1(t), \dots, \alpha_r(t)$ — неизвестные в месте приема неинформационные (мешающие) параметры сигнала; $u_{ш}(t)$ — аддитивная помеха.

Сообщение $x(t)$ в общем случае является вектором вида $x(t) = \langle x_1(t), \dots, x_j(t), \dots, x_N(t) \rangle$. Например, в многоканальной СПИ $x_j(t)$ — сообщение, передаваемое по j -му каналу системы. В радиолокации сообщениями $x_1(t), x_2(t), \dots$ могут являться дальность до цели $r(t)$, ее угол места $\alpha(t)$, азимут $\beta(t)$ и производные от этих координат ($\dot{r}, \dot{\alpha}, \dot{\beta}$).

Каждая компонента $x_j(t)$ вектора $x(t)$ может быть случайной дискретной величиной, случайной непрерывной величиной или случайным процессом. Первый вид она имеет в задачах обнаружения и распознавания сигналов, второй — при оценке постоянных за время наблюдения параметров сигнала, а третий — при фильтрации.

Наибольшие математические трудности при синтезе оптимальной по помехоустойчивости обработки (оптимального приемника) возникают в задачах фильтрации. В общем случае алгоритм фильтрации оказывается нелинейным и задача синтеза сводится к отысканию алгоритма оптимальной нелинейной фильтрации и соответствующей этому алгоритму точности воспроизведения сообщений (потенциальной точности, потенциальной помехоустойчивости). В настоящее время существуют два основных математических метода решения этой задачи. Первый основан на аппроксимации сообщения $x(t)$ гауссовским процессом, а второй — компонентами многомерного марковского процесса¹. В обоих случаях

¹ В некоторых задачах сообщение $x_j(t)$ целесообразно аппроксимировать детерминированной функцией времени с конечным (обычно небольшим) числом n неизвестных параметров $a_{j1}, \dots, a_{jn}$, например в виде полинома $x_j(t) = a_{j1} +$

задачу удается решить до конца обычно лишь при дополнительном допущении о том, что значение среднего квадрата $M[\delta^2(t)]$ ошибки воспроизведения сообщений достаточно мало, т. е. система и в условиях помех работает в режиме высокой точности воспроизведения сообщений.

Гауссовской и особенно марковской теории оптимальной нелинейной фильтрации посвящены десятки книг, основными из которых явились [23, 25, 29, 30] и др. В наиболее доступной и краткой форме основы этих теорий изложены в [5, 24]. Полная теория оптимальной нелинейной фильтрации весьма многообразна и сложна. Поэтому ограничимся рассмотрением лишь следующего частного, но достаточно важного случая [5].

Смесь сигнала и помех имеет вид вектора (3.21), причем

$$y_i(t) = u_{ci}(t, x) + u_{шi}(t), \quad i = \overline{1, m}, \quad (3.23)$$

где $u_{ci}(t, x)$ — сигнал на i -м выходе антенной системы, известный в месте приема точно, за исключением воспроизводимого сообщения x , одинакового для всех m выходов антенны; $u_{шi}(t), i = \overline{1, m}$, — независимые нормальные белые шумы со спектральными плотностями $N_{01}, \dots, N_{0i}, \dots, N_{0m}$.

Например, для двухканального фазового пеленгатора, измеряющего малые угловые отклонения θ от оси его равносигнальной зоны,

$$x = \theta, \quad u_{c1}(t, x) = U_1(t) \cos [\omega t + h\theta(t) + \varphi(t)],$$

$$u_{c2}(t, x) = U_2(t) \cos [\omega t - h\theta(t) + \varphi(t)],$$

где $h = 2\pi d/\lambda$; d — расстояние между фазовыми центрами антенн A_1 и A_2 ; λ — длина волны принимаемого сигнала; $U_1(t), U_2(t)$ и $\varphi(t)$ — точно известные в месте приема функции.

Сообщение $x(t)$ является нормальным случайным процессом (гауссовская аппроксимация) с известными математическим ожиданием $m_x(t)$ и корреляционной функцией $R_x(t_1, t_2)$ (определяемой для центрированных сообщений). Точность воспроизведения сообщения оценивается средним квадратом $M[\delta^2(t)]$ ошибки его воспроизведения и предполагается высокой. Модуляция сигнала $u_c(t, x)$ сообщением $x(t)$ — прямая, т. е. сообщение $x(t)$ входит в выражение для сигнала не под знаком оператора $K(p)$, а непосредственно (в частности, это имеет место при амплитудной или фазовой модуляции сигнала).

При сделанных допущениях структура оптимального приемника имеет вид, изображенный на рис. 3.2, т. е. состоит из оптимального дискриминатора ДИС и оптимального линейного фильтра Φ , охваченных обратной связью. На выход фильтра включается

$+ a_{j2}t^2 + a_{j3}t^3 + \dots + a_{jn}t^{n-1}$. При этом задача фильтрации каждого сообщения $x_j(t)$ сводится к совместной оценке n неизвестных параметров, постоянных за время T воспроизведения сообщений. Если за время T сообщение $x_j(t)$ изменяется незначительно, то число оцениваемых параметров можно ограничить двумя (линейная аппроксимация) или тремя (квадратичная аппроксимация) и алгоритм оптимальной обработки существенно упрощается.

априорно известное математическое ожидание $m_x(t)$ сообщения $x(t)$. Так как все сигналы $u_{ci}(t, x)$, $i = \overline{1, m}$, модулированы одним и тем же сообщением $x(t)$, дискриминатор, m -канальный по входу, оказывается по своему выходу одноканальным. Структура оптимального дискриминатора, т. е. закон преобразования входной реализации $y(t)$ в выходное напряжение $z(t)$ дискриминатора, определяется соотношением

$$z(t) = - \left[\frac{\partial Q(y, x, t)}{\partial x} \right]_{x=x^*} \quad (3.24)$$

где

$$Q(y, x, t) = \sum_{i=1}^m \frac{1}{N_{oi}} [y_i(t) - u_{oi}(t, x)]^2. \quad (3.25)$$

При этом напряжение на выходе дискриминатора

$$z(t) = k [e(t) + \Delta \varepsilon_n(t)], \quad (3.26)$$

где

$$e(t) = x(t) - x^*(t) \quad (3.27)$$

— ошибка воспроизведения сообщения; $\Delta \varepsilon_n(t)$ — нормальный белый шум с односторонней спектральной плотностью

$$g_{\text{ш0}} = 1 / \sum_{i=1}^m \frac{1}{N_{oi}} \left[\frac{\partial u_{oi}(t, x)}{\partial x} \right]^2, \quad (3.28)$$

волнистая линия обозначает усреднение по времени за достаточно большое время T (строго говоря, $T \rightarrow \infty$).

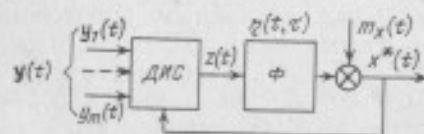


Рис. 3.2

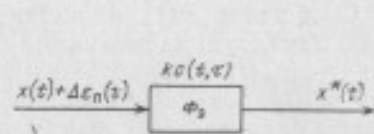


Рис. 3.3

Крутизна характеристики дискриминатора k связана со спектральной плотностью $g_{\text{ш0}}$ простым соотношением

$$k = 2/g_{\text{ш0}}. \quad (3.29)$$

Структура оптимального линейного фильтра Φ определяется его импульсной реакцией $\eta(t, \tau)$, которая в общем случае находится решением интегрального уравнения

$$k \int_{\tau}^t c(t, s) \eta(s, \tau) ds + c(t, \tau) = \eta(t, \tau). \quad (3.30)$$

При этом вспомогательная функция $c(t, \tau)$ является, в свою очередь, решением интегрального уравнения

$$k \int_{t_0}^t c(t, s) R_x(s, \tau) ds + c(t, \tau) = R_x(t, \tau), \quad (3.31)$$

где t_0 — момент начала воспроизведения сообщения.

Можно показать (см., например, [5]), что $kc(t, \tau)$ — импульсная реакция эквивалентного линейного фильтра Φ_* , осуществляющего оптимальную (линейную) фильтрацию сообщения $x(t)$ на фоне аддитивного независимого белого шума $\Delta \varepsilon_n(t)$ (рис. 3.3). Поэтому, если решение этой задачи линейной фильтрации для данного вида сообщения $x(t)$ известно, решать интегральное уравнение (3.31) не требуется.

Если фильтр Φ оказывается стационарным [т. е. реакция $c(t, \tau)$ зависит лишь от разности $(t - \tau)$], а спектр сообщения $x(t)$ может быть описан дробно-рациональной функцией, то не требуется решать и интегральное уравнение (3.30), так как передаточная функция $K_1(j\omega)$ фильтра Φ может быть в этом случае определена по формуле [22]

$$K_1(j\omega) = \frac{K_*(j\omega)}{k[1 - K_*(j\omega)]}, \quad (3.32)$$

где $K_*(j\omega)$ — передаточная функция фильтра Φ_* , соответствующая его импульсной реакции $kc(t, \tau)$.

Например, если сообщение $x(t)$ имеет энергетический спектр

$$S_x(\omega) = \frac{2\sigma_0^2 T}{1 + (\omega T)^2}, \quad (3.33)$$

то, как показано в [22],

$$K_1(j\omega) = \frac{g_{\text{ш0}} \rho}{2(\sqrt{1 + \rho} + 1)} \frac{1}{(1 + j\omega T)}, \quad (3.34)$$

где $\rho = 4\sigma_0^2 T / g_{\text{ш0}}$, т. е. оптимальный линейный фильтр Φ состоит из безынерционного усилителя с коэффициентом передачи $g_{\text{ш0}} \rho / (2(\sqrt{1 + \rho} + 1))$ и инерционного звена с постоянной времени T .

Оптимальный приемник обеспечивает получение минимального значения среднего квадрата ошибки воспроизведения сообщения, которое оказывается равным

$$M[e^2] = c(t, t), \quad (3.35)$$

где $c(t, t)$ — значение функции $c(t, \tau)$ при $\tau = t$. При этом математическое ожидание ошибки равно нулю, так что дисперсия ошибки D_e совпадает с ее средним квадратом. Если спектр сообщения имеет вид (3.33), то [22]

$$D_e = M[e^2] = \frac{\sigma_0^2}{\sqrt{1 + \rho}}. \quad (3.36)$$

Из схемы на рис. 3.2 следует, что в оптимальном приемнике дискриминатор ДИС является оптимальным в том смысле, что обеспечивает воспроизведение «сигнала ошибки» $e(t)$ с минималь-

ным значением спектральной плотности $g_{ш}$ шума (ошибки) $\Delta \varepsilon_n(t)$, соответствующего этому воспроизведению. Фильтр же Φ обеспечивает наилучшее подавление помехи $k\Delta \varepsilon_n(t)$, прошедшей на выход дискриминатора, и тем самым — получение наибольшей точности (наименьшего среднего квадрата ошибки) воспроизведения сообщения $x(t)$.

Из изложенного следует, что при сделанных выше предположениях о характере смеси сигнала и помех оптимальное приемное устройство имеет вид следящей [за сообщением $x(t)$] системы, состоящей из дискриминатора и линейного фильтра. В более общих случаях (например, при негауссовском характере сообщения и не очень высокой точности воспроизведения сообщения) четкое разбиение системы на дискриминатор и фильтр отсутствует или хотя и существует, но оптимальный фильтр оказывается нелинейным. Алгоритм оптимальной нелинейной фильтрации во многих практических задачах оказывается настолько сложным, что требует его реализации на ЭВМ, обладающих большим быстродействием и памятью. Поэтому часто приходится заменять оптимальные алгоритмы квазиоптимальными (в смысле помехоустойчивости). При этом ориентиром является проигрыш в точности предлагаемого квазиоптимального алгоритма по сравнению с потенциальной точностью [потенциальной помехоустойчивостью, соответствующей оптимальному (по помехоустойчивости) алгоритму].

3.5. ОСОБЕННОСТИ СИНТЕЗА СЛЕДЯЩИХ УСТРОЙСТВ

В § 3.4 рассматривалась задача синтеза оптимального (по помехоустойчивости) устройства воспроизведения сообщения, являющегося функцией времени. При этом оказалось, что оно должно быть следящим устройством, состоящим из дискриминатора и линейного фильтра, охваченных обратной связью (рис. 3.2). В ряде практических задач заранее задается, что синтезируемое радиоприемное устройство должно быть следящим, например осуществлять автоматическое слежение за принимаемым сигналом по частоте, фазе или по направлению прихода (по угловым координатам источника сигнала).

Рассмотрим, например, задачу синтеза оптимального (по помехоустойчивости) следящего угломерного устройства. Оно состоит из радиопеленгатора, содержащего антенную систему A и приемник, и блока управления угловыми координатами оси z_A равносигнальной зоны пеленгатора (рис. 3.4). Для простоты будем рас-

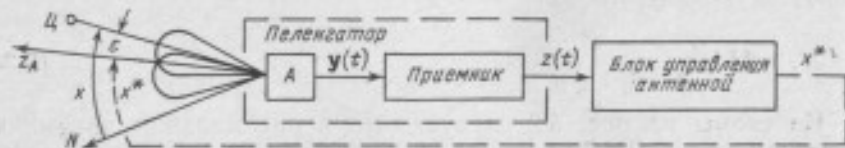


Рис. 3.4

смагивать пеленгацию в одной плоскости. Пусть N есть фиксированное направление в этой плоскости, от которого отсчитывается угловая координата x направления на цепь Π (пеленгуемый объект) и угловая координата x^* оси z_A равносигнальной зоны. Тогда ошибка слежения $\varepsilon(t) = x(t) - x^*(t)$.

В общем случае, когда антенная система имеет m выходов, смесь $y(t)$ сигнала и помех на ее выходе имеет вид вектора (3.21). Если принять те же допущения относительно характера сигналов и помех, что и в § 3.4 (сигнал известен точно, за исключением несомого им сообщения, а единственной помехой являются аддитивные независимые шумы на выходе каждой из антенн), то

$$y_i(t) = u_{ci}(t, \varepsilon) + u_{шi}(t), \quad i = \overline{1, m}. \quad (3.37)$$

Это выражение отличается от (3.23) лишь тем, что в данном случае сигнал на выходе антенны модулирован не непосредственно воспроизводимым сообщением $x(t)$, а сигналом ошибки $\varepsilon(t)$. Задача следящей системы состоит в том, чтобы обеспечить малость ошибки $\varepsilon(t)$ слежения за сигналом (за угловой координатой x). Если предположить, как и в § 3.4, что средний квадрат ошибки слежения достаточно мал, то напряжение $z(t)$ на выходе пеленгатора можно представить в виде

$$z(t) = k[\varepsilon(t) + \Delta \varepsilon_n(t)], \quad (3.38)$$

где k — крутизна дискриминационной характеристики пеленгатора, а $k\Delta \varepsilon_n(t)$ — помеха на выходе пеленгатора.

Если величина z достаточно мала, то блок управления можно считать линейным по отношению ко входному воздействию $z(t)$ и, следовательно, описывать некоторой импульсной реакцией $\eta(t, \tau)$. При этом структурная схема следящей системы совпадает со структурной схемой оптимальной приемной системы на рис. 3.2, если под дискриминатором понимать приемник пеленгатора, а под линейным фильтром Φ — блок управления антенной. Необходимо сделать лишь следующие замечания.

Во-первых, в схеме рис. 3.2 на выход фильтра Φ подается внешнее воздействие, равное математическому ожиданию $m_x(t)$ сообщения $x(t)$ (которое полагается априори известными). Поэтому в случае следящей угломерной системы для того, чтобы эта система была оптимальной, необходимо также вводить на выход блока управления антенной внешнее (программное) воздействие, равное математическому ожиданию угловой координаты цели $x(t)$ (если это математическое ожидание не равно нулю).

Во-вторых, как уже отмечалось, в схеме на рис. 3.2 смесь $y_i(t)$ определяется соотношением (3.23), а в случае следящего устройства — выражением (3.37), т. е. в следящем устройстве сигнал u_{ci} модулирован не самим сообщением $x(t)$, а сигналом ошибки $\varepsilon(t) = x(t) - x^*(t)$. Как показывает анализ [23], это отличие приводит лишь к тому, что при синтезе оптимального следящего устройства в формулах (3.24), (3.25) и (3.28) следует заменить x на ε и соответственно полагать

$$z(t) = - \left[\frac{\partial Q(y, \varepsilon, t)}{\partial \varepsilon} \right]_{\varepsilon=0}, \quad (3.39)$$

где

$$Q(y, \varepsilon, t) = \sum_{i=1}^m \frac{1}{N_{oi}} [y_i(t) - u_{oi}(t, \varepsilon)]^2, \quad (3.40)$$

$$g_{\omega 0} = 1 / \sum_{i=1}^m \frac{1}{N_{oi}} \left[\frac{\partial u_{oi}(t, \varepsilon)}{\partial \varepsilon} \right]^2. \quad (3.41)$$

В-третьих, при выводе формул, определяющих структуру оптимального приемника (т. е. при решении задачи оптимальной нелинейной фильтрации), на структуру и параметры синтезируемого (отыскиваемого) устройства не накладывалось никаких ограничений (за исключением естественного требования физической реализуемости). Поэтому формулы (3.39)–(3.41) для синтеза оптимального пеленгатора и соотношения (3.30)–(3.36) для оптимального блока управления остаются справедливыми, если при синтезе на структуру следящего устройства и значения его параметров не накладывается никаких ограничений (кроме условия физической реализуемости).

В действительности при синтезе следящего устройства заранее вводится ограничение, состоящее в том, что структура этого устройства должна иметь вид, изображенный на рис. 3.4. Однако, как было показано ранее, эта структура полностью совпадает со структурой оптимальной системы (рис. 3.2). Поэтому данное ограничение не приводит к потере оптимальности (по меньшей мере в тех задачах, где оптимальный приемник распадается на оптимальный дискриминатор и линейный фильтр, охваченные обратной связью). Поэтому, если нет никаких дополнительных ограничений, структура оптимального следящего устройства и даваемое им (потенциальное) значение среднего квадрата ошибки $M[\varepsilon^2(t)]$ могут определяться по приведенным в § 3.4 формулам (с указанной заменой в них x на ε), если сигнал известен точно, а единственной помехой являются аддитивные независимые нормальные белые шумы $u_{oi}(t)$.

Но в некоторых задачах синтеза следящих устройств могут иметь место дополнительные ограничения, например блок управления может быть полностью задан и в таких условиях требуется синтезировать лишь оптимальный пеленгатор (дискриминатор). Если при этом остается справедливым допущение о малости среднего квадрата $M[\varepsilon^2]$ ошибки и линейности блока управления, то структура и параметры (крутизна характеристики) оптимального пеленгатора (дискриминатора) могут быть найдены, как и прежде, по формулам (3.39)–(3.41) и останется справедливым соотношение (3.29). Действительно, дискриминатор, описываемый этими формулами, как отмечалось в § 3.4, оптимален в том смысле, что обеспечивает (по сравнению с другими возможными дискриминаторами) воспроизведение сигнала ошибки $\varepsilon(t)$ с минимальным

значением $g_{\omega 0}$ спектральной плотности помехи (ошибки) на своем выходе. Но при линейном блоке управления (линейном фильтре Φ) спектральная плотность помехи на выходе пеленгатора (дискриминатора) полностью характеризует качество пеленгатора — чем меньше значение этой спектральной плотности, тем меньше средний квадрат $M[\varepsilon^2]$. Поэтому, если блок управления (фильтр Φ) задан в классе линейных устройств и неоптимален, то пеленгатор останется оптимальным и спектральная плотность помехи на его выходе может определяться по формуле (3.41). При этом средний квадрат $M[\varepsilon^2]$ ошибки слежения увеличится по сравнению с ее потенциальным значением (3.35) лишь за счет неоптимальности блока управления (фильтра Φ).

Если характеристики смеси $y(t)$ сигнала и помех изменяются по сравнению с рассмотренными или приходится учитывать дополнительные ограничения структуры и параметров синтезируемого устройства или системы, отыскание оптимальной (по помехоустойчивости) системы (устройства) может весьма существенно усложниться. Наибольшие трудности возникают при синтезе систем радиоуправления.

3.6. ОСОБЕННОСТИ СИНТЕЗА СИСТЕМ РАДИОУПРАВЛЕНИЯ

Системы радиоуправления отличаются большим разнообразием и сложностью. Они могут быть полностью автоматическими или автоматизированными, содержать один или несколько управляемых объектов, единицы или большое число радиолиний и т. д. В проектировании каждой системы радиоуправления могут участвовать десятки, сотни или тысячи инженеров различных специальностей. Обеспечение помехоустойчивости таких систем, особенно при наличии организованных помех, является весьма сложной задачей.

Для иллюстрации главных особенностей, возникающих при синтезе помехоустойчивой системы радиоуправления, рассмотрим типичную функциональную схему такой системы, изображенную на рис. 3.5. Она содержит управляемый объект УО, управляющее устройство УУ и измерительное устройство ИУ. Здесь $x_2(t)$ — вектор параметров, характеризующих состояние управляемого объекта и называемых координатами его состояния или фазовыми координатами¹. Если управляемым объектом является летательный аппарат (ЛА), то его фазовые координаты — параметры движения центра масс аппарата (координаты, скорости и ускорения) и поворотов корпу-

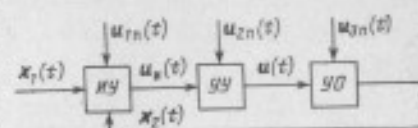


Рис. 3.5

¹ Здесь термин «фаза» понимается в смысле «состояние».

са ЛА вокруг его центра масс (углы поворота, угловые скорости и ускорения)¹.

Характер изменения во времени вектора $x_2(t)$ фазовых координат объекта будем называть для краткости *характером фазовой траектории* объекта. Задача управления состоит в обеспечении наилучшего (в смысле некоторого заранее выбранного критерия оптимальности) вида фазовой траектории управляемого объекта. В случае программного управления требуемый вид $x_{tr}(t)$ фазовой траектории заранее известен и задается программным механизмом. Однако в общем случае управление должно осуществляться не по заранее установленной программе, а в соответствии с состоянием некоторой совокупности $x_1(t)$ параметров внешней среды (внешнего объекта), которые будем называть фазовыми координатами внешнего объекта, а их изменение в процессе управления — фазовой траекторией внешнего объекта.

Например, при наведении одного ЛА (самолета, ракеты, космического аппарата) на другой ЛА (для стыковки с ним или его уничтожения) требуемая фазовая траектория управляемого ЛА зависит от вида фазовой траектории того ЛА (цели), на который он наводится. В этом случае $x_1(t)$ есть не что иное, как вектор фазовых координат цели, так как в качестве внешнего объекта выступает цель.

Итак, в общем случае задача управления состоит в том, чтобы обеспечить наилучшее (в заранее заданном смысле) соответствие фазовой траектории $x_2(t)$ управляемого объекта той фазовой траектории $x_1(t)$, которую имеет внешний объект, играющий роль задающего объекта. При этом все или часть фазовых координат, входящих в состав векторов $x_1(t)$ и $x_2(t)$, заранее неизвестны и, следовательно, подлежат измерению в процессе управления. Их измерение осуществляется измерительным устройством ИУ, выдающим результаты измерения в виде некоторого вектора $u_n(t)$, содержащего информацию об измеряемых процессах $x_1(t)$ и $x_2(t)$. В некоторых задачах управления не требуется отдельно измерять векторы $x_1(t)$ и $x_2(t)$ — достаточно измерять лишь некоторый вектор $e(t)$, характеризующий степень рассогласования между процессами $x_2(t)$ и $x_1(t)$ и называемый *вектором рассогласования*. Например, если система управления должна быть следящей, т. е. обеспечивать выполнение равенства $x_2(t) = x_1(t)$, то вектор рассогласования имеет вид $e(t) = x_2(t) - x_1(t)$, т. е. требуется измерять лишь отклонения фазовых координат управляемого объекта от соответствующих фазовых координат внешнего (задающего) объекта.

¹ Различные фазовые координаты могут иметь несовпадающие размерности (например, углы и угловые скорости). Поэтому при образовании вектора $x_2(t)$ необходимо произвести соответствующую нормировку его составляющих (приведение к безразмерному виду или к единой размерности), либо рассматривать вектор $x_2(t)$ лишь как упорядоченную совокупность входящих в него составляющих. В этом же смысле следует понимать и все остальные приводимые в данном параграфе векторы $[x_1(t), e(t), h(t)$ и др.] Эта особенность рассматриваемых векторов обсуждается далее на примере вектора ошибки h .

На основе полученной информации управляющее устройство УУ вырабатывает команды управления $u(t)$, вызывающие соответствующее изменение фазовых координат $x_2(t)$ управляемого объекта. Действие различных источников помех проявляется в мешающих воздействиях $u_{1n}(t)$, $u_{2n}(t)$ и $u_{3n}(t)$, приложенных соответственно к блокам ИУ, УУ и УО системы.

Радиосредства, входящие в состав системы управления, выполняют функции всех или части измерительных устройств (т. е. устройств извлечения информации), а также, при необходимости, используются для передачи информации. Например, при управлении космическим аппаратом по командам с Земли одна часть измерительных и управляющих устройств расположена на Земле, а другая — на управляемом космическом аппарате. При этом требуется передавать по радио измерительную информацию (по радиолиниям «Земля—Борт» и «Борт—Земля») и команды управления (по радиолинии «Земля—Борт»). Эти радиолинии передачи информации на рис. 3.5 для простоты не выделены, и их можно считать включенными в состав блоков ИУ и УУ. Радиосредства включаются в систему управления для повышения точности управления, которая характеризуется отклонением вектора $x_2(t)$ фазовых координат управляемого объекта от требуемого значения $x_{tr}(t)$. В каждый данный момент времени это отклонение

$$h(t) = x_2(t) - x_{tr}(t) \quad (3.42)$$

и называется *вектором ошибки управления*.

В ряде задач управления необходимо обеспечить малое значение этой ошибки не в произвольный момент времени t , а лишь в некоторый конечный момент t_k , т. е. важно обеспечить малость лишь конечной ошибки

$$h(t_k) = x_2(t_k) - x_{tr}(t_k). \quad (3.43)$$

Такое управление называется *терминальным*. (Например, в задачах наведения одного ЛА на другой часто важно обеспечить малость лишь конечной ошибки, имеющей место в тот момент, когда управляемый аппарат находится на минимальном или достаточно близком к минимальному расстоянию от цели). При следящем управлении $x_{tr}(t) = x_1(t)$ и соотношения (3.42) и (3.43) принимают соответственно следующий вид:

$$h(t) = x_2(t) - x_1(t), \quad (3.42')$$

$$h(t_k) = x_2(t_k) - x_1(t_k). \quad (3.43')$$

На первый взгляд кажется, что приемлемой мерой точности управления в случае (3.42) является средний квадрат $M[h^2(t)]$ модуля текущей ошибки, а при терминальном управлении — средний квадрат $M[h^2(t_k)]$ модуля конечной ошибки. Однако в общем случае характеризовать вектор ошибки h его модулем совершенно недостаточно или даже неприемлемо. Действительно, различные составляющие вектора $h(h_1, h_2, \dots, h_m)$ могут иметь даже различные размерности. Например, при управлении сближением космического

аппарата (КА) с долговременной орбитальной станцией (ДОС) важно обеспечить в конце этапа радиоуправления малость отклонений не только координат, но и скорости КА от соответствующих координат и скорости ДОС; поэтому даже в простейшем случае вектор ошибки имеет вид

$$\mathbf{h} = \langle r, V_r \rangle,$$

где r — расстояние между центрами масс КА и ДОС, V_r — скорость сближения центров масс.

Очевидно, в данном случае модуль

$$h = \sqrt{r^2 + V_r^2}$$

и соответственно его средний квадрат $M[h^2]$ не имеют физического смысла и в качестве меры точности следует брать взвешенный средний квадрат вида

$$h_a^2 = M[r^2 + aV_r^2] = M[r^2] + aM[V_r^2], \quad (3.44)$$

где a — некоторый весовой коэффициент (вес), имеющий размерность $[c^2]$. Значение этого веса следует выбирать тем большим, чем больше опасность возникновения больших значений скорости сближения V_r по сравнению с большими значениями расстояния r . Однако дать четкое количественное обоснование значения веса a не всегда возможно. Еще сложнее обосновать меру точности управления, когда вектор ошибки $\mathbf{h}$ содержит не две координаты различной размерности, а значительно больше.

Поскольку мерой помехоустойчивости системы радиоуправления является точность управления при наличии действия помех, соответствующие трудности возникают и при оценке помехоустойчивости системы управления. Дополнительные трудности при оценке качества систем радиоуправления возникают вследствие необходимого учета ряда дополнительных ограничений. В частности, даже на начальных этапах синтеза помехоустойчивой системы управления весьма желательно учесть ограниченность ресурсов, которые могут быть израсходованы системой в процессе управления (например, ограниченность запасов топлива, которые могут быть израсходованы в процессе управления, или ограниченность допустимого поперечного ускорения управляемого ЛА).

С учетом изложенного при математическом синтезе помехоустойчивых систем управления показатель качества управления в ряде случаев приходится выбирать в следующем виде ([26] и др.):

$$I_k = M \left\{ \mathbf{h}^T(t_k) \Gamma(t_k) \mathbf{h}(t_k) + \int_{t_0}^t \mathbf{h}^T(\tau) L(\tau) \mathbf{h}(\tau) d\tau + \int_{t_0}^{t_k} \mathbf{u}^T(\tau) K^{-1} \mathbf{u}(\tau) d\tau \right\}. \quad (3.45)$$

Здесь $\Gamma(t_k)$, $L(\tau)$ и K^{-1} — матрицы весов (подобных рассмотренному выше весу a), учитывающих относительную опасность больших значений различных составляющих векторов $\mathbf{h}(\tau)$ и $\mathbf{u}(\tau)$ (ес-

ли бы все составляющие имели одинаковую размерность и были одинаково опасными, то матрицы весов были бы единичными); $*$ обозначает транспонированный вектор; t_0 , t_k и t — соответственно начальный, конечный и текущий моменты управления.

Очевидно, первое слагаемое в (3.45) характеризует конечную ошибку, второе слагаемое — интегральное качество переходного процесса в течение всего времени управления, а третье — затраты на управление (поскольку $\mathbf{u} = \langle u_1, \dots, u_r \rangle$ и с ростом значений $u_1^2, \dots, u_r^2$ расходы ресурсов на управление, например расходы топлива, возрастают).

Показатель качества вида (3.45) называют *взвешенным квадратичным показателем* (поскольку при единичных весах каждое из слагаемых имеет квадратичный характер). При терминальном управлении второе слагаемое отсутствует ($L=0$).

Математический синтез помехоустойчивой системы радиоуправления весьма сложен даже при простом виде показателя качества. Это объясняется тем, что в общем случае необходимо одновременно синтезировать четверку $\{\text{ИУ, УУ, ОУ, } u_a(t)\}$ (рис. 3.5), т. е. найти оптимальные (по помехоустойчивости) принципы построения (алгоритмы действия) измерительного и управляющего устройств и самого управляемого объекта при наиболее опасной (или наиболее вероятной) совокупности $u_a(t)$ помех, которая будет иметь место при таких принципах (алгоритмах) построения системы управления.

Возникающие при этом трудности обусловлены в первую очередь тем, что, не зная характера помех $u_a(t)$, нельзя найти помехоустойчивые принципы построения системы; сформулировать же математически зависимость характеристик всех возможных помех от всех возможных принципов построения системы также не представляется возможным. Такого рода трудность возникает и при синтезе систем передачи или извлечения информации, работающих автономно, а не в составе системы управления. Однако при синтезе систем управления задача дополнительно осложняется тем, что кроме синтеза устройств извлечения и передачи информации (входящих в состав измерительного устройства ИУ) приходится одновременно синтезировать еще две сложные части системы — управляющее устройство УУ и управляемый объект УО.

Практически эти трудности преодолеваются использованием метода последовательных приближений. Применительно к системам радиоуправления он может иметь несколько модификаций, например следующую. Четверку $\{\text{ИУ, УУ, ОУ, } u_a(t)\}$ рассматривают как пару $(\text{СУ, } u_a(t))$, где $\text{СУ} = (\text{ИУ, УУ, ОУ})$ — система управления в целом.

На первом этапе на основе имеющихся априорных данных задаются приближенным (и обычно достаточно простым) видом совокупности (вектора) помех $u_a(t) = \langle u_{a1}(t), u_{a2}(t), u_{an}(t) \rangle$. Для этой совокупности помех находят оптимальную по помехоустойчивости структуру СУ_1 системы управления. Для найденной структуры СУ_1 уточняют характеристики вектора помех $u_a(t)$ и находят

второе приближение СУ₂ для структуры системы управления и т. д. При этом на каждом этапе синтез помехоустойчивой системы управления также производят методом последовательных приближений, например в следующем порядке.

Система управления рассматривается как тройка СУ — (УУ, ИУ, УО). На первом этапе производится синтез помехоустойчивой пары (УУ, ИУ)₁ для заданного в качестве первого приближения объекта УО₁. Для найденной пары (УУ, ИУ)₁ производится синтез помехоустойчивого объекта, т. е. находится второе приближение УО₂ для структуры (и параметров) управляемого объекта. Для найденного УО₂ производится заново синтез помехоустойчивой пары (ИУ, УУ)₂ и т. д. Отсюда следует, что в процессе проектирования системы управления приходится решать задачи синтеза как пары (ИУ, УУ) при данном УО, так и синтез УО при заданной паре (ИУ, УУ). Очевидно, возможен и другой порядок проектирования, при котором решаются, например, задачи синтеза ИУ при заданной паре (УУ, УО) и синтеза пары (УУ, УО) при заданном ИУ.

В решении этих частных задач (подзадач) синтеза системы управления принимают участие специалисты различных профилей. Например, радиоспециалисты чаще всего решают задачи синтеза ИУ при заданной паре (УУ, УО) или синтеза пары (ИУ, УУ) при заданном УО. Специалисты по технической кибернетике (управлению) проводят синтез пары (УУ, УО) при данном ИУ или синтез УУ при данных ИУ и УО.

Разбивка задачи совместного синтеза четверки {ИУ, УУ, УО, $u_n(t)$ } на ряд более частных подзадач приводит, рассуждая чисто формально, к потере строгой оптимальности решения. Однако, поскольку строгий оптимальный синтез четверки {ИУ, УУ, УО, $u_n(t)$ } практически невозможен даже при использовании самых последних достижений аналитической и вычислительной математики (ЭВМ), разбивка общей задачи на ряд подзадач и применение метода последовательных приближений оказываются в той или иной степени совершенно необходимыми.

Достаточно строгое в математическом смысле решение задачи синтеза системы радиуправления удастся получить лишь в некоторых простейших случаях. Рассмотрим в качестве иллюстрации несколько таких примеров.

Пример 1. Пусть синтезируемая система управления некоторым объектом должна обеспечить автоматическое равенство фазовой координаты $x_2(t)$ этого объекта фазовой координате $x_1(t)$ внешнего задающего объекта, т. е. обеспечить равенство нулю параметра рассогласования $e(t) = x_2(t) - x_1(t)$. На синтезируемую систему и управляемый объект не накладывается никаких ограничений, кроме условия их физической реализуемости. Измерительное устройство измеряет параметр рассогласования $e(t)$ таким образом, что на его входе образуется смесь $y(t)$ сигнала и помех, описываемая соотношениями (3.21) и (3.37). Точность (помехоустойчивость) управления оценивается средним квадратом $M[e^2(t)]$ параметра рассогласования — требуется при данных характеристиках

помех обеспечить минимальное значение этого среднего квадрата в каждый момент времени t , причем значение $M[e^2(t)]$ должно быть весьма малым. При таких условиях указанная задача сводится к рассмотренной в § 3.5 задаче синтеза оптимальной (по помехоустойчивости) следящей системы и решается на основе теории оптимальной нелинейной фильтрации. При этом роль оптимального измерительного устройства ИУ выполняет оптимальный дискриминатор ДИС (рис. 3.2), а совокупность блоков УУ и УО должна обладать характеристиками, совпадающими с характеристиками оптимального линейного фильтра Φ в схеме на рис. 3.2.

Следовательно, в рассматриваемом случае схема на рис. 3.5 системы управления преобразуется к виду, показанному на рис. 3.6. Входная часть ИУ в данном случае полагается заданной, а оптимальная структура остальной части системы управления определяется соотношениями (3.30), (3.31), (3.39) — (3.41). Потенциальное значение помехоустойчивости (точности) управления, т. е. минимальное значение $M[e^2(t)]$, определяется при этом соотношениями (3.29), (3.31), (3.35), (3.41).

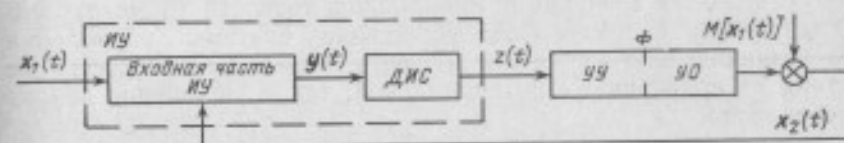


Рис. 3.6

Пример 2. Пусть задача отличается от рассмотренной в примере 1 лишь тем, что требуется обеспечить минимальное значение $M[e^2(t)]$ не в произвольный момент t , а только в некоторый конечный момент t_n , т. е. задача сводится к терминальному управлению. Очевидно, что рассмотренная в предыдущем примере система, обеспечивающая минимум $M[e^2(t)]$ в каждый момент времени t , будет обеспечивать минимум и величины $M[e^2(t_n)]$, что требуется при терминальном управлении. Поэтому потенциальная точность системы терминального управления будет в данном случае такой же, как в примере 1. Однако структура оптимальной системы в этом случае может быть и более простой (поскольку требуется обеспечить минимум СКО не в произвольный, а лишь в конечный момент времени).

Пример 3. Пусть задача отличается от рассмотренной в примере 1 лишь тем, что объект управления задан, т. е. его характеристики не могут варьироваться в процессе синтеза. При этом известно, что поведение объекта описывается линейными дифференциальными уравнениями с постоянными коэффициентами, т. е. объект может рассматриваться как некоторый линейный стационарный фильтр. В этом случае, как было показано в § 3.5, оптимальный дискриминатор (а следовательно, и оптимальное измерительное устройство ИУ в целом) остается таким же, как в примерах 1 и 2. Однако вследствие неоптимальности линейного фильтра система управления в целом оказывается неоптимальной (по помехоустойчивости). Для того чтобы найти структуру оптимальной системы управления, в этом случае требуется произвести синтез оптимального (по помехоустойчивости) управляющего устройства УУ. Современная статистическая теория

оптимального управления¹ позволяет получить строгое математическое решение такой задачи (см., например, [1, 26, 28]). Однако для реализации полученных алгоритмов требуется проведение большого объема вычислений в реальном масштабе времени. Учитывая это обстоятельство, а также то, что при математической формулировке решаемой задачи приняты весьма существенные упрощения, часто применяют вместо найденных при этом оптимальных решений более простые, так называемые квазиоптимальные решения.

Как уже отмечалось, по совокупности показателей качества упрощенные алгоритмы обработки информации часто могут оказаться значительно лучше более сложных алгоритмов, оптимальных только по одному показателю (помехоустойчивости), к тому же найденных при грубых идеализациях характера помех и управляемого объекта (например, при учете помех только типа внутреннего шума приемника и в предположении линейности управляемого объекта). Ориентиром при замене оптимальной (по помехоустойчивости) структуры системы управления на квазиоптимальную является значение потенциальной точности (помехоустойчивости) — если существенное упрощение системы управления лишь незначительно ухудшает точность (помехоустойчивость) ее действия по сравнению с потенциальной, такое упрощение вполне целесообразно.

3.7. ОСОБЕННОСТИ СИНТЕЗА ПОМЕХОУСТОЙЧИВЫХ СИСТЕМ И УСТРОЙСТВ ПРИ НЕПОЛНЫХ АПРИОРНЫХ ДАННЫХ

Основным математическим аппаратом, используемым при синтезе помехоустойчивых радиосистем, является теория статистических решений. Априорная информация, необходимая для получения оптимальных статистических решений, считается полной, если известно совместное априорное распределение $P(x, y)$ воспроизводимых сообщений $x(t)$ и смеси $y(t)$ сигнала [несущего сообщения $x(t)$] и помех, а также функция потерь $I(x, x^*)$ (см. § 2.2., 2.3). Так как

$$P(x, y) = P(x)P(y|x), \quad (3.46)$$

то требуется знание априорного распределения сообщений $P(x)$ и функции правдоподобия $P(y|x)$. При наличии такой полной информации оптимальный алгоритм обработки смеси $y(t)$, т. е. оптимальный закон

$$x^* = \Gamma(y(t)) \quad (3.47)$$

преобразования входной смеси $y(t)$ в принимаемое системой решение $x^*(t)$ о значении воспроизводимого сообщения, находится из условия¹

$$R = \int \int I(x, \Gamma(y)) P(x, y) dx dy = \min_{\Gamma(y)} \quad (3.48)$$

¹ Иногда она не очень удачно называется теорией оптимального стохастического управления.

т. е. из условия минимизации среднего риска R^1 . Однако в большинстве реальных задач распределение $P(x, y)$ известно неполностью или невозможно строго обосновать вид функции потерь $I(x, x^*)$. В таких случаях определить средний риск R принципиально невозможно. Но даже если $P(x, y)$ и $I(x, x^*)$ известны, выражение (3.48) может оказаться настолько сложным, что получить математическое решение задачи (3.48) окажется практически невозможным.

Существует много различных путей преодоления этих трудностей. Они рассмотрены, в частности, в [30, 31, 34—40]. Поэтому ограничимся лишь краткой характеристикой этих путей. Они зависят от того, что именно $[P(x), P(y|x)$ или $I(x, x^*)]$ неизвестно и в какой степени, а также от решаемой задачи — передача или извлечение информации, вид воспроизводимых сообщений $x(t)$ и др.

Случай неизвестных $P(x)$ и $I(x, x^*)$. Если неизвестно априорное распределение сообщений $P(x)$, то применяют методы, уже рассмотренные в § 2.2 и 2.3. Если неизвестен вид функции потерь $I(x, x^*)$, то, как указывалось в § 2.1, при дискретных сообщениях полагают все весовые коэффициенты одинаковыми ($c_1 = \dots = c_m = 1$), т. е. принимают все виды ошибок воспроизведения сообщений одинаково опасными. При непрерывных сообщениях, если вид функции $I(x, x^*)$ неизвестен, обычно полагают

$$I(x, x^*) = (x - x^*)^2 \quad (3.49)$$

(квадратичная функция потерь) либо принимают

$$I(x, x^*) = \text{const} - \delta(x - x^*) \quad (3.50)$$

(простая функция потерь). В первом случае обеспечивается минимум среднего квадрата ошибки воспроизведения сообщений, а во втором — максимум (по x) апостериорной плотности вероятности $P(x|y)$ воспроизводимого сообщения. Функция потерь вида (3.49) обычно удовлетворительно соответствует назначению системы и в то же время удобна с точки зрения простоты отыскания оптимального алгоритма $\Gamma_{\text{опт}}(y)$ и его реализации. Функция (3.50) во многих случаях еще более упрощает процедуру отыскания оптимального алгоритма, но может менее удовлетворительно соответствовать назначению системы. Однако в ряде работ показано (см., например, [5]), что при высоких требованиях к точности воспроизведения непрерывных сообщений (т. е. помехоустойчивости системы) вид априорного распределения $P(x)$ и функции потерь $I(x, x^*)$ мало влияют на результаты синтеза.

Случай неполноты известной функции правдоподобия $P(y|x)$. Наибольшие трудности возникают при неполной известности функции правдоподобия $P(y|x)$, т. е. закона распределения смеси сиг-

¹ Здесь предполагается для простоты, что алгоритм (3.47) является детерминированным (нерандомизированным), т. е. каждой реализации $y(t)$ соответствует вполне определенное значение воспроизводимого сообщения $x^*(t)$ (т. е. вполне определенное, а не случайное решение).

нала и помех $y(t)$ при данном (фиксированном) значении сообщения x . Здесь характерны следующие случаи.

1. Распределение $P(y|x)$ известно с точностью до некоторого вектора $\theta = (\theta_1, \dots, \theta_m)$ параметров, а распределение $P(\theta)$ известно, т. е. известны распределения $P(y|x, \theta)$ и $P(\theta)$. Тогда функцию правдоподобия можно вычислить по формуле

$$P(y|x) = \int_{\theta_1} \dots \int_{\theta_m} P(y|x, \theta_1, \dots, \theta_m) P(\theta_1, \dots, \theta_m) d\theta_1 \dots d\theta_m.$$

Однако практическое вычисление этого интеграла во многих задачах может представить большие или даже непреодолимые трудности. Тогда проще считать распределение $P(\theta)$ неизвестным, т. е. перейти к случаю, рассматриваемому далее.

2. Вид условного распределения $P(y|x, \theta)$ известен, но закон распределения $P(\theta)$ вектора параметров θ неизвестен. Здесь возможны следующие пути решения:

а) применяется минимаксный метод — ищется наименее благоприятное распределение $P_{\max}(\theta)$, т. е. такое, при котором минимальное значение $R_{\min}$ среднего риска R имеет максимальное значение. Это распределение $P_{\max}(\theta)$ и принимается за истинное. Такому пути свойственны два основных недостатка — математический [трудно найти распределение $P_{\max}(\theta)$] и принципиальный [синтезированная система оказывается оптимальной лишь в предположении, что вектор θ имеет наиболее опасное (в смысле помехоустойчивости) распределение, тогда как в действительности оно может быть более благоприятным (менее опасным)];

б) полагается $P(\theta) = \begin{cases} \text{const} & \text{при } \theta \in \Theta, \\ 0 & \text{вне этой области.} \end{cases}$ Такое допущение резко упрощает процедуру синтеза. Кроме того, равномерное распределение часто оказывается наименее благоприятным;

в) к синтезируемой системе предъявляется требование, чтобы она в процессе функционирования производила оценку неизвестного вектора параметра θ , т. е. требование воспроизведения сообщения x расширяется до требования воспроизведения пары (x, θ) . При этом в (3.48) полагается, что распределение $P(x, y) = P(x)P(y|x)$ известно, так как единственный неизвестный при этом вектор параметров θ считается известным и равным его оценке θ^* , полученной в процессе функционирования системы. Если распределение $P(x)$ неизвестно, то оценка параметра θ производится методом максимального правдоподобия, т. е. исходя из условия $P(y|x, \theta) = \max_{\theta}$;

г) применение (если это технически возможно) режима обучения (см. далее).

3. Не является полностью известным даже вид распределения $P(y|x)$, т. е. это распределение нельзя считать известным за исключением лишь вектора параметров θ . Такой случай называется **непараметрическим**.

Непараметрические задачи. Конечно, какие-то априорные сведения о распределении $P(y|x)$ необходимы (так как если о помехах и сигналах абсолютно ничего неизвестно, то задача обеспечения помехоустойчивости вообще теряет смысл), но в непараметрическом случае они могут быть весьма скудными. Например, о смеси $y(t)$ сигнала и помех может быть известно лишь то, что отсчеты $y(t_1), \dots, y(t_n)$, производимые в моменты $t_1, \dots, t_n$, статистически независимы, а распределение $P(y|x)$ при отсутствии сигнала ($x=x_2$) четное, а при наличии сигнала ($x=x_1$) становится нечетным. В непараметрических задачах применяются следующие основные методы.

1. Ограничение класса синтезируемых систем таким образом, чтобы для синтеза в этом классе не требовалось знания всех статистических характеристик смеси $y(t)$. Например, для решения задачи оптимальной фильтрации сигнала в классе линейных систем по критерию минимума среднего квадрата ошибки воспроизведения требуется знать лишь математические ожидания и корреляционные функции сигнала и помехи и не требуется знания полного распределения $P(y|x)$.

2. Применение критерия близости. Если сигнал известен точно (за исключением несомого им сообщения), т. е. $u_c = u_c(t, x)$, то можно применять критерий близости, т. е. считать, что передавалось то значение $x^*(t)$ сообщения $x(t)$, при котором сигнал $u_c(t, x^*)$ наиболее близок в некотором смысле к принятой реализации $y(t)$ смеси сигнала и помех, т. е. «расстояние» d между $u_c(t, x^*)$ и $y(t)$ минимально.

Важным частным случаем этого метода является так называемый метод наименьших квадратов (МНК). Он состоит в том, что приемное устройство выдает на выходе решение $\gamma = x^*(t)$, где $x^*(t)$ является решением задачи

$$d = \int_0^T [y(t) - u_c(t, x)]^2 dt = \min_x \quad (3.51)$$

Очевидно, в данном случае «расстояние» d с точностью до постоянного множителя $(1/T)$ совпадает со средним (за время T) квадратом отклонения сигнала $u_c(t, x)$ от принятой реализации $y(t)$. Основное достоинство МНК состоит в том, что не требуется знания закона распределения смеси $y(t)$ — используется лишь априорная информация о том, что реализация $y(t)$ достаточно близка к $u_c(t, x)$, т. е. что малые значения помех более вероятны, чем большие.

Алгоритм $\Gamma(y)$, найденный этим методом, в общем случае не является строго оптимальным [поскольку в общем случае при решении задачи (3.51) может использоваться далеко не вся реально существующая априорная информация о сигнале и помехах]. Однако доказано, что он совпадает с оптимальным алгоритмом, найденным при наличии полной априорной информации, по меньшей мере в любом из следующих случаев [46]:

а) единственная помеха — аддитивный независимый гауссов-

ский белый шум с известной спектральной плотностью. Критерием оптимальности алгоритма (приемника) является максимум правдоподобия, а при равномерном априорном распределении $P(x)$ — максимум апостериорной вероятности (плотности вероятности) $P(x)$ сообщения.

Как уже отмечалось, при высокой точности воспроизведения сообщений такой же результат получается и при других разумных критериях оптимальности [других видах функции потерь $I(x, x^*)$], особенно если сообщение можно считать не изменяющимся за время его воспроизведения;

б) помеха и сообщение аддитивны, независимы и имеют гауссовское распределение, а функция потерь квадратична;

в) отсчеты $y = y(t_i)$ ($i = \overline{1, n}$, $n = T/\Delta t$, Δt — интервал времени между соседними отсчетами) между собой статистически независимы, алгоритм ищется в классе линейных, а критерием оптимальности является минимум среднего квадрата ошибки воспроизведения сообщений. [В справедливости сказанного в случае а) нетрудно убедиться, если сравнить (3.51) с выражением (3.6) для апостериорной вероятности (плотности вероятности): из (3.6) видно, что максимуму по x функции $P(y|x)$ соответствует минимум интеграла $\int_0^T [y(t) - u_c(t, x)]^2 dt$ и наоборот.]

В ряде задач алгоритм, найденный на основе МНК, дает значительно худшие результаты, чем алгоритм, оптимальный для полной информации. Следовательно, то обстоятельство, что в МНК не учитывается никакая априорная информация (кроме информации о близости) может в некоторых случаях привести к значительному ухудшению помехоустойчивости. В этих случаях, как указано в [46], можно несколько видоизменить МНК, введя в него дополнительную известную априорную информацию.

3. Применение каких-либо других (кроме основанных на критерии близости) непараметрических методов. Известно большое число различных непараметрических методов. Однако они разработаны главным образом лишь для задач обнаружения сигнала и в предположении, что имеется большое число n независимых выборочных значений $y_1, \dots, y_n$ смеси $y(t)$ сигнала и помех. Основная идея этих методов заключается в следующем. Исходную совокупность $\langle y_1, \dots, y_n \rangle$ выборочных значений, называемую исходной статистикой, преобразуют к такому виду, чтобы закон распределения преобразованной статистики был известен и не зависел от неизвестных характеристик исходной статистики. Затем ищут такой алгоритм обработки преобразованных статистических данных, который давал бы приемлемое качество обнаружения. В некоторых задачах удается найти алгоритмы, которые при $n \rightarrow \infty$ обеспечивают такое же качество обнаружения, как оптимальные при полностью известном законе распределения $P(y|x)$ исходной статистики. В этом случае найденные алгоритмы называют *асимптотически* (при $n \rightarrow \infty$) *оптимальными*.

При некоторых допущениях удается обосновать строгий метод отыскания асимптотически оптимальных алгоритмов обнаружения сигнала. [Основные допущения состоят в том, что размер выборки n полагается весьма большим ($n \rightarrow \infty$) и с ростом n энергия сигнала, требуемая для его надежного обнаружения, остается неизменной, так что амплитуда сигнала уменьшается пропорционально $1/\sqrt{n}$. Поэтому оказывается возможным применение аппроксимаций, справедливых для весьма малых отношений сигнал-шум.]

Основным достоинством непараметрических методов является то, что они позволяют найти алгоритмы обработки входной смеси $y(t)$ сигнала и помех при весьма малых априорных данных о статистических характеристиках этой смеси. В задачах обнаружения сигнала, как было показано, обеспечивается, кроме того, полная независимость (инвариантность) вероятности ложной тревоги $P_{\text{лт}}$ от характеристик входной смеси. Это свойство полученных алгоритмов весьма полезно, так как в процессе работы реального обнаружителя характеристики помех могут меняться в широких пределах и выбор в таких условиях значения порога является одной из основных проблем при проектировании обнаружителя.

Однако скудность априорной информации, на основе которой приходится выносить окончательное решение о воспроизводимом сообщении, не может не сказаться отрицательно на качестве этого решения. Как уже отмечалось, качество решения обычно получается удовлетворительным, лишь если входная реализация $y(t)$ содержит весьма большое число статистически независимых отсчетов, т. е. может наблюдаться в течение длительного времени T . Отсюда естественным образом вытекают следующие два пути преодоления недостаточности априорных данных — применение принципов обучения и адаптации.

Применение принципов обучения и адаптации. Принцип обучения применяется в тех случаях неполной априорной информации, когда время T , отводимое на воспроизведение сообщений, недостаточно для обеспечения требуемого качества воспроизведения, но можно организовать дополнительную (обучающую) реализацию (в течение дополнительного времени ΔT — времени обучения) $\langle y'_1, \dots, y'_n \rangle$, путем обработки которой можно получить недостающие статистические сведения о рабочей реализации $\langle y_1, \dots, y_n \rangle$. Полученные в процессе обучения сведения используются затем в алгоритме обработки рабочей реализации.

Различают обучение с учителем (с классификацией) и самообучение (без классификации, без учителя). В первом случае в процессе обучения используется дополнительная (по сравнению с рабочим режимом) априорная информация о характеристиках смеси $y(t)$. Во втором случае такая дополнительная информация отсутствует. Например, в задачах обнаружения сигнала при обучении с учителем указывается, присутствует ли в данной обучающей реализации полезный сигнал (и если присутствует, то какие он имеет параметры) или отсутствует. Очевидно, обучение с учителем при прочих равных условиях эффективнее, чем самообуче-

ние. Однако режим обучения даже без учителя не всегда можно создать. Действительно, для того, чтобы обучение (или самообучение) улучшало качество воспроизведения сообщений, необходимо, чтобы статистические характеристики реализаций были достаточно близки к статистическим характеристикам рабочих реализаций (в противном случае система будет обучаться не тому, чему следует, и обучение принесет не пользу, а вред). Для этого обычно требуется, чтобы обучающие реализации не были слишком удалены по времени от рабочих реализаций. Например, может потребоваться неоднократное чередование во времени рабочих и обучающих реализаций.

Принцип адаптации (приспособления) состоит в том, что недостающая априорная информация о смеси $y(t)$ сигнала и помех извлекается самой радиосистемой (радиоустройством) в процессе ее эксплуатации и используется при формировании алгоритма воспроизведения сообщений для улучшения качества воспроизведения. Очевидно, адаптация может иметь место как при наличии обучающих реализаций, так и при их отсутствии — в первом случае недостающая априорная информация может извлекаться как из рабочих, так и из обучающих реализаций, а во втором — только из рабочих. С другой стороны, система с обучением не обязательно является адаптивной, так как обучение может производиться и до начала эксплуатации системы — в процессе ее разработки и производства.

Как отмечено Р. Л. Стратонович в [36], главными особенностями, свойственными принципу адаптации, являются следующие.

1. Расширение числа оцениваемых (измеряемых) системой параметров сигналов и помех путем включения параметров, характеризующих неопределенность априорного распределения. Например, если сигнал $u_c(t, x, \theta)$ содержит совокупность (вектор) неинформационных параметров $\theta = \{\theta_1, \dots, \theta_2\}$, априорное распределение которых неизвестно, то можно потребовать, чтобы система оценивала (измеряла) не только полезное сообщение x , но и все неизвестные параметры $\theta_1, \dots, \theta_2$.

2. Использование полученных системой оценок неинформационных параметров сигналов и помех для повышения качества воспроизведения полезного сообщения x . Это достигается при формировании оптимальных алгоритмов обработки информации путем замены априорных вероятностей, математических ожиданий и других параметров априорных распределений их эмпирическими [т. е. полученными в результате обработки смеси $y(t)$] оценками. *

3. Широкое использование стохастических итерационных методов, приводящих к рекуррентным, т. е. многошаговым алгоритмам. При этом каждый параметр оценивается в дискретном времени и каждая последующая (во времени) оценка θ_{i+1}^* формируется с учетом оценки θ_i^* , полученной на предыдущем шаге. С течением времени (с увеличением числа шагов) качество оценки улучшается и (если алгоритм обладает свойством асимптотической эффективности) стремится к тому значению, которое имело бы место

при наличии полной априорной информации. Принципы реализации таких алгоритмов и анализ их качества рассматривается, в частности, в теории стохастической аппроксимации (см., например, [5, 38, 39]).

* * *

Из изложенного следует, что существует много методов (путей) преодоления априорной неопределенности. Наибольшие трудности возникают при неполностью известных статистических характеристиках сигналов и помех. В таких задачах высокое качество воспроизведения сообщений можно обеспечить обычно лишь при больших размерах выборки n смеси $y(t)$ сигнала и помех ($n \gg 1$). Поэтому, если по условиям задачи размер рабочей выборки невозможно сделать достаточно большим, желательно рассмотреть возможность получения дополнительных выборок хотя бы в процессе обучения и повысить качество воспроизведения сообщений (помехоустойчивость системы) применением обучения — желательно с «учителем» (с классификацией), а в случае невозможности — без «учителя» (самообучение).

При большом размере выборки (рабочей или совокупности рабочей и обучающей) часто целесообразно употребить принцип адаптации. Однако принципы обучения и адаптации могут привести к существенному усложнению системы. Кроме того, их применение существенно затрудняется при нестационарном характере помех. Особенно большие трудности возникают при необходимости обеспечения высокой помехоустойчивости в условиях действия организованных помех. В таких задачах основными математическими методами синтеза оптимальных алгоритмов являются методы, основанные на теории игр ([5, 25] и др.). Синтезированный на основе этой теории алгоритм обеспечивает минимум среднего риска R для такой помехи, при которой этот минимум оказывается максимальным (т. е. применяется минимаксный критерий оптимальности системы). Принципиальным недостатком такого метода является обеспечение наилучшей защиты не от реальных помех, которые будут действовать на систему, а лишь от теоретически наиболее опасных. Кроме того, синтез оптимальных алгоритмов на основе теории игр часто наталкивается на большие математические трудности.

Из изложенного следует, что математический и инженерный синтез систем, оптимальных по помехоустойчивости, является одним из сложных, если не самым сложным этапом проектирования, и требует, как правило, использования в процессе синтеза не одного, а нескольких методов преодоления недостаточности априорных сведений.

Литература: основная [5, 40, 42, 43], дополнительная [14—17, 20, 23, 29—31, 34—39, 41, 44].

Глава 4

ПРОПУСКНАЯ СПОСОБНОСТЬ РАДИОСИСТЕМ

4.1. ВВОДНЫЕ ЗАМЕЧАНИЯ

Пропускная способность системы передачи или извлечения информации соответственно характеризует максимально допускаемую этой системой скорость передачи или извлечения информации с заданной степенью достоверности (точности). Для количественной характеристики пропускной способности применяются различные показатели в зависимости от вида и назначения системы. Так, если СПИ содержит один канал или несколько однородных каналов (например, только телефонные каналы), то ее пропускную способность удобно оценивать в битах в секунду. Если же СПИ содержит разнородные каналы (например, телефонные и телевизионные), то в общем случае объединять потоки передаваемой по ним качественно различной информации в один общий поток недопустимо и приходится оценивать пропускную способность системы в целом не одним числом, а несколькими числами — отдельно для каждой группы каналов. При этом можно оценивать пропускную способность системы числами каналов каждого вида. Однако в ряде случаев потоки на выходах всех разнородных каналов имеют цифровой код в двоичной системе исчисления и суммируются в общий цифровой поток двоичных символов для последующей обработки в ЭВМ и/или последующей передачи, осуществляемой другой системой. В таких случаях полезно оценивать пропускную способность системы в целом также одним числом, измеряемым в битах в секунду.

Аналогичная ситуация имеет место и в системах извлечения информации (СИИ). В ряде случаев СИИ, например радиолокатор, имеет несколько каналов, извлекающих качественно различные виды информации (о наличии или отсутствии целей, о числе и типах целей, их координатах, скоростях и т. п.). При этом суммировать все качественно различные виды информации в один общий поток и измерять его в битах в общем случае недопустимо — приходится оценивать пропускную способность СИИ другим способом, например числом целей данного вида, обслуживаемых в единицу времени или одновременно. Однако если потоки информации на выходе всех каналов СИИ имеют цифровой вид и объединяются в единый двоичный цифровой поток для последующего ввода в ЭВМ и/или дальнейшей передачи информации, то целесообразно характеризовать суммарную пропускную способность СИИ не только числом обслуживаемых целей, но и показателем, выраженным в битах в секунду.

Таким образом, в зависимости от характера и назначения системы ее пропускную способность можно оценивать одним или

несколькими показателями и измерять в битах в секунду или в других единицах. В дальнейшем будем для простоты рассматривать лишь первый случай. При этом под пропускной способностью C системы (СПИ или СИИ) понимается максимальное (по всем возможным источникам I информации) значение средней скорости R передачи или извлечения информации, допускаемое этой системой при заданных требованиях к достоверности (точности) передачи (извлечения) информации:

$$C = \max_{(I)} R \quad \text{при } \varepsilon \leq \varepsilon_{\text{доп}}. \quad (4.1)$$

Здесь

$$R = J_T/T; \quad (4.2)$$

J_T — среднее количество информации (в битах), передаваемой или извлекаемой системой за время T (в секундах); ε — показатель достоверности (точности) передачи (извлечения) информации ($P_{\text{ош}}$, δ^2 и т. п.).

В процессе проектирования приходится сравнивать различные варианты S построения системы и выбирать наилучший из некоторого множества M_S допустимых вариантов. Поэтому представляет интерес определение максимального значения $C_{\text{макс}}$ пропускной способности, являющегося решением задачи

$$C_{\text{макс}} = \max_{S \in M_S} C \quad \text{при } \varepsilon \leq \varepsilon_{\text{доп}}. \quad (4.3)$$

При этом в процессе отыскания максимума величины C может варьироваться не только структура системы, но и значения ее параметров. Чем шире множество M_S , т. е. чем менее жесткими являются ограничения, накладываемые в процессе решения задачи (4.3), тем большим (лучшим) будет значение величины $C_{\text{макс}}$. То предельное значение величины $C_{\text{макс}}$ (а значит, и величины C), которое может быть достигнуто при весьма слабых ограничениях, будем называть *потенциальной* (предельной теоретически достижимой) *пропускной способностью* и обозначать C_0 . Следовательно, по определению

$$C_0 = \max_{S \in M_{S0}} C \quad \text{при } \varepsilon \leq \varepsilon_{\text{доп}}. \quad (4.4)$$

где M_{S0} — достаточно широкое множество (класс) рассматриваемых вариантов S построения системы.

Очевидно, величина C_0 будет зависеть от того, что конкретно будет пониматься под «достаточно широким множеством M_{S0} » или соответственно под «достаточно слабыми ограничениями вариантов S построения системы». Поэтому потенциальная пропускная способность системы, равно как и другие потенциальные характеристики (например, потенциальная помехоустойчивость), существенно зависит от исходных данных, принятых при ее определении.

При определении пропускной способности (C или C_0) в общем случае следует учитывать действие помех и других источников

ошибок. Поэтому между показателями пропускной способности, точности и помехоустойчивости существует взаимная связь — пропускная способность (C и C_0) зависит, как правило, от допустимых значений показателей точности и помехоустойчивости.

4.2. ОПРЕДЕЛЕНИЕ ПРОПУСКНОЙ СПОСОБНОСТИ СПИ В КЛАССИЧЕСКОЙ ТЕОРИИ ИНФОРМАЦИИ

Структура СПИ. Классическая теория информации, основы которой были заложены К. Шенноном [47], обычно изучается в курсе «Системы передачи информации». Поэтому мы рассмотрим лишь основные результаты этой теории, необходимые для понимания последующего материала.

Система передачи информации полагается состоящей из трех частей (рис. 4.1) — канала связи, а также кодера и декодера,

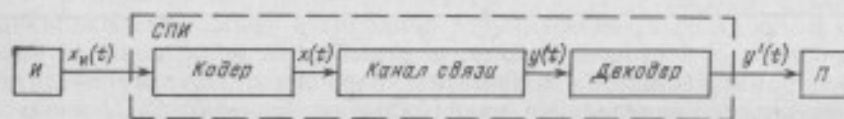


Рис. 4.1

согласующих канал с источником информации И и ее получателем (адресатом) П. В частном случае кодер и декодер могут отсутствовать. Канал может быть дискретным или непрерывным, а источник И может выдавать на своем выходе дискретные или непрерывные сообщения.

Передача дискретных сообщений по дискретному каналу. При этом источник вырабатывает информацию со средней скоростью

$$R_n = V_n H_n(x_n), \quad (4.5)$$

где V_n — средняя скорость выработки символов источником; $H_n(x_n)$ — среднее количество информации в битах, приходящееся на один символ, называемое также *энтропией* или *неопределенностью источника*. Величина R_n называется *производительностью источника*. Если символы источника статистически независимы, то

$$H_n(x_n) = - \sum_{i=1}^{n_n} P(x_i) \log_2 P(x_i), \quad (4.6)$$

где n_n — число видов символов источника; $P(x_i)$ — вероятность появления символа x_i .

Если между символами источника имеется статистическая связь, то $H_n(x_n)$ уменьшается по сравнению с (4.6). Энтропия $H_n(x_n)$ максимальна, когда все символы статистически независимы и равновероятны. При этом, как следует из (4.6),

$$H_n(x_n) = H_{\max}(x_n) = \log_2 n_n. \quad (4.7)$$

В отсутствие шумов дискретный канал характеризуется числом n_n видов передаваемых им символов и допустимой средней скоростью V_n следования этих символов.

Предположим сначала, что кодер (и декодер) отсутствует, т. е. источник И непосредственно подключен ко входу канала. Тогда пропускную способность канала нетрудно найти из общего соотношения (4.1). В соответствии с этим скорость передачи информации максимальна, если источник обладает следующими характеристиками:

$$n_n = n_n, \quad V_n = V_n; \quad H_n(x_n) = H_{\max}(x_n) = \log_2 n_n. \quad (4.8)$$

Поэтому

$$C = V_n \log_2 n_n. \quad (4.9)$$

Так как источники искажений в кодере, декодере и канале полагаются отсутствующими, то информация передается без искажений, т. е. $\varepsilon = 0$.

Если источник не удовлетворяет условиям (4.8), т. е. не согласован с каналом, то для обеспечения передачи информации от этого источника со скоростью, равной пропускной способности канала (4.9), необходимо включить согласующие элементы — кодер и декодер. При этом по отношению к каналу источником информации является кодер. Поэтому на выходе кодера должны выполняться условия (4.8), в которых под n_n , V_n , H_n следует понимать значения соответствующих параметров на выходе кодера.

Как доказал К. Шеннон, если производительность источника И не превышает C , осуществить такое кодирование символов источника в принципе всегда возможно. При этом источник должен вырабатывать символы со скоростью $V_n = C/H_n$.

При наличии помех дискретный канал характеризуется не только параметрами n_n и V_n , но и матрицей $\|P(y_j|x_i)\|$ вероятностей перехода символа x_i на входе канала в символ x_j на его выходе ($i = \overline{1, n_n}$; $j = \overline{1, n_n}$). При этом символ x_j обозначается здесь через y_j для того, чтобы подчеркнуть, что он появляется на выходе канала, т. е. в составе процесса $y(t)$, а не на его входе, где действует $x(t)$. Пропускная способность такого канала

$$C = V_n \max_{(x)} [H(x) - H(x|y)], \quad (4.10)$$

где $H(x|y)$ — условная энтропия, определяемая соотношением

$$H(x|y) = - \sum_{i=1}^{n_n} \sum_{j=1}^{n_n} P(x_i, y_j) \log_2 P(x_i|y_j), \quad (4.11)$$

и характеризующая уменьшение скорости передачи информации за счет действия помех в канале [в отсутствие помех $H(x|y) = 0$].

Соотношение (4.11), равно как и (4.6), справедливо в предположении, что статистическая связь между символами x_i (а также между символами y_j) с различными индексами отсутствует.

Как доказал К. Шеннон, передачу информации по каналу со скоростью, равной его пропускной способности (4.10), можно обес-

печать при сколь угодно малой вероятности ошибки ($P_{\text{ош}} \rightarrow 0$), если выполняется условие $R_k \leq C$ и осуществляется идеальное кодирование (и декодирование) информации, поступающей от источника И. Идеальное кодирование (и декодирование) предполагает совместную обработку последовательности сколь угодно большого числа символов и соответственно требует неограниченно большой задержки τ_z в передаче информации ($\tau_z \rightarrow \infty$).

Частным, но весьма важным случаем дискретного канала является *двоичный симметричный канал без памяти*, т. е. канал, в котором $n_k = 2$, вероятности ошибок одинаковы

$$P(0|1) = P(1|0) = P_{\text{ош},k}$$

и осуществляется независимый прием каждого символа (так называемый посимвольный или элементарный прием). В этом случае в соответствии с (4.6), (4.9) и (4.10)

$$C = V_k [1 + P_{\text{ош},k} \log_2 P_{\text{ош},k} + (1 - P_{\text{ош},k}) \log_2 (1 - P_{\text{ош},k})]. \quad (4.12)$$

При этом следует иметь в виду различие между вероятностями $P_{\text{ош},k}$ и $P_{\text{ош}}$, т. е. соответственно вероятностью ошибочного приема каждого символа в канале и вероятностью ошибки воспроизведения информации на выходе декодера.

Благодаря наличию идеального кодирования и декодирования вероятность $P_{\text{ош}}$ может быть сколь угодно малой даже при большой вероятности ошибки $P_{\text{ош},k}$, т. е. и при весьма плохом в смысле помехоустойчивости канале. Однако даже при идеальном кодировании действие помех может весьма сильно снижать пропускную способность системы. Например, из (4.12) следует, что при $P_{\text{ош},k} \rightarrow 0,5$ пропускная способность C стремится к нулю. По мере уменьшения действия помех в канале, т. е. при уменьшении вероятности $P_{\text{ош},k}$, пропускная способность (4.12) стремится к $V_k \cdot 1$, т. е. к тому значению, которое она имела бы при полном отсутствии помех. Из (4.12) нетрудно убедиться, что если

$$P_{\text{ош},k} \leq 10^{-2}, \text{ то } C \geq 0,92 V_k. \quad (4.13)$$

т. е. при этом пропускная способность уменьшается за счет действия помех менее чем на 10%.

Анализ общих соотношений (4.8)–(4.10) показывает, что в общем случае дискретного канала выполнение условия (4.13) достаточно для того, чтобы пропускную способность канала можно было рассчитывать по той же формуле, что и в отсутствие помех, т. е. вместо (4.10) можно полагать

$$C \approx V_k \log_2 n_k, \text{ если } P_{\text{ош},k} \leq 10^{-2}. \quad (4.14)$$

Передача дискретных сообщений по непрерывному каналу. Непрерывным называется канал, способный передавать сигналы, непрерывные как по времени, так и по значениям в каждый момент времени. Полагается, что передаваемые по каналу сигналы имеют спектр, ширина которого равна полосе пропускания Δf канала и длительность T , такую, что $\Delta f T \gg 1$. При этом передаваемые сигналы $u(t)$ полностью характеризуются своими отсчетами, взя-

тыми в дискретные моменты времени с временным интервалом $\Delta t = 1/2\Delta f$. Поэтому передачу сигналов по непрерывному каналу можно рассматривать как передачу отсчетов, следующих со скоростью $V_k = 1/\Delta t = 2\Delta f$.

Значение каждого отсчета непрерывно, т. е. может иметь бесконечное число (точнее — континуум) значений даже в пределах сколь угодно малого интервала Δt этих значений. Поэтому непрерывный канал можно рассматривать как предел дискретного канала, в котором роль передаваемых символов играют отсчеты, а число n_k возможных видов (значений) этих символов стремится к бесконечности. Отсюда следует, что все соотношения, полученные для дискретного канала, остаются справедливыми и для непрерывного канала, если в них положить $V_k = 2\Delta f$ и $n_k \rightarrow \infty$. В частности, из (4.9) следует, что в отсутствие шумов пропускная способность непрерывного канала стремится к бесконечности и, следовательно, имеет смысл определять пропускную способность непрерывного канала лишь в присутствии помех (шума). При этом формула (4.10) принимает вид

$$C = 2 \Delta f \max_{(y)} [h(x) - h(x|y)], \quad (4.15)$$

$$\left. \begin{aligned} \text{где } h(x) &= - \int_{-\infty}^{\infty} W(x) \log_2 W(x) dx, \\ h(x|y) &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(x, y) \log_2 W(x|y) dx dy \end{aligned} \right\} \quad (4.16)$$

— так называемые *дифференциальные энтропии*, а $W(x)$ и $W(x|y)$ — безусловная и условная плотность вероятности величины x .

Формулы (4.16) получены из (4.8) и (4.10) и поэтому справедливы лишь при отсутствии статистической связи между отсчетами. В (4.15) максимум ищется по всем источникам И, т. е. по всем возможным видам распределения $W(x)$.

Если дисперсия σ_x^2 величины x фиксирована, то максимум энтропии $h(x)$ имеет место при нормальном (гауссовском) законе распределения $W(x)$ и равен

$$h_{\text{макс}}(x) = \log_2 \sqrt{2\pi e \sigma_x^2}, \quad (4.17)$$

где $e = 2,71828...$

Если единственным источником искажений в канале является аддитивный независимый нормальный белый шум со спектральной плотностью N_0 , то из общей формулы (4.15) получается

$$C = \Delta f \log_2 (1 + P_c / N_0 \Delta f), \quad (4.18)$$

где P_c — средняя мощность сигнала.

Формула (4.18), называемая далее *основной формулой Шеннона*, показывает, какую пропускную способность при передаче дискретных сообщений имеет непрерывный канал с полосой пропускания Δf , в котором действует аддитивный независимый нор-

мальный белый шум с фиксированной спектральной плотностью N_0 и задана лишь средняя мощность P_c сигнала, т. е. ограничений на его пиковую мощность не накладывается. Формулу (4.18) часто удобно записывать в следующем виде:

$$\beta_E = \beta_{\Delta f} (\exp[(\ln 2)/\beta_{\Delta f}] - 1), \quad (4.19)$$

$$\text{где } \beta_E = P_c/N_0 C \quad (4.20)$$

$$\text{и } \beta_{\Delta f} = \Delta f/C \quad (4.21)$$

— нормированные (безразмерные) величины, называемые соответственно *удельным* (на единицу пропускной способности) *расходом* средней мощности сигнала и *полосы пропускания* системы. По формуле (4.19) построена зависимость, изображенная на рис. 4.2 сплошной линией.

Чем меньше каждая из величин $(\beta_E, \beta_{\Delta f})$, тем лучше система при прочих равных условиях. Поэтому удельные расходы β_E и $\beta_{\Delta f}$ можно рассматривать как показатели качества системы передачи информации, а зависимость (4.19) (и соответствующую ей кривую на рис. 4.2) — как диаграмму обмена между этими показателями: она показывает, насколько можно улучшить (уменьшить) каждый из этих показателей качества, если допустить то или иное ухудшение (увеличение) второго показателя качества.

