

Д В. Гаскаров,
А А. Дахнович

**ОПТИМИЗАЦИЯ
ТЕХНОЛОГИЧЕСКИХ
ПРОЦЕССОВ
В ПРОИЗВОДСТВЕ
ЭЛЕКТРОННЫХ
ПРИБОРОВ**

Допущено
Министерством высшего
и среднего специального
образования СССР
в качестве учебного пособия
для студентов вузов,
обучающихся по специальности
«Промышленная электроника»

Москва «Высшая школа» 1986



ББК 32.85
Г22
УПК 621.38

Рецензенты:

кафедра электронного машиностроения и робототехники Московского института электронного машиностроения (зав. кафедрой проф., д-р техн. наук А. Т. Александрова), гл. инженер КБ НПО «Светлана» канд. техн. наук Р. Я. Рощин

Гаскаров Д. В., Дахнович А. А.

Г22 Оптимизация технологических процессов в производстве электронных приборов: Учеб. пособие для студентов вузов спец. «Промышленная электроника». — М.: Высш. шк., 1986. — 191 с.: ил.

В книге изложены методологические основы оптимизации технологических процессов производства изделий электронной техники (в том числе процесса испытаний), основы теории планирования эксперимента, рассмотрены регрессионный анализ, ортогональное планирование, метод Бокса — Вильсона, симплекс- и эволюционное планирование, оптимизация свойств многокомпонентных материалов и др.

Г $\frac{240300000-266}{001(01)-86}$ 135-86

ББК 32.85
6Ф0.3

© Издательство «Высшая школа», 1986

ПРЕДИСЛОВИЕ

Методы оптимизации как в теоретическом, так и в практическом плане получили интенсивное развитие и в приложении к широкому внедрению вычислительной техники на этапах проектирования производства, испытаний и эксплуатации изделий. В системах автоматизированного проектирования (САПР), автоматизированного управления технологическими процессами (АСУ ТП) и технического диагностирования (СТД) в той или иной форме решаются разнообразные задачи оптимизации, так как получение оптимального результата позволяет достигнуть существенного экономического эффекта. Особенно роль алгоритмов и программ оптимизации возросла с разработкой гибких автоматизированных производств (ГАП).

Разработка проблемы оптимизации началась в 50-х годах, большой вклад в нее внесли работы многих советских и зарубежных ученых, среди которых можно отметить В. А. Бункина, Г. К. Круга, Л. В. Канторовича, Е. В. Маркову, В. В. Налимова, А. А. Первозванского, В. В. Федорова, Д. Данцига, Дж. Бокса, К. Вальсона и др.

Книга представляет собой введение в теорию и практику методов оптимизации технологических процессов (ТП) производства изделий электронной техники. Основное внимание уделено изучению объекта оптимизации — ТП в целом или отдельных его операций, выбору метода оптимизации, построению модели процесса и ее анализу. Рассмотренные методы иллюстрируются практическими примерами.

В основу учебного пособия положен материал, который читается авторами в Ленинградском электротехническом институте им. В. И. Ульянова (Ленина) по курсам «Технология электронных приборов», «Оптимизация технологических процессов». В небольшом по объему пособии не представляется возможным рассмотреть все методы оптимизации ТП и особенности их применения.

ГЛАВА 1. СУЩНОСТЬ ЗАДАЧИ ОПТИМИЗАЦИИ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА

§ 1.1. ОБЩИЕ ВОПРОСЫ ПРОБЛЕМЫ ОПТИМИЗАЦИИ ТП

Задача оптимизации ТП. Критерии оптимальности. Под *оптимизацией* ТП понимают комплекс мероприятий, направленный на достижение экстремального уровня (и дальнейшее поддержание на этом уровне) некоторого обобщенного количественного показателя качества ТП. Обобщенный количественный показатель качества ТП называют *критерием оптимальности, параметром оптимизации или целевой функцией*.

Для одного и того же ТП могут ставиться как частные задачи оптимизации, связанные с оптимизацией отдельных технологических операций или даже с их этапами, так и более общие задачи, основанные на оптимизации некоторой группы взаимосвязанных операций или всего ТП в целом. Выбираемые при этом критерии оптимальности могут отличаться большим разнообразием. Различают две группы критериев: технические, которые для изделий электронной техники (ИЭТ) связаны с электрическими параметрами, конструкцией или надежностными характеристиками выпускаемых изделий, и экономические, зависящие от стоимости выпускаемых изделий, производительности труда или эффективности использования технологического оборудования.