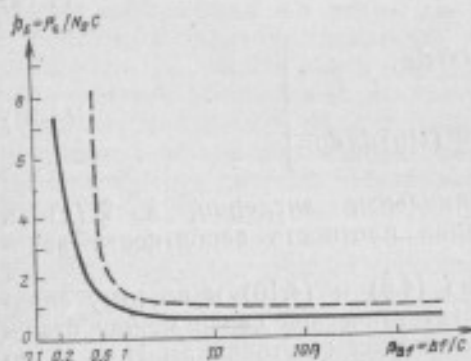


Рис. 4.2

Поскольку, как уже отмечалось, непрерывный канал можно рассматривать как предел дискретного канала, имеющий место при $n_k \rightarrow \infty$, а с ростом n_k пропускная способность дискретного канала возрастает, справедливо следующее утверждение: если в дискретном канале с полосой пропускания Δf единственным источником искажений сообщений является аддитивный независимый нормальный белый шум со спектральной плоскостью N_0 и ограничена (значением P_c) лишь средняя мощность сигнала, то при увеличении числа n_k видов символов пропускная способность канала возрастает, стремясь в пределе (при $n_k \rightarrow \infty$) к величине, определяемой формулой (4.18). Однако даже в двоичном симметричном канале без памяти ($n_k = 2$) проигрыш в величине C (по сравнению со случаем $n_k \rightarrow \infty$) получается сравнительно небольшим, если полоса пропускания канала относительно велика ($\Delta f \gg C$). В общем случае, т. е. при произвольном отношении $\beta_{\Delta f} = \Delta f/C$, различие между пропускными способностями непрерывного канала и двоичного симметричного канала без памяти характеризуется кривыми, приведенными на рис. 4.2: сплошной линией

для непрерывного канала и штриховой — для двоичного. При этом зависимость для двоичного симметричного канала без памяти (штриховая линия) получена из общей формулы (4.12) при следующих предположениях. Каждый из двоичных символов передается сигналом длительностью τ_0 , имеющим ширину спектра $\Delta f = 1/2\tau_0$, т. е. $V_k = 1/\tau_0 = 2\Delta f$. Форма сигналов и способ их обработки выбраны оптимальными по помехоустойчивости, т. е. обеспечивающими минимум полной вероятности ошибки приема символов при заданной средней мощности сигнала P_c . Как отмечалось в гл. 3, при двоичной передаче сообщений этому соответствуют противоположные сигналы $[u_c(t, x_2) = -u_c(t, x_1)]$ и корреляционный способ приема. В таком оптимальном по помехоустойчивости канале полная вероятность ошибки

$$P_{\text{ош.к}} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-z^2/2} dz, \quad (4.22)$$

$$\text{где } \alpha = \frac{2Q}{N_0} = \frac{P_c \tau_0}{N_0}.$$

Подстановка (4.21) и (4.22) в (4.12) и приводит к зависимости, изображенной на рис. 4.2 штриховой линией (см., например, [48]).

Из рис. 4.2 видно, что при $\beta_{\Delta f} \geq 0,6$ проигрыш γ в требуемой средней мощности сигнала в двоичном канале по сравнению с непрерывным невелик: $\gamma \leq 2$. В пределе (при $\beta_{\Delta f} \rightarrow \infty$)

$$\gamma = \pi/2. \quad (4.23)$$

Соответствующие зависимости $\beta_E = f(\beta_{\Delta f})$ для дискретных каналов с большим числом видов символов ($n_k > 2$) должны располагаться между штриховой и сплошной линиями, стремясь при $n_k \rightarrow \infty$ к сплошной линии. Отсюда следует, что основным преимуществом непрерывного канала по сравнению с дискретным является возможность существенного уменьшения удельного расхода средней мощности при малых значениях удельного расхода полосы ($\beta_{\Delta f} < 0,5$). Чем больше $\beta_{\Delta f}$, тем меньший выигрыш в значении β_E может дать увеличение числа видов символов.

К. Шенноном было доказано, что пропускную способность (4.15), а следовательно, и (4.18) можно в принципе реализовать (при передаче дискретных сообщений) со сколь угодно малой вероятностью ошибки ($P_{\text{ош.к}} \rightarrow 0$), если применить идеальное помехоустойчивое кодирование, основанное на преобразовании входной последовательности $x_n(t)$ символов в весьма длительные ($T \rightarrow \infty$) реализации $x(t)$ передаваемого по каналу сообщения и соответственно требующее весьма большой задержки ($\tau_s \rightarrow \infty$) в передаче информации от источника к потребителю.

Передача непрерывных сообщений. Непрерывные сообщения можно передавать как по непрерывному, так и по дискретному

каналу (в последнем случае¹ требуется конечно применение преобразователей аналог — цифра и цифра — аналог). Однако для краткости ограничимся случаем передачи по непрерывному каналу (поскольку, как было показано, различие в пропускных способностях непрерывного и дискретного каналов при $\beta_{\Delta f} \geq 0,6$ невелико). Предполагается, что непрерывное сообщение $x_n(t)$ имеет спектр, ограниченный высшей частотой F_n , и поэтому точно описывается последовательностью отсчетов, следующих во времени со скоростью $V_n = 1/\Delta t$, где $\Delta t = 1/2F_n$, т. е. $V_n = 2F_n$. Так как сообщение непрерывно, то каждый отсчет может принимать бесконечное число значений, т. е. $n_n \rightarrow \infty$. Поэтому производительность такого источника бесконечно велика ($R_n \rightarrow \infty$). Пропускная же способность непрерывного канала при наличии шумов конечна. Поэтому передавать сообщение такого источника по каналу с шумами со сколь угодно высокой точностью невозможно даже при наличии идеального помехоустойчивого кодирования (и декодирования).

К. Шеннон доказал, что при наличии идеального помехоустойчивого кодирования (и декодирования) можно обеспечить передачу непрерывных сообщений по каналу с пропускной способностью C лишь с относительным средним квадратом ошибки δ^2 , определяемым из соотношения

$$R_{n\delta} = C, \quad (4.24)$$

где

$$R_{n\delta} = F_n \log_2(1 + 1/\delta^2). \quad (4.25)$$

Здесь по определению

$$\delta^2 = (y' - x_n)^2 / x_n^2 = e / x_n^2$$

— средний квадрат e ошибки передачи сообщений, отнесенный к среднему квадрату передаваемого сообщения. Величина $R_{n\delta}$ называется e -производительностью источника, чтобы подчеркнуть, что она соответствует точной (с $P_{\text{ош}} \rightarrow 0$) передаче по каналу не истинного сообщения $x_n(t)$ источника, а сообщения $y'(t)$, отличающегося от $x_n(t)$ на величину $(y' - x_n)$, средний квадрат которой равен заданной (достаточно малой) величине e .

При выводе (4.25) предполагалось, что сообщение $x_n(t)$ имеет нормальный закон распределения и равномерный в пределах полосы $[0, F_n]$ частотный спектр². Из (4.24) и (4.25) следует, что при оптимальном кодировании сообщения $x_n(t)$, имеющего нор-

¹ При передаче непрерывных сообщений по непрерывному каналу для обеспечения оптимального кодирования преобразователи аналог — цифра и цифра — аналог также включаются в состав кодера и декодера.

² В литературе вместо (4.25) приводится формула $R_{n\delta} = F_n \log_2 1/\delta^2$ бит/с. Однако при выводе этой формулы К. Шенноном и другими авторами была допущена неточность. В этом легко убедиться, так как при $\delta^2 > 1$ получается $R_{n\delta} < 0$, что невозможно. Практического значения эта неточность не имеет, так как на практике всегда $\delta^2 \leq 0,1$, и можно полагать $1 + 1/\delta^2 \approx 1/\delta^2$.

мальный закон распределения и спектр, равномерный в полосе $[0, F_n]$ и равный нулю за ее пределами, возможно передавать сообщение по каналу с относительным средним квадратом ошибки δ^2 , определяемым соотношением

$$\log_2(1 + 1/\delta^2) = C/F_n. \quad (4.26)$$

Смысл этого соотношения, равно как и условия (4.24), вполне понятен: если бы пропускная способность C была бесконечно большой, то при идеальном кодировании можно было бы передавать сообщение $x_n(t)$ практически безошибочно ($\delta^2 \rightarrow 0$, $P_{\text{ош}} \rightarrow 0$). При конечной пропускной способности канала обеспечить такой результат невозможно даже при идеальном кодировании — чем меньше пропускная способность C канала, тем с большим средним квадратом ошибки возможно передавать по нему сообщения. Понятно также, что при этом должно играть роль не абсолютное значение пропускной способности, а ее отношение к высшей частоте F_n в спектре сообщения.

В рассмотренном важном случае гауссовского канала, т. е. при справедливости основной формулы Шеннона (4.18), выражение (4.26) принимает вид

$$\begin{aligned} \log_2(1 + 1/\delta^2) &= (\Delta f / F_n) \log_2(1 + P_c / N_0 \Delta f) = \\ &= \log_2(1 + P_c / N_0 \Delta f)^{\Delta f / F_n}, \end{aligned} \quad (4.27)$$

$$\text{т. е. } 1/\delta^2 = (1 + P_c / N_0 \Delta f)^{\Delta f / F_n} - 1. \quad (4.28)$$

Отсюда следует, что при $P_c / N_0 \rightarrow \infty$, т. е. в отсутствие помех, $\delta^2 \rightarrow 0$. По мере уменьшения отношения P_c / N_0 ошибка δ^2 монотонно возрастает.

По аналогии с удельными расходами (4.20) и (4.21) при передаче непрерывных сообщений удобно ввести удельные расходы вида

$$\beta'_E = P_c / N_0 F_n, \quad (4.29)$$

$$\beta'_{\Delta f} = \Delta f / F_n. \quad (4.30)$$

характеризующие расходы средней мощности сигнала и полосы пропускания канала на один герц полосы сообщения. В этих обозначениях соотношение (4.28) принимает следующий вид:

$$1 + 1/\delta^2 = (1 + \beta'_E / \beta'_{\Delta f})^{\beta'_{\Delta f}} \quad (4.31)$$

или

$$\beta'_E = \beta'_{\Delta f} [(1 + 1/\delta^2)^{1/\beta'_{\Delta f}} - 1]. \quad (4.32)$$

Соотношение (4.32) является эквивалентом зависимости (4.19), приведенной для дискретных сообщений. Принципиальным отличием (4.32) от (4.19) является наличие зависимости от допустимой ошибки δ^2 . По формуле (4.32) построены кривые, приведенные на рис. 4.3 для трех значений величины δ^2 . Графики на рис. 4.2 и 4.3 показывают связь между удельными расходами средней мощ-

ности сигнала и полосы пропускания и называются поэтому *диаграммами обмена* между этими важными показателями качества системы. Они показывают, насколько можно в принципе улучшить (уменьшить) один из этих показателей, если допустить ухудшение (увеличение) другого показателя.

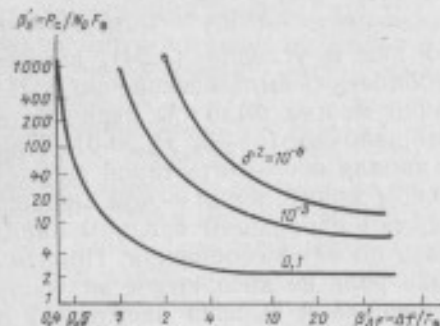


Рис. 4.3

В заключение отметим, что равенство (4.24) не только позволяет определить достижимое значение среднего квадрата ошибки δ^2 , но и является условием согласования источника, обладающего верхней частотой сообщений F_a , с каналом, имеющим пропускную способность C :

«Если источник при данной ошибке δ^2 имеет ϵ -производительность R_{ϵ} , то можно в принципе закодировать сообщения на выходе источника и передавать их по каналу с пропускной способностью C при погрешности воспроизведения, как угодно близкой к δ^2 , если только $R_{\epsilon} \leq C$. Это невозможно, если $R_{\epsilon} > C$ » (теорема Шеннона для непрерывного источника).

4.3. ПОТЕНЦИАЛЬНАЯ ПРОПУСКНАЯ СПОСОБНОСТЬ СПИ

Выясним, в какой мере изложенные ранее результаты классической теории информации могут быть использованы для оценки потенциальной пропускной способности СПИ. Рассмотрим сначала основную формулу Шеннона (4.18) [а следовательно, и соответствующие ей зависимости (4.19) и сплошную кривую на рис. 4.2]. Из § 4.2 следует, что формула (4.18) определяет в некотором смысле потенциальную (предельно теоретически достижимую) пропускную способность системы. Этот «некоторый смысл» состоит в следующем: при заданных значениях параметров P_c , N_0 и Δf системы ни при каком источнике и ни при каком принципе построения системы невозможно передавать по ней информацию от источника со скоростью, большей C , со сколь угодно малой вероятностью ошибки, если источник выдает дискретные сообщения, и с относительным средним квадратом ошибки δ^2 , если источник выдает непрерывные сообщения с нормальным законом распределения и прямоугольным частотным спектром шириной $[0, F_a]$. При этом δ^2 определяется соотношением (4.28).

Действительно, при получении этих результатов на систему передачи информации не накладывалось никаких ограничений, кроме задания значений P_c , N_0 , Δf , и при таких сравнительно слабых ограничениях искался максимум величины C по всем возможным способам кодирования и последующего декодирования сообщений

$x_n(t)$ источника. В такой широкой постановке задачи под способами кодирования и декодирования понимались по-существу все возможные способы преобразования сообщений $x_n(t)$ в сигналы и обратного преобразования смеси $y(t)$ сигнала и помех в сообщение $y'(t)$, выдаваемое получателю. Следовательно, максимум пропускной способности C искался по всем возможным принципам временной обработки при заданных P_c , N_0 и Δf .

Если снять ограничение на величину полосы Δf , то, как следует из (4.18), максимум величины C получается при $\Delta f \rightarrow \infty$ и равен¹

$$C_{\infty} = P_c / N_0 \ln 2 \approx 1,44 P_c / N_0. \quad (4.33)$$

Этому соответствует минимально возможный удельный расход средней мощности сигнала

$$\beta_{E\infty} = P_c / N_0 C_{\infty} = \ln 2 \approx 0,69. \quad (4.34)$$

В общем случае, т.е. при произвольном значении удельного расхода $\beta_{\Delta f}$, потенциальные возможности СПИ при передаче дискретных сообщений характеризуются диаграммой обмена (4.19) и соответствующей ей сплошной кривой на рис. 4.2. Эта кривая соответствует непрерывному каналу, т.е. случаю $n_k \rightarrow \infty$. По мере уменьшения числа n_k видов символов, передаваемых системой, потенциальные возможности СПИ монотонно ухудшаются и при $n_k = 2$ характеризуются штриховой кривой на рис. 4.2.

При передаче непрерывных сообщений, как отмечалось, помимо ограничений P_c , N_0 и Δf необходимо учитывать также допустимое значение среднего квадрата δ^2 ошибки воспроизведения сообщений. При этом диаграмма обмена между удельными расходами β'_E и $\beta'_{\Delta f}$, мощности и полосы пропускания определяется формулой (4.32) и соответствующими ей кривыми, изображенными на рис. 4.3.

4.4. РЕАЛЬНАЯ ПРОПУСКНАЯ СПОСОБНОСТЬ СПИ

Рассмотрим определение реальной пропускной способности системы и соответственно реальных диаграмм обмена между удельными расходами мощности и полосы пропускания. При этом ради краткости ограничимся случаем передачи независимых равновероятных двоичных сообщений.

Пропускная способность реальной СПИ оказывается при этом меньше (хуже) найденного ранее потенциального значения (4.19) по целому ряду причин, из которых основными являются следующие.

1. Кроме аддитивного независимого нормального белого шума со спектральной плотностью N_0 могут присутствовать и другие помехи.

¹ Поскольку в (4.18), (4.19), (4.28), (4.32) — (4.34) входит потенциальное значение пропускной способности, величины C , β_E , $\beta_{\Delta f}$, β'_E , $\beta'_{\Delta f}$ следовало бы снабжать индексом 0. Однако для упрощения записи этот индекс здесь и далее не ставится.

2. Кроме ограничений средней мощности P_c и полосы пропускания Δf необходимо учитывать и другие ограничения, в первую очередь допустимого значения τ_a запаздывания в передаче информации и пиковой мощности сигнала. Кроме того, поскольку при ограниченном значении τ_a невозможно осуществить идеальное кодирование, обеспечивающее $P_{\text{ош}} \rightarrow 0$, вероятность ошибки $P_{\text{ош}}$ оказывается конечной и чем меньше допустимое значение этой вероятности, тем большей будет требуемая для ее обеспечения мощность сигнала. Наконец, в ряде реальных задач из-за ограничений массы, габаритных размеров, надежности и других показателей системы не только невозможно осуществить идеальное (по помехоустойчивости) кодирование (и декодирование), но вообще приходится отказаться от кодирования (если не включать в понятие «кодирование» — декодирование» обычную модуляцию — демодуляцию сигнала).

Рассмотрим влияние ряда указанных факторов на пропускную способность системы и требуемую мощность сигнала. При этом ради краткости и простоты положим, что единственным источником искажений сообщений в системе является действие аддитивного независимого гауссовского белого шума со спектральной плотностью N_0 .

Влияние ограничения пиковой мощности сигнала. Наличие кроме ограничения средней мощности сигнала ограничения и его пиковой мощности P_n может лишь ухудшить пропускную способность системы. Поэтому при

$$P_n \leq P_{\text{макс}} \quad (4.35)$$

имеет место неравенство

$$C \leq \Delta f \log_2 (1 + P_{\text{макс}}/N_0 \Delta f). \quad (4.36)$$

Кроме того, К. Шенноном были получены следующие результаты [47]:

$$а) \quad C \geq \Delta f \log_2 (1 + 2 P_{\text{макс}}/\pi^2 e^2 N_0 \Delta f); \quad (4.37)$$

б) если $P_{\text{макс}}/N_0 \Delta f \rightarrow 0$, то

$$C \rightarrow \Delta f \log_2 (1 + P_{\text{макс}}/N_0 \Delta f). \quad (4.38)$$

Из соотношений (4.36) — (4.38) следует, что в некоторых случаях [например, в случае б)] ограничение пиковой мощности значением $P_{\text{макс}}$ (в дополнение к ограничению средней мощности, таким же значением $P_{\text{макс}}$) не приводит к уменьшению потенциальной пропускной способности; однако в ряде случаев уменьшение пропускной способности может быть существенным.

Влияние неидеальности кодирования (и декодирования). Рассмотрим сначала случай, в котором кодирование вообще отсутствует. Пусть на вход канала от источника сообщений поступают независимые равновероятные двоичные символы, длительность каждого из которых равна τ_0 . В канале эти символы передаются

оптимальным способом — посредством фазовой манипуляции (на π) несущего синусоидального колебания в передатчике и оптимального поэлементного (посимвольного) приема в приемнике. Такой канал является двоичным симметричным каналом без памяти. Его пропускная способность (в присутствии аддитивного независимого нормального белого шума) определяется соотношением (4.12), а при $P_{\text{ош.к}} \leq 10^{-2}$ (что обычно имеет место) — формулой (4.14), т. е.

$$C \approx V_n \log_2 n_k = (1/\tau_0) \log_2 2 = 1/\tau_0. \quad (4.39)$$

Поскольку кодирование (декодирование) отсутствует, полная вероятность ошибки $P_{\text{ош}}$ совпадает с $P_{\text{ош.к}}$ и определяется (при $P_{\text{ош}} \leq 10^{-2}$) соотношением

$$Q/N_0 \approx \ln(1/P_{\text{ош}}) - 3 \approx 2,3 \lg(1/P_{\text{ош}}) - 3, \quad (4.40)$$

где $Q = P_c \tau_0$

— энергия сигнала, расходуемая на передачу одного символа. Полоса пропускания системы

$$\Delta f \approx 1/\tau_0 = C. \quad (4.41)$$

При этом из формул (4.40) и (4.41) получается

$$\beta_{\Delta f} = (\Delta f/C) = 1, \quad \beta_E = P_c/N_0 C = 2,3 \lg(1/P_{\text{ош}}) - 3. \quad (4.42)$$

При идеальном кодировании (включая использование непрерывного канала, т. е. при $n_k \rightarrow \infty$) справедливы соотношение (4.19) и соответствующая ему сплошная кривая на рис. 4.2. Из них получаются следующие результаты:

если $\beta_{\Delta f} = 1$, то $\beta_E = 1$;

если $\beta_{\Delta f} \rightarrow \infty$, то $\beta_E = \ln 2 = 0,69$. (4.43)

Из сравнения (4.42) с (4.43) следует, что при сделанных допущениях применение идеального кодирования (включая оптимальный непрерывный канал) дает следующие значения выигрыша γ в требуемой средней мощности сигнала (величине β_E):

а) при том же значении полосы ($\beta_{\Delta f} = 1$)

$$\gamma \approx 2,3 \lg(1/P_{\text{ош}}) - 3 \quad \text{при} \quad P_{\text{ош}} \leq 10^{-2}. \quad (4.44)$$

В частности, $\gamma = 1,6 \approx 2$ дБ, если $P_{\text{ош}} = 10^{-2}$ и $\gamma = 11 \approx 10,4$ дБ при $P_{\text{ош}} = 10^{-6}$;

б) при возможности значительного расширения полосы (в случае применения кодирования), т. е. при $\beta_{\Delta f} \geq 10$,

$$\gamma \approx 1,44 [2,3 \lg(1/P_{\text{ош}}) - 3] \quad (4.45)$$

и составляет примерно 3,5 дБ, если $P_{\text{ош}} = 10^{-2}$ и около 12 дБ при $P_{\text{ош}} = 10^{-6}$.

Отсюда следует, что выигрыш, даваемый применением идеального кодирования, возрастает по мере уменьшения допустимой вероятности $P_{\text{ош}}$ ошибки передачи сообщений и при $P_{\text{ош}} \geq 10^{-6}$ не превышает 10...12 дБ. При применении реальных устройств коди-

рования и декодирования из-за наличия ряда ограничений (конечности запаздывания τ_c и числа n_k видов передаваемых символов, ограничения сложности кодера и декодера и др.) выигрыш, даваемый применением кодирования, при $P_{\text{ош}} \geq 10^{-6}$ обычно не превосходит 5...7 дБ (если единственной помехой является указанный выше белый шум). Это означает, что при $P_{\text{ош}} \geq 10^{-6}$ к идеальному результату (4.34) можно реально приблизиться с точностью порядка 1...5 дБ (в зависимости от допустимой вероятности $P_{\text{ош}}$).

Рассмотрим теперь систему, в которой осуществляется блочное кодирование ортогональными сигналами. Пусть по-прежнему сообщение $x_n(t)$, поступающее на вход системы, представляет собой последовательность независимых равновероятных двоичных символов, каждый из которых имеет длительность τ_0 . В кодере накапливается блок из m таких символов и передается по каналу одним из M ортогональных сигналов длительностью

$$\tau_c = m \tau_0. \quad (4.46)$$

Очевидно, число возможных реализаций сообщения в блоке равно 2^m ; поэтому число ортогональных сигналов должно быть

$$M = 2^m. \quad (4.47)$$

При этом каждой возможной реализации сообщения в блоке соответствует свой ортогональный сигнал. Прием закодированных таким способом сообщений производится поблочно — в течение каждого интервала времени $\Delta t = \tau_c$ передачи блока в приемнике принимается решение о том, какой из M ортогональных сигналов был передан на данном интервале времени, и декодер выдает соответствующую этому ортогональному сигналу реализацию сообщения¹. Поэтому полная вероятность $P_{\text{ош}}$ ошибки воспроизведения кодовой комбинации сообщений совпадает с полной вероятностью ошибки распознавания в приемнике M ортогональных сигналов². Так как все значения сообщения приняты равновероятными, то и ортогональные сигналы будут равновероятны.

Пусть все ортогональные сигналы имеют одинаковую энергию $Q = P \tau_c$ и единственной помехой, действующей в канале, является аддитивный независимый гауссовский шум с неизменной в полосе канала спектральной плотностью N_0 . Тогда, как следует из (3.17), энергия сигнала, требуемая для обеспечения заданной вероятности $P_{\text{ош}}$ (при $P_{\text{ош}} \leq 10^{-2}$)

$$Q/N_0 \approx 2 [\ln(1/P_{\text{ош}}) + \ln(M-1)]. \quad (4.48)$$

¹ Предполагается, что в системе имеется идеальная синхронизация, т. е. моменты начала и окончания каждого блока известны в месте приема точно.

² Часто удобно характеризовать достоверность передачи сообщений также средней вероятностью P искажения одного (равновероятного) двоичного символа, т. е. одного бита передаваемой информации. Как показано в [48], $P = P_{\text{ош}} \cdot 2^m / (2^m - 1) \approx P_{\text{ош}} / 2$, т. е. в каждой передаваемой (ортогональным сигналом) кодовой комбинации искажается в среднем примерно 50% двоичных символов.

Известно (см., например, [49]), что максимально возможное число M ортогональных сигналов, которое можно разместить в полосе Δf на интервале времени τ_c ,

$$M = a \Delta f \tau_c, \quad (4.49)$$

где a — коэффициент, зависящий от того, что именно понимается под полосой пропускания Δf канала и длительностью τ_c ортогонального сигнала (т. е. каковы формы частотной характеристики канала и огибающей сигнала и на каком уровне затухания частотной характеристики и огибающей определяются соответственно полоса Δf и длительность сигнала τ_c). Приближенно можно полагать $a \approx 1...2$.

Из (4.49) следует, что для реализации описанного принципа передачи сообщений канал должен иметь полосу

$$\Delta f = M/a \tau_c. \quad (4.50)$$

Очевидно, что при передаче сообщений будет иметь место задержка на время $\tau = m \tau_0$ в передатчике [в связи с тем, что до передачи по каналу необходимо образовать («накопить») блок из m элементарных символов первичного сообщения] и на такое же значение в приемнике (поскольку в оптимальном приемнике решение о том, какой именно из M ортогональных сигналов присутствует во входной смеси сигнала с шумом, должно выноситься в момент окончания ожидаемого ортогонального сигнала, т. е. через интервал времени $\Delta t = \tau_c = m \tau_0$ после его посылки). Поэтому общая задержка передачи сообщений, вызванная необходимостью кодирования и декодирования сообщений,

$$\tau_s = 2m \tau_0 = 2 \tau_c. \quad (4.51)$$

Очевидно, описанный канал является дискретным M -ичным каналом ($n_k = M$) с ортогональными сигналами, так как на каждом интервале времени $\Delta t = \tau_c$ сигнал может иметь не любой вид, а лишь один из M возможных видов. В системе реализуется блочное кодирование сообщений, поскольку передаваемые первичные символы $x_n(t)$ в течение интервала времени $\Delta t = m \tau_0$ накапливаются и передаются в соответствии с выбранным кодом одним из M ортогональных сигналов. Определим пропускную способность такой системы передачи сообщений. Первичный источник вырабатывает информацию со средней скоростью $R_n = V_n H_n$, где $V_n = 1/\tau_0$ — скорость выработки этим источником двоичных символов, а H_n — энтропия источника, т. е. количество информации, передаваемой (в среднем) каждым символом. Поскольку эти символы двоичны, независимы и равновероятны, то

$$H_n = \log_2 2 = 1 \text{ бит/символ и } R_n = 1/\tau_0. \quad (4.52)$$

Кодер превращает двоичные символы в ортогональные сигналы. Так как можно полагать, что он не вносит искажений в передаваемую информацию, вся поступающая в него информация с та-

кой же скоростью (но с задержкой $m\tau_0$) поступает на вход канала. Поэтому скорость передачи информации в канал

$$R = R_n = 1/\tau_0. \quad (4.53)$$

С учетом (4.46) и (4.47)

$$R = (\log_2 M)/\tau_0.$$

Поэтому в отсутствие помех пропускная способность канала была бы

$$C = R = (\log_2 M)/\tau_0. \quad (4.54)$$

При действии шума пропускная способность канала будет несколько меньше. Однако, как отмечалось в § 4.2, при $P_{\text{ош}} \leq 10^{-2}$ уменьшение пропускной способности из-за действия помех незначительно. Поэтому при $P_{\text{ош}} \leq 10^{-2}$ можно считать соотношение (4.54) приближенно справедливым и при наличии помех. Найдем теперь соотношение между удельными расходами β_E и $\beta_{\Delta f}$ для описанного канала. Так как по определению $\beta_{\Delta f} = \Delta f/C$, то из (4.50) и (4.54) получаем

$$\beta_{\Delta f} = M/a \log_2 M. \quad (4.55)$$

Удельный расход средней мощности

$$\beta_E = P_0/N_0 C. \quad (4.56)$$

Поэтому, принимая во внимание (4.46) и (4.54), имеем $\beta_E = Q/N_0 \log_2 M$, откуда с учетом (4.47) следует

$$\beta_E = (\ln 2) \left[\frac{\lg(M-1)}{\lg M} + \frac{2 \lg(1/P_{\text{ош}})}{\lg M} \right]. \quad (4.57)$$

В дальнейшем примем для конкретности

$$M \geq 8, \quad a = 1. \quad (4.58)$$

Тогда получим следующие основные соотношения:

$$\beta_{\Delta f} = M/\log_2 M = 0,3 M/\lg M, \quad (4.59)$$

$$\beta_E = (\ln 2) [1 + 2 \lg(1/P_{\text{ош}})/\lg M], \quad (4.60)$$

$$\Delta f = \beta_{\Delta f} C \approx 0,3 MC/\lg M, \quad (4.61)$$

$$\tau_0 = 2\tau_0 \log_2 M; \quad (4.62)$$

$$M = 2^m, \text{ т. е. } m = \log_2 M \approx 3,3 \lg M. \quad (4.63)$$

Анализ этих соотношений позволяет сделать следующие заключения. Если выполняется условие

$$\lg M \gg 2 \lg(1/P_{\text{ош}}), \quad (4.64)$$

то

$$\beta_E = \ln 2 = 0,69 = \beta_{E \text{ мин}}, \quad (4.65)$$

где $\beta_{E \text{ мин}}$ — наименьший теоретический возможный удельный расход средней мощности сигнала, вытекающий из основной формулы (4.18) Шеннона (при $\Delta f \rightarrow \infty$).

Следовательно, при выполнении условия (4.64) блочное кодирование ортогональными сигналами обеспечивает столь же малый удельный расход энергии, как идеальное кодирование. Этот результат может быть достигнут при сколь угодно малой (но конечной) вероятности ошибки $P_{\text{ош}}$, если число M ортогональных сигналов неограниченно возрастает. Но при этом, как следует из (4.61) и (4.62), также неограниченно возрастают требуемая полоса пропускания Δf и запаздывание τ_0 . Кроме того, из-за увеличения M неограниченно возрастает и сложность устройств кодирования и декодирования сообщений. Поэтому представляет интерес выяснить, насколько блочное кодирование ортогональными сигналами приближается по результатам к идеальному кодированию при конечном, но достаточно малом значении допустимой вероятности ошибки. Например, $P_{\text{ош}} = 10^{-6}$. Тогда из (4.60) получаем

$$\beta_E = 0,69 (1 + 12/\lg M). \quad (4.65')$$

На основании (4.59) и (4.65') на рис. 4.4 построена зависимость β_E от $\beta_{\Delta f}$ в виде последовательности точек. Дискретный характер этой зависимости обусловлен тем, что в соответствии с (4.63) величина M может иметь лишь дискретные значения 2, 4, 8, 32, 64, 128, ... [при этом следует иметь в виду, что зависимость (4.60), а следовательно, и (4.65) достаточно точны лишь при $M \geq 8$]. Для сравнения на рис. 4.4 штриховой линией построена зависимость β_E от $\beta_{\Delta f}$, соответствующая непрерывному каналу с идеальным кодированием [т. е. соотношению (4.19)].

Как видно из рис. 4.4, энергетический проигрыш $\gamma = \beta_E/\beta_{E \text{ мин}}$ данного способа кодирования по сравнению с идеальным уменьшается от 11 дБ при $M=8$ до 5,5 дБ при $M=32768$ (т. е. при $m=15$). Однако рассматривать систему со столь большим числом ортогональных сигналов в настоящее время не представляется возможным. Кроме того, как следует из (4.61) и (4.62), предел увеличе-

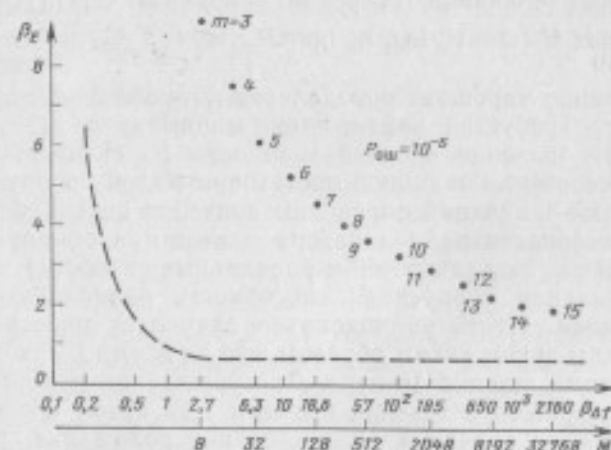


Рис. 4.4

нию M может ставить чрезмерное расширение полосы (при большой пропускной способности C) и/или запаздывания τ_z . Из рис. 4.4 следует также, что рассмотренная система кодирования не позволит получить малые значения удельного расхода полосы $\beta_{\Delta f}$, даже если пойти на существенное увеличение удельного расхода мощности.

В настоящее время существует большое число различных видов кодирования сообщений, включая и такие, которые допускают получение малых значений $\beta_{\Delta f}$. Однако при всех разумных видах кодирования (и $P_{\text{ош}} \leq 10^{-2}$) соответствующая зависимость $\beta_{\Delta f}$ от $P_{\text{ош}}$ (диаграмма обмена) расположена правее и выше левой нижней границы Шеннона (штриховой кривой на рис. 4.4) и, как уже указывалось, энергетический проигрыш по сравнению с этой границей составляет примерно 1...5 дБ (в зависимости от допустимого значения полной вероятности ошибки $P_{\text{ош}}$).

4.5. ПОТЕНЦИАЛЬНАЯ ПРОПУСКНАЯ СПОСОБНОСТЬ СИСТЕМ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ

В СИИ невозможно идеальное кодирование сообщений источника. Соответственно невозможно (при конечном значении $\beta_{\Delta f}$) и извлечение информации со сколь угодно малой вероятностью $P_{\text{ош}}$ ошибки даже при дискретном источнике информации. Поэтому потенциальная пропускная способность СИИ определяется при условии, что вероятность ошибки конечна, но достаточно мала (например, $P_{\text{ош}} \leq 10^{-2}$).

Так как изображенные на рис. 4.1 кодер и декодер в данном случае отсутствуют, то вероятность ошибки на выходе канала и на выходе системы совпадают и, следовательно, малое значение $P_{\text{ош}}$ должно быть обеспечено в самом канале. При этом (т. е. в случае извлечения дискретной информации) в формуле (4.10) можно пренебречь величиной $H(x|y)$ по сравнению с $H(x)$ и полагать

$$C = V_k \max_{(x)} H(x) = V_k \log_2 n_k \text{ при } P_{\text{ош}} \leq P_{\text{ош, доп}}, \quad (4.66)$$

что существенно упрощает определение потенциальной пропускной способности. Требуется найти лишь минимальное (теоретически достижимое) значение средней мощности P_c сигнала, необходимой для обеспечения заданной достаточно малой вероятности $P_{\text{ош}}$. С учетом этих замечаний определим значения потенциальной пропускной способности радиолокатора в режимах обнаружения целей и измерения их дальностей и радиальных скоростей.

Потенциальная пропускная способность радиолокатора обнаружения целей. Пусть радиолокатор сканирует пространство по азимуту узким лучом таким образом, что за время T луч смещается на угол, равный ширине $\Delta\theta$ луча. Для упрощения можно полагать, что луч неподвижен в течение времени T , а затем скачкообразно перемещается на значение $\Delta\theta$ в соседнее положение, в котором также находится в течение времени T , и т. д. Пусть все обнаруживаемые цели находятся на примерно одинаковой дальности от

локатора и имеют приблизительно одинаковые эффективные поверхности рассеяния. Предположим далее, что в каждом секторе обзора шириной $\Delta\theta$ может появляться только одна цель и притом независимо от того, имеются или нет цели в других секторах обзора. Решение о наличии или отсутствии цели в данном секторе должно приниматься в течение времени T . Единственной помехой является аддитивный независимый гауссовский белый шум со спектральной плотностью N_0 .

При сделанных допущениях процесс обнаружения состоит из последовательности элементарных (частных) сеансов обнаружения, длительность каждого из которых равна T . В течение каждого сеанса решается задача о наличии или отсутствии сигнала в данном секторе обзора. Поэтому в формуле (4.66) следует полагать $n_k = 2$ (решается задача бинарного обнаружения), а $V_k = 1/T$, так как роль двоичного символа играет сигнал (нулевой или ненулевой), имеющий в каждом последовательном сеансе длительность T . Поэтому формула (4.66) принимает вид

$$C = (\log_2 2)/T = 1/T \text{ [бит/с] при } P_{\text{ош}} \leq P_{\text{ош, доп}}. \quad (4.67)$$

Определим теперь пропускную способность, выраженную не в битах в секунду, а в целях в секунду (ц/с).

В данной задаче на обнаружение одной цели тратится время, равное T (через время T луч радиолокатора переходит в соседнее положение, где может находиться другая цель). Поэтому максимальное число целей, которые такой локатор может обнаружить в единицу времени,

$$C' = 1/T \text{ [ц/с] при } P_{\text{ош}} \leq P_{\text{ош, доп}}. \quad (4.68)$$

Из сравнения (4.67) и (4.68) следует, что в данной задаче C' и C выражаются одним и тем же соотношением. Это объясняется тем, что максимальное количество информации, извлекаемой при обнаружении каждой (одной) цели, равно одной двоичной единице (одному биту).

Соотношение (4.67) [равно как и (4.68)] справедливо как для реального, так и для идеального локатора. Поэтому для нахождения потенциального (наибольшего) значения пропускной способности C (или C') требуется найти минимальное (по всем возможным принципам построения локатора) значение времени T , затрачиваемого на обнаружение сигнала с заданными вероятностями ошибок, в данном случае вероятностями $P_{\text{пр}}$ и $P_{\text{лт}}$ пропуска сигнала и ложной тревоги.

Будем искать минимум T при условии, что отраженный от цели сигнал принимается на фоне аддитивного независимого нормального белого шума со спектральной плотностью N_0 , а средняя (за время T) мощность сигнала равна P_c . При этом значение T может быть минимизировано лишь за счет выбора соответствующей формы зондирующего сигнала и способа его обработки в при-

емнике¹. Как следует из [5], в рассматриваемой задаче бинарного обнаружения наилучшие результаты достигаются, если в месте приема эхо-сигнал $u_c(t)$ может считаться известным точно (неизвестен лишь факт его наличия или отсутствия). При этом форма сигнала не влияет на качество обнаружения и при оптимальной обработке получается [5] (при $P_{пр} \leq 10^{-2}$, $P_{лт} \leq 10^{-2}$)

$$P_c T/N_0 = (\sqrt{\ln 1/P_{лт}} + \sqrt{\ln 1/P_{пр}})^2, \quad (4.69)$$

где $P_{лт}$ и $P_{пр}$ — вероятности ложной тревоги и пропуска сигнала. Соотношение (4.69) определяет минимальное значение времени T при фиксированных значениях вероятностей ошибок ($P_{лт}$ и $P_{пр}$) и характеристик канала (системы) P_c и N_0 . Поэтому в соответствии с (4.68) и (4.69) потенциальная пропускная способность локатора в рассматриваемом режиме

$$C' = P_c/N_0 (\sqrt{\ln 1/P_{лт}} + \sqrt{\ln 1/P_{пр}})^2 \text{ [ц/с]}. \quad (4.70)$$

Например, при $P_{пр} = P_{лт} = 10^{-2}$ $C' = P_c/18,4 N_0$ [ц/с]. При этом минимальная требуемая полоса пропускания локатора

$$\Delta f \approx 1/T. \quad (4.71)$$

Из (4.69), (4.70) и (4.71) следует, что удельный расход мощности

$$\beta_E = P_c/N_0 C' \quad (4.72)$$

равен

$$\beta_E = (\sqrt{\ln 1/P_{лт}} + \sqrt{\ln 1/P_{пр}})^2, \quad (4.73)$$

а удельный расход полосы

$$\beta_{\Delta f} = \Delta f/C' \quad (4.74)$$

равен единице:

$$\beta_{\Delta f} = 1. \quad (4.75)$$

Потенциальная пропускная способность радиолокатора при измерении дальностей и радиальных скоростей целей. Пусть луч локатора шириной $\Delta\theta$ перемещается в пространстве по тому же закону, что и в предыдущем случае, т. е. находится в каждом угловом секторе в течение интервала времени, равного T . Однако в каждом таком положении луча требуется не обнаруживать цель (полагается, что она уже обнаружена), а определять дальность r до цели (или ее радиальную скорость V_r). При этом положим, что цель (единственная в данном угловом секторе) может находиться на любой дальности в пределах диапазона

$$\Delta r = r_{\max} - r_{\min} \quad (4.76)$$

причем $\Delta r/r_{\min} \ll 1$, так что приближенно можно считать амплитуду сигнала на входе приемника не зависящей от дальности.

¹ Как и при выводе формулы Шеннона (4.18), полагается, что вид пространственной обработки сигналов задан и оптимум ищется лишь вариацией временной обработки. Кроме того, не рассматривается случай последовательного обнаружения.

Пусть $u_{ci}(t)$ — эхо-сигнал на входе приемника локатора, порожденный отражением от цели, находящейся на i -м интервале (в i -й ячейке) дальности ($i = \overline{1, M}$) шириной $\delta r = \Delta r/M$. Тогда, если в приемнике будет определяться, какой из сигналов $u_{c1}(t), \dots, u_{cM}(t)$ присутствует на его входе (на данном интервале времени $[0, T]$ измерения), то тем самым будет определяться номер ячейки дальности, в которой находится цель¹. Если при этом полагать дальность до цели соответствующей середине данной ячейки, то дальность до цели будет определяться с ошибкой, не превышающей $\delta r/2 = \Delta r/2M$, точнее, вероятность того, что ошибка превысит $\delta r/2$, будет равна полной вероятности $P_{ош}$ ошибки распознавания сигналов $u_{ci}(t)$ ($i = \overline{1, M}$), которая должна быть обеспечена достаточно малой (например, $P_{ош} \leq 10^{-2}$).

Итак, для измерения дальности до цели с ошибкой, не превышающей [с большой вероятностью $(1 - P_{ош})$] $\delta r/2 = \Delta r/2M$, следует обеспечить распознавание того, какой именно из M возможных эхо-сигналов $u_{c1}(t), \dots, u_{cM}(t)$ присутствует на входе приемника (в течение данного сеанса измерения $[0, T]$). При этом полная вероятность $P_{ош}$ ошибки распознавания должна быть достаточно малой. Так как на измерение дальности до каждой из целей, расположенных в различных угловых секторах обзора, тратится время, равное T , то пропускная способность системы, измеряемая в целях в секунду,

$$C' = 1/T \text{ при } P_{ош} \leq P_{ош, доп}. \quad (4.77)$$

Найдем теперь пропускную способность системы C в битах в секунду в предположении, что $P_{ош} \leq 10^{-2}$. При этом можно приближенно полагать, что пропускная способность системы при наличии помех остается такой же, как в их отсутствии, и достигается в том случае, когда значения дальностей до целей (в различных секторах обзора по углу) независимы и равновероятны. Тогда в соответствии с (4.66) пропускная способность

$$C = (1/T) \log_2 M. \quad (4.78)$$

Из сравнения (4.77) и (4.78) следует, что в данном случае значения C' и C различаются не только по размерности, но и по виду определяющей формулы. Из (4.77) и (4.78) следует, что связь между C' и C можно представить в виде

$$C = C' H_{\max}, \quad (4.79)$$

где $H_{\max} = \log_2 M$ — максимальное значение энтропии H источника информации.

Очевидно, зависимость (4.79) обусловлена тем фактором, что среднее количество информации, извлеченной за некоторое время, равно числу «обслуженных» (т. е. обеспеченных измерением

¹ Предполагается, что за время измерения T цель не может переместиться на расстояние большее чем $\delta r/2$, т. е. должно выполняться условие $V_r T < \delta r/2 = \Delta r/2M$, где V_r — радиальная скорость цели.

дальности) целей, умноженному на среднее количество информации, извлекаемой при измерении дальности до одной цели.

Из сравнения соотношений (4.77) и (4.78) вытекают следующие заключения.

1. В рассмотренном примере допустимо определение пропускной способности как в целях в секунду, так и в битах в секунду.

2. Так как при сравнении вариантов построения системы (локатора) величина M , характеризующая точность измерения, задана, т. е. неизменна, то варианту с наибольшим значением C' будет соответствовать максимальное значение и величины C . Поэтому при отыскании оптимального варианта построения системы безразлично, какую из этих мер пропускной способности применять.

Из (4.77) и (4.78) следует, что для отыскания потенциально-го значения пропускной способности C' (или C) требуется найти минимальное (по всем вариантам построения локатора) время T , затрачиваемое на распознавание того, какой из M возможных сигналов $u_{c1}(t), \dots, u_{cM}(t)$ присутствует на входе локатора в течение интервала $[0, T]$ с заданной достаточно малой полной вероятностью $P_{\text{ош}}$ ошибки.

Будем при этом полагать, что средняя (за время T) мощность P_{ci} каждого i -го сигнала ($i=1, M$) задана, а его пиковая мощность не ограничивается; единственной помехой является аддитивный независимый нормальный белый шум со спектральной плотностью N_0 . С математической точки зрения удобнее искать сначала не минимум времени T при заданных значениях P_{ci} ($i=1, M$), N_0 и $P_{\text{ош}}$, а минимум значений $Q_i = P_{ci}T$ ($i=1, M$) требуемых энергий сигналов при заданных значениях T , N_0 и $P_{\text{ош}}$. При этом, как отмечалось, все сигналы полагаются независимыми и равновероятными. При таких условиях весьма близки к оптимальным (при $M \geq 8$) ортогональные сигналы с равными энергиями и требуемая энергия с достаточной точностью (при $P_{\text{ош}} \leq 10^{-2}$, $M \geq 8$) определяется соотношением

$$P_c T / N_0 = 2 \ln 1/P_{\text{ош}} + \ln M. \quad (4.80)$$

Это соотношение показывает, какое минимальное значение P_c можно получить при фиксированных значениях показателей T и $P_{\text{ош}}$. Но из (4.80) следует, что зависимость каждого из показателей, которые можно рассматривать как стандартные показатели качества системы [58], от значений любого из остальных трех показателей является монотонно убывающей. Поэтому найденное соотношение является по-существу диаграммой обмена и дает зависимость минимального значения каждого из показателей P_c , T , $1/N_0$ и $P_{\text{ош}}$ от значений остальных трех показателей. В частности, из (4.80) следует, что минимальное значение времени T определяется соотношением

$$T = (N_0/P_c) (2 \ln 1/P_{\text{ош}} + \ln M). \quad (4.81)$$

При этом с учетом (4.77) и (4.78) потенциальная пропускная способность C' и C определяется по следующим формулам:

$$C' = \frac{1}{T} = \frac{P_c/N_0}{2 \ln 1/P_{\text{ош}} + \ln M} \text{ [ц/с]}, \quad (4.82)$$

$$C = \frac{1}{T} \log_2 M = \frac{(P_c/N_0) \log_2 M}{2 \ln 1/P_{\text{ош}} + \ln M} = \frac{P_c/N_0}{(\ln 2) [1 + 2(\lg 1/P_{\text{ош}})/\lg M]} \text{ [бит/с]}. \quad (4.83)$$

При $M \rightarrow \infty$ получается

$$C \rightarrow (P_c/N_0)/\ln 2 \text{ [бит/с]}, \quad (4.84)$$

$$C' \rightarrow (P_c/Q_0)/\ln M \text{ [ц/с]} \quad (4.85)$$

при сколь угодно малой (но конечной) вероятности $P_{\text{ош}}$ ошибки.

Для образования M ортогональных сигналов на интервале времени $[0, T]$ требуется, чтобы ширина спектра сигналов, а следовательно, и полоса пропускания системы Δf удовлетворяли условию [49] $\Delta f T \geq M$. Поэтому минимальное значение требуемой полосы пропускания

$$\Delta f = M/T. \quad (4.86)$$

Следовательно, потенциальные удельные расходы средней мощности сигнала и полосы равны:

$$\beta_E = P_c/N_0 C = (\ln 2) [1 + 2(\lg 1/P_{\text{ош}})/\lg M], \quad (4.87)$$

$$\beta_{\Delta f} = \Delta f/C = M/\log_2 M. \quad (4.88)$$

При $M \rightarrow \infty$

$$\beta_E \rightarrow \ln 2, \beta_{\Delta f} \rightarrow \infty \quad (4.89)$$

при сколь угодно малой $P_{\text{ош}}$.

Нетрудно убедиться, что соотношение (4.89) совпадает по форме с формулой Шеннона (4.18), если в ней положить $\Delta f \rightarrow \infty$. При этом для СПИ, рассмотренной К. Шенноном, P_c — средняя мощность сигнала, полученная усреднением за достаточно длительное время передачи информации. В рассмотренном случае радиолокации под P_c понимается мощность одного из ортогональных сигналов, усредненная за время его действия T . Однако, поскольку все ортогональные сигналы при этом имеют одинаковую среднюю мощность, мощность сигнала, полученная усреднением по любому числу сеансов измерения дальности, также будет равна P_c . Поэтому можно считать, что параметр P_c в формулах (4.18) и (4.82) — (4.86) имеет один и тот же смысл.

На практике обеспечить весьма большое значение M невозможно не только из-за усложнения системы, но и из-за резкого возрастания требуемой полосы пропускания Δf . Действительно, в соответствии с условием, отмеченным в сноске к стр. 113, должно быть $T < \Delta r/2V_r M$. Поэтому из (4.86) получается

$$\Delta f > 2V_r M^2/\Delta r.$$

Ранее была рассмотрена система измерения дальности до цели. Но, очевидно, такие же результаты получатся, если вместо дальности r измеряется радиальная скорость цели V_r путем образования M_V ортогональных эхо-сигналов, соответствующих M_V значениям скорости в некотором интервале ΔV_r , так что $M_V = \Delta V_r / \delta V_r$, где δV_r — элементарный интервал значений скорости, которому соответствует элементарный интервал δF_d значений доплеровского сдвига F_d частоты сигнала, отраженного от цели. Ортогональность сигналов, соответствующих различным элементарным интервалам значений скорости, может быть приближено обеспечена соответствующим выбором формы зондирующего сигнала.

Если локатор должен в течение интервала $[0, T]$ одновременно измерять дальность и радиальную скорость цели с точностями, требующими образования M_r ячеек по дальности, а для каждой ячейки дальности — M_V ячеек по радиальной скорости, то общее число требуемых ортогональных сигналов

$$M = M_r M_V \quad (4.90)$$

и формулы (4.77), (4.78), (4.82) — (4.86) для потенциальной пропускной способности остаются справедливыми, если под M понимать величину (4.90).

* * *

В сжатом виде основные результаты изложенного в данной главе сводятся к следующему:

1. Потенциальная пропускная способность СПИ в случае передачи дискретных сообщений имеет наибольшее значение при непрерывном канале передачи сообщений. Если канал дискретный, то она уменьшается по мере уменьшения числа n_k видов символов, передаваемых каналом. Однако даже при $n_k = 2$ проигрыш в величине $P_c/N_0 C$ получается сравнительно небольшим (в $\pi/2 \dots 2$ раза), если удельная полоса пропускная не слишком мала ($\Delta f/C \geq 0,6$). Для увеличения потенциальной пропускной способности системы необходимо увеличивать отношение P_c/N_0 сигнала к шуму и полосу пропускания Δf системы. Однако при $\Delta f > C$ дальнейшее увеличение полосы уже не дает существенного увеличения потенциальной пропускной способности. Благодаря наличию идеального кодирования-декодирования потенциальная пропускная способность реализуется со сколь угодно малой вероятностью ошибки передачи дискретных сообщений.

Передача непрерывных сообщений со скоростью, равной потенциальной пропускной способности C системы, возможна лишь с некоторым относительным средним квадратом ошибки δ^2 . При нормальном законе распределения и равномерном спектре сообщения, ограниченном частотой F_h , значение δ^2 определяется соотношением (4.26). Если потенциальная пропускная способность C определяется основной формулой Шеннона (4.18), то соотношение (4.26)

приводит к диаграмме обмена (4.32) между удельным расходом средней мощности сигнала $\beta' = P_c/N_0 F_h$ и удельным расходом полосы $\beta'_{\Delta f} = \Delta f/F_h$. Из соответствующих ей графиков на рис. 4.3 следует, что чем меньше допустимое значение δ^2 , тем больше требуемые удельные расходы средней мощности сигнала и полосы пропускания системы.

2. Реальная пропускная способность СПИ меньше ее потенциального значения за счет ряда факторов, из которых основными являются следующие:

недопустимость сколь угодно больших значений задержки t_z , полосы Δf и числа n_k видов символов;

наличие кроме аддитивного независимого нормального белого шума со спектральной плотностью N_0 еще и других помех и источников ошибок передачи сообщений;

неидеальность помехоустойчивой обработки информации в передатчике и приемнике даже при конечном времени задержки (в частности, наличие потерь, вызванных необходимостью разделения каналов в многоканальных СПИ);

ограниченность не только средней, но и пиковой мощности сигнала.

Поэтому основные меры приближения реальной пропускной способности к ее потенциальному значению должны быть направлены на возможное ослабление указанных факторов. При принятии соответствующих мер, как показывает опыт проектирования реальных СПИ, удельный расход средней мощности сигнала β_k (или β'_k) может превышать свое потенциальное значение не более чем на 5...6 дБ, если допустим достаточно большой удельный расход полосы $\beta_{\Delta f}$ (или $\beta'_{\Delta f}$).

3. В системах извлечения информации потенциальная пропускная способность определяется при не сколь угодно малой, а конечной (но допустимо малой) вероятности ошибки (ошибок) и удельный (на единицу пропускной способности) расход средней мощности сигнала β_k тем больше (при данном удельном расходе полосы $\beta_{\Delta f}$), чем меньше допустимое значение этой вероятности.

Потенциальная пропускная способность радиолокатора в режиме обнаружения целей определяется соотношениями (4.67) — (4.69), а потенциальные удельные расходы мощности и полосы пропускания — формулами (4.71) и (4.75). Для реализации потенциальных значений (4.67), (4.69) форма сигнала не играет роли, но обработка принимаемого сигнала должна быть оптимальной. Если единственным источником искажения сообщений является аддитивный независимый нормальный белый шум, то соответствующие соотношения для реального локатора могут быть весьма близкими к потенциальным.

В режиме измерения дальностей (или радиальных скоростей) целей потенциальные соотношения описываются формулами (4.82) — (4.86). При этом с ростом требуемого числа ячеек по дальности (или по скорости) пропускная способность, выражен-

ная в целях в секунду, уменьшается, а выраженная в битах в секунду — несколько возрастает. Это обусловлено тем обстоятельством, что с увеличением M количество информации в битах в секунду, извлекаемой в каждом акте измерения дальности до цели, возрастает пропорционально $\log_2 M$. При неограниченном увеличении M удельный (на единицу пропускной способности, выраженной в битах в секунду) расход средней мощности сигнала уменьшается, стремясь к предельному значению, равному $\ln 2$. Однако при увеличении числа ячеек M система усложняется, а требуемая полоса пропускания возрастает пропорционально M^2 .

Для приближения реальной пропускной способности локатора к потенциальной (в режиме измерения дальностей и радиальных скоростей) требуется выполнение ряда условий, в первую очередь следующих:

не должно быть существенных помех воспроизведению сообщений, за исключением аддитивного независимого нормального белого шума с известной спектральной плотностью N_0 ;

зондирующие сигналы должны иметь такую форму, чтобы эхосигналы, соответствующие M различным ячейкам дальности (или радиальной скорости), можно было считать близкими к ортогональным. Кроме того, требуется, чтобы ограничениями пиковой мощности сигнала в передатчике можно было пренебречь при заданных ограничениях его средней мощности;

обработка эхо-сигналов в приемнике должна быть близка к оптимальной.

В ряде практических задач эти условия удается приближенно выполнить и значения пропускной способности и удельных расходов средней мощности оказываются близкими к потенциальным.

Литература: *основная* [47, 48], *дополнительная* [49].

Глава 5

ЭЛЕКТРОМАГНИТНАЯ И ЭКОЛОГИЧЕСКАЯ СОВМЕСТИМОСТЬ

5.1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОБЛЕМЫ ЭЛЕКТРОМАГНИТНОЙ СОВМЕСТИМОСТИ

Под электромагнитной совместимостью (ЭМС) радиосредств понимают их способность совместно функционировать в реальных условиях эксплуатации с требуемыми (допустимыми) значениями показателей качества при воздействии всевозможных непреднамеренных помех и не создавать помехи недопустимого уровня другим радиосредствам. Очевидно, проблема обеспечения ЭМС имеет много общего с проблемой помехозащищенности. Действительно, как для обеспечения ЭМС, так и для обеспечения помехоза-

щищенности требуется уменьшить до допустимого уровня действие различных видов помех. Однако между этими проблемами имеются и весьма существенные различия, в первую очередь следующие.

1. При рассмотрении проблемы ЭМС обычно не учитывается действие преднамеренных (организованных) помех. В первую очередь анализируется действие помех от радиостанций и технических средств, расположенных как вне системы, так и внутри ее. Внутренний шум, флуктуации сигнала и атмосферные помехи принимаются во внимание лишь в той мере, в какой это необходимо для правильной оценки дополнительного действия указанных выше помех.

В отличие от этого, при рассмотрении проблемы помехозащищенности обязательно учитывается действие организованных помех (если имеется существенная вероятность наличия таких помех). Действие же непреднамеренных помех от радиостанций и других технических средств обычно не рассматривается. Следовательно, первое различие между проблемами ЭМС и помехозащищенности состоит в различном (хотя и перекрывающемся) составе учитываемых помех.

2. Проблема обеспечения ЭМС состоит не только в защите данной РС от действующих на нее помех, но и в недопущении слишком большого мешающего действия данной РС на другие РС.

3. Из сказанного следует, что для обеспечения ЭМС можно принимать соответствующие меры (технические и организационные) и в самих источниках помех, тогда как возможность таких мер для обеспечения помехозащищенности во многих случаях (в частности, при преднамеренных помехах) отсутствует.

4. Многие источники помех, учитываемые при определении ЭМС, могут быть расположены столь близко к рецепторам (устройствам, на которые они действуют), что для правильной оценки действия таких помех приходится учитывать кроме основных и нежелательные (побочные и внеполосные) излучения радиопередающих устройств, побочные каналы приема радиоприемных устройств и другие каналы проникновения помех от их источников к рецепторам.

Из изложенного следует, что проблема обеспечения ЭМС, перекрываясь в некоторой мере с проблемой помехозащищенности, имеет ряд весьма существенных отличий в постановке задачи и соответственно в путях ее решения.

По мере развития человеческого общества число и интенсивность помех, влияющих на ЭМС, быстро возрастает. Особенно остро проблема обеспечения ЭМС встает тогда, когда на небольшой площади или в небольшом объеме (например, на борту корабля, самолета, искусственного спутника Земли, на территории аэродрома) приходится размещать десятки радиостанций и других технических средств, которые могут создавать помехи. Однако часто приходится учитывать помехи и между радиостанциями,

установленными на больших расстояниях друг от друга, даже находящимися в разных странах и в космосе. Поэтому проблема обеспечения ЭМС является, с одной стороны, весьма актуальной, а с другой — чрезвычайно сложной, требующей для своего решения совместных усилий инженеров не только в пределах одной страны, но и в пределах всего мира.

Материал, изложенный в данной главе, можно рассматривать лишь как весьма краткое введение в современную теорию ЭМС, при этом основная часть приведенного материала базируется на книгах [32, 51, 52].

В настоящее время различают следующие виды помех, учитываемых при определении ЭМС.

1. Межсистемные электромагнитные помехи (МЭМП):

а) первого вида — помехи между системами различных классов, например между спутниковыми системами связи и радиорелейными линиями связи;

б) второго вида — помехи между системами одного и того же класса, например помехи между радиолокационными системами одного и того же типа (класса), но расположенными на различных самолетах, совершающих совместный полет (например, в составе авиаэскадры или авиаотряда).

2. Внутрисистемные электромагнитные помехи (ВЭМП):

а) первого вида — помехи между различными радиоприемными одной и той же системы (здесь и далее радиоприемными — это совокупность передатчика и принимающего его сигнала приемника или только приемника в случае пассивной радиосистемы). Например, в спутниковой системе связи (ССС) через один и тот же ретранслятор, установленный на спутнике, передаются сигналы различных радиостанций и нужно обеспечить отсутствие взаимных помех между радиостанциями;

б) второго вида — помехи внутри данного радиопередающего или радиоприемного устройства, например помехи усилителям от источников питания, синтезаторов частоты, разверток электронно-лучевых трубок и т. д.

5.2. ПОМЕХИ ОТ РАДИОСТАНЦИЙ

Радиостанции создают помехи за счет излучения их передающих устройств, а также гетеродинов приемников. Излучения радиопередающих устройств делятся на основные (т. е. обеспечивающие выполнение радиопередающими устройствами их полезных функций) и нежелательные. Нежелательные излучения, в свою очередь, подразделяют на внеполосные и побочные.

К внеполосным относят излучения в полосах частот, примыкающих к необходимой полосе¹, обусловленной модуляцией (ма-

нипуляцией) сигнала. Обычно необходимая полоса меньше занимаемой ширины полосы частот радионизлучения (ШПИ), т. е. полосы, в которой сосредоточено 99% средней мощности всего излучения. Это объясняется тем, что передатчики излучают за пределами необходимой полосы обычно более 1% средней мощности. Для уменьшения необходимой полосы Международный консультативный комитет по радиосвязи (МККР) рекомендует:

уменьшать девиацию частоты для ЧМ излучений;

использовать, если возможно, режим одной боковой полосы (ОБП);

при импульсном излучении обеспечивать скругление фронтов импульсов (для прямоугольного импульса ШПИ в несколько раз больше необходимой полосы);

по возможности уменьшать верхнюю частоту спектра передаваемых сообщений.

К внеполосным относят также шумовые излучения, создаваемые паразитной модуляцией несущего колебания напряжением шума передатчика. Эти излучения имеют ширину спектра, в десятки или сотни раз превышающую ШПИ. Уровень шумового излучения обычно на 60...80 дБ ниже уровня полезного излучения, но он может на десятки децибел превышать уровень внутреннего шума расположенных близко радиоприемных устройств.

Природа шума в передатчиках та же, что и в приемниках (дробовой шум, наведенный шум, шум токораспределения и тепловой). Но в передатчиках часто приходится применять электронные приборы с большим уровнем шума (например, радиувеличения генерируемой мощности или КПД). Следует также учитывать увеличение шума при применении умножителей частоты.

К побочным относят излучения на гармониках и субгармониках (они характерны для передатчиков с умножителями частоты), комбинационные излучения (в случаях, когда колебания основного излучения формируются путем нелинейных преобразований вспомогательных колебаний), паразитные колебания (т. е. побочные колебания, не связанные с образованием несущих колебаний — например, вызванные наличием паразитных колебательных контуров), а также интермодуляционные излучения (побочные излучения, возникающие в данном передатчике при воздействии на него излучений других передатчиков, т. е. обусловленные взаимной модуляцией передатчиков).

Излучения гетеродинов приемников могут воздействовать через антенну или через металлическое шасси. Помехи от гетеродинов могут воздействовать на соседние приемники также через цепи питания и коммутации. Эти помехи могут быть существенными, даже если между антенной и смесительным каскадом включаются каскады усиления радиочастоты и фильтры. В некоторых случаях приходится учитывать нежелательные излучения (внеполосные и побочные) и излучения гетеродинов не только по основному лепестку диаграммы направленности радиопередающего устройства, но и по боковым лепесткам.

¹ Необходимая полоса — это минимальное значение полосы пропускания спектра модулированного сигнала, при котором искажения сообщений в отсутствие помех не превышают допустимых значений.

Помехи от радиостанций могут воздействовать на радиоприемные устройства как по основному, так и по побочным каналам приема этих устройств.

Побочные каналы разделяются на комбинационные и интермодуляционные. Комбинационные каналы приема обусловлены взаимодействием в смесителе (преобразователе частоты) напряжений сигнала и гетеродина и образованием комбинационных частот вида $f = |mf_c \pm nf_r|$, где f_c и f_r — соответственно частоты напряжений сигнала и гетеродина. Как известно, основными комбинационными каналами являются зеркальный канал (канал симметричной станции) и канал прямого прохождения сигнала (канал промежуточной частоты). Интермодуляционные каналы приема возникают за счет взаимодействия нескольких мешающих сигналов (от двух или более радиостанций) и напряжения гетеродина. К побочным каналам приема относится также прием мешающих излучений по боковым лепесткам диаграммы направленности приемной антенны. Действие помех в приемнике может проявляться не только в искажении полезных сообщений, но и в их подавлении.

5.3. ИНДУСТРИАЛЬНЫЕ ПОМЕХИ

К промышленным относят помехи, создаваемые любыми техническими устройствами и установками, кроме радиостанций¹. Такими устройствами и установками являются:

а) высокочастотные, генерирующие высокочастотные колебания не для информационных, а каких-либо, других («энергетических») целей — облучения живых организмов (в медицине), закаливания стали, сушки древесины, обработки пищевых продуктов, питания синхротронов и т. д.;

б) не предназначенные для генерации высокочастотных колебаний, но создающие мешающие электрические, магнитные и электромагнитные поля, например электротранспорт, системы зажигания двигателей внутреннего сгорания, реактивные двигатели, создающие плазменные излучения, линии электропередач, электротехнические и коммутационные устройства, пылесосы, стиральные машины, движущиеся (трущиеся о грунт) гусеницы тракторов и танков и многие другие.

Помехи, вызванные устройствами группы б), имеют, как правило, непериодический характер, интенсивность их спектра убывает с ростом частоты, и на дециметровых и более коротких волнах спектральная плотность таких помех падает ниже уровня внутренних шумов радиоприемных устройств.

Индустриальные помехи могут попадать в радиоприемное устройство различными путями: в результате электромагнитного излучения, проникающего в антенну приемника; воздействия элек-

трических и магнитных полей на цепи и контуры приемника, не защищенные электрическими и магнитными экранами; проникновения через цепи электропитания приемника.

5.4. ВНУТРИСИСТЕМНЫЕ ПОМЕХИ ВТОРОГО ВИДА

Перечисленные помехи от радиостанций и промышленные относятся к межсистемным, а также внутрисистемным помехам первого вида. Поэтому остановимся дополнительно лишь на внутрисистемных помехах второго вида, т. е. таких, источники которых находятся внутри самой радиоаппаратуры. Прежде всего к таким источникам можно было бы отнести внутренние шумы радиоприемных устройств. Однако вследствие того, что эти источники существуют во всех без исключения радиоприемных устройствах и определяют их предельную чувствительность, создаваемые ими помехи выносятся в отдельный класс и, следовательно, не относят к внутрисистемным помехам второго вида.

Основными источниками внутрисистемных помех второго вида являются [51, 52]:

внутренние шумы радиопередающих устройств¹;

при питании от переменного тока — напряжение питания и его гармоники;

при применении преобразования частоты в приемнике или передатчике (однократного или многократного, с умножением или делением частоты или образованием промежуточной частоты разностного, суммарного или комбинационного типа) — напряжение комбинационных частот различного вида и происхождения. Помехи такого вида особенно существенны при применении синтезаторов частоты и в приемниках с многократным преобразованием частоты;

перекрестные помехи между различными каналами системы, вызываемые нелинейными явлениями, паразитными связями и другими факторами;

нестабильные контакты в кабельных разъемах, заземлениях и других электрических и механических цепях;

электромагнитные и электронные реле, электротехнические переключатели и выключатели;

электрические двигатели и генераторы;

тиристорные регуляторы в источниках питания.

¹ Внутренние шумы радиопередающих устройств, вызывающие шумовую составляющую неосновного излучения передатчика, могут создавать также и межсистемные помехи, и внутрисистемные помехи первого вида. Однако они особенно опасны в тех случаях, когда по условиям работы радиопередающее устройство расположено в непосредственной близости от радиоприемного устройства или, тем более, если оба эти устройства работают на общую приемно-передающую антенну. Поэтому внутренние шумы передатчиков могут быть существенным источником и внутрисистемных помех второго вида.

¹ Здесь, как и ранее, под радиостанцией понимают приемопередающее, приемное или передающее устройство информационного назначения, т. е. обеспечивающее прием или передачу информации (сообщений).

5.5. ОСНОВНЫЕ ПУТИ ОБЕСПЕЧЕНИЯ ЭМС

Обеспечение ЭМС достигается комплексом мер, принимаемых как в отношении источников помех, так и в рецепторах. Из описанного характера главных источников помех вытекают следующие основные меры для ослабления их мешающего действия.

Для помех от радиостанций:

- уменьшение мощности основного излучения радиопередающих устройств;

- уменьшение длительности основного излучения;

- уменьшение нестабильности частоты передатчиков и гетеродинов приемников;

- сокращение необходимой полосы спектра излучения¹;

- увеличение направленности излучения (при сохранении неизменной излучаемой мощности);

- ослабление уровня боковых лепестков диаграммы направленности антенной системы;

- применение волн такой длины, при которой излучение распространяется лишь в пределах прямой видимости, и притом возможно более коротких;

- ослабление нежелательных видов излучения (внеполосных и побочных);

- достаточно большое разнесение несущих частот радиостанций, особенно расположенных на небольших расстояниях друг от друга.

Для индустриальных помех:

- применение (при прочих примерно равных условиях) таких принципов действия устройств и установок, при которых интенсивность их излучения в области высоких частот была бы возможно меньшей;

- применение фильтров нижних частот в цепях централизованного питания установок (для предотвращения распространения помех от этих установок по сети питания);

- применение экранирования (электрического и магнитного) установок и их частей;

- устранение нестабильности контактов и их искрения;

- применение заземлений.

Для ослабления внутрисистемных помех второго вида могут применяться самые разнообразные меры (в зависимости от характера источника помех), включая упомянутые применительно к индустриальным помехам.

¹ В гл. 2 указывалось, что если вблизи радиопередающего устройства нет радиоприемных устройств, то применение сложных сигналов с большой базой, занимающих большую полосу частот, может не повредить, а понизить их мешающее действие вследствие уменьшения спектральной плотности сигнала при том же значении его средней мощности. Однако в общем случае, особенно при наличии близко расположенных радиоприемных устройств, расширение спектра сигнала приводит к увеличению его мешающего действия.

Для устранения возможности создания различными источниками помех мешающих излучений недопустимого уровня принимаются специальные организационно-технические меры как внутри страны, так и международного значения: создаются специальные регламентирующие и контролирующие национальные и международные организации и нормативные документы. Одним из основных назначений этих организаций и документов является распределение частот (полос частот) для отдельных стран, районов и служб. Наиболее крупной международной организацией по распределению частот является МСЭ — Международный союз электросвязи (являющийся органом ООН). В него входят МКРЧ (Международный комитет регистрации частот) и МККР (Международный консультативный комитет по радиосвязи). Основным международным документом по распределению частот — Регламент радиосвязи.

Для распределения полос частот весь земной шар разбивается на три зоны:

- все страны Европы (включая европейские части СССР и Турции), азиатские части СССР и Турции, Монгольская народная республика и Африка;

- Северная и Южная Америка, Гренландия;

- Азия (кроме СССР, Монголии и Турции), Австралия и Океания. Различают распределение, присвоение и выделение частот (полос частот): распределение относится к службам, выделение — к зонам и странам, присвоение — к станциям.

Регламент радиосвязи распределяет частоты между службами и выделяет частоты зонам и странам, устанавливает нормы на допустимые радионизлучения, меры борьбы с радиопомехами, правила работы радиостанций и др. При этом учитываются следующие основные виды служб: Радионавигация, Радиолокация, Радиовещание, Морская, Подводная, Любители (радиолюбители), Воздушная радионавигация, Морская радионавигация, Радиометеорология, Наземная подвижная, Радиоастрономия, Космические исследования, Космическая связь с помощью ИСЗ, Радионавигация с помощью ИСЗ, Радиометеорология с помощью ИСЗ. Распределение частот между этими службами для первой зоны, приведено, например, в [51]. При этом предусматриваются следующие области применения (виды служб) для основных диапазонов радиоволн:

- глобальная радиосвязь для декаметровых и декаметровых волн;

- связь и звуковое вещание — для километровых, гектометровых и декаметровых волн;

- связь, астрономия — для метровых волн;

- локация, навигация, управление, связь, телевидение, астрономия и др. — для дециметровых и сантиметровых волн;

- космическая связь и волноводные линии — для миллиметровых и децимиллиметровых волн.

Кроме отмеченных мер, принимаемых в отношении источников помех, весьма важными для обеспечения ЭМС являются также:

повышение помехоустойчивости радиоприемных устройств по основным каналам приема (применение пространственной, поляризационной, частотной, временной, амплитудной и других видов селекции);

ослабление неосновных каналов приема;

применение фильтров в цепях питания;

экранирование и заземление;

устранение нестабильности контактов и их искрения;

обеспечение необходимой развязки между передающими и приемными антеннами и между отдельными излучателями приемных антенн.

5.6. НОРМЫ И РЕКОМЕНДАЦИИ ПО ОГРАНИЧЕНИЮ УРОВНЯ ПОМЕХ

К нормам и рекомендациям по ограничению уровня помех относятся, в частности, следующие.

1. Нормы на необходимую ширину полосы Δf_n :

АМ-одноканальная телеграфия

$$\Delta f_n = k B,$$

где B — скорость передачи телеграфных знаков в бодах; k — коэффициент, который для линий без замирания выбирается равным 3;

АМ-телефония (передача речи, музыки)

$$\Delta f_n = 2 F_n;$$

телефония с одной боковой полосой

$$\Delta f_n = F_n - F_{\text{н}},$$

где F_n и $F_{\text{н}}$ — соответственно высшая и низшая частоты спектра сообщения;

ЧМ-телефония

$$\Delta f_n = 2 (\Delta f_{\text{макс}} + F_n),$$

где $\Delta f_{\text{макс}}$ — девиация частоты.

2. Норма на ширину Δf_n полосы излучения (ШПИ):

АМ-телеграфия

$$\Delta f_n = \Delta f_n + B/2;$$

АМ-телефония

$$\Delta f_n = 3 F_n;$$

ЧМ-телефония

$$\Delta f_n = 2 (\Delta f_{\text{макс}} + F_n) + F_n$$

(при индексе модуляции $m = \Delta f_{\text{макс}}/F_n \leq 5$).

3. Нормы на стабильность частоты передатчиков приведены в [51], табл. 6.1. Например, для радиовещательных станций нестабильность Δf частоты передатчика не должна превышать 10 ... 100 Гц (в зависимости от диапазона используемых частот); для станции морской подвижной службы допускается $\Delta f/f_0 \leq 3 \cdot 10^{-4}$ и т. д.

4. Нормы на побочные излучения передатчиков (на входе антенны передатчика):

при $f \leq 30$ МГц средняя мощность побочного излучения (по отношению к средней мощности основного излучения) не более —40 дБ, но ≤ 50 МВт;

для передатчиков со средней мощностью более 50 кВт ≤ -60 дБ;

для передатчиков с мощностью менее 5 Вт ≤ -30 дБ.

5. Нормы на некоторые параметры радиорелейных линий связи (РРЛ) и спутниковых систем связи.

Поскольку РРЛ и ССС работают в общем диапазоне частот (3 ... 10 ГГц), установлены жесткие нормы, обеспечивающие их совместную работу:

плотность потока мощности у поверхности Земли от излучения передатчика, установленного на ИСЗ, в полосе 4 кГц не должна превышать $(-152 + \theta/15)$ дБ·Вт/м², где θ — угол прихода волны относительно местной горизонтали в градусах. Например, при $\theta = 15^\circ$ мощность излучения в полосе 4 кГц, приходящаяся на 1 м² поверхности Земли, не должна превышать —151 дБ относительно 1 Вт, т. е. $0,8 \cdot 10^{-15}$ Вт;

наземная станция, входящая в состав ССС, во избежание помех радиорелейным линиям должна излучать в полосе 4 кГц не более 55 дБ·Вт в любом направлении, т. е. не более $3 \cdot 10^5$ Вт. Соответствующие нормы устанавливаются и на излучение радиорелейных линий.