Минимизация токов утечки в полупроводниковой интегральной микросхеме (ИМС), разброса номиналов резисторов, отклонений от заданного профиля легирования, частоты отказов ИЭТ, максимизация напряжения пробоя транзистора — примеры задач оптимизации с использованием технического критерия оптимальности.

Минимизация процента брака, затрат на единицу выпускаемой продукции, содержания драгоценных металлов в выпускаемых приборах, длительности испытаний, максимизация процента выхода годных изделий, процента при-

ВВЕДЕНИЕ

Научно-технический прогресс в электронном приборостроении привел к всестороннему усложнению изделий электронной техники (ИЭТ) и особенно технологии их производства. С усложнением ТП производства возникла проблема их оптимизации, так как при сравнительно простых ТП для достижения оптимальных условий их протекания было достаточно интуиции и опыта технолога. Появилась необходимость определения многомерных количественных зависимостей выходных показателей ИЭТ от режимов и условий производства или оценки их качественного влияния на параметры приборов.

Успешное решение проблемы оптимизации ТП предполагает рассмотрение ее с двух точек зрения: 1) анализ физической картины явлений, лежащих в основе процессов и объектов оптимизации (фундаментальная сторона проблемы, требующая рассмотрения); 2) математико-статистическое описание ТП с целью построения многофакторной модели (прикладная сторона, требующая принятия конкретного решения).

Следует подчеркнуть, что задачу оптимизации нужно рассматривать и решать непрерывно на всех стадиях создания, разработки и применения изделий. Только в этом случае можно достичь существенного эффекта и оптимизации их качества. На стадии проектирования необходимо оптимизировать показатели конструкции приборов, на стадии производства — технологические факторы, режимы работы оборудования и свойства используемых материалов. В процессе испытаний и применения уже готовых изделий нужно найти оптимальные условия эксплуатации с целью обеспечения требуемой долговечности и надежности.

роста производительности труда — примеры задач оптимизации с использованием экономического критерия оптимальности.

При выборе конкретного критерия оптимальности необходимо соблюдать принцип однозначности: в каждой задаче оптимизации целесообразно минимизировать или максимизировать одну и только одну целевую функцию. Кроме того, необходимо определить накладываемые на задачу ограничения, т. е. любая задача оптимизации должна формулироваться следующим образом: добиться максимума (минимума) целевой функции y при следующих ограничениях (далее следует перечень ограничений).

Пример 1.1. Максимизировать напряжение пробоя $U_{з.пр}$ затворного оксида полевого транзистора при следующих ограничениях: толщина оксида $d \leq 2000 \text{ \AA}$; температура печи $T \leq 1300 \text{ }^\circ\text{C}$; время выращивания оксида $t \leq 4 \text{ ч}$.

В приведенном примере в процессе производства оптимизируется $U_{з.пр}$ — один из электрических параметров транзистора, ограничения накладываются на один из конструктивных параметров транзистора и две управляемые переменные ТП.

На практике, однако, приходится анализировать несколько параметров, которые следует оптимизировать одновременно. Так, в примере 1.1 операция формирования затворного оксида определяет еще один важный оптимизируемый параметр — пороговое напряжение $U_{пор}$ отпирания транзистора. В таких случаях для соблюдения принципа однозначности задача оптимизации формулируется одним из следующих способов: а) из всех параметров ИЭТ выбирают важнейший, который принимают за параметр оптимизации (целевую функцию), остальные параметры переводят в разряд ограничений. Например, для примера 1.1 можно ставить задачу максимизации $U_{з.пр}$ при $U_{пор} \leq 3 \text{ В}$; б) строят обобщенную целевую функцию y , в которую интересующие параметры входят таким образом, что оптимизация функции автоматически приводит к значениям отдельных параметров, близким к требуемым.

Способы построения обобщенных целевых функций будут рассматриваться в гл. 4.

Структурная и параметрическая оптимизации ТП. Существует два вида оптимизации ТП: структурная и параметрическая.