6. Нормы на интенсивность промышленных помех приведены, например, в [51].

5.7. ПОКАЗАТЕЛИ И ПАРАМЕТРЫ ЭМС

Поскольку ЭМС зависит от большого числа разнообразных источников помех, оказывающих мешающее действие многими существенно различными путями, для ее количественной характеристики приходится применять совокупность большого числа показателей и параметров. К ним, в частности, относятся следующие.

Для радиопередающих устройств:

частота основного излучения (рабочая частота);

частоты гармоник;

частоты комбинационных излучений;

диапазон частот или сетка частот;

сочетание частот двух и более передатчиков, определяющее интермодуляционные помехи (во всем диапазоне или во всей сетке частот этих передатчиков);

мощность основного излучения (плотность потока мощности, т. е. произведение мощности передатчика на усиление его антенны); относительная мощность излучения на гармониках и комбинационных излучений;

относительная мощность взаимно модуляционных излучений; ширина основного луча диаграммы направленности антенны; сектор качания луча (обзора); уровни боковых лепестков диаграммы направленности; вид поляризации излучения; коэффициенты связи между антеннами передатчиков и с антеннами приемников;

занимаемая ширина полосы частот радиоизлучения; скорость убывания внеполосных излучений; интенсивность шумовых излучений; нестабильность рабочей частоты; ширина полосы излучения гармоник.

Для радиоприемных устройств:

частота основного канала (рабочая частота); диапазон частот или сетка дискретных частот; промежуточная частота (частоты); частота (частоты) гетеродина; полоса пропускания основного канала; коэффициент прямоугольности частотной характеристики; коэффициент шума; чувствительность (пороговая чувствительность) по основному каналу;

частоты и уровни восприимчивости побочных каналов (канала промежуточной частоты, зеркального и других комбинационных каналов);

характеристики многоканальной избирательности (по блокированию, по перекрестным искажениям и по интермодуляции);

ширина основного луча диаграммы направленности и сектор обзора;

уровни боковых лепестков диаграммы направленности; вид поляризации и потери на рассогласование поляризаций приемной и передающей антенн;

интенсивность излучения гетеродина (гетеродинов);

нестабильность частоты гетеродина (гетеродинов);

частотная характеристика антенны;

шум антенны и космического пространства;

общий уровень промышленных помех.

5.8. РАСЧЕТ ПРОСТРАНСТВЕННОГО И ЧАСТОТНОГО РАЗНОСА, НЕОБХОДИМОГО ДЛЯ ОБЕСПЕЧЕНИЯ ЭМС

На рис. 5.1 изображены диаграммы направленности передатчика и приемника данной радиолнии и мешающего передатчика, расположенных соответственно в пунктах А, В и С. Мощ-

ность помехи на входе приемника (на выходе приемной антенны) [51]

$$P_{пр.п} = P_{пер.п} + G_{пер.п} + G_{пр} - K_{вч.п} - K_p - K_{вч.пр} - K_{пол} - K_{сн} - K_{вр} - K_{п}. \quad (5.1)$$

Здесь все величины выражены в децибелах и приняты следующие обозначения: $P_{пер.п}$ — мощность мешающего передатчика; $G_{пер.п}$ и $G_{пр}$ — коэффициенты усиления антенн передатчика помехи и приемника в направлении ВС (рис. 5.1); $K_{вч.п}$ — потери в высокочастотном тракте передатчика помехи; K_p — потери при распространении на линии передатчик помехи — приемник; $K_{вч.пр}$ — потери в высокочастотном тракте приемника; $K_{пол}$ — коэффициент, учитывающий ослабление помехи из-за несовпадения поляризаций антенн передатчика помехи и приемника; $K_{сн}$ — коэффициент, учитывающий ослабление помехи из-за несовпадения рабочих частот и несовпадения спектра помехи с шириной полосы приемника; $K_{вр}$ — коэффициент, учитывающий изменение мешающего действия помехи из-за изменения усиления антенн мешающей и рабочей станций при вращении диаграмм направленности их антенн; $K_{п}$ — коэффициент, учитывающий неравномерность распределения мощности сигнала и помехи по спектру частот.



Рис. 5.1

Необходимо, чтобы выполнялось условие

$$K_{\Phi} \geq K_{тр}, \quad (5.2)$$

где $K_{тр}$ — требуемое отношение сигнала к помехе (по мощности) на входе приемника, обеспечивающее нормальное воспроизведение сообщений в приемнике; K_{Φ} — фактическое (расчетное) отношение сигнал-помеха при данном пространственном разнорасе. Оно равно

$$K_{\Phi} = P_{пр.с} / P_{пр.п}$$

или в децибелах

$$K_{\Phi} = P_{пр.с} - P_{пр.п}. \quad (5.3)$$

Здесь $P_{пр.с}$ — мощность полезного сигнала на входе приемника. Следовательно, в децибелах должно выполняться условие [при знаке равенства в (5.2)]

$$P_{пр.с} - P_{пр.п} = K_{тр},$$

которое с учетом (5.1) принимает вид

$$K_p = P_{пер.п} - P_{пр.с} + K_{тр} + G_{пер.п} + G_{пр} - K_{вч.п} - K_{вч.с} - K_{пол} - K_{сн} - K_{вр} - K_{п}. \quad (5.4)$$

При распространении в свободном пространстве (в противном случае следует учесть дополнительные потери из-за затухания радиоволн в несвободном пространстве)

$$K_p = 10 \lg (4 \pi R/\lambda)^2 = 20 \lg (4 \pi R/\lambda), \quad (5.5)$$

где R — расстояние от приемника до мешающего передатчика (ВС на рис. 5.1).

Из соотношений (5.4) и (5.5) можно определить значение R , обеспечивающее выполнение условия (5.4), т. е. требуемое пространственное разнесение между приемником и мешающим передатчиком. Очевидно, аналогичный расчет следует сделать и для случая, когда станцию A , принятую ранее за рабочую, считают мешающей, а мешающую станцию — рабочей (ибо необходимо обеспечить допустимый уровень взаимных помех). Если применяются радиоволны, распространяющиеся лишь в пределах прямой видимости, то следует определить также дальность прямой видимости в километрах

$$R_{пр} \approx 4,1 (\sqrt{h_1} + \sqrt{h_2}), \quad (5.6)$$

где h_1 и h_2 — высоты приемной и передающей антенн в м. Если расстояние R , найденное из соотношений (5.4) и (5.5), удовлетворяет условию

$$R \leq R_{пр}, \quad (5.7)$$

то пространственное разнесение следует выбрать равным R , а в противном случае $R_{пр}$. Однако при $R > 100$ км следует учесть тропосферное распространение.

Если расстояние между радиостанциями заданы, то для обеспечения ЭМС необходимо выбрать требуемое частотное разнесение между рабочей и мешающей станциями. Его значение также можно определить на основании соотношений (5.4) и (5.5), но теперь заданным следует считать расстояние R , а искомым — значение коэффициента $K_{сп}$. При этом требуемым будет такое разнесение частот, которое обеспечит найденное значение величины $K_{сп}$.

5.9. АНАЛИЗ ЭМС

Для проверки выполнения требований ЭМС необходимо определить соответствующие показатели и параметры ЭМС, в том числе указанные ранее. Приближенная оценка и учет целого ряда из этих показателей и параметров необходимы уже на начальных стадиях проектирования, т. е. при выборе основных принципов построения системы. Однако многие параметры и показатели можно достаточно точно определить лишь на более поздних стадиях проектирования путем моделирования и испытания макетов. Наконец, наиболее полная проверка выполнения требований ЭМС происходит при испытаниях опытных образцов в условиях эксплуатации системы.

Проверка выполнения требований ЭМС, включающая определение всех необходимых показателей и параметров ЭМС, называется *анализом ЭМС*. Важной составной частью этого анализа является изучение электромагнитной обстановки, в которой будет функционировать (или уже функционирует) данная радиосистема. При этом под электромагнитной обстановкой понимается совокупность электромагнитных излучений, действующих на данную радиосистему (радиоприемное устройство) и создаваемых другими радиосистемами и источниками (за исключением, как отмечалось, преднамеренных помех). При этом необходимо учитывать электромагнитную обстановку (ЭО) как для группы радиоэлектронных средств, входящих в данную систему, так и для индивидуального радиосредства. Электромагнитная обстановка делится на внешнюю (по отношению к данной системе) и внутреннюю. В первом случае помехи действуют в основном через антенные системы, а во втором — по всем возможным путям.

Для анализа ЭМС ввиду его большой сложности широко применяется моделирование, в частности создаются следующие виды моделей:

- для анализа ЭО;
- для определения побочных каналов приема;
- для анализа совместной работы радиосредств;
- отборочная.

Отборочная модель предназначена для выделения (отбора) из большого числа радиоэлектронных средств (например, из нескольких тысяч) только тех, которые могут работать без взаимных помех.

Из приведенного рассмотрения следует, что проблема обеспечения ЭМС и расчет ее параметров и показателей являются одними из самых сложных задач при проектировании радиосистем и их частей. В то же время актуальность проблемы обеспечения ЭМС с каждым годом быстро возрастает.

5.10. ЭКОЛОГИЧЕСКАЯ СОВМЕСТИМОСТЬ РАДИОСИСТЕМ

Под экологической совместимостью (ЭКС) радиосистемы понимается возможность создания этой системы, ее функционирования и утилизации без нанесения недопустимого ущерба среде обитания людей и непосредственно самим людям¹. При этом ущерб среде обитания считается недопустимым, если он может существенно ухудшить существование людей данного или последующих поколений.

При создании, функционировании и утилизации радиосистемы ущерб людям может наноситься самыми разнообразными путями, в первую очередь следующими:

¹ Ущерб, который может быть нанесен людям использованием военных радиосистем в агрессивных целях, не включают в проблему ЭКС, так как он должен являться предметом специальных исследований.

генерацией интенсивных электромагнитных излучений или высоких электрических напряжений, наносящих вред здоровью людей, животному или растительному миру;

какими-либо другими нарушениями требований техники безопасности (например, повышенной взрывоопасностью);

чрезмерным расходом ограниченных природных ресурсов — энергии и веществ (особенно редких);

загрязнением среды в процессе утилизации системы (т. е. после окончания срока ее службы).

Отсюда следует, что требования ЭКС включают в себя все требования, предъявляемые правилами техники безопасности, но не сводятся к ним. Показатели, которые должны количественно учитывать степень удовлетворения требований ЭКС, весьма разнообразны. В первую очередь к ним относятся:

потребляемая энергия;

коэффициент полезного действия;

средняя и пиковая мощность излучения на различных частотах и длительность излучения;

надежность электрических и магнитных экранов, препятствующих созданию в окружающей среде сильных электрических и магнитных полей;

надежность экранов, предотвращающих проникновение в окружающую среду побочных (т. е. вне рабочего диапазона волн) вредных излучений (рентгеновского, гамма-излучения и др.);

взрывобезопасность;

массы редких веществ, расходуемых на создание, функционирование и утилизацию данной системы (золота, платины, серебра, других редких металлов и других дефицитных веществ);

показатели, характеризующие степень засорения окружающей среды в результате ликвидации системы после окончания срока ее службы.

С каждым годом требования к ЭКС радиосистем становятся все более жесткими. Это вызывается рядом факторов, основными из которых являются следующие:

резкое увеличение числа РС различных видов и назначений (только телевизоров и радиоприемников в СССР выпускается в год более 15 млн. экз);

усложнение функций, выполняемых радиосистемами, приводящее к необходимости создания систем, потребляющих большую энергию и большое количество редких веществ (включая золото, серебро и платину) и требующих для своего функционирования сильных электрических и магнитных полей;

необходимость использования для извлечения и передачи информации все более коротковолновых излучений (включая оптическое, рентгеновское и гамма-излучения), представляющих повышенную опасность;

применение интегральной технологии изготовления блоков, затрудняющее ликвидацию системы после окончания срока ее службы.

Поэтому, если в прошлом при проектировании радиосистем требования ЭКС сводились в основном к стандартным требованиям техники безопасности и не представляло больших трудностей их выполнение, в настоящее время проблема обеспечения ЭКС является одной из важных и сложных проблем, которые должны решать разработчики РС.

Литература: основная [32, 51], дополнительная [33, 52, 53].

Глава 6 НАДЕЖНОСТЬ

6.1. ВЕРОЯТНОСТЬ БЕЗОТКАЗНОЙ РАБОТЫ СИСТЕМЫ

Как известно, надежность является сложным понятием, включающим безотказность, долговечность, ремонтпригодность и сохраняемость. Вопросы надежности подробно рассматриваются в курсах конструирования и производства радиоэлектронной аппаратуры. Поэтому напомним лишь наиболее существенные положения теории надежности, которые особенно важно учитывать на начальных этапах проектирования.

Основной характеристикой надежности объекта (системы, устройства) является вероятность $P(t)$ его безотказной работы в течение времени t . Для определения $P(t)$ удобно использовать понятие интенсивности $\lambda(t)$ отказов¹, т. е. числа отказов в единицу времени:

$$P(t) = \exp \left[- \int_0^t \lambda(\tau) d\tau \right]. \quad (6.1)$$

Типичная зависимость $\lambda(t)$ (рис. 6.1) содержит три участка: приработки (I); нормального функционирования (II) и износа (III). Обычно в процессе производства аппаратуру подвергают

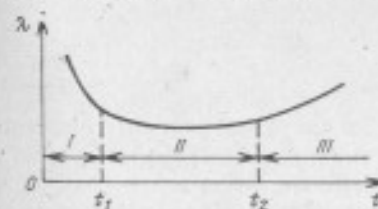


Рис. 6.1

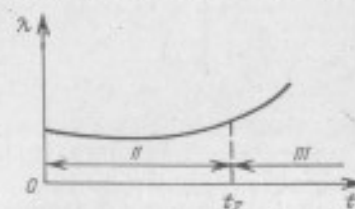


Рис. 6.2

¹ Использование λ -характеристики существенно упрощает расчет, но не позволяет полностью учесть все факторы, влияющие на надежность, например частоту включения аппаратуры.

тренировке для выхода на участок II. Поэтому, если за начало отсчета принять момент t_1 начала нормальной эксплуатации системы, зависимость $\lambda(t)$ приобретает вид, изображенный на рис. 6.2. На участке II можно полагать $\lambda(t) = \text{const}$ и зависимость (6.1) принимает экспоненциальный вид

$$P(t) = \exp(-\lambda t). \quad (6.2)$$

Плотность вероятности $W(t)$ связана с вероятностью отказа соотношением

$$W(t) = -dP(t)/dt. \quad (6.3)$$

Поэтому при экспоненциальном законе (6.2) получаем

$$W(t) = \lambda \exp(-\lambda t). \quad (6.4)$$

При этом среднее время безотказной работы (средняя наработка до первого отказа, средняя наработка на отказ)

$$T = \int_0^{\infty} t W(t) dt = \frac{1}{\lambda}. \quad (6.5)$$

Экспоненциальная аппроксимация (6.2) и (6.4) весьма удобна для расчетов. В частности, при таком законе просто определять надежность системы, состоящей из n элементов (частей). Действительно, если в каждом i -м элементе системы вероятность отсутствия отказа

$$P_i(t) = \exp(-\lambda_i t), \quad (6.6)$$

отказы элементов статистически независимы и отказ любого из них приводит к отказу всей системы, то вероятность отсутствия отказа в системе (за время t)

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

где N число элементов.

С учетом (6.6) получаем

$$P(t) = \exp(-\lambda_c t), \quad (6.7)$$

где

$$\lambda_c = \sum_{i=1}^N \lambda_i, \quad (6.8)$$

т. е. закон распределения отказов для системы в целом оказывается также экспоненциальным и интенсивность отказов λ_c равна сумме интенсивностей отказов всех элементов.

Однако из рис. 6.2 следует, что полагать $\lambda(t) = \text{const}$ (чему соответствует экспоненциальный вид закона надежности) можно лишь приближенно и лишь в пределах участка II. Для того чтобы учесть возрастание интенсивности отказов $\lambda(t)$ при подходе к участку III и на самом этом участке, необходимо использовать более сложные аппроксимации закона распределения, например распре-

деление Вейбулла, гамма-распределение и др. [54, 55]. Одна из наиболее простых аппроксимаций (рис. 6.3) имеет следующий вид:

$$P(t) = \begin{cases} \exp(-\lambda t) & \text{при } t \in [0, T], \\ 0 & \text{вне этих границ;} \end{cases} \quad (6.9)$$

Значение параметра T выбирается примерно равным моменту t_2 перехода на участок III λ -характеристики (т. е. на участок износа).

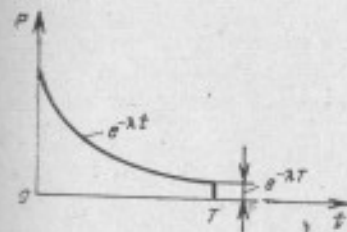


Рис. 6.3

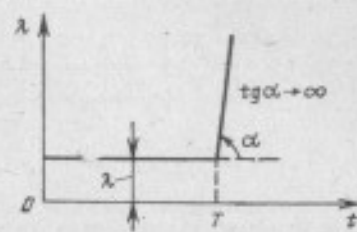


Рис. 6.4

Очевидно, закону (6.9) соответствует аппроксимация λ -характеристики, изображенной на рис. 6.2, зависимостью, приведенной на рис. 6.4, т. е. предположение, что в момент $t = T \approx t_2$ скачкообразно наступает полный износ системы. Иначе говоря, предполагается, что при $t \leq T$ отказы имеют экспоненциальный закон распределения, а в момент $t = T$ отказ наступает с вероятностью, равной единице (вследствие истощения ресурса T системы).

Из (6.3) следует, что закону распределения $P(t)$, изображенному на рис. 6.3, соответствует дифференциальный закон

$$W(t) = \begin{cases} \lambda \exp(-\lambda t) + \exp(-\lambda T) \delta(t - T) & \text{при } t \in [0, T], \\ 0 & \text{вне этих границ.} \end{cases} \quad (6.10)$$

здесь $\delta(t)$ — функция Дирака.

6.2. ОСОБЕННОСТИ ОПРЕДЕЛЕНИЯ НАДЕЖНОСТИ СИСТЕМ С ВОССТАНОВЛЕНИЕМ

В большинстве систем в случае наступления отказа предусматривается возможность восстановления работоспособности этих систем путем замены отказавших элементов (частей) резервными или путем ремонта. В таких случаях в течение всего срока t эксплуатации (функционирования) системы имеет место последовательность чередующихся участков нормальной работы и участков восстановления (рис. 6.5, где ξ_i — длительность i -го участка нормальной работы; θ_j — длительность j -го участка восстановления). Надежность такой системы, кроме вероятности $P(t)$ безотказной работы, характеризуется также средним значением T_*

времени восстановления системы и коэффициентом K_r готовности системы. По определению

$$T_a = M[\theta_i], \quad (6.11)$$

а под K_r понимается вероятность того, что в произвольный момент t система будет находиться в состоянии нормального функционирования (а не отказа). Коэффициент готовности наиболее просто определяется в установившемся режиме (т. е. при $t \rightarrow \infty$) и при

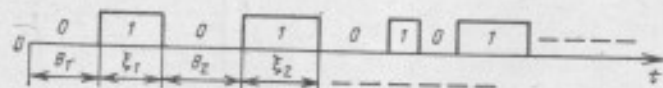


Рис. 6.5

следующих допущениях: величины θ_i и θ_j статически независимы и имеют экспоненциальные законы распределения; при этом

$$K_r = T_{cp} / (T_{cp} + T_a), \quad (6.12)$$

где $T_{cp} = 1/\lambda$ — среднее время безотказной работы; T_a — среднее время восстановления.

Из (6.12) следует, что при данном значении T_{cp} между показателями надежности K_r и T_a существует взаимно-однозначная связь: если задано допустимое значение $T_a = T_{a, \max}$, то минимально допустимое значение коэффициента готовности

$$K_{r, \min} = T_{cp} / (T_{cp} + T_{a, \max}). \quad (6.13)$$

Если, наоборот, задано минимально-допустимое значение коэффициента готовности, то в соответствии с (6.12)

$$T_{a, \max} = T_{cp} \left(\frac{1}{K_{r, \min}} - 1 \right). \quad (6.14)$$

В одних практических задачах лимитирующим (т. е. заданным) оказывается значение $T_{a, \max}$, а в других — значение $K_{r, \min}$.

6.3. СРОК СЛУЖБЫ СИСТЕМЫ

При оценке надежности системы важное значение имеет понятие «срок службы системы» T_{cl} , под которым понимают календарное время ее эксплуатации до наступления предельного состояния. Иначе говоря, при $t > T_{cl}$ система снимается с эксплуатации. Предельное состояние вызывается одним из следующих факторов:

- моральным (по истечении времени $T_{m, n}$ с начала эксплуатации);
- техническим (по истечении времени $T_{t, n}$);
- экономическим (по истечении времени $T_{э, n}$).

Моральный фактор обусловлен тем, что в связи с развитием техники к моменту $t = T_{m, n}$ появляются системы, настолько превосходящие данную по своим показателям качества, что оставлять ее в эксплуатации оказывается нецелесообразным. Технический фактор вызван тем, что к моменту $t = T_{t, n}$ система становится технически неработоспособной или по меньшей мере резко возрастает вероятность ее отказа. Экономический фактор проявляется в том, что к моменту $t = T_{э, n}$ система, даже если она и сохраняет работоспособность, настолько часто отказывает и соответственно требует столь много средств на ремонт, что экономически целесообразно заменить ее новым экземпляром. Очевидно, срок службы T_{cl} определяется той из величин $T_{m, n}$, $T_{t, n}$, $T_{э, n}$, которая оказывается наименьшей, т. е.

$$T_{cl} = \min [T_{m, n}, T_{t, n}, T_{э, n}]. \quad (6.15)$$

Поскольку срок службы T_{cl} является случайной величиной, в качестве показателей надежности используются назначенный срок службы $T_{cl, n}$, средний срок службы $T_{cl, cp}$ и гамма-процентный срок службы $T_{cl, \gamma}$ (равный продолжительности от начала эксплуатации объекта, в течение которой он не достигнет предельного состояния с заданной вероятностью γ , выраженной в процентах). Величина $T_{cl, n}$ выбирается из условия

$$T_{cl, n} = t_s, \quad (6.16)$$

где t_s — назначенный срок эксплуатации системы. Из рис. 6.3 и 6.4 следует, что при выполнении условия

$$T \geq T_s \quad (6.17)$$

(где T определяется в календарном времени) приближенно справедлив экспоненциальный закон надежности (6.2), т. е. вероятность безотказной работы системы за время t , равна

$$P(t_s) = \exp(-\lambda t_s). \quad (6.18)$$

6.4. ОСНОВНЫЕ ПУТИ ПОВЫШЕНИЯ НАДЕЖНОСТИ

- Основными путями повышения надежности являются:
- применение более надежных элементов;
 - облегчение режимов работы аппаратуры;
 - совершенствование технологии производства аппаратуры, в частности применение интегральной технологии;
 - замена механических переключателей и управляющих устройств электронными;
 - замена аналоговой обработки цифровой;
 - упрощение (при прочих примерно равных условиях) принципов действия системы, т. е. устранение излишнего ее усложнения;
 - предварительная (до ввода в эксплуатацию) тренировка системы (для выхода на участок II λ -характеристики);
 - регламентные работы;

максимальное использование функциональных резервов;
введение структурного резервирования;
применение и совершенствование систем обнаружения и устранения неисправностей (отказов);
повышение квалификации обслуживающего персонала.

Остановимся несколько подробнее на использовании функционального и структурного резервирования для повышения надежности. Функциональные резервы в системе проявляются в том, что в случае выхода из строя (отказа) частей системы, обеспечивающих выполнение некоторых функций, система в целом сохраняет способность решать поставленные перед ней основные задачи, быть может, с некоторым допустимым ухудшением качества решения. Приведем несколько примеров наличия функциональных резервов.

Пример 1. В системе с многоэлементной ФАР выход из строя нескольких процентов элементов ФАР не приводит к существенному ухудшению показателей качества системы.

Пример 2. В голографической оптоэлектронной системе нарушение значительной части (например, 20%) голограммы не вызывает недопустимого ухудшения показателей качества.

Пример 3. В ряде систем радиоуправления космическими аппаратами предусматриваются независимые измерения шести радиолокационных параметров летательного аппарата: дальности r , двух угловых координат (например, азимута и угла места) α и β и их производных $\dot{r}$, $\dot{\alpha}$ и $\dot{\beta}$. Однако в некоторых задачах качество управления может остаться удовлетворительным даже при выходе из строя нескольких измерителей указанных параметров.

Пример 4. В спутниковой системе связи, содержащей девять спутников, вышли из строя три. Если система состоит из специализированных спутников, то оставшиеся спутники не смогут взять на себя функции вышедших из строя. Например, если из девяти спутников только три в состоянии ретранслировать телевизионные программы, а остальные шесть могут передавать лишь телефонную и телеграфную информацию, то выход из строя трех телевизионных спутников приведет к полной невозможности передавать телевизионные программы.

В отличие от этого, если в системе все из девяти спутников являются универсальными, т. е. каждый из них способен передавать телевизионную, телеграфную и телефонную информацию, то выход из строя трех спутников может привести не к полному отказу системы, а лишь к некоторому уменьшению ее пропускной способности. Очевидно, в первом случае система связи не обладает функциональным резервом, а во втором случае — обладает.

Из приведенных примеров видно, что в некоторых системах (см. примеры 1 и 2) функциональные резервы являются органическими, т. е. неизбежно (внутренне, органически) присущими данному классу систем (устройств). В других системах (примеры 3 и 4) они не являются органическими, т. е. системы данного класса можно, вообще говоря, строить и так, что они не будут обладать функциональными резервами.

Из примеров следует также, что, если даже система данного класса в принципе (потенциально) обладает функциональными резервами, для того, чтобы эти резервы действительно повышали надежность действия системы, могут потребоваться специальные технические меры, обеспечивающие использование указанных резервов.

Наконец, необходимо учитывать следующие обстоятельства. Система, обладающая функциональным резервом, обеспечивает по сравнению с системой, лишенной такого резерва, большую надежность лишь при прочих равных условиях, в частности при одинаковой возможности использования других путей повышения надежности системы. Однако в ряде случаев прочие условия оказываются неравными. Например, в активной многоэлементной ФАР допустим выход из строя 20% элементов, но вероятность выхода из строя более 20% элементов за время t , может быть равна, например, 10^{-2} , тогда как вероятность выхода из строя за то же время t , зеркальной антенны может оказаться меньшей 10^{-2} . Следует также учесть, что системы, не обладающие функциональными резервами, могут иметь меньшие массы, габаритные размеры и стоимость, что позволит применить другие меры повышения надежности (например, структурное резервирование) и получить в итоге более высокую надежность. Поэтому вопросы целесообразности или нецелесообразности применения принципа функционального резервирования для повышения надежности системы следует решать конкретно в каждой задаче проектирования и с учетом не только показателей надежности, но и других показателей качества системы. Однако, если в системе функциональное резервирование в принципе уже имеется, его использование для повышения надежности требует, как правило, небольших дополнительных затрат и поэтому является целесообразным.

Рассмотрим теперь повышение надежности применением структурного резервирования. Этот вопрос подробно рассматривается в курсе конструирования и производства радиоаппаратуры, поэтому приведем лишь некоторые главные результаты [54—57].

Существуют различные виды резервирования: нагруженное, облегченное и ненагруженное; постоянное (без перестройки структуры объекта при возникновении отказа его элемента) и динамическое (с перестройкой структуры, в частности замещением отказавших элементов резервными); общее (резервирование объекта в целом), раздельное (поэлементное) и скользящее; смешанное и др.

Постоянное резервирование применяется лишь на самых низких иерархических уровнях системы, поэтому далее оно не рассматривается.

При нагруженном резервировании резервные элементы находятся в тех же условиях, что и основные (работающие); при облегченном резервировании резервные элементы находятся в облегченных условиях, а при ненагруженном резервные элементы

включаются в работу лишь после обнаружения отказа работающего элемента. Скользящее резервирование применимо лишь при наличии в системе одинаковых элементов (частей) и состоит в том, что число резервных элементов выбирается сравнительно небольшим, так как предусматривается, что каждый из резервных элементов может заменить любой из отказавших элементов.

При оценке надежности резервированных систем в первом приближении часто полагают, что устройства обнаружения отказов и переключения с отказавшего элемента на резервный работают идеально, т. е. срабатывают мгновенно и безошибочно. При таких предположениях получаются, в частности, следующие результаты [54]¹.

1. Раздельное (поэлементное) резервирование дает большее повышение надежности системы, чем общее резервирование (т. е. резервирование системы в целом). В частности, при нагруженном резервировании, если основная (рабочая) часть системы состоит из n равнонадежных элементов с вероятностью отказа каждого элемента q и каждая из m резервных частей системы содержит n таких же элементов, то вероятность отказа $P_{отк.с}$ резервированной системы при общем резервировании

$$P'_{отк.с} = [1 - (1 - q)^n]^{m+1}, \quad (6.19)$$

а при раздельном резервировании

$$P''_{отк.с} = 1 - (1 - q^{m+1})^n. \quad (6.20)$$

Если $nq \ll 1$ (т. е. надежность каждого элемента высока), то

$$P'_{отк.с} \approx n^{m+1} q^{m+1}, \quad P''_{отк.с} \approx nq^{m+1}, \quad (6.21)$$

$$P'_{отк.с} / P''_{отк.с} \approx n^m, \quad (6.22)$$

т. е. при раздельном резервировании вероятность отказа системы уменьшается в n^m раз. При однократном ($m=1$) резервировании, т. е. дублировании, этот выигрыш равен n . С увеличением кратности резервирования выигрыш возрастает.

2. Ненагруженное резервирование дает выигрыш по сравнению с нагруженным (при предположении, что вероятность отказа резервного элемента, находящегося в ненагруженном состоянии, т. е. до его включения вместо отказавшего элемента, равна нулю). Пусть на каждый рабочий элемент приходится m резервных элементов и вероятность отказа каждого из элементов (рабочего и резервного) за одинаковое время t одинакова и равна $q(t)$. Тогда вероятность отказа с учетом возможности замены отказавшего элемента резервным:

при нагруженном резервировании

$$P'_{отк}(t) = q^{m+1}(t); \quad (6.23)$$

¹ Из-за различия в обозначениях при сравнении приведенных далее формул с [54] следует заменить в них $(m+1)$ на m .

при ненагруженном резервировании формула для вероятности отказа получается весьма громоздкой, даже если $q(t)$ имеет экспоненциальное распределение (т. е. при $q(t) = 1 - e^{-\lambda t}$). Однако при весьма малых значениях q (т. е. при высоконадежных элементах) приближенно можно полагать

$$P'_{отк}(t) \approx q^{m+1}(t)/(m+1)!. \quad (6.24)$$

Из (6.23) и (6.24) следует, что при сделанных допущениях

$$P'_{отк}(t)/P''_{отк}(t) \approx (m+1)!, \quad (6.25)$$

т. е. при ненагруженном резервировании вероятность отказа в $(m+1)!$ раз меньше. При однократном ($m=1$) резервировании (дублировании) этот выигрыш минимален и равен двум, т. е. сравнительно незначителен. С учетом того, что в действительности (вопреки принятому допущению) при ненагруженном резервировании вероятность отказа резервного элемента до его включения в работу не равна нулю, выигрыш надежности, обусловленный применением ненагруженного резервирования вместо нагруженного, оказывается еще меньшим. Однако следует иметь в виду, что при ненагруженном резервировании элементы, находящиеся в состоянии резерва, не потребляют энергии, а при нагруженном резервировании они могут потреблять значительную энергию (например, если этими элементами являются мощные каскады радиопередающих устройств или устройства питания).

Сделанное допущение о том, что устройства обнаружения отказов и подключения резервных элементов вместо отказавших срабатывают мгновенно и безошибочно, является грубым. С учетом неидеальности этих устройств приведенные ранее соотношения и утверждения могут существенно измениться. В частности, может оказаться более целесообразным применять динамическое резервирование не на самых низких иерархических уровнях системы, а лишь на уровне довольно крупных блоков.

Вероятность отказа резервированной системы может оказаться значительно большей, чем это следует из приближенных соотношений (6.19) — (6.21), (6.23) и (6.24). Кроме того, необходимо учитывать, что применение резервирования приводит к существенному увеличению массы, габаритных размеров и стоимости системы. Вследствие этого, в частности, нередко приходится ограничиваться лишь однократным резервированием (дублированием), да и то не всех блоков системы.

Отсюда следует также, что хотя резервирование является одним из наиболее радикальных способов повышения надежности, оно должно, как правило, сочетаться с другими способами повышения надежности, в частности с перечисленными выше. Поэтому целесообразность применения проектируемой системы (устройства) следует обосновывать с учетом не только показателей надежности, но и других показателей качества системы, в первую очередь таких, как масса, габаритные размеры и стоимость.

6.5. ПРИЕМО-СДАТОЧНЫЕ ИСПЫТАНИЯ НА НАДЕЖНОСТЬ

Принципиальная трудность испытания на надежность заключается в невозможности определить статистические характеристики надежности за конечное время с полной достоверностью: при конечном времени испытаний (и конечном числе испытываемых образцов) всегда имеются какие-то вероятности ошибок. Например, при испытаниях, связанных с контролем средней наработки на отказ, разработчик и потребитель аппаратуры согласуют приемочное T_0 и браковочное T_1 значения средней наработки на отказ (очевидно, $T_0 > T_1$) и достаточно малые значения α и β вероятностей ошибок, где α — вероятность неприемки (браковки) партии изделий с приемочным значением T_0 , называемая *риском поставщика*; β — вероятность приемки партии изделий с браковочным значением T_1 , которая называется *риском потребителя*. По согласованным значениям T_0 , T_1 , α и β можно выбрать соответствующие параметры (код) испытаний.

6.6. ПРИМЕР РАСЧЕТА НАДЕЖНОСТИ ФУНКЦИОНИРОВАНИЯ СПУТНИКОВОЙ СИСТЕМЫ СВЯЗИ

Пусть для обеспечения нормального функционирования полностью развернутой спутниковой системы связи требуется, чтобы на геостационарной орбите вокруг Земли вращались n идентичных спутников — ретрансляторов сигналов. Срок эксплуатации (функционирования) системы $t_s = 10$ лет. Будем полагать, что нарушение функционирования ССС обусловлено лишь выходом из строя (отказами) аппаратуры спутников и закон распределения отказов приблизительно описывается соотношением (6.2), т. е. выполняется условие (6.17).

Будем считать, что отказ аппаратуры на каждом спутнике приводит к отказу всей ССС. Тогда с учетом независимости отказов на различных спутниках можно полагать в соответствии с (6.7) и (6.8), что распределение отказов для системы из n спутников имеет вид

$$P(t) = \exp(-\lambda_0 t), \quad (6.26)$$

где

$$\lambda_0 = \sum_{i=1}^n \lambda_i = n\lambda.$$

При этом вероятность безотказной работы системы в течение всего времени t_s ее эксплуатации (при отсутствии замены отказавших спутников)

$$P(t_s) = \exp(-\lambda_0 t_s) = \exp(-n\lambda t_s).$$

Обычно значения n и λ оказываются такими, что вероятность $P(t_s)$ недостаточна мала. Например, если $\lambda = 0,1$ 1/год, то при $t_s = 10$ лет и $n = 4$ $P(t_s) = e^{-4} = 0,018$. Поэтому для обеспечения нормального функционирования ССС в течение $t_s = 10$ лет необходимо обеспечить замену выходящих из строя спутни-

ков резервными. Пусть резервные спутники находятся на Земле в состоянии ненагруженного резерва и в этом состоянии вероятность их отказа равна нулю. Система обнаружения отказов спутников работает идеально. При отказе любого из n работающих на орбите спутников с Земли запускается на орбиту резервный спутник (вероятность одновременного выхода из строя двух или более работающих спутников можно полагать пренебрежимо малой). Этот спутник запускается ракетой-носителем, причем полагается, что каждая ракета запускает всего один спутник. Резервный спутник после выхода на требуемую орбиту включается в работу.

Пусть вероятность успешного запуска резервного спутника на орбиту и включения его в нормальное функционирование равна P_0 . В процессе дальнейшего функционирования этот спутник может отказаться еще до окончания срока t_s эксплуатации системы. Поэтому для его замены может потребоваться запуск дополнительного спутника из находящихся в состоянии ненагруженного резервирования, который, в свою очередь, может в дальнейшем отказаться. В таких условиях для обеспечения нормального функционирования ССС в течение времени t_s может потребоваться N запусков резервных спутников, где N — случайное число.

Действительно, существует некоторая вероятность $P(t_s)$ того, что за t_s лет ни один из работающих n спутников не выйдет из строя и, следовательно, не потребуются ни одного запуска резервного спутника ($N=0$). Но с вероятностью $1 - P(t_s)$ за время t_s выйдет из строя один или более спутников, а для замены даже одного спутника требуется один или более запусков резервных спутников. Следовательно, возможны (хотя и весьма маловероятны) и такие ситуации, когда потребуются весьма много (например, 100 или более) запусков резервных спутников. Спрашивается, как в такой ситуации выбрать требуемое число N_{tr} комплектов резервных спутников и запускающих их ракет-носителей (каждый комплект состоит из одного спутника и одной ракеты-носителя). Здесь возможны, в частности, следующие два подхода.

1. Выбираем

$$N_{tr} = M[N], \quad (6.27)$$

где $M[N]$ — математическое ожидание (среднее статистическое значение) числа N требуемых запусков.

2. Выбираем

$$N_{tr} = N_p, \quad (6.28)$$

где N_p — число запусков, обеспечивающее с вероятностью P_0 нормальное функционирование системы в течение времени t_s .

Преимущества и недостатки каждого из этих подходов мы отметим далее, после того как будет рассмотрена методика вычисления $M[N]$ и N_p .

Очевидно,

$$M[N] = M[N_s]/P_0, \quad (6.29)$$

где $M[N_s]$ — математическое ожидание требуемого числа замен спутников, отказавших в процессе их функционирования на орбите. Как уже отмечалось, можно полагать, что число замен равно числу отказов в системе из n спутников. Поэтому $M[N_s]$ равняется математическому ожиданию числа отказов в системе из n спутников. При принятом предположении об экспоненциальном законе распределения отказов (6.26)

$$M[N_s] = \lambda_c t_s \quad (6.30)$$

и, следовательно,

$$M[N] = \lambda_c t_s / P_s \quad (6.31)$$

При этом (6.27) принимает вид

$$-N_{\text{т.р.}} = \lambda_c t_s / P_s \quad (6.32)$$

Например, при $P_s = 0,8$ и $\lambda_c t_s = 4$ получается

$$N_{\text{т.р.}} = 5 \quad (6.33)$$

Определим теперь значение N_p — число запусков, обеспечивающее с вероятностью P_Φ нормальное функционирование системы в течение t_s лет. Эту задачу будем решать в следующем порядке.

Сначала найдем зависимость вероятности P_Φ от располагаемого числа запусков N_p , т. е. зависимость вида

$$P_\Phi = f(N_p) \quad (6.34)$$

Затем определим значение аргумента N_p , обеспечивающего требуемое значение функции P_Φ .

Итак, пусть имеется N_p резервных комплектов спутников и ракет-носителей, обеспечивающих возможность осуществления N_p запусков резервных спутников. Определим, с какой вероятностью P_Φ при этом будет обеспечено нормальное функционирование ССС из n спутников. Очевидно, нормальное функционирование ССС будет обеспечено, если в течение времени t_s потребуются запустить не более N_p резервных спутников (так как в противном случае отказавший спутник не удастся заменить резервным). Поэтому

$$P_\Phi = \sum_{j=0}^{N_p} P_j \quad (6.35)$$

где P_j — вероятность того, что за t_s лет потребуется произвести ровно j запусков. Но

$$P_j = \sum_{i=0}^j P_i P_{ji} \quad (6.36)$$

где P_i — вероятность того, что в ССС за время t_s произойдет ровно i отказов, а P_{ji} — вероятность того, что для замены i отказавших спутников потребуется ровно j запусков резервных спутников. Поскольку было принято, что отказы в ССС подчиняются экспоненциальному закону распределения (6.26), вероятность P_i подчиняется закону Пуассона, т. е.

$$P_i = \frac{(\lambda_c t_s)^i}{i!} \exp(-\lambda_c t_s) \quad (6.37)$$

Поэтому остается вычислить входящую в (6.36) условную вероятность P_{ji} . Покажем, что в рассматриваемой задаче

$$P_{ji} = C^{i-1}_{j-1} P_s^{i-1} (1-P_s)^{j-i} \quad (6.38)$$

где C^{i-1}_{j-1} — число сочетаний по $(i-1)$ элементов из $(j-1)$ элементов, т. е.

$$C^{i-1}_{j-1} = \frac{(j-1)!}{(i-1)! [(j-1)-(i-1)]!} = \frac{(j-1)!}{(i-1)! (j-i)!} \quad (6.39)$$

Действительно, P_{ji} есть вероятность события, состоящего в том, что для замены i отказавших спутников потребуется ровно j запусков резервных спутников. Очевидно, такое событие имеет место при совместном наступлении следующих двух событий:

имели место i успешных запусков;

имели место $(j-i)$ неуспешных запусков (т. е. отказов в процессе вывода резервного спутника на орбиту и включения его в нормальное функционирование).

Поскольку отказы при различных запусках статистически независимы, вероятность первого события равна P_s^i , второго $(1-P_s)^{j-i}$ и совместная вероятность обоих событий равна $P_s^i (1-P_s)^{j-i}$. Наличие в (6.38) кроме указанного выражения еще множителя C^{i-1}_{j-1} учитывает то обстоятельство, что при данных i и j число (несовместимых) ситуаций, при которых возможны ровно i успешных и $(j-i)$ неуспешных запусков, равно C^{i-1}_{j-1} .

Чтобы убедиться в этом, рассмотрим следующий простой пример: $i=2$, $j=3$, т. е. найдем вероятность события, состоящего в том, что для обеспечения замены двух отказавших на орбите спутников потребуется ровно три запуска с Земли резервных спутников. Возможные при этом ситуации изображены на рис. 6.6, где цифры на оси (1, 2, 3) указывают номера запусков резервных спутников, а 1 и 0 обозначают соответствующие этим запускам успешный (1) и неуспешный (0) исход: при успешном исходе спутник выводится на орбиту и начинает нормально функционировать вместо отказавшего на орбите; при неуспешном — происходит отказ в процессе вывода резервного спутника на орбиту или при включении его (сразу после вывода) в нормальное функционирование. На рис. 6.6,а изображена ситуация, в которой двумя успешными запусками ($i=2$) оказались 2-й и 3-й; на рис. 6.6,б — ситуация, в которой двумя успешными запусками оказались 1-й и 3-й. На рис. 6.6,в изображена ситуация, в которой двумя успешными запусками оказались 1-й и 2-й. Но эта ситуация должна быть исключена из рассмотрения. Действительно, если первые два запуска были успешными, то 3-го запуска не потребуется (так как $i=2$, т. е. требуется произвести всего две замены). Поэтому ситуация на рис. 6.6,в соответствует случаю, когда для осуществления двух замен ($i=2$) требуется два запуска ($j=2$), а не три ($j=3$). Мы же ищем вероятность события, при котором для двух замен ($i=2$) требуется ровно три запуска ($j=3$).

Следовательно, в рассматриваемом случае ($i=2$, $j=3$) получается, что возможны всего две ситуации (ситуации 1 и 2), при которых требуется три запуска для обеспечения двух замен. Формула (6.39) для этого случая ($i=2$, $j=3$) дает именно такой результат:

$$C^{i-1}_{j-1} = \frac{(3-1)!}{(2-1)! (3-2)!} = 2.$$

Из приведенного примера понятно, почему в (6.38) входит число сочетаний, равное не C^i_j , а C^{i-1}_{j-1} : это обусловлено тем, что последний запуск не может

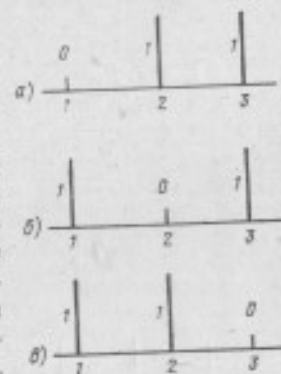


Рис. 6.6

быть неуспешным — он обязательно должен быть успешным (именно по этой причине пришлось исключить ситуацию 3 на рис. 6.6,в). Действительно, если (последний) запуск производился, значит он был необходим; но если он оказался неуспешным, значит потребуется следующий запуск.

На основе аналогичных рассуждений нетрудно убедиться, что в (6.39) следует полагать

$$C^{-1}_{-1}=1, C^{-1}_{-1}=0, C^{-1}_{j-1}=0 \text{ при } j>0. \quad (6.40)$$

Подставляя найденные выражения (6.36) — (6.38) в (6.35), получаем

$$P_0 = \sum_{j=0}^{N_p} \sum_{i=0}^j \frac{(\lambda_c t_0)^i}{i!} e^{-\lambda_c t_0} C_{j-1}^{i-1} P_2^i (1-P_2)^{j-i}, \quad (6.41)$$

где

$$C_{j-1}^{i-1} = \frac{(j-1)!}{(i-1)!(j-i)!},$$

причем $C^{-1}_{-1}=1, C^{-1}_{-1}=0, C^{-1}_{j-1}=0$ при $j>0$.

Из (6.41) следует, что в искомой зависимости (6.34) $P_0(N_p)$ параметрами являются значения λ_c и t_0 и характеристика надежности устройства вывода резервных спутников на орбиту (P_2). Поэтому при известных λ_c, t_0 и P_2 по формуле (6.41) можно построить зависимость P_0 от N_p и по этой зависимости определить значение N_p , требуемое для обеспечения заданной вероятности P_0 .

Для иллюстрации по формуле (6.41) вычислены значения P_0 , приведенные в табл. 6.1 для $P_2=0,8; \lambda_c t_0=4$. Из таблицы следует, что для обеспечения вероятности нормального функционирования ССС, не меньшей 0,9, требуется располагать числом запусков резервных спутников $N_{rp}=N_p=9$. Если же потребовать, чтобы вероятность нормального функционирования ССС была не менее 0,99, то получится $N_{rp}=12$. Поскольку запуск каждого спутника обходится весьма дорого, желательно обосновать, каким значением вероятности P_0 целесообразно задаваться.

Таблица 6.1

N_p	0	1	2	4	5	6	8	9	10	11	12	∞
P_0	0,018	0,067	0,170	0,474	0,611	0,726	0,876	0,920	0,960	0,982	0,993	1

Если выбрать P_0 большой, например 0,99, то значительно возрастут расходы на изготовление резервных комплектов. В то же время с большой вероятностью фактически за t_0 лет будет израсходовано значительно меньшее число спутников и, следовательно, запас спутников окажется слишком большим. Если же выбрать P_0 недостаточно большой, например $P_0=0,6$, то будет слишком велика вероятность срыва нормального функционирования системы за $t_0=10$ лет. С учетом этих соображений, по-видимому, в большинстве случаев близким к оптимальному является значение $P_0=0,8 \dots 0,9$. При этом вероятность того, что изготовленных заранее резервных комплектов спутников и столько же ракет-носителей не хватит на весь длительный срок $t_0=10$ лет функ-

ционирования системы, сравнительно невелика (0,1...0,2). Кроме того, срыв функционирования системы при этом с весьма высокой вероятностью может быть предотвращен, если предусматривать возможность изготовления в течение ближайших трех — пяти лет дополнительных комплектов спутников и ракет-носителей, если окажется, что за эти годы пришлось израсходовать уже большую часть заранее изготовленных комплектов.

Следовательно, допустимое значение P_0 зависит от того, целесообразно ли и в какой мере предусмотреть возможность пополнения (в случае необходимости) числа резервных комплектов спутников и ракет-носителей в процессе длительного времени t_0 функционирования системы.

Итак, в рассмотренном примере при $P_0=0,8 \dots 0,9$

$$N_{rp}=8 \dots 9. \quad (6.42)$$

Для тех же данных ($P_2=0,8; \lambda_c t_0=4$) выше было найдено, что математическое ожидание требуемого числа спутников $N_{rp}=M[N]=5$.

Из табл. 6.1 следует, что если выбрать число резервных запусков (комплектов) равными математическому ожиданию, то вероятность того, что такого числа запусков (комплектов) хватит для замены отказывающих спутников в течение всего срока $t_0=10$ лет, будет равна всего 0,611, т. е. недостаточно высока. Поэтому определять требуемое число запусков по формуле (6.27) в общем случае (в частности, при невозможности изготовления или покупки в случае необходимости дополнительных резервных комплектов) недопустимо.

Основным преимуществом использования формулы (6.27) вместо (6.28) является значительно большая простота вычислений. Например, при принятой экспоненциальной аппроксимации закона распределения отказов (6.27) принимает весьма простой вид (6.32), тогда как в случае применения (6.28) приходится использовать выражение (6.41). Кроме того, следует иметь в виду, что в ряде случаев экспоненциальная аппроксимация закона распределения отказов может оказаться слишком грубой (например, если условие $T \gg t_0$ не выполняется). При этом существенно усложняется вычисление математического ожидания $M[N]$ и особенно вероятности P_0 . Например, для закона (6.10) формулу для $M[N]$ получить удалось [57], а для P_0 соответствующее выражение пока не получено. В то же время, как следует из (6.33) и (6.42) и других аналитических соотношений, обычно $N_p/M[N] < 2$. Поэтому на начальных этапах проектирования, если не удастся вычислить N_p , в ряде задач допустимо принимать в качестве первого приближения

$$N_{rp} \approx (1,5 \dots 2) M[N]. \quad (6.43)$$

Поскольку в рассмотренном примере для обеспечения нормального функционирования системы предусматривается замена отказавших спутников резервными, т. е. восстановление системы, необходимо определить среднее время восстановления T_v и коэффициент готовности K_r . Как уже отмечалось, T_v и K_r в первом приближении связаны соотношениями (6.12) — (6.14), причем в данном случае

$$T_{cp} = 1/\lambda_c = 1/n\lambda.$$

Время восстановления T_v при запуске резервных спутников с Земли определяется в основном временем, необходимым для вывода спутника на требуемую (геостационарную) орбиту (и притом в заданный участок этой орбиты), и в зависимости от места и условий запуска может составлять несколько дней

или более. Пусть, например, $T_p = 10$ дней; тогда при $\lambda_0 = 0,1$ 1/год оказывается $T_{cp} = 10 \text{ лет} = 3650$ дней и $K_r = 1/(1+1/365) = 0,997$, т. е. $1 - K_r \approx 0,003$.

Как видно, в данном случае коэффициент готовности сам по себе достаточно высок; однако соответствующее ему среднее время восстановления $T_p = 10$ дней для ряда применений ССС может быть недопустимо большим. Поэтому в данном примере определяющим параметром является не K_r , а T_p .

Наиболее радикальным путем уменьшения T_p является размещение резервных спутников не на Земле, а на той же орбите (и на тех же ее участках), что и рабочие спутники; при этом время восстановления может быть уменьшено на два порядка и более. Однако при этом могут возрасти расходы на эксплуатацию системы. Действительно, если резервные спутники находятся на Земле, то запуск каждого резервного спутника на орбиту производится только в случае отказа функционирующего спутника. Поэтому с большой вероятностью за время t_0 потребуются фактически запустить не все резервные спутники и, более того, с некоторой (правда малой) вероятностью удастся обойтись вообще без запуска. Если же резервные спутники должны находиться на орбите, их запуск производится заранее, независимо от того, будут или не будут иметь место отказы функционирующих спутников.

Из приведенного примера видно, что как расчет надежности системы, так и выбор способов ее повышения могут быть весьма сложными и не всегда существует строгое и однозначное решение.

Литература: основная [54], дополнительная [55—57].

Глава 7

СТОИМОСТЬ, МАССА И ОБЪЕМ АППАРАТУРЫ

7.1. ВВЕДЕНИЕ

Расходы на проектирование, производство и эксплуатацию радиоэлектронных систем и устройств являются одной из самых крупных статей расходов как в СССР, так и в других индустриальных странах мира. С каждым годом доля этих расходов возрастает в связи с быстрым расширением областей применения радиосредств и усложнением выполняемых ими функций. При этом основная часть расходов приходится на два класса радиоаппаратуры: сравнительно недорогая, но массовая аппаратура и весьма сложные и дорогие системы мелкосерийного или даже уникального типа. Примером аппаратуры первого класса являются радиовещательные приемники и телевизоры. В СССР их выпускается более 15 млн. экз. в год на несколько миллиардов рублей. Типичные примеры аппаратуры второго класса — радиолокационные и радионавигационные системы и спутниковые системы связи.

Снижение стоимости радиоэлектронной аппаратуры всего на 5—10% дает государству экономию в несколько миллиардов рублей в год. Поэтому уменьшение стоимости является одним из важнейших требований при проектировании радиоэлектронных систем и устройств. Очевидно, что стоимость должна снижаться без ущерба для остальных показателей качества радиосистем (или радиоустройств), т. е. должна решаться следующая задача:

$$\text{обеспечить } \min_{S \in M_{SD}} \Pi \text{ при } K_r \in M_{K_d}. \quad (7.1)$$

Здесь Π — полные (приведенные) затраты на проектирование, производство, эксплуатацию и утилизацию системы; M_{SD} — множество возможных (допустимых исходными данными) вариантов построения системы; M_{K_d} — множество допустимых (приемлемых с точки зрения назначения системы) значений вектора K_r технического качества системы, т. е. совокупности $\langle k_1, \dots, k_{m-1} \rangle$ всех существенных показателей качества, кроме стоимости.

Задача (7.1) называется задачей обеспечения технико-экономической эффективности системы, так как при ее решении обеспечивается выполнение системой поставленных перед ней задач (т. е. достижение заданного технического эффекта) при минимальных затратах. Повышение технико-экономической эффективности радиосистем и устройств является одной из важнейших составляющих повышения технико-экономической эффективности народного хозяйства в целом.

Расчеты затрат на проектирование, производство, эксплуатацию и утилизацию радиосистем нужны для нескольких целей.

1. Для решения задачи (7.1), т. е. для выбора оптимального варианта построения проектируемой системы.
2. Для оценки годового экономического эффекта (ГЭЭ), достигаемого при применении найденного оптимального варианта построения системы вместо прототипа (прототипов), т. е. наилучшего из уже существующих (или ранее спроектированных) вариантов. Очевидно, эта задача может решаться лишь в случаях, когда прототип удается найти.
3. Для определения стоимости уже изготовленного изделия с целью установления его оптовой цены.

Как будет ясно из дальнейшего, приведенные затраты Π на систему в значительной мере зависят от принципов ее построения; поэтому задачу (7.1) лучше решать, хотя бы весьма приближенно, на ранних этапах проектирования, когда выбираются принципы построения системы. Оценка ГЭЭ также весьма желательна при оптимальном выборе варианта построения системы.

Стоимость уже изготовленного объекта можно определить лишь после завершения проектирования или в самом его конце, и к точности экономических расчетов при этом предъявляются наибольшие требования.

С учетом специфики указанных трех видов экономических расчетов первый из них [т. е. решение задачи (7.1)] решается

в основном радиоинженерами (с привлечением экономистов); последний — в основном экономистами, а второй (определение ГЭЭ) — радиоинженерами и экономистами совместно. В дальнейшем изложении рассматриваются лишь первые два вида расчетов.

7.2. ОПРЕДЕЛЕНИЕ ПРИВЕДЕННЫХ ЗАТРАТ

Как уже отмечалось, полные затраты включают стоимость разработки (проектирования), производства, эксплуатации и утилизации системы (устройства). Затраты на разработку предшествуют началу процесса производства (выпуску продукции), и обычно их включают в состав общих капитальных вложений K , которые должны быть сделаны для обеспечения последующего процесса производства. Поэтому затраты на разработку в конечном счете входят (в составе капитальных вложений K) в общую стоимость производства.

Затраты на утилизацию системы значительно меньше совокупности затрат на ее производство и эксплуатацию. Поэтому на начальных этапах проектирования при выборе принципов построения системы этими затратами можно пренебречь. Однако на дальнейших этапах проектирования их необходимо принимать во внимание. Таким образом, при выборе оптимального варианта построения системы, проводимого на начальных этапах проектирования, в первом приближении достаточно учитывать лишь два вида затрат — на производство системы (включая разработку) и ее эксплуатацию.

Рассмотрим сначала полные (приведенные) затраты на производство массовой аппаратуры (например, телевизоров). Их можно разбить на три основные составляющие:

капитальные вложения K ;

текущие годовые расходы v на зарплату работникам, участвующим в процессе производства аппаратуры в количестве A единиц (изделий) в год (не включая оплату работников, участвовавших в разработке системы, поскольку, как уже отмечалось выше, эти затраты включают в состав капитальных вложений K); прочие текущие годовые расходы c_0 на производство A единиц в год (расходы на материалы, детали, топливо, электроэнергию, текущий и капитальный ремонт и др.).

Здесь по определению K — это вложения, которые делаются до начала процесса производства (выпуска) аппаратуры. В дальнейшем время будем отсчитывать от момента начала производства, принятого равным нулю. K — это вложения, сделанные до момента $t=0$, а v и c_0 — после этого момента.

Вложения K должны обеспечить процесс производства аппаратуры в течение всего срока T функционирования данного производства. В дальнейшем для простоты будем полагать процесс производства стационарным, т. е. считать, что объем производства A и затраты c_0 и v по годам не изменяются в течение всех

T лет. Капитальные вложения K имеют две основные составляющие:

$$K = K_{ос} + K_{об}, \quad (7.2)$$

где $K_{ос}$ и $K_{об}$ — вложения в основные производственные фонды и в оборотные соответственно. Обычно $K_{об}$ значительно (в несколько раз) меньше $K_{ос}$, поэтому в интересующей нас задаче (выбор оптимального варианта на начальном этапе проектирования) можно в первом приближении полагать

$$K \approx K_{ос}, \quad (7.3)$$

где $K_{ос}$ — затраты на разработку (проектирование) выпускаемых изделий, строительство зданий и сооружений, необходимых для обеспечения производства A изделий в год, закупку, установку и наладку необходимых для той же цели машин и оборудования¹.

Обычно отдельные составляющие затрат $K_1, \dots, K_n$, входящие в K , производятся в различное время. Однако приближенно их можно свести к единым суммарным затратам K , производимым в момент $t=0$, полагая

$$K = q_1 K_1 + \dots + q_i K_i + \dots + q_n K_n, \quad (7.4)$$

где q_i ($i = \overline{1, n}$) — некоторые весовые коэффициенты, учитывающие влияние фактора времени (т. е. неодновременности осуществления затрат) и называемые *коэффициентами приведения капитальных вложений по фактору времени*. Методика определения коэффициентов q_i будет указана далее.

Теперь остается выяснить, как следует суммировать разовые капитальные вложения K (осуществляемые в момент $t=0$) с ежегодными текущими годовыми расходами c_0 и v . Очевидно, простое суммирование затрат K с c_0 и v недопустимо, так как они не только различны по виду, но и имеют различные размерности: K измеряется в рублях, а c_0 и v — в рублях в год. Поэтому, принимая приближенно, что все затраты можно привести к единому показателю, называемому *приведенными затратами* Π и являющемуся линейной комбинацией затрат c_0 , v и K , полагают

$$\Pi = c_0 + v + \gamma K, \quad (7.5)$$

где γ — некоторый весовой коэффициент, учитывающий различный характер капитальных (разовых) вложений K и текущих (годовых) затрат $c_0 + v$.

При определении приведенных затрат Π часто оказывается удобным использовать понятие себестоимости с годового производства изделий, где по определению

$$c = c_0 + v + P_a K. \quad (7.6)$$

¹ Если некоторая часть основных фондов будет использоваться для производства не только вновь разработанных, но и «старых» (т. е. выпускавшихся ранее) изделий, то стоимость этих фондов должна соответственно распределяться между затратами на новые и «старые» изделия.

Здесь c_0 , v и K имеют тот же смысл, что и в (7.5), а P_a — коэффициент (доля) годовых амортизационных отчислений на реновацию (полное восстановление) основных производственных фондов [в силу принятого допущения (7.3) здесь полагается, что $K \approx K_{ос}$]. Очевидно, $P_a K$ является годовым амортизационным отчислением в государственный фонд, обеспечивающим в течение срока T действия капитальных вложений их полное восстановление (реновацию). Коэффициент P_a полагается неизменным во времени.

С учетом (7.6) выражение (7.5) для приведенных затрат принимает вид

$$P = c + E_n K, \quad (7.7)$$

где

$$E_n = \gamma - P_a \quad (7.8)$$

— нормативный коэффициент эффективности капитальных вложений, определяемый на основе анализа экономических закономерностей развития народного хозяйства. Для определения приведенных затрат рекомендуется¹, например, полагать

$$E_n = 0,15 \text{ 1/год}, \quad (7.9)$$

$$P_a = \frac{E}{(1+E)^T - 1} \approx \frac{1}{T}, \quad (7.10)$$

$$E = 0,1 \text{ 1/год}, \quad (7.11)$$

где T — срок действия капитальных вложений (в годах).

Из (7.8) следует, что

$$\gamma = E_n + P_a. \quad (7.12)$$

Соотношение (7.7) позволяет дополнительно прояснить смысл приведенных затрат P . Действительно, входящая в (7.7) себестоимость c учитывает, как следует из (7.6), все виды затрат, осуществляемых ежегодно в процессе производства продукции A , включая взносы $P_a K$ в государственный фонд, обеспечивающие реновацию капитальных вложений. Поэтому, если бы в народном хозяйстве в целом ресурсы, отводимые на капитальные вложения, были неограниченными, то оптимальным следовало бы считать тот вариант производимой системы, при котором ее себестоимость c минимальна, т. е. полагать $P = c$. Однако ограниченность указанных ресурсов приводит к появлению в составе приведенных затрат P дополнительного слагаемого $E_n K$, играющего роль дополнительного «штрафа» за отвлечение из общегосударственных фондов ресурса K . Очевидно, этот «штраф» должен быть тем большим, чем больше ценность (эффективность) для народного хозяйства отвлекаемого ресурса. Поэтому коэффициент E_n действительно характеризует эффективность капитальных вложений.

¹ Методика определения экономической эффективности использования в народном хозяйстве новой техники, изобретений и рационализаторских предложений. — М.: Экономика, 1977. — 41 с.

Итак, если не учитывать затрат, осуществляемых в процессе эксплуатации произведенных изделий (продукции), то оптимальным следует считать тот вариант, при котором приведенные затраты P , определяемые соотношением (7.7) или эквивалентным ему соотношением (7.5), оказываются минимальными. Здесь P — затраты на годовую продукцию в объеме A единиц. Если у сравниваемых вариантов объемы годовой продукции не одинаковы, то приходится сравнивать удельные приведенные затраты

$$P' = P/A, \quad (7.13)$$

т. е. приведенные затраты на единицу продукции.

Если сравниваемые варианты системы (изделия) отличаются затратами, которые необходимо производить в процессе эксплуатации произведенной продукции (изделий), то для выбора оптимального варианта необходимо использовать такое выражение приведенных затрат P_p , которое учитывает затраты не только в сфере производства, но и при эксплуатации. В общем случае выражение для результирующих приведенных затрат P_p оказывается весьма громоздким и еще более грубым, чем для затрат P , поскольку приходится принимать целый ряд дополнительных гипотез и идеализаций [59, 61]. Поэтому ограничимся лишь двумя простейшими, но важными случаями.

1. Требуется выбрать оптимальный вариант построения изделия (системы, устройства), выпускаемого (производимого) в объеме A изделий в год и рассчитанного на большой срок эксплуатации T_p ($T_p \gg 1$ г.). При этом значение объема A производства и срока эксплуатации T_p при различных вариантах построения изделия заданы и одинаковы. В процессе эксплуатации никаких дополнительных капитальных вложений не требуется. Текущие годовые расходы в процессе эксплуатации равны I и по годам эксплуатации не изменяются, но могут быть различными для различных вариантов построения изделия. Здесь предполагается, что расходы I включают все виды текущих затрат в процессе эксплуатации, но не включают годовые амортизационные отчисления на реновацию. При сделанных допущениях выражение для результирующих приведенных затрат P_p на производство и эксплуатацию одного изделия можно получить следующим образом. Для обеспечения эксплуатации данного изделия в течение T_p лет необходимо:

а) произвести изделие, т. е. осуществить затраты $P' = P/A$;

б) производить в процессе эксплуатации ежегодные затраты v_a на рабочую силу и прочие годовые расходы c_{0a} . При этом с точки зрения обеспечения процесса эксплуатации затраты P' играют роль предварительных (разовых) капитальных вложений, а ежегодные затраты c_{0a} , v_a — роль текущих вложений, осуществляемых равномерно в течение времени T_p , начиная с момента начала эксплуатации. Следовательно, необходимо суммировать разовые (капитальные) вложения P' с текущими годовыми вложениями c_{0a} , v_a .

Очевидно, эта задача полностью аналогична рассмотренной применительно к процессу производства задаче суммирования капитальных вложений K и текущих годовых вложений c_0 и v . Поэтому в данном случае по аналогии с (7.5) можно полагать

$$P_p = c_{00} + v_0 + \gamma_0 P', \quad (7.14)$$

где в соответствии с (7.12) и (7.10)

$$\gamma_0 = P_{a0} + E_n, \quad (7.15)$$

$$P_{a0} = \frac{E}{(1+E)^{T_0}-1} \approx \frac{1}{T_0}. \quad (7.16)$$

Обозначим через

$$H = c_{00} + v_0, \quad (7.17)$$

годовые расходы на эксплуатацию одного изделия (не включающие отчисления на реновацию капитальных вложений). Тогда (7.14) принимает вид

$$P_p = H + (P_{a0} + E_n) P' = \left(P' + \frac{H}{P_{a0} + E_n} \right) (P_{a0} + E_n). \quad (7.18)$$

Так как при выводе соотношения (7.18) было принято, что у всех сравниваемых вариантов срок эксплуатации T_0 одинаков, то значение P_{a0} (как и E_n) не будет изменяться при переходе от одного варианта к другому. Поэтому минимум P_p будет совпадать с минимумом величины

$$P'_p = P' + H/(P_{a0} + E_n). \quad (7.19)$$

С учетом (7.5), (7.12) и (7.16) формулу (7.19) можно представить в следующем развернутом виде:

$$P'_p = \frac{c_0 + v + (P_a + E_n) K}{A} + \frac{c_{00} + v_0}{P_{a0} + E_n}, \quad (7.20)$$

где

$$E_n = 0,15; \quad P_a = 0,1/(1,1^T - 1); \quad P_{a0} = 0,1/(1,1^{T_0} - 1). \quad (7.21)$$

Рассмотрим в качестве иллюстрации следующий модельный пример. Требуется выбрать лучший из двух вариантов построения телевизора. В обоих вариантах телевизоры будут иметь одинаковые технические и (эстетические) характеристики, включая одинаковый срок службы $T_0 = 10$ лет. Ежегодно должно выпускаться 100 тыс. телевизоров данного класса, т. е. $A = 10^6$ штук в год. В первом варианте требуются следующие затраты:

$$K = 150 \cdot 10^6 \text{ р.}, \quad v = 5 \cdot 10^6 \text{ р.}, \quad c_0 = 10 \cdot 10^6 \text{ р.}, \quad c_{00} = 40 \text{ р.}, \quad v_0 = 20 \text{ р.}$$

Во втором варианте

$$K = 200 \cdot 10^6 \text{ р.}, \quad v = 3 \cdot 10^6 \text{ р.}, \quad c_0 = 7 \cdot 10^6 \text{ р.}, \quad c_{00} = 20 \text{ р.}, \quad v_0 = 10 \text{ р.}$$

Срок действия капитальных вложений T в обоих случаях равен 10 лет. Так как в данном случае $T_0 = T = 10$ лет, то $P_a = P_{a0} = 0,063$. При этом для первого варианта из (7.20) получается, что $P'_p = 800$ р., а во втором варианте $P'_p = 690$ р. Следовательно, второй вариант лучше первого, несмотря на то,

что во втором варианте капитальные вложения K ($200 \cdot 10^6$ р.) и общие приведенные затраты на производство одного изделия (530 р.) больше, чем в первом варианте ($150 \cdot 10^6$ р. и 480 р. соответственно). Правда, следует иметь в виду, что при выводе формулы (7.20) сделан целый ряд допущений. Кроме того, нужно учитывать, что на начальном этапе проектирования значения ожидаемых затрат K , v , c_0 , c_{00} и v_0 также могут быть определены лишь приближенно. Влияние этого последнего фактора будет рассмотрено далее (§ 7.3—7.6).

2. Требуется выбрать оптимальный вариант построения уникального изделия (системы, устройства). Целый ряд важных радиосистем изготавливается в единственном экземпляре или всего в нескольких экземплярах. Типичными примерами таких радиосистем и радиоустройств являются спутниковые системы связи, система управления воздушным движением страны, уникальный радиотелескоп или какой-либо другой уникальный радиоизмерительный прибор и др. При этом невозможно или нецелесообразно считать, что производство состоит в ежегодном изготовлении A единиц (изделий). Создание каждого экземпляра такой системы может занимать 3—5 лет или более. Поэтому целесообразно рассматривать процесс производства системы как последнюю стадию ее разработки (проектирования), а процесс последующей длительной эксплуатации T_0 (функционирования) созданной системы считать производством (этой системой) услуг в объеме A единиц в год.

В дальнейшем ради краткости и наглядности такой случай проанализируем на примере спутниковой системы связи, создаваемой в единственном экземпляре. Процесс эксплуатации (функционирования) этой системы можно рассматривать как производство услуг (отдельным лицам, организациям и обществу в целом) в объеме A единиц в год. Здесь единицей услуги («выпускаемой продукции») может служить бит (двоичная единица) информации, канал-час или какая-либо другая подходящая единица измерения количества («объема») передаваемой информации. При этом под приведенными затратами P следует понимать затраты, необходимые для обеспечения передачи A единиц информации в год в течение всего срока T_0 эксплуатации (функционирования) системы. Эти затраты состоят из капитальных вложений K_0 , реализуемых до начала эксплуатации системы, и текущих ежегодных расходов v_0 и c_{00} , осуществляемых в процессе эксплуатации соответственно на зарплату и прочих (затрат на электроэнергию, топливо, текущий и капитальный ремонт и других, исключая амортизационные отчисления на реновацию капитальных вложений).

В капитальные вложения K_0 входят затраты на проектирование системы, строительство зданий и сооружений, покупку и установку необходимых радиопередающих и радиоприемных устройств и другого оборудования, монтаж, настройку, регулировку и испытания системы, включая испытательные запуски спутников.

Поскольку K_0 , v_0 и c_0 рассматриваются в данном случае как затраты на производство продукции в объеме A единиц в год, для их суммирования можно использовать выражение (7.5) или (7.7) для приведенных затрат на производство продукции в объеме A единиц в год, в котором c_0 , v_0 , K и c заменены соответственно на c_{00} , v_0 , K_0 и c_0 . Следовательно, приведенные затраты на ССС, обеспечивающую передачу информации в объеме A единиц (информации) в год,

$$P = c_{00} + v_0 + \gamma K_0 \quad (7.22)$$

или, что эквивалентно,

$$P = c_0 + E_K K_0, \quad (7.23)$$

где

$$c_0 = c_{00} + v_0 + P_{00} K_0, \quad (7.24)$$

$$P_{00} = E/[1 + E)^{T_0} - 1] \quad (7.25)$$

— коэффициент годовых отчислений на реновацию системы. Коэффициенты γ , E_K и E имеют те же значения, что и ранее.

Из отмеченного состава капитальных вложений K_0 следует, что эти вложения $K_{10}, \dots, K_{n0}$ осуществляются практически не в один момент времени $t=0$, непосредственно предшествующий началу функционирования системы, а в течение ряда лет, предшествующих этому моменту. Поэтому для вычисления суммарных капитальных вложений K_0 необходимо, как следует из (7.4), суммировать отдельные составляющие вложений с соответствующими весами, т. е. полагать

$$K_0 = q_1 K_{10} + \dots + q_n K_{n0}, \quad (7.26)$$

где

$$q_i = (1 + E)^{t_i} \approx 1,1^{t_i};$$

t_i — число лет, отделяющих момент вложения K_{i0} от момента $t=0$ (т. е. число предшествующих лет). Если t_i — нецелое число, то его обычно округляют до ближайшего целого.

Итак, в рассмотренном случае ССС затраты на ее эксплуатацию (функционирование) являются по существу затратами на производство этой системой продукции в объеме A единиц в год (например, A бит информации в год). При этом можно поставить вопрос: а не следует ли в этом случае включить в состав приведенных затрат также те дополнительные затраты, которые будут связаны с эксплуатацией выпускаемой «продукции», т. е. с использованием переданной через систему связи информации? Очевидно, эти дополнительные затраты зависят от того, в каком техническом виде ССС выдает информацию: в виде закодированного или незакодированного, буквенного или цифрового текста и т. п. Поэтому, строго говоря, если в процессе проектирования системы связи вид, в котором должна выдаваться информация, жестко не задан, т. е. может варьировать-

ся, такие дополнительные затраты действительно желательно учитывать при выборе оптимального варианта построения системы. Однако в большинстве практических задач вид, в котором должна выдаваться информация, задан исходными данными жестко (входит в состав совокупности условий работы системы), т. е. не подлежит выбору (вариации). При этом вводить указанные дополнительные затраты в состав полных приведенных затрат P на систему, используемых для выбора оптимального варианта ее построения, не требуется. Следовательно, в большинстве задач, связанных с проектированием уникальных (или почти уникальных) систем, приведенные затраты, определяемые соотношениями (7.23) — (7.26), достаточно полно учитывают все существенные виды затрат.

7.3. ОБЩАЯ ХАРАКТЕРИСТИКА ВЛИЯНИЯ СЛУЧАЙНОСТИ ЗАТРАТ

Пусть P_K и P — истинное (фактическое) и расчетное значения приведенных затрат соответственно. В процессе проектирования определяется лишь расчетное значение P . Поэтому важно оценить, в какой мере истинное значение может отличаться от расчетного. Это отличие можно характеризовать как разностью

$$\Delta P = P_K - P, \quad (7.27)$$

так и отношением

$$\eta = P_K/P. \quad (7.28)$$

При малых отличиях P_K от P удобнее оценивать отличие ошибкой ΔP , а при больших — отношением η . Из (7.27) и (7.28) следует, что

$$\eta = 1 + \Delta P/P. \quad (7.29)$$

Истинное (фактическое) значение приведенных затрат является случайной величиной, так как определяется совокупностью большого числа факторов, многие из которых должны или могут рассматриваться как случайные (например, увеличение затрат, вызываемое истощением природных ресурсов, ростом расходов на борьбу с загрязнением среды, необходимостью реконструкции заводов, неопределенностью срока морального износа аппаратуры и др.).

Ошибка ΔP расчета также случайна, поскольку обусловлена совокупным действием многих случайных факторов. Поэтому затраты P_K , P , ΔP и η , входящие в соотношения (7.27) и (7.28), являются непрерывными случайными величинами. В процессе проектирования задача учета случайности затрат обычно сводится в первую очередь к следующей: по известному расчетному значению P затрат оценить в вероятностном смысле истинное (фактическое) значение P_K . Наиболее полная информация о P_K при этом содержится в условном распределении $W(P_K|P)$, т. е. в

плотности вероятности величины P_n при данном (фиксированном, известном) значении P . Действительно, зная распределение $W(P_n|P)$, нетрудно найти вероятность

$$\alpha = P(P_n \leq \eta_\alpha P) \quad (7.30)$$

того, что P_n не превысит расчетного значения P более чем в η_α раз, где η_α — любое наперед заданное (положительное) число. Из (7.29) и (7.30) следует, что

$$\alpha = P(P_n \leq \eta_\alpha P) = \int_0^{\eta_\alpha P} W(P_n|P) dP_n. \quad (7.31)$$

Условное распределение $W(P_n|P)$ можно определить по формуле

$$W(P_n|P) = \frac{W(P, P_n)}{W(P)} = \frac{W(P, P_n)}{\int_0^\infty W(P, P_n) dP_n}. \quad (7.32)$$

Следовательно, для вычисления распределения $W(P_n|P)$ нужно знать совместное распределение (совместную плотность вероятности) $W(P, P_n)$.

Если из опыта предыдущих разработок систем, близких к проектируемой, известен достаточно представительный ансамбль данных о значениях фактических затрат P_n на эти системы и соответствующих расчетных данных P , то распределение $W(P, P_n)$ можно найти приближенно известными методами математической статистики [7]. Однако для упрощения практического применения весьма желательно свести задачу отыскания двумерного распределения $W(P, P_n)$ к определению некоторого одномерного распределения. Это может быть приближенно сделано следующим образом.

В (7.31) заменим переменную P_n на η в соответствии с соотношением (7.28). Тогда получим

$$\alpha = P(P_n \leq \eta_\alpha P) = \int_0^{\eta_\alpha} W(\eta|P) d\eta. \quad (7.33)$$

В большинстве практических задач можно полагать

$$W(\eta|P) \approx W(\eta), \quad (7.34)$$

т. е. считать, что распределение отношения $\eta = P_n/P$ не зависит от абсолютного значения затрат P (в пределах представляющих интерес значений P).

Так, если для системы данного класса можно полагать с высокой вероятностью, например равной 0,99, что при расчетном значении затрат $P = 100 \times 10^6$ р. фактические затраты P_n не превысят расчетные более чем в 2 раза (т. е. не превзойдут 200×10^6 р.), то обычно есть основание полагать, что примерно с той же вероятностью (0,99) при расчетной величине $P = 30 \cdot 10^6$ р. фактические затраты не превзойдут $30 \cdot 10^6 - 60 \cdot 10^6$ р.

С учетом допущения (7.34) формула (7.33) принимает вид

$$\alpha = P(P_n \leq \eta_\alpha P) = \int_0^{\eta_\alpha} W(\eta) d\eta, \quad (7.35)$$

т. е. для определения вероятности α достаточно знать лишь одномерное распределение $W(\eta)$, определить которое по выборочным данным значительно проще, чем двумерное распределение $W(P, P_n)$.

Следовательно, если удастся найти распределение $W(\eta)$ отношения $\eta = P_n/P$, то по (7.35) можно вычислить вероятность α выполнения неравенства

$$P_n \leq \eta_\alpha P \quad (7.36)$$

для заданного параметра η_α или, что эквивалентно, найти значение параметра η_α , соответствующего заданной (достаточно большой) вероятности α . При этом α называется *доверительной вероятностью*, а η_α — *коэффициентом доверия*, соответствующим данной доверительной вероятности α . Из (7.36) следует, что коэффициент доверия η_α позволяет найти максимально возможное значение

$$P_{n, \max \alpha} = \eta_\alpha P \quad (7.37)$$

фактических затрат при данном расчетном значении затрат P . Вероятность $(1-\alpha)$ превышения величиной P_n этого максимального значения достаточно мала, например равна 0,01, если $\alpha = 0,99$.

Рассуждая аналогично, можно получить следующие соотношения для оценки (с вероятностью α) возможного минимального значения истинных затрат:

$$\alpha = P\left(P_n \geq \frac{P}{\eta_\alpha}\right) = \int_{1/\eta_\alpha}^\infty W(\eta) d\eta. \quad (7.38)$$

Из этого выражения, задавая допустимым значением доверительной вероятности α , можно найти соответствующее значение коэффициента доверия η_α . Тем самым определяется, что с вероятностью α будет выполняться неравенство

$$P_n \geq P/\eta_\alpha, \quad (7.39)$$

т. е.

$$P_{n, \min \alpha} = P/\eta_\alpha. \quad (7.40)$$

Следовательно, коэффициент доверия η_α позволяет определить (с вероятностью α) минимально возможное значение истинных затрат при данном расчетном значении затрат.

Итак, если найдены коэффициенты доверия η_α^* и $\bar{\eta}_\alpha$, то тем самым определяются (с заданной достаточно высокой вероятностью α) максимальное и минимальное значения истинных затрат, соответствующие найденному расчетному значению P . Очевидно, вероятность β совместного выполнения неравенств (7.36) и (7.39), т. е. выполнения условия

$$P/\bar{\eta}_\alpha \leq P_n \leq \eta_\alpha^* P, \quad (7.41)$$

определяется соотношением

$$(1-\beta) = (1-\alpha) + (1-\alpha), \text{ т. е. } \beta = 2\alpha - 1. \quad (7.42)$$

Например, если $\alpha = 0,99$, то $\beta = 0,98$.

Интервал $[P/\bar{\eta}_\alpha, P\eta_\alpha^*]$, в котором с вероятностью β находится значение P_n (при данном значении P расчетных затрат) называется *доверительным*. Этим понятием удобно пользоваться при высокой точности расчета, когда коэффициенты доверия $\bar{\eta}_\alpha$ и η_α^* примерно одинаковы и близки к единице. Однако в общем случае величины η_α^* и $\bar{\eta}_\alpha$ могут существенно отличаться как между собой, так и от единицы. При этом более полно характеризуют значение P_n два отдельных неравенства (7.36) и (7.39) и соответственно два коэффициента доверия η_α^* и $\bar{\eta}_\alpha$.

7.4. ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ ДОВЕРИЯ

Рассмотрим сначала случай, когда значение P можно рассчитать с высокой точностью, т. е. выполняются условия

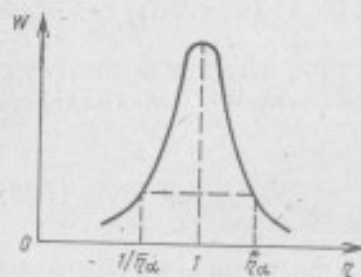


Рис. 7.1

$$M[\eta] = 1, \quad (7.43)$$

$$D_\eta \ll 1, \quad (7.44)$$

где $M[\eta]$ и D_η — соответственно математическое ожидание и дисперсия величины η .

В этом случае в первом приближении допустимо аппроксимировать $W(\eta)$ нормальным распределением (рис. 7.1), т. е. полагать

$$W(\eta) = (1/\sqrt{2\pi D_\eta}) \exp [-(\eta-1)^2/2 D_\eta]. \quad (7.45)$$

Подстановка этого выражения в (7.35) и (7.38) дает

$$\eta_\alpha^* = 1 + t_\alpha \sqrt{D_\eta}, \quad \bar{\eta}_\alpha = 1/(1 - t_\alpha \sqrt{D_\eta}), \quad (7.46)$$

где t_α определяется для заданного значения α из соотношения

$$\frac{1}{\sqrt{2\pi}} \int_{t_\alpha}^{\infty} e^{-t^2/2} dt = \alpha \quad (7.47)$$

или из эквивалентного ему соотношения

$$\left. \begin{aligned} \alpha &= [1 + \Phi(t_\alpha)]/2, \\ \text{где } \Phi(t_\alpha) &= \frac{1}{\sqrt{2\pi}} \int_{-t_\alpha}^{t_\alpha} e^{-t^2/2} dt \end{aligned} \right\} \quad (7.48)$$

— табулированный интеграл вероятностей.

В частности, значениям α , равным 0,6; 0,9 и 0,99, соответствуют следующие значения t_α : 0,26, 1,28 и 2,32. В дальнейшем будем полагать, что

$$\alpha \leq 0,99, \quad (7.49)$$

так как для использования более высоких значений вероятности α необходимо знать распределение $W(\eta)$ с недостижимой в большинстве практических задач точностью. Тогда, как следует из (7.44) и (7.48), $t_\alpha \sqrt{D_\eta} \ll 1$, и вместо (7.46) можно полагать

$$\eta_\alpha^* \approx \bar{\eta}_\alpha = 1 + t_\alpha \sqrt{D_\eta}, \quad (7.50)$$

т. е. коэффициенты доверия $\bar{\eta}_\alpha$ и η_α^* оказываются приблизительно одинаковыми и близкими к единице. Из (7.48) и (7.50) следует, в частности, что при $\sqrt{D_\eta} = 0,08$ получается

$$\eta_{0,6} = 1,02; \quad \eta_{0,9} = 1,1; \quad \eta_{0,99} = 1,2. \quad (7.51)$$

При этом

$$\eta_{0,99} \approx \eta_{0,9}^2. \quad (7.52)$$

Из соотношений (7.50) — (7.52) следует, что коэффициенты доверия η_α^* и $\bar{\eta}_\alpha$ зависят от двух факторов: точности расчета затрат (дисперсии D_η) и доверительной вероятности α , т. е. требуемой достоверности оценки истинного значения затрат по его расчетному значению. В дальнейшем будем называть требования $\alpha = 0,6$; $\alpha = 0,9$ и $\alpha = 0,99$ соответственно требованиями минимальной, умеренной и высокой достоверности оценки истинных затрат.

В § 7.5 будет выяснено, из каких соображений следует выбирать эти требования. Пока же будем считать значение вероятности α заданным и рассмотрим методику определения коэффициентов доверия η_α^* и $\bar{\eta}_\alpha$ для тех случаев, когда высокую точность расчета затрат невозможно обеспечить. На начальных этапах проектирования такая ситуация наиболее характерна, особенно если проектируемая система существенно отличается от ранее известных и не рассчитана на массовое производство. При этом не только не выполняются условия (7.43) и (7.44), но может быть неизвестен и вид закона распределения $W(\eta)$.

Тогда значения коэффициентов доверия η_α^* и $\bar{\eta}_\alpha$ можно определить приближенно с помощью экспертных оценок. Для этого целесообразно разделить все возможные ситуации, различающиеся точностью расчета затрат и требованиями к достоверности

оценки (прогнозирования) фактических затрат по результатам расчета, на несколько дискретных ситуаций, например, следующим образом. Введем три градации точности (высокая, средняя и низкая) и две градации требуемой достоверности (умеренная, $\alpha \approx 0,9$, и высокая, $\alpha \approx 0,99$). Разбивка значений η_{α}^* и $\bar{\eta}_{\alpha}$ между этими градациями проводится так, как указано в табл. 7.1, составленной на основе следующих предположений.

Таблица 7.1

Достоверность	Точность					
	высокая ($\sqrt{D\eta} = 0,08$)		средняя		низкая	
	$\bar{\eta}_{\alpha}$	η_{α}^*	$\bar{\eta}_{\alpha}$	η_{α}^*	$\bar{\eta}_{\alpha}$	η_{α}^*
Умеренная ($\alpha \approx 0,9$)	1,1	1,1	1,2	1,4	1,4	2
Высокая ($\alpha \approx 0,99$)	1,2	1,2	1,4	2	2	4

1. Считается, что не только при высокой, но и при средней и низкой точности расчета должно приблизительно выполняться соотношение (7.52), т. е. повышение требований к достоверности оценки от умеренных к высоким должно проявляться в возведении соответствующего коэффициента доверия (η_{α}^* или $\bar{\eta}_{\alpha}$) в квадрат.

2. При высокой точности, как было доказано ранее, можно полагать $\bar{\eta}_{\alpha} \approx \eta_{\alpha}^*$. В остальных случаях целесообразно полагать

$$\eta_{\alpha}^* \approx (\bar{\eta}_{\alpha})^2, \quad (7.53)$$

т. е. коэффициент η_{α}^* выбирать существенно большим, чем $\bar{\eta}_{\alpha}$, поскольку при невысокой точности значительно более вероятно, что фактическое значение затрат P_{Π} превысит расчетное значение P (т. е. будет $\eta > 1$), а не окажется меньше его ($\eta < 1$).

Такая ситуация имеет место на практике потому, что в процессе проектирования, производства и эксплуатации системы появляется ряд дополнительных источников существенных затрат, которые невозможно учесть на начальных этапах проектирования. (Например, сравнительно часто оказывается, что фактические затраты в 1,5—2 раза больше расчетных, и значительно более редки случаи, в которых фактические затраты в 1,5—2 раза меньше расчетных.) Выбор зависимости между η_{α}^* и $\bar{\eta}_{\alpha}$ в виде (7.53), аналогичном (7.52); удобен потому, что при этом уменьшается число градаций коэффициентов доверия, соответственно упроща-

ется задача экспертов и вследствие этого уменьшается вероятность принятия ими ошибочных решений.

3. Принято, что даже при низкой точности расчета можно полагать практически невероятным превышение фактических затрат по сравнению с расчетными более чем в 4 раза, т. е. сделано допущение, что при низкой точности расчета $\eta_{0,99}^* = 4$.

4. Для высокой точности расчета приняты значения η_{α}^* и $\bar{\eta}_{\alpha}$, определяемые соотношениями (7.51), так как при значениях η и $\bar{\eta}_{\alpha}$, близких к единице, т. е. при очень высокой точности расчета, коэффициенты доверия можно вычислить по формулам (7.35) и (7.38) и, следовательно, прибегать к экспертным оценкам нет необходимости.

При использовании табл. 7.1 эксперты должны ответить на следующие два вопроса:

какие требования к достоверности следует предъявить — умеренные ($\alpha \approx 0,9$) или высокие ($\alpha \approx 0,99$);

для установленной степени достоверности (т. е. для выбранной строки в табл. 7.1) решить, к какому из трех классов следует отнести точность расчета затрат, т. е. какими из приведенных в данной строке следует принять значения коэффициентов η_{α}^* и $\bar{\eta}_{\alpha}$.

При этом ответы на первый и второй вопросы должны давать различные по составу группы экспертов. Действительно, при ответе на первый вопрос решающее слово должно принадлежать инженерам-разработчикам аппаратуры и ее заказчикам, а при решении второго вопроса (оценки точности расчета) — экономистам и (на начальных этапах проектирования) разработчикам. Для повышения надежности экспертных оценок следует производить соответствующую статистическую обработку оценок, даваемых отдельными экспертами [58].

Необходимо иметь в виду, что дискретизация возможных значений коэффициентов доверия, приведенная в табл. 7.1, в значительной мере условна. В частности, можно было бы, например, ввести в эту таблицу не две, а три градации достоверности, т. е. добавить случай, когда достаточна минимальная достоверность ($\alpha \approx 0,6$). Однако экспертам решить, какими должны быть значения $\eta_{0,6}^*$ и $\bar{\eta}_{0,6}$ при средней и низкой точности расчета, в ряде практических задач может быть слишком сложно.

7.5. ВЫБОР ДОВЕРИТЕЛЬНОЙ ВЕРОЯТНОСТИ

Рассмотрим выбор требуемой доверительной вероятности α применительно к высокой точности расчета, так как только в этом случае вероятность α действительно можно вычислить. При средней и низкой точности, строго говоря, можно говорить лишь о высокой, умеренной или минимальной достоверности, не конкретизируя совершенно определенно ее количественного значения. Одна-

ко в качественном отношении соображения, которые будут высказаны для высокой точности расчета, останутся приближенно справедливыми и при средней или низкой точности расчета.

Рассмотрим выбор доверительной вероятности при сравнении двух вариантов S_1 и S_2 построения системы. Обозначим $K'_T = \langle k'_1, \dots, k'_m \rangle$, $K''_T = \langle k''_1, \dots, k''_m \rangle$ значения вектора K_T технических показателей качества для вариантов S_1 и S_2 . Тогда можно записать

$$K' = \langle P_1, K'_T \rangle, K'' = \langle P_2, K''_T \rangle, \quad (7.54)$$

т. е. качество каждого варианта характеризуется значениями вектора K_T технических показателей и затрат P .

Рассмотрим сначала случай, когда значения векторов K'_T и K''_T либо полностью совпадают, либо если и различаются, то невозможно отдать предпочтение какому-либо из этих значений. Это означает, что лучший из вариантов S_1 и S_2 следует выбирать на основании сравнения только затрат P_1 и P_2 . Если бы затраты P_1 и P_2 были бы известны точно, т. е. имели место равенства $P_1 = P_{1н}$, $P_2 = P_{2н}$, то при $P_2 < P_1$ следовало отдать предпочтение варианту S_2 , а при $P_2 > P_1$ — варианту S_1 . Однако в действительности разработчик системы располагает лишь расчетными значениями P_1 и P_2 , а их истинные значения $P_{1н}$ и $P_{2н}$ могут отличаться от расчетных. Предположим для простоты, что затраты на вариант S_1 (например, прототип) известны достаточно точно, так что можно полагать

$$P_1 = P_{1н}; \quad (7.55)$$

затраты же P_2 определены неточно, так что известны лишь коэффициенты доверия η_α и $\bar{\eta}_\alpha$. Это значит, что с вероятностью α можно утверждать, что

$$P_{2н} \leq \eta_\alpha P_2. \quad (7.56)$$

Очевидно, вариант S_2 будет лучше S_1 при выполнении условия

$$P_{2н} < P_1. \quad (7.57)$$

Поэтому, если будет выполняться неравенство

$$\eta_\alpha P_2 < P_1,$$

$$\text{т. е. } P_2 < P_1 / \eta_\alpha, \quad (7.58)$$

то с вероятностью, не меньшей α , вариант S_2 лучше S_1 и с вероятностью, не большей $(1-\alpha)$, он не лучше S_1 .

Пусть величина η_α определена для $\alpha=0,6$. Тогда при выполнении неравенства (7.58) с вероятностью, не меньшей $\alpha=0,6$, вариант S_2 лучше S_1 и со значительно меньшей вероятностью — не большей $(1-\alpha)=0,4$ — он будет не лучше S_1 . Если возможностей для уточнения затрат не имеется и решение о том, какой из двух вариантов, S_1 или S_2 , лучше, должно быть принято обязательно, то

в таких условиях следует отдать предпочтение варианту S_2 , ибо более вероятно, что он лучше S_1 .

Пусть, например, в соответствии с (7.51) $\eta_{0,6}=1,02$. Тогда решение о предпочтительности варианта S_2 должно быть принято, если оказалось, что

$$P_2 < P_1 / 1,02. \quad (7.59)$$

Рассмотрим теперь случай, когда значение вектора K_T технических показателей качества у варианта S_2 несколько хуже, чем у варианта S_1 , хотя и остается в допустимых пределах¹. В этом случае для того, чтобы отдать предпочтение варианту S_2 , недостаточно выполнения условия (7.58). Необходимо, чтобы переход к варианту S_2 давал существенный годовой экономический эффект $\Delta P_{г.н}$, где по определению

$$\Delta P_{г.н} = P_{1н} - P_{2н} = P_1 - P_{2н}. \quad (7.60)$$

Но истинное значение затрат $P_{2н}$ неизвестно — известно лишь расчетное значение P_2 . При этом в соответствии с (7.56) можно лишь утверждать, что с вероятностью α будет $P_{2н} \leq \eta_\alpha P_2$.

Пусть выполняется условие $\eta_\alpha P_2 \leq P_1$, т. е.

$$P_2 \leq P_1 / \eta_\alpha. \quad (7.61)$$

Тогда с вероятностью α экономический эффект $\Delta P_{г.н}$ будет неотрицательным и его значение будет удовлетворять неравенству

$$\Delta P_{г.н} \geq P_1 - \eta_\alpha P_2. \quad (7.62)$$

т. е. с вероятностью α будет гарантировано, что экономический эффект будет не меньше

$$\Delta P_{г.н. мин \alpha} = P_1 - \eta_\alpha P_2. \quad (7.63)$$

Очевидно, для того, чтобы вариант S_2 был предпочтительнее, необходимо, чтобы эффект $\Delta P_{г.н. мин \alpha}$ был не слишком мал и гарантирован с достаточно высокой вероятностью α , например равной 0,9.

¹ При технико-экономическом сравнении вариантов S_1 и S_2 желательно предварительно привести их к сопоставимому виду, т. е. обеспечить равенство $K'_T = K''_T$ (например, сравнивать стоимости P_1 и P_2 вариантов S_1 и S_2 построения системы передачи информации при условии, что в обоих вариантах такие показатели, как пропускная способность C , вероятность ошибки $P_{ош}$, вероятность отказа $P_{отк}$ и другие одинаковы). Однако практически обеспечить сопоставимость вариантов по всем без исключения техническим показателям качества $k_1, \dots, k_n$ не удастся (даже путем пересчета некоторых технических показателей в эквивалентные им затраты); поэтому в большинстве практических задач считать варианты S_1 и S_2 полностью (точно) совершенно эквивалентными, за исключением их стоимостей P_1 и P_2 , недопустимо — почти всегда найдутся какие-то, хотя бы и не очень существенные, различия между техническими показателями систем S_1 и S_2 .

Выбирать $\alpha=0,6$ в этом случае уже недопустимо. Как следует из (7.51), при $\alpha=0,9$ получается $\eta_{\alpha}=1,1$ и из (7.63) имеем

$$\Delta P_{г.л.мин 0,9} = P_1 - 1,1 P_2. \quad (7.64)$$

Если, например, следует считать достаточным уменьшение затрат при переходе от S_1 к S_2 на 10%, т. е. полагать $\Delta P_{г.л.мин 0,9} = 0,1 P_1$, то из (7.64) получим, что необходимо обеспечить

$$P_2 \leq P_1/1,2. \quad (7.65)$$

Следовательно, в рассматриваемом случае для предпочтения варианта S_2 достаточным будет выполнение условия (7.65), значительно более жесткого, чем (7.59).

Рассмотрим теперь случай, когда вектор технических показателей K_T варианта S_2 значительно хуже, чем у варианта S_1 , хотя и остается допустимым (с точки зрения назначения системы). В этом случае для того, чтобы отдать предпочтение варианту S_2 , необходимо, чтобы получаемый экономический эффект составлял не 10% от P_1 , а значительно больше, например 50%, и был гарантирован не с вероятностью $\alpha \approx 0,9$, а с более высокой вероятностью $\alpha = 0,99$. При этом из (7.51) следует, что $\eta_{\alpha} = 1,2$, и с учетом (7.62) получается, что должно быть $P_1 - 1,2 P_2 > 0,5 P_1$, т. е.

$$P_2 < P_1/2,4. \quad (7.66)$$

Следовательно, переход к варианту S_2 будет оправдан лишь в том случае, если расчетное значение затрат P_2 будет по крайней мере в 2,4 раза меньше, чем у варианта S_1 (прототипа).

Из этих примеров видно, что допустимое значение доверительной вероятности α зависит от того, сколь существенно изменяется вектор технических показателей K_T при переходе от варианта S_1 к S_2 :

если этот вектор не изменяется, то можно полагать $\alpha \approx 0,6$ (минимальная достоверность оценки);

если изменяется, но незначительно, то можно полагать $\alpha \approx 0,9$, т. е. требовать умеренной достоверности оценки затрат;

если изменяется значительно, то необходима высокая достоверность оценки ($\alpha \approx 0,99$).

7.6. ВЛИЯНИЕ СЛУЧАЙНОСТИ НЕИЗМЕННОЙ (ОДИНАКОВОЙ) ДЛЯ ВСЕХ ВАРИАНТОВ СОСТАВЛЯЮЩЕЙ ЗАТРАТ

Затраты P_1 и P_2 для вариантов S_1 и S_2 можно представить в виде $P_1 = P_0 + P_{1\sim}$, $P_2 = P_0 + P_{2\sim}$, где P_0 — составляющая затрат, одинаковая для обоих вариантов (например, входящие в состав радиосистем S_1 и S_2 передающие устройства одинаковы, а разные лишь приемные устройства).

В общем случае могут иметь место ошибки определения всех составляющих затрат, т. е. $P_0 = P_{0н} + \Delta P_0$, $P_{1\sim} = P_{1\sim н} + \Delta P_{1\sim}$, $P_{2\sim} = P_{2\sim н} + \Delta P_{2\sim}$. Поэтому

$$\begin{aligned} P_1 &= P_{0н} + \Delta P_0 + \Delta P_{1\sim н} + \Delta P_{1\sim}, \\ P_2 &= P_{0н} + \Delta P_0 + \Delta P_{2\sim н} + \Delta P_{2\sim} \end{aligned} \quad (7.67)$$

и

$$\Delta P_r = P_1 - P_2 = (P_{1\sim н} - P_{2\sim н}) + (\Delta P_{1\sim} - \Delta P_{2\sim}), \quad (7.68)$$

т. е. неточность ΔP_0 расчета неизменной составляющей затрат P_0 не влияет на результат сравнения вариантов S_1 и S_2 , если это сравнение производится только на основе годового экономического эффекта ΔP_r — существенны ошибки лишь в определении переменных (зависящих от варианта S) составляющих P_1 и P_2 . Однако, как уже отмечалось, при переходе от варианта S_1 к S_2 может изменяться значение не только затрат P , но и вектора K_T технических параметров. В этом случае для вынесения правильного решения о предпочтительности варианта S_2 (или S_1) может потребоваться знание не только абсолютного экономического эффекта $|\Delta P_r|$, но и относительного эффекта $|\Delta P_r|/P_1$ (или $|\Delta P_r|/P_2$).

Пусть, например, вариант S_2 спутниковой системы связи дешевле варианта S_1 на $\Delta P_r = 10^6$ р. в год, но имеет несколько худшие (хотя и допустимые) значения технических показателей качества ($k_1, \dots, k_m$). Экономия $|\Delta P_r|$ по абсолютному значению велика, но она может составлять всего несколько процентов или даже менее от полной стоимости P_1 системы. Пойти на заметное ухудшение вектора K_T столь дорогой (а следовательно, и важной) системы ради уменьшения ее стоимости всего на несколько процентов может оказаться нецелесообразным. Поэтому для принятия правильного решения важно знать не только абсолютный экономический эффект $|\Delta P_r|$, но и относительный эффект $|\Delta P_r|/P_1$.

При этом на результат сравнения вариантов может повлиять и ошибка ΔP_0 определения неизменной составляющей затрат P_0 . Однако, поскольку эта ошибка обычно не приводит к изменению значения P_0 на порядок или более даже при низкой точности расчета, ее влияние на результаты сравнения вариантов, как правило, значительно меньше влияния ошибок расчета переменных составляющих. Поэтому при выборе оптимального варианта построения системы допустимо в первом приближении учитывать ошибки расчета лишь переменных составляющих затрат, т. е. полагать $\Delta P_0 = 0$.

7.7. ПУТИ УМЕНЬШЕНИЯ СТОИМОСТИ

В дальнейшем приведенные затраты любого вида (P , P' , P'') будем называть для краткости стоимостью и обозначать буквой B .

Стоимость проектируемой системы зависит от предъявляемых к ней требований и способов их реализации. Совокупность требований, предъявляемых к системе, определяется количественно совокупностью D исходных данных на проектирование, которые, как уже отмечалось, включают условия U работы системы, ограничения O , на структуру системы и значения ее параметров, со-

став с. К вектора качества системы и ограничения O_k на допустимое значение этих показателей. Очевидно, минимальное (по всем строго допустимым системам) значение $B_{\min}$ стоимости спроектированной системы при некоторых исходных данных D будет тем меньшим, чем менее жесткими (более простыми) являются эти исходные данные: чем уже диапазон условий, в которых должна работать система и чем проще эти условия; чем менее жесткими (более слабыми) являются ограничения O_k ; чем меньше показателей входит в состав с. К вектора качества K и чем больше предельно допустимые значения k_i макс отдельных (стандартных) составляющих k_i этого вектора.

Это весьма важное обстоятельство необходимо учитывать при обосновании исходных данных, т. е. на этапе внешнего проектирования системы, и по возможности делать эти данные менее жесткими (более простыми). Однако при «облегчении» исходных данных качество системы в какой-то мере ухудшится, что может быть оправдано лишь существенным уменьшением минимального значения $B_{\min}$ стоимости. Но $B_{\min}$ можно определить лишь в процессе выбора принципов построения системы и значений ее параметров, т. е. на этапе внутреннего проектирования. Отсюда следует, что на этапе внутреннего проектирования весьма желательно определить зависимость значения $B_{\min}$ от исходных данных D и, если это будет признано целесообразным заказчиками и конструкторами системы, соответствующим образом скорректировать исходные данные.

К сказанному необходимо сделать следующее важное примечание. Корректировка исходных данных D путем облегчения возможных условий U работы системы приводит к тому, что спроектированная система не сможет удовлетворительно работать в более трудных условиях и, следовательно, в какой-то мере снижается качество системы. Уменьшение состава учитываемых при проектировании показателей качества приводит к тому, что некоторые из неучитываемых второстепенных показателей могут оказаться слишком плохими, и, следовательно, также снижается в какой-то мере качество системы. К явному снижению качества системы приводит и увеличение предельно допустимых значений (стандартных) показателей качества.

В отличие от этого корректировка исходных данных D ослаблением ограничений O_k на структуру и параметры проектируемой системы не только не приводит к ухудшению качества системы, но может даже повысить его, благодаря тому что расширяется диапазон, в котором осуществляется поиск оптимальной системы (расширяется исходное множество $M_{\text{сх}}$ строго допустимых систем).

Итак, первым возможным путем снижения стоимости системы является соответствующая корректировка исходных данных D до начала внутреннего проектирования и (или) в процессе этого проектирования. Однако основным путем является предельно возможное снижение стоимости при фиксированных исходных дан-

ных без ущерба (или без существенного ущерба) для качества системы. Иначе говоря, задача внутреннего проектирования состоит в том, чтобы найти значение $B_{\min}$ и определить структуру системы и значения ее параметров, обеспечивающие получение такого минимального значения стоимости. Очевидно, для этого в допустимых исходными данными пределах необходимо варьировать структуру системы и значения ее параметров и выяснить, в какой мере эти вариации сказываются на стоимости B системы (приведенных затрат).

На стоимость системы в той или иной мере влияют и ее структура, и значения всех параметров. Поэтому достаточно полное и строгое решение указанной задачи технико-экономической оптимизации системы является весьма сложным. Отсюда вытекает необходимость выяснить, какие части системы, факторы и параметры настолько существенно влияют на стоимость радиосистемы, что ими нельзя пренебрегать даже на начальных этапах проектирования.

Основными факторами, влияющими на стоимость системы, являются:

- принципы организации связи;
- принципы пространственно-временной обработки информации;
- принципы обеспечения надежности;
- состав элементной базы;
- конструкция системы;
- технология производства системы;
- методы, используемые при проектировании системы.

Здесь принципы организации связи и пространственно-временной обработки понимаются в смысле, указанном в § 3.1. Ограничения, накладываемые на эти принципы, входят в состав O_k , задаваемых исходными данными. Чем менее жесткими они являются, тем больше удастся снизить стоимость путем выбора соответствующих принципов.

Различные способы обеспечения (повышения) надежности системы были рассмотрены в § 6.4. Нужно выбирать такие из них, которые обеспечивают заданную надежность системы при минимальных затратах.

Состав элементной базы системы влияет на стоимость как непосредственно (т. е. путем изменения стоимости самой элементной базы), так и косвенно, поскольку от вида элементной базы в значительной мере зависят и другие факторы, определяющие стоимость, в частности способы обеспечения надежности, конструкция системы, методы проектирования и технология производства.

Методы, используемые при проектировании, влияют на стоимость системы также двояко: изменением как стоимости самого процесса проектирования, так и остальных составляющих стоимости системы. Например, более широкое использование при проектировании стандартных узлов, блоков и подсистем обычно существенно снижает стоимость как самого процесса проектирова-

ния, так и отдельных (выбранных стандартными) частей системы. Однако при проектировании совершенно новой системы чрезмерное использование стандартных частей может привести не к снижению, а к увеличению стоимости, так как может исключить возможность применения более прогрессивных (с точки зрения минимизации стоимости) принципов организации связи и пространственно-временной обработки.

Автоматизация проектирования обычно существенно снижает стоимость проектирования (а следовательно, и стоимость системы). Однако существует оптимальная степень автоматизации, превышение которой приводит не к уменьшению, а к увеличению стоимости проектирования.

Принципы, закладываемые в конструкцию системы, оказывают влияние на стоимость системы многими путями, в том числе следующими:

- от конструкции системы зависит ее надежность и, следовательно, затраты, необходимые для обеспечения требуемой надежности;

- от конструкции весьма существенно зависит технология производства системы, состав и количество расходуемых материалов и деталей;

- конструкция системы в значительной мере определяет массу и габариты аппаратуры. Увеличение же массы и габаритов аппаратуры, особенно бортовой (например, размещенной на космическом летательном аппарате), может резко увеличить стоимость системы (например, вывод каждого килограмма аппаратуры на орбиту спутника Земли может стоить тысячи или даже десятки тысяч рублей в зависимости от вида орбиты).

Большое влияние на стоимость системы технологии ее производства очевидно. Стоимость системы, особенно выпускаемой в больших количествах, часто можно существенно снизить, используя стандартные и более дешевые материалы, детали и узлы, а также автоматизацией производства. Однако, как и при проектировании, обычно существуют оптимальные степени стандартизации и автоматизации, превышение которых приводит не к уменьшению, а к увеличению стоимости.

Выясним теперь, какие части радиосистемы в наибольшей степени влияют на ее стоимость. К ним обычно относятся:

- мощные каскады радиопередающих устройств (включая необходимые для них источники питания);

- антенные устройства (передающие и приемные), особенно остроуправленные с несколькими сканирующими в пространстве лучами;

- входные каскады радиоприемных устройств, определяющие их шумовую температуру, особенно при малых значениях этой температуры;

- устройства фазовой, частотной и временной синхронизации, синтезаторы частот, особенно при высоких требованиях к диапазону и точности синхронизации;

ЭВМ, осуществляющие обработку информации, при больших объемах и скорости обработки;

кодирующие и особенно декодирующие устройства в системах (устройствах) передачи информации при высоких требованиях к их эффективности.

Условно все приведенные части можно разбить на две большие группы:

- устройства, в первую очередь определяющие дальность действия системы или, как говорят, обеспечивающие энергетику радиолиний. К ним относятся первые три из упомянутых шести частей;

- все остальные части, осуществляющие в основном временную обработку (некоторые из них, например ЭВМ, могут кроме «чистой» временной производить и пространственно-временную обработку).

Удельное значение стоимости каждой из этих двух групп или их отдельных частей в общей стоимости системы может варьироваться в зависимости от вида и назначения системы в больших пределах. Однако в большинстве случаев стоимость устройств, обеспечивающих энергетику системы, составляет около половины общей стоимости системы.

Выясним теперь, от каких параметров системы зависит ее стоимость. Таких параметров, вообще говоря, существует очень много. Поэтому перечислим лишь те из них, которые в большинстве случаев можно считать оказывающими на стоимость наибольшее влияние. Из изложенного следует, что к таким параметрам можно отнести, в частности, следующие:

- число радиопередающих устройств (РПДУ) и отдельно число бортовых РПДУ;

- число радиоприемных устройств (РПУ) и отдельно число бортовых РПУ;

- рабочая длина волны (длины волн);

- кратность резервирования отдельных устройств;

- число антенно-фидерных устройств каждого типа (зеркальных, линзовых, ФАР, с управляемыми и неуправляемыми диаграммами и т. п.);

- средняя выходная мощность $P_{ср}$ каждого радиопередающего устройства;

- площадь раскрытия S_A каждой антенны (приемной и передающей каждого типа);

- шумовая температура $T_{ш}$ собственных шумов каждого приемного устройства;

- состав и число устройств временной обработки; кратность резервирования каждого из них.

Для оценки влияния энергетических параметров часто можно использовать, по меньшей мере в первом приближении, следующие аппроксимирующие зависимости [61—66]:

$$B_{мп} = c_1 P_{ср}^{\alpha_1}, B_A = c_2 S_A^{\alpha_2}, B_{ш} = c_3 (1/T_{ш})^{\alpha_3}. \quad (7.69)$$

Здесь $B_{\text{ш}}$, B_A и $B_{\text{ш}}$ — соответственно стоимость мощных каскадов передатчика, антенны и входной части приемника, определяющей его шумовую температуру; a_1 , a_2 и a_3 — положительные безразмерные числа. В первом приближении можно полагать

$$a_1 = a_2 = a_3 = 1. \quad (7.70)$$

Коэффициенты c_1 , c_2 и c_3 существенно зависят от типа (класса) соответствующего устройства, рабочей длины волны, кратности резервирования и ряда других факторов; поэтому они должны определяться для каждого конкретного случая отдельно.

В общем случае параметры (c_1, a_1) , (c_2, a_2) и (c_3, a_3) аппроксимаций (7.69) желательно определять путем статистического анализа существующих выборочных данных о стоимостях изделий данного класса. Очевидно, чем шире диапазон значений варьируемых параметров ($P_{\text{ср}}$, S_A , $T_{\text{ш}}$), тем более грубыми будут аппроксимации (7.69), особенно если в них полагать $a_1 = a_2 = a_3 = 1$.

Однако если не налагать упрощающего условия (7.70), то аппроксимации (7.69) могут оказаться удовлетворительными (в качестве первого приближения) даже при изменениях параметров в 10 ... 100 раз (например, при $P_{\text{ср}} = 10 \dots 1000$ Вт, $S_A = 1 \dots 100$ м², $T_{\text{ш}} = 30 \dots 3000$ К). Типичные значения параметров c_i , a_i ($i = \overline{1, 3}$) приводятся также в ряде литературных источников [4, 64—66 и др.].

Наиболее сложно установить достаточно простые и универсальные зависимости стоимости устройств временной обработки в целом или отдельных их частей от значений технических параметров этих устройств. Это объясняется рядом факторов, в первую очередь весьма большим разнообразием таких устройств и бурным развитием их элементной базы. Поэтому обычно даже для грубой оценки стоимости таких устройств требуется предварительно произвести их эскизный расчет — отдельно для каждого конкретного варианта построения того или иного устройства.

С учетом этой трудности часто оказывается целесообразной следующая процедура минимизации стоимости системы (информационной). Стоимость системы в целом представляется в виде

$$B = B_{\text{эл}} + B_{\text{УВО}}, \quad (7.71)$$

где $B_{\text{эл}}$ и $B_{\text{УВО}}$ — соответственно стоимости энергетической (в указанном выше смысле) части системы и устройств временной обработки (УВО).

Рассматривается несколько конкретных видов (классов)* УВО₁, УВО₂, УВО₃, ..., УВО_N. Для каждого данного вида УВО_i определяется путем эскизных расчетов (или каким-либо другим способом) приближенное значение его стоимости $B_{\text{УВО } i}$.

Далее для каждого вида УВО_i находится путем оптимизации энергетической части системы минимальное значение ее стоимости $B_{\text{эл. мин } i}$. Тогда для каждого (i -го) варианта построения УВО минимальное значение стоимости всей системы

$$B_{i \text{ мин}} = B_{\text{эл. мин } i} + B_{\text{УВО } i}. \quad (7.72)$$

Выполняя такие расчеты для всех видов ($i = \overline{1, N}$) УВО, получаем соответствующие значения $B_{1 \text{ мин}}, \dots, B_{N \text{ мин}}$ условных минимальных стоимостей системы. Условность каждой из стоимостей $B_{i \text{ мин}}$ состоит в том, что она является минимальной лишь для данного (i -го) вида УВО.

Очевидно, оптимальной будет такая УВО_i (и соответственно система в целом), для которой значение $B_{i \text{ мин}}$ окажется минимальным, т. е. оптимальный вариант построения системы в целом находится из условия

$$B_{\text{мин}} = \min_i B_{i \text{ мин}} \quad i = \overline{1, N}. \quad (7.73)$$

При определении минимального значения стоимости системы часто оказывается полезным предварительно вычислить так называемую *потенциальную* стоимость $B_{\text{п}}$, т. е. минимальное, теоретически предельно достижимое значение стоимости. Для этого следует принять целый ряд идеализаций, из которых основной является допущение о бесплатности УВО (т. е. о равенстве нулю составляющей $B_{\text{УВО}}$ полной стоимости системы).

7.8. ПОСТАНОВКА ЗАДАЧИ ОПРЕДЕЛЕНИЯ ПОТЕНЦИАЛЬНОЙ СТОИМОСТИ

Методику определения потенциальной стоимости и использования найденных потенциальных значений при проектировании рассмотрим на примере радиолинии передачи информации. Аналогичную методику можно применить (с некоторыми коррективами) и для радиолиний извлечения информации [62], а также и для радиосистем в целом.

Желательно, чтобы потенциальная стоимость удовлетворяла тем же общим требованиям, что и рассмотренные ранее потенциальные помехоустойчивость (точность) и пропускная способность. Эти требования заключаются в следующем:

1. Потенциальное значение стоимости должно быть наименьшим, т. е. таким, которое не может быть превзойдено (в сторону уменьшения) ни в одной реальной системе (в пределах сформулированных исходных данных).

2. Должны существовать такие реальные задачи проектирования, в которых реальная стоимость весьма близко (асимптотически) приближается к потенциальной. В большинстве других уже решенных задач реальная стоимость должна отличаться от потенциальной не более чем на порядок. Это означает, что потенциальная стоимость должна быть для реальной стоимости асимптотически точной или, во всяком случае, не слишком грубой нижней гранью.

3. Потенциальная стоимость должна зависеть от возможно меньшего числа параметров, чтобы ее вычисление требовало минимальной априорной информации и поэтому было возможно на самых ранних этапах проектирования.

Как и в случае информационных потенциальных характеристик, целесообразно сформулировать не один, а несколько различных подходов к определению потенциальной стоимости (в пределах сформулированных трех основных требований), отличающихся различной степенью идеализации. В частности, предоставляется целесообразным различать следующие виды потенциальной стоимости.

Безусловная потенциальная стоимость (БПС) — минимальная стоимость радиолинии, достигаемая в предположении, что все преобразования сообщений и сигналов в радиолинии выполняются в информационном отношении оптимально (т. е. в соответствии с требованиями достижения потенциальных значений информационных характеристик — точности и пропускной способности) и бесплатно. Это означает, что учитывается стоимость обеспечения лишь таких основных параметров радиолинии, входящих в уравнение дальности ее действия, как средняя мощность P_n передатчика, площади S_n и S_{pr} передающей и приемной антенн и шумовая температура $T_{ш}$ приемника. При этом никакие ограничения на значения P_n , S_n , S_{pr} и $T_{ш}$ (кроме естественных ограничений положительности и уравнения дальности) не учитываются.

Стоимость минимизируется соответствующим подбором значений варьируемых параметров P_n , S_n , S_{pr} и $T_{ш}$.

Условная потенциальная стоимость (УПС) 1-го рода отличается от БПС лишь тем, что при определении из уравнения дальности требуемой мощности P_n учитываются потери за счет отличия реальных значений точности (помехоустойчивости) и скорости передачи информации от их потенциальных значений. Иначе говоря, преобразования сообщений, сигналов и помех в радиолинии полагаются не идеальными (оптимальными), а реальными, но стоимость этих преобразований, как и в случае БПС, полагается равной нулю. Это означает, что если в случае БПС преобразования сообщений и сигналов считаются идеальными и бесплатными, то в случае УПС 1-го рода — реальными и бесплатными.

Условная потенциальная стоимость 2-го рода отличается от БПС лишь тем, что учитываются ограничения на предельно допустимые значения варьируемых параметров:

$$P_n \leq P_{n, \max}, S_n \leq S_{n, \max}, S_{pr} \leq S_{pr, \max}, T_{ш} \geq T_{ш, \min}. \quad (7.74)$$

Условная потенциальная стоимость УПС 3-го рода отличается от БПС тем, что учитывается как неидеальность обработки информации, так и ограничения (7.74). Но обработка по-прежнему считается бесплатной. Следовательно, УПС-3 отличается от реальной стоимости лишь тем, что обработка информации считается бесплатной. Кроме того, при определении УПС-3 полагается, что в рамках ограничений (7.74) параметры P_n , S_n , S_{pr} и $T_{ш}$ выбираются по критерию минимума стоимости.

Если обозначить значения стоимости B , найденной при упомянутых выше четырех подходах, соответственно $B_{п1}$, $B_{п2}$, $B_{п3}$ и $B_{п4}$, то должны выполняться следующие очевидные неравенства:

$$B_{п1} \leq B_{п2}, B_{п3}, B_{п4}, B_{п2} \leq B_{п4}, B_{п3} \leq B_{п4}, \quad (7.75)$$

$$B_p \geq B_{п1}, B_{п2}, B_{п3}, B_{п4}, \quad (7.76)$$

где B_p — реальное значение стоимости.

Докажем, что все сформулированные виды потенциальной стоимости (БПС, УПС-1, УПС-2 и УПС-3) удовлетворяют всем указанным основным требованиям к потенциальным характеристикам.

Во-первых, из (7.76) следует, что любая из потенциальных характеристик $B_{п1}$, $B_{п2}$, $B_{п3}$, $B_{п4}$ является нижней гранью для реальной стоимости B_p . Во-вторых, очевидно, что при большой дальности действия d радиолинии возрастание этой дальности будет вызывать опережающее возрастание стоимости реализации больших значений P_n , S_n и S_{pr} и малой $T_{ш}$ по сравнению с ростом стоимости обработки информации. Поэтому отличие потенциальной стоимости от реальной, вызванное пренебрежением стоимостью обработки информации, по мере роста дальности действия радиолинии асимптотически стремится к нулю, т. е. $B_p \rightarrow B_{п4}$.

Таким образом, существуют задачи, в которых реальная стоимость асимптотически совпадает с потенциальной (по меньшей мере с $B_{п4}$). Как показали расчеты, во многих задачах величины $B_{п1}$, $B_{п2}$, а следовательно, и $B_{п3}$ отличаются от B_p менее чем на порядок. Иначе говоря, $B_{п1}$, $B_{п2}$, $B_{п3}$ и $B_{п4}$ действительно удовлетворяют сформулированным выше первому и второму основным требованиям. Вследствие пренебрежения стоимостью устройств обработки информации они удовлетворяют также и третьему основному требованию. Действительно, устройства обработки информации в передатчике и приемнике столь разнообразны по принципам построения и элементной базе, что оценить их потенциальную или реальную стоимость на ранних этапах проектирования системы практически невозможно. В это же время приближенно оценить стоимость реализации тех или иных значений параметров P_n , S_n , S_{pr} и $T_{ш}$ вполне возможно даже на самых ранних этапах проектирования. Поэтому сформулированные определения потенциальных стоимостей БПС и УПС-1, -2, -3 представляются вполне оправданными. Конечно, в некоторых случаях значение потенциальной стоимости (особенно БПС) может оказаться более чем на порядок меньше ее реального значения. Но даже в этих случаях знание БПС и УПС-1, -2, -3 может оказаться полезным, так как позволит лучше понять, какие факторы и в какой степени препятствуют снижению стоимости реальной системы.

Сформулированный подход к определению потенциальной стоимости радиолинии можно назвать энергетическим. Действительно, он предполагает, что в процессе минимизации стоимости варьируются лишь те параметры радиолинии (P_n , S_n , S_{pr} и $T_{ш}$), которые определяют ее энергетику.

Вообще говоря, возможны и другие виды оптимизации (минимизации) стоимости, которые можно назвать соответственно информационной, конструкторской, технологической и эксплуатаци-

онной оптимизацией. При информационной оптимизации стоимости варьируются лишь структура и параметры системы обработки информации. Конструкторская оптимизация стоимости обеспечивается вариацией лишь конструктивных принципов и параметров (массы, габаритов, технологической надежности). Технологическая оптимизация стоимости достигается выбором комплектующих изделий (элементной базы), включая детали и материалы, и технологию изготовления аппаратуры. При эксплуатационной оптимизации стоимость минимизируется соответствующим выбором режима эксплуатации и значением соответствующих параметров (например, периода профилактических проверок). Наконец, наиболее полной является комплексная оптимизация, т. е. совместная оптимизация двух или более из упомянутых пяти видов.

Далее ограничимся лишь энергетической оптимизацией стоимости и укажем некоторые пути ее совмещения с другими видами оптимизации.

Рассмотрим элементарную СВЧ радиолинию без ретрансляции и с пассивной ретрансляцией. Эта радиолиния предназначена для передачи информации и может иметь один или более каналов.

Радиолиния передачи информации без ретрансляции. В этом случае уравнение дальности можно представить в следующем виде:

$$P_n = d^2 \lambda^2 L_1^2 L_2 P_{\text{пр}} / S_n S_{\text{пр}}. \quad (7.77)$$

Здесь P_n — средняя мощность передатчика, требуемая для обеспечения заданной дальности действия d ; $P_{\text{пр}}$ — средняя мощность на входе приемника (т. е. на выходе приемной антенны), требуемая для обеспечения воспроизведения передаваемой информации; λ — рабочая длина волны; S_n и $S_{\text{пр}}$ — эффективные площади передающей и приемной антенн соответственно; L_1 — потери при распространении энергии в открытом пространстве. Если этим пространством является идеальный вакуум, то $L_1 = 1$. В остальных случаях $L_1 > 1$ и зависит от характера и состояния среды (атмосферы, тропосферы, ионосферы) и от длины волны λ ; L_2 — потери при распространении электромагнитной энергии в антенных и фидерных устройствах радиопередающего и радиоприемного устройств (из-за побочных лепестков, из-за несовпадения направления максимума главного лепестка с направлением передатчик — приемник), поляризационные потери, затухание в фидерных линиях и др. В любом случае $L_2 > 1$, причем в большинстве реальных случаев $L_2 \approx 5 \dots 10$.

Требуемую мощность приемника $P_{\text{пр}}$ удобно представить в следующем виде:

$$P_{\text{пр}} = \xi P_{\text{пр.л}}, \quad (7.78)$$

где $P_{\text{пр.л}}$ — требуемая средняя мощность сигнала на входе приемника в идеальных условиях, т. е. при наличии помехи лишь в виде аддитивного белого шума и идеальной (в смысле Шеннона)

обработке информации; ξ — потери за счет неидеальности обработки информации и (возможно) действия других видов помех. В любом случае $\xi \geq 1$, а в реальных радиолиниях обычно

$$\xi \approx 2 \dots 10. \quad (7.79)$$

При определении БПС следует полагать $\xi = 1$, а в случае вычисления УПС-1 ξ равна своему реальному значению. Если на начальном этапе проектирования достаточно точные данные о значении ξ отсутствуют, можно полагать $\xi \approx 5 \dots 10$.

Величина $P_{\text{пр.л}}$ определяется следующими соотношениями:

$$P_{\text{пр.л}} = (P_c / N_0) N_0, \quad (7.80)$$

где

$$N_0 = k T_{\text{ш}} \quad (7.81)$$

— спектральная плотность аддитивного белого шума, имеющего шумовую температуру $T_{\text{ш}}$ (в градусах Кельвина); $k = 1,38 \times 10^{-23}$ Дж/град — постоянная Больцмана; P_c / N_0 — отношение сигнал-шум, требуемое для обеспечения скорости передачи информации C бит/с (со сколь угодно малой вероятностью ошибки). Оно определяется формулой (4.18) и построенной на ее основе кривой, изображенной на рис. 4.2 сплошной линией. По этой кривой для заданных значений C и Δf , т. е. для данного значения β_E , определяется требуемое значение коэффициента β_E . Тогда в соответствии с (4.20) требуемое значение отношения сигнал-шум

$$P_c / N_0 = \beta_E C. \quad (7.82)$$

Будем в дальнейшем обозначать требуемое отношение сигнал-шум P_c / N_0 через η :

$$\eta = P_c / N_0. \quad (7.83)$$

Тогда с учетом (7.78), (7.80) — (7.83) уравнение дальности (7.77) можно представить в следующем виде:

$$P_n = d^2 \lambda^2 L_1 L_2 \xi \eta k T_{\text{ш}} / S_n S_{\text{пр}}, \quad (7.84)$$

где

$$\eta = \beta_E C. \quad (7.85)$$

При определении потенциальной стоимости варьируются лишь энергетические параметры P_n , S_n , $S_{\text{пр}}$ и $T_{\text{ш}}$. Поэтому удобно ввести следующие обозначения:

$$x_1 = P_n, x_2 = S_n, x_3 = S_{\text{пр}}, x_4 = 1/T_{\text{ш}}, \quad (7.86)$$

$$A_1 = d^2 \lambda^2 L_1 L_2 k \xi \eta. \quad (7.87)$$

Тогда уравнение дальности (7.84) принимает вид

$$x_1 x_2 x_3 x_4 = A_1. \quad (7.88)$$

При оптимизации (минимизации) стоимости радиолинии это уравнение играет роль ограничения типа связи.

Радиолиния с пассивной ретрансляцией. В этом случае уравнение дальности радиолинии вместо (7.77) имеет следующий вид:

$$P_{\text{п}} = \frac{4\pi d_1^2 d_2^2 \lambda^2 L_1 L_2}{S_{\text{э}} S_{\text{п}} S_{\text{пр}}} P_{\text{пр}}, \quad (7.89)$$

где d_1 и d_2 — соответственно расстояния от передатчика и приемника до объекта, осуществляющего пассивное отражение падающей волны и имеющего эффективную поверхность рассеяния (ЭПР) $S_{\text{э}}$. После преобразований, аналогичных приведенным выше, это уравнение принимает вид

$$x_1 x_2 x_3 x_4 = A_2, \quad (7.90)$$

которое отличается от (7.88) лишь тем, что вместо A_1 , входит эквивалентная константа

$$A_2 = 4\pi d_1^2 d_2^2 \lambda^2 L_1 L_2 k \xi \eta / S_{\text{э}}. \quad (7.91)$$

Здесь L_1 , L_2 , k , ξ и η имеют тот же смысл, что и в формуле (7.87).

7.9. ОПРЕДЕЛЕНИЕ ПОТЕНЦИАЛЬНЫХ СТОИМОСТЕЙ БПС И УПС-1

Реальная стоимость B_p радиолинии складывается из следующих составляющих:

$$B_p = B_{\text{мп}} + B_{\text{Ап}} + B_{\text{оп}} + B_{\text{Апр}} + B_{\text{ш}} + B_{\text{о.пр}}, \quad (7.92)$$

где $B_{\text{мп}}$, $B_{\text{Ап}}$ и $B_{\text{оп}}$ — соответственно стоимости мощных каскадов передатчика, его антенны и остальных частей; $B_{\text{Апр}}$, $B_{\text{ш}}$ и $B_{\text{о.пр}}$ — стоимости антенны приемника, его частей, определяющих значение температуры шума и остальных частей. Обозначим

$$B_0 = B_{\text{оп}} + B_{\text{о.пр}} \quad (7.93)$$

и будем называть B_0 стоимостью обработки информации в радиолинии. Тогда полную стоимость (7.92) можно представить в виде

$$B_p = B_0 + B_{\text{э}}, \quad (7.94)$$

где $B_{\text{э}} = B_{\text{мп}} + B_{\text{Ап}} + B_{\text{Апр}} + B_{\text{ш}}$ будем называть стоимостью обеспечения энергетики радиолинии или, короче, стоимостью энергетик.

При определении потенциальной стоимости значение B_0 полагается равным нулю и минимизируется лишь значение $B_{\text{э}}$. Составляющие $B_{\text{мп}}$, $B_{\text{Ап}}$, $B_{\text{Апр}}$ и $B_{\text{ш}}$ являются возрастающими функциями соответствующих энергетических параметров: $B_{\text{мп}} = f_1(x_1)$, $B_{\text{Ап}} = f_2(x_2)$, $B_{\text{Апр}} = f_3(x_3)$, $B_{\text{ш}} = f_4(x_4)$. Поэтому потенциальное значение стоимости (БПС и УПС-1) определяется следующими соотношениями:

$$B_{\text{э}} = \min_x [f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4)], \quad (7.95)$$

где x_1 , x_2 , x_3 и x_4 удовлетворяют уравнению дальности

$$x_1 x_2 x_3 x_4 = A. \quad (7.96)$$

Здесь $x = \langle x_1, x_2, x_3, x_4 \rangle$ — вектор варьируемых параметров.

Для радиолинии без ретрансляции

$$A = A_1, \quad (7.97)$$

а для радиолинии с пассивной ретрансляцией

$$A = A_2. \quad (7.98)$$

Для получения аналитического выражения потенциальной стоимости необходимо выбрать соответствующую аппроксимацию функций $f_1(x_1)$, $f_2(x_2)$, $f_3(x_3)$ и $f_4(x_4)$. Целесообразной является следующая аппроксимация:

$$\begin{aligned} B_{\text{мп}} &= c_1 x_1^{a_1}, B_{\text{Ап}} = c_2 x_2^{a_2}, \\ B_{\text{Апр}} &= c_3 x_3^{a_3}, B_{\text{ш}} = c_4 x_4^{a_4}, \end{aligned} \quad (7.99)$$

где c_i и a_i ($i = \overline{1, 4}$) — некоторые положительные числа (не обязательно целые); при этом потенциальная стоимость является минимальным значением функции

$$B_{\text{э}} = c_1 x_1^{a_1} + c_2 x_2^{a_2} + c_3 x_3^{a_3} + c_4 x_4^{a_4}, \quad (7.100)$$

при вариациях (положительных) переменных x_1 , x_2 , x_3 и x_4 , удовлетворяющих уравнению дальности (7.96), т. е. задача определения потенциальной стоимости сводится к виду

$$\left. \begin{aligned} B_{\text{э}} &= c_1 x_1^{a_1} + c_2 x_2^{a_2} + c_3 x_3^{a_3} + c_4 x_4^{a_4} = \min_x \\ \text{при } x_1 x_2 x_3 x_4 &= A. \end{aligned} \right\} \quad (7.101)$$

Задача (7.101), так же как и (7.95), (7.96), является задачей нелинейного программирования. Однако выбор целевой функции в виде (7.100), т. е. в виде так называемого *позинома* [63], позволяет получить аналитическое решение. Действительно, из уравнения дальности следует, что

$$x_4 = A (x_1 x_2 x_3)^{-1}. \quad (7.102)$$

Поэтому

$$B_{\text{э}} = c_1 x_1^{a_1} + c_2 x_2^{a_2} + c_3 x_3^{a_3} + c_4' (x_1 x_2 x_3)^{-a_4}, \quad (7.103)$$

$$\text{где } c_4' = c_4 A^{a_4}. \quad (7.104)$$

Единственный минимум функции (7.103) находится приравниванием нулю частных производных

$$\frac{\partial B_{\text{э}}}{\partial x_1} = 0, \quad \frac{\partial B_{\text{э}}}{\partial x_2} = 0, \quad \frac{\partial B_{\text{э}}}{\partial x_3} = 0,$$

решением полученной таким образом системы из трех уравнений и подстановкой найденных оптимальных значений x_1^* , x_2^* , x_3^* переменных в (7.103). После несложных, но громоздких выкладок получаются следующие результаты:

$$B_{\Pi} = B_{\text{мин}} = \sqrt{\left(\frac{c_1 a_1 \mu}{a_4}\right)^{a_1} \left(\frac{c_2 a_2 \mu}{a_4}\right)^{a_2} \left(\frac{c_3 a_3 \mu}{a_4}\right)^{a_3} c_4' \mu},$$

где $\mu = 1 + \frac{a_1}{a_1} + \frac{a_2}{a_2} + \frac{a_3}{a_3}$; $c_4' = c_4 A^{a_4}$,

(7.105)

$$\left. \begin{aligned} x_1^* &= (B_{\Pi}/\gamma c_1)^{1/a_1}, \quad x_2^* = (B_{\Pi} a_1/\gamma c_2 a_2)^{1/a_2}, \\ x_3^* &= (B_{\Pi} a_1/\gamma c_3 a_3)^{1/a_3}, \quad x_4^* = (B_{\Pi} a_1/\gamma c_4 a_4)^{1/a_4}. \end{aligned} \right\}$$

(7.106)

Здесь

$$\gamma = 1 + \frac{a_1}{a_2} + \frac{a_1}{a_3} + \frac{a_1}{a_4}.$$

(7.107)

Для контроля точности численных расчетов может служить соотношение

$$x_1^* x_2^* x_3^* x_4^* = A.$$

(7.108)

Пусть B_{Π}^* , $B_{A\Pi}^*$, $B_{A\Pi\Pi}^*$ и $B_{\Pi\Pi}^*$ — значения стоимостей B_{Π} , $B_{A\Pi}$, $B_{A\Pi\Pi}$ и $B_{\Pi\Pi}$ при оптимальных значениях энергетических параметров [т. е. при значениях, определяемых соотношениями (7.106)]. Тогда оказывается, что справедливы соотношения

$$\frac{B_{A\Pi}^*}{B_{\Pi}^*} = \frac{a_1}{a_2}, \quad \frac{B_{A\Pi\Pi}^*}{B_{\Pi}^*} = \frac{a_1}{a_3}, \quad \frac{B_{\Pi\Pi}^*}{B_{\Pi}^*} = \frac{a_1}{a_4}$$

(7.109)

и выражение (7.105) можно представить в виде

$$B_{\Pi} = B_{\text{мин}} = B_{\Pi}^* \left(1 + \frac{a_1}{a_2} + \frac{a_1}{a_3} + \frac{a_1}{a_4}\right) = \gamma B_{\Pi}^*.$$

(7.110)

Из (7.109) и (7.110) следует, что относительные вклады, вносимые в потенциальную стоимость стоимостями обеспечения энергетических параметров P_{Π} , S_{Π} , $S_{\Pi\Pi}$ и T_{Π} , определяются лишь отношениями показателей степени a_1 , a_2 , a_3 и a_4 в частных целевых функциях (7.99): чем больше показатель степени у соответствующего параметра, тем меньше вклад этого параметра в потенциальную стоимость.

Если применить аппарат геометрического программирования [63], то все приведенные результаты (7.105) — (7.110) можно получить весьма просто, без громоздких выкладок, как показано в приложении 1. Эти результаты получаются особенно простыми и наглядными, если принять $a_1 = a_2 = a_3 = a_4 = 1$, т. е. считать, что стоимости обеспечения данных значений энергетических параметров P_{Π} , S_{Π} , $S_{\Pi\Pi}$ и $1/T_{\Pi}$ пропорциональны значениям этих параметров. Тогда, как следует из (7.109) и (7.110), вклады всех этих параметров в потенциальную стоимость одинаковы, т. е. $B_{\Pi}^* = B_{A\Pi}^* = B_{A\Pi\Pi}^* = B_{\Pi\Pi}^*$, и общие соотношения (7.105), (7.106) принимают следующий вид:

$$B_{\Pi} = B_{\text{мин}} = 4 \sqrt[4]{c_1 c_2 c_3 c_4 A},$$

(7.111)

$$\begin{aligned} x_1^* &= B_{\Pi}/4 c_1, \quad x_2^* = B_{\Pi}/4 c_2, \\ x_3^* &= B_{\Pi}/4 c_3, \quad x_4^* = B_{\Pi}/4 c_4. \end{aligned}$$

(7.112)

Для контроля может по-прежнему служить уравнение дальности (7.96).

С учетом соотношений (7.87), (7.91) выражение (7.111) принимает вид:

$$B_{\Pi} = 4 \sqrt[4]{c_1 c_2 c_3 c_4 d^2 \lambda^2 L_1 L_2 k \xi \eta};$$

(7.113)

для радиолинии с пассивной ретрансляцией

$$B_{\Pi} = 4 \sqrt[4]{c_1 c_2 c_3 c_4 4 \pi d_1^2 d_2^2 \lambda^2 L_1 L_2 k \xi \eta / S_0}.$$

(7.114)

Если в (7.113) и (7.114) положить $\xi = 1$, то получается выражение для БПС (т. е. для $B_{\Pi 1}$), а при реальных значениях ξ эти выражения будут давать значения УПС-1 (т. е. $B_{\Pi 2}$). Так как для реальных радиолиний значение ξ обычно удовлетворяет соотношению (7.79), то при этом УПС-1 превышает значение БПС не более чем в 1,8 раза.

Из (7.113) и (7.114) наглядно видна зависимость потенциальной стоимости и от дальности действия (d или d_1 и d_2) радиолинии. Проследить зависимость B_{Π} от λ значительно сложнее, так как в общем случае от значения λ могут зависеть коэффициенты стоимости c_1 , c_2 , c_3 и c_4 , потери L_1 и L_2 и эффективная поверхность рассеяния $S_{\text{эф}}$. Поэтому если длина волны λ не задана, следует вычислять потенциальную стоимость для ряда представляющих интерес значений λ .

Рассмотрим в качестве иллюстрации следующий пример расчета для радиолинии «Борт — Земля» без ретрансляции, в котором стоимость выражается в некоторых условных единицах стоимости е. с¹.

Задано

$$a_1 = a_2 = a_3 = a_4 = 1, \quad c_1 = 10^3 \text{ е. с./бит}, \quad c_2 = 10^4 \text{ е. с./бит}, \quad c_3 = 10^3 \text{ е. с./м}^2, \\ c_4 = 10^3 \text{ е. с./град}, \quad \lambda = 0,1 \text{ м}; \quad d = 1,2 \cdot 10^3 \text{ м}, \quad L_1 L_2 = 10, \quad \Delta f = 10^6 \text{ Гц}, \quad C = 10^6 \text{ бит/с},$$

Из рис. 4.2 находим, что $\beta_{\Pi} = 1$, и из (7.85) получаем $\eta = 10^6$ л/с.

Для вычисления БПС в формуле (7.113) следует положить $\xi = 1$. Тогда $B_{\Pi 1} = 8500$ е. с. При этом в соответствии с (7.86) и (7.112) оптимальные значения энергетических параметров равны

$$P_{\Pi}^* = 2,1 \text{ Вт}, \quad S_{\Pi}^* = 0,21 \text{ м}^2, \quad S_{\Pi\Pi}^* = 2,1 \text{ м}^2, \quad T_{\Pi}^* = 47 \text{ К}.$$

При реальной обработке, которой соответствует $\xi = 10$ (УПС-1) $B_{\Pi 2} = B_{\Pi 1} \sqrt[4]{\xi} = 15300$ е. с. При этом в соответствии с (7.112) оптимальные зна-

¹ Этот пример является гипотетическим.

чения всех энергетических параметров также возрастает в $\sqrt[4]{\xi}$ раз, т. е. в 1,8 раза; отсюда

$$P^*_{\Sigma}=3,8 \text{ Вт}, S^*_{\Sigma}=0,38 \text{ м}^2, S^*_{\text{пр}}=3,8 \text{ м}^2, T_{\Sigma}=26 \text{ К}.$$

7.10. МЕТОДИКА ОПРЕДЕЛЕНИЯ ПОТЕНЦИАЛЬНЫХ СТОИМОСТЕЙ С УЧЕТОМ ОГРАНИЧЕНИЙ (УПС-2 И УПС-3)

В § 7.9 были найдены значения потенциальной стоимости радиолонии без учета ограничений энергетических параметров ($x_1=P_{\Sigma}$, $x_2=S_{\Sigma}$, $x_3=S_{\text{пр}}$, $x_4=1/T_{\Sigma}$) сверху. Для вычисления условных потенциальных стоимостей УПС-2 и УПС-3 следует найти минимум целевой функции (7.103) с учетом ограничений (7.74) на варьируемые параметры, т. е. обеспечить минимум величины

$$B_{\Sigma}=c_1 x_1^{\alpha_1} + c_2 x_2^{\alpha_2} + c_3 x_3^{\alpha_3} + c'_4 (x_1 x_2 x_3)^{-\alpha_4} \quad (7.115)$$

при условиях

$$x_1 \leq x_{1\text{ макс}}, x_2 \leq x_{2\text{ макс}}, x_3 \leq x_{3\text{ макс}}, x_4 = A (x_1 x_2 x_3)^{-1} \leq x_{4\text{ макс}}. \quad (7.116)$$

Очевидно, эта задача имеет решение лишь при выполнении условия $A(x_{1\text{ макс}} x_{2\text{ макс}} x_{3\text{ макс}})^{-1} < x_{4\text{ макс}}$ или эквивалентного ему условия

$$x_{1\text{ макс}} x_{2\text{ макс}} x_{3\text{ макс}} x_{4\text{ макс}} > A. \quad (7.117)$$

Если вместо (7.117) выполняется условие $x_{1\text{ макс}} x_{2\text{ макс}} x_{3\text{ макс}} x_{4\text{ макс}} \leq A$, то дальность действия обеспечивается лишь в том случае, когда значения всех энергетических параметров выбраны максимально допустимыми, т. е. оптимизация этих параметров по критерию минимума стоимости невозможна.

Если же вместо (7.117) имеет место обратное неравенство, т. е. $x_{1\text{ макс}} x_{2\text{ макс}} x_{3\text{ макс}} x_{4\text{ макс}} < A$, то заданная дальность вообще не может быть обеспечена, т. е. исходные данные противоречивы и требуют корректировки. Поэтому оптимизацию по критерию стоимости при наличии ограничений (7.116) следует начинать с проверки выполнения условия (7.117) и проводить лишь при выполнении этого условия.

Методику решения задачи (7.115), (7.116) ради простоты и наглядности рассмотрим на примере радиолонии без ретрансляции, причем используем значения параметров, приведенные в § 7.9.

При идеальной обработке информации и без учета ограничений энергетических параметров сверху для этой радиолонии были получены следующие результаты:

$$\left. \begin{aligned} B_{\Sigma 1} &= 8500 \text{ е. с.}, x^*_{10} = 2,1 \text{ Вт}, x^*_{20} = 0,21 \text{ м}^2, \\ x^*_{30} &= 2,1 \text{ м}^2, x^*_{40} = (1/47) \text{ 1/град.} \end{aligned} \right\} \quad (7.118)$$

При реальной обработке ($\xi=10$)

$$\left. \begin{aligned} B_{\Sigma 2} &= 15300 \text{ е. с.}, x^*_{10} = 3,8 \text{ Вт}, x^*_{20} = 0,38 \text{ м}^2, \\ x^*_{30} &= 3,8 \text{ м}^2, x^*_{40} = (1/26) \text{ 1/град.} \end{aligned} \right\} \quad (7.119)$$

Пусть теперь заданы следующие ограничения:

$$\left. \begin{aligned} x_1 &\leq x_{1\text{ макс}} = 5 \text{ Вт}, x_2 \leq x_{2\text{ макс}} = 0,25 \text{ м}^2, x_3 \leq x_{3\text{ макс}} = 100 \text{ м}^2, \\ x_4 &\leq x_{4\text{ макс}} = (1/300) \text{ 1/град.} \end{aligned} \right\} \quad (7.120)$$

Проверим, не препятствуют ли эти ограничения оптимизации стоимости, т. е. выполняется ли неравенство (7.117). Его левая часть

$$x_{1\text{ макс}} x_{2\text{ макс}} x_{3\text{ макс}} x_{4\text{ макс}} = 5 \cdot 0,25 \cdot 100 \cdot 1/300 = 0,42 \text{ Вт} \cdot \text{м}^4/\text{град.} \quad (7.121)$$

В данном случае $A = A_1 = d^2 \lambda^2 L_1 L_2 \xi \eta K$. Поэтому при идеальной обработке ($\xi=1$)

$$A = 1,5 \cdot 10^6 \cdot 10^{-2} \cdot 10 \cdot 1,38 \cdot 10^{-23} \cdot 10^6 \approx 0,02 \text{ Вт} \cdot \text{м}^4/\text{град.}, \quad (7.122)$$

а при реальной

$$A = 0,2 \text{ Вт} \cdot \text{м}^4/\text{град.} \quad (7.122')$$

Из сравнения (7.121), (7.122) и (7.122') следует, что неравенство (7.117) выполняется как при идеальной, так и при реальной обработке и, следовательно, оптимизация (минимизация) стоимости варьированием энергетических параметров возможна в обоих этих случаях. Определим сначала потенциальную стоимость $B_{\Sigma 2}$ (УПС-2), т. е. положим $\xi=1$ и, следовательно, $A=0,02 \text{ Вт} \cdot \text{м}^4/\text{град.}$ Из сопоставления найденных на первом этапе экстремальных значений (7.118) с их верхними границами (7.120) видим, что одна из переменных, а именно x_4 , вышла за пределы своей верхней границы. Поэтому найденные значения (7.118) с учетом ограничений не являются оптимальными и необходимо перейти ко второму этапу поиска оптимального решения. Для этого полагаем

$$x^*_{40} = x_{4\text{ макс}} = (1/300) \text{ 1/град.} \quad (7.123)$$

При этом минимизируемая целевая функция вместо (7.103) принимает следующий вид:

$$B_{\Sigma} = c_4 x_{4\text{ макс}}^{\alpha_4} + B'_{\Sigma}, \quad (7.124)$$

где

$$B'_{\Sigma} = c_1 x_1^{\alpha_1} + c_2 x_2^{\alpha_2} + c'_3 (x_1 x_2)^{-\alpha_3}, \quad (7.125)$$

$$c'_3 = c_3 (A/x_{4\text{ макс}})^{\alpha_4}. \quad (7.126)$$

Результаты ее безусловной (т. е. без учета ограничений переменных x_1 , x_2 и x_3) минимизации проще всего найти методом геометрического программирования, как показано в приложении 2. Эти результаты имеют следующий вид:

$$\left. \begin{aligned} B_{\Sigma} &= B_{\Sigma\text{ мин}} = \sqrt[3]{(c_1 a_1 \mu'/a_3)^{a_1/a_2} (c_3 a_2 \mu'/a_3)^{a_2/a_2} c'_3 \mu'}, \\ \text{где } \mu' &= 1 + a_1/a_1 + a_2/a_2; \quad x^*_{11} = (B_{\Sigma}/\mu' c_1)^{1/a_1}; \quad x^*_{21} = (B_{\Sigma} a_1/\mu' c_2 a_2)^{1/a_2}; \\ x^*_{40} &= x_{4\text{ макс}}; \quad x^*_{31} = A(x^*_{11} x^*_{21} x^*_{40})^{-1}; \quad \mu' = 1 + a_1/a_2 + a_1/a_3. \end{aligned} \right\} \quad (7.127)$$

В рассматриваемом примере, когда $a_1=a_2=a_3=a_4=1$, эти выражения резко упрощаются и с учетом (7.126) принимают вид

$$\left. \begin{aligned} B_{\Sigma\text{ мин}} &= 3 \sqrt[3]{c_1 c_2 c_3 A/x_{4\text{ макс}}}, \\ x^*_{11} &= B_{\Sigma\text{ мин}}/3c_1, \quad x^*_{21} = B_{\Sigma\text{ мин}}/3c_2, \\ x^*_{31} &= B_{\Sigma\text{ мин}}/3c_3, \quad x^*_{40} = x_{4\text{ макс}}. \end{aligned} \right\} \quad (7.128)$$

Подставляя численные значения параметров, получаем $B_{\text{мин}}^* = 11\,700 \text{ е.с.}$, $x_{11}^* = 3,9 \text{ Вт}$, $x_{21}^* = 0,39 \text{ м}^2$, $x_{31}^* = 3,9 \text{ м}^2$.

Сравнивая эти значения переменных с их верхними границами (7.120), видим, что одна из них, а именно x_{21}^* , выходит за эти границы. Поэтому необходимым третьим этапом поиска оптимального решения, на котором полагаем $x_2^* = x_{2 \text{ макс}} = 0,25 \text{ м}^2$. При этом целевая функция принимает вид

$$B_{\text{н}} = c_2 x_2^{a_2} + c_4 x_4^{a_4} + B_{\text{н}}^*, \quad (7.129)$$

где $B_{\text{н}}^* = c_1 x_1^{a_1} + c_3 x_3^{a_3}$, но $x_3 = A(x_1 x_2 x_4)^{-1}$, поэтому

$$B_{\text{н}}^* = c_1 x_1^{a_1} + c_3 x_1^{-a_3}, \quad (7.130)$$

где $c_3' = c_3(A/x_2 x_4)^{a_3}$.

В рассматриваемом примере $a_1 = a_3 = 1$ и соотношения (7.130) принимают вид

$$B_{\text{н}}^* = c_1 x_1 + c_3' x_1^{-1}, \quad (7.131)$$

где $c_3' = c_3(A/x_2 x_4)$.