При *структурной* оптимизации для достижения экстремальных значений целевой функции аппарат оптимизации применяют к самой структуре ТП, т. е. изменяют последовательность технологических операций, методы проведе-

ния отдельных операций, состав и тип технологического оборудования и т. п. В какой последовательности проводить обезгаживание деталей собранного мощного электровакуумного прибора (ЭВП), применять водородный или вакуумный отжиг, вести диффузию из жидкого или твердого источника, а может быть использовать полную имплантацию вместо диффузии - такие вопросы приходится решать при структурной оптимизации ТП. Очевидно, что задачи структурной оптимизации возникают только на этапе

проектирования или модернизации ТП.

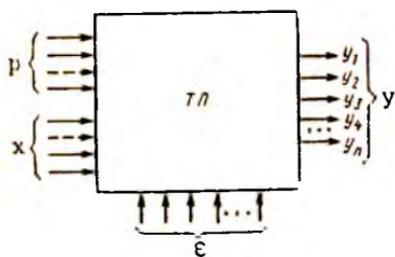


Рис 1.1

Чаще оптимизацию проводят при заданной структуре ТП, когда для достижения экстремума целевой функции можно менять только режимы (параметры) технологических операций. Такую оптимизацию называют *параметрической**.

При параметрической оптимизации ТП представляют в виде «черного ящика» (рис 1.1) с набором входных и выходных переменных, среди которых выделяют: y — вектор выходных переменных; x — вектор входных управляемых переменных, p — вектор входных контролируемых, но не управляемых переменных; ϵ — вектор случайных возмущений. Тогда параметрическая оптимизация ТП заключается в нахождении и реализации таких способов варьирования управляемых переменных $x_{opt}(t)$, при которых в рамках установленных технологических ограничений на переменные y и x обеспечивается экстремум целевой функции, несмотря на изменения контролируемых переменных p и воздействие случайных возмущений ϵ .

При решении поставленной задачи оптимизации выделяют два этапа:

1) нахождение математического описания или модели ТП, т. е. такой совокупности математических величин y , x , p и их отношений, которая адекватно отображала бы свойства рассматриваемого объекта оптимизации;

2) разработка алгоритма оптимального управления, т. е. такой последовательности обработки текущей инфор-

* В данном пособии основное внимание уделяется параметрической оптимизации ТП.

мации о ТП и такого порядка изменения уровней управляемых переменных x , которые обеспечивали бы поддержание целевой функции на экстремальном уровне при соблюдении технологических ограничений и несмотря на воздействие неуправляемых факторов p и случайных возмущений ε .

Следует отметить, что между этими двумя этапами существует глубокая внутренняя взаимосвязь: с одной стороны, оптимальное управление невозможно без знания модели объекта, а с другой — выбор того или иного вида модели ТП во многом определяется тем, какой метод управления предполагается применить. Например, использование алгоритма оптимизации Бокса—Вильсона (см. гл. 4) возможно только, если располагают локальным описанием целевой функции $y(x)$ вблизи текущей точки x . Однако если предполагается проводить оптимизацию методом Бокса—Вильсона, то нет необходимости строить сложную модель (например, полиномиальную высокого порядка) для описания $y(x)$ во всей области технологических ограничений. Достаточно линейной модели, удовлетворительно описывающей $y(x)$ в малой окрестности текущей точки x .

Непрерывные и дискретные ТП. Специфика производства ИЭТ. Как модель, так и алгоритм управления зависят от характера ТП. Так, в химической и металлургической промышленности в большинстве случаев ТП имеют непрерывный характер. Их описывают динамическими моделями с непрерывным временем:

$$\frac{dy_i}{dt} = \varphi_i(y_1, \dots, y_m; x_1, \dots, x_n; t) \quad i = \overline{1, m}. \quad (1.1)$$

Часто предпочтительнее перейти к дискретному времени. Тогда модель (1.1) вырождается в систему разностных уравнений.

Задачи динамической оптимизации решают с помощью вариационных методов или используя динамическое программирование.

Особенностью ТП изготовления ИЭТ в электронной промышленности является их дискретный характер. Отдельные операции ТП разделены во времени, иногда между ними проходят дни и даже недели. Такие ТП описывают следующим образом:

$$y_i = f_i(x_1, \dots, x_n; p_1, \dots, p_k) \quad i = \overline{1, m}. \quad (1.2)$$

Воздействие случайных возмущений ϵ приводит к тому, что соотношения (1.2) носят регрессионный характер, т. е. позволяют предсказать математическое ожидание выходных переменных. Поэтому зависимость (1.2) обязательно дополняется оценками дисперсии $D\{y_i\}$.

После определения зависимости (1.2) оптимизация дискретного ТП сводится к нахождению $x_{opt}(p)$, которое при известных значениях параметров p и ограничениях, накладываемых на переменные x , обеспечивает экстремум функции $y(x, p)$. В дальнейшем для упрощения записи математических выражений неуправляемые переменные p в (1.2) опускаем.

Существует два подхода к построению модели дискретного ТП:

1) физический, когда соотношения между входными и выходными переменными (1.2) определяют из анализа физико-химических процессов, происходящих в рассматриваемой технологической операции. После построения физической модели (1.2) используют тот или иной алгоритм нахождения x_{opt} , обеспечивающего достижение экстремума целевой функции;

2) экспериментально-статистический, базирующийся на статистической обработке данных, получаемых в результате эксперимента на объекте оптимизации. При этом модель объекта выражают в виде $y(x)$, т. е. в виде зависимости целевой функции y от переменных управления x .

Кроме того, можно сочетать физические и экспериментально-статистические методы построения модели ТП. В этом случае рассматривают физику процесса, пытаются установить принципиальный вид искомого модели, а затем экспериментально уточняют ее параметры.

Переменные управления x_1, \dots, x_n называют *факторами*, а значения, которые может принимать фактор, — *уровнями данного фактора*. Многомерное пространство, в котором каждому набору факторов x_1, \dots, x_n соответствует точка $x = (x_1, \dots, x_n)$, называют *факторным пространством*. С помощью технологических ограничений $Q_i(x)$ можно выделить из этого пространства некоторую допустимую область Q . Зависимость между x и соответствующими значениями параметра оптимизации y исследуют в пределах области Q . Функцию, связывающую параметр оптимизации y с факторами x , называют *функцией отклика*, а ее геометрический образ — *поверхностью отклика*. Качественный вид допустимой области Q и поверхности отклика для двумерного факторного пространства представлен на рис. 1.2,

а, б. Вид поверхности отклика может изменяться при изменении неуправляемых параметров p . Кроме того, влияние возмущений ϵ на поверхность отклика приводит к тому, что она как бы «размывается», флуктуирует подобно поверхности воды при ветре. Мерой этих флуктуаций служит дисперсия отклика или дисперсия воспроизводимости.

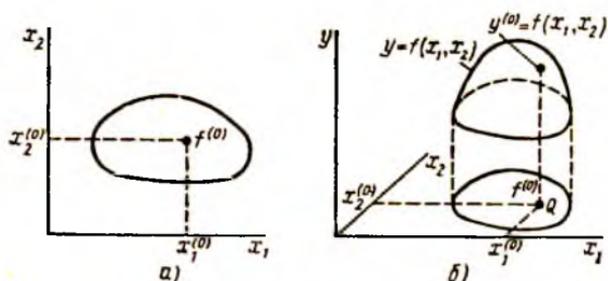


Рис. 1.2

При рассмотрении конкретных задач поверхность отклика удобно представлять в виде проекций линий равного уровня ($y = \text{const}$) на плоскость факторного пространства (рис. 1.3, а, б), где в процентах показан уровень выходного параметра от максимально возможного.

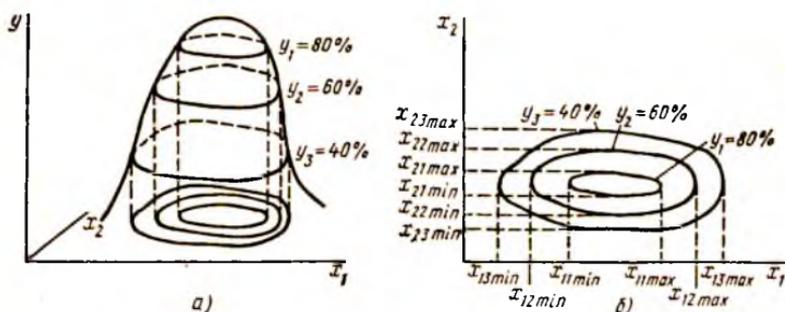


Рис. 1.3

Необходимо отметить, что для отдельных ТП поверхность отклика может иметь несколько максимумов. В этом случае в процессе эксперимента следует найти главный экстремум.