Минимум (безусловный) функции (7.131) легко найти из условия $dB_{\text{н}}^*/dx_1 = c_1 - c_3'/x_1^2 = 0$, откуда

$$x_{12}^* = \sqrt{c_3'/c_1} = \sqrt{c_3 A/c_1 x_2 x_4},$$

$$x_{32}^* = A/x_{12}^* x_2 x_4,$$

$$B_{\text{н}}^* = 2\sqrt{c_1 c_3 A/x_2 x_4}. \quad (7.132)$$

Подставляя численные значения параметров c_1 , c_3 , $x_2 \text{ макс.}$, $x_4 \text{ макс.}$ и $A = 0,02 \text{ Вт} \cdot \text{м}^4/\text{град}$ (для $\xi = 1$), получаем $x_{12}^* = 4,9 \text{ Вт}$, $x_{32}^* = 4,9 \text{ м}^2$. Эти значения переменных удовлетворяют ограничениям (7.120). Следовательно, при $\xi = 1$ имеем $x_1^* = P_{\text{н}} = 4,9 \text{ Вт}$, $x_2^* = S_{\text{н}} = 0,25 \text{ м}^2$, $x_3^* = S_{\text{нр}} = 4,9 \text{ м}^2$, $x_4^* = 1/T_{\text{н}} = (1/300) \cdot 1/\text{град}$. При этом из (7.129) и (7.132) следует

$$B_{\text{н2}} = c_2 x_2 + c_4 x_4 + 2\sqrt{c_1 c_3 A/x_2 x_4}, \quad (7.133)$$

и для найденных численных значений параметров $B_{\text{н2}} = 12\,600 \text{ е.с.}$

Для определения $B_{\text{н3}}$ (УПС-3) нужно принять $\xi = 10$, т. е. $A = 0,2 \text{ Вт} \cdot \text{м}^4/\text{град}$. При этом $x_{12}^* = 4,9\sqrt{10} = 16 \text{ Вт}$, $x_{32}^* = 4,9\sqrt{10} = 16 \text{ м}^2$.

Сравнивая эти результаты с ограничениями (7.120), видим, что при реальной обработке ($\xi = 10$) одна из переменных, а именно x_{12} , выходит за допустимую для нее границу. Поэтому следует принять $x_1^* = x_{1 \text{ макс}} = 5 \text{ Вт}$. При этом x_3^* определится из уравнения дальности: $x_3^* = A(x_1 x_2 x_4)^{-1} = 4,8 \text{ м}^2$.

Как и следовало ожидать, значение x_3^* получилось допустимым ($x_3^* < x_{3 \text{ макс}} = 100 \text{ м}^2$), так как в противном случае должно быть $x_1 \text{ макс.} x_2 \text{ макс.} x_4 \text{ макс.} < A$, что противоречит установленному перед началом оптимизации факту выполнения обратного неравенства (7.117).

Таким образом, при реальной обработке ($\xi = 10$) потенциальная стоимость (УПС-3) оказывается равной

$$B_{\text{н4}} = c_1 x_1^* + c_2 x_2^* + c_3 x_3^* + c_4 x_4^* = 55\,800 \text{ е.с.}$$

Следовательно, в рассмотренном примере значения потенциальной стоимости таковы:

- безусловная потенциальная стоимость (БПС) $B_{\text{н1}} = 8500 \text{ е.с.}$;

условная потенциальная стоимость 1-го рода (УПС-1) $B_{\text{н2}} = B_{\text{н1}} \sqrt[4]{\xi} = 15\,300 \text{ е.с.}$;

условная потенциальная стоимость 2-го рода (УПС-2) $B_{\text{н3}} = 12\,600 \text{ е.с.}$;

условная потенциальная стоимость 3-го рода (УПС-3) $B_{\text{н4}} = 55\,800 \text{ е.с.}$

Из этих результатов наглядно видно, насколько увеличивается потенциальная стоимость за счет неидеальности обработки информации (на 6800 е.с.), за счет ограничений энергетических параметров (на 4100 е.с.) и за счет совместного влияния неидеальности обработки и ограничений (на 47\,200 е.с.). Чем более жесткими являются ограничения, тем сильнее сказывается неидеальность обработки, и наоборот, чем больше потери за счет неидеальности обработки, тем сильнее сказывается наличие ограничений.

Этот же пример показывает важность проведения оптимизации энергетических параметров по критерию стоимости, если неравенство (7.117) выполняется с достаточно большим запасом.

7.11. ОБЛАСТИ ПРИМЕНЕНИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Ранее была изложена методика оптимизации стоимости радиолинии путем соответствующего выбора ее энергетических параметров (x_1 , x_2 , x_3 и x_4). При этом полная реальная стоимость радиолинии полагалась состоящей из двух частей: $B_{\text{р}} = B_0 + B_{\text{н}}$, где $B_{\text{н}}$ является некоторой функцией варьируемых энергетических параметров, а B_0 не зависит от этих параметров и называется стоимостью обработки информации (в радиопередающем и радиоприемном устройствах).

Очевидно, лучшие результаты минимизации реальной стоимости можно получить, если варьировать не только энергетические, но и другие параметры и принципы (информационные, конструктивные, технологические и эксплуатационные). При этом изложенную методику определения потенциальных значений стоимости $B_{\text{н1}}$, $B_{\text{н2}}$, $B_{\text{н3}}$ и $B_{\text{н4}}$ можно применить следующими путями.

Первый путь состоит в том, что найденные значения $B_{\text{н1}}$, $B_{\text{н2}}$, $B_{\text{н3}}$ и $B_{\text{н4}}$ используются в основном лишь для определения главных направлений минимизации значений $B_{\text{р}}$, т. е. наиболее перспективной области поиска оптимального решения. Сам же поиск в этой области осуществляется непосредственно по значению $B_{\text{р}}$, т. е. без учета формул потенциальной стоимости.

Второй путь состоит в том, чтобы при минимизации значения $B_{\text{р}}$ воспользоваться приведенными ранее формулами для $B_{\text{н мин}}$. Поскольку наиболее близким к реальной стоимости является значение $B_{\text{н4}}$ (т. е. значение $B_{\text{н мин}}$, найденное для реального значения ξ и с учетом ограничений варьируемых параметров), то при наличии необходимой априорной информации желательно использовать величину $B_{\text{н4}}$ и лишь при отсутствии необходимой информации (на самых начальных этапах проектирования) $B_{\text{н3}}$, $B_{\text{н2}}$ или $B_{\text{н1}}$. Ранее был найден минимум величины $B_{\text{н}}$ при вариациях лишь энергетических параметров x_1 , x_2 , x_3 , x_4 . Но само значение $B_{\text{н мин}}$ ($B_{\text{н3}}$) при этом зависит от коэффициентов стоимости

c_i ($i = \overline{1, n}$), показателей степени a_i ($i = \overline{1, n}$) и параметра A , в который в качестве множителя входит коэффициент ξ , характеризующий потери за счет неидеальности обработки информации. Кроме того, значение $B_{\text{мин}}$ может зависеть от всех или от некоторых из ограничений $x_{1\text{макс}}, x_{2\text{макс}}, x_{3\text{макс}}, x_{4\text{макс}}$. Поэтому воздействие на $B_{\text{мин}}$ прочих (неэнергетических) факторов можно учесть, соответствующим образом скорректировав значения $c_i, a_i, x_{i\text{макс}}$ ($i = \overline{1, n}$) и ξ .

Например, при вариациях принципов построения и параметров системы обработки информации их влияние на значение B_p можно учесть, во-первых, изменив величину $B_{\text{мин}}$ за счет коррекции ξ и, если это окажется необходимым, ограничений $x_{i\text{макс}}$ и, во-вторых, приняв во внимание стоимость B_0 устройства обработки информации. Изменения, вносимые вариациями технологии изготовления, можно учесть коррекцией значений параметров c_i, a_i и $x_{i\text{макс}}$ и т. д.

Однако следует иметь в виду, что комплексная оптимизация радиолинии по критерию стоимости является многомерной и весьма трудоемкой. Методы такой оптимизации (особенно на ранних стадиях проектирования) требуют еще значительного совершенствования.

7.12. ГАБАРИТНЫЕ РАЗМЕРЫ АППАРАТУРЫ

По своей конфигурации все виды радиоаппаратуры можно разбить на объемные и плоские. К плоским принадлежит лишь сравнительно небольшая группа крупных частей радиосистем: панели солнечных батарей, плоские антенные решетки и некоторые другие. Для их пространственной характеристики в первом приближении достаточно одного показателя — площади A ; для более полной характеристики требуется учитывать два габаритных размера: длину l_1 и ширину l_2 , а иногда и три (длину l_1 , ширину l_2 и толщину d). Большинство других частей радиоаппаратуры имеет объемную конфигурацию, и для их пространственной характеристики в первом приближении достаточен один показатель — объем V . Для более полной характеристики могут потребоваться три габаритных размера (длина l_1 , ширина l_2 и высота h).

На начальных этапах проектирования с точки зрения методики учета пространственных показателей (l_1, l_2, d, h, A и V) могут иметь место следующие характерные случаи.

1. Достаточно учесть лишь ограничения сверху на один или несколько габаритных размеров (l_1, l_2, d, h), например обеспечить выполнение условий

$$\begin{aligned} l_1 &\leq l_{1\text{макс}}, \quad l_2 \leq l_{2\text{макс}}, \\ d &\leq d_{\text{макс}}, \quad h \leq h_{\text{макс}}. \end{aligned}$$

(7.134)

2. Необходимо в процессе проектирования обеспечить минимизацию одной из пространственных характеристик, например

$$\min_{S \in M_{\text{сд}}} V \quad (7.135)$$

$$\text{или } \min_{S \in M_{\text{сд}}} A, \quad (7.136)$$

$$\text{или } \min_{S \in M_{\text{сд}}} [\max(l_1, l_2, h)]. \quad (7.137)$$

При этом, очевидно, минимизируемый пространственный параметр считается главным показателем качества системы. Такой случай может иметь место, если существующие системы удовлетворяют их потребителей по всем основным показателям, за исключением размеров, например не могут быть размещены на летательном аппарате заданного типа.

Минимизация пространственных показателей (например, площади A) может потребоваться и в некоторых других задачах проектирования. Например, при установке антенны на ракете или самолете от ее площади может существенно зависеть лобовое сопротивление встречного воздушного потока и желательно выбрать конструкцию антенны такой, чтобы это сопротивление было меньшим.

Однако в большинстве задач проектирования достаточно выполнить лишь условия (7.134) и улучшать не пространственные, а другие показатели качества, например стоимость B или массу m . При учете влияния пространственных показателей на качество частей системы и системы в целом необходимо также иметь в виду, что это влияние в случае антенно-фидерных устройств, солнечных батарей и некоторых других устройств оказывается иным, чем для большинства устройств прочих видов. Действительно, от площади раскрыва антенны непосредственно зависят важнейшие показатели ее качества — коэффициент усиления и коэффициент направленного действия. Площадь солнечной батареи определяет количество падающей на нее солнечной энергии и, следовательно, также непосредственно влияет на ее главный показатель качества — вырабатываемую мощность питания.

В отличие от этого основные показатели качества большинства каскадов передатчика и приемника, ЭВМ и др. зависят от размеров лишь косвенно — увеличение допустимых размеров аппаратуры позволяет отыскивать варианты построения системы, оптимальные в смысле остальных показателей, в более широкой области и, следовательно, получать в принципе лучшие значения этих показателей.

Все эти обстоятельства и приводят к тому, что в большинстве задач проектирования достаточно выполнить лишь условия (7.134).

7.13. МАССА АППАРАТУРЫ

Масса является весьма важным показателем качества системы (устройства), характеризующим это качество в следующих отношениях.

Во-первых, масса в силу ее физической сущности характеризует сопротивляемость системы к неравномерным (во времени) перемещениям. Это свойство может существенно затруднить транспортировку и эксплуатацию аппаратуры, если масса велика.

Во-вторых, в соответствии с законом тяготения сила притяжения, действующая на данную аппаратуру (например, со стороны планеты, на которой или вблизи которой она функционирует), пропорциональна ее массе. Это обстоятельство не только может затруднить эксплуатацию системы, но и значительно увеличить ее стоимость. Например, стоимость доставки с Земли на существенно удаленную от нее орбиту (или планету) аппаратуры массой 1 кг может на порядок или более превысить стоимость производства и эксплуатации этой аппаратуры. В меньшей мере, но существенно это обстоятельство может сказываться и при размещении аппаратуры на самолетах, судах и даже на наземных подвижных объектах.

В-третьих, масса аппаратуры характеризует расход веществ на ее изготовление. Но с увеличением расхода веществ (при их неизменном составе) возрастают объем аппаратуры и ее стоимость.

Из сказанного следует, в частности, что в отличие от большинства других показателей качества масса аппаратуры не косвенно, а непосредственно влияет на два других важнейших показателя качества аппаратуры: ее объем V и стоимость B . Поэтому приближенно можно полагать, что

$$V \approx c_1 m^{a_1}, \quad (7.138)$$

$$B \approx c_2 m^{a_2}, \quad (7.139)$$

где c_1, c_2, a_1, a_2 имеют вполне определенные положительные числовые значения (для аппаратуры данного класса и назначения и на данном уровне развития науки и техники). В первом приближении часто можно полагать

$$a_1 = 1, a_2 = 0,5 \dots 1. \quad (7.140)$$

Однако следует всегда помнить, что соотношения (7.138) и (7.139) носят приближенный характер и в некоторых случаях пользоваться ими недопустимо (например, в задаче минимизации объема или стоимости аппаратуры при заданной ее массе). Необходимо также иметь в виду, что, несмотря на наличие приближенной простой связи (7.139), между учетом массы и стоимости в процессе проектирования имеются и существенные различия, в частности следующие. Почти всегда можно говорить о стоимости системы в целом и минимизировать именно эту стоимость. В то же время говорить о массе системы во многих случаях невозможно, например, если в составе системы имеются назем-

ная и бортовая части. При этом в первую очередь обычно требуется минимизировать массу бортовой аппаратуры. Далее, иногда некоторое превышение стоимости системы по сравнению с заранее заданной, например в 1,5 ... 2 раза, может оказаться все же допустимым, если остальные показатели качества системы достаточно хороши. В то же время такое же (в 1,5 ... 2 раза) превышение массы бортовой части системы может сделать систему совершенно непригодной вследствие невозможности разместить ее на соответствующем объекте (например, на ракете). Иначе говоря, ограничения массы системы сверху обычно являются более жесткими, чем ограничения стоимости.

С другой стороны, дальнейшая минимизация массы аппаратуры, если эта масса уже достаточно мала, в ряде случаев может быть ненужной или даже бесполезной или вредной. В то же время уменьшение стоимости системы (при сохранении допустимыми всех остальных показателей качества) практически всегда полезно. Это объясняется более универсальным характером влияния стоимости, т. е. затрат труда: стоимость (труд), высвобожденная в процессе проектирования данной радиосистемы, можно использовать практически в любой отрасли народного хозяйства, например для дополнительного производства продуктов питания. В отличие от этого масса, сэкономленная при проектировании данной системы, не может быть использована в любой другой отрасли народного хозяйства.

С учетом этих особенностей при проектировании в большинстве случаев показатель массы целесообразно переводить в разряд ограничения $m \leq m_{\max}$, а показатель стоимости рассматривать как главный показатель качества, т. е. требовать его минимизации.

Однако в некоторых специальных случаях может представлять интерес и обратная задача: минимизация массы аппаратуры при наличии ограничений (сверху) стоимости и других (стандартных) показателей качества. Очевидно, решение такой задачи особенно актуально, если существующие системы требуемого класса имеют допустимые значения всех основных показателей качества, за исключением массы.

7.14. ПУТИ УМЕНЬШЕНИЯ МАССЫ И ГАБАРИТНЫХ РАЗМЕРОВ АППАРАТУРЫ

Эти пути можно разбить на три большие группы:

- облегчение исходных данных;
- совершенствование конструкции и технологии производства системы;
- оптимизация (по критерию минимума массы или объема) принципов построения системы и значений ее параметров.

Первый и третий пути имеют тот же характер, что и при минимизации стоимости. Поэтому остановимся лишь на втором. При этом целесообразно исключить из рассмотрения антенно-фидерные устройства, панели солнечных батарей и некоторые другие устройства, размеры которых непосредственно (а не косвенно) влияют

на главные технические показатели этих устройств (коэффициент усиления, коэффициент направленного действия, вырабатываемую мощность питания и др.) и поэтому определяются в первую очередь требуемыми значениями таких показателей. Тогда можно утверждать, что для остальных частей системы минимальные значения их объема лимитируются тремя основными факторами [1]:

минимальными конструктивно и технологически осуществимыми размерами деталей, узлов и соединений между ними; возрастанием паразитных обратных связей; увеличением нагрева аппаратуры из-за повышения мощностей рассеяния на единицу объема.

Последовательными этапами уменьшения массы и габаритных размеров радиоаппаратуры были уменьшение габаритных размеров радиоламп и пассивных элементов, появление транзисторов и печатных схем, создание микромодульных конструкций и интегральных микросхем, повышение степени интеграции микросхем.

За последние 20—25 лет объем малоомощной радиоаппаратуры (лучших ее образцов) сократился примерно в $10^4 \dots 10^5$ раз, но существующие возможности здесь использованы еще далеко не полностью. Это объясняется тем, что значительную часть массы и объема современной аппаратуры составляют не основные (активные или пассивные) элементы, а вспомогательные — соединительные контакты, устройства теплоотвода, герметизации, механической и другой защиты и т. п. Отсюда вытекают основные пути дальнейшей микроминиатюризации радиоаппаратуры: повышение степени интеграции (т. е. замена совокупностей интегральных микросхем едиными интегральными микросхемами с большим числом элементов); создание элементов с минимальным потреблением энергии и устройств с максимальным КПД; использование бескорпусного монтажа (т. е. корпусирование не каждого отдельного элемента или узла, а совокупности большого числа элементов и узлов в целом); применение таких принципов действия и построения системы, при которых можно максимально реализовать возможности современной интегральной технологии.

Для уменьшения массы аппаратуры, кроме указанных мероприятий, следует по возможности заменять материалы с большой удельной массой на материалы с малой удельной массой (например, сталь на пластмассы).

Одним из путей снижения массы и габаритных размеров бортового оборудования является комплексный подход к его проектированию, заключающийся в замене ряда независимых (функционально, динамически и конструктивно) устройств единым совмещенным устройством (комплексом). Однако при таком совмещении необходимо избегать существенного ухудшения надежности (в частности, ремонтпригодности) системы и увеличения ее стоимости (например, вследствие уменьшения возможности использовать в составе системы стандартные модули).

Литература: *основная* [59], *дополнительная* [57, 60—66].

8.1. ОБЩАЯ ХАРАКТЕРИСТИКА ИСХОДНЫХ ДАННЫХ ДЛЯ ПРОЕКТИРОВАНИЯ

Как отмечалось в гл. 1, проектируемая система S должна удовлетворять совокупности исходных данных

$$D = (Y, O_s, \text{с. К, кр. К}). \quad (8.1)$$

Здесь $Y = (y_1, \dots, y_p)$ — совокупность условий работы системы (характеристики полезных сообщений, сигналов и помех, диапазон рабочих температур, давлений, влажности и др.); $O_s = (O_{s_1}, \dots, O_{s_r})$ — совокупность ограничений структуры системы S и значений ее параметров. Ограничения структуры в зависимости от решаемой задачи могут варьироваться от слабых (нежестких) до сильных (жестких). К слабым относится, в частности, ограничение класса систем, например требование, чтобы система передачи информации имела свободный доступ, не содержала обратного (служебного) канала и ретрансляторов. При жестком ограничении структура системы может быть задана полностью и в процессе проектирования могут варьироваться (выбираться) лишь значения отдельных параметров.

Ограничения параметров $x_1, \dots, x_n$ системы могут быть типа равенств ($x_i = x_{i0}$), неравенств ($x_i \leq x_{i \text{ макс}}$ или $x_{i \text{ мин}} \leq x_i \leq x_{i \text{ макс}}$), дискретности, ($x_i = 1, 2, 3, \dots$), связи [$\Phi_i(x_1, \dots, x_n) \leq 0$ или $\Phi_j(x_1, \dots, x_n) = 0$] и иного характера. Вектор $K = \langle k_1, \dots, k_m \rangle$ включает совокупность всех показателей качества $k_1, \dots, k_m$, которые должны учитываться в процессе проектирования. При формулировке исходных данных определяется лишь состав с. К этой совокупности, т. е. указывается, что именно следует понимать под k_1, k_2 и т. д.; численные же значения составляющих $k_1, \dots, k_m$ вектора K зависят от структуры и параметров системы и в процессе проектирования варьируются.

Число m показателей, которые следует учитывать при проектировании, может составлять несколько десятков или даже более. Однако на начальных этапах проектирования принимают во внимание лишь важнейшие из них, обычно не более 10 ... 20. Это в первую очередь показатели точности, пропускной способности, помехоустойчивости, надежности, массы, габаритных размеров и стоимости.

Исходные показатели качества k_{in} ($i = \overline{1, m}$) могут иметь неодинаковые размерности; поэтому составляющими вектора K должны быть не исходные, а нормированные значения показателей

$$k_i = k_{in} / k_{i \text{ н.о.}} \quad (8.2)$$

где $k_{i,0}$ — некоторое опорное (эталонное) значение показателя $k_{i,0}$. Однако если единицы измерения всех показателей $k_{i,0}$ ($i = \overline{1, m}$) фиксированы, то можно выбрать значения $k_{i,0}$ ($i = \overline{1, m}$) единичными. Тогда численные значения показателей $k_{i,0}$ будут совпадать с их нормированными значениями k_i . Например, если $k_{1,0} = 10$ кг — масса системы в килограммах, а $k_{2,0} = 2000$ р. — приведенные затраты на систему в рублях, то, положив $k_{1,0} = 1$ кг, $k_{2,0} = 1$ р., получим $k_1 = 10$, $k_2 = 2000$. В дальнейшем показатели k_i всюду полагаются нормированными.

Вообще говоря, к показателям качества системы можно отнести любой параметр системы, монотонно связанный с ее качеством, т. е. обладающий следующим свойством: чем меньше (или чем больше) данный показатель k_i , тем лучше система при прочих равных условиях, т. е. при фиксированных (неизменных) значениях остальных $(m-1)$ показателей качества (и неизменных исходных данных). Например, стоимость системы B и ее пропускная способность C удовлетворяют этим условиям и, следовательно, являются показателями качества. Утверждать же, что чем больше (или чем меньше) рабочая длина волны λ , тем лучше система (при прочих равных условиях), как правило, нельзя, и, следовательно, этот параметр системы, хотя и является весьма важным, не может быть отнесен к показателям качества, т. е. включен в состав вектора K . Таким образом, все показатели качества можно разбить на две группы:

монотонно убывающие, т. е. такие, что чем меньше k_i , тем лучше система (при прочих равных условиях);

монотонно возрастающие, т. е. такие, что чем больше k_i , тем лучше система (при прочих равных условиях).

Например, масса M , стоимость B , вероятность ошибки $P_{\text{ош}}$ и средний квадрат ошибки $\bar{\delta}^2$ являются монотонно убывающими показателями, а пропускная способность C и вероятность безотказной работы $P_{\text{б.о.}}$ — монотонно возрастающими.

Так как вектор K служит для сравнения систем, удобно, чтобы все входящие в его состав показатели k_i ($i = \overline{1, m}$) имели одинаковый характер, т. е. все были или монотонно убывающими, или монотонно возрастающими. Обычно большинство исходных показателей качества имеют монотонно убывающий характер. Это объясняется тем, что качество системы удобно характеризовать мерой ее отклонения от идеальной системы и, следовательно, чем меньше это отклонение, тем лучше система. Например, точность воспроизведения сообщений характеризуют вероятностью ошибки или средним квадратом ошибки, так как у идеальной (по точности) системы ошибка должна отсутствовать (быть равной нулю). Поэтому в большинстве случаев удобнее приводить все исходные показатели $k_1, \dots, k_m$ к монотонно убывающему (а не к монотонно возрастающему) виду. Кроме того, как выяснится из дальнейшего, удобно выбирать показатели k_i ($i = \overline{1, m}$) такими, чтобы у идеальной (по качеству) системы

значение этих показателей равнялось нулю. В связи с этим в дальнейшем будем полагать, что все исходные показатели $k_1, \dots, k_m$ приводятся к так называемому стандартному виду, т. е. удовлетворяют следующим условиям:

$$k_i > 0;$$

чем меньше k_i , тем лучше система (при прочих равных условиях);

у идеальной (в смысле показателя k_i) системы $k_i \rightarrow 0$.

Нетрудно убедиться, что большинство исходных показателей качества (масса M , стоимость B , вероятность ошибки $P_{\text{ош}}$, средний квадрат ошибки $\bar{\delta}^2$ и др.) удовлетворяют этим условиям, т. е. имеют стандартный вид, и, следовательно, их приведения к стандартному виду не требуется. Однако некоторые исходные показатели качества (пропускная способность C , вероятность безотказной работы $P_{\text{б.о.}}$ и др.) может потребоваться привести к стандартному виду, что можно легко осуществить без какого-либо ущерба для результатов проектирования.

Действительно, пусть, например, неотрицательный исходный показатель качества $k_{i,0}$ таков, что может изменяться в пределах

$$k_{i,0,\text{мин}} \leq k_{i,0} \leq k_{i,0,\text{макс}} \quad (8.3)$$

и чем больше значение $k_{i,0}$, тем лучше система (при прочих равных условиях). Тогда можно выбрать эквивалентный ему (с точки зрения результатов проектирования) стандартный показатель качества

$$k_i = k_{i,0,\text{макс}} - k_{i,0} \quad (8.4)$$

Если $k_{i,0,\text{макс}} \rightarrow \infty$, то вместо (8.4) можно полагать

$$k_i = 1/k_{i,0,\text{макс}} \quad (8.5)$$

Наконец, если (неотрицательный) показатель $k_{i,0}$ удовлетворяет условиям (8.3), но чем он меньше, тем лучше система (при прочих равных условиях), то можно полагать

$$k_i = k_{i,0} - k_{i,0,\text{мин}} \quad (8.6)$$

В частном случае, когда $k_{i,0,\text{мин}} = 0$, получается $k_i = k_{i,0}$, т. е. исходный показатель $k_{i,0}$ является стандартным и, следовательно, преобразовывать его не требуется. Пусть, например, $k_{1,0} = C$ — пропускная способность системы. У идеальной системы $C \rightarrow \infty$, т. е. $k_{1,0,\text{макс}} \rightarrow \infty$. Поэтому в соответствии с (8.5) следует полагать $k_1 = 1/C$.

Пусть $k_{2,0} = P_{\text{б.о.}}$ — вероятность безотказной работы (за время t_3). Тогда $k_{2,0,\text{макс}} = 1$ и в соответствии с (8.4) следует полагать $k_2 = 1 - k_{2,0} = 1 - P_{\text{б.о.}} = P_{\text{отк}}$, где $P_{\text{отк}}$ — вероятность отказа (за время t_3).

Таким образом, все исходные показатели качества $k_{1,0}, \dots, k_{m,0}$ путем элементарных преобразований всегда можно привести к безразмерному стандартному виду. В дальнейшем будем пола-

гать, что вектор качества имеет именно такой стандартный вид. При этом сравнение систем (вариантов их построения) удобно производить в m -мерном пространстве R_m показателей качества (рис. 8.1). В этом пространстве каждой совокупности $\langle k_1, \dots, k_m \rangle$ показателей качества соответствует m -мерный вектор $K = \langle k_1, \dots, k_m \rangle$, проведенный из начала координат. При этом системе, идеальной (наилучшей) в смысле всех m показателей качества, соответствует нулевое значение вектора K , т. е. точка, совпадающая с началом координат. Однако у любой реальной системы вектор K отличен от нуля и его концом является некоторая точка A в R_m , не совпадающая с началом координат. При этом каждой точке в R_m соответствует некоторое значение вектора K .

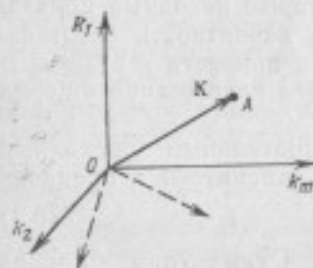


Рис. 8.1

Выясним, в каком соответствии находятся множество M_S систем и множество M_K соответствующих им векторов (точек) в R_m . Характер этого соответствия зависит от того, что именно вкладывается в понятия «система S » и ее «вектор качества K ». Действительно, качество реальной (т. е. уже изготовленной) системы невозможно совершенно полно охарактеризовать конечным числом показателей. Например, даже одно свойство цветного телевизора, а именно качество даваемого им изображения, строго говоря, невозможно охарактеризовать конечным числом m показателей, всегда найдется еще некоторый дополнительный $(m+1)$ -й показатель k_{m+1} , который будет дополнительно характеризовать качество изображения. Кроме того, многие показатели качества данной вполне определенной системы зависят от условий ее функционирования, которые лишь весьма приближенно можно сформулировать в виде исходных данных. Следовательно, даже для вполне определенной технической системы вектор ее качества $K = \langle k_1, \dots, k_m \rangle$, строго говоря, не может быть вполне определенным ни по своему составу (т. е. по числу и виду учитываемых показателей), ни по численному значению каждой составляющей. Поэтому для обеспечения практической сравнимости таких систем приходится по существу заменять условия U работы системы и значения ее вектора качества K их моделями (т. е. приближенным описанием) — полагать, что система S будет функционировать во вполне определенных условиях U и иметь в этих условиях вполне определенное значение конечно-мерного вектора качества K .

В таком приближении каждой реальной системе S будет соответствовать одно вполне определенное значение вектора качества K . При этом, вообще говоря, возможны такие случаи, в которых несколько систем, различающихся принципами построения и (или) значениями отдельных параметров, будут иметь практически одинаковые значения вектора K . Однако, если число

m учитываемых показателей не слишком мало, такое совпадение значений вектора K будет весьма редким исключением. Поэтому при сравнении реальных технических систем можно практически полагать, что между системой S и вектором ее качества K имеется взаимно-однозначное соответствие.

Однако в процессе проектирования, особенно на его начальных этапах, разработчикам приходится сравнивать не реальные системы, а лишь возможные варианты их построения. При этом каждый вариант S построения системы можно описать лишь приближенно, т. е. вместо реальной системы разработчику приходится иметь дело с некоторой моделью системы, являющейся ее приближенным описанием. На начальных этапах проектирования, и особенно на стадии научных исследований, модель системы может быть весьма упрощенной. Например, при математическом синтезе линейного стационарного фильтра, дающего на выходе максимальное отношение сигнал-шум, под фильтром понимают любую линейную стационарную цепь, обладающую определенной формой комплексной передаточной функции $K_\Phi(j\omega)$. При этом отвлекаются от целого ряда других характеристик фильтра: коэффициента усиления на опорной частоте, массы, габаритных размеров, стоимости и др.

Очевидно, одной и той же модели S системы может соответствовать целый класс реальных систем S_p , различающихся свойствами, не учтенными при формировании модели¹. Однако на каждом этапе проектирования (за исключением заключительного, когда система уже изготовлена полностью) достаточно знать лишь соответствующую модель системы и доводить задачу сравнения систем до выбора наилучшей модели S системы, т. е. до выбора класса (множества) реальных систем, соответствующих данной модели.

По мере перехода от начальных этапов проектирования к заключительным этот класс реальных систем последовательно сужается. Поэтому в дальнейшем под системой S понимается соответствующая (данному этапу проектирования или научным исследованиям) модель системы. Но при этом возникает вопрос: имеется ли взаимно-однозначное соответствие между моделью системы и ее вектором качества K . Причем под вектором качества понимается, как было принято, конечно-мерный вектор $K = \langle k_1, \dots, k_m \rangle$. Вообще говоря, одной и той же модели S системы может соответствовать не одно, а несколько значений вектора качества K .

Для того чтобы убедиться в этом, рассмотрим приведенный в [58] пример синтеза оптимальной по помехоустойчивости формы передаточной функции $K_\Phi(j\omega)$ линейного стационарного фильтра при действии на его входе суммы сигнала и шума $y(t) = u_c(t) + u_m(t)$, где

¹ Очевидно также, что одной и той же реальной системе соответствует не единственное ее приближенное описание, т. е. не единственная модель.

$$u_e(t) = \begin{cases} U_{m.e} \cos(\omega_e t + \Psi_e) & \text{при } t \in [0, T], \\ 0 & \text{при } t < 0, t > T; \end{cases}$$

$T \gg 2\pi/\omega_e$; $u_{\Sigma}(t)$ — белый шум с энергетическим спектром N_0 .

Под системой S здесь понимается определенная (с точностью до постоянного множителя) форма передаточной функции $K_{\Phi}(j\omega)$. Вектор качества K этой системы нужно выбрать таким, чтобы, во-первых, он правильно характеризовал помехоустойчивость этой системы (фильтра) и, во-вторых, позволял достаточно просто отыскать оптимальную по этому показателю систему. Рассматриваемая система (фильтр) является частью некоторой более сложной системы, например радиоприемного устройства. Поэтому помехоустойчивость фильтра надо оценивать таким вектором K , чтобы его лучшему значению соответствовала более высокая помехоустойчивость не отдельно взятого фильтра, а радиоприемного устройства в целом (более строго — даже еще более высокой по иерархии системы, в состав которой входит радиоприемное устройство).

Если фильтр включен в каскадах приемника, предшествующих детектору, то, как показано, например, в [5], во многих задачах достаточно простой и правильной оценкой помехоустойчивости фильтра является отношение шума к сигналу на выходе фильтра вида $U_{\Sigma}/U_{m.e}$, где U_{Σ} — среднеквадратическое значение выходного напряжения шума, а $U_{m.e}$ — максимальное значение выходного напряжения сигнала. Поэтому вектор качества K можно выбрать в виде

$$K = k_1, \quad (8.7)$$

где $k_1 = U_{\Sigma}/U_{m.e}$, т. е. по существу заменить векторный показатель качества скалярным.

Решение этой задачи имеет, как известно, следующий вид (см., например, [5]). Наилучшим (минимальным) значением показателя k_1 обладает фильтр с передаточной функцией

$$K_{\Phi}(j\omega) = a S_e(-j\omega) \exp(-j\omega t_0), \quad (8.8)$$

где $S_e(j\omega)$ — комплексный спектр входного сигнала $u_e(t)$; a — произвольная константа; t_0 — момент времени, в который полезный сигнал на выходе фильтра достигает наибольшего значения. Величина t_0 может выбираться любой, но не меньшей, чем T .

При решении данной задачи можно было бы выбрать вектор качества и не в виде (8.7), а в следующем более сложном виде:

$$K = \langle k_1, k_2, k_3 \rangle, \quad (8.9)$$

где $k_1 = U_{\Sigma}/U_{m.e}$; $k_2 = 1/N_0$, N_0 — максимально допустимая для системы интенсивность (спектральная плотность) помехи.

Действительно, не только k_1 , но и k_2 и k_3 удовлетворяют требованиям, предъявляемым к (стандартным) показателям качества: чем меньше каждый из них, тем лучше система при неизменных значениях остальных показателей. Например, чем меньше k_3 (т. е. чем больше N_0) при фиксированных значениях отношения $k_1 = U_{\Sigma}/U_{m.e}$ и амплитуды сигнала $U_{m.e}$, тем лучше система, так как позволяет получать такую же точность (такое же значение k_1) при большей интенсивности N_0 шума без увеличения амплитуды (и длительности) сигнала.

Однако в рассматриваемой задаче дополнительные показатели качества k_2 и k_3 являются излишними, так как не влияют на результат решения задачи. Действительно, здесь, во-первых, требуется синтезировать фильтр приемника, а показатели $k_2 = U_{\Sigma}/U_{m.e}$ и $k_3 = 1/N_0$ характеризуют качество других частей более

высокой по иерархии (по сравнению с фильтром) системы: радиопередатчика ($U_{m.e}$) и входных каскадов приемника (N_0). Во-вторых, для решения поставленной задачи (отыскания оптимальной по критерию k_1 формы передаточной функции фильтра) учета показателей k_2 и k_3 не требуется: из (8.8) следует, что результат синтеза (оптимальная форма передаточной функции фильтра) зависит не от значений k_2 и k_3 , а лишь от вида комплексного спектра $S_e(j\omega)$ сигнала, т. е. от формы входного сигнала $u_e(t)$ (амплитуда $U_{m.e}$ этого сигнала не влияет, так как постоянный множитель a может быть произвольным).

Таким образом, в рассматриваемой задаче синтеза фильтра одной и той же системе (фильтру) могут соответствовать совершенно различные даже по составу векторы качества: $K = k_1$ и $K = \langle k_1, k_2, k_3 \rangle$. Однако при этом показатели k_2 и k_3 являются избыточными, так как не влияют на результат синтеза и лишь усложняют процесс его получения.

Поэтому при формулировании задачи проектирования (синтеза, оптимизации) желательно устранять появление избыточных показателей качества. В дальнейшем будем полагать, что избыточные показатели качества отсутствуют. При этом каждой системе (модели системы) S будет соответствовать единственное значение вектора качества K . Обратное утверждение справедливо не всегда: данным значением вектора качества K может обладать, вообще говоря, не одна вполне определенная система (модель), а некоторое множество таких систем. Однако наличие множества систем (моделей) — скорее исключение, чем правило, и это исключение является особенно редким, если вектор качества K имеет не одну, а несколько составляющих.

С учетом изложенного можно утверждать следующее.

1. В общем случае между реальной системой (или ее моделью) S и ее вектором качества K взаимно-однозначного соответствия нет: одной и той же системе (модели) могут соответствовать различные векторы K , а данным значением K может обладать множество систем (моделей).

2. Если состав вектора K фиксирован (и число показателей конечно), то каждой реальной системе соответствует единственное значение вектора K .

3. Если избыточные показатели качества отсутствуют, то каждая модель системы обладает единственным значением вектора K .

4. Чем больше показателей качества содержит вектор K , тем уже класс систем (или моделей систем), обладающих одним и тем же значением вектора K .

5. Из сказанного следует, что между системой S и ее вектором качества K существует взаимно-однозначное соответствие при совместном выполнении следующих условий:

состав вектора K фиксирован и включает конечное число показателей (при сравнении реальных систем) или отсутствуют избыточные показатели качества (при сравнении моделей систем); под системой (моделью системы) понимается класс систем, обладающих одним и тем же значением вектора K .

В дальнейшем будем полагать, что эти условия выполняются и, следовательно, множества M_S и M_K систем и их показателей качества эквивалентны: каждой системе S соответствует единственная, вполне определенная точка в M_K (т. е. в R_m), и наоборот, каждой точке из M_K (R_m) соответствует вполне определенная система (класс систем) S .

Обозначим

$$D' = (Y, O_S, c, K) \quad (8.10)$$

— часть исходных данных D [см. (8.1)], не включающую критерий качества кр. K . Будем называть систему (класс систем) S , удовлетворяющую исходным данным D' , *допустимой* (возможной) *системой*.

В общем случае существует не одна допустимая система, а некоторое множество M_{dS} таких систем. В пространстве R_m этим системам будет соответствовать множество M_{dK} допустимых значений вектора качества K (допустимых точек). Так как между системами (классами систем) S и значениями векторов их качества K существует взаимно-однозначное соответствие, множества M_{dS} и M_{dK} эквивалентны: каждому элементу из M_{dS} соответствует вполне определенный элемент (определенная точка) в M_{dK} , и наоборот. Поэтому в дальнейшем оба эти множества будем одинаково обозначать M_d и полагать, что речь идет о допустимом множестве (систем или точек в R_m).

Каждая точка в R_m , не принадлежащая M_d , является недопустимой (невозможной), т. е. не удовлетворяет исходным данным (8.8). Пусть, например, рассматривается вопрос о выборе обнаружителя сигнала по двум показателям качества: вероятности P_{np} пропуска сигнала и вероятности $P_{лт}$ ложной тревоги, т. е. $K = \langle k_1, k_2 \rangle$, где $k_1 = P_{np}$; $k_2 = P_{лт}$. Тогда при фиксированных исходных данных D' (при фиксированных характеристиках сигнала и помех и заданных ограничениях структуры и параметров обнаружителя) множество M_d допустимых точек (в пространстве R_m)

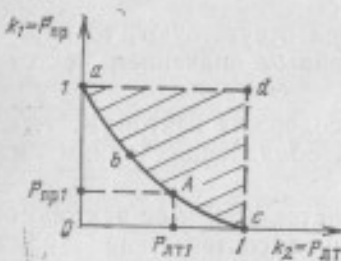


Рис. 8.2

будет иметь вид, изображенный на рис. 8.2 в виде заштрихованной области $abcd$: любая точка этой области (любые значения вектора $K = \langle k_1, k_2 \rangle$, принадлежащие этой области) является допустимой, т. е. может быть реализована при сформулированных исходных данных D' . Если ослабить требования, входящие в исходные данные, например увеличить энергию сигнала или уменьшить интенсивность помехи, то область M_d расширится (за счет приближения к началу координат границы abc). Если, напротив, ужесточить эти требования (потребовать обнаружения сигнала, имеющего малую энергию, на фоне весьма интенсивной помехи), то область M_d сузится.

8.2. ВИДЫ КРИТЕРИЕВ КАЧЕСТВА

Исходные данные D' определяют множество M_d допустимых точек (систем), из которых следует выбрать наилучшую или по меньшей мере приемлемую точку (систему). Чтобы осуществить такой выбор, необходимо сформулировать критерий качества кр. K системы, входящий в полные исходные данные D [см. (8.1)].

Сформулировать критерий качества системы так, чтобы он годился «на все случаи жизни», т. е. для любых видов систем и при любых условиях их применения, невозможно, поскольку все явления в природе взаимосвязаны и находятся в состоянии изменения, а множество этих явлений неограничено: то, что хорошо для одних условий, на одном пространственно-временном интервале, может оказаться неудовлетворительным для других условий, на другом пространственно-временном интервале. Поэтому любой критерий качества является в какой-то мере условным и имеет ограниченную область применения. В практике проектирования наиболее распространено его разбиение на критерий приемлемости и критерий предпочтения. Критерий приемлемости устанавливает границу между приемлемыми и неприемлемыми значениями вектора качества K : по одну сторону от этой границы (в направлении начала координат) значения вектора K считаются с точки зрения назначения системы приемлемыми, а по другую сторону — неприемлемыми. Например, при двух показателях качества эта граница может иметь вид кривой abc на рис. 8.3: все значения

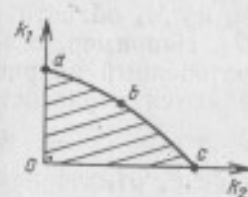


Рис. 8.3

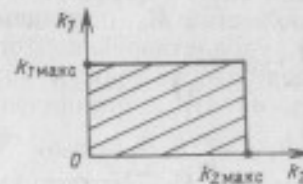


Рис. 8.4

вектора K (все точки, являющиеся концами этого вектора), принадлежащие области $Oabc$ (заштрихованной), считаются приемлемыми, а не принадлежащие этой области — неприемлемыми (неудовлетворительными). В общем случае уравнение этой области можно записать в виде¹

$$f(k_1, k_2, \dots, k_m) \leq 0. \quad (8.11)$$

При этом приемлемое значение каждого показателя качества k_i зависит от значений остальных показателей. Например, на рис.

¹ В (8.11) знак равенства соответствует границе области, а знак меньше или равно указывает, что приемлемые точки должны находиться внутри этой области или на ее границе.

8.3 приемлемое значение показателя k_1 тем больше, чем меньше k_2 , и, наоборот, приемлемое значение k_2 уменьшается с ростом k_1 . Для упрощения критерия приемлемости большей частью принимают, что приемлемое значение каждого показателя k_i не зависит от значений остальных показателей, т. е. заменяют одно сложное (функциональное) неравенство (8.11) системой ограничений вида

$$k_1 \leq k_{1 \text{ макс}}, \quad k_2 \leq k_{2 \text{ макс}}, \dots, \quad k_m \leq k_{m \text{ макс}}, \quad (8.12)$$

где $k_{1 \text{ макс}}, \dots, k_{m \text{ макс}}$ — максимальные (предельно допустимые) значения показателей качества¹ $k_1, \dots, k_m$. При этом криволинейная граница (рис. 8.3) заменяется линейно-ломаной (рис. 8.4).

Обозначим: O_h — совокупность ограничений показателей качества вида (8.12) [или (8.11)], устанавливающая границу между приемлемыми (с точки зрения назначения системы) и неприемлемыми (неудовлетворительными) значениями вектора качества системы;

$$D^* = (D', O_h) = (V, O_s \text{ с. К. } O_h) \quad (8.13)$$

— совокупность исходных данных, отличающихся от полной совокупности D [см. (8.1)] только тем, что критерий качества кр. К учен лишь введением критерия приемлемости O_h .

Каждая система, удовлетворяющая совокупности исходных данных (8.13), является не только допустимой [т. е. удовлетворяющей исходным данным (8.10)], но и приемлемой. Такую систему будем называть *строго допустимой*. В общем случае существует не одна строго допустимая система, а некоторое множество M_{cd} таких систем (точек). Очевидно, множество M_{cd} является частью множества M_d , получаемой выделением из M_d области (множества), удовлетворяющей ограничениям O_h . Например, если M_d имеет вид, изображенный на рис. 8.2 и повторенный на рис. 8.5 (область $abcd$), а ограничения O_h определяются неравенствами

$$k_1 \leq k_{1 \text{ макс}}, \quad k_2 \leq k_{2 \text{ макс}}, \quad (8.14)$$

то множеству M_{cd} соответствует область $a'b'c'd'$, отмеченная на рис. 8.5 штриховкой.

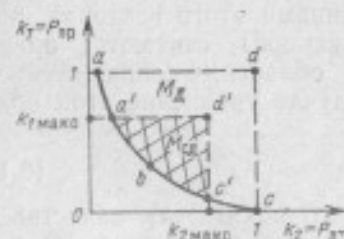


Рис. 8.5

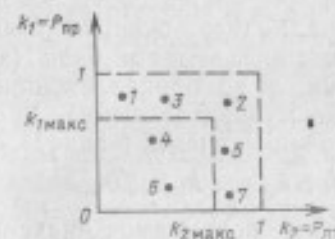


Рис. 8.6

Множества M_d и M_{cd} , изображенные на рис. 8.5, содержат континуум точек. Однако в некоторых случаях эти множества могут содержать лишь конечное число точек [например, на рис. 8.6 множество M_d содержит семь точек, из которых только две (4 и 6) являются строго допустимыми, т. е. образуют множество M_{cd}]. Это имеет место, если ограничения O_h , накладываемые на структуру системы и значения ее параметров, являются предельно жесткими: допускают лишь конечное число N вполне определенных систем (или вариантов построения системы), имеющих фиксированные значения всех параметров. В этом случае множество M_d содержит всего N точек, из которых в M_d входят лишь точки, удовлетворяющие ограничениям O_h показателей качества.

Этот случай, когда требуется сравнивать лишь конечное число систем (вариантов построения систем), называется в дальнейшем *задачей дискретного выбора систем*. В отличие от него при синтезе структуры системы и (или) оптимизации значений ее параметров множества M_d и M_{cd} содержат континуум точек. При этом в зависимости от вида ограничений O_h эти множества могут быть выпуклыми (при достаточно слабых ограничениях) или невыпуклыми (при более жестких ограничениях).

Как известно, множество M называется выпуклым, если все точки отрезка прямой, соединяющего любые две точки множества M , принадлежат этому множеству; в противном случае множество M называется невыпуклым. Например, множество M_{cd} , изображенное на рис. 8.7, а, выпуклое, а на рис. 8.7, б — невыпуклое: у отрезка прямой bc , соединяющего точки b и c этого множества, не все точки (а лишь граничные точки b и c) принадлежат этому множеству.

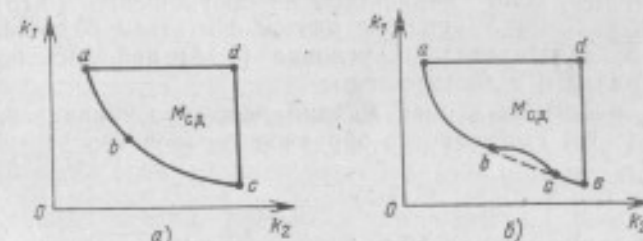


Рис. 8.7

Все строго допустимые системы (соответственно все точки множества M_{cd}) с точки зрения их применения являются приемлемыми. Однако желательно из всего множества M_{cd} таких систем выбрать в каком-то смысле наилучшую или по меньшей мере отсеять безусловно худшие варианты построения системы. Для решения этой задачи необходимо сформулировать соответствующий критерий предпочтения одной системы (одного значения вектора K ее качества) другой системе (другому значению вектора K).

¹ Здесь, как и ранее, показатели $k_1, \dots, k_m$ полагаются имеющими стандартный вид.

Рассмотрим сначала так называемый *безусловный критерий предпочтения* (БКП)¹. Пусть сравниваются две системы (два варианта их построения) S' и S'' , которым соответствуют векторы качества $K' = \langle k'_1, \dots, k'_i, \dots, k'_m \rangle$ и $K'' = \langle k''_1, \dots, k''_i, \dots, k''_m \rangle$; имеющие стандартный вид, т. е. чем меньше k_i , тем лучше система при неизменных значениях остальных ($m-1$) показателей. Очевидно, необходимое и достаточное условие того, что система S'' лучше, чем S' , имеет вид

$$k''_i \leq k'_i \text{ для всех } i = \overline{1, m}, \quad (8.15)$$

и по меньшей мере для одного номера $i=i_0$ выполняется строгое неравенство

$$k''_{i_0} < k'_{i_0}.$$

Условие (8.15) сокращенно записывается так:

$$K'' \leq K'. \quad (8.16)$$

Действительно, из (8.15) или (8.16) следует, что у системы S'' все показатели качества не хуже, чем у системы S' , и по меньшей мере один из них не только не хуже, а лучше. Но это означает, что система S'' безусловно лучше, чем S' . Конечно, если у системы S'' будет меньше (лучше) не один, а несколько показателей качества, то она также будет безусловно лучше, чем S' , но это условие, будучи достаточным, не является необходимым. С другой стороны, если какие-либо из соотношений (8.15) не выполняются, система S'' уже не может считаться безусловно лучшей, чем S' . Следовательно, условие (8.15) действительно является необходимым и достаточным.

Путем аналогичных рассуждений нетрудно убедиться, что если вместо (8.16) выполняется обратное неравенство

$$K'' \geq K', \quad (8.17)$$

т. е. в развернутом виде $k''_i \geq k'_i$ для всех $i = \overline{1, m}$, и по меньшей мере для одного значения $i=i_0$ выполняется строгое неравенство $k''_{i_0} > k'_{i_0}$, то система S'' безусловно хуже, чем S' . Наконец, если

$$K'' = K', \quad (8.18)$$

т. е. $k''_i = k'_i$ для всех $i = \overline{1, m}$, то системы S'' и S' эквивалентны по качеству.

Таким образом, БКП может быть сформулирован следующим образом:

¹ Его называют также критерием Парето, по имени ученого, впервые предложившего этот критерий в 1896 г.

$$\left. \begin{array}{l} \text{если } K'' \leq K', \text{ то система } S'' \text{ безусловно лучше,} \\ \text{чем } S', \\ \text{если } K'' \geq K', \text{ то система } S'' \text{ безусловно хуже,} \\ \text{чем } S', \\ \text{а при равенстве } K'' = K' \text{ системы } S'' \text{ и } S' \text{ экви-} \\ \text{валентны по качеству.} \end{array} \right\} \quad (8.19)$$

Однако возможны такие случаи, когда для векторов K'' и K' не будет выполняться ни одно из соотношений (8.19). Например, если $K' = \langle 1, 10 \rangle$, а $K'' = \langle 2, 5 \rangle$, то $K'' \neq K'$, но несправедливы и неравенства $K'' \leq K'$ и $K'' \geq K'$. В таких случаях системы S'' и S' не могут быть признаны ни эквивалентными, ни безусловно превосходящими одна другую по качеству: они оказываются по БКП *несравнимыми*. Такие несравнимые по БКП системы, как будет показано в гл. 9, относятся к классу так называемых *нехудших* систем. Для их сравнения необходимо ввести какой-либо другой критерий предпочтения, называемый *условным критерием предпочтения* (УКП), ибо необходимо дополнительно условиться, по какому принципу одной комбинации показателей качества $\langle k'_1, \dots, k'_m \rangle$ должно отдаваться предпочтение по сравнению с другой комбинацией $\langle k''_1, \dots, k''_m \rangle$ этих показателей, т. е. одному значению вектора качества K' по сравнению с другим его значением K'' .

Для этого вводится результирующая скалярная функция векторного аргумента

$$K_{\text{рез}} = f_{\text{рез}}(k_1, \dots, k_m) = f_{\text{рез}}(K), \quad (8.20)$$

где $K_{\text{рез}}$ — скалярная величина, называемая *результирующим (обобщенным) показателем качества*; $f_{\text{рез}}(k_1, \dots, k_m)$ — некоторая вполне определенная функция показателей качества $k_1, \dots, k_m$, называемая *результирующей целевой функцией*; ее вид выбирается (обосновывается) исходя из назначения системы. Например, иногда оказывается возможным и целесообразным полагать, что

$$K_{\text{рез}} = c_1 k_1 + c_2 k_2 + \dots + c_m k_m, \quad (8.21)$$

где c_i — весовые коэффициенты, удовлетворяющие условиям

$$c_i > 0, \quad \sum_{i=1}^m c_i = 1.$$

Более подробно вопрос о выборе вида результирующей целевой функции $f_{\text{рез}}(k_1, \dots, k_m)$ и ее применении при проектировании рассматривается в гл. 9. Пока же отметим лишь некоторые общие свойства этой функции. Вид функции $f_{\text{рез}}(\cdot)$ можно выбрать таким, что чем меньше $K_{\text{рез}}$, тем лучше система. Такая результирующая целевая функция называется *функцией потерь*: она как бы назначает потери, соответствующие той или иной комбинации показателей $(k_1, \dots, k_m)$: чем меньше потери, тем лучше система. Можно выбрать вид функции $f_{\text{рез}}(\cdot)$ и таким, что чем больше

величина $K_{\text{рез}}$, тем лучше система. Такая результирующая целевая функция называется *функцией полезности* (выгоды).

Поскольку частные показатели качества $k_1, \dots, k_m$ мы условились выбирать таким образом, что чем меньше каждый из них, тем лучше система (при прочих равных условиях), то удобно и результирующий показатель качества $K_{\text{рез}}$ выбрать таким, чтобы улучшению системы соответствовало уменьшение (а не увеличение) значения $K_{\text{рез}}$, т. е. выбрать функцию $f_{\text{рез}}(\cdot)$ в классе функций потерь, а не функций полезности. При этом, если функция $f_{\text{рез}}(\cdot)$ выбрана вполне корректно, то она должна монотонно возрастать по каждому аргументу [при фиксированных значениях остальных $(m-1)$ аргументов]. Действительно, при возрастании каждого аргумента, например k_2 (при фиксированных значениях остальных аргументов $k_1, k_3, \dots, k_m$), качество системы ухудшается (это следует из определения каждого частного показателя качества k_i). Но это означает, что и величина, характеризующая качество системы (потери), должна также возрастать. Нетрудно убедиться, что, например, функция (8.21) удовлетворяет этому требованию. Как будет показано в гл. 9, иногда ради упрощения процесса сравнения систем (отыскания оптимальной системы) приходится отступать от требования полной корректности результирующей целевой функции и выбирать ее в виде не монотонно возрастающей, а лишь неубывающей функции.

Поскольку результирующий показатель качества $K_{\text{рез}}$ является скалярной величиной, а не векторной, введение функции (8.20) означает по-существу сведение векторной задачи сравнения (выбора, оптимизации) систем к скалярной — к сравнению по единственному, скалярному показателю $K_{\text{рез}}$. Такая скаляризация задачи позволяет в ряде случаев не только существенно упростить процесс ее решения, но и сравнить между собой такие системы, которые по БКП оказываются принципиально несравнимыми. Действительно, при сравнении двух вариантов S' и S'' систем по их результирующему показателю качества $K_{\text{рез}}$ возможны лишь следующие случаи: $K''_{\text{рез}} < K'_{\text{рез}}$, $K''_{\text{рез}} > K'_{\text{рез}}$, $K''_{\text{рез}} = K'_{\text{рез}}$. В первом случае система S'' лучше, чем S' , во втором — хуже, чем S' , а в третьем — варианты S'' и S' эквивалентны. Следовательно, при применении УКП, в отличие от БКП, несравнимых вариантов быть не может, что является основным преимуществом УКП. Однако это преимущество УКП сопровождается весьма серьезным недостатком: требуется обосновать вид результирующей целевой функции (8.20), что, как правило, возможно лишь со значительной степенью условности, произвола, субъективности. Например, в ряде случаев невозможно достаточно строго обосновать, почему эта функция должна быть линейной комбинацией (8.21) показателей качества и каковы должны быть значения весовых коэффициентов $c_1, \dots, c_m$. Поэтому критерии вида (8.20) в отличие от безусловного критерия (8.19) названы условными.

Однако следует отметить, что критерий (8.19), являясь по

сравнению с (8.20) безусловным, строго говоря, также не может считаться абсолютно безусловным, так как в природе вообще ничего абсолютно безусловного не существует. Условность критерия (8.19) состоит в том, что при его применении система характеризуется вполне определенным и притом обычно весьма ограниченным числом m показателей качества $k_1, \dots, k_m$. Пусть по этим показателям система S'' оказалась безусловно лучшей S' . Но может существовать некоторый дополнительный показатель качества k_{m+1} , который у системы S'' окажется хуже, чем у S' (т. е. $k''_{m+1} > k'_{m+1}$). Тогда по этим $(m+1)$ показателям система S'' уже не может быть признана безусловно лучшей S' . Поэтому утверждение «система S'' безусловно лучше (или безусловно хуже, или эквивалентна), чем S' », всегда справедливо лишь при выполнении некоторых предварительных условий (т. е. для некоторых вполне определенных исходных данных) и, следовательно, в какой-то мере условно. Однако условность такого рода (вызванная условностью исходных данных) имеет место и при формулировке УКП. Поэтому принятое деление критериев предпочтения на безусловный и условный, хотя и является само по себе условным, оправдано тем, что при введении УКП (по сравнению с БКП) появляется дополнительная и притом весьма существенная условность, связанная с необходимостью выбора (обоснования) вида функции потерь (8.20).

Таким образом, введение УКП вместо БКП обычно позволяет весьма существенно упростить процедуру сравнения (выбора, оптимизации) различных вариантов построения системы и избежать возможности появления несравнимых вариантов. Однако это преимущество достигается путем введения ряда дополнительных условностей и произвола. С учетом этих особенностей БКП применяется, как правило, на стадии научных исследований и на начальных этапах проектирования, когда допустимо учитывать лишь сравнительно небольшое число показателей качества ($m \leq 5 \dots 10$) и в то же время особенно нежелательно вводить произвол, связанный с конкретизацией вида функции потерь. С другой стороны, на заключительных этапах проектирования, когда необходимо принимать во внимание большое число показателей качества и производить сравнение, включая варианты, оказавшиеся по БКП несравнимыми, применение какого-либо конкретного УКП оказывается не только оправданным, но и необходимым.

В ряде случаев достаточно убедительно обосновать применение УКП ко всей совокупности $k_1, \dots, k_m$ исходных показателей качества не удается, но оказывается возможным дать необходимое обоснование для отдельных групп показателей. Вводимые для таких групп результирующие показатели качества, в отличие от результирующего показателя $K_{\text{рез}}$ вида (8.20), объединяющего все m показателей, будем называть *приведенными* и обозначать

$$K_{\text{г}} = f_{\text{г}}(k_j, \dots, k_l), \quad (8.22)$$

где $(k_j, \dots, k_l)$ — группа показателей качества из общей совокуп-

ности ($k_1, \dots, k_m$) этих показателей; $f_n(\cdot)$ — некоторая вполне определенная функция показателей $k_1, \dots, k_i$, которую следует обосновать исходя из назначения системы и условий ее функционирования. В большинстве случаев наиболее просто обосновать сведение группы показателей к одному приведенному показателю в тех случаях, когда эта группа состоит из *однородных* показателей качества, т. е. показателей, характеризующих одно и то же свойство системы, например ее точность или помехоустойчивость. К таким приведенным показателям относится, в частности, целый ряд уже описанных показателей: приведенные затраты P , средний риск R и др.

Использование приведенных показателей позволяет сократить исходное число m показателей до значительно меньшего числа m' приведенных показателей. Действительно, как отмечалось в гл. 2 и 7, точность воспроизведения дискретных сообщений в общем случае характеризуется совокупностью (матрицей) большого числа условных вероятностей ошибок, а стоимость системы — большим числом затрат различных видов; поэтому введение приведенных показателей P и R позволяет сократить число показателей, характеризующих два свойства системы (точность и стоимость), с нескольких десятков или даже сотен показателей до двух.

После введения приведенных показателей качество системы характеризуется совокупностью (вектором) приведенных показателей $K_n = \langle k_{n1}, \dots, k_{nj}, \dots, k_{nm_n} \rangle$ и на ее основе можно ввести безусловный критерий предпочтения, в котором показатели $k_{n1}, \dots, k_{nm_n}$ полностью заменяют первичные (элементарные) показатели $k_1, \dots, k_m$. При этом критерий предпочтения в целом является смешанным — условно-безусловным, так как каждый приведенный показатель k_{nj} образуется путем введения соответствующей условной целевой функции вида (8.22).

Для того чтобы критерий предпочтения был смешанным, не обязательно, чтобы все составляющие вектора K_n были приведенными показателями — достаточно, чтобы приведенным был хотя бы один из показателей, учитываемых в составе вектора K_n . Например, смешанный вектор качества может иметь вид $K_n = \langle R, P, M \rangle$, где R — средний риск (приведенный показатель точности); P — приведенные затраты на систему (приведенный показатель стоимости); M — масса системы (являющаяся первичным, а не приведенным показателем).

Для сокращения числа учитываемых показателей качества применяется также приведение сравниваемых вариантов построения системы в сопоставимый вид. Очевидно, сравнение различных вариантов S' и S'' построения системы S было бы наиболее простым, если бы все составляющие вектора качества $K = \langle k_1, \dots, k_m \rangle$ системы были у этих вариантов одинаковыми, за исключением одного, которому присвоим номер один. Тогда лучшим будет тот вариант, которому соответствует минимальное значение

показателя k_1 . При этом, очевидно, задача векторного синтеза будет сведена к скалярному синтезу, т. е. предельно упрощена. Такие варианты, у которых значения всех показателей, кроме одного, совпадают, называются *полностью сопоставимыми*. Например, если вектор качества имеет вид

$$K = \langle P_{\text{ош}}, C, P \rangle, \quad (8.23)$$

где $P_{\text{ош}}$ — полная вероятность ошибки; C — пропускная способность; P — приведенные затраты, и сравниваемые варианты различаются лишь значением приведенных затрат P , то такие варианты полностью сопоставимы. Если у всех сравниваемых вариантов совпадают значения не $(m-1)$ показателей, а меньшего (но не меньше одного) числа показателей, то такие варианты называются *частично сопоставимыми*. Например, если вектор качества имеет вид (8.23) и у всех вариантов совпадают значения лишь вероятности ошибки $P_{\text{ош}}$, то такие варианты лишь частично сопоставимы. Очевидно, что при сравнении систем те показатели качества k_i , которые у всех вариантов системы совпадают, могут не включаться в состав вектора K ¹. Поэтому приведение систем в сопоставимый вид позволяет сократить число учитываемых показателей качества даже в случае частичной сопоставимости. Методы приведения систем к сопоставимому виду подробно рассмотрены в [61]. Рассмотрим лишь несколько иллюстрирующих примеров.

Пример 1. Сравниваются по показателю (8.23) системы передачи дискретной информации с АМ и ЧМ при действии внутреннего шума с интенсивностью N_0 . Их удобно сравнивать при одинаковых значениях показателей $P_{\text{ош}}$ и C (и одинаковых значениях интенсивности N_0 внутреннего шума). Но при одинаковых значениях $P_{\text{ош}}$ и C в системе АМ требуется большее значение мощности P_0 входного сигнала, что приводит к соответствующему увеличению затрат P ; с другой стороны, система АМ проще. Это, при прочих равных условиях, позволяет сократить затраты P . Поэтому, сравнивая результирующие затраты в вариантах АМ и ЧМ, можно отдать предпочтение тому варианту, в котором эти затраты окажутся меньшими. Следовательно, в данном примере сравнение вариантов по затратам P при одинаковых значениях $P_{\text{ош}}$ и C обеспечивает их полную сопоставимость [в рамках вектора качества (8.23)].

Пример 2. Те же, что и в примере 1, варианты ЧМ и АМ сравниваются по большему числу показателей, а именно вектор K имеет вид

$$K = \langle P_{\text{ош}}, C, \Delta f, P_0, P \rangle, \quad (8.24)$$

где Δf — требуемая полоса пропускания, а P_0 — требуемая мощность сигнала (передатчика).

Полоса Δf учитывает требование ЭМС (чем меньше Δf , тем лучше ЭМС), а P_0 — комплексный показатель — его уменьшение улучшает (при прочих равных условиях) целый ряд свойств системы (затраты P , массу M , ЭМС и ЭКС).

Сравнение вариантов АМ и ЧМ можно производить по-прежнему при одинаковых значениях $P_{\text{ош}}$ и C . Но при этом у них в общем случае будут различные значения показателей Δf , P_0 и P . Изъять показатель P_0 , переведя его в

¹ Они включаются лишь в состав ограничений.

соответствующее изменение затрат P , в данном примере нельзя, так как здесь величина P_0 характеризует не только затраты, но и ЭМС, и ЭКС. Можно, правда, попытаться изъять P_0 как отдельный показатель качества, пересчитав увеличение мощности P_0 в дополнительные затраты, которые потребуются для обеспечения ЭМС и ЭКС при большей мощности. Однако такой пересчет не всегда допустим. Например, если мощность P_0 превысит международную норму на ЭМС, то это нарушение, как правило, нельзя перевести на деньги. Увеличение мощности может привести также к увеличению массы M , которое в случае бортового передатчика часто совершенно недопустимо и поэтому не может быть переведено в рост затрат P . Часто нельзя свести к увеличению затрат и увеличение полосы Δf (в случае ЧМ). Таким образом, в этом случае, как правило, можно обеспечить лишь частичную сопоставимость. Но и такая сопоставимость весьма полезна, так как позволяет сравнивать варианты лишь по трем показателям вместо пяти.

Из рассмотренных примеров следует, что приведение систем в сопоставимый вид позволяет сократить число учитываемых показателей. Однако, как правило (особенно при большом числе первичных показателей), удастся обеспечить лишь частичную сопоставимость — число показателей, которые необходимо учитывать, остается большим, и, следовательно, для обеспечения сравнения требуется применить безусловный, условный или смешанный критерий предпочтения.

Выясним теперь, в какой мере решение задачи сравнения систем, производимое применением БКП или УКП, может считаться однозначным (единственным). Как уже отмечалось, каждой конкретной комбинации показателей $\langle k_1, \dots, k_m \rangle$, т. е. каждому значению вектора K , соответствует не одна система, а некоторый класс их. Поэтому и БКП, и УКП не позволяют сравнить системы «внутри» данного класса: для осуществления такого сравнения, т. е. для сужения класса систем до единственной совершенно определенной системы, потребовалось бы ввести, строго говоря, бесконечное число показателей качества. Однако практически, если число m показателей достаточно велико, можно считать класс систем, обладающих данным значением вектора K , настолько узким, что его можно рассматривать как некоторую достаточно определенную систему. Поэтому в дальнейшем будем полагать решение задачи сравнения систем единственным, если его удалось довести до выбора вполне определенного значения вектора K (и, следовательно, до выбора вполне определенного достаточно узкого класса систем).

При применении БКП, как будет показано в § 9.1, в общем случае сравнение систем (векторов K) удастся довести до выбора не одного значения вектора K (одной точки в пространстве R_m), а множества $M_{\text{нх}}$ нехудших векторов (точек, систем). Лишь в том частном случае, когда множество $M_{\text{нх}}$ оказывается вырожденным, решение доводится до выбора единственного значения вектора K , которое при этом оказывается безусловно лучшим, т. е. безусловно оптимальным (при определенных исходных данных D'').

В этом частном случае БКП является безусловным критерием оптимальности (критерием безусловной оптимальности). В остальных же случаях (т. е. при невырожденном множестве $M_{\text{нх}}$) БКП является лишь критерием предпочтения, а не оптимальности.

При использовании УКП системы сравниваются по результирующему показателю качества $K_{\text{рез}}$, а исходные (частные) показатели $k_1, \dots, k_m$ при этом рассматриваются лишь как некоторые варьируемые параметры, влияющие на $K_{\text{рез}}$ в соответствии с видом выбранной результирующей целевой функции (8.20). При этом, как уже отмечалось, решение задачи сравнения систем всегда приводит к некоторому единственному значению $K_{\text{рез}}$ — минимальному (по всему множеству $M_{\text{сх}}$) $K_{\text{рез. мин}}$, которое считается при этом наилучшим возможным, т. е. оптимальным.

Как следует из (8.20), величина $K_{\text{рез}}$ является функцией вектора качества K . Поэтому минимум значения $K_{\text{рез}}$ (по всем строго допустимым значениям вектора K и соответствующим классам систем) может иметь место либо при одном вполне определенном значении вектора K , либо при двух или более значениях этого вектора. Будем говорить, что решение задачи сравнения систем в первом случае однозначно в смысле значения вектора K , а во втором случае — неоднозначно в смысле вектора K . В том, что решение задачи может быть неоднозначно, нетрудно убедиться на следующем простом примере. Пусть сравниваются три варианта построения системы, обладающие следующими значениями показателей качества: для S_1 $k_1=1, k_2=6$; для S_2 $k_1=1,25, k_2=5$; для S_3 $k_1=5, k_2=1$.

На основе БКП ни одному из вариантов нельзя отдать предпочтение — они все являются нехудшими. Пусть введен результирующий показатель качества вида $K_{\text{рез}} = 0,8k_1 + 0,2k_2$. Такой УКП дает следующие результаты: для S_1 $K_{\text{рез}} = K_{\text{рез. 1}} = 2$, для S_2 $K_{\text{рез}} = K_{\text{рез. 2}} = 2$; для S_3 $K_{\text{рез}} = K_{\text{рез. 3}} = 4,2$.

Для первых двух вариантов значение $K_{\text{рез}}$ оказалось одинаковым ($K_{\text{рез. 1}} = K_{\text{рез. 2}} = 2$) и лучшим, чем для варианта S_3 . Следовательно, на основе данного УКП довести результат сравнения систем до выбора единственного (в смысле значения вектора K) варианта невозможно.

Таким образом, при применении как БКП, так и УКП результат сравнения систем может быть как однозначным, так и неоднозначным. Если результат сравнения оказывается однозначным, т. е. приводит к выбору единственной (наилучшей, оптимальной) системы, то данный критерий предпочтения превращается в критерий оптимальности.

Из всего изложенного следует, что при использовании БКП однозначное решение сравнительно редко. При применении же УКП, напротив, исключением является неоднозначное решение. Поэтому, как правило, УКП, а не БКП является критерием оптимальности. Однако при применении УКП эта оптимальность следст-

ПРИМЕНЕНИЕ БЕЗУСЛОВНОГО И УСЛОВНОГО КРИТЕРИЕВ ПРЕДПОЧТЕНИЯ

ние условности самого критерия предпочтения оказывается весьма условной.

В заключение рассмотрим, как поступать в тех случаях, когда на основе принятого критерия предпочтения (БКП или УКП) решение оказалось неоднозначным (в смысле значения вектора K). Если неоднозначность возникла при использовании БКП, то на последующих этапах проектирования ее можно устранить, применив соответствующий УКП. Если надлежащий УКП достаточно объективно обосновать не удастся, то остаются следующие возможности:

1. Сузить множество нехудших систем соответствующей корректировкой исходных данных D'' (например, введением дополнительных ограничений O_8), если целесообразность такой корректировки может быть обоснована.

2. Выбрать для дальнейших этапов проектирования любой из нехудших вариантов.

3. Отказаться от выбора единственного варианта и оставить для дальнейшего проектирования (или даже для изготовления) несколько нехудших вариантов.

Например, если окажется, что невозможно обоснованно предпочесть вариант системы со сложным сигналом варианту с простым сигналом, то в ряде случаев может оказаться целесообразным выпускать системы с различными (не перестраиваемыми в процессе функционирования) сигналами — каждая из них найдет впоследствии свою целесообразную область применения. Так, промышленность выпускает радиолы нескольких классов. Радиолы высшего класса обладают значительно лучшим качеством воспроизведения звука, но имеют большую массу, габаритные размеры и стоимость. Радиолы низшего класса, напротив, имеют более низкое (но во многих случаях приемлемое) качество воспроизведения сообщений, но зато значительно лучшие значения таких показателей, как стоимость, масса и габаритные размеры.

Если неоднозначность возникла при применении УКП, то возможны следующие решения:

1) проверить обоснованность данного УКП, т. е. возможность и целесообразность корректировки вида УКП. Некоторое изменение УКП, осуществляемое путем более глубокого его обоснования, может привести к устранению неоднозначности решения;

2) провести соответствующую корректировку исходных данных D'' ;

3) принять в качестве оптимальной любую из систем, обладающих данным значением $K_{рез} = K_{р.мин}$, несмотря на различия в значениях вектора K ;

4) отказаться от выбора единственной системы и принять решение о проектировании (или производстве) систем нескольких классов.

Литература: основная [58].

9.1. НЕХУДШИЕ И ХУДШИЕ СИСТЕМЫ. ДИАГРАММЫ ОБМЕНА

На основе БКП вводятся понятия «нехудшая и худшая системы (точки)». Система S , имеющая вектор качества K и принадлежащая множеству $M_{сд}$, считается нехудшей, если в этом множестве не существует ни одной безусловной лучшей системы, т. е. не существует системы S' , обладающей вектором качества K' , таким, что $K' \leq K$; в противном случае (т. е. если такая система S' существует) система S считается худшей. Соответственно нехудшим или худшим считается и значение вектора качества K системы S , т. е. соответствующая точка в пространстве R_m . Следовательно, каждая точка в R_m является либо нехудшей, либо худшей¹. Поэтому последовательное применение БКП ко всем точкам множества $M_{сд}$ разбивает это множество на два непересекающихся подмножества — $M_{нх}$ нехудших систем и M_x худших систем.

Пусть, например, множество $M_{сд}$ строго допустимых систем имеет такой же вид, как на рис. 8.5, и изображено вновь на рис. 9.1 в виде сектора, ограниченного линиями $a'b'c'd'$ и включающего эти границы. Рассмотрим некоторую точку A внутри этой области. Очевидно, точки A' и B_1 являются по сравнению с точкой A безусловно лучшими, так как каждой из них соответствует меньшее (лучшее) значение одного из показателей качества при таком же значении второго показателя, например точке A' соответствует меньшее (чем у точки A) значение показателя k_2 при таком же значении показателя k_1 . Следовательно, в соответствии с данны-

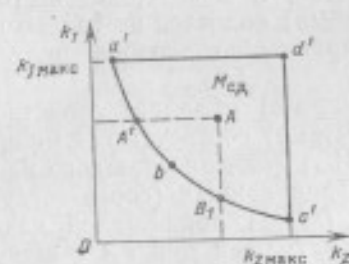


Рис. 9.1

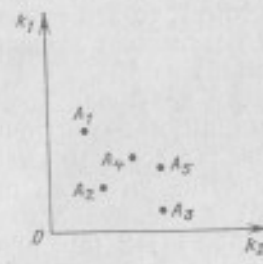


Рис. 9.2

¹ Нехудшие точки в литературе по технической кибернетике называют также эффективными точками или точками Парето. Однако в литературе по радиосистемам более распространен термин «нехудшие точки» [1, 46, 58 и др.].

ми определениями нехудших и худших точек точка A является худшей.

Нетрудно убедиться, что для любой точки, принадлежащей $M_{сд}$, но не находящейся на левой нижней границе $a'bc'$, на этой границе будут существовать две безусловно лучшие точки. Поэтому любая из точек множества $M_{сд}$, за исключением принадлежащих границе $a'bc'$, является худшей, как и соответствующая этой точке система. Любая же из точек границы $a'bc'$ является нехудшей, так как ни для одной из них не существует ни одной безусловно лучшей точки в пределах всего множества $M_{сд}$ (включая все его граничные точки).

Таким образом, в рассмотренном примере $M_{нх}$ — множество всех точек левой нижней границы $a'bc'$, а M_x — множество всех остальных точек области $a'bc'd$. Предоставляем читателю самому доказать, что в случае, когда множество $M_{сд}$ состоит из дискретных точек A_1, A_2, A_3, A_4 и A_5 (рис. 9.2), точки A_1, A_2, A_3 (и соответствующие им системы) являются нехудшими, а точки A_4 и A_5 (и соответствующие им системы) — худшими.

Рассмотрим основные свойства нехудших точек (систем). При этом будем полагать, что множество $M_{сд}$ замкнутое. В подавляющем большинстве практических задач это условие выполняется. В противном случае можно рассматривать вместо самого множества $M_{сд}$ его замыкание¹. Все свойства нехудших точек (систем) однозначно вытекают из определения нехудшей точки (системы) и сводятся к следующему [58]:

1. Всякое непустое замкнутое множество $M_{сд}$ содержит по меньшей мере одну нехудшую точку (систему), т. е. множество $M_{нх}$ непусто.
2. Для каждой точки множества M_x можно найти в множестве $M_{нх}$ по меньшей мере одну безусловно лучшую точку.
3. Ни одна из точек множества $M_{нх}$ не может быть признана безусловно лучшей или худшей по сравнению с какой-либо другой точкой этого множества.
4. Если множество $M_{нх}$ вырождено, т. е. содержит всего одну точку (систему), то эта точка (система) является не только нехудшей, но и безусловно лучшей по сравнению с любой другой точкой (системой) из $M_{сд}$.
5. Каждая нехудшая точка (система) обладает свойством m -кратного минимума: этой точке (системе) соответствует минимальное (в пределах всего множества $M_{сд}$) значение каждого из показателей качества $k_1, \dots, k_m$ при фиксированных (соответствующих этой точке) значениях остальных $(m-1)$ показателей качества. Например, при $m=2$ (см. рис. 8.2) нехудшей точке A с координатами $k_1=P_{пр1}, k_2=P_{пр1}$ соответствует минимальное значение показателя k_1 (т. е. вероятности $P_{пр}$ пропуска сигнала) при фиксиро-

ванном (равном $k_2=P_{пр1}$) значении показателя k_2 (вероятности ложной тревоги). Этой же точке A соответствует минимальное значение показателя k_2 (вероятности $P_{пр}$) при фиксированном (равном $P_{пр1}$) значении k_1 (вероятности $P_{пр}$).

6. В пространстве R_m геометрическим местом всех нехудших точек является некоторая многомерная поверхность, связывающая между собой нехудшие значения показателей качества $k_1, \dots, k_m$. Она называется иногда *оптимальной поверхностью* [58] и может быть описана любым из следующих m соотношений:

$$\begin{aligned} k_1 &= f'_{нх}(k_2, k_3, \dots, k_m), \quad k_2 = f''_{нх}(k_1, k_3, \dots, k_m), \dots, \\ k_m &= f^{(m)}_{нх}(k_1, k_2, \dots, k_{m-1}). \end{aligned} \quad (9.1)$$

Например, если множеством нехудших точек является совокупность точек левой нижней границы $a'bc'$ множества $M_{сд}$ (рис. 9.1), то многомерная оптимальная поверхность имеет вид кривой $a'bc'$, которая может быть описана соотношением $k_1 = f'_{нх}(k_2)$, чему соответствует зависимость, приведенная на рис. 9.3,а, или в эквивалентном виде $k_2 = f''_{нх}(k_1)$, которому соответствует рис. 9.3,б. Очевидно, функция $f''_{нх}(k_1)$ является обратной по отношению к $f'_{нх}(k_2)$, а рис. 9.3,б может быть получен из рис. 9.3,а соответствующим поворотом осей координат.

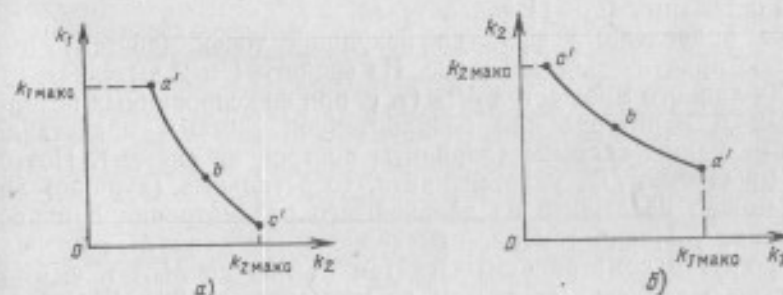


Рис. 9.3

Каждая i -я зависимость из (9.1) может быть в принципе получена из любой другой из них, если эту другую зависимость рассматривать как уравнение относительно показателя качества k_i ; например, если первую зависимость (9.1) рассматривать как уравнение относительно k_m и решать его при произвольных значениях $k_1, \dots, k_{m-1}$, то получится самая последняя (m) -я зависимость. Можно доказать (см., например, [58]), что любая из функций $f'_{нх}, \dots, f^{(m)}_{нх}$ является монотонно убывающей по каждому из $(m-1)$ ее аргументов [при фиксированных значениях остальных $(m-2)$ аргументов]. Поскольку все m соотношений (9.1) для рассматриваемой многомерной оптимальной поверхности эквиваленты, в дальнейшем будем использовать запись

$$k_1 = f_{нх}(k_2, \dots, k_m) \quad (9.2)$$

¹ Строгое определение понятий «замкнутое множество» и «замыкание множества», а также возможность замены исходного незамкнутого множества его замыканием даны, например, в [58].

и называть ее *m*-мерной диаграммой обмена между показателями качества $k_1, \dots, k_m$ системы. Действительно, в силу отмеченного ранее свойства *m*-кратного минимума соотношение (9.2) показывает, как зависит минимально достижимое (наилучшее возможное, потенциальное) значение каждого из *m* показателей качества от значений всех остальных (*m*—1) показателей. Например, кривая $k_1 = f_{\text{нх}}(k_2)$ на рис. 9.3,а показывает, как минимально возможное (в пределах $M_{\text{сд}}$) значение показателя k_1 зависит от значения показателя k_2 ; и эта же кривая показывает, как k_1 влияет на минимально достижимое значение показателя k_2 . Следовательно, в этом примере уравнение $k_1 = f_{\text{нх}}(k_2)$ соответствует двумерной диаграмме обмена между показателями k_1 и k_2 .

С учетом изложенного шестое свойство нехудших систем можно сформулировать следующим образом. В пространстве R_m множество нехудших точек имеет вид многомерной оптимальной поверхности, являющейся *m*-мерной диаграммой обмена между показателями качества $k_1, \dots, k_m$ систем. Эта диаграмма обмена строго монотонна (каждый из показателей k_i монотонно убывает при возрастании любого другого из (*m*—1) показателей [если при этом остальные (*m*—2) показателя остаются фиксированными]). Она показывает, как потенциальное (наилучшее достижимое в пределах $M_{\text{сд}}$) значение каждого показателя k_i ($i = 1, m$) зависит от значений остальных (*m*—1) показателей.

Все приведенные свойства нехудших точек (систем) имеют большое практическое значение. Из свойств 1 и 2 вытекает, что в рамках данного множества $M_{\text{сд}}$ (т. е. при фиксированных исходных данных D'') интерес для разработчиков системы представляют лишь нехудшие системы (варианты построения систем). Поэтому, если множество $M_{\text{нх}}$ удастся найти, то остальные (худшие) варианты можно исключить из дальнейшего рассмотрения при проектировании¹. Иначе говоря, отыскание множества $M_{\text{нх}}$ позволяет сузить круг рассматриваемых систем от $M_{\text{сд}}$ до $M_{\text{нх}}$, т. е. от любых строго допустимых систем до нехудших систем. В ряде случаев это сужение области поиска лучших (оптимальных) систем может быть весьма существенным (например, в случае, изображенном на рис. 9.1, вся область $a'b'c'd'$ сокращается до ее левой нижней границы $a'bc'$).

Из свойства 4 следует, что когда множество $M_{\text{нх}}$ оказывается вырожденным, т. е. содержит в R_m всего одну точку, эта точка (и соответствующая ей система) оказывается не только нехудшей, но и безусловно лучшей по сравнению с любыми другими возможными в рамках $M_{\text{сд}}$ (т. е. при фиксированных исходных данных D'') точками (системами). Найденную в этом случае систе-

¹ Как отмечалось в § 8.2, при изменении исходных данных, в частности при учете дополнительных показателей качества, системы, худшие в смысле *m* показателей, могут оказаться нехудшими в смысле *m'* показателей (где $m' > m$). Поэтому пренебрежение худшими вариантами допустимо лишь в пределах фиксированных исходных данных.

му будем называть *совместно-оптимальной*, так как ей соответствуют одновременно (совместно) наилучшие значения всех *m* показателей качества. Будем также говорить, что, когда исходные данные приводят к вырожденному $M_{\text{нх}}$, существует *совместно-оптимальное* решение задачи векторной оптимизации.

Пример вырожденного множества $M_{\text{нх}}$ приведен на рис. 9.4. Оно содержит единственную точку — точку А. (Точка А будет оставаться единственно нехудшей точкой и при деформациях множества $M_{\text{сд}}$, если это множество не будет выходить за пределы квадранта BAF или лишаться точки А.) При этом диаграмма обмена между показателями качества $k_1, \dots, k_m$ состоит всего из одной точки. Это означает, что обмен между показателями качества отсутствует — показатель k_1 (например, k_2), соответствующий этой точке, невозможно уменьшить, увеличивая (ухудшая) какой-либо другой (другие) из (*m*—1) показателей.

Иначе говоря, в тех случаях, когда исходным данным D'' соответствует вырожденное множество $M_{\text{нх}}$, минимальное значение любого показателя качества $k_1, \dots, k_m$ невозможно улучшить ослаблением требований к каким-либо другим из этих *m* показателей (вплоть до полного их игнорирования). И наоборот, если множество $M_{\text{нх}}$ невырождено (см., например, рис. 9.1), то предел уменьшению каждого из показателей $k_1, \dots, k_m$ ставит увеличение (ухудшение) других показателей. Поэтому в случае вырожденного $M_{\text{нх}}$ показатели качества $k_1, \dots, k_m$ будем называть *согласованными* (совместно-улучшаемыми), а в случае невырожденного $M_{\text{нх}}$ — *несогласованными* или *антагонистическими* (ухудшение каждого из показателей препятствует улучшению других показателей). Например, на рис. 9.4, минимальное (в пределах $M_{\text{сд}}$) значение показателя k_1 равно k_{1A} независимо от того, каково допустимое значение показателя k_2 , а минимальное значение показателя k_2 равно k_{2A} независимо от допустимого значения показателя k_1 . В случае же, изображенном на рис. 9.1, минимальное значение k_1 тем меньше, чем больше (хуже) значение показателя k_2 и наоборот.

Случай согласованных показателей качества (т. е. вырожденного $M_{\text{нх}}$) является редким и обычно может иметь место лишь в

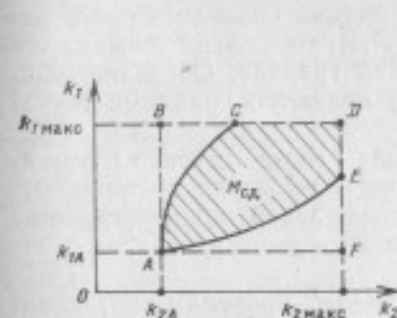


Рис. 9.4

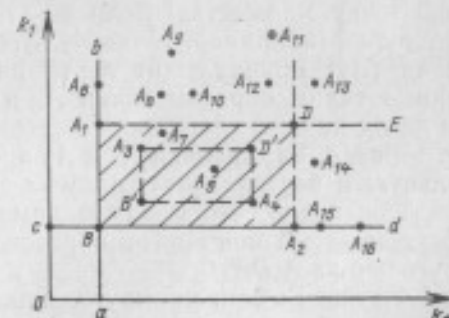


Рис. 9.5

задачах с однородными показателями $k_1, \dots, k_m$ (т. е. различными показателями, характеризующими одно и то же свойство системы, например точность ее действия). Важный пример такого частного случая рассмотрен в § 11.3. В большинстве же практических задач множество $M_{\text{нх}}$ оказывается невырожденным, т. е. содержит не одну, а несколько или даже континуум точек. При этом, как следует из свойства 3, ни одной из нехудших систем нельзя отдать предпочтение, оставаясь в рамках БКП: для дальнейшего сужения круга рассматриваемых систем, в частности для выбора единственной (оптимальной) системы, требуется применить какой-либо условный критерий предпочтения. Однако и в таких случаях множество $M_{\text{нх}}$, как будет показано в гл. 10 и 11, весьма полезно найти, особенно на стадии научных исследований и на начальных этапах проектирования.

9.2. МЕТОДЫ ОТЫСКАНИЯ НЕХУДШИХ СИСТЕМ

Рассмотрим сначала случай дискретного выбора систем (или вариантов их построения). При этом множество $M_{\text{сд}}$ строго допустимых точек в R_m конечно и отыскать среди них множество $M_{\text{нх}}$ нехудших точек нетрудно. Рассмотрим в качестве иллюстрации две характерные ситуации.

1. Число N точек, содержащихся в $M_{\text{сд}}$, велико, но учитываются всего два показателя качества (этот случай наиболее характерен для научных исследований и начального этапа проектирования). Тогда удобно изобразить все точки на плоскости показателей качества, так как это сделано, например, на рис. 9.5. Чтобы выяснить, какие из строго допустимых точек $A_1, \dots, A_{16}$ являются нехудшими, можно применить описанный в [58] метод прямоугольников. Он состоит из следующих этапов.

1-й этап. Проводим вертикальную прямую ab через самую левую точку множества. Если на этой прямой оказалось более одной точки (например, точки A_1 и A_6), то из этих точек самая нижняя точка (A_1) будет принадлежать искомой левой нижней границе множества (области) строго допустимых точек.

2-й этап. Проводим горизонтальную прямую через самую нижнюю точку множества. Если на этой прямой оказалось более одной точки (например, точки A_2 , A_{13} и A_{16}), то из этих точек самая левая (A_2) принадлежит левой нижней границе. Очевидно, найденные таким образом точки A_1 и A_2 являются крайними точками левой нижней границы.

3-й этап. Через точки A_1 и A_2 проводим соответственно горизонтальную и вертикальную прямые до их пересечения (в точке D). Тогда очевидно, что все остальные точки левой нижней границы могут лежать лишь внутри образовавшегося заштрихованного прямоугольника A_1DA_2B .

4-й этап. Внутри прямоугольника A_1DA_2B проводим вертикальную прямую через самую левую точку (точки) и горизонтальную прямую через самую нижнюю точку (точки). Тогда самая нижняя

точка на вертикальной прямой (точка A_3) и самая левая точка на горизонтальной прямой (точка A_4) будут также точками левой нижней границы.

5-й этап. Проведем через точку A_3 горизонтальную прямую, а через точку A_4 вертикальную прямую. Тогда все остальные точки левой нижней границы могут быть расположены лишь внутри прямоугольника $A_3D'A_4B'$. Далее процедура продолжается в том же порядке.

В рассматриваемом примере (рис. 9.5) в ее продолжении нет необходимости, так как внутри прямоугольника $A_3D'A_4B'$ оказалась всего одна точка (A_5), которая и является последней искомой точкой левой нижней границы. Таким образом, в рассматриваемом случае искомое множество $M_{\text{нх}}$ состоит из точек A_1, A_2, A_3, A_4 и A_5 .

2. Число m учитываемых показателей качества сравнительно велико (например, $m > 5$), а число N сравниваемых вариантов невелико (например, $N \leq 5$). Этот случай характерен для этапов проектирования, близких к заключительному. При этом множество $M_{\text{нх}}$ можно искать методом, состоящим из следующих этапов.

1-й этап. Присваиваем всем сравниваемым вариантам номера по следующему принципу: меньший номер — тому варианту, которому соответствует меньшее значение показателя k_1 (поэтому, в частности, номер 1 присваивается варианту с наименьшим значением показателя k_1). Если при этом окажется, что два или более варианта имеют одинаковое значение показателя k_1 , то меньший номер присваивается тому из них, которому соответствует меньшее значение показателя k_2 . Если окажется, что у некоторых вариантов одинаковы значения не только k_1 , но и k_2 , то меньший номер присваивается тому из них, которому соответствует меньшее значение показателя k_3 , и т. д. Наконец, если окажется, что два или более варианта имеют одинаковые значения всех m показателей качества (что, очевидно, может иметь место на практике лишь весьма редко), то всем этим вариантам присваивается один и тот же номер и они рассматриваются как единая система, точнее — класс систем. Сравнение вариантов внутри этого класса производится на последующих этапах проектирования путем более точного определения уже учтенных показателей качества (в результате этого уточнения показатели, считавшиеся одинаковыми, могут оказаться различными) и (или) путем учета дополнительных показателей качества.

2-й этап. В соответствии с присвоенными номерами составляем для всех (строго допустимых) вариантов таблицу значений их показателей качества; например, табл. 9.1, позволяет сравнивать пять вариантов по шести показателям качества.

3-й этап. Рассматривая таблицу, выявляем нехудшие варианты по принципу: если данный вариант имеет меньшее, чем все остальные варианты, значение хотя бы одного из показателей качества, то он нехудший. По этому принципу находим, что наименьшее значение показателя k_2 имеет вариант 5, показателя k_3 — ва-

Таблица 9.1

Вариант	Показатель					
	k_1	k_2	k_3	k_4	k_5	k_6
1	1	3	2	5	1	2
2	1	4	7	1	5	2
3	3	2	1	4	2	2
4	4	2	8	7	10	6
5	6	1	3	2	1	9

риант 3, показателя k_4 — вариант 2. Следовательно, варианты 2, 3 и 5 — нехудшие. Вариант 1 — заведомо нехудший, так как ему был присвоен минимальный номер. Действительно, значение показателя k_1 минимально у двух вариантов — 1 и 2, но у варианта 1 меньшее значение показателя k_2 . Следовательно, вариант 2 не может быть признан безусловно лучшим, чем 1. Ни один из остальных вариантов (3, 4 и 5) также не может быть безусловно лучшим варианта 1, так как у этих вариантов больше (хуже) значения показателя k_1 . Следовательно, в рассматриваемом множестве вариантов (точек в R_m) нет ни одного безусловно лучшего, чем вариант 1, и вариант 1 — нехудший.

Таким образом установлено, что варианты 1, 2, 3 и 5 — нехудшие. Наименьшее значение показателя k_3 имеют варианты 1 и 5, но это обстоятельство уже не дает дополнительной информации, так как было установлено, что эти варианты нехудшие. Также не дает дополнительной информации то, что показатель k_6 минимален у варианта 1. Следовательно, к концу данного этапа осталось выяснить, нехудшим или худшим является вариант 4.

4-й этап. Выясним, к какому множеству (M_{nx} или M_x) следует отнести оставшиеся неклассифицированными варианты. Поскольку показатели качества больше, чем сравниваемых вариантов, после 3-го этапа останется неклассифицированным, как правило, весьма небольшое число вариантов. В данном примере таковым остался лишь вариант 4. Поэтому для его классификации (худший он или нехудший) проще попытаться найти среди остальных четырех вариантов хотя бы один безусловно лучший, чем вариант 4. Очевидно, при этом поиск следует производить лишь среди вариантов с меньшими номерами, чем данный, так как более высокие номера имеют большее (худшее) значение показателя k_1 . Таким образом, в данном примере следует выяснить, не является ли какой-либо из вариантов 1, 2 и 3 безусловно лучшим, чем вариант 4. Вариант 1 не является безусловно лучшим, так как имеет большее значение показателя k_2 . Вариант 2 имеет большее, чем у варианта 4, значение показателя k_2 и, следовательно, также не может быть признан безусловно лучшим. У варианта 3 все показатели не больше, чем у варианта 4, в том числе показатель k_1 меньше (у него меньше и показатели k_3 , k_4 , k_5 и k_6 , но это уже не дает дополнительной информации для классификации варианта

4). Следовательно, вариант 3 безусловно лучше, чем 4, и поэтому вариант 4 — худший.

В результате проведения всех этапов выяснено, что вариант 4 является худшим, а остальные — нехудшими, т. е. множество M_{nx} состоит из вариантов (точек) 1, 2, 3 и 5.

Рассмотрим теперь методы отыскания множества M_{nx} в задачах синтеза структуры и (или) оптимизации параметров системы. В таких задачах множество M_{cx} содержит не конечное число, а континуум точек и процесс отыскания множества M_{nx} оказывается значительно более сложным. Существуют различные методы отыскания этого множества. Рассмотрим кратко лишь три из них, описанные подробно в [58]: метод рабочих характеристик, весовой метод и комбинированный метод.

Метод рабочих характеристик состоит в следующем. Все показатели качества, кроме одного, которому присваивается первый номер, переводятся в разряд ограничений типа равенств и ищется минимум (по всем строго допустимым системам) показателя k_1 , т. е. решается задача:

$$\text{обеспечить } \min_{S \in M_{cx}} k_1 \text{ при } k_2 = \underline{k}_2, k_3 = \underline{k}_3, \dots, k_m = \underline{k}_m. \quad (9.3)$$

Здесь и далее обозначение $S \in M_{cx}$ означает, что рассматриваются (варьируются) лишь те варианты построения системы, которые удовлетворяют исходным данным D'' , в частности ограничениям O_s структуры и параметров системы и ограничениям O_k показателей качества $k_1, \dots, k_m$. Например, если структура системы задана полностью, за исключением параметров $x_1, \dots, x_j, \dots, x_n$, которые должны находиться в пределах $x_{j \min} \leq x_j \leq x_{j \max}$ ($j = \overline{1, n}$), а ограничения O_k заданы в виде (8.12), то задача (9.3) в развернутом виде имеет следующий вид: при заданных условиях $\mathcal{U}$ функционирования системы и заданной ее структуре

$$\left. \begin{aligned} &\text{обеспечить } \min_{(x_1, \dots, x_j, \dots, x_n)} k_1 \\ &\text{при } x_{j \min} \leq x_j \leq x_{j \max} \quad (j = \overline{1, n}), \\ &k_i \leq k_{i \max} \quad (i = \overline{1, m}), \quad k_2 = \underline{k}_2, \dots, k_m = \underline{k}_m. \end{aligned} \right\} \quad (9.4)$$

Здесь $k_i = f_i(x_1, \dots, x_j, \dots, x_n)$ ($j = \overline{1, n}$) — некоторая определенная функция варьируемых параметров.

В (9.3) и (9.4) и далее черта снизу под показателями качества k_i ($i = \overline{2, m}$) означает, что в процессе отыскания минимума k_1 эти показатели должны поддерживаться постоянными (неизменными, фиксированными, «замороженными»).

Пусть результатом решения задачи (9.3) является некоторая система (класс систем) S_p , а соответствующее ей минимальное значение показателя k_1 равно $k_{1 \min}$. Очевидно, результаты решения задачи (9.3), т. е. вид системы S_p и значение величины $k_{1 \min}$, будут в общем случае зависеть от фиксированных значений $k_2, \dots$

..., k_m показателей качества, переведенных в разряд ограничений типа равенств, т. е.

$$S_p = S_p(k_2, \dots, k_m), \quad (9.5)$$

$$k_{1 \min} = f_p(k_2, \dots, k_m), \quad (9.6)$$

где $f_p(k_2, \dots, k_m)$ — некоторая функция показателей $k_2, \dots, k_m$, которую будем называть *рабочей поверхностью*.

Зависимости (9.5) и (9.6) можно найти, варьируя значения показателей $k_2, \dots, k_m$ в допустимых для них ограничениях (8.12) пределах. Кроме того, в соответствии с (9.3) необходимо учитывать ограничение и показателя k_1 ($k_1 \leq k_{1 \max}$), т. е. исключать из рассмотрения те комбинации параметров $k_2, \dots, k_m$, при которых значение $k_{1 \min}$ выходит за пределы ограничения k_1 . Будем называть все такие возможные комбинации ($k_2, \dots, k_m$) показателей, переведенных в разряд ограничений типа равенств, *допустимыми комбинациями* показателей $k_2, \dots, k_m$. Тогда всему множеству допустимых комбинаций будут соответствовать некоторое множество M_p систем S_p и некоторая рабочая поверхность (9.6) в пространстве R_m . Каждой точке этой поверхности соответствует некоторая система S_p , и наоборот.

В [58] доказано, что необходимым и достаточным условием совпадения множества M_p с искомым $M_{\text{нх}}$ и соответственно совпадения рабочей поверхности (9.6) с оптимальной (диаграммой обмена) (9.2) является монотонно убывающий характер функции (9.6) по каждому из ее аргументов [при любых фиксированных комбинациях значений остальных ($m-2$) аргументов]. То, что это условие является необходимым, очевидно, ибо ранее отмечалось, что оптимальная поверхность (диаграмма обмена) имеет именно такой монотонно убывающий характер. Достаточность же указанного условия требует доказательства, которое и приведено в [58]. Там же доказано, что если рабочая поверхность (9.6) не обладает указанным свойством строгой монотонности, то она также содержит все точки множества $M_{\text{нх}}$, но, кроме того, в ней имеются и худшие точки. Следовательно, можно утверждать, что найденное

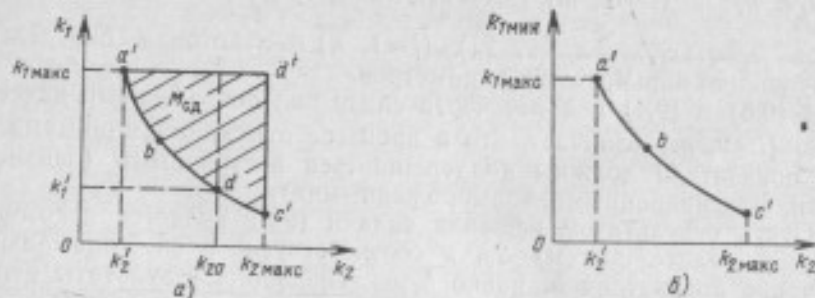


Рис. 9.6

указанным способом множество M_p систем (или точек в R_m) всегда включает множество $M_{\text{нх}}$, а при строгой монотонности рабочей поверхности M_p совпадает с $M_{\text{нх}}$. Рассмотрим в качестве иллюстрации этих положений следующие два примера.

Пример 1. Множество $M_{\text{ед}}$ в R_m имеет вид, изображенный на рис. 9.6, а, т. е. такой же, как на рис. 9.1, причем, как уже было показано, множество $M_{\text{нх}}$ есть совокупность точек левой нижней границы $a'bc'$. Все остальные точки множества $M_{\text{ед}}$ (т. е. области $a'bc'd'$) — худшие. Выясним, какой результат даст в этом случае применение указанного выше метода рабочих характеристик.

В соответствии с этим методом должна решаться задача:

$$\begin{aligned} &\text{обеспечить } \min k_1 \\ &\quad S = M_{\text{ед}} \\ &\text{при } k_2 = k_2 \end{aligned} \quad (9.7)$$

для всех допустимых значений показателя k_2 .

Из рис. 9.6, а следует, что эти допустимые значения таковы:

$$k_2' \leq k_2 \leq k_{2 \max}. \quad (9.8)$$

Действительно, из ограничения $k_2 \leq k_{2 \max}$ следует, что должно быть и $k_2 \leq k_{2 \max}$; из ограничения же $k_1 \leq k_{1 \max}$ и рис. 9.6, а следует, что должно быть $k_2 \geq k_2'$.

Пусть, например, в этой области выбрано $k_2 = k_{20}$. Из рис. 9.6, а следует, что при таком значении показателя k_2 минимально возможное значение показателя k_1 равно k_1' и ему соответствует точка d на левой нижней границе области $M_{\text{ед}}$. Поэтому в результате решения задачи (9.7) при $k_2 = k_{20}$ будет найдена точка d искомого множества $M_{\text{нх}}$ (и соответствующая этой точке система S_p). При другом допустимом [неравенством (9.8)] значении показателя k_2 получится другая точка левой нижней границы, т. е. множества $M_{\text{нх}}$. Поэтому, решая задачу (9.7) последовательно для всех допустимых [неравенством (9.8)] значений показателя k_2 , находим тем самым все точки левой нижней границы $a'bc'd'$, т. е. все множество $M_{\text{нх}}$. Следовательно, в данном примере M_p совпадает с $M_{\text{нх}}$. Зависимость $k_{1 \min} = f_p(k_2)$, т. е. рабочая поверхность в данном примере имеет вид, изображенный на рис. 9.6, б, и полностью совпадает с оптимальной поверхностью (диаграммой обмена). Как видно из рис. 9.6, б, рабочая поверхность оказалась монотонно убывающей, что и должно быть при совпадении M_p с $M_{\text{нх}}$.

Пример 2. В этом случае (рис. 9.7) граница abc множества $M_{\text{ед}}$ не является монотонно убывающей. Поэтому множество $M_{\text{нх}}$ содержит лишь точки, расположенные на участке ab кривой (любая из точек участка bc , не совпадающая с точкой b , является худшей, так как ей соответствует большее значение показателя k_2 при том же значении показателя k_1).

Предоставляем читателю самому доказать (тем же способом, что и в примере 1), что рабочая поверхность в этом случае имеет вид, изображенный на рис. 9.8, т. е. совпадает с границей abc множества $M_{\text{ед}}$ и, следовательно, содержит как все точки множества $M_{\text{нх}}$ (участок ab), так и ряд худших точек (на отрезке bc). Поэтому в данном случае множество M_p оказывается шире, чем $M_{\text{нх}}$, и после нахождения множества M_p требуется затем исключить из него худшие точки.

Таким образом, метод рабочих характеристик сводится к следующему этапу.

1. Все показатели, кроме одного (k_1), переводятся в разряд ограничений типа равенств, т. е. формулируется задача (9.3).

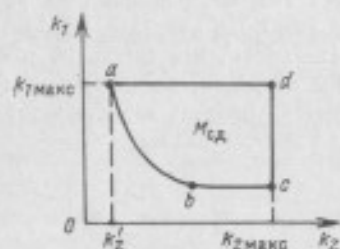


Рис. 9.7

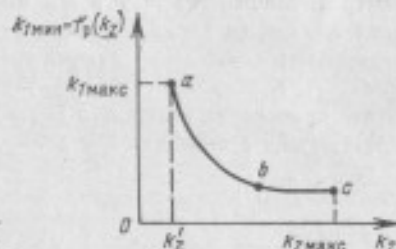


Рис. 9.8

2. Задача (9.3) решается последовательно для всех допустимых [ограничениями (8.12)] комбинаций ($k_2, \dots, k_m$) показателей качества, переведенных в разряд ограничений, в результате чего находится рабочая поверхность (9.6) (и соответствующее ей множество M_r рабочих систем S_r). Если оказывается возможным аналитическое решение этой задачи, то нет необходимости задавать численные значения показателей в комбинациях ($k_2, \dots, k_m$) — достаточно решить задачу в общем (аналитическом) виде для произвольной, но удовлетворяющей ограничениям (8.12) фиксированной комбинации ($k_2, \dots, k_m$). Если же аналитического решения найти не удастся, то зависимость (9.6) можно определить лишь приближенно — не для всего континуума допустимых комбинаций, а лишь для выбранных с достаточно малым интервалом допустимых численных значений таких комбинаций.

3. Если зависимость (9.6) оказывается монотонно убывающей по каждому из ее аргументов, то она полностью совпадает с искомой диаграммой обмена — геометрическим местом всех точек множества M_{ex} . В противном случае из найденного множества M_r точек необходимо исключить худшие. Однако во многих практических задачах множество M_r совпадает с M_{ex} и такого исключения не требуется.

Рассмотрим теперь весовой метод отыскания M_{ex} . При применении этого метода минимизируется взвешенная сумма

$$k'_a = \lambda_1 k_1 + \dots + \lambda_i k_i + \dots + \lambda_m k_m \quad (9.9a)$$

показателей качества, в которой веса $\lambda_1, \dots, \lambda_m$ выбираются произвольными в пределах

$$\lambda_i > 0, \quad \sum_{i=1}^m \lambda_i = 1, \quad (9.9b)$$

т. е. решается задача:

$$\text{обеспечить } \min_{S \in M_{cd}} k'_a \quad (9.9b)$$

при некоторой фиксированной комбинации ($\lambda_1, \dots, \lambda_m$) весовых коэффициентов.

Нетрудно убедиться, что тот же результат получится, если вместо суммы вида (9.9a) минимизировать взвешенную сумму вида

$$k_a = k_1 + a_1 k_2 + a_2 k_3 + \dots + a_{i-1} k_i + \dots + a_{m-1} k_m, \quad (9.10a)$$

$$\text{где } 0 < a_i < \infty, \quad (9.10b)$$

т. е. решать задачу

$$\text{обеспечить } \min_{S \in M_{cd}} k_a. \quad (9.10b)$$

Действительно, из (9.9) и (9.10) следует, что $k'_a = \lambda_1 k_a$, если $a_{i-1} = \lambda_i / c_i$. Так как в процессе минимизации k_a (или k'_a) значения весовых коэффициентов λ_i не изменяются, то из минимума k_a следует минимум k'_a и наоборот. Таким образом, записи постановки задачи при весовом методе в виде соотношений (9.9) и (9.10) эквивалентны и в дальнейшем будут применяться оба эти вида записи. Однако при рассмотрении основных свойств весового метода запись вида (9.10) удобнее, так как при этом число весовых коэффициентов на единицу меньше.

Итак, при весовом методе решается задача (9.10) при некоторой фиксированной комбинации ($a_1, \dots, a_{m-1}$) весовых коэффициентов. В результате решения этой задачи находится соответствующая система (класс систем) S_a и соответствующие ей значения $k_{1a}, \dots, k_{ma}$ показателей, обеспечивающие получение минимального значения взвешенной суммы (9.9), т. е.

$$k_{a, \min} = k_a + a_1 k_{2a} + \dots + a_{m-1} k_{ma}. \quad (9.11)$$

Очевидно, что как вид S_a системы S [и (или) значения ее параметров], так и соответствующие значения ее показателей $k_{1a}, \dots, k_{ma}$ зависят от конкретного значения выбранной комбинации $a_1, \dots, a_{m-1}$ весов. Поэтому, варьируя эти комбинации весов во всех допустимых неравенствами (9.10b) пределах, можно в принципе найти зависимости вида

$$S_a = S_a(a_1, \dots, a_{m-1}), \quad (9.12)$$

$$k_{1a} = f_{1a}(a_1, \dots, a_{m-1}), \dots, k_{ma} = f_{ma}(a_1, \dots, a_{m-1}). \quad (9.13)$$

Решая систему уравнений (9.13) относительно весовых коэффициентов $a_1, \dots, a_{m-1}$, можно в принципе получить уравнение, связывающее показатели $k_{1a}, \dots, k_{ma}$. Запишем его в виде

$$k_{1a} = f_a(k_{2a}, \dots, k_{ma}) \quad (9.14)$$

и назовем зависимость (9.14) *весовой поверхностью*. Каждой точке этой поверхности соответствует некоторая система (класс сис-

тем) S_k , и наоборот. Множество точек этой поверхности обозначим M_k . В соответствии с (9.10б) задача (9.10в) должна быть решена для континуума значений весов $a_1, \dots, a_{m-1}$; поэтому ее точное решение возможно лишь в тех случаях, когда его удастся найти аналитическим путем, т. е. в общем виде, не прибегая к числовым расчетам. В тех случаях, когда аналитического решения найти не удастся, можно, задаваясь конкретными численными значениями весов $a_1, \dots, a_{m-1}$, найти лишь некоторые опорные дискретные точки весовой поверхности. В [58] показано, что множество M_k обладает следующими свойствами.

1. Каждая точка из M_k — нехудшая, но часть нехудших точек может не войти в состав M_k , т. е. $M_k \subset M_{nk}$.

2. Если множество M_{cd} выпуклое, то M_k совпадает с M_{nk} .

3. Если решение задачи (9.10в) существует для всех допустимых (в указанном выше смысле) комбинаций показателей $k_1, \dots, k_m$, то M_k совпадает с M_{nk} .

4. Если решения задачи (9.10в) для некоторых допустимых комбинаций ($k_1, \dots, k_m$) показателей качества не существует, то этим комбинациям показателей в R_m могут соответствовать как нехудшие, так и худшие точки и для таких комбинаций ответ (худшей или нехудшей точке они соответствуют) можно получить лишь каким-либо другим методом, например методом рабочих характеристик. Поясним эти положения на следующих простых примерах.

Пример 1. Множество M_{cd} имеет вид, приведенный на рис. 9.6,а, и повторно изображено на рис. 9.9. В этом случае множество M_{nk} состоит из всех точек кривой $a'bc'$. Выясним, какие результаты даст весовой метод. Взвешенная сумма (9.10а) показателей качества в данном случае имеет вид

$$k_a = k_1 + ak_2, \quad (9.15)$$

$$\text{где } 0 < a < \infty. \quad (9.16)$$

На рис. 9.9 эта зависимость изображается прямой l :

$$k_1 = k_a - ak_2. \quad (9.17)$$

образующей с (отрицательной) осью абсцисс угол α , где

$$\operatorname{tg} \alpha = a. \quad (9.18)$$

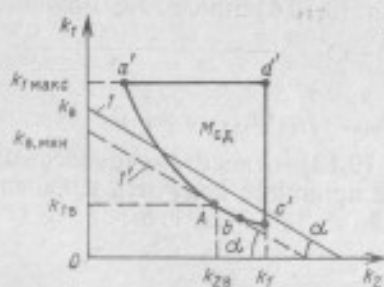


Рис. 9.9

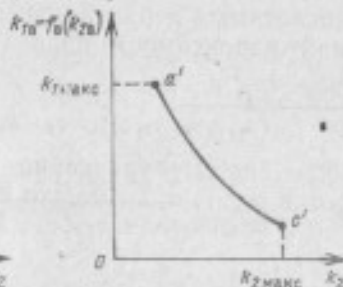


Рис. 9.10

На оси ординат прямая (9.17) отсекает отрезок, равный k_b . Очевидно, значение k_b имеет минимум, когда прямая (9.17) совпадает с касательной к левой нижней границе $a'bc'$, т. е. принимает положение l' , изображенное на рис. 9.9 штриховой линией.

Система S_k , которой соответствует минимальное значение k_b , должна принадлежать как множеству M_{cd} (т. е. области $a'bc'd'$), так и прямой $k_{1b} = k_b - ak_2 = k_{1b}$, т. е. прямой l' . Следовательно, системе S_k соответствует точка A касания прямой l' к границе $a'bc'$, а координаты этой точки равны k_{1b} и k_{2b} . Но найденная таким образом точка A принадлежит множеству M_{nk} . Следовательно, при данном значении весового коэффициента a , т. е. угла наклона α , найдена одна из точек искомого множества M_{nk} . Если несколько увеличить a , т. е. угол наклона прямой l , то точка ее касания к границе $a'bc'$ будет расположена выше точки A , а при уменьшении a она будет перемещаться по границе $a'bc'$ от точки A к точке c' . Поэтому, придавая весовому коэффициенту a все возможные значения (в пределах $0 < a < \infty$), можно найти все точки кривой $a'bc'$, т. е. все множество M_{nk} . Так как каждой найденной точке A (k_{1b} , k_{2b}) соответствуют вполне определенные значения координат k_{1b} , k_{2b} , то будет найдена весовая поверхность

$$k_{1b} = f_1(k_{2b}), \quad (9.19)$$

которая в данном примере полностью совпадает с левой нижней границей $a'bc'$, т. е. с диаграммой обмена, и имеет вид, изображенный на рис. 9.10.

Рассматриваемое множество M_{cd} выпуклое; поэтому, варьируя вес a , удается найти все точки диаграммы обмена, т. е. весовая поверхность (9.19) оказывается определенной для всех допустимых значений показателя k_2 и, следовательно, множество M_k совпадает с искомым множеством M_{nk} . Эти результаты полностью соответствуют изложенным общим свойствам весового метода.

Пример 2. Множество M_{cd} имеет вид, изображенный на рис. 9.11 (с границей $gbcdef$), т. е. не является выпуклым. При этом диаграммой обмена является левая нижняя граница $gbcde$, содержащая множество M_{nk} нехудших точек, т. е. остальные точки множества M_{cd} — худшие.

Применяя весовой метод аналогично тому, как это было сделано в примере 1, получаем следующие результаты. При значении весового коэффициента $a = a_{кр}$, т. е. $\alpha = \alpha_{кр}$, прямая (9.17) касается левой нижней границы области M_{cd} в двух точках, b и d , т. е. решение задачи (9.10в) при таком a позволяет

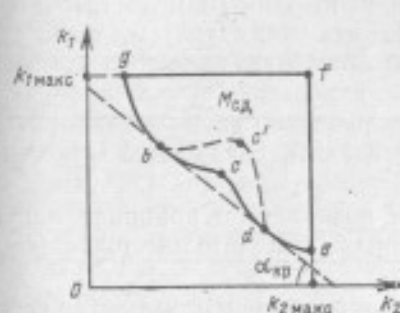


Рис. 9.11

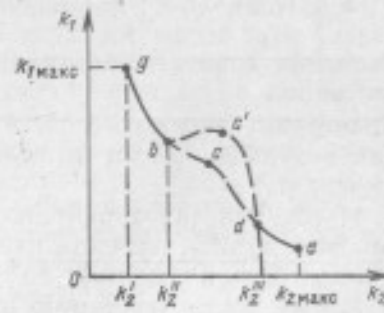


Рис. 9.12

найти нехудшие точки b и d [т. е. координаты (k_{1a}, k_{2a}) , соответствующие этим точкам].

При $a_{np} < a < \infty$ будут найдены все точки участка gb диаграммы обмена, а при $0 < a < a_{np}$ — все точки участка db диаграммы обмена. Но ни одна из внутренних точек участка bcd диаграммы обмена при этом найдена не будет. Поэтому весовая поверхность в данном случае имеет вид, изображенный на рис. 9.12 сплошной линией. На участках gb и de она совпадает с соответствующими участками диаграммы обмена, а на участке bd (т. е. при $k''_2 < k_{2a} < k'''_2$) эта поверхность (в данном двумерном случае — кривая) не существует (не определена). Следовательно, в данном случае весовой метод позволил найти не все множество M_{nx} нехудших точек, а лишь его часть. Если бы граница множества M_{ed} на участке bd не была монотонно убывающей (например, соответствовала штриховой кривой $bc'd$), то соответствующие ей точки были бы не худшими, а худшими, но они также не были бы определены весовым методом. Поэтому на том участке, где весовая поверхность оказалась несуществующей, могут иметься как худшие, так и нехудшие точки, и, следовательно, рассматривая такую неполностью определенную весовую поверхность, нельзя установить, какие именно точки оказались пропущенными — нехудшие или худшие. Ответ на этот вопрос может дать лишь применение какого-либо другого метода отыскания нехудших систем, например метода рабочих характеристик. Нетрудно убедиться, что результаты, полученные в этом примере, также полностью соответствуют сформулированным выше общим свойствам весового метода.

Некоторые дополнительные свойства этого метода даны в [58]. Одно из них состоит в следующем. Если полученное весовым методом решение [система S_a и (или) значения ее показателей качества $k_{1a}, \dots, k_{ma}$] оказывается не зависящим от значений весовых коэффициентов $a_1, \dots, a_{m-1}$ (или $\lambda_1, \dots, \lambda_m$), то это означает, что множество M_{nx} является вырожденным — содержит всего одну точку (систему, класс систем).

Справедливо и обратное положение: если M_{nx} вырождено, т. е. содержит всего одну точку, то эта точка при применении весового метода не будет пропущена и полученное этим методом решение не будет зависеть от значений весов $a_1, \dots, a_{m-1}$ (или $\lambda_1, \dots, \lambda_m$).

Таким образом, при применении весового метода независимость полученного решения от значений весовых коэффициентов $a_1, \dots, a_{m-1}$ (или $\lambda_1, \dots, \lambda_m$) является необходимым и достаточным условием вырожденности множества нехудших систем, т. е. существования совместно-оптимального решения задачи векторной оптимизации.

Сравнение основных свойств изложенных двух методов отыскания нехудших систем позволяет сделать следующие основные заключения.

1. Метод рабочих характеристик позволяет в принципе найти все множество M_{nx} , но часть найденных этим методом точек (систем) могут быть и худшими: $M_p \supset M_{nx}$.

2. Любая точка (система), найденная весовым методом, всегда нехудшая, но часть нехудших точек может быть при этом пропущена: $M_v \subset M_{nx}$.

3. Во многих практических задачах множество точек (систем), найденное как методом рабочих характеристик, так и весовым методом, полностью совпадает с M_{nx} , т. е. недостатки, отмеченные в пп. 1 и 2, отсутствуют. Об отсутствии этих недостатков проще всего судить по виду найденной рабочей поверхности (при применении метода рабочих характеристик) или весовой поверхности (при весовом методе): рабочая поверхность должна быть монотонно убывающей по каждому из ее аргументов $k_2, \dots, k_m$, а весовая поверхность должна существовать во всем диапазоне допустимых комбинаций показателей $k_2, \dots, k_m$.

4. Если применение какого-либо из указанных двух методов не позволяет получить решение в аналитическом виде, то удаётся точно найти лишь конечное число точек искомого множества M_{nx} (искомой диаграммы обмена).

Рассмотрим теперь комбинированный метод отыскания M_{nx} (диаграммы обмена). Он отличается от метода рабочих характеристик тем, что в разряд ограничений типа равенств переводится не $(m-1)$ показателей качества, а меньше, т. е. вектор качества представляется в виде $K = \langle k_1, \dots, k_{m'}; \underline{k}_{m'+1}, \dots, \underline{k}_m \rangle$, где черта снизу, как и ранее, обозначает, что подчеркнутые ею показатели качества рассматриваются как некоторые фиксированные (замороженные) величины. Для оставшихся нефиксированными m' показателей качества $k_1, \dots, k_{m'}$ ищется множество M'_{nx} систем, нехудших в смысле этих m' показателей. Это множество можно найти любым методом, например весовым, и тем самым получить диаграмму обмена между m' показателями качества вида

$$k_1 = f'_{nx}(k_2, \dots, k_{m'}; \underline{k}_{m'+1}, \dots, \underline{k}_m). \quad (9.20)$$

Эта диаграмма обмена зависит от значений $\underline{k}_{m'+1}, \dots, \underline{k}_m$ фиксированных показателей, что и отмечено в записи (9.20). Диаграмма обмена (9.20) ищется для всех допустимых [неравенствами (8.12)] значений показателей качества $k_1, \dots, k_m$. Затем проверяется зависимость функции (9.20) от значений каждого из ее фиксированных параметров $\underline{k}_{m'+1}, \dots, \underline{k}_m$. Как показано в [58], если эта зависимость окажется монотонно убывающей по всем этим параметрам (во всем диапазоне их допустимых значений), то в выражении (9.20) черта снизу под показателями $\underline{k}_{m'+1}, \dots, \underline{k}_m$ может быть ликвидирована и функция (9.20), рассматриваемая как функция всех показателей качества, совпадает с искомой диаграммой обмена между m (а не только m') показателями качества.

Следовательно, при монотонно убывающей зависимости показателя k_1 от $\underline{k}_{m'+1}, \dots, \underline{k}_m$ в выражении (9.11) можно формально отбросить не только черту снизу под показателями $\underline{k}_{m'+1}, \dots, \underline{k}_m$ («разморозить» значения этих показателей), но и штрих в обозначении f'_{nx} , т. е. записать (9.20) в виде

$$k_1 = f_{nx}(k_2, \dots, k_{m'}; k_{m'+1}, \dots, k_m), \quad (9.21)$$

где $f_{nx}(\cdot)$ — уравнение искомой диаграммы обмена между всеми m показателями качества.

При этом проверять наличие монотонной зависимости k_1 от показателей $k_2, \dots, k_m$ не требуется, так как эта зависимость заведомо монотонна вследствие того, что уравнение (9.21) является уравнением диаграммы обмена между показателями $k_1, \dots, k_m$. Например, если при $m'=2, m=3$ было найдено, что функция (9.20) имеет вид

$$k_1 = \ln \frac{1}{k_2} / k_3, \quad (9.22)$$

то, поскольку зависимость k_1 от k_3 — монотонно убывающая, уравнение искомой диаграммы обмена между показателями k_1, k_2 и k_3 можно получить из (9.22), отбрасывая черту снизу под показателем k_3 :

$$k_1 = \ln \frac{1}{k_2} / k_3.$$

Если в выражении (9.20) зависимость k_1 хотя бы от одного из показателей $k_{m'+1}, \dots, k_m$ окажется не монотонно убывающей, то зависимость $k_1 = f_{\text{нх}}(k_2, \dots, k_m; k_{m'+1}, \dots, k_m)$, получаемая из (9.20) отбрасыванием черты снизу, в пространстве R_m содержит не только все нехудшие (в смысле m показателей) точки, но и некоторые худшие. Следовательно, множество $M_{\text{нх}}$ точек, найденных комбинированным методом, включает искомое множество $M_{\text{нх}}$, а в частном (но нередко имеющем место на практике) случае совпадает с $M_{\text{нх}}$.

Изложенный метод называется комбинированным (смешанным), так как в общем случае он не совпадает ни с методом рабочих характеристик (при $m' > 1$), ни с весовым методом (при $m' < m$). Очевидно, его имеет смысл применять лишь при $m \geq 3$, так как при $m=2$ получается $m'=1$, и комбинированный метод сводится к методу рабочих характеристик.

Основное преимущество комбинированного метода состоит в том, что при $M_{\text{нх}} = M_{\text{нх}}$ он позволяет свести задачу отыскания множества систем, нехудших в смысле m показателей, к отысканию множества $M'_{\text{нх}}$ систем, нехудших в смысле меньшего числа m' показателей. Особенно большое упрощение получается тогда, когда множество $M'_{\text{нх}}$ уже было найдено в результате предыдущих исследований и ряд дополнительных показателей качества $k_{m'+1}, \dots, k_m$ при отыскании $M'_{\text{нх}}$ был переведен в разряд ограничений типа равенств. Тогда для «автоматического» распространения уже известной для m' показателей диаграммы обмена (9.20) на все m показателей достаточно лишь убедиться в том, что зависимость k_1 от дополнительных показателей $k_{m'+1}, \dots, k_m$ является монотонно убывающей.

9.3. ПРИМЕНЕНИЕ УСЛОВНОГО КРИТЕРИЯ ПРЕДПОЧТЕНИЯ

В гл. 8 отмечалось, что условный критерий предпочтения основан на введении некоторого результирующего скалярного показателя качества $K_{\text{рез}}$, являющегося известной (выбранной) функ-

цией (8.20) первичных показателей $k_1, \dots, k_m$, называемой результирующей целевой функцией (РЦФ). Поскольку возможно, вообще говоря, неограниченное множество видов РЦФ, число возможных видов УКП, в отличие от БКП, неограничено.

В гл. 8 отмечалось также, что РЦФ можно выбрать либо в виде функции потерь (чем меньше $K_{\text{рез}}$, тем лучше система), либо в виде функции полезности (чем больше $K_{\text{рез}}$, тем лучше система). В дальнейшем для конкретности полагается, что $K_{\text{рез}}$ выбрана в классе функций потерь. Наконец, отмечалось, что РЦФ может быть корректной или не вполне корректной. В первом случае зависимость $K_{\text{рез}}$ от каждого из (стандартных) показателей $k_1, \dots, k_m$ является монотонно возрастающей [при любых интересующих нас комбинациях значений остальных $(m-1)$ показателей], а во втором — лишь неубывающей. Нетрудно доказать (см. также [58]), что, когда выбранная РЦФ вполне корректна, найденная на ее основе система обязательно относится к множеству нехудших. Если же РЦФ не вполне корректна, то в отношении результата оптимизации, полученного при ее применении, можно утверждать следующее [58]: если значение $K_{\text{рез}}$ минимально лишь при единственном значении вектора $\mathbf{K} = \langle k_1, \dots, k_m \rangle$, то этому минимуму соответствует нехудшее значение вектора $\mathbf{K}$, т. е. найденная оптимальная система принадлежит к классу нехудших; если же минимуму $K_{\text{рез}}$ соответствует не одно, а два или более значений вектора $\mathbf{K}$, то по меньшей мере одно из них будет нехудшим, а остальные могут быть и худшими.

Для иллюстрации этого положения рассмотрим следующий пример. Пусть РЦФ выбрана в виде

$$K_{\text{рез}} = f_{\text{рез}}(k_1, \dots, k_m) = \begin{cases} k_1 & \text{при } k_2 \leq k_{2 \text{ макс}}, \dots, k_m \leq k_{m \text{ макс}}, \\ \infty & \text{вне этой области} \end{cases} \quad (9.23)$$

или, что эквивалентно, $K_{\text{рез}} = k_1$ при условии

$$k_2 \leq k_{2 \text{ макс}}, \dots, k_m \leq k_{m \text{ макс}} \quad (9.24)$$

(т. е. значения $k_i > k_{i \text{ макс}}$, $i=2, \dots, m$, считаются заведомо недопустимыми).

Иначе говоря, в РЦФ вида (9.23) все показатели, кроме одного (k_1), которому присваивается номер 1, переводятся в разряд ограничений типа неравенств и ищется (с учетом этих ограничений) минимум единственного (главного) показателя k_1 .

Из (9.23) следует, что в интересующей нас области показателей качества ($k_i \leq k_{i \text{ макс}}$, $i=2, \dots, m$) зависимость $K_{\text{рез}}$ от аргументов $k_1, \dots, k_m$ имеет вид, изображенный на рис. 9.13: зависимость от показателя k_1 является возрастающей, а от остальных показателей — лишь неубывающей. Поэтому РЦФ вида (9.23) является не вполне корректной.

Пусть, например, $m=2$ и множество $M_{\text{сд}}$ строго допустимых показателей качества имеет вид, изображенный на рис. 9.14 (об-

ласть abc). Тогда применение РЦФ вида (9.23) [или (9.24)] означает, что ищется минимум величины k_1 при условии, что $k_2 \leq k_{2\text{ макс}}$. Поэтому, как следует из рис. 9.14, в результате применения такой РЦФ будет выбрана точка b , т. е. одна из нехудших точек (поскольку области abc соответствует множество $M_{\text{нх}}$ нехудших точек в виде всех точек кривой ab).

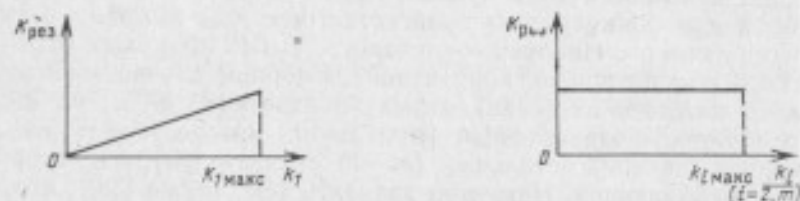


Рис. 9.13

Пусть теперь множество $M_{\text{сд}}$ имеет вид области $abcd$, изображенной на рис. 9.15. В этом случае применение той же РЦФ, т. е. отыскание минимума показателя $K_{\text{рез}} = k_1$ при условии, что $k_2 \leq k_{2\text{ макс}}$, приведет к тому, что выбранному критерию будут удовлетворять все точки отрезка bc прямой, из которых только точка b нехудшая, а все остальные точки — худшие (так как имеют большие, чем у точки b , значения показателя k_2 при том же значении показателя k_1).

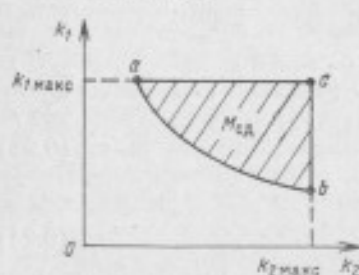


Рис. 9.14

Из этих примеров видно, что в первом случае (рис. 9.14) решение задачи оптимизации оказалось единственным (в смысле значения вектора K) и привело к выбору нехудшей точки (системы). Во втором случае (рис. 9.15) применение той же РЦФ привело к нахождению континуума значений вектора K (соответствующих континууму точек отрезка bc) и лишь одно из этих значений оказалось нехудшим. Эти результаты полностью соответствуют сформулированным общим положениям о возможных результатах оптимизации при применении не вполне корректной РЦФ.

Очевидно, тогда, когда решение задачи оптимизации оказывается не единственным (в смысле вектора K), из найденного мно-

жества значений вектора K следует отсеять все худшие значения, оставив лишь нехудшие (или лишь одно из них). Этот недостаток не вполне корректной РЦФ по сравнению с вполне корректной часто окупается большей возможностью ее обоснования. Пусть, например, помехоустойчивость радиолокатора обнаружения цели оценивается двумя первичными показателями качества — условными вероятностями пропуска сигнала и ложной тревоги, т. е.

$$k_1 = P_{\text{пр}}, \quad k_2 = P_{\text{лт}}. \quad (9.25)$$

Тогда можно ввести результирующий показатель качества в виде (8.21), т. е. положить

$$K_{\text{рез}} = c_1 P_{\text{пр}} + c_2 P_{\text{лт}}, \quad (9.26)$$

где $c_1 > 0$, $c_2 > 0$, $c_1 + c_2 = 1$.

Результирующая целевая функция (9.26) возрастает по каждому из ее аргументов ($P_{\text{пр}}$ и $P_{\text{лт}}$) и, следовательно, вполне корректна. Однако для ее применения необходимо обосновать значение веса c_1 (или c_2), что в большинстве случаев осуществить не удастся. Если вместо этого выбрать РЦФ вида (9.24), т. е. положить $K_{\text{рез}} = P_{\text{пр}}$ при $P_{\text{лт}} \leq P_{\text{лт макс}}$, то критерий оптимальности принимает вид:

$$\text{обеспечить } \min_S P_{\text{пр}} \text{ при } P_{\text{лт}} \leq P_{\text{лт макс}}. \quad (9.27)$$

При этом достаточно обосновать лишь максимально допустимое (с точки зрения назначения радиолокатора) значение $P_{\text{лт макс}}$ вероятности ложной тревоги. Такое обоснование, хотя бы приближенное, в большинстве практических задач возможно и более убедительно, чем обоснование веса c_1 (или c_2).

Ранее было рассмотрено деление возможных РЦФ на функции потерь и функции полезности, а также на вполне и не вполне корректные. Кроме того, возможные РЦФ целесообразно делить на объективные и субъективные. РЦФ называется объективной [58], если ее вид может быть достаточно объективно обоснован исходя из назначения системы. В противном случае РЦФ называется субъективной.

Рассмотрим в качестве иллюстрации следующий важный случай, при котором можно достаточно объективно обосновать РЦФ. Проектируемая система S с показателями качества $k_1, \dots, k_m$ является частью более сложной системы $S_{\text{сд}}$, характеризующейся единственным (стандартным) показателем качества $K_{\text{сд}}$, и известна зависимость

$$K_{\text{сд}} = f_{\text{сд}}(k_1, \dots, k_m) \quad (9.28)$$

этого общего показателя от показателей (стандартных) $k_1, \dots, k_m$. Если показатели качества $k_1, \dots, k_m$ выбраны вполне корректно [т. е. чем меньше каждый из них при фиксированных значениях остальных $(m-1)$ показателей, тем лучше система], то зависимость (9.28) будет монотонно возрастающей по каждому из по-

казателей. При этом, очевидно, эта зависимость и должна быть выбрана в качестве РЦФ. Следовательно, можно принять

$$K_{\text{рез}} = K_{\text{сн}}$$

$$\text{т. е. } K_{\text{рез}} = f_{\text{сн}}(k_1, \dots, k_m), \quad (9.29)$$

и оптимальной будет та система $S = S^*$, которая обеспечит минимум (по всем $S \in M_{\text{сн}}$) значения $K_{\text{рез}}$.

Пусть, например, синтезируется система наведения зенитного снаряда по двум основным показателям качества:

$$k_1 = C_{\text{н}}, \quad k_2 = k_{\text{эф}}, \quad (9.30)$$

где $C_{\text{н}}$ — стоимость системы наведения, а $k_{\text{эф}}$ — эффективное (среднеквадратическое) значение ошибки наведения (промаха). При этом роль сложной системы $S_{\text{сн}}$ играет управляемый снаряд в целом, характеризуемый двумя основными показателями

$$k_{1 \text{ сн}} = C_{\text{с}}, \quad k_{2 \text{ сн}} = 1 - P_{\text{н}} = P_{\text{нп}}, \quad (9.31)$$

где $C_{\text{с}}$ — стоимость снаряда в целом, а $P_{\text{н}}$ — обеспечиваемая им вероятность поражения цели.

Если считать, что стоимость $C_{\text{с}}$ фиксирована, т. е.

$$C_{\text{с}} = C_{\text{с0}}, \quad (9.32)$$

то снаряд в целом характеризуется единственным показателем качества

$$K_{\text{сн}} = k_{2 \text{ сн}} = P_{\text{нп}}. \quad (9.33)$$

Приближенно можно полагать [22], что

$$P_{\text{н}} = 1 / (1 + h_{\text{эф}}^2 / R_{\text{эф}}^2), \quad (9.34)$$

где $R_{\text{эф}}$ — эффективный радиус поражающего действия снаряда. Приближенно можно полагать также, что

$$R_{\text{эф}}^2 = AC_0, \quad (9.35)$$

где $C_0 = C_{\text{с0}} - C_{\text{н}}$ — стоимость остальной (т. е. за исключением системы самонаведения) части снаряда.

Тогда с учетом (9.31), (9.34) и (9.35) получается, что

$$P_{\text{нп}} = [1 + A(C_{\text{с0}} - C_{\text{н}}) / h_{\text{эф}}^2]^{-1}, \quad (9.36)$$

$$\text{где } C_{\text{н}} < C_{\text{с0}}. \quad (9.37)$$

С учетом (9.29) и (9.33) выражение (9.36) можно записать также в следующем виде:

$$K_{\text{сн}} = [1 + A(C_{\text{с0}} - k_1) / k_2^2]^{-1}, \quad (9.38)$$

$$\text{где } k_1 < C_{\text{с0}}. \quad (9.39)$$

Нетрудно убедиться, что зависимость (9.38) является монотонно возрастающей по каждому из аргументов (k_1 и k_2). В соответствии с изложенным эту зависимость и можно принять в качестве РЦФ, т. е. следует положить

$$K_{\text{рез}} = [1 + A(C_{\text{с0}} - k_1) / k_2^2]^{-1}, \quad k_1 < C_{\text{с0}}, \quad (9.40)$$

и считать оптимальной такую систему наведения снаряда (такую комбинацию показателей k_1 и k_2 ее качества), при которой значение $K_{\text{рез}}$ минимально.

Однако рассмотренный случай, в котором возможно объективное обоснование РЦФ, имеет место далеко не всегда, так как сформулированные в нем допущения могут не выполняться:

проектируемая система S может быть автономной, т. е. не являться частью какой-либо более сложной системы (например, может проектироваться автономный измерительный прибор);

проектируемая система (подсистема) может быть универсальной, т. е. входить не в какую-либо определенную сложную систему, а в любую из целого класса различных сложных систем;

зависимость показателя качества $K_{\text{сн}}$ сложной системы от показателей качества $k_1, \dots, k_m$ проектируемой системы S может быть на данной стадии проектирования системы S неизвестна;

качество сложной системы (в состав которой входит проектируемая система S) часто характеризуется не одним, а двумя или более показателями качества.

Возможны и некоторые другие, более частные, случаи, при которых можно достаточно объективно обосновать вид РЦФ [см., например, [58]]. Однако чаще всего такое обоснование невозможно, и, следовательно, выбранная РЦФ является в значительной мере субъективной (произвольной, волевой). Тогда для повышения объективности обоснования приходится привлекать к нему коллектив экспертов, имеющих большой опыт проектирования и (или) эксплуатации систем подобного или близких классов [58].

Если до применения УКП удалось найти множество $M_{\text{нх}}$ наилучших систем, то УКП целесообразно применять к этому множеству (а не к исходному множеству $M_{\text{сн}}$ строго допустимых систем), так как при этом, как правило, существенно облегчается процедура отыскания оптимальной (в смысле данного УКП) системы. Кроме того, если не вполне корректная РЦФ применяется к множеству $M_{\text{нх}}$ (а не к $M_{\text{сн}}$), исключается возможность выбора на ее основе какой-либо из худших систем.

Литература: основная [58].

Глава 10

ПОТЕНЦИАЛЬНЫЕ ХАРАКТЕРИСТИКИ РАДИОСИСТЕМ — ДИАГРАММЫ ОБМЕНА

10.1. ОДНОМЕРНЫЕ ПОТЕНЦИАЛЬНЫЕ ХАРАКТЕРИСТИКИ РАДИОСИСТЕМ

Одномерной потенциальной характеристикой радиосистемы (или радиоустройства) называется потенциальное, т. е. наилучшее теоретически достижимое значение одного из показате-

лей ее качества (элементарного или приведенного) при определенных сформулированных математически исходных данных. Ранее уже рассматривались некоторые из таких характеристик — потенциальная точность, потенциальная пропускная способность и потенциальная стоимость. Поскольку потенциальная точность определялась при наличии действия помех, она называлась также потенциальной помехоустойчивостью.

При отыскании одномерной потенциальной характеристики, т. е. определении минимального (наилучшего) значения $k_{\min}$ одного из показателей качества (элементарного или приведенного), применяется сочетание двух приемов — оптимизации и идеализации. Оптимизация состоит в том, что отыскивается система (класс систем), обладающая наименьшим (наилучшим) значением $k_{\min}$ (скалярного) показателя качества k среди всех допустимых систем, т. е. систем, удовлетворяющих сформулированным математически исходным данным D' [см. (8.10)]. Идеализация заключается, во-первых, в том, что вектор качества полагается содержащим всего одну составляющую:

$$K = k \quad (10.1)$$

(здесь k — элементарный или приведенный показатель качества). Остальные показатели качества переводятся в разряд ограниченного типа равенств, а часть из них вообще не учитывается (полностью игнорируется). Во-вторых, учитываются не все реальные ограничения структуры системы и значений ее параметров, а лишь некоторые из них. В-третьих, облегчаются по сравнению с реальными условия U работы системы.

Очевидно, все эти три вида идеализаций позволяют получить меньшее (лучшее) значение показателя k по сравнению с реальной системой. Поэтому найденное для таких исходных данных значение $k_{\min}$ и называется потенциальным (т. е. наилучшим теоретически, а не реально достижимым). Например, при отыскании потенциальной точности (с учетом действия помех) определяется минимальное значение среднего риска R (или полной вероятности ошибки $P_{\text{ош}}$, или среднего квадрата ошибки $\overline{\delta^2}$), т. е. единственного (приведенного) показателя качества. Все остальные показатели качества при этом переводятся в разряд ограничений типа равенств (пропускная способность, интенсивность помех, мощность передатчика и др.) либо вообще не учитываются (масса, габаритные размеры, стоимость, показатели надежности и др.).

Из ограничений O_8 на искомую систему учитываются лишь некоторые самые общие, например система ищется в пределах класса одноканальных физически реализуемых систем. Условия работы системы существенно облегчаются: учитываются не все реально возможные помехи, а лишь некоторые важнейшие из них (например, только внутренний шум и флуктуации сигнала); некоторые классы мешающих воздействий (например, механические нагрузки, колебания температуры, давления, влажности и др.) вообще не учитываются.

На основании материала гл. 4 и 7 читателю предлагается самому убедиться, что указанные три вида идеализаций применяются при отыскании и потенциальной пропускной способности, и потенциальной стоимости.

Характер идеализаций, принимаемых при отыскании данной потенциальной характеристики, может быть самым разнообразным. Соответственно могут существенно различаться и найденные при этом потенциальные значения $k_{\min}$ показателя качества. Например, потенциальная точность (при наличии действия помех) существенно зависит от числа и вида учитываемых помех (только внутренний шум, внутренний шум и флуктуации сигнала и т. п.). Строго говоря, множество возможных постановок задачи (видов исходных данных D') при отыскании данной потенциальной характеристики неограниченно велико и имеет мощность типа континуума или даже большую. Поэтому для дальнейших научных исследований в этой области, впрочем как и в любой другой, имеются неограниченные возможности.

Применение идеализаций при отыскании потенциальной характеристики полезно не только тем, что позволяет найти теоретический предел для данного показателя качества, но и тем, что весьма существенно упрощает процесс оптимизации. Кроме того, потенциальное значение $k_{\min}$ оказывается зависящим от значительно меньшего числа параметров, чем для реальной системы. Например, потенциальная пропускная способность (C) системы зависит всего от трех параметров (P_c , N_0 и Δf), тогда как реальная пропускная способность может зависеть от нескольких десятков или даже сотен параметров системы и условий ее работы. На начальных этапах проектирования такое множество параметров не может быть оценено даже приближенно, тогда как возможные (или требуемые) значения P_c , N_0 , Δf легко оценить даже в самом начале проектирования. В то же время, если идеализации, принятые при определении потенциальной характеристики, не слишком велики, то найденное потенциальное значение $k_{\min}$ показателя k отличается от реального значения k_r этого показателя обычно менее чем на порядок, а часто всего в 2...3 раза. Поэтому на начальных этапах проектирования, и особенно при научных исследованиях, для ориентировочных расчетов часто можно использовать формулы для соответствующих потенциальных значений показателей, делая в них, если это необходимо, соответствующий запас (в 2...5 раз).

Из изложенного следует, что одномерные потенциальные характеристики можно использовать при научных исследованиях и на начальных этапах проектирования для следующих основных целей.

1. Сравнение значения k_r показателя качества реальной системы данного класса с потенциальным значением $k_{\min}$ этого показателя. Если окажется, что k_r близко к $k_{\min}$, то это будет означать, что реальная система по данному показателю близка к оптимальной и дальнейшее улучшение показателя k_r возможно лишь путем

облегчения исходных данных, принятых при его отыскании (например, при определении пропускной способности C путем увеличения полосы Δf или средней мощности P_c или путем уменьшения спектральной плотности N_0 внутреннего шума).

Если $k_p \gg k_{p, \min}$, то это означает, что рассматриваемая система по данному показателю качества далека от оптимальной и необходимо исследовать возможность ее улучшения (т. е. приближения k_p к $k_{\min}$). При этом может быть полезным сравнение структур и значений параметров реальной (S_p) и оптимальной по этому показателю (S^*) систем. Это исследование покажет, что является основной причиной отклонения k_p от $k_{\min}$: неиспользованные в реальной системе возможности или слишком сильные идеализации, принятые при определении $k_{\min}$.

2. Сравнение различных вариантов S построения системы по значению потенциальной характеристики $k_{\min}$ с целью приближенного выбора варианта, оптимального по этому показателю.

Из сказанного следует, что потенциальные характеристики позволяют выяснить принципиальную возможность улучшения системы и облегчить отыскание путей ее совершенствования.

10.2. МНОГОМЕРНЫЕ ПОТЕНЦИАЛЬНЫЕ ХАРАКТЕРИСТИКИ — ДИАГРАММЫ ОБМЕНА

Качество системы, как правило, можно достаточно полно охарактеризовать лишь совокупностью $k_1, \dots, k_i, \dots, k_m$ нескольких показателей качества (все или некоторые из них могут быть не элементарными, а приведенными). Поэтому важно знать потенциальное значение $k_{i, \min}$ каждого из этих показателей. Но эти потенциальные значения невозможно вычислить независимо. Действительно, в гл. 8 было показано, что минимальное (по всем допустимым системам) значение $k_{i, \min}$ данного показателя качества, например $k_{i, \min}$, зависит от значений остальных $(m-1)$ показателей ($k_2, \dots, k_m$). Эта зависимость имеет вид рабочей поверхности (9.6). Если она оказывается строго монотонной (т. е. убывающей по каждому показателю $k_2, \dots, k_m$), то совпадает с диаграммой обмена (9.2). В противном случае отдельные точки рабочей поверхности оказываются худшими (соответствуют худшим системам) и, следовательно, при сформулированных исходных данных (D' или D'') должны быть исключены из дальнейшего рассмотрения. Следовательно, в любом случае нужно учитывать зависимость $k_{i, \min}$ от остальных $(m-1)$ показателей лишь для тех значений этих показателей, которые удовлетворяют уравнению (9.2) диаграммы обмена. Но в силу указанного в (9.1) свойства m -кратного минимума диаграмма обмена дает зависимость минимального значения каждого из m показателей от значений остальных $(m-1)$ показателей. Следовательно, если найдена диаграмма обмена, то тем самым найдены потенциальные значения $k_{1, \min}, \dots, k_{m, \min}$ каждого из m показателей $k_1, \dots, k_m$ и, более того, найдена зависимость каждого из этих значений $k_{i, \min}$ от значений остальных $(m-1)$ пока-

зателей. Например, как указывалось в § 9.1, двумерная диаграмма обмена, изображенная на рис. 9.3, а, показывает зависимость минимального значения показателя k_1 от значения показателя k_2 и она же представляет собой зависимость минимального значения показателя k_2 от показателя k_1 . Отсюда следует, что m -мерная диаграмма обмена является наиболее полной потенциальной характеристикой данного класса систем (т. е. систем, удовлетворяющих исходным данным D' или D'' , сформулированным при отыскании данной диаграммы обмена): она дает зависимость минимального значения каждого из m показателей качества $k_1, \dots, k_m$ от значений остальных $(m-1)$ показателей. Каждой точке этой диаграммы (т. е. каждой нехудшей точке) соответствует некоторая нехудшая система (класс систем) S . Следовательно, при отыскании диаграммы обмена одновременно определяются и системы, обладающие соответствующими потенциальными значениями отдельных показателей (нехудшие системы). Поэтому m -мерную диаграмму обмена (9.2) можно называть также *многомерной (m -мерной) потенциальной характеристикой системы*.

Докажем, что рассмотренная в § 10.1 одномерная потенциальная характеристика является, как правило, скрытой формой многомерной диаграммы обмена, т. е. многомерной потенциальной характеристики. Действительно, при определении одномерной потенциальной характеристики, как уже отмечалось, все показатели качества, кроме одного, переводятся в разряд ограничений типа равенств или полностью игнорируются (никак не учитываются). Но полностью игнорировать все показатели, кроме одного, недопустимо, так как при этом минимально возможное значение $k_{i, \min}$ единственного минимизируемого показателя окажется равным нулю, т. е. решаемая задача вырождается в тривиальную — такой тривиальный результат можно будет получить и без всякого анализа, просто на основе здравого смысла.

Пусть, например, требуется определить минимально возможное значение вероятности $P_{\text{ош}}$ ошибки воспроизведения сообщений при аддитивной независимой помехе. Если при этом не учитывать ни прямо, ни косвенно никакой из показателей качества, кроме точности, можно и без всякого анализа сказать, что значение $P_{\text{ош}}$ может быть сделана сколь угодно малым, например, за счет соответствующего увеличения мощности (или энергии) сигнала. Аналогично можно утверждать, что и стоимость системы можно сделать сколь угодно малой, если никак не учитывать при этом получающиеся значения каких-либо других показателей качества. Предоставляем читателю доказать, что это утверждение справедливо и для любого другого из показателей качества (показателей надежности, ЭМС, массы и др.).

Следовательно, при минимизации данного показателя качества, например k_1 , необходимо принимать во внимание значение по меньшей мере еще одного антагонистического (т. е. препятствующего уменьшению k_1 до нуля) показателя качества, который обозначим k_2 . При этом показатель k_2 можно учесть, переведя его в

разряд ограничений типа равенства или неравенства, т. е. должна решаться задача:

$$\text{обеспечить } \min_{S \in M_{\text{сд}}} k_1 \text{ при } k_2 = \underline{k}_2 \quad (10.2)$$

либо задача

$$\text{обеспечить } \min_{S \in M_{\text{сд}}} k_1 \text{ при } k_2 \leq k_{2 \text{ макс.}} \quad (10.3)$$

Рассмотрим сначала решение задачи (10.2). Если показатель k_2 фиксирован совершенно жестко, т. е. имеет единственное вполне определенное численное значение, то в результате решения задачи (10.2) будет найдено некоторое конкретное численное потенциальное значение $k_{1 \text{ мин}}$ показателя k_1 . Но при научных исследованиях или на начальных этапах проектирования, т. е. именно тогда, когда знание потенциального (а не реального) значения k_1 может быть полезным, задавать ограничение k_2 в таком жестком виде, как правило, нецелесообразно — важно выяснить, как будет зависеть значение $k_{1 \text{ мин}}$ от показателя k_2 . Это значит, что практически задача (10.2) должна решаться при различных значениях показателя k_2 . Но в этом случае она совпадает с описанной в § 9.2 задачей отыскания рабочей поверхности (9.6). Следовательно, при решении задачи (10.2) будет по-существу найдена рабочая поверхность (в данном случае рабочая кривая) (9.6).