Для плохо обусловленных Т-систем сильное влияние случайных возмущений ϵ_j не позволяет осуществить декомпозицию в модели ТП, так как эффекты управляемых факторов первых звеньев системы, например x_1 , оказываются при приближении к конечным операциям практически полностью заглушены случайными возмущениями на промежуточных операциях. В этом случае можно использовать второй подход для решения задачи оптимизации,

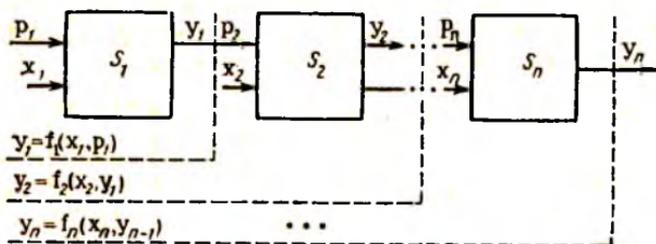


Рис. 1.5

который заключается в том, что каждая j -я операция ТП оптимизируется самостоятельно с использованием локальной модели $f_j(x_j, p_j)$, в которой результаты предыдущих операций учитываются через известный на момент проведения j -й операции вектор параметров p_j (рис. 1.6). При таком подходе оптимизация ТП достигается путем последовательной оптимизации отдельных операций, следующих друг за другом. При всей кажущейся простоте подобная оптимизация ТП на практике усложняется из-за необходимости формирования большого числа частных критериев оптимальности, не противоречащих друг другу. Действительно, попытка оптимизировать некоторый неудачно выбранный промежуточный параметр может затруднить или даже сделать невозможной оптимизацию конечных параметров последующей технологической цепочки. Поэтому следует подчеркнуть особую важность знания физических основ функционирования выпускаемых приборов, а также физико-химических основ используемых технологических методов. Только умелое сочетание физического моделирования и математического аппарата оптимизации оказывается достаточно эффективным средством для оптимизации технологических систем.

Типовые задачи оптимизации. Несмотря на многообразие задач оптимизации ТП, большинство из них решают с помощью общих алгоритмов, т. е. выделяют несколько типовых задач, алгоритмы решения которых едины.

Выделим три основные задачи оптимизации ТП ЭП:

1) оптимизация ТП или отдельных операций в случае, когда необходимо максимизировать (минимизировать) выходные параметры всего процесса или отдельной его операции;

2) оптимизация свойств многокомпонентных материалов (смесей);

3) оптимизация производственных испытаний готовых ЭП.

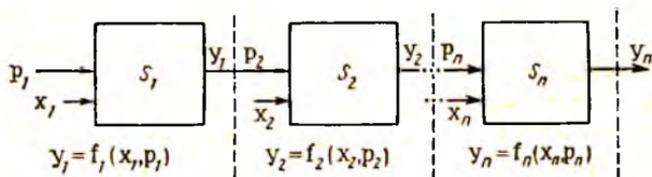


Рис. 1.6

Первая задача является наиболее общей, так как в качестве варьируемых факторов выбирают технологические режимы, характеристики конструкции и т. п., а в качестве параметров оптимизации — показатели качества приборов. Варьирование факторов может быть любым. Решать эту задачу можно и на стадии проектирования приборов для оптимизации их конструкции. Алгоритм решения задачи зависит от процесса (объекта) оптимизации.

Вторая задача, часто называемая задачей планирования на смесях, имеет особенность принципиального характера. Дело в том, что при создании многокомпонентной смеси она, количественно выражаясь в весовых или объемных единицах, должна как целое условно равняться 1 или 100 %. Очевидно, что содержание в смеси каждого компонента представляет собой определенную часть или процент от целого, а целое принимают за 1 или 100 %. Аналитически это условие записывают следующим образом $\sum_{j=1}^k x_j = 1$.