Так как показатель k_2 по условию является антагонистическим, т. е. препятствующим уменьшению k_1 , то зависимость (9.6) получается монотонно убывающей. Но при этом, как следует из свойств рабочей поверхности, она совпадает с диаграммой обмена между k_1 и k_2 . Следовательно, решая задачу (10.2) отыскания потенциального значения единственного показателя качества k_1 , мы тем самым находим диаграмму обмена между показателями k_1 и k_2 . Эта диаграмма обмена представляет собой (в силу свойства m -кратного минимума) не только зависимость $k_{1 \text{ мин}}(k_2)$, но и $k_{2 \text{ мин}}(k_1)$. Иначе говоря, стремясь найти одномерную потенциальную характеристику ($k_{1 \text{ мин}}$), мы находим по-существу двумерную потенциальную характеристику.

Если бы при отыскании минимума показателя k_1 в разряд ограничений типа равенств были переведены не только показатель k_2 , но и показатели $k_3, \dots, k_m$ и полученная в результате минимизации зависимость $k_{1 \text{ мин}}$ от $k_2, \dots, k_m$ оказалась монотонно убывающей по каждому из этих показателей (т. е. показатели $k_2, \dots, k_m$ оказались антагонистическими), то найденная рабочая поверхность (9.6) совпала бы с m -мерной диаграммой обмена. Следовательно, переводя [при отыскании значения $k_{1 \text{ мин}}$] ($m-1$) показателей в разряд ограничений типа равенств, мы находим, как правило (т. е. при антагонистическом характере показателей $k_2, \dots, k_m$) m -мерную диаграмму обмена между $k_1, \dots, k_m$.

Пусть теперь вместо (10.2) решается задача (10.3), причем величина $k_{2 \text{ макс}}$ зафиксирована жестко: задано ее вполне опре-

деленное численное значение. В таком виде задача (10.3) в отличие от (10.2) может иметь смысл и на начальном этапе проектирования, а иногда и при научных исследованиях. В этом случае ее решение дает следующие результаты [58]. Если это решение оказывается единственным (в смысле значения вектора K), то оно является нехудшим, т. е. находится одна (крайняя) точка множества $M_{\text{нх}}$ (диаграммы обмена между k_1 и k_2). Эта точка имеет координаты $(k_{1 \text{ мин}}, k_{2 \text{ макс}})$.

Если решению задачи (10.3) соответствует не одно, а множество значений вектора $K = \langle k_1, k_2 \rangle$, то среди этих значений могут быть и худшие, но по меньшей мере одно из них будет нехудшим. [Например, если $M_{\text{нх}}$ имеет вид, изображенный на рис. 9.7, то, как читатель может сам доказать, в результате решения задачи (10.3) будет найдено множество точек, образующих отрезок (bc) прямой, из которых только одна (точка b) нехудшая]. Однако если показатель k_2 является антагонистическим (по отношению к k_1), то решение оказывается единственным (в смысле вектора K) и, следовательно, в результате решения задачи (10.3), как правило, найдется одна (крайняя) точка диаграммы обмена. Если задачу (10.3) решать не для жестко фиксированного значения ограничения $k_{2 \text{ макс}}$, а для различных значений $k_{2 \text{ макс}}$, т. е. определять зависимость $k_{1 \text{ мин}} = f(k_{2 \text{ макс}})$, то можно доказать [58], что при антагонистическом характере показателя k_2 найденная зависимость будет совпадать с диаграммой обмена между k_1 и k_2 или с частью этой диаграммы (если диапазон изменения значений $k_{2 \text{ макс}}$ недостаточно широк).

В значительной мере аналогичные, но значительно более сложные соотношения получатся, если в разряд ограничений типа неравенств переводится не один, а несколько показателей качества (т. е. $m > 2$).

Таким образом, доказано, что в большинстве практических случаев при отыскании одномерной потенциальной характеристики $k_{1 \text{ мин}}$ находится по-существу m -мерная диаграмма обмена между показателями $k_1, k_2, \dots, k_m$, где $k_2, \dots, k_m$ — показатели, переведенные (при отыскании значения $k_{1 \text{ мин}}$) в разряд ограничений типа равенств.

10.3. ПРИМЕРЫ МНОГОМЕРНЫХ ДИАГРАММ ОБМЕНА

Пример 1. В. А. Котельников еще в 1946 г. получил следующее выражение для оценки точности воспроизведения непрерывного сообщения, постоянного в течение времени T , введенного на его воспроизведение [14]:

$$\bar{\delta}^2 = N_0 / 2T \left[\frac{\partial u_c(t, x)}{\partial x} \right]^2, \quad (10.4)$$

где $\bar{\delta}^2$ — средний квадрат ошибки воспроизведения сообщения (нормированный к максимальному значению сообщения); N_0 — спектральная плотность аддитивного независимого нормального белого шума, являющегося единственным источ-

ником ошибки; $\left[\frac{\partial u_c(t, x)}{\partial x} \right]^2$ — средний квадрат частной производной от сигнала $u_c(t, x)$ по сообщению x . Выражение (10.4) было названо В. А. Котельниковым формулой, определяющей *потенциальную помехоустойчивость*¹ радиоприема указанного класса сообщений. При этом мерой помехоустойчивости являлся средний квадрат $\overline{\delta^2}$ ошибки воспроизведения сообщений в условиях действия помехи (белого шума) с фиксированной интенсивностью N_0 .

Действительно, при фиксированном значении N_0 чем меньше $\overline{\delta^2}$, тем меньше сказываются действие помехи и, следовательно, тем выше помехоустойчивость системы (приемника). Поэтому величина

$$k_1 = \overline{\delta^2} \quad (10.5)$$

в указанных условиях может считаться стандартным показателем качества (показателем помехоустойчивости). В. А. Котельников нашел минимальное значение величины k_1 . Поэтому формулу (10.4) можно записать в виде

$$k_{1 \min} = N_0/2T \left[\frac{\partial u_c(t, x)}{\partial x} \right]^2. \quad (10.6)$$

В процессе минимизации величины k_1 , т. е. оптимизации системы, В. А. Котельниковым был принят ряд идеализаций, каждая из которых по отдельности и все они совместно могут приводить лишь к преуменьшению величины k_1 по сравнению с ее реальным значением. Наиболее важными идеализациями при этом были следующие:

не учитывалось действие каких-либо других источников ошибок (в частности, других помех), кроме белого шума, а форма сигнала $u_c(t, x)$ считалась полностью известной в месте приема (за исключением значения воспроизводимого сообщения x);

сообщение за время T считалось неизменяющимся;

на структуру возможного синтезируемого приемника не накладывалось никаких ограничений, кроме того, что он должен выносить решение за заранее установленное время T .

Следовательно, при получении формулы (10.5) В. А. Котельниковым были выполнены все условия, необходимые для того, чтобы найденная величина $k_{1 \min}$ действительно могла считаться потенциальным значением показателя k_1 : была проведена оптимизация системы (вариантов ее построения) при идеализированных условиях, причем все идеализации были таковы, что могли лишь уменьшить (а не увеличить) величину k_1 . Итак, В. А. Котельников нашел потенциальное значение показателя качества k_1 при фиксированных значениях интенсивности N_0 помехи и времени T воспроизведения сообщения.

Но N_0 и T также можно рассматривать как показатели качества системы (приемника). Действительно, чем больше N_0 при данных (фиксированных) значениях k_1 и T , тем лучше система (приемник), так как позволяет воспроизвести сообщение x с той же точностью ($k_1 = \overline{\delta^2} = \text{const}$) и за то же время ($T = \text{const}$) при большей интенсивности помехи. Таким образом, при $k_1 = \text{const}$ и $T = \text{const}$ величину N_0 можно принять за показатель помехоустойчивой систе-

мы, а $k_1 = \overline{\delta^2}$ — за показатель ее точности. Поэтому в стандартном виде показателем помехоустойчивости может быть величина

$$k_2 = 1/N_0. \quad (10.7)$$

Наконец,

$$k_3 = T \quad (10.8)$$

является показателем быстродействия системы, а обратная ей нестандартная величина

$$1/k_3 = 1/T \quad (10.9)$$

— показателем пропускной способности системы. Поэтому (10.6) можно записать также в следующем виде:

$$k_{1 \min} = 1/2k_2k_3 \left[\frac{\partial u_c(t, x)}{\partial x} \right]^2. \quad (10.10)$$

т. е. В. А. Котельников по существу нашел минимальное значение одного из показателей качества ($k_1 = \overline{\delta^2}$) при фиксированных значениях двух других (k_2 и k_3). При этом значения k_2 и k_3 не были фиксированы жестко (численно) — В. А. Котельникову удалось получить зависимость $k_{1 \min}$ от k_2 и k_3 в аналитическом виде, т. е. при фиксированных, но произвольных значениях показателей k_2 и k_3 .

Следовательно, выражение (10.10) дает зависимость минимального значения одного из показателей ($k_{1 \min}$) от фиксированных, но произвольных значений двух других показателей (k_2 и k_3), т. е. является уравнением рабочей поверхности вида (9.6).

Из (10.10) следует, что зависимость k_1 от k_2 и k_3 — монотонно убывающая. Следовательно, найденная В. А. Котельниковым зависимость (10.10) является диаграммой обмена между показателями k_1 , k_2 и k_3 :

она указывает не только минимальное значение показателя k_1 при фиксированных значениях k_2 и k_3 , но и минимальное значение показателя k_2 (т. е. максимальное значение N_0) при данных k_1 и k_3 и, наконец, минимальное значение показателя k_3 при данных k_1 и k_2 . Иначе говоря, если формулу (10.4) переписать в виде

$$N_0 = \overline{\delta^2} T \cdot 2 \left[\frac{\partial u_c(t, x)}{\partial x} \right]^2, \quad (10.11)$$

то можно утверждать, что она определяет потенциальное значение помехоустойчивости (максимальное теоретически достижимое значение спектральной плотности N_0 помехи, при которой обеспечивается воспроизведение сообщения x за заданное время T с заданным средним квадратом ошибки $\overline{\delta^2}$). Если же формулу (10.4) переписать в виде

$$T = N_0/2\overline{\delta^2} \left[\frac{\partial u_c(t, x)}{\partial x} \right]^2, \quad (10.12)$$

то можно утверждать, что она оценивает потенциальное значение быстродействия T системы (или, что эквивалентно, пропускной способности $1/T$) при заданных интенсивности помехи N_0 и точности воспроизведения сообщения $\overline{\delta^2}$.

Из этого примера следуют два важных вывода.

¹ В. А. Котельников впервые ввел понятие потенциальной помехоустойчивости и определил ее значение для широкого класса сообщений и сигналов [14].

1. Один и тот же показатель (δ^2) может выступать в двух видах — как показатель помехоустойчивости (при действии помехи заданного вида и заданной интенсивности) и как требуемая точность (если задана точность и ищется допустимая интенсивность помехи). В тех случаях, когда δ^2 выступает в качестве показателя точности (т. е. при $\delta^2 = \text{const}$) показателем помехоустойчивости может служить допустимая (с точки зрения обеспечения заданной точности) интенсивность помехи (в рассматриваемом примере интенсивность N_0 шума). Поэтому можно говорить о связи между показателями точности и помехоустойчивости даже в тех случаях, когда единственной причиной неточности (ошибок) является действие помех.

2. Отыскивая минимум единственного (скалярного) показателя k_1 , В. А. Котельников нашел по-существу диаграмму обмена между тремя показателями: k_1 , k_2 и k_3 , т. е. между показателями точности (δ^2), помехоустойчивости (N_0) и быстродействия T (или пропускной способности $1/T$). Иначе говоря, отыскивая одномерную потенциальную характеристику, В. А. Котельников нашел по-существу трехмерную потенциальную характеристику. Поэтому можно с равным правом утверждать, что он нашел не только потенциальную помехоустойчивость системы, но и ее потенциальные точность и быстродействие (пропускную способность).

Пример 2. В гл. 4 было получено соотношение (4.51), которое с учетом (4.47) можно записать в следующем виде:

$$C \approx \frac{P_0}{N_0} / (\ln 2) \left(1 + \frac{2 \log_2 1/P_{0\text{ш}}}{\log_2 M} \right), \quad (10.13)$$

где M определяется из условия

$$M / \log_2 M = \Delta / C. \quad (10.14)$$

При заданных значениях P_0 и Δ эти соотношения связывают между собой следующие (стандартные) показатели качества:

$$k_1 = P_{0\text{ш}}, \quad k_2 = 1/C, \quad k_3 = 1/N_0. \quad (10.15)$$

Их можно рассматривать соответственно как показатели точности, быстродействия (пропускной способности) и помехоустойчивости.

Исторически соотношения (10.13), (10.14) были получены путем отыскания системы (варианта ее построения), обеспечивающей минимальное значение вероятности $P_{0\text{ш}}$ ошибки при фиксированных (но произвольных) значениях показателей k_2 и k_3 . При этой оптимизации был принят ряд идеализаций, которые могут лишь уменьшить (а не увеличить) значение k_1 . Следовательно, соотношения (10.13), (10.14) определяют (в неявном виде) значение потенциальной точности воспроизведения допустимых сообщений в условиях действия помехи (белого шума). При этом минимальное значение одного из показателей (k_1) оказывается зависящим от двух других показателей k_2 и k_3 . Поэтому соотношения (10.13) и (10.14) можно выразить в следующем виде:

$$\frac{1}{k_2} = P_0 k_3 / (\ln 2) \left[1 + \frac{2 \log_2 1/k_1 \text{ мин}}{\log_2 M} \right], \quad (10.16)$$

$$\text{где } M / \log_2 M = \Delta / k_3. \quad (10.17)$$

Они представляют собой записанную в неявном виде рабочую поверхность вида (9.2).

Путем анализа соотношений (10.16) и (10.17) нетрудно убедиться, что зависимость $k_1 \text{ мин}$ как от k_2 , так и от k_3 — монотонно убывающая. Поэтому найденная рабочая поверхность совпадает с диаграммой обмена между показателями k_1 , k_2 и k_3 . Следовательно, соотношения (10.16), (10.17) определяют потенциальные значения не только показателя k_1 (точности), но и показателей k_2 (быстродействия или пропускной способности) и k_3 (помехоустойчивости): они показывают, какую минимальную вероятность ошибки $P_{0\text{ш}}$ можно получить при данных C и N_0 , или же какую максимальную пропускную способность C можно получить при данных $P_{0\text{ш}}$ и N_0 , или, наконец, каково максимальное значение интенсивности помехи, при котором еще можно обеспечить заданные значения C и $P_{0\text{ш}}$.

Таким образом, как и в примере 1, отыскание одномерной потенциальной характеристики (величины $P_{0\text{ш. мин}}$) привело по-существу к нахождению многомерной потенциальной характеристики — диаграммы обмена между показателями точности $P_{0\text{ш}}$, помехоустойчивости N_0 и пропускной способности C .

Пример 3. Как отмечалось в гл. 4, К. Шенноном была получена формула (4.18), определяющая потенциальное значение пропускной способности C системы при фиксированном значении интенсивности N_0 помехи. Было показано, что обладающая такой максимальной пропускной способностью оптимальная система обеспечивает также сколь угодно малую вероятность ошибки:

$$P_{0\text{ш}} \rightarrow 0.$$

При фиксированных значениях P_0 и N_0 формулу (4.18) можно представить в виде

$$k_1 \text{ мин} = 1/\Delta f \log_2 \left(1 + \frac{P_0}{\Delta f} k_2 \right), \quad (10.18)$$

$$\text{где } k_2 = 1/N_0, \text{ а } k_1 \text{ мин} = 1/C \text{ мин}. \quad (10.19)$$

— минимальное (по всем допустимым системам) значение быстродействия системы. Следовательно, соотношение (10.18) является уравнением рабочей поверхности (в данном случае — рабочей кривой), определяющей зависимость минимального значения одного показателя качества (k_1) от значения второго показателя (k_2), переведенного в разряд ограничения типа равенства. Эта зависимость, как следует из (10.18), — монотонно убывающая; поэтому она совпадает с диаграммой обмена между показателями k_1 и k_2 и ее можно записать в виде

$$k_1 = 1/\Delta f \log_2 \left(1 + \frac{P_0}{\Delta f} k_2 \right). \quad (10.20)$$

Она показывает не только, какое можно получить минимальное значение показателя k_1 (т. е. максимальное значение пропускной способности C) при данной интенсивности помехи N_0 , но и каково максимально допустимое значение интенсивности N_0 помехи при данной пропускной способности C .

Следовательно, К. Шеннон, отыскивая потенциальное значение пропускной способности, нашел по-существу диаграмму обмена между пропускной способностью C и помехоустойчивостью N_0 системы, т. е. не одномерную, а двумерную потенциальную характеристику системы.

В двух предыдущих примерах была найдена диаграмма обмена между показателями точности, помехоустойчивости и пропускной способности. Здесь же показатель точности не вошел в диаграмму обмена, поскольку при идеализации,

принятой Шенноном (возможность идеального кодирования сообщений со сколь угодно большой задержкой t_d), оказалось возможным получить сколь угодно высокую точность при любых (конечных) значениях пропускной способности системы и интенсивности помехи. При идеализации, более близкой к реальности, т. е. при конечном времени задержки в передаче информации, точность воспроизведения сообщений (например, полная вероятность ошибки $P_{\text{ош}}$ или средний квадрат ошибки $\overline{\delta^2}$) не может быть сколь угодно высокой и наибольшее достижимое значение пропускной способности системы зависит не только от интенсивности помехи (N_0), но и от требуемой точности воспроизведения сообщений (например, от $P_{\text{ош}}$). При этом отыскание потенциальной пропускной способности приводит к нахождению диаграммы обмена между показателями пропускной способности, помехоустойчивости и точности, как это имело место, например, в рассмотренном примере 2.

Таким образом, случай, рассмотренный Шенноном, является вырожденным: в нем третий показатель качества (показатель точности) исчезает, так как может быть получен сколь угодно близким к идеальному при любых значениях двух остальных показателей качества. Вырожденность этого случая приводит также к тому, что в отличие от примеров 1 и 2 помехоустойчивость системы может оцениваться только интенсивностью помехи (N_0), а не каким-либо показателем точности (например, величинами $P_{\text{ош}}$ или $\overline{\delta^2}$).

Действительно, если предположить, что помехоустойчивость системы оценивается точностью воспроизведения сообщений при наличии помех, то результат, полученный К. Шенноном, означал бы, что помехоустойчивость системы при идеальном кодировании может быть сколь угодно высока — при любой (конечной) интенсивности помехи N_0 может быть получена сколь угодно высокая точность воспроизведения сообщений. Однако из (4.18) следует, что при возрастании интенсивности помехи пропускная способность системы монотонно убывает. Поэтому нельзя обеспечить требуемую не слишком малую пропускную способность C при сколь угодно большой интенсивности помехи (и не слишком большой мощности P_0 сигнала). Следовательно, даже в весьма идеализированном (вырожденном) случае, рассмотренном К. Шенноном, помехоустойчивость системы оказывается не столь угодно высокой, а ограниченной и эта ограниченность правильно учитывается, если помехоустойчивость системы оценивается показателем не точности, а интенсивности помехи.

Таким образом, на основе рассмотренных примеров можно сделать следующие основные выводы.

Помехоустойчивость системы в невырожденном случае можно оценивать как допустимой интенсивностью помехи (N_0) при заданной точности ($P_{\text{ош}}$ или $\overline{\delta^2}$), так и допустимой точностью ($\overline{\delta^2}$ или $P_{\text{ош}}$) при заданной интенсивности помехи (N_0).

Кроме того, при определении как точности системы (в условиях действия помех), так и допустимой интенсивности помехи необходимо учитывать также требование к пропускной способности (или быстродействию) системы. В противном случае показатели точности и помехоустойчивости могут быть получены сколь угодно близкими к идеальным. Поэтому, если нужно найти потенциальное значение любого из трех показателей (точности, помехоустойчивости и пропускной способности или быстродействия), то придется фиксировать значения остальных двух показателей и в результате будет найдена диаграмма об-

мена между всеми тремя показателями. Таким образом, как правило, отыскивая одномерную потенциальную характеристику (потенциальную точность, потенциальную помехоустойчивость, потенциальную пропускную способность), мы будем находить тем самым трехмерную потенциальную характеристику — диаграмму обмена между показателями точности, помехоустойчивости и пропускной способности.

Пример 4. В гл. 7 было получено выражение (7.113) для потенциального (минимально допустимого) значения B_n стоимости радиолинии передачи информации. При идеальной (в смысле К. Шеннона) обработке информации в этом выражении следует полагать $\xi=1$, а $\eta=P_0/N_0$ — требуемое отношение сигнала к шуму, определяемое при заданных значениях Δf и C формулой Шеннона (4.18). Поэтому выражение (7.113) можно записать в виде

$$k_1 = k_2 = k_3 = \sqrt[4]{P_0 k_2} \quad (10.21)$$

где $P_0 k_2$ определяется из соотношения

$$\log_2(1 + P_0 k_2 / \Delta f) = 1 / \Delta f k_2 \quad (10.22)$$

Здесь

$$k_1 = B \quad (10.23)$$

— стоимость радиолинии (точнее, ее варьируемой части);

$$k_2 = 1/N_0 \text{ — показатель помехоустойчивости системы;}$$

$$k_3 = 1/C; \quad (10.24)$$

$$g = 4 \sqrt[4]{c_1 c_2 c_3 c_4 d^2} L_1 L_2 k \quad (10.25)$$

— некоторая константа, не зависящая от показателей качества k_1 , k_2 и k_3 . Выражение (10.21) было получено минимизацией показателя k_1 при фиксированных значениях показателей k_2 и k_3 . Предоставляем читателю самому доказать, что это выражение является по существу диаграммой обмена между показателями k_1 , k_2 , k_3 , т. е. между показателями стоимости, помехоустойчивости и быстродействия (пропускной способности).

Если рассматривать обработку информации, оптимальную не в смысле К. Шеннона (т. е. с идеальным кодированием), а более близкую к реальной, в диаграмму обмена войдет также и показатель точности (например, величина $P_{\text{ош}}$ или $\overline{\delta^2}$).

10.4. МЕТОДЫ УМЕНЬШЕНИЯ РАЗМЕРНОСТИ ДИАГРАММЫ ОБМЕНА

Число m показателей качества $k_1, \dots, k_m$, которые необходимо учитывать в процессе проектирования, обычно весьма велико: несколько десятков или больше. Найти диаграмму обмена для столь большого числа показателей весьма сложно или практически невозможно. Если даже диаграмму обмена удастся найти, то при большом числе m ее сложно и неудобно использовать для принятия на ее основе каких-либо решений (в процессе проектирования или при научных исследованиях), а именно для этого и предназначена в конечном счете всякая диаграмма обмена. Например, для графического изображения диаграммы обмена

при $m=3$ требуется построить поверхность в трехмерном пространстве (k_1, k_2, k_3) или семейство большого числа кривых на плоскости; при $m=4$ потребуется построить уже больший набор таких семейств кривых; при $m \geq 4$ можно графически изобразить лишь отдельные участки (сечения) диаграммы обмена. Поэтому весьма актуальным является выявление методов уменьшения размерности диаграммы обмена на этапе отыскания и (или) применения этих диаграмм. Рассмотрим несколько таких методов.

Игнорирование второстепенных показателей. Поскольку диаграмма обмена применяется лишь на начальных этапах проектирования (или при научных исследованиях), при ее отыскании и использовании допустимо, как уже отмечалось, учитывать лишь важнейшие показатели качества и полностью пренебречь всеми остальными (с тем, чтобы учесть их лишь на последующих этапах проектирования). Этот метод является наиболее радикальным, так как позволяет полностью исключить из рассмотрения большую часть показателей качества как при отыскании диаграммы обмена, так и при ее использовании. Однако с этой радикальностью связан и весьма существенный недостаток метода — при его применении отброшенные (игнорированные) показатели качества ни в какой мере не учитываются при построении и использовании диаграммы обмена. Поэтому отбрасывать можно лишь действительно второстепенные показатели и число показателей m , которые оказывается необходимым учитывать, часто остается значительно большим единицы. Поэтому обычно необходимо дополнять этот метод другими, позволяющими еще в большей мере уменьшить размерность диаграммы обмена.

Использование приведенных показателей качества. Этот метод был уже рассмотрен в § 8.2. При этом отмечалось, что его применение позволяет сократить число показателей качества с нескольких десятков или сотен до нескольких единиц. Однако его применение связано с необходимостью обоснования зависимости каждого приведенного показателя от заменяемых им элементарных показателей, что в большинстве случаев может быть сделано лишь весьма приближенно. Поэтому во многих задачах даже после применения рассмотренных методов число показателей качества [элементарных и (или) приведенных], которые в дальнейшем необходимо учитывать, остается еще значительно большим единицы. Будем далее обозначать эти показатели качества $k_1, \dots, k_m$ и называть их первичными. Описываемые далее методы предназначены для уменьшения их числа.

Перевод ряда первичных показателей в разряд ограничений типа равенств. Пусть из состава $k_1, \dots, k_m$ первичных показателей m_1 показателей переведены в разряд ограничений типа равенств, т. е. принято, что

$$K = \langle k_1, \dots, k_{m'}; \underline{k_{m'+1}}, \dots, \underline{k_m} \rangle, \quad (10.26)$$

$$\text{где } m' = m - m_1. \quad (10.27)$$

При этом ищется диаграмма обмена

$$k_1 = f_{\max}(k_1, \dots, k_{m'}; \underline{k_{m'+1}}, \dots, \underline{k_m}) \quad (10.28)$$

лишь между оставшимися m' показателями, а зафиксированные показатели $k_{m'+1}, \dots, k_m$ играют при этом роль некоторых параметров, от которых зависит вид диаграммы. Если показатели $k_{m'+1}, \dots, k_m$ зафиксированы жестко, т. е. приняты равными некоторым вполне определенным численным значениям, то может упроститься не только отыскание диаграммы, но и ее интерпретация: достаточно анализировать (графически или аналитически) зависимость лишь между m' , а не m показателями качества. Однако на начальных этапах проектирования, и особенно при научных исследованиях, часто нельзя фиксировать значения $k_{m'+1}, \dots, k_m$ жестко — необходимо искать зависимость (10.28) в достаточно широком диапазоне изменения «замороженных» показателей $k_{m'+1}, \dots, k_m$. Но это равносильно тому, что приходится учитывать взаимную зависимость между всеми m первичными показателями, т. е. упрощения интерпретации получаемой диаграммы обмена не получается. Как следует из изложенного в § 9.2 комбинированного метода отыскания нехудших систем, если зависимость (10.28) оказывается по каждому из аргументов $k_{m'+1}, \dots, k_m$ монотонно убывающей, то она совпадает с m -мерной (а не m' -мерной) диаграммой обмена. Таким образом, в большинстве случаев этот метод полезен для отыскания m -мерной диаграммы обмена, но не позволяет облегчить ее интерпретацию.

Перевод ряда первичных показателей в разряд ограничений типа неравенств. При этом методе ищется множество систем, нехудших в смысле вектора

$$K = \langle k_1, \dots, k_{m'} \rangle,$$

при наличии, кроме прочих, ограничений вида

$$k_{m'+1} \leq k_{(m'+1) \max}, \dots, k_m \leq k_{m \max}.$$

Если при этом максимально допустимые значения $k_{(m'+1) \max}, \dots, k_{m \max}$ показателей $k_{m'+1}, \dots, k_m$ приняты равными вполне определенным численным значениям, то перевод $(m-m')$ показателей в разряд ограничений типа неравенств позволяет существенно упростить интерпретацию полученной (по m' показателям) диаграммы обмена, а иногда упростить и отыскание диаграммы обмена. Однако на начальных этапах проектирования, и особенно при научных исследованиях, в ряде задач нежелательно столь жестко фиксировать максимально допустимые значения показателей $k_{m'+1}, \dots, k_m$. В таких случаях найденная (по m' показателям качества) диаграмма обмена окажется зависящей от $(m-m')$ параметров $k_{m'+1}, \dots, k_m$ и, следовательно, ее интерпретация существенно не упростится (по сравнению с интерпретацией m -мерной диаграммы обмена).

Введение обобщенных показателей качества. Этот метод применяется только для упрощения интерпретации уже найденной m -мерной диаграммы обмена путем преобразования ее в обобщенную диаграмму обмена меньшей размерности (см. § 10.5).

10.5. УМЕНЬШЕНИЕ РАЗМЕРНОСТИ ДИАГРАММЫ ОБМЕНА ПУТЕМ ЕЕ ПРЕОБРАЗОВАНИЯ В ОБОБЩЕННУЮ ДИАГРАММУ ОБМЕНА

Для простоты начнем рассмотрение этого метода со следующего конкретного примера. Выражение (4.18) для пропускной способности радиолинии передачи информации можно рассматривать как диаграмму обмена между следующими (стандартными) показателями качества:

$$k_1 = 1/C, \quad k_2 = \Delta f, \quad k_3 = P_c, \quad k_4 = 1/N_0. \quad (10.29)$$

Действительно, в рассмотренном (§ 10.3) примере 3 полоса Δf и средняя мощность P_c сигнала рассматривались как некоторые фиксированные параметры, но по существу они являются стандартными показателями качества, а зависимость k_1 от k_2, k_3 и k_4 — монотонно убывающая и, следовательно, совпадает с диаграммой обмена вида

$$k_1 = 1/k_2 \log_2(1 + k_3 k_4/k_2). \quad (10.30)$$

Диаграмма обмена (10.30) — четырехмерная и неудобна для интерпретации. В § 4.2 было показано, что формулу К. Шеннона (4.18), а следовательно, и зависимость (10.30) можно свести к двумерной зависимости вида (4.19) введением обобщенных показателей качества

$$\beta_K = P_c/N_0 C, \quad \beta_{\Delta f} = \Delta f/C. \quad (10.31)$$

С учетом обозначений (10.29) и (10.31) зависимость (4.18), а следовательно, и (10.30) сводятся к следующему виду:

$$K_{0.61} = K_{0.62} \{ \exp [(\ln 2)/K_{0.62}] - 1 \}, \quad (10.32)$$

где

$$K_{0.61} = \beta_K = P_c/N_0 C = k_1 k_3 k_4; \quad K_{0.62} = \beta_{\Delta f} = \Delta f/C = k_1 k_2. \quad (10.33)$$

Зависимость (10.32) была изображена на рис. 4.2 (сплошной линией) и повторена на рис. 10.1. Рассмотрим, какими свойствами она обладает. Нетрудно убедиться, что путем подстановки в (10.32) вместо обобщенных показателей $K_{0.61}$ и $K_{0.62}$ их выражений через первичные показатели k_1, k_2, k_3 и k_4 [в соответствии с (10.33)] и последующих однозначных преобразований можно получить первичную диаграмму обмена (10.30). Следовательно, при переходе от первичной диаграммы (10.30) к обобщенной (10.32) никакой потери информации о первичной диаграмме не происходит. Все нехудшие (в смысле первичных показателей $k_1, \dots, k_n$) точки удовлетворяют как уравнению (10.30), так и (10.32) и, следовательно, отображаются в соответствующие точки обобщенной диаграммы (10.32).

Но при этом каждой точке обобщенной диаграммы соответствует не одна, а некоторое множество точек первичной диаграммы [например, точке А обобщенной диаграммы с координатами $(K'_{0.61}, K'_{0.62})$ соответствуют любые точки пер-

вичной диаграммы, удовлетворяющие равенствам $k_1 k_3 k_4 = K'_{0.61}$, $k_1 k_2 = K'_{0.62}$]. Любая худшая (в смысле первичных показателей) точка не удовлетворяет соотношению (10.30) и соответственно не удовлетворяет и (10.32), т. е. в пространстве обобщенных показателей $(K_{0.61}, K_{0.62})$ (рис. 10.1) расположена вне обобщенной диаграммы и притом всегда левее и выше ее (например, точка В на рис. 10.1). В этом нетрудно убедиться, если учесть свойства худших точек. Таким образом, все нехудшие (в смысле первичных показателей) точки расположены на обобщенной диаграмме обмена, а все худшие — правее и выше нее. Но для каждой худшей точки (например, точки В) на обобщенной диаграмме обмена найдутся две точки (B' и B'') безусловно лучшие в смысле обобщенных показателей $K_{0.61}$, $K_{0.62}$. Следовательно, все точки, худшие в смысле первичных показателей, оказываются худшими и в смысле обобщенных показателей.

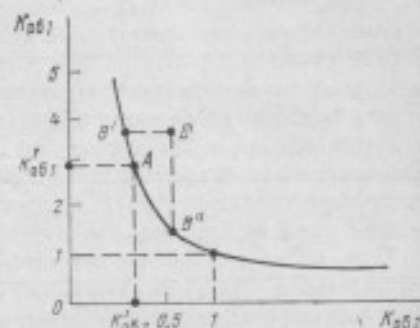


Рис. 10.1

Аналогичными рассуждениями нетрудно убедиться, что все точки, нехудшие в смысле первичных показателей, оказываются нехудшими и в смысле обобщенных показателей. Поэтому зависимость вида (10.32), т. е. свернутое отображение первичной диаграммы обмена, является в то же время диаграммой обмена между обобщенными показателями $K_{0.61}$, $K_{0.62}$ и, следовательно, обладает в отношении этих показателей свойством двукратного минимума: каждой точке этой диаграммы (например, точке А) соответствует минимальное (по всем допустимым системам) значение показателя $K_{0.61}$ при данном (соответствующем этой точке) значении показателя $K_{0.62}$ и минимальное значение показателя $K_{0.62}$ при данном (соответствующем этой точке) значении показателя $K_{0.61}$. Очевидно, что двумерная зависимость (10.32), изображенная на рис. 10.1, значительно удобнее для использования, чем исходная четырехмерная диаграмма обмена (10.22), которую можно было бы наиболее просто изобразить графически лишь в виде набора семейств кривых.

Построение обобщенной диаграммы обмена было рассмотрено на конкретном примере преобразования четырехмерной диаграммы обмена радиолинии передачи информации, полученной К. Шенноном, в обобщенную двумерную диаграмму. Вообще говоря, может быть полезным сокращение размерности (при $m > 2$) любой первичной диаграммы обмена и необязательно путем ее сведения к двумерной (если $m > 3$), но наибольшее упрощение достигается при преобразовании исходной m -мерной диаграммы

$$k_1 = f_{\text{пх}}(k_2, \dots, k_m) \quad (10.34)$$

в двумерную обобщенную

$$K_{0.61} = f_{0.6}(K_{0.62}). \quad (10.35)$$

Однако такое преобразование возможно не всегда¹, а лишь при выполнении следующей совокупности условий.

1. Зависимости

$$K_{об1} = f_1(k_1, \dots, k_l), \quad (10.36)$$

$$K_{об2} = f_2(k_1, \dots, k_l) \quad (10.37)$$

Вводимых обобщенных показателей от первичных должны обладать следующими свойствами:

каждый из m первичных показателей $k_1, \dots, k_m$ должен быть аргументом хотя бы одного из обобщенных показателей;

состав первичных показателей, входящих в (10.36) и (10.37), должен быть не идентичным (т. е. составы должны различаться хотя бы на один показатель);

функции (10.36) и (10.37) должны быть монотонно возрастающими по каждому из их аргументов.

2. Вводимые обобщенные показатели (10.36), (10.37) должны обеспечивать преобразование первичной диаграммы (10.34) в (10.35) и однозначное обратное преобразование обобщенной диаграммы (10.35) в первичную (10.34). При этом зависимость (10.35) должна быть монотонно убывающей.

Легко убедиться в том, что в рассмотренном примере введенные обобщенные показатели (10.33) удовлетворяют всем сформулированным условиям.

Нетрудно доказать, что при выполнении этих условий обобщенная диаграмма (10.35) обладает всеми общими свойствами, доказанными применительно к диаграмме К. Шеннона вида (10.30).

10.6. СВЯЗЬ ДИАГРАММЫ ОБМЕНА СИСТЕМЫ С ДИАГРАММАМИ ОБМЕНА ЕЕ ПОДСИСТЕМ

Пусть система S должна удовлетворять исходным данным D'' и характеризуется вектором качества $K = \langle k_1, \dots, k_m \rangle$. Требуется найти множество нехудших систем M_{nx} , удовлетворяющих этим исходным данным, и соответствующую им диаграмму обмена

$$k_1 = f_{nx}(k_2, \dots, k_m). \quad (10.38)$$

Решить эту задачу, не разбивая систему S на подсистемы, часто слишком сложно или даже практически невозможно. Поэтому рассмотрим случаи, когда для нахождения множества M_{nx} и диаграммы обмена (10.38) приходится разбивать систему S на подсистемы. Пусть, например, приходится разбивать систему S на две подсистемы S' и S'' . Для этого на основе исходных данных D'' необходимо сформулировать исходные данные D''_1 и D''_2 для каждой

подсистемы S' и S'' соответственно, в частности обосновать состав векторов качества $K' = \langle k'_1, \dots, k'_m \rangle$ и $K'' = \langle k''_1, \dots, k''_m \rangle$ этих подсистем. Предположим, что такая декомпозиция произведена. Тогда для подсистемы S' можно найти множество M'_{nx} подсистем, нехудших с точки зрения исходных данных D''_1 . Соответственно для подсистемы S'' можно найти множество M''_{nx} подсистем, нехудших для исходных данных D''_2 .

Возникает вопрос, в какой мере найденное решение (структуры подсистем и значения их параметров) можно считать нехудшими, а отброшенные при этом худшие решения действительно худшими с точки зрения системы в целом. Для ответа на него достаточно рассмотреть значение (ценность) множеств M'_{nx} и M''_{nx} нехудших и худших решений для подсистемы S' , найденных на основе исходных данных D''_1 . Прежде всего очевидно, что не всякое решение, нехудшее с точки зрения исходных данных D''_1 , будет нехудшим и с точки зрения исходных данных D'' для системы в целом. Для доказательства этого положения рассмотрим простейший случай, когда система S характеризуется скалярным показателем качества, т. е. $K = k_1$ ($m=1$). Тогда система, нехудшая в смысле вектора K , есть не что иное, как система, обладающая минимальным значением показателя k_1 , и эта система является не только нехудшей, но и наилучшей (оптимальной).

Пусть качество подсистемы S' по-прежнему характеризуется вектором $K' = \langle k'_1, \dots, k'_m \rangle$. При этом различным нехудшим подсистемам (решениям) $S' \in M'_{nx}$ будут соответствовать в общем случае различные значения как вектора K' , так и показателя качества $k = k_1$ системы в целом. Но это означает, что среди нехудших подсистем (решений) $S' \in M'_{nx}$ будут и такие, которые дают не наименьшее возможное значение показателя k_1 , т. е. являются с точки зрения системы S в целом безусловно худшими. Итак, в общем случае лишь часть нехудших подсистем $S' \in M'_{nx}$ войдет в состав нехудших систем $S \in M_{nx}$; остальные же подсистемы $S' \in M'_{nx}$ будут признаны с точки зрения системы S в целом (т. е. исходных данных D'') худшими и, следовательно, исключены из дальнейшего рассмотрения.

Выясним теперь следующий вопрос. Основным смыслом отыскания множества M'_{nx} нехудших подсистем S' состоит в том, чтобы исключить из дальнейшего рассмотрения множество систем M'_{xh} , худших в смысле исходных данных D''_1 ; но можно ли считать, что подсистема (вариант решения) S' , признанная худшей при исходных данных D''_1 , останется худшей и с точки зрения системы S в целом (т. е. исходных данных D''). Иначе говоря, не отбросим ли мы при этом варианты, которые с точки зрения системы S в целом отнюдь не являются худшими и в конечном счете могут оказаться даже наилучшими (оптимальными).

Очевидно, ответ на этот вопрос должен быть следующим. Если подсистема S' , признанная с точки зрения исходных данных D''_1 худшей, может оказаться с точки зрения системы в целом не-

¹ Вообще говоря, всегда можно преобразовать диаграмму (10.34), в двумерную (10.35), если ввести следующие обобщенные показатели качества: $K_{об1} = k_1$, $K_{об2} = 1/f_{nx}(k_2, \dots, k_m)$. При этом в соответствии с (10.34) получается $K_{об2} = 1/K_{об1}$, т. е. любую m -мерную диаграмму обмена можно преобразовать в двумерную вида гиперболы. Однако, поскольку вид полученной таким путем обобщенной диаграммы оказывается идентичным для любых первичных диаграмм, рассмотрение этой диаграммы не может дать никакой полезной информации — она является в этом смысле тривиальной. В данном же параграфе рассматриваются лишь нетривиальные преобразования, но их не всегда можно реализовать, т. е. не при любой форме первичной диаграммы.

худшей, то это означает, что разбиение системы S на подсистемы S' и S'' (соответственно разбиение исходных данных D'' на D''_1 и D''_2) произведено не вполне правильно или даже совершенно неправильно, ибо смысл декомпозиции S на подсистемы S' и S'' (соответственно разбиение исходных данных D'' на D''_1 и D''_2) как раз в том и состоит, чтобы в дальнейшем проектировать (оптимизировать) каждую подсистему независимо, не интересуясь ее влиянием на другие подсистемы и систему в целом.

Соответственно декомпозицию системы S на подсистемы S' и S'' (исходных данных D'' на D''_1 и D''_2) можно считать вполне правильной (совершенной) лишь в том случае, если решения (варианты) относительно S' , признанные на основе исходных данных D''_1 худшими, будут худшими и с точки зрения системы в целом. Для реализации совершенной декомпозиции необходимо так разбить исходные данные D'' для системы S на данные D''_1 и D''_2 для подсистем S' и S'' , чтобы они полностью и точно учитывали все виды связи подсистемы S' с системой в целом (включая связи с остальными подсистемами системы S). Но эти связи, вообще говоря, бесконечно разнообразны; поэтому учесть их полностью и точно принципиально невозможно, да и не нужно, ибо исходные данные D'' для системы S в целом также можно обосновать лишь неполностью и приближенно. Поэтому практически можно считать декомпозицию совершенной, если она правильно и с приемлемой точностью учитывает все основные (главные) виды связей подсистемы с системой в целом.

К ним относятся в первую очередь динамические, конструктивные и экономические связи. Под динамическими понимаются связи между процессами, протекающими в различных частях системы, в частности обмен полезной информацией и помехами. Конструктивные связи обусловлены тем, что все или часть подсистем должны объединяться в общую конструкцию с вполне определенными массой, объемом и габаритными размерами. Экономические связи вызваны тем, что экономические (стоимостные) затраты на систему в целом ограничены и зависят от соответствующих затрат на отдельные подсистемы. На ранних этапах проектирования, особенно на стадии научных исследований, учесть достаточно полно и точно связи подсистемы с системой в целом обычно не удастся. Например, часто не учитываются конструктивные и экономические связи и часть динамических. При такой несовершенной декомпозиции варианты подсистемы S' , признанные с точки зрения (неточных) исходных данных D''_1 худшими, могут оказаться с точки зрения системы в целом (данных D'') нехудшими или даже наилучшими (оптимальными).

В качестве иллюстрации рассмотрим задачу оптимизации системы передачи информации, решенную К. Шенноном (§ 10.3). К. Шеннон, как уже отмечалось, нашел диаграмму обмена между четырьмя показателями качества $1/C$, $1/N_0$, Δf и P_c и соответствующее этой диаграмме множество $M_{\text{нх}}$ нехудших систем. Каждая из этих систем является нехудшей в том смысле, что обеспечива-

ет минимум каждого из этих четырех показателей качества при любых фиксированных значениях остальных трех показателей. Однако при этом Шеннон не принимал во внимание такие важные показатели качества, как время задержки передачи информации, стоимость, надежность, масса системы и др.

Сравним с этими нехудшими системами некоторую реальную систему, которая имеет несколько (например, на 30%) худшее значение всех учтенных Шенноном показателей качества, но конечное время задержки, приемлемую массу, надежность, стоимость и другие практически важные показатели. С точки зрения совокупности показателей, принятых Шенноном, эта система будет худшей, тогда как в действительности, т. е. с учетом более полных и точных данных, она может оказаться нехудшей и даже наилучшей (оптимальной). Однако это не лишает результаты, полученные Шенноном, всякой ценности, так как дополнительные факторы (уточнение исходных данных) могут привести лишь к ухудшению значений учтенных Шенноном показателей качества. Поэтому во всяком случае результаты Шеннона можно рассматривать как нижние грани потенциальных значений принятых им (стандартных) показателей качества.

Из изложенного можно сделать следующие выводы:

1. Если исходные данные D''_1 для подсистемы S' системы S достаточно полны и корректны (совершенны), то варианты подсистемы S' , признанные при этих данных худшими, будут худшими и с точки зрения системы S' в целом. Поэтому с точки зрения системы в целом представляют интерес лишь варианты подсистемы S' , принадлежащие множеству $M'_{\text{нх}}$ систем, нехудших в смысле данных D''_1 (в частности, в смысле вектора K'). При этом в состав систем S , нехудших в смысле исходных данных D'' (вектора K), т. е. в состав множества $M_{\text{нх}}$, войдут не все нехудшие варианты $S' \in M'_{\text{нх}}$, а лишь часть из них.

2. Если исходные данные D''_1 для подсистемы S' недостаточно полны и корректны, система S' , признанная, исходя из D''_1 (K'), худшей, может оказаться с точки зрения системы в целом нехудшей или даже оптимальной. Поэтому в таких случаях на более поздних этапах проектирования (т. е. при более полном и точном учете связей подсистемы в целом) может потребоваться корректировка (пересмотр) решений, принятых на более ранних этапах.

3. На самых ранних этапах проектирования, особенно на стадии научных исследований, исходные данные для синтезируемой системы (подсистемы) обычно весьма неполны, т. е. имеет место случай, отмеченный в п. 2. При этом основная ценность отыскания множества нехудших вариантов (подсистем) состоит не в возможности отбрасывания худших вариантов построения системы (подсистемы), поскольку такое отбрасывание может оказаться неправильным, а в нахождении нижней грани для потенциальных значений учитываемых показателей качества. Однако для уверенности в том, что найдены действительно нижние грани,

нужно следить за тем, чтобы идеализации (упрощения), осуществляемые при обосновании исходных данных D''_1 для отыскания данной диаграммы обмена, улучшали, а не ухудшали соответствующие показатели качества (диаграмму обмена).

Для иллюстрации изложенного рассмотрим простой (гипотетический) пример. Системой S является зенитный реактивный снаряд, управляемый установленной на нем системой самонаведения. Его основными показателями качества являются вероятность $P_{\text{нз}}$ непоражения цели и стоимость B , т. е.

$$k_1 = P_{\text{нз}} = 1 - P_{\text{п}}, \quad k_2 = B, \quad (10.39)$$

где $P_{\text{п}}$ — вероятность поражения цели одним снарядом.

Требуется найти диаграмму обмена между потенциальными значениями этих показателей $k_1 = f_{\text{нз}}(k_2)$.

Систему S можно разбить на две подсистемы: S' — систему самонаведения и S'' — остальную часть снаряда, содержащую, кроме прочего, неконтактный взрыватель [22]. Основными показателями качества подсистемы S' можно считать средний квадрат $R^2_{\text{сф}}$ промаха снаряда и стоимость B_T системы управления, т. е.

$$k'_1 = R^2_{\text{сф}}, \quad k'_2 = B_T. \quad (10.40)$$

Будем полагать действие взрывателя идеальным. Тогда главными показателями подсистемы S'' являются средний квадрат $R^2_{\text{сф}}$ радиуса поражающего действия снаряда и стоимость $B_{\text{св}}$ остальной части снаряда, т. е.

$$k''_1 = 1/R^2_{\text{сф}}, \quad k''_2 = B_{\text{св}}. \quad (10.41)$$

Очевидно,

$$B_T + B_{\text{св}} = B, \quad (10.42)$$

т. е.

$$k'_2 + k''_2 = k_2. \quad (10.43)$$

Кроме того, как показано в [22],

$$P_{\text{п}} = 1/(1 + R^2_{\text{сф}}/R^2_{\text{сф}}), \quad (10.44)$$

и, следовательно,

$$P_{\text{нз}} = 1/(1 + R^2_{\text{сф}}/k^2_{\text{сф}}), \quad (10.45)$$

т. е.

$$k_1 = 1/(1 + 1/k'_1 k''_1). \quad (10.46)$$

Итак, известны соотношения (10.43) и (10.45), связывающие показатели качества $\langle k_1, k_2 \rangle$ системы с показателями $\langle k'_1, k'_2 \rangle$ и $\langle k''_1, k''_2 \rangle$ ее подсистем.

Пусть для каждой из подсистем найдем диаграммы обмена (рис. 10.2)

$$\left. \begin{aligned} k'_1 &= f_{\text{нз}1}(k'_2), \text{ т. е. } R^2_{\text{сф}} = f_{\text{нз}1}(B_T) \\ \text{и} \\ k''_1 &= f_{\text{нз}2}(k''_2), \text{ т. е. } 1/R^2_{\text{сф}} = f_{\text{нз}2}(B_{\text{св}}). \end{aligned} \right\} \quad (10.46)$$

Этим диаграммам обмена соответствуют множества $M'_{\text{нз}}$ и $M''_{\text{нз}}$ нехудших систем. Будем сначала полагать, что разбиение системы S на подсистемы S' и S'' является практически совершенным, т. е. совокупность выбранных показателей качества подсистем $\langle k'_1, k'_2, k''_1, k''_2 \rangle$ достаточно полно характеризует связи между подсистемами и их роль в выполнении функций системы

в целом. Тогда, как уже отмечалось, варианты подсистем, худшие в смысле их показателей качества (т. е. на входящие в множества $M'_{\text{нз}}$ и $M''_{\text{нз}}$), можно исключить из рассмотрения при отыскании множества $M_{\text{нз}}$ нехудших вариантов построения системы в целом и соответственно искомой диаграммы обмена между показателями k_1 и k_2 системы. Отсюда следует, что множество

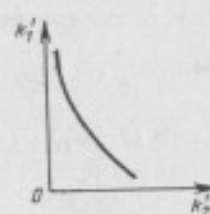


Рис. 10.2

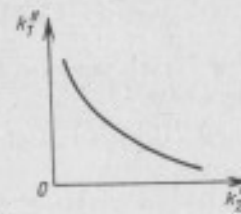


Рис. 10.3

$M_{\text{нз}}$ нехудших систем можно найти на основе рассмотрения лишь множеств $M'_{\text{нз}}$ и $M''_{\text{нз}}$ нехудших подсистем. Применим для нахождения $M_{\text{нз}}$ метод рабочих характеристик, т. е. найдем минимум величин k_1 при условии, что k_2 имеет фиксированное (но произвольное) значение, т. е. решим задачу

$$\text{обеспечить } \min_{S \in M_S} k_1 \text{ при } k_2 = \text{const.} \quad (10.47)$$

Здесь M_S — множество систем, которые могут быть образованы путем различных комбинаций нехудших подсистем S' и S'' .

Но для нехудших подсистем справедливы соотношения (10.46) между их показателями качества. Поэтому с учетом соотношений (10.43), (10.45) и (10.46) задача (10.47) принимает вид:

$$\left. \begin{aligned} \text{найти минимум величин } k_1 &= 1/(1 + 1/k'_1 k''_1) \\ \text{при условии } k'_2 + k''_2 &= k_2, \end{aligned} \right\} \quad (10.48)$$

где $k'_1 = f_{\text{нз}1}(k'_2)$, $k''_1 = f_{\text{нз}2}(k''_2)$.

Для решения этой задачи удобно аппроксимировать диаграммы обмена подсистем полиномами

$$k'_1 = a_1 (k'_2)^{-\gamma_1}, \quad k''_1 = a_2 (k''_2)^{-\gamma_2}, \quad (10.49)$$

где a_1 , γ_1 и a_2 , γ_2 — коэффициенты, найденные при аппроксимации зависимостей на рис. 10.2 полиномами (10.49). Подстановка соотношений (10.49) в (10.48) приводит задачу к следующему виду:

найти безусловный минимум функции

$$k_1 = 1/[1 + (a_1 a_2)^{-1} (k'_2)^{\gamma_1} (k_2 - k'_2)^{\gamma_2}] \quad (10.50)$$

при вариации ее аргумента k'_2 (здесь $k'_2 < k_2$).

Минимум находится из условия $dk_1/dk'_2 = 0$, откуда

$$k'_2 = k_2 / (1 + \gamma_2 / \gamma_1). \quad (10.51)$$

Подставляя это значение в (10.50), получаем

$$k_{1\text{нз}} = 1/[1 + (a_1 a_2)^{-1} (1 + \gamma_2 / \gamma_1)^{-\gamma_1} (1 + \gamma_1 / \gamma_2)^{-\gamma_2} k_2^{\gamma_1 + \gamma_2}]. \quad (10.52)$$

Это уравнение является рабочей характеристикой системы в целом. Но зависимость (10.52) — монотонно убывающая (с ростом k_2 значение k_1 или убывает) и, следовательно, совпадает с уравнением диаграммы обмена между показателями k_1 и k_2 системы. Поэтому из (10.52) следует, что при сделанных допущениях диаграмма обмена для системы в целом имеет вид

$$k_1 = 1/[1 + (a_1 a_2)^{-1} (1 + \gamma_2/\gamma_1)^{-\gamma_1} (1 + \gamma_1/\gamma_2)^{-\gamma_2} k_2^{\gamma_1 + \gamma_2}]. \quad (10.53)$$

При этом показатель качества k'_2 подсистемы S' должен удовлетворять условию (10.51).

Возвращаясь к первоначальным обозначениям, эти результаты можно записать в следующем виде:

$$P_{\text{вн}} = 1/[1 + (a_1 a_2)^{-1} (1 + \gamma_2/\gamma_1)^{-\gamma_1} (1 + \gamma_1/\gamma_2)^{-\gamma_2} B^{\gamma_1 + \gamma_2}], \quad (10.54)$$

причем должно быть

$$\left. \begin{aligned} B_7 &= B/(1 + \gamma_2/\gamma_1) \\ \text{и, следовательно,} \\ B_{07} &= B - B_7 = B/(1 + \gamma_1/\gamma_2). \end{aligned} \right\} \quad (10.55)$$

Если $\gamma_1 = \gamma_2$, то

$$P_{\text{вн}} = 1/[1 + (a_1 a_2)^{-1} (B/2)^{2\gamma_1}], \text{ причем } B_7 = B_{07} = B/2. \quad (10.56)$$

Следовательно, при $\gamma_1 = \gamma_2$ оптимальным будет такое распределение затрат B на систему в целом, при котором половина их расходуется на систему управления и столько же на остальную часть снаряда. Если $\gamma_1 > \gamma_2$, то из (10.55) получается $B_7 > B/2$ и $B_{07} < B/2$. Если же $\gamma_1 < \gamma_2$, то $B_7 < B/2$, а $B_{07} > B/2$.

Из этого примера видно, что не любые комбинации из нехудших вариантов подсистем управления и остальной части снаряда приводят к нехудшим вариантам для системы в целом, а лишь такие, при которых выполняется условие (10.51); остальные комбинации дают худшие (а не нехудшие) решения. Тем самым пример подтверждает общее утверждение о том, что не любой нехудший вариант подсистемы является нехудшим и с точки зрения системы в целом.

Худшие варианты построения подсистем были исключены из рассмотрения на основании предположения, что совокупность показателей ($k_{\text{вн}}$, $1/R_{\text{вн}}$, B_7 , B_{07}) полностью учитывает связи между подсистемами и их роль в системе в целом. Однако это предположение грубое. Действительно, в принятой совокупности показателей не принимались во внимание такие важные для бортовых подсистем показатели, как масса и габаритные размеры. С их учетом полученные результаты могут потребовать корректировки, поскольку без этого некоторые варианты построения подсистем можно ошибочно отнести к худшим, хотя на самом деле они нехудшие (так как могут иметь меньшее значение хотя бы одного из дополнительных показателей качества: массы или габаритных размеров). Однако, несмотря на то, что полученные результаты (10.54) — (10.56) могут потребовать на последующих этапах проектирования корректировки, они являются полезными, так как представляют собой нижние грани для всех точек диаграммы обмена между $P_{\text{вн}}$ и B : в результате корректировки вероятность $P_{\text{вн}}$ непоражения цели при каждом данном значении стоимости B снаряда может лишь увеличиться и соответственно возрастут затраты B , необходимые для достижения того или иного значения вероятности $P_{\text{вн}}$. Иначе говоря, при корректных идеализациях вместо истинной диаграммы

обмена находится ее нижняя граница (в рассмотренном примере двух показателей качества — ее левая нижняя граница).

Следует отметить, что задача декомпозиции системы на подсистемы, возникающая как при отыскании диаграмм обмена, так и вообще в процессе проектирования, весьма сложна и нередко не имеет достаточно точного и однозначного решения, особенно на ранних этапах проектирования. Более того, в ряде случаев при определении некоторых показателей качества или диаграмм обмена системы приходится вообще отказываться от декомпозиции и находить их сразу для системы в целом.

Ответы на ряд следующих кардинальных вопросов, возникающих при декомпозиции: можно ли с приемлемой точностью осуществить декомпозицию и что при этом понимать под приемлемой точностью; если декомпозиция возможна, то является ли она целесообразной, т. е. упростит или ускорит процесс проектирования; какими именно показателями качества следует оценивать каждую подсистему и какие связи между подсистемами необходимо учитывать — часто настолько сложны, особенно на ранних этапах проектирования, что могут быть даны в значительной мере лишь на основе опыта и инженерной интуиции коллективов разработчиков. В связи с этим приходится корректировать результаты, полученные путем декомпозиции, в процессе перехода от начального этапа проектирования к конечному.

Из приведенного рассмотрения вытекают следующие основные принципы получения и применения потенциальных характеристик радиосистем.

1. Потенциальные характеристики радиосистемы (радиоподсистемы) могут быть одномерными или многомерными. Многомерная потенциальная характеристика — это диаграмма обмена между показателями качества $k_1, \dots, k_m$ системы. В пространстве R_m этих показателей она является геометрическим местом множества $M_{\text{нх}}$ всех нехудших точек, соответствующих множеству нехудших систем (вариантов их построения). Каждая точка диаграммы обмена дает потенциальное (наилучшее достижимое при сформулированных исходных данных D'') значение любого из m показателей качества k_i ($i = \overline{1, m}$) при фиксированных (соответствующих этой точке) значениях остальных $(m-1)$ показателей. Поэтому диаграмма обмена показывает зависимость потенциального значения каждого из m показателей от значения остальных $(m-1)$ показателей. Одномерную потенциальную характеристику ($m=1$) можно рассматривать как вырожденную диаграмму обмена.

2. Определяемая диаграммой обмена зависимость $k_1 = f_{\text{нх}}(k_2, \dots, k_m)$ между показателями качества, является монотонно убывающей по каждому из аргументов. Если это уравнение решить относительно любого из аргументов $k_2, \dots, k_m$, например относительно k_2 , то полученная зависимость $k_2 = f'_{\text{нх}}(k_1, k_3, k_4, \dots, k_m)$ также будет монотонно убывающей по каждому из ее аргументов и представлять в пространстве R_m ту же самую диаграмму обмена.

3. Каждая диаграмма обмена вычисляется для вполне определенных (сформулированных математически) исходных данных D'' (8.13). Поэтому ее вид зависит от этих данных и число возможных видов диаграмм обмена неограничено велико.

4. При отыскании диаграмм обмена (потенциальных характеристик) идеализации, допускаемые в процессе составления (математического ожидания) исходных данных D'' , должны быть направлены лишь на облегчение (уменьшение состава учитываемых показателей качества, ослабление ограничений и условий работы системы); в противном случае полученные характеристики нельзя будет считать потенциальными (наилучшими теоретически достижимыми). Однако, с другой стороны, принимаемые идеализации не должны быть чрезмерными, так как в противном случае найденные потенциальные характеристики будут слишком далеки от реально достижимых.

5. Для отыскания диаграммы обмена можно применять любые методы определения множества $M_{\text{нх}}$ нехудших точек (систем), в частности метод рабочих характеристик, весовой или комбинированный.

6. При отыскании одномерной потенциальной характеристики, как правило, приходится переводить ряд (антагонистических) показателей качества в разряд ограничений типа равенств и решать задачу оптимизации в широком диапазоне значений фиксированных («замороженных») показателей. При этом находится по существу многомерная диаграмма обмена между всеми показателями качества, включая «замороженные». В частности, при определении потенциальной точности, или потенциальной помехоустойчивости, или потенциальной пропускной способности находится, как правило, диаграмма обмена между показателями точности, помехоустойчивости и пропускной способности.

7. Для облегчения отыскания диаграмм обмена и их интерпретации весьма желательно сокращать их размерность, а также применять такие методы, как игнорирование второстепенных показателей качества, использование приведенных показателей качества, перевод ряда показателей в разряд ограничений типа равенств или неравенств. Каждый из этих методов имеет свои преимущества и недостатки. В частности, перевод ряда показателей в разряд ограничений (типа равенств или неравенств) обычно облегчает лишь отыскание, но не интерпретацию диаграмм обмена.

Для удобства интерпретации получаемой m -мерной диаграммы желательно, если это возможно, вводить обобщенные показатели качества, преобразующие первичную m -мерную диаграмму обмена в обобщенную диаграмму обмена меньшей размерности, по возможности в двумерную. При этом следует соблюдать условия (§ 10.5), которым должны удовлетворять вводимые обобщенные показатели качества для преобразования первичной диаграммы обмена произвольной размерности в двумерную.

8. Диаграммы обмена подсистем можно использовать для построения диаграммы обмена системы лишь в том случае, если раз-

биение системы на подсистемы проведено достаточно корректно — при формулировке исходных данных D'' ; для каждой подсистемы S_i учитывались все основные виды ее связей с другими подсистемами и системой в целом. Если некоторые важные связи при этом не учитывались, но очевидно, что пренебрежение ими может привести лишь к завышению качества системы в целом (например, если не учитывались конструктивные или экономические связи), то найденная диаграмма обмена для системы в целом должна рассматриваться лишь как предел (в сторону улучшения) более близкой к реальной диаграммы обмена (между теми же показателями качества), которая получилась бы при меньших (менее грубых) идеализациях. Знание такого предела может быть полезно при научных исследованиях и на начальном этапе проектирования.

9. Диаграммы обмена желательно использовать на начальных этапах проектирования, и особенно при научных исследованиях, так как в этих случаях идеализации, принимаемые при их отыскании, не только допустимы, но и полезны — позволяют получить количественные оценки при минимальных априорных данных.

Основными областями применения диаграмм обмена являются: определение принципиальных возможностей улучшения показателей качества реальных систем;

облегчение установления путей совершенствования систем — выявление основных факторов, препятствующих улучшению систем, отыскание структуры и (или) значений параметров нехудших или оптимальных систем;

сравнение вариантов построения систем по их потенциальным характеристикам в тех случаях, когда определить реальные характеристики трудно или даже невозможно (из-за отсутствия на данном этапе проектирования или исследования необходимых априорных данных).

Литература: основная [58], дополнительная [14, 47].

Глава 11

ПРИМЕРЫ ОПТИМИЗАЦИИ РАДИОУСТРОЙСТВ И РАДИОСИСТЕМ

11.1. ОПТИМИЗАЦИЯ ОБНАРУЖИТЕЛЯ СИГНАЛА ПО ДВУМ И ПО ТРЕМ ПОКАЗАТЕЛЯМ КАЧЕСТВА

На вход обнаружителя (рис. 11.1) поступает сумма $y(t)$ независимых сигнала $u_c(t)$ и шума $u_{\text{ш}}(t)$ или один только шум, что можно записать в виде

$$y(t) = u_c(t, x) + u_{\text{ш}}(t) \quad (11.1)$$

где $u_c(t, x_1) = u_c(t)$; $u_c(t, x_2) = 0$. Обнаружитель, анализируя сумму $y(t)$ в течение заданного времени T , должен вынести решение γ о наличии или отсутствии сигнала, т. е. принять решение $\gamma = x_1$ (сигнал есть) или $\gamma = x_2$ (сигнала нет), причем $u_{ш}(t)$ — нормальный белый шум с заданной спектральной плотностью N_0 ; сигнал имеет вид

$$u_c(t) = \begin{cases} U \cos(\omega t + \psi) & \text{при } t \in [0, T], \\ 0 & \text{вне этих пределов.} \end{cases} \quad (11.2)$$

Амплитуда U и фаза ψ сигнала в течение времени анализа T неизменны, а от одного опыта к другому изменяются по случайным законам — амплитуда U имеет рэлеевское распределение, а фаза ψ распределена равномерно в пределах $0 \dots 2\pi$, т. е. сигнал является медленно флуктуирующим.

Требуется найти оптимальную структуру обнаружителя.

Сформулированные исходные данные представляют собой совокупность $\{V, O_s\}$ условий работы оптимизируемой (синтезируемой) системы (в данном случае — устройства обнаружения).

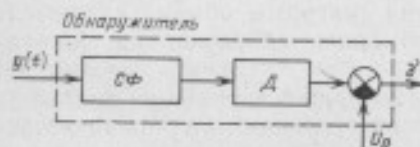


Рис. 11.1

Для того чтобы эти данные были полными, необходимо, кроме того, выбрать состав показателей качества и критерий качества. Главными показателями качества обнаружения являются вероятности $P_{пр}$ и

$P_{лт}$ пропуска сигнала и ложной тревоги и требуемая (удельная) средняя энергия сигнала

$$Q_{ср} = P_{ср} T, \quad (11.3)$$

$$\text{где } P_{ср} = M[U^2/2] \quad (11.4)$$

— требуемая (удельная) средняя мощность сигнала.

Важными показателями качества являются также требуемое время T обнаружения сигнала, определяющее быстродействие обнаружителя, и максимально допустимая спектральная плотность помехи N_0 . Однако в соответствии с исходными данными величины T и N_0 полагаются заданными или, что эквивалентно, переведенными в разряд ограничений типа равенства. Следовательно, в рассматриваемой задаче показатели качества

$$k_1 = P_{пр}, \quad k_2 = P_{лт}, \quad (11.5)$$

$$k_3 = Q_{ср}. \quad (11.6)$$

Все эти показатели имеют стандартный вид, поэтому их преобразование не требуется.

Кроме показателей k_1, k_2, k_3 качество обнаружителя, вообще говоря, должно характеризоваться также показателями надежности, массы, габаритных размеров, стоимости и др. Однако в данном примере для простоты рассматриваются задачи синтеза (оптимизации) обнаружителя только по двум или по трем показателям ка-

чества. Поэтому полученные при этом значения учитываемых показателей (k_1, k_2, k_3) можно считать лишь потенциальными, а найденную структуру — оптимальной лишь в смысле совокупности этих главных показателей.

Сначала рассмотрим задачу синтеза обнаружителя только по двум показателям k_1 и k_2 . Для этого необходимо задать значение средней энергии сигнала или, что в данном случае эквивалентно, отношение

$$k'_3 = q = Q_{ср}/N_0 \quad (11.7)$$

сигнала к шуму. Иначе говоря, при синтезе по вектору $K = \langle k_1, k_2 \rangle$ показатель k_3 (или k'_3) полагается заранее переведенным в разряд ограничений типа равенства.

Решим задачу оптимизации сначала на основе БКП, т. е. найдем множество $M_{\text{вх}}$ систем, нехудших в смысле вектора K , и соответствующую им диаграмму обмена между показателями k_1 и k_2 . В [58] эта задача решена весовым методом. Для разнообразия решим ее методом рабочих характеристик¹, т. е. найдем минимальное значение k_1 при фиксированном (но произвольном) значении k_2 . Иначе говоря, требуется определить минимальное значение вероятности $P_{пр}$ пропуска сигнала при фиксированном (но произвольном) значении вероятности $P_{лт}$ ложной тревоги, т. е. для известного в теории обнаружения критерия Неймана — Пирсона. Решение этой задачи для рассматриваемого обнаружителя медленно флуктуирующего сигнала известно (см., например [5],) и имеет вид (при $P_{пр} \leq 0,1$; $P_{лт} \leq 0,1$)

$$P_{пр, \text{мин}} \approx \frac{\ln 1/P_{лт}}{q}, \quad (11.8)$$

т. е. в обозначениях (11.5)

$$k_{1 \text{ мин}} = (\ln 1/k_2)/q. \quad (11.8')$$

Обнаружитель, обладающий такими показателями качества, изображен на рис. 11.1 и состоит из согласованного фильтра СФ, линейного детектора огибающей Д и устройства сравнения с порогом U_0 . Фильтр СФ имеет передаточную функцию

$$K_{\Phi}(j\omega) = AS_c(-j\omega) \exp(-j\omega t_0), \quad (11.9)$$

где $S_c(j\omega)$ — комплексный спектр сигнала (11.2); A — произвольная константа; t_0 — произвольное запаздывание сигнала, удовлетворяющее условию $t_0 \geq T$. Обычно целесообразно полагать $t_0 = T$ [5].

Пороговое напряжение

$$U_0 = U_{ш} \sqrt{2 \ln 1/P_{лт}} = U_{ш} \sqrt{2 \ln 1/k_2}, \quad (11.10)$$

где $U_{ш}$ — среднеквадратическое значение напряжения шума на выходе фильтра, пропорциональное величине A .

¹ Так как в данной задаче множество $M_{\text{вх}}$ выпуклое, то решения, полученные весовым методом и методом рабочих характеристик, совпадают.

Зависимость (11.8'), справедливая для любых (положительных) значений показателя k_2 , представляет собой рабочую характеристику. Как монотонно убывающая, она совпадает с геометрическим местом нехудших точек (в пространстве показателей k_1, k_2), т. е. с диаграммой обмена между показателями k_1 и k_2 . Поэтому искомая диаграмма обмена (рис. 11.2) имеет вид

$$k_1 = (\ln 1/k_2)/q, \quad (11.11)$$

где $k_1 \leq 0,1$; $k_2 \leq 0,1$.

Каждая ее точка, например точка A , является нехудшей, и ей соответствует некоторая нехудшая система S_A с показателями ($P_{пр.0}$ и $P_{лт.0}$).

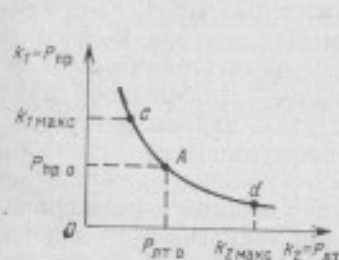


Рис. 11.2

Из рис. 11.1 и соотношений (11.9), (11.10) следует, что одна нехудшая система отличается от другой лишь значением порогового напряжения U_0 . Например, точке A соответствует система (обнаружитель), изображенная на рис. 11.1, у которой

$$U_0 = U_{ш} \sqrt{2 \ln 1/P_{лт.0}}. \quad (11.12)$$

Если предельно допустимые значения показателей качества равны $k_{1.макс} = P_{пр.макс}$ и $k_{2.макс} = P_{лт.макс}$, то строго допустимыми будут лишь те нехудшие точки (системы), которые расположены на участке cd (рис. 11.2). Таким образом, синтез обнаружителя по двум показателям качества ($k_1 = P_{пр}$, $k_2 = P_{лт}$) в рамках безусловного критерия предпочтения привел к семейству нехудших обнаружителей, структура которых определяется рис. 11.1 и соотношением (11.9) и которые отличаются друг от друга лишь значениями порогового напряжения U_0 . Значения показателей качества этих обнаружителей определяются диаграммой обмена (11.11).

Чтобы из найденного семейства (множества) нехудших обнаружителей выбрать один (оптимальный), необходимо ввести какой-либо условный критерий предпочтения. В радиолокации, как известно, наиболее распространенным является критерий Неймана — Пирсона: оптимальным считается такой обнаружитель, который обеспечивает минимум показателя k_1 при заданном достаточно малом значении $k_{20} = P_{лт.0}$ показателя k_2 . При этом k_{20} обычно выби-

рается максимально допустимым, т. е. полагается $k_{20} = k_{2.макс} = P_{лт.макс}$. В этом случае, как следует из рис. 11.2, оптимальной является крайняя правая точка диаграммы обмена cd , т. е. точка d . Пороговое напряжение U_0 в оптимальной системе

$$U_{0.опт} = U_{ш} \sqrt{2 \ln 1/P_{лт.макс}}. \quad (11.13)$$

Рассмотрим теперь синтез обнаружителя по трем показателям качества k_1, k_2, k_3 , определяемым соотношениями (11.5), (11.6). Так как для двух показателей (k_1 и k_2) при фиксированном (но произвольном) значении третьего (k_3) задача уже решена, для отыскания множества $M_{нх}$ точек (систем), нехудших в смысле трех показателей (k_1, k_2, k_3), проще всего применить комбинированный метод, описанный в § 9.2. В соответствии с этим методом сначала надо перевести показатель $k_3 = q$ в разряд ограничений типа равенств и найти множество $M'_{нх}$ точек (систем), нехудших в смысле показателей (k_1, k_2). Но это множество уже было найдено и ему соответствует диаграмма обмена (11.11), которая с учетом (11.7) принимает вид

$$k_1 = \frac{\ln 1/k_2}{k_3}, \quad (11.14)$$

где черта снизу показывает, что k_3 рассматривается здесь лишь как некоторая фиксированная (но произвольная) константа.

Так как зависимость k_1 от k_2 монотонно убывающая, то это означает, что уравнение (11.14) совпадает с уравнением диаграммы обмена между всеми тремя показателями (k_1, k_2, k_3), т. е. искомая диаграмма обмена (геометрическое место точек, нехудших в смысле совокупностей трех показателей k_1, k_2 и k_3) имеет вид

$$k_1 = (\ln 1/k_2)/k_3, \quad \text{т. е. } P_{пр} = (\ln 1/P_{лт})/q. \quad (11.15)$$

На плоскости показателей k_1, k_2 она изображается семейством кривых, параметром которых является k_3 (рис. 11.3). Каждой точке (например, точке A) этого семейства соответствует некоторая нехудшая система S_A , обладающая показателями $P_{пр.0}$, $P_{лт.0}$, q_0 . Из предыдущего следует, что, как и при двух показателях, одна нехудшая система отличается от другой лишь значениями порогового напряжения U_0 , определяемого соотношением (11.10) [так как передаточная функция (11.9) СФ и параметры D не зависят от координат нехудшей точки A , т. е. от значений показателей k_1, k_2, k_3].

Для того чтобы выбрать из найденного множества нехудших точек (систем) одну (оптимальную), нужно задаться каким-либо УКП. При радиолокационных применениях обнаружителя наиболее распространенным является следующий УКП: оптимальной считается система, обеспечивающая минимальное значение показателя k_3 (т. е. в конечном счете — требуемую среднюю энергию $Q_{ср}$ сигнала) при предельно допустимых значениях вероятностей $P_{пр}$ и $P_{лт}$ ошибок обнаружения, т. е. при

$$P_{пр} = P_{пр.макс}, \quad P_{лт} = P_{лт.макс}. \quad (11.16)$$

При этом, полагая

$$k_1 = P_{\text{лт. макс}} = k_{1 \text{ макс}}, \quad P_{\text{пр}} = P_{\text{пр. макс}} = k_{2 \text{ макс}},$$

из (11.15) находим

$$k_{3 \text{ мин}} = q_{\text{мин}} = \frac{\ln 1/P_{\text{лт. макс}}}{P_{\text{пр. макс}}} = \frac{\ln 1/k_{1 \text{ макс}}}{k_{1 \text{ макс}}}. \quad (11.17)$$

Тем самым определены все координаты ($k_{1 \text{ макс}}, k_{2 \text{ макс}}, k_{3 \text{ мин}}$) оптимальной точки — им соответствует крайняя точка B диаграммы обмена (рис. 11.3) и оптимальной является система (обнаружитель), у которой порог определяется соотношением (11.13), а минимально допустимое отношение $q_{\text{мин}}$ сигнала к шуму на входе определяется соотношением (11.7).

Оценим смысл полученных результатов. При определении нехудших систем и оптимальной системы принимались во внимание только три показателя качества $k_1 = P_{\text{пр}}, k_2 = P_{\text{лт}}$ и $k_3 = Q_{\text{ср}}/N_0$. С учетом дополнительных показателей качества (например, массы, стоимости и др.) и дополнительных условий (например, действия дополнительных видов помех) нехудшими и оптимальными могут оказаться системы, отличные от изображений на рис. 11.1 [например, может оказаться более целесообразным иметь частотную характеристику фильтра отличной от (11.9)]. Однако найденные для приведенных ранее идеализированных условий и ограниченно-го числа показателей результаты сохраняют свою ценность, так как определяют потенциальные (наилучшие теоретически возможные) значения каждого из показателей (k_1, k_2, k_3) при фиксированных значениях остальных двух показателей. Например, учет дополнительных видов помех и дополнительных показателей качества может привести лишь к увеличению минимального требуемого значения $q_{\text{мин}}$ отношения сигнал-шум на входе при заданных значениях вероятностей $P_{\text{пр}}$ и $P_{\text{лт}}$ ошибок обнаружения сигнала.