Следовательно, для решения рассматриваемой задачи необходимо применять модели и планы оптимизации, принципиально отличающиеся от тех, которые используют для решения первой задачи. К подобным задачам необходимо отнести оптимизацию свойств многокомпонентных материалов при производстве резисторов, конденсаторов, ферритов, разнообразных припоев и др. Например, трехкомпонентная композиция контактола создается на основе компаунда К-139, наполненного мелкодисперсным порошком

серебра. Твердая фаза получается с помощью отвердителя аминафенольного типа АФ-2. Пятикомпонентный феррит содержит Co, Ni, Al, Mn, Cu, трехкомпонентный припой — Cu, Ba, In.

Третья задача характерна тем, что варьируемым фактором является время испытаний t . Оно отличается от тех факторов времени, которые могут фигурировать в технологических операциях (время термообработки, время напыления и т. д.) и могут изменяться в необходимых пределах от опыта к опыту. Время t , начиная от начала испытаний ($t = 0$), возрастает и не поддается рандомизации, что следует учитывать при составлении планов и построении моделей испытаний. При этом данный алгоритм оптимизации испытаний отличается от алгоритма оптимизации первой и второй задач. Если необходимо оценить влияние внешних (климатических) факторов и времени испытаний на параметр оптимизации, то нужно найти зависимость

$$y = f(x_1, x_2, \dots, x_k, t),$$

где $x_j, j = \overline{1, k}$ — воздействующие климатические факторы.

Влияние на y факторов x можно определить, как и в первой задаче; зависимость y от времени t оценивается с помощью отдельного алгоритма.

§ 1.3. КЛАССИФИКАЦИЯ МЕТОДОВ ОПТИМИЗАЦИИ

Оптимизация ТП стала необходима вследствие усложнения Т-систем, а также исследовательских и производственных процессов, которые требовали минимизации затрат и негативных сторон исследуемого явления и максимизации положительного эффекта. При этом особых сложностей не вызывало исследование одномерной зависимости $y = f(x)$, когда в диапазоне $x_{\min} - x_{\max}$ аргумент принимает дискретные значения и характер зависимости выясняется полностью (рис. 1.7), а поиск оптимального значения отклика (в данном случае максимального) достаточно прост. Анализ зависимости выходного параметра, когда число факторов превышает два или три, усложняется, так как необходимо анализировать многомерную зависимость вида (1.2). В результате возникает вопрос — каким образом исследовать многофакторную зависимость $y = f(x)$ и какие пути приемлемы для оптимизации ТП.

До появления математических методов исследования многомерных зависимостей в эксперименте при оптимизации

нельзя определить направление возрастания функции, поэтому производят следующий опыт, например в точке 2. Функция отклика уменьшается, следовательно, следующий опыт выполняют сначала в точке 3 (функция растет), а затем в точке 4. Так как функция уменьшается, возвращаются в точку 3, в которой $y = y_{\max}$. Останавливаются на соответствующем этой точке значении $x_1 = \text{const}$, изменяют фактор x_2 и анализируют зависимость $y^{(2)} =$

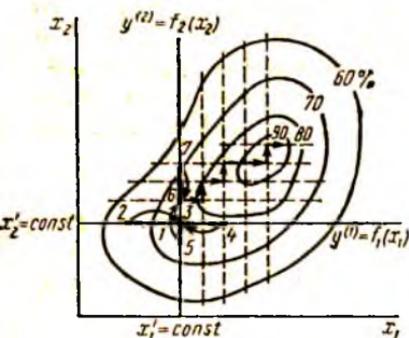


Рис. 1.8

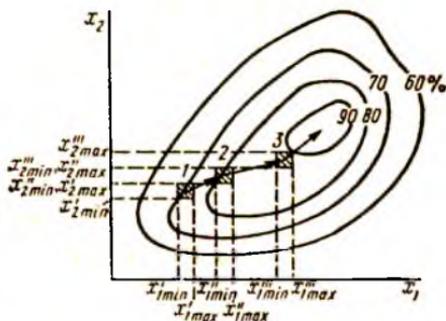


Рис. 1.9

$= f_2(x_2)$. Аналогичным образом проводят опыты 5–7, из которых опыт 6 соответствует максимальному значению функции отклика. Затем вновь изменяют первый, второй и т. д. факторы, постепенно приближаясь к оптимальному значению параметра оптимизации. Очевидны трудности, возникающие при использовании такого метода, если учесть, что в реальных задачах число факторов обычно больше двух. Таким образом, отличительной особенностью оптимизации методом однофакторного эксперимента является исследование множества одномерных зависимостей, когда варьируемые факторы принимают множество различных значений. Недостатком этого метода является невозможность практически оценить характер совместного влияния на функцию отклика всей совокупности факторов, что наиболее важно при экспериментальных исследованиях.