11.2. ОПТИМИЗАЦИЯ УСТРОЙСТВА ВОСПРОИЗВЕДЕНИЯ ДИСКРЕТНЫХ СООБЩЕНИЙ ПО m ПОКАЗАТЕЛЯМ КАЧЕСТВА

На вход устройства воспроизведения дискретных сообщений поступает смесь

$$y(t) = u_0(t, x, \alpha_1, \dots, \alpha_r) + u_{\text{ш}}(t) \quad (11.18)$$

сигнала $u_0(\cdot)$ и шума $u_{\text{ш}}(t)$.

Кроме полезного сообщения x сигнал содержит также ряд неизвестных в месте приема случайных параметров $\alpha_1, \dots, \alpha_r$, не содержащих информации о x (так называемых паразитных, мешающих или неинформационных параметров). В течение интервала наблюдения (воспроизведения) $[0, T]$ сообщение x (являющееся векторной или скалярной величиной) может иметь одно из конечного числа n возможных значений $x_1, \dots, x_j, \dots, x_n$ с априорными вероятностями $P(x_1), \dots, P(x_j), \dots, P(x_n)$ соответственно, причем

$$\sum_{i=1}^n P(x_i) = 1. \quad (11.19)$$

На основании наблюдения (анализа) смеси $y(t)$ в течение интервала $[0, T]$ к концу этого интервала должно быть вынесено решение γ о том, какое именно из n возможных значений имело сообщение x в данном опыте (интервале наблюдения).

Предполагается, что распределение $W_{\alpha}(\alpha_1, \dots, \alpha_r)$ и априорное распределение $W_{\text{ш}}(u_{\text{ш}})$ шума известны, а распределение $P(x_i)$ ($i = \overline{1, n}$) может быть либо известным, либо неизвестным. Вследствие действия шума воспроизведение сообщения x_j сопровождается ошибками вида $(x_i | x_j)$ с вероятностями

$$P_{ij} = P(x_i | x_j), \quad i = \overline{1, n}, \quad j = \overline{1, n}, \quad i \neq j, \quad (11.20)$$

где P_{ij} — вероятность принятия решения $\gamma = x_i$, когда в действительности сообщение имеет вид x_j . При этом общее число видов ошибок $m = n(n-1)$.

Следовательно, качество (помехоустойчивость) приема характеризуется в общем случае m показателями (условными вероятностями ошибок)

$$k_{ij} = P_{ij}, \quad i = \overline{1, n}, \quad j = \overline{1, n}, \quad i \neq j. \quad (11.21)$$

Найдем множество $M_{\text{нх}}$ систем, нехудших в смысле вектора качества

$$K = \langle k_1, \dots, k_m \rangle. \quad (11.22)$$

Применим для отыскания $M_{\text{нх}}$ весовой метод, т. е. найдем минимум по $S \in M_{\text{сн}}$ взвешенной суммы показателей вида

$$k'_{\text{в}} = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} k_{ij}, \quad (11.23)$$

где веса λ_{ij} удовлетворяют условиям

$$\lambda_{ij} > 0, \quad \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} = 1, \quad \lambda_{ii} = 0. \quad (11.24)$$

Любая из систем, минимизирующая величину $k'_{\text{в}}$, является нехудшей, а множество $M_{\text{нх}}$ таких систем, соответствующих различным комбинациям весов λ_{ij} , удовлетворяет условию $M_{\text{нх}} \subset M_{\text{сн}}$.

В большинстве задач, решаемых в теории оптимального по помехоустойчивости приема, ограничения $O_{\text{н}}$ являются настолько слабыми, что множество $M_{\text{сн}}$ строго допустимых систем оказывается выпуклым. В этом случае, как отмечалось в § 9.2, множество $M_{\text{нх}}$ совпадает с $M_{\text{сн}}$. В дальнейшем будем полагать, что имеет место именно этот случай. Тогда все элементы (точки, системы) множества $M_{\text{нх}}$ можно найти весовым методом, т. е. путем минимизации (по $S \in M_{\text{сн}}$) взвешенной суммы (11.23).

Из (11.21) и (11.23) следует, что

$$k'_{\text{в}} = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} P_{ij}, \quad \lambda_{ii} = 0. \quad (11.25)$$

Очевидно, что

$$P_{ij} = \int_{\Gamma_i} P(y | x_j) dy, \quad (11.26)$$

где $P(y | x_j)$ — плотность вероятности входной смеси $y(t)$ при условии, что $x = x_j$; Γ_i — область принятия решения $\gamma = x_i$.

Из (11.18) следует, что функцию правдоподобия $P(y|x)$ можно определить через заданные распределения $W_\alpha(a_1, \dots, a_r)$ и $W_\pi(u_\pi)$ по формуле:

$$P(y|x) = \int \dots \int P(y|x, a_1, \dots, a_r) W_\alpha(a_1, \dots, a_r) da_1, \dots, da_r,$$

в которой

$$P(y|x, a_1, \dots, a_r) = W_\pi(y - u_\pi(i, x, a_1, \dots, a_r)).$$

Из (11.25) и (11.26) следует, что наилучшие системы можно найти из условия:

$$\text{обеспечить } \min_{S \in M_{\text{ед}}} k'_n, \quad (11.27)$$

где

$$k'_n = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} \int_{\Gamma_i} P(y|x_j) dy, \quad \lambda_{ii} = 0. \quad (11.28)$$

Из (11.28) видно, что минимизацию величины k'_n можно осуществить лишь выбором надлежащей совокупности $\Gamma = \langle \Gamma_1, \dots, \Gamma_i, \dots, \Gamma_n \rangle$ областей принятия решений. Выражение (11.28) можно представить в виде

$$k'_n = \sum_{i=1}^n \int_{\Gamma_i} \sum_{j=1}^n \lambda_{ij} P(y|x_j) dy, \quad \lambda_{ii} = 0. \quad (11.28')$$

Так как все весовые коэффициенты λ_{ij} и плотности $P(y|x_j)$ неотрицательны, эта сумма минимальна, если области Γ_i принятия решений $y = x_i$ выбираются следующим образом. Каждая данная реализация $y(t)$ относится к области с тем номером $i = i_n$, которому соответствует наименьшее (по всем значениям индекса i) значение суммы

$$\Phi_i = \sum_{j=1}^n \lambda_{ij} P(y|x_j), \quad i = \overline{1, n}, \quad \lambda_{ii} = 0. \quad (11.29)$$

Это означает, что алгоритм действия каждой наилучшей системы, соответствующей данной комбинации весовых коэффициентов λ_{ij} ($i = \overline{1, n}, j = \overline{1, n}, \lambda_{ii} = 0$), должен быть таким, как показано на рис. 11.4.

Для принятой реализации $y(t)$ определяются значения функции правдоподобия $P(y|x_j)$, соответствующие всем возможным значениям $x_1, \dots, x_j, \dots, x_n$ сообщения x :

$$z_1 = P(y|x_1), \dots, z_j = P(y|x_j), \dots, z_n = P(y|x_n); \quad (11.30)$$

затем вычисляются суммы вида (11.29) для всех значений $i = \overline{1, n}$:

$$\Phi_i = \sum_{j=1}^n \lambda_{ij} P(y|x_j), \dots, \Phi_i = \sum_{j=1}^n \lambda_{ij} P(y|x_j), \dots, \Phi_n = \sum_{j=1}^n \lambda_{nj} P(y|x_j), \text{ где } \lambda_{ii} = 0.$$

В схеме сравнения и отбора наименьшего ССО величины $\Phi_1, \dots, \Phi_n$ сравниваются между собой, определяется, какая из них является наименьшей, и вы-

дается решение $y = x_{i_n}$, где i_n — индекс той из сравниваемых величин $\Phi_1, \dots, \Phi_n$, которая оказалась наименьшей. Из (11.29) и (11.30) следует, что

$$\Phi_i = \sum_{j=1}^n \lambda_{ij} z_j, \quad i = \overline{1, n}, \quad \lambda_{ii} = 0,$$

т. е. вычислитель величин Φ_i ($i = \overline{1, n}$) состоит из n сумматоров взвешенных значений величин $z_1, \dots, z_n$, поступающих на его вход. Так как величины z_j — это плотности вероятности $P(y|x_j)$, то они неотрицательны и, следовательно, $z_j > 0$ ($j = \overline{1, n}$). Одна наилучшая система отличается от другой только значениями весовых коэффициентов λ_{ij} , т. е. лишь структурой вычислителя величин Φ_i (рис. 11.4), так как структура вычислителя значений $P(y|x_j)$ функции правдоподобия и устройства ССО не зависит от значений λ_{ij} .

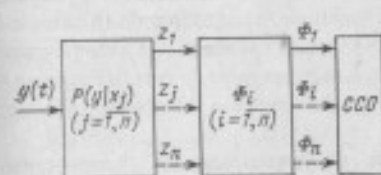


Рис. 11.4

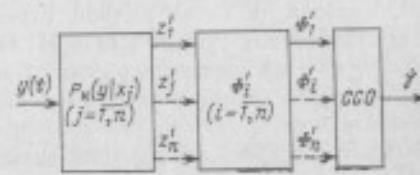


Рис. 11.5

Пусть функция правдоподобия представима в виде

$$P(y|x) = \Phi_0(y) \Phi(y|x), \quad (11.31)$$

где $\Phi_0(y)$ — некоторый произвольный положительный функционал от входной реализации $y(t)$, не зависящий от сообщения x . Тогда в соответствии с (11.30)

$$z_j = \Phi_0(y) z'_j, \quad (11.32)$$

где $z'_j = \Phi(y|x_j) = P(y|x_j) / \Phi_0(y)$.

Действие наилучшей системы основано на отборе наименьшей из величин $\Phi_i = \sum_{j=1}^n \lambda_{ij} z_j$ ($i = \overline{1, n}$). Поэтому вероятности P_{ij} ошибочных решений не изменятся, если на вход вычислителя величин Φ_i подавать вместо значений z_j ($j = \overline{1, n}$) значения z'_j (так как множитель $\Phi_0(y)$ не зависит от номера j). Это означает, что в наилучшей системе (рис. 11.4) вычислитель значений $P(y|x_j)$ функции правдоподобия можно заменить без ущерба для ее качества (помехоустойчивости) вычислителем величин z'_j , т. е. значений нормированной функции правдоподобия

$$P_n(y|x) = P(y|x) / \Phi_0(y). \quad (11.33)$$

При этом схема на рис. 11.4 принимает вид, изображенный на рис. 11.5. Здесь вычислитель величин Φ'_i образует взвешенные суммы вида

$$\Phi'_i = \sum_{j=1}^n \lambda_{ij} z'_j. \quad (11.34)$$

¹ Полагается, что шум $u_\pi(t)$ и сигнал $u_s(\cdot)$ статистически независимы.

Пусть $P(y|0)$ — функция правдоподобия $P(y|x)$ в отсутствие сигнала [$u_0(t)=0$], т. е. при наличии только шума $u_n(t)$. Тогда, если помеха такова, что для всех возможных реализаций $y(t)$ выполняется условие

$$P(y|0) > 0, \quad (11.35)$$

то можно положить $\Phi_0(y) = P(y|0)$ и из (11.33) получим

$$P_n(y|x) = P(y|x)/P(y|0) = \Lambda(y|x), \quad (11.36)$$

где $\Lambda(y|x)$ — отношение правдоподобия. Следовательно, при выполнении (11.35) вычислитель значений нормированной функции правдоподобия можно заменить вычислителем отношения правдоподобия.

Таким образом, любую нехудшую (по помехоустойчивости) систему воспроизведения дискретных сообщений можно представить в виде, изображенном на рис. 11.5, т. е. состоящей из вычислителя значений нормированной функции правдоподобия (в более частном случае — отношения правдоподобия), вычислителя взвешенных сум Φ'_i вида (11.34) и устройства сравнения и отбора наименьшей из этих взвешенных сум. При этом одна нехудшая система отличается от другой лишь значениями весов λ_{ij} , закладываемых в алгоритм (11.34) вычислителя Φ'_i .

Поскольку схема на рис. 11.5 является весьма универсальной, рассмотрим ее несколько подробнее. Вычислитель $P_n(y|x_j)$ определяет в общем случае m неотрицательных чисел $z'_1, \dots, z'_j, \dots, z'_n$, дальнейшая обработка которых сводится к сравнительно несложным операциям образования взвешенных сум (11.34) и отбору наименьшей из них. Поэтому основной частью нехудшей системы является вычислитель значений $z'_1, \dots, z'_n$ нормированной функции правдоподобия $P_n(y|x)$ [или отношения правдоподобия $\Lambda(y|x)$]. В общем случае он представляет собой n -канальное приемное устройство, имеющее n выходов и выдающее n -мерный вектор $z' = \langle z'_1, \dots, z'_j, \dots, z'_n \rangle$, составляющими которого являются неотрицательные числа $z'_1, \dots, z'_n$. При этом каждый канал номера j вычисляет функционал $P_n(y|x_j)$, т. е. согласуется с ожидаемым сигналом $u_0(t, x_j, \alpha_1, \dots, \alpha_r)$ (в ряде хорошо известных случаев этот канал содержит, например, коррелятор или согласованный фильтр и детектор отгибающей).

В частных случаях вектор z' может иметь не n составляющих, а меньше. Например, в достаточно важном случае вектор z' может содержать всего одну составляющую, даже если $n \gg 1$. Действительно, пусть смесь $y(t)$ сигнала и помех такова, что среди составляющих $z'_1, \dots, z'_n$ существует такая составляющая z'_r , зная которую в принципе можно однозначно определить все остальные $(n-1)$ составляющие с учетом только априорных данных, т. е. без использования значений реализации $y(t)$ входной смеси. Это означает, что для принятия решения достаточно иметь апостериорную информацию лишь об этой единственной составляющей z'_r , и, следовательно, схему на рис. 11.5 можно свести к виду, изображенному на рис. 11.6.

Решающее устройство (РУ) на основании поступающей в него информации о неотрицательной величине z'_r (и имеющейся априорной информации) должно вывести одно из n возможных решений ($\gamma = x_1, \dots, \gamma = x_n$). Так как область возможных решений в пространстве векторов z' имеет в данном случае вид полупрямой $z'_r > 0$, показанной на рис. 11.7 (поскольку в данном случае вектор z' вырождается в скалярную величину z'_r), и должна содержать n областей

$\Gamma_1, \dots, \Gamma_n$ принятия решений, то границами между этими областями могут быть лишь некоторые точки (а не линии или поверхности) $z_{10}, \dots, z_{n-1,0}$. Иначе говоря, в этом случае решение о значении сообщения x должно приниматься путем сравнения неотрицательного напряжения (z'_r) на выходе вычислителя $P_n(y|x)$ [или $\Lambda(y|x)$] с пороговыми величинами $z_{10}, \dots, z_{n-1,0}$.

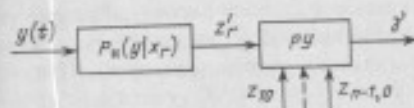


Рис. 11.6

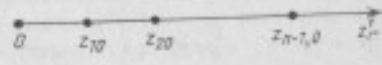


Рис. 11.7

Таким образом, в рассматриваемом частном случае n -канальный вычислитель $P_n(y|x)$ [или $\Lambda(y|x)$] сводится к одноканальному, а совокупность вычислителей величин Φ_i ($i = \overline{1, n}$) и устройства ССО заменяются РУ, сравнивающим величину z'_r с $(n-1)$ пороговыми величинами $z_{10}, \dots, z_{n-1,0}$. При этом одна нехудшая система отличается от другой лишь значениями этих пороговых величин.

Рассмотрим для иллюстрации следующий простой пример. На интервале наблюдения $[0, T]$ сигнал имеет вид синусоиды $a \cos(\omega t + \varphi)$ с постоянной амплитудой a и случайной равномерно распределенной в интервале $[0, 2\pi]$ начальной фазой φ . Амплитуда a может принимать лишь одно из трех значений a_1, a_2, a_3 . Требуется определить, какой именно сигнал (с какой амплитудой) присутствует на данном интервале $[0, T]$, если $u_n(t)$ — аддитивный нормальный белый шум со спектральной плотностью N_0 .

Очевидно, в данном случае $x = a$, $x_1 = a_1$, $x_2 = a_2$, $x_3 = a_3$ и $n = 3$. Функция правдоподобия $P(y|x)$ при этом имеет вид [5]

$$P(y|x) = c \exp \left(-\frac{1}{N_0} \int_0^T y^2 dt \right) \exp \left(-\frac{x^2 T}{2N_0} \right) I_0 \left(\frac{2x M(y)}{N_0} \right), \quad (11.37)$$

где $I_0(\eta)$ — модифицированная функция Бесселя;

$$M(y) = \sqrt{X^2 + Y^2}; \quad X = \int_0^T y(t) \cos \omega t dt; \quad Y = \int_0^T y(t) \sin \omega t dt. \quad (11.38)$$

Из (11.31) и (11.37) следует, что в данном случае можно положить

$$\Phi_0(y) = c \exp \left(-\frac{1}{N_0} \int_0^T y^2(t) dt \right) = P(y|0)$$

и

$$P_n(y|x) = \Lambda(y|x) = \exp \left(-x^2 T / 2N_0 \right) I_0(2xM(y)/N_0). \quad (11.39)$$

Поэтому в соответствии с (11.32), (11.33) и (11.39)

$$\begin{aligned} z'_1 &= P_n(y|x_1) = \exp \left(-x_1^2 T / 2N_0 \right) I_0(2x_1M(y)/N_0), \\ z'_2 &= P_n(y|x_2) = \exp \left(-x_2^2 T / 2N_0 \right) I_0(2x_2M(y)/N_0), \\ z'_3 &= P_n(y|x_3) = \exp \left(-x_3^2 T / 2N_0 \right) I_0(2x_3M(y)/N_0), \end{aligned} \quad (11.40)$$

где $x_1 = a_1$; $x_2 = a_2$; $x_3 = a_3$.

Из (11.40) следует, что в данном случае значение лишь одной из трех величин z'_1, z'_2, z'_3 позволяет однозначно (с учетом априорных данных) опре-

делить две остальные величины. Пусть, например, известна величина z'_1 . Тогда, учитывая, что величины x_1 , T и N_0 априори известны, а $I_0(\eta)$ — монотонная функция своего аргумента, первое из соотношений (11.40) позволяет однозначно определить величину $M(y)$, а второе и третье соотношения (11.40) — величинами z'_2 и z'_3 [поскольку значения x_2 и x_3 также полагаются априори известными — неизвестно лишь, какие именно из трех значений (x_1 , x_2 , x_3) имеют место на данном интервале наблюдения $[0, T]$]. Таким образом, в рассматриваемом случае не требуется применять n -канальный (3-канальный) вычислитель и структура нехудших систем определяется одноканальной схемой (рис. 11.6). При этом в качестве z' можно принять любую из величин z'_1 , z'_2 , z'_3 , например

$$z'_1 = \exp(-x_1^2 T / 2N_0) I_0(2x_1 M(y) / N_0). \quad (11.42)$$

В РУ эта величина должна сравниваться с двумя пороговыми z_{10} , z_{20} , и выносятся решение

$$y = \begin{cases} x_1, & \text{если } z'_1 < z_{10}, \\ x_2, & \text{если } z_{10} < z'_1 < z_{20}, \\ x_3, & \text{если } z'_1 > z_{20} \end{cases} \quad (11.43)$$

(здесь принято для определенности, что $x_1 < x_2 < x_3$).

Из (11.42) и (11.43) следует, что вероятности ошибочных решений (P_{12} , P_{13} , P_{23} , P_{21} , P_{31} , P_{32}) не изменятся, если сравнивать не z'_1 с z_{10} и z_{20} , а $z''_1 = M(y)$ с надлежащим образом скорректированными пороговыми значениями z''_{10} и z''_{20} .

Но из (11.38) видно, что величина $M(y)$ пропорциональна огибающей напряжения на выходе линейного фильтра, согласованного с сигналом, имеющим вид отрезка (на интервале $[0, T]$) синусоиды $a \cos(\omega t + \varphi)$. Поэтому в рассматриваемом примере схема на рис. 11.6 упрощается и принимает вид, изображенный на рис. 11.8, т. е. состоит из согласованного фильтра СФ, детектора огибающей Д и решающего устройства РУ, в кото-

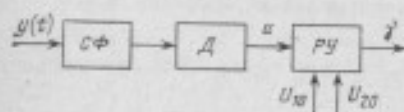


Рис. 11.8

ром выходное напряжение детектора сравнивается с двумя пороговыми напряжениями U_{10} и U_{20} , и выносятся решение

$$y = \begin{cases} x_1, & \text{если } u < U_{10}, \\ x_2, & \text{если } U_{10} < u < U_{20}, \\ x_3, & \text{если } u > U_{20} \end{cases}$$

При этом одна нехудшая система отличается от другой лишь значениями U_{10} и U_{20} .

Приведенный анализ основан на применении лишь безусловного критерия предпочтения. Поэтому удалось найти не оптимальную (в смысле какого-либо УКП) систему, а множество нехудших систем. Однако оптимальной (в смысле любого разумного УКП) может быть лишь одна из нехудших систем¹, а все нехудшие системы, как было показано, различаются только значениями весовых коэффициентов λ_{ij} в вычислителе Φ_i или пороговыми напряжениями на выходе. Следовательно, на основе БКП, т. е. не вводя никаких условных критериев предпочтения и не делая поэтому никаких предположений о значениях априорных вероятностей $P(x_1), \dots, P(x_n)$, удалось получить достаточно полное

представление о структуре оптимальных (по помехоустойчивости) устройств воспроизведения сообщений. Кроме того, поскольку каждая нехудшая система обладает свойством m -кратного минимума, т. е. минимальным значением каждого показателя качества $k_1, \dots, k_m$ [при фиксированных значениях остальных $(m-1)$ показателей], найденные нехудшие системы обладают потенциальными (многомерными) характеристиками помехоустойчивости.

Чтобы выбрать из найденного множества нехудших систем одну (оптимальную), т. е. уточнить (конкретизировать) значения весовых коэффициентов λ_{ij} (а в более частном случае — пороговых значений $z_{10}, \dots, z_{n-1,0}$), необходимо применить какой-либо УКП.

Как отмечалось в § 8.2, таких УКП может быть сколь угодно много. Одним из наиболее распространенных (применительно к оценке помехоустойчивости устройств воспроизведения дискретных сообщений) является критерий минимума среднего риска (2.8). Так как в данном случае $m=n$, то выражение (2.8) принимает вид

$$R = \sum_{i=1}^n \sum_{j=1}^n c_{ij} P_j P_{ij}, \quad (11.44)$$

где $P_j = P(x_j)$; $P_{ij} = P(x_i | x_j)$; $c_{ij} > 0$; $c_{ii} = 0$.

Из сравнения (11.44) с (11.25) следует, что системе, оптимальной в смысле минимума среднего риска R (при заданных значениях априорных вероятностей P_j и потерь c_{ij} , $i, j = \overline{1, n}$, $c_{ii} = 0$) соответствует нехудшая система, у которой

$$\lambda_{ij} = P_j c_{ij}, \quad i = \overline{1, n}, \quad j = \overline{1, n}, \quad (11.45)$$

причем $c_{ii} = 0$.

Так как одна нехудшая система отличается от другой лишь значениями весовых коэффициентов λ_{ij} , учитываемыми в алгоритме действия вычислителя взвешенных сумм Φ_i , то соотношение (11.45) позволяет определить конкретные значения λ_{ij} и тем самым выбрать из множества нехудших систем единственную (оптимальную в смысле минимума среднего риска). Однако, как следует из (11.45), для такого выбора необходимо знать априорные вероятности P_j ($j = \overline{1, n}$) и потери c_{ij} ($i = \overline{1, n}$, $j = \overline{1, n}$, $c_{ii} = 0$), соответствующие различным видам ($x_i | x_j$) ошибок воспроизведения сообщений.

Во многих задачах допустимо полагать, что потери c_{ij} не зависят от индекса i , т. е.

$$c_{ij} = \begin{cases} c_j & \text{при } i \neq j, \\ 0 & \text{при } i = j. \end{cases} \quad (11.46)$$

Это означает, что одинаково опасны ошибки воспроизведения вместо x_j любого другого (неважно какого именно) сообщения x_i ($i \neq j$). С учетом (11.46) выражение (11.44) принимает вид

$$R = \sum_{j=1}^n c_j P_j P_{j\pi}, \quad P_{j\pi} = \sum_{i=1}^n P_{ij}, \quad i \neq j, \quad (11.47)$$

где $P_{j\pi}$ — вероятность приема вместо сообщения x_j какого-либо другого (неважно какого именно) из сообщений $x_1, \dots, x_n$. Очевидно,

$$P_{j\pi} = 1 - P_{jj}, \quad (11.48)$$

где P_{jj} — вероятность того, что переданное сообщение x_j будет принято правильно

¹ При заданном составе показателей качества.

но (т. е. будет принято решение $\gamma = x_j$). Иначе говоря, P_{jj} — условная (при условии передачи x_j) вероятность правильного воспроизведения сообщения x_j , а P_{jk} — условная вероятность неправильного воспроизведения этого сообщения. Из (11.47) и (11.48) следует, что минимуму среднего риска R соответствует максимум взвешенной суммы

$$P_* = \sum_{j=1}^n c_j P_j P_{jj}. \quad (11.49)$$

Определим структуру системы, максимизирующую величину P_* . Так как $P_{jj} = \int_{\Gamma_j} P(y|x_j) dy$, то с учетом (11.49) максимум величины P_* будет иметь место, если максимальна сумма

$$\sum_{j=1}^n \int_{\Gamma_j} c_j P_j P(y|x_j) dy.$$

Отсюда следует, что оптимальная система должна вычислять величины

$$\Phi''_j = c_j P_j P(y|x_j), \quad j = \overline{1, n}, \quad (11.50)$$

выбирать ту из них, которая окажется наибольшей, и полагать $\gamma = x_{j^*}$, где j^* — порядковый номер наибольшей из сумм Φ''_j . Так как при этом решение основано на сравнении и отборе наибольшего значения, вместо сравнения величин (11.50) можно сравнивать величины

$$\Phi'''_j = c_j P_j z'_j, \quad j = \overline{1, n}, \quad (11.51)$$

где z'_j определяется сообщением (11.32). Поэтому структуру оптимальной системы можно представить в виде, изображенном на рис. 11.9, где ССО' — схема сравнения и отбора наибольшей из входных величин (Φ'''_j) и принятия решения $\gamma = x_{j^*}$.

Рассмотрим теперь еще более частный, но также распространенный на практике случай критерия среднего риска, когда вместо (11.46) выполняется условие

$$c_{ij} = \begin{cases} 1 & \text{при } i \neq j, \\ 0 & \text{при } i = j. \end{cases} \quad (11.52)$$

т. е. все виды ошибок ($x_i|x_j$) одинаково опасны. При этом, как следует из (11.48), (11.49), критерий минимума среднего риска сводится к критерию минимума полной вероятности $P_{\text{ош}}$ ошибки

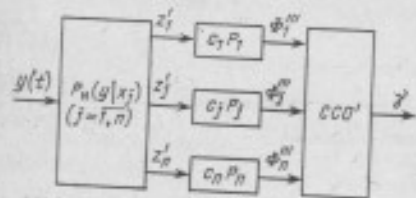


Рис. 11.9

или эквивалентному ему критерию максимума полной вероятности правильного воспроизведения сообщений и оптимальная система имеет вид, изображенный на рис. 11.9, но веса $c_1, \dots, c_n$ должны быть приняты равными единице.

Из сравнения схем на рис. 11.5 и 11.9 следует, что применение условного критерия предпочтения — критерия минимума полной вероятности ошибки

позволяет преобразовать общую схему наилучшей системы (рис. 11.5) в схему оптимальной системы (рис. 11.9, в котором $c_1 = \dots = c_n = 1$), отличающуюся лишь более простым алгоритмом преобразования выходных величин $z'_1, \dots, z'_n$ вы-

числителя нормированной функции правдоподобия (или отношения правдоподобия).

Если кроме условия (11.52) выполняется дополнительное условие $P_1 = \dots = P_n$, т. е. априорные вероятности передачи всех видов $x_1, \dots, x_n$ сообщения одинаковы, то схема оптимальной системы (рис. 11.9) еще более упрощается, так как в ней можно положить $c_j P_j = 1$, т. е. исключить вычислители величин Φ'''_j и производить сравнение и отбор наибольшей из величин z'_j . При этом, очевидно, вероятности ошибок не изменятся, если вместо сравнения величин z'_j (и отбора наибольшей из них) сравнивать

$$z''_j = f_j(z'_j), \quad j = \overline{1, n}, \quad (11.53)$$

где $f_j(z'_j)$ — любая монотонная функция. Например, часто проще сравнивать не сами нормированные функции правдоподобия $P_n(y|x_j)$ [или отношения правдоподобия $\Lambda(y|x_j)$], а их логарифмы.

11.3. ОПТИМИЗАЦИЯ СИСТЕМЫ ВОСПРОИЗВЕДЕНИЯ НЕПРЕРЫВНЫХ СООБЩЕНИЙ ПО m ПОКАЗАТЕЛЯМ КАЧЕСТВА

Пусть смесь $y(t)$ сигнала и помех на входе приемного устройства (системы воспроизведения сообщений) имеет вид $y(t) = u_c(t, x_1, \dots, x_n) + u_m(t)$, где $x_1, \dots, x_1, \dots, x_n$ — подлежащие воспроизведению сообщения (параметры сигнала), постоянные за время воспроизведения сообщения $[0, T]$; $u_m(t)$ — независимая помеха с известным распределением $W_m(u_m)$. Сообщения x_j ($j = \overline{1, n}$) — непрерывные величины, которые могут принимать любые значения в некотором интервале $[x_{\min}, x_{\max}]$; x^*_j ($j = \overline{1, n}$) — результат воспроизведения сообщения x_j иско- мой системой.

Ошибка воспроизведения j -го сообщения

$$\Delta x_j = x^*_j - x_j. \quad (11.54)$$

Пусть показателями качества являются вторые моменты ошибок, т. е. система показателей качества имеет вид

$$k_{ij} = M[\Delta x_i \Delta x_j] = M[(x^*_i - x_i)(x^*_j - x_j)], \quad j = \overline{1, n}, \quad i = \overline{1, n}, \quad (11.55)$$

где M — знак математического ожидания.

Так как

$$k_{ij} = k_{ji}, \quad (11.56)$$

то общее число (несовпадающих) показателей качества

$$m = n + n(n-1)/2 = n(n+1)/2. \quad (11.57)$$

При $i = j$ $k_{ij} = M[(\Delta x_j)^2]$ — средний квадрат ошибки воспроизведения сообщения x_j ; при $i \neq j$ величина k_{ij} учитывает статистическую связь между ошибками Δx_i и Δx_j .

Итак, качество системы определяется совокупностью m показателей качества вида (11.55), характеризующей точность (помехоустойчивость) воспроизведения сообщений. Найдем множество $M_{\text{нх}}$ систем, наилучших в смысле совокупности этих m показателей. Для этого применим весовой метод, т. е. в соответ-

ствии с (9.9а) будем искать минимум взвешенной суммы показателей k_{ij} вида

$$k'_n = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} k_{ij}, \quad (11.58)$$

где $\lambda_{ij} > 0$; $\lambda_{ij} = \lambda_{ji}$ [в соответствии с (11.56)]; $\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} = 1$.

С учетом (11.55) выражение для k'_n можно записать иначе:

$$k'_n = \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} M[(x^*_i - x_i)(x^*_j - x_j)]. \quad (11.59)$$

Так как операция взятия математического ожидания M линейна, то в (11.59) операции суммирования и математического ожидания можно поменять местами. Тогда

$$k'_n = M \left[\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} (x^*_i - x_i)(x^*_j - x_j) \right]. \quad (11.60)$$

Обозначим

$$x = \langle x_1, \dots, x_n \rangle, \quad x^* = \langle x^*_1, \dots, x^*_n \rangle, \quad (11.61)$$

$$B = \|\lambda_{ij}\| \quad (11.62)$$

— матрица весовых коэффициентов (симметричная, так как $\lambda_{ij} = \lambda_{ji}$, и невырожденная, так как ее определитель не равен нулю).

Тогда (11.60) можно записать более компактно:

$$k'_n = M[I(x, x^*)], \quad (11.63)$$

$$\text{где } I(x, x^*) = (x^* - x)' B (x^* - x) \quad (11.64)$$

— некоторая скалярная функция векторных случайных аргументов x и x^* ; штрих означает знак транспонирования.

Пусть совместная плотность вероятности векторов x и x^* есть $P(x, x^*)$. Тогда в соответствии с (11.63) и (11.64)

$$k'_n = \int \int I(x, x^*) P(x, x^*) dx dx^* = \int \int (x^* - x)' B (x^* - x) P(x, x^*) dx dx^*. \quad (11.65)$$

Здесь интегрирование должно производиться по всем возможным комбинациям значений векторов x и x^* .

Вектор x^* — результат воспроизведения системой вектора сообщений x путем соответствующей обработки входной смеси $y(t)$. В векторном виде входную смесь можно записать так:

$$y(t) = u_c(t, x) + u_{ш}(t), \quad x = \langle x_1, \dots, x_n \rangle. \quad (11.66)$$

При этом алгоритм действия искомой системы воспроизведения сообщений можно представить в виде

$$x^*_i(t) = \Gamma_i(y), \quad (11.67)$$

где $\Gamma(y) = \langle \Gamma_1(y), \dots, \Gamma_j(y), \dots, \Gamma_n(y) \rangle$ — совокупность операций, проводимых системой над смесью $y(t)$ для получения вектора оценки x^* , т. е. оператор системы.

Векторной записи (11.67) в развернутом виде соответствует совокупность n скалярных операторов:

$$x^*_1(t) = \Gamma_1(y), \dots, x^*_j(t) = \Gamma_j(y), \dots, x^*_n(t) = \Gamma_n(y). \quad (11.68)$$

Будем искать $\Gamma(y)$ в классе *нерандомизированных* операторов, т. е. таких, которые каждой реализации вектора $y(t)$ ставят в соответствие вполне определенное значение вектора $x^*(t)$. Тогда справедливо соотношение

$$P(x, x^*) dx dx^* = P(x, y) dx dy,$$

где $x^* = \Gamma(y)$.

В соответствии с этим выражение (11.65) можно записать в виде

$$k'_n = \int \int (x^* - x)' B (x^* - x) P(x, y) dx dy, \quad (11.69)$$

где $x^* = \Gamma(y)$.

В соответствии с (9.9а) каждая нехудшая система находится из условия минимума величины k'_n при некотором конкретном значении весовых коэффициентов λ_{ij} , входящих в матрицу B , т. е. при некотором конкретном (фиксированном) значении матрицы B . При этом минимум находится вариацией структуры системы, т. е. в данном случае вариацией вида оператора $\Gamma(y)$ системы. В [23] показано, что минимум выражения (11.69) при вариациях оператора $\Gamma(y)$ имеет место при

$$\Gamma(y) = \Gamma^*(y) = \langle \Gamma^*_1(y), \dots, \Gamma^*_j(y), \dots, \Gamma^*_n(y) \rangle, \quad (11.70)$$

$$\text{где } \Gamma^*_j(y) = \frac{\int \dots \int x_j P(x_1, \dots, x_j, \dots, x_n; y) dx_1, \dots, dx_j, \dots, dx_n}{\int \dots \int P(x_1, \dots, x_j, \dots, x_n; y) dx_1, \dots, dx_j, \dots, dx_n}, \quad j = \overline{1, n}.$$

Так как

$$P(x, y) = P(x) P(y|x), \quad (11.71)$$

то при известном априорном распределении $P(x)$ воспроизводимых сообщений x и при известном распределении $W_{ш}(u_{ш})$ помехи $u_{ш}(t)$ все функции, входящие в (11.70), можно в принципе вычислить и, следовательно, найти и оператор $\Gamma^*(y)$, определяющий алгоритм действия нехудшей системы¹. Практическое вычисление в общем случае наталкивается на непреодолимые математические трудности, и обычно приходится вводить ряд упрощающих приближений. Однако уже сам вид оператора (11.70) позволяет сделать ряд весьма важных для практики выводов. Действительно, из (11.70) следует, что искомый оператор $\Gamma^*(y)$, т. е. алгоритм действия нехудшей системы, не зависит от матрицы B весовых коэффициентов λ_{ij} , а следовательно, и от значений этих коэффициентов. Отсюда, в свою очередь, следует (см. § 9.2), что при сформулированных исходных данных множество $M_{нх}$ нехудших систем оказывается вырожденным — содержит всего одну систему (класс систем), определяемую выражением (11.70).

Принцип действия этой системы иллюстрируется рис. 11.10. Система имеет n выходов $1, \dots, j, \dots, n$, на которых выдаются оптимальные оценки $x^*_1, \dots, x^*_j, \dots, x^*_n$ значения воспроизводимых сообщений $x_1, \dots, x_n$, содержащихся во входной смеси $y(t)$ сигнала и помех. Следовательно, в данном примере удастся най-

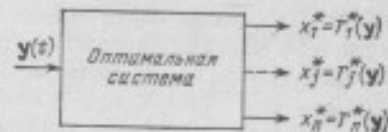


Рис. 11.10

¹ Здесь и далее распределение $P(x, y)$ считается таким, что решение вида (11.70) существует (в математическом смысле).

ти оптимальную систему, не прибегая к какому-либо условному критерию предпочтения: алгоритм (11.70) действия системы оказывается не только нехудшим, но и безусловно лучшим [в смысле совокупности всех m показателей качества вида (11.55)].

При этом, как отмечалось в § 9.1, найденная система (алгоритм ее действия) совместно-оптимальна, т. е. обеспечивает одновременный (совместный) минимум значений всех m показателей k_{ij} ¹. Это означает также, что при отыскании минимального значения данного показателя качества, например k_{11} , снижение требований ко всем остальным ($m-1$) показателям (вплоть до полного их игнорирования) не позволит улучшить (уменьшить) минимальное значение ($k_{11} \min$) данного показателя.

Пусть, например, будет признано, что из всех сообщений $x_1, \dots, x_n$ желательно воспроизводить лишь сообщение x_1 . Тогда в системе на рис. 11.10 вместо n выходов $x^*_1, \dots, x^*_n$ останется всего один выход $x^*_1 = \Gamma^*_1(y)$. При этом оптимальный алгоритм $x^*_1 = \Gamma^*_1(y)$ получения оценки останется неизменным [так как в соответствии с (11.70) он не зависит от конкретных требований к показателям качества], а следовательно, останется неизменным и качество этой оценки.

Таким образом, отказ от оценок некоторых из m показателей качества или снижение требований к ним могут упростить систему (в случае отказа — ликвидировать часть выходов системы), но не могут улучшить качество остальных оценок. Для того чтобы пояснить значение этого результата для синтеза систем воспроизведения сообщений (оптимальных по показателям помехоустойчивости), рассмотрим несколько конкретных случаев, когда входную смесь сигнала и помех можно представить в виде (11.66).

1. Все параметры $x_1, \dots, x_n$ сигнала являются полезными сообщениями (например, при передаче сообщений от n различных источников в n -канальной системе передачи информации). В этом случае полученный результат свидетельствует о том, что снижение требований к точности воспроизведения одних сообщений не позволит повысить точность воспроизведения других.

2. Совокупность $x_1, \dots, x_n$ параметров, входящих в выражение (11.66), состоят из двух групп параметров: «полезных» $x_1, \dots, x_j$, т. е. безусловно подлежащих воспроизведению, и паразитных (мешающих) $x_{j+1}, \dots, x_n$, не несущих никакой информации о полезных параметрах. В этом случае воспроизведение на выходе системы (приемника) значений $x^*_{j+1}, \dots, x^*_n$ параметров $x_{j+1}, \dots, x_n$ вообще говоря, не требуется — оно могло быть целесообразным, лишь если бы позволяло повысить точность воспроизведения полезных параметров. Однако алгоритм (11.70) обеспечивает наибольшую точность воспроизведения сообщений $x_1, \dots, x_j$ независимо от того, с какой точностью воспроизводятся параметры $x_{j+1}, \dots, x_n$. Следовательно, точность воспроизведения параметров $x_1, \dots, x_j$ не ухудшится, если даже полностью отказаться от воспроизведения мешающих параметров $x_{j+1}, \dots, x_n$ (и соответственно ликвидировать выходы $x^*_{j+1}, \dots, x^*_n$).

¹ Получаемые на основе алгоритма (11.70) совместно-оптимальные оценки в общем случае отличаются от совместно-эффективных оценок, выведенных Р. Фишером и Г. Крамером [7], так как при отыскании совместно-эффективных оценок вводятся несколько иные показатели качества оценок, а функция правдоподобия $P(y|x)$ полагается удовлетворяющей некоторым ограничивающим условиям. Однако можно показать, что в тех случаях, когда совместно-эффективные оценки существуют, они являются также и совместно-оптимальными.

т. е. ограничиться интегрированием по этим параметрам, производимым при получении оценок $x^*_1, \dots, x^*_j$ в соответствии с алгоритмом (11.70).

3. Смесь $y(t)$ на входе приемника имеет вид

$$y(t) = u_c(t, x(t)) + u_m(t), \quad (11.72)$$

где $x(t)$ — сообщение, являющееся случайной функцией времени. В соответствии с теоремой Котельникова можно приближенно полагать, что это сообщение полностью определяется своими n отсчетами $x_1, \dots, x_i, \dots, x_n$, где $x_i = x(i\Delta t)$; $n = 2F_n T \gg 1$; F_n — высшая частота в спектре сообщения; T — время его воспроизведения. При этом смесь (11.72) приводится к виду (11.66), в котором $x_1, \dots, x_n$ являются отсчетами сообщения $x(t)$, и, следовательно, оказываются справедливыми сформулированные ранее результаты. В частности, снижение требований к точности воспроизведения одной части отсчетов не позволит повысить точность воспроизведения другой части отсчетов. Например, в случае терминального управления объектами (например, ракетами) требуется обеспечить высокую точность воспроизведения сообщений лишь в конечный момент управления. При этом нецелесообразно (как это часто принимается в теории оптимальной фильтрации) требовать возможно более точного воспроизведения сообщения в каждый (а не только в конечный) момент времени, т. е. требовать минимальной погрешности воспроизведения всех отсчетов $x_1, \dots, x_n$, а не только конечного отсчета x_n . Однако при рассмотренных исходных данных, т. е. при существовании совместно-оптимальной системы (11.70), любое снижение требований к точности воспроизведения даже всех отсчетов, кроме x_n , не позволит повысить точность воспроизведения отсчета x_n — оно может привести лишь к упрощению системы.

Путем рассуждений, аналогичных приведенным в пп. 1, 2 и 3, нетрудно распространить результаты и на значительно более общий случай, когда векторный параметр сигнала $x = \langle x_1, \dots, x_n \rangle$ содержит компоненты, соответствующие искомым полезным сообщениям $x_1(t), \dots, x_M(t)$ и нескольким мешающим параметрам $x_{M+1}(t), \dots, x_N(t)$, $N \leq n$ (поскольку каждую из этих функций времени можно характеризовать некоторой совокупностью $x_1, \dots, x_n$ дискретных отсчетов, входящих в состав вектора x): если решение вида (11.70) существует, то отказ от оценки части случайных процессов (полезных или мешающих) или от оценки всех (или некоторых) этих процессов в любой момент времени (из заданного интервала $[0, T]$) не позволит повысить точность оценки любого временного отсчета любого из процессов $x_1(t), \dots, x_M(t)$ (такой отказ может привести лишь к упрощению системы).

Однако необходимо иметь в виду, что при изменении исходных данных по сравнению с принятыми в этом параграфе (например, при другом виде показателей качества (11.55)) множество $M_{\text{вх}}$ может оказаться невырожденным и, следовательно, совместно-оптимальная система — несуществующей. При этом сделанные здесь выводы будут несправедливыми.

11.4. ОСНОВНЫЕ ЭТАПЫ ПРОЕКТИРОВАНИЯ СПУТНИКОВОЙ СИСТЕМЫ СВЯЗИ

Спутниковая система связи предназначена для осуществления связи между различными пунктами пространства с помощью ретрансляторов, установленных на ИСЗ. В зависимости от

исходных данных связываемые пункты могут быть расположены целиком на поверхности Земли, целиком в атмосфере, в космосе или в нескольких из этих сред. В каждом из этих пунктов устанавливается обычно приемно-передающая связная радиостанция (одноканальная или многоканальная), а на спутниках — ретрансляторы, принимающие радиосигналы от корреспондентов и ретранслирующие эти сигналы другим корреспондентам. В простейшем случае ретрансляция сводится к усилению мощности входных сигналов и переносу их спектров на другие несущие частоты. Однако в ряде задач в ретрансляторе должна производиться более сложная обработка сигналов, чтобы уменьшить перекрестные помехи между сигналами от различных радиостанций и повысить помехоустойчивость системы. В общем случае для обеспечения требуемой связи между всеми пунктами (корреспондентами) ретрансляторы приходится размещать на нескольких спутниках, вращающихся на различных орбитах.

Спутниковые системы связи различают в первую очередь по степени глобальности и универсальности обслуживания корреспондентов. Степень глобальности ССС характеризуется размером зоны обслуживания, а универсальности ССС — набором типов корреспондентов (самолеты, космические аппараты и др.) и числом видов предоставляемой связи: телефонная, телеграфная и др. ССС различаются существенно и по организации связи. В связи с таким разнообразием ССС их структура также может быть различной. Однако, как правило, ССС состоит из следующих основных частей (подсистем):

ИСЗ, содержащие ретрансляторы и вспомогательное оборудование (устройства питания, управления и др.);

ракетно-космический комплекс, включающий ракеты-носители и другое оборудование, необходимое для вывода спутников на требуемую орбиту (а в случае геостационарной орбиты — и в заданную точку этой орбиты);

станции спутниковой связи, содержащие радиоприемные, радиопередающие и другие устройства, обеспечивающие связь через ретрансляторы между заданными пунктами пространства;

командно-измерительный комплекс, осуществляющий контроль за спутниками и коррекцию их орбит, контроль исправности аппаратуры, замену отказавших спутников и другие операции;

система управления связью, обеспечивающая управление каналами и направлениями связи для более эффективного использования пропускной способности системы;

прочие устройства и системы (наведение антенн на ИСЗ, служебная связь и др.).

Спутниковая система связи, даже не являющаяся глобальной и универсальной, принадлежит к классу весьма сложных и дорогих радиосистем. Ее проектирование ведется большим коллективом специалистов разных профилей, работающих в различных организациях. Процесс проектирования такой системы весьма

сложен и длителен. Поэтому остановимся лишь весьма кратко на его основных этапах.

Проектирование начинается с обоснования исходных данных, выполняемого совместно заказчиками и исполнителями. Исходные данные, как отмечалось в гл. 8, включают совокупность условий работы системы, ограничений, накладываемых на структуру системы и значения ее параметров, состав показателей качества и предъявляемые к ним требования. Для столь сложной системы, как ССС, эти данные весьма многообразны и включают сотни различных параметров, показателей, ограничений и т. п. К ним, в частности, относятся:

направления связи (из каких пунктов в какие должна быть обеспечена связь);

требуемые пропускные способности по каждому направлению; виды связи (телефонная, телеграфная, телевизионная, дуплексная, симплексная и т. д.) по каждому направлению;

ограничения на организацию связи (например, допустим или нет отказ от закрепленных каналов и использование свободного доступа, должна ли связь быть радиальной, кольцевой, радиально-кольцевой и др.);

требования к показателям помехозащищенности, надежности, ЭМС, ЭКС, массе, габаритным размерам, стоимости и др.;

срок, отводимый на проектирование системы, ее производство и размещение в пространстве.

Исходные данные в процессе дальнейшего проектирования подвергаются корректировке, если выясняется их противоречивость (например, система оказывается слишком дорогой или не может быть спроектирована и реализована в требуемые сроки).

Последующие этапы проектирования, как отмечалось в § 1.2, включают определение принципов построения системы, аппаратный синтез, конструирование, разработку технологии изготовления и разработку испытательной аппаратуры. Ради краткости остановимся лишь на первом из этих этапов. При этом будем полагать, что все станции должны быть расположены неподвижно в заданных пунктах поверхности Земли, и будем называть их земными станциями (ЗС).

К принципам построения такой ССС в первую очередь относятся:

число спутников на геостационарной орбите и их вид — универсальные или специализированные. Если спутники универсальные, то они идентичны и каждый из них обеспечивает все требуемые виды связей. Если спутники специализированные, то они неидентичны и могут различаться по обслуживаемым направлениям связи, видам связи, используемым диапазонам волн и другим характеристикам;

число спутников, расположенных на других орбитах, параметры этих орбит и типы спутников;

виды сетей связи (если они заранее не заданы);

типы ЗС (центральная, узловая, периферийная) и число ЗС каждого типа;

используемые диапазоны волн и их распределение внутри ССС; принципы пространственно-временной обработки информации, включая принципы уплотнения — разделения станций (т. е. многостанционного доступа) и каналов, модуляции и демодуляции, кодирования и декодирования. При этом разделение станций может осуществляться как путем пространственной селекции, так и разделением сигналов по времени, частоте или форме;

принципы обеспечения надежности;

принципы обеспечения ЭМС.

Выбор и обоснование всех указанных принципов можно производить методом последовательных приближений, например в следующей последовательности. Сначала определяют минимальное число спутников, которые необходимо разместить на геостационарной и эллиптических орбитах для того, чтобы обеспечить географическое перекрытие всех заданных пунктов связи, т. е. создать условия прямой видимости с каждого спутника на соответствующие географические пункты в течение необходимых интервалов времени.

На втором этапе число спутников на каждой орбите уточняют, принимая во внимание заданную пропускную способность на каждом направлении связи в соответствии с уравнениями радиосвязи. При этом для обеспечения суммарной пропускной способности ССС могут потребоваться такие большие значения мощностей бортовых передатчиков и полос пропускания, что их не удастся реализовать при выбранном на первом этапе минимальном числе спутников (вследствие ограниченности массы и габаритных размеров спутников). В результате может оказаться необходимым увеличить число спутников.

Чтобы в следующем приближении уточнить требуемое число спутников, необходимо, кроме прочего, выбрать вид спутников (индивидуальные или специализированные), используемые в них диапазоны волн и виды диаграмм направленности антенных систем. Поскольку на столь раннем этапе проектирования такой выбор сделать однозначно, как правило, не удастся, приходится дальнейшее проектирование вести одновременно для нескольких вариантов и лишь на конечных этапах проектирования сократить число вариантов до одного-двух. При этом учитывается лишь, небольшая часть показателей качества, включая показатели пропускной способности, помехозащищенности, ЭМС и надежности и ограничения энергетических параметров ретрансляторов и земных станций, причем помехоустойчивость определяется только по отношению к внутренним шумам приемных устройств, а требования ЭМС — лишь при выборе диапазонов волн и ограничений максимальных мощностей излучения бортовых и наземных передающих устройств. Требования надежности учитываются также лишь в первом приближении — при оценке массы и габаритных

размеров спутников принимается во внимание необходимость по меньшей мере дублирования большинства бортовых устройств.

На третьем этапе уточняются принципы пространственно-временной обработки сигналов на всех направлениях связи, в первую очередь исходя из требований обеспечения заданной пропускной способности в условиях действия не только внутреннего шума, но и всех прочих существенных помех, а также требования скрытности действия системы.

На четвертом этапе более полно учитывают требования надежности. В частности, определяют запас спутников и ракет-носителей, необходимый для обеспечения формирования системы (при выводе спутников на орбиту) и ее последующего функционирования в течение заданного срока эксплуатации.

Стоимость (приведенные затраты) системы на первых четырех этапах проектирования количественно обычно не учитывается и заменяется более общим качественным понятием «сложность» (т. е. отбрасываются лишь заведомо слишком сложные варианты). Оптимизация системы на этих этапах обычно производится путем сочетания следующих методов:

математический синтез и широкое использование потенциальных характеристик (диаграмм обмена);

весьма приближенные аналитические методы анализа;

анализ и синтез с помощью ЭВМ в режиме диалога «человек — ЭВМ».

На пятом этапе принципы построения системы уточняются с учетом показателя стоимости (приведенных затрат), т. е. производится технико-экономическая оптимизация системы — сравниваются по стоимости различные варианты построения системы при значении параметров каждого варианта, оптимальных с точки зрения стоимости. При этом все остальные показатели качества системы должны оставаться в допустимых исходными данными пределах. В результате, во-первых, уменьшается число вариантов, оставляемых для дальнейшего проектирования, и, во-вторых, уточняются оптимальные значения параметров системы для каждого варианта.

На этом основной этап определения принципов построения системы завершается. Однако эти принципы могут корректироваться по результатам последующих этапов (аппаратурного синтеза, конструирования, разработки технологии производства и контрольно-испытательной аппаратуры).

Литература: основная [5, 58], дополнительная [2, 23].

ПРИЛОЖЕНИЕ 1

При использовании метода геометрического программирования минимизируемую функцию (принадлежащую к классу полиномов) следует представить в виде суммы

$$B_n = U_1 + U_2 + \dots + U_n,$$

(П.1)

где каждое слагаемое имеет вид

$$U_i = c_i x_1^{a_{i1}}, \dots, x_m^{a_{im}}, \quad (П.2)$$

m — число варьируемых переменных.

В нашем случае, как следует из (7.103)

$$\left. \begin{aligned} n=4, \quad U_1 &= c_1 x_1^{a_1} x_2^{a_2} x_3^{a_3}, \quad U_2 = c_2 x_1^{a_1} x_2^{a_2} x_3^{a_3}, \\ U_3 &= c_3 x_1^{a_1} x_2^{a_2} x_3^{a_3}, \quad U_4 = c_4 x_1^{-a_4} x_2^{-a_4} x_3^{-a_4}. \end{aligned} \right\} \quad (П.3)$$

По применяемой в теории геометрического программирования терминологии [63] эта задача имеет степень трудности $n - (m+1) = 4 - (3+1) = 0$. Ее решение таково:

$$B_{\text{мин}} = \prod_{i=1}^n \left(\frac{c_i}{\delta_i} \right)^{\delta_i} \quad (П.4)$$

(здесь роль c_i играет коэффициент c'_i), где $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$ — так называемый двойственный вектор. Для его отыскания составляется матрица показателей степени переменных, входящих в слагаемые U_i ($i=1, n$). В данном случае из (П.3) следует, что эта матрица содержит $m=3$ столбцов и $n=4$ строк и имеет вид

$$\begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \\ -a_4 & -a_4 & -a_4 \end{bmatrix}.$$

Ее нужно преобразовать так, чтобы в ее верхней части образовалась единичная матрица $m \times m$. Для этого достаточно разделить все элементы 1-го столбца на a_1 , 2-го — на a_2 и 3-го — на a_3 . Тогда получится матрица вида

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_4/a_1 & -a_4/a_2 & -a_4/a_3 \end{bmatrix} \begin{matrix} a_4/a_1 \\ a_4/a_2 \\ a_4/a_3 \\ 1 \end{matrix}$$

(справа этой матрицы, т. е. в 4-м столбце, записан ортогональный вектор, смысл которого разъясняется далее).

Искомый вектор $\delta = (\delta_1, \delta_2, \delta_3, \delta_4)$ должен удовлетворять двум условиям: быть ортогонален любому вектору-столбцу матрицы показателей; быть нормирован так, что

$$\sum_{i=1}^n \delta_i = 1. \quad (П.5)$$

Вектор, удовлетворяющий первому условию, указан в столбце справа от матрицы показателей (т. е. в 4-м столбце). Он получается путем сочетания двух приемов: продолжением вправо вниз главной диагонали единичной матрицы показателей и симметричным относительно этой диагонали отображением других элементов матрицы показателей (с переменной знака).

Сумма элементов найденного ортогонального вектора

$$\mu = 1 + a_4/a_1 + a_4/a_2 + a_4/a_3.$$

Для того чтобы этот вектор удовлетворял условию нормировки (П.5), все его элементы следует поделить на μ . Поэтому искомый двойственный вектор

$$\delta = \begin{bmatrix} a_4/a_1 \mu \\ a_4/a_2 \mu \\ a_4/a_3 \mu \\ 1/\mu \end{bmatrix}$$

и искомый минимум функции $B_{\text{мин}}$ в соответствии с (П.4)

$$B_{\text{мин}} = \left(\frac{c_1 a_1 \mu}{a_4} \right)^{\frac{a_4}{a_1 \mu}} \left(\frac{c_2 a_2 \mu}{a_4} \right)^{\frac{a_4}{a_2 \mu}} \left(\frac{c_3 a_3 \mu}{a_4} \right)^{\frac{a_4}{a_3 \mu}} (c'_4 \mu)^{1/\mu},$$

что совпадает с (7.105).

Оптимальные значения переменных x^*_1 , x^*_2 и x^*_3 находятся из условия

$$U_i^*/U_j^* = \delta_i/\delta_j.$$

В данном случае эти условия имеют вид

$$\frac{U_1^*}{U_2^*} = \frac{c_1 (x_1^*)^{a_1}}{c_2 (x_2^*)^{a_2}} = \frac{a_4/a_1 \mu}{a_4/a_2 \mu} = \frac{a_2}{a_1},$$

$$\frac{U_1^*}{U_3^*} = \frac{c_1 (x_1^*)^{a_1}}{c_3 (x_3^*)^{a_3}} = \frac{a_4/a_1 \mu}{a_4/a_3 \mu} = \frac{a_3}{a_1},$$

$$\frac{U_1^*}{U_4^*} = \frac{c_1 (x_1^*)^{a_1}}{c'_4 (x_1^*)^{-a_4} (x_2^*)^{-a_4} (x_3^*)^{-a_4}} = \frac{a_4/a_1 \mu}{1/\mu} = \frac{a_4}{a_1}.$$

Решая систему из этих трех уравнений, нетрудно получить выражения для x^*_1 , x^*_2 и x^*_3 (при этом $x^*_4 = A_1/x^*_1 x^*_2 x^*_3$), которые после соответствующих преобразований приводятся к более удобному для вычислений виду (7.106).

ПРИЛОЖЕНИЕ 2

Для целевой функции (7.125), содержащей всего две переменные (x_1 и x_2) и три слагаемых, матрица показателей степени имеет вид

$$\begin{bmatrix} a_1 & 0 \\ 0 & a_2 \\ -a_3 & -a_3 \end{bmatrix}.$$

После деления элементов 1-го столбца на a_1 и 2-го на a_2 она преобразуется следующим образом:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -a_3/a_1 & -a_3/a_2 \end{bmatrix} \begin{matrix} a_3/a_1 \\ a_3/a_2 \\ 1 \end{matrix}.$$

Здесь в 3-м столбце записан вектор, ортогональный к обоим столбцам матрицы. Сумма его элементов

$$\mu' = 1 + a_3/a_1 + a_3/a_2.$$

Двойственный вектор δ получается из ортогонального делением всех его элементов на μ' :

$$\delta = \begin{bmatrix} a_3/a_1 \mu' \\ a_3/a_2 \mu' \\ 1/\mu' \end{bmatrix}.$$

Поэтому искомый минимум функции (7.125)

$$B'_{\text{мин}} = \prod_{i=1}^3 \left(\frac{c_i}{\delta_i} \right)^{\delta_i} = \left(\frac{c_1 a_1 \mu'}{a_3} \right)^{\frac{a_3}{a_1 \mu'}} \left(\frac{c_2 a_2 \mu'}{a_3} \right)^{\frac{a_3}{a_2 \mu'}} \left(c_3 \mu' \right)^{\frac{1}{\mu'}}.$$

[здесь роль c_3 выполняет коэффициент c'_3 при третьем слагаемом функции (7.125)].

Экстремальные значения переменных x^*_{11} и x^*_{21} находятся из уравнений $U^*_{11}/U^*_{21} = \delta_1/\delta_2$, $U^*_{11}/U^*_{31} = \delta_1/\delta_3$.

Так как $U^*_1 = c_1 (x^*_{11})^{a_1}$, $U^*_2 = c_2 (x^*_{21})^{a_2}$, $\delta_1 = a_3/a_1 \mu'$, $\delta_2 = a_3/a_2 \mu'$ и $\delta_3 = 1/\mu'$, а

$$B'_{\text{мин}} = U^*_1 + U^*_2 + U^*_3 = U^*_1 (1 + \delta_2/\delta_1 + \delta_3/\delta_1) = U^*_1 \gamma',$$

где $\gamma' = 1 + a_1/a_2 + a_1/a_3$, то получается $x^*_{11} = (B'_{\text{мин}}/\gamma' c_1)^{1/a_1}$, $x^*_{21} = (B'_{\text{мин}}/\gamma' c_2 a_2)^{1/a_2}$.

При этом, поскольку $x^*_{31} = x_{31 \text{ мин}}$, из уравнения дальности следует, что

$$x^*_{31} = A (x^*_{11} x^*_{21} x^*_{31})^{-1}.$$

Эти результаты соответствуют формуле (7.127).

СПИСОК ЛИТЕРАТУРЫ

1. Гуткин Л. С. Современная радиоэлектроника и ее проблемы. — 2-е изд., — М.: Сов. радио, 1980.
2. Основы технического проектирования систем связи через ИСЗ/А. Д. Фортуненко, Г. В. Аскинази, В. Л. Быков и др. — М.: Связь, 1970. — 331 с.
3. Мановцев А. П. Основы теории радиотелеметрии. — М.: Энергия, 1973. — 592 с.
4. Конторов Д. С., Голубев-Новожилов Ю. С. Введение в радиолокационную системотехнику. — М.: Сов. радио, 1971. — 367 с.
5. Гуткин Л. С. Теория оптимальных методов радиоприема при флуктуационных помехах. — М.: Сов. радио, 1972. — 441 с.
6. Системы связи и радиорелейные линии/Под ред. Н. И. Калашникова. — М.: Связь, 1977. — 392 с.
7. Крамер Г. Математические методы статистики: Пер. с англ./Под ред. А. Н. Колмогорова. — 2-е изд., — М.: Мир, 1975. — 648 с.
8. Энциклопедия кибернетики/Под ред. В. М. Глушкова и др. Т. I. — Киев: Гл. ред. Укр. сов. энциклопедии, 1974. — 607 с.
9. Теория автоматического регулирования. Кн. 2/Под ред. В. В. Солодовникова. — М.: Машиностроение, 1967. — 679 с.
10. Бобнев М. П., Кривяцкий Б. Х., Ярлыков М. С. Комплексные системы радиоавтоматики. — М.: Сов. радио, 1968. — 232 с.
11. Гуткин Л. С., Тилугин В. Н., Пестряков В. Б. Радиоуправление. — М.: Сов. радио, 1970. — 324 с.
12. Гришин Ю. П., Казаринов Ю. М., Катиков В. М. Микропроцессоры в радиотехнических системах. — М.: Радио и связь, 1982. — 279 с.

13. Помехозащищенность радиосистем со сложными сигналами/Под ред. Г. И. Тузова. — М.: Радио и связь, 1985. — 264 с.
14. Котельников В. А. Теория потенциальной помехоустойчивости. — М.: Энергоиздат, 1956. — 151 с.
15. Варакин Л. Е. Системы связи с шумоподобными сигналами. — М.: Радио и связь, 1985. — 384 с.
16. Шумоподобные сигналы/Под ред. В. Б. Пестрякова. — М.: Сов. радио, 1973. — 424 с.
17. Чердынец В. А. Проектирование радиотехнических систем со сложными сигналами. — Минск: Высшая школа, 1979. — 192 с.
18. Теоретические основы радиолокации/Под ред. Я. Д. Ширмана. — М.: Сов. радио, 1970. — 560 с.
19. Пространственно-временная обработка сигналов/Под ред. И. Я. Кремера. — М.: Радио и связь, 1983. — 22 с.
20. Фалькович С. Е. Оценка параметров сигналов. — М.: Сов. радио, 1970. — 334 с.
21. Ширман Я. Д. Разрешение и сжатие сигналов. — М.: Сов. радио, 1974. — 360 с.
22. Радиоуправление реактивными снарядами и космическими аппаратами/Л. С. Гуткин, Ю. П. Борисов, А. А. Валуев и др. — М.: Сов. радио, 1968. — 679 с.
23. Вопросы статистической теории радиолокации/Под ред. Г. П. Тартаковского. В 2-х т. — М.: Сов. радио, Т. 1, 1963. — 424 с. Т. 2, 1964. — 1079 с.
24. Перачев С. В., Валуев А. А., Чиликин В. М. Статистическая динамика сложных систем. — М.: Сов. радио, 1973. — 488 с.
25. Математические основы современной радиоэлектроники/Под ред. Л. С. Гуткина. — М.: Сов. радио, 1968. — 206 с.
26. Максимов М. В., Горгонов Г. И. Радиоэлектронные системы самонаведения. — М.: Сов. радио, 1982. — 304 с.
27. Прикладные математические методы анализа в радиотехнике/Под ред. Г. В. Обрезкова. — М.: Высшая школа, 1985. — 343 с.
28. Казаков И. Б. Статистическая теория систем управления в пространстве состояний. — М.: Наука, 1975. — 432 с.
29. Тихонов В. И., Кульман Н. К. Нелинейная фильтрация в квазикогерентный прием сигналов. — М.: Сов. радио, 1975. — 304 с.
30. Сосулин Ю. Г. Теория обнаружения и оценивания стохастических сигналов. — М.: Сов. радио, 1978. — 320 с.
31. Левин Б. Р. Теоретические основы статистической радиотехники. Кн. 2-я. — М.: Сов. радио, 1975. — 391 с.
32. Князев А. Д. Элементы теории и практики обеспечения электромагнитной совместимости радиоэлектронных средств. — М.: Радио и связь, 1984. — 336 с.
33. Электромагнитная совместимость радиоэлектронных средств и систем/Под ред. Н. М. Царькова. — М.: Радио и связь, 1985. — 280 с.
34. Теория обнаружения сигналов/Под ред. П. А. Бакута. — М.: Радио и связь, 1984. — 302 с.
35. Репин В. Г., Тартаковский Г. П. Статистический синтез при априорной неопределенности и адаптации информационных систем. — М.: Сов. радио, 1977. — 432 с.
36. Стратонович Р. Л. Принципы адаптивного приема. — М.: Сов. радио, 1973. — 141 с.
37. Шахгильдян В. В., Лохвицкий М. С. Методы адаптивного приема сигналов. — М.: Связь, 1974. — 159 с.
38. Ципкин Я. З. Адаптация и обучение в автоматических системах. — М.: Наука, 1968. — 399 с.
39. Ципкин Я. З. Основы теории обучающихся систем. — М.: Наука, 1970. — 251 с.
40. Ширман Я. Д., Манжос В. Н. Теория и техника обработки радиолокационной информации на фоне помех. — М.: Радио и связь, 1981. — 416 с.
41. Кловский Д. Д., Сойфер В. А. Обработка пространственно-временных сигналов. — М.: Связь, 1976. — 207 с.

42. Вакин С. А., Шустов Л. Н. Основы радиопротиводействия и радиотехнической разведки. — М.: Сов. радио, 1968. — 444 с.
43. Защита от радиопомех/Под ред. М. В. Максимова. — М.: Сов. радио, 1976. — 495 с.
44. Сосулин Ю. Г. Методы оптимальной обработки сигналов на фоне комплекса помех. — Радиотехника и электроника, 1982, № 6, с. 1171—1181.
45. Сосулин Ю. Г. Оптимальное комплексирование измерителей. — В кн.: Эффективность применения цифровых устройств в радиолокации/МАИ. — М., 1982, с. 4—16.
46. Березин Л. В., Вейцель В. А. Теория и проектирование радиосистем. — М.: Сов. радио, 1977. — 448 с.
47. Шеннон К. Работы по теории информации и кибернетике: Пер. с англ. — М.: ИЛ, 1963. — 829 с.
48. Тепляков И. М., Калашников И. Д., Рошин Б. Р. Радиолинии космических систем передачи информации. — М.: Сов. радио, 1975. — 390 с.
49. Возенкрафт Д., Джекобс И. Теоретические основы техники связи: Пер. с англ. — М.: Мир, 1969. — 640 с.
50. Справочник по радиолокации в 4-х т./Под ред. Р. Сокольника. — М.: Сов. радио, 1975—1978.
51. Книзев А. Д., Пчелкин В. Ф. Проблемы обеспечения совместной работы радиоэлектронной аппаратуры. — М.: Сов. радио, 1971. — 200 с.
52. Уайт Р. Ж. Электромагнитная совместимость радиоэлектронных средств и непреднамеренные помехи: Пер. с англ. — М.: Сов. радио, вып. 1, 1977. — 348 с.; вып. 2, 1978. — 272 с.
53. Пестряков В. Б., Кузнецов В. Д. Радиотехнические системы. — М.: Радио и связь, 1985. — 376 с.
54. Перротте А. И., Сторчак М. А. Вопросы надежности РЭА. — М.: Сов. радио, 1976. — 185 с.
55. Надежность технических систем: Справочник/Под ред. И. А. Ушакова. — М.: Радио и связь, 1985. — 608 с.
56. Левин Б. Р. Теория надежности радиотехнических систем. — М.: Сов. радио, 1978. — 263 с.
57. Губонин Н. С. Среднее число замен однотипного оборудования с усеченным экспоненциальным распределением вероятности времени его безотказной работы. — Автоматика и телемеханика, 1982, № 3, с. 170, 171.
58. Гуткин Л. С. Оптимизация радиоустройств по совокупности показателей качества. — М.: Сов. радио, 1975. — 367 с.
59. Ефимов К. А., Львов Л. С. Эффективность новой техники. — М.: Экономика, 1979. — 144 с.
60. Оптимизация функционирования социалистической экономики/Под ред. С. С. Шаталина. — М.: МГУ, 1980. — 288 с.
61. Юрлов Ф. Ф. Техно-экономическая эффективность сложных радиоэлектронных систем. — М.: Сов. радио, 1980. — 278 с.
62. Гуткин Л. С. Потенциальная стоимость радиолиний извлечения информации. — Изв. вузов. Радиотехника, 1980, № 6, с. 55—60.
63. Зенер К. Геометрическое программирование и техническое проектирование. — М.: Мир, 1973. — 108 с.
64. Богомолов А. Ф., Соколов А. Г., Попереченко Б. А. и др. О проблеме создания комплекса современных экономических радиотелескопов. — В кн.: Антенны/Под ред. А. А. Пистолькорса. — М.: Связь, 1976, вып. 24, с. 106—123.
65. Окунев Ю. Б., Плотников В. Г. Принципы системного подхода к проектированию в технике связи. — М.: Связь, 1976. — 183 с.
66. Талызин Н. В., Кантор Л. Я., Манякин С. А., Паянский Ю. М. Об оптимальных параметрах и экономической эффективности систем спутниковой связи. — Радиотехника, 1969, № 11, с. 5—13.
67. Петров А. В., Яковлев А. А. Анализ и синтез радиотехнических комплексов. — М.: Радио и связь, 1984. — 246 с.

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	4
Глава 1. Общая характеристика задач и методов проектирования	9
1.1. Основные виды задач проектирования	9
1.2. Основные этапы проектирования	10
1.3. Организация проектирования	12
1.4. Общая характеристика методов проектирования	13
1.5. Системный подход как основа проектирования	17
Глава 2. Точность воспроизведения сообщений	20
2.1. Основные характеристики воспроизводимых сообщений	20
2.2. Показатели точности воспроизведения дискретных сообщений	21
2.3. Показатели точности воспроизведения непрерывных сообщений, постоянных в процессе их воспроизведения	24
2.4. Показатели точности воспроизведения сообщений, изменяющихся во времени	30
2.5. Источники ошибок	36
2.6. Суммирование ошибок, вызванных различными источниками	40
2.7. Общая характеристика путей уменьшения ошибок	45
2.8. Принцип компенсации	47
2.9. Принцип инвариантности	51
2.10. Принцип комплексирования	56
Глава 3. Помехозащищенность	57
3.1. Общая характеристика показателей помехозащищенности и путей их улучшения	57
3.2. Синтез оптимальной по помехоустойчивости временной обработки (пары «сигнал — фильтр»)	62
3.3. Синтез помехоустойчивой пространственно-временной обработки в приемнике	66
3.4. Синтез помехоустойчивой временной обработки в приемнике	70
3.5. Особенности синтеза следящих устройств	74
3.6. Особенности синтеза систем радиуправления	77
3.7. Особенности синтеза помехоустойчивых систем и устройств при неполных априорных данных	84
Глава 4. Пропускная способность радиосистем	92
4.1. Вводные замечания	92
4.2. Определение пропускной способности СПИ в классической теории информации	94
4.3. Потенциальная пропускная способность СПИ	102
4.4. Реальная пропускная способность СПИ	103
4.5. Потенциальная пропускная способность систем извлечения информации	110
Глава 5. Электромагнитная и экологическая совместимость	118
5.1. Общая характеристика проблемы электромагнитной совместимости	118
5.2. Помехи от радиостанций	120
5.3. Индустриальные помехи	122
5.4. Внутрисистемные помехи второго вида	123
5.5. Основные пути обеспечения ЭМС	124
5.6. Нормы и рекомендации по ограничению уровня помех	126
5.7. Показатели и параметры ЭМС	127
5.8. Расчет пространственного и частотного разброса, необходимого	

для обеспечения ЭМС	128
5.9. Анализ ЭМС	130
5.10. Экологическая совместимость радиосистем	131
Глава 6. Надежность	133
6.1. Вероятность безотказной работы системы	133
6.2. Особенности определения надежности систем с восстановлением	135
6.3. Срок службы системы	136
6.4. Основные пути повышения надежности	137
6.5. Прием-сдаточные испытания на надежность	142
6.6. Пример расчета надежности функционирования спутниковой системы связи	142
Глава 7. Стоимость, масса и объем аппаратуры	148
7.1. Введение	148
7.2. Определение приведенных затрат	150
7.3. Общая характеристика влияния случайности затрат	157
7.4. Определение коэффициентов доверия	160
7.5. Выбор доверительной вероятности	163
7.6. Влияние случайности неизменной (одинаковой для всех вариантов) составляющей затрат	166
7.7. Пути уменьшения стоимости	167
7.8. Постановка задачи определения потенциальной стоимости	173
7.9. Определение потенциальных стоимостей БПС и УПС-1	178
7.10. Методика определения потенциальных стоимостей с учетом ограничений (УПС-2 и УПС-3)	182
7.11. Области применения полученных результатов	185
7.12. Габаритные размеры аппаратуры	186
7.13. Масса аппаратуры	188
7.14. Пути уменьшения массы и габаритных размеров аппаратуры	189
Глава 8. Особенности учета совокупности показателей качества	191
8.1. Общая характеристика исходных данных для проектирования	191
8.2. Виды критериев качества	199
Глава 9. Применение безусловного и условного критериев предпочтения	211
9.1. Нехудшие и худшие системы. Диаграммы обмена	211
9.2. Методы отыскания нехудших систем	216
9.3. Применение условного критерия предпочтения	228
Глава 10. Потенциальные характеристики радиосистем — диаграммы обмена	233
10.1. Одномерные потенциальные характеристики радиосистем	233
10.2. Многомерные потенциальные характеристики — диаграммы обмена	236
10.3. Примеры многомерных диаграмм обмена	239
10.4. Методы уменьшения размерности диаграммы обмена	245
10.5. Уменьшение размерности диаграммы обмена путем ее преобразования в обобщенную диаграмму обмена	248
10.6. Связь диаграммы обмена системы с диаграммами обмена ее подсистем	250
Глава 11. Примеры оптимизации радиоустройств и радиосистем	259
11.1. Оптимизация обнаружителя сигнала по двум и по трем показателям качества	259
11.2. Оптимизация устройства воспроизведения дискретных сообщений по m показателям качества	264
11.3. Оптимизация системы воспроизведения непрерывных сообщений по m показателям качества	273
11.4. Основные этапы проектирования спутниковой системы связи	277
Приложение 1	281
Приложение 2	283
Список литературы	284