Методы многофакторного эксперимента и соответственно оптимизации лишены указанного недостатка, отличительной их особенностью является одновременное варьирование всеми факторами и оценка многомерной зависимости

$$y = f(x_1, x_2, \dots, x_k); x_1, x_2, \dots, x_k = \text{var.} \quad (1.5)$$

Изменение факторов осуществляют на ограниченном числе уровней, движение к оптимуму производят кривым путем в градиентном направлении (рис. 1.9). В многофакторной области 1 после выполнения предварительного эксперимента определяют градиентное направление и в этом направлении перемещаются до области 2, в которой повторяется многофакторный эксперимент. При этом факторы варьируют на двух или ограниченном числе уровней.

Таким образом, принципиальное отличие многофакторных экспериментов от однофакторных заключается в том, что: 1) варьируются одновременно все k факторов; 2) число уровней варьирования невелико.

Преимущество многофакторных экспериментов заключается в их эффективности, т. е. многомерная зависимость (1.2) определяется при меньшем объеме эксперимента и меньших затратах времени.

За последние годы получил развитие ряд методов оптимизации, среди которых можно выделить следующие:

а) методы вариационного исчисления, с помощью которых производят выбор конечного числа операторов или функций, или единственного оптимального из них;

б) методы параметрической оптимизации, с помощью которых находят оптимальные значения выходных параметров технологической системы при заданной ее структуре.

Задачи, решаемые этими методами, характеризуются определенным видом функции, в частности вида (1.2), или функционала, подлежащего оптимизации. Методы вариационного исчисления применяют для оптимизации, когда структура системы не задана. Однако ТП как объект оптимизации, обычно задан с точностью до отдельного технологического действия, поэтому при исследовании технологических систем, представляющих собой сложные стохастические системы, когда необходимо получить оптимальные значения одного или совокупности параметров, используют методы параметрической оптимизации.

Методы параметрической оптимизации можно разделить на: а) методы безусловной оптимизации; б) методы математического программирования; в) методы статистической оптимизации.

Для стохастических технологических систем наиболее приемлемы методы статистической оптимизации.

Методы безусловной оптимизации. Эти методы, как правило, не ставят своей целью математическое описание зависимости (1.2), а выбирают только такие варианты опытов,

которые при варьировании x приводят к изменению y в сторону экстремума. Кроме того, эти методы не применяют специальных планов варьирования x . Среди них можно отметить следующие: а) метод сканирования; б) метод градиента; в) метод наискорейшего спуска; г) метод тяжелого шарика.

Метод сканирования (слепой поиск) заключается в последовательном переборе вариантов сочетаний факторов x из множества допустимых, выборе экстремального значения целевой функции $y_i = \text{ekst } y(x_i)$, $i = \overline{1, N}$, и совокупности факторов, приведших к этому экстремальному значению. В соответствии с этим алгоритмом выбранный вариант заменяется только в том случае, когда новое значение целевой функции окажется ближе к экстремуму. Преимущество метода — независимость поиска от вида и характера целевой функции, недостатки — неэкономичность, ограниченность применения с ростом размерности факторного пространства.

Метод градиента является одним из самых распространенных методов безусловной оптимизации. Суть его заключается в том, что поиск оптимума осуществляют в градиентном направлении, т. е. в направлении вектора градиента целевой функции

$$\text{grad } y = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \dots, \frac{\partial y}{\partial x_k} \right).$$

Поскольку вектор градиента определяет направление наибольшего возрастания целевой функции, данный метод обеспечивает движение при поиске оптимума (максимума) в наилучшем направлении. Поиск по методу градиента состоит из двух этапов: 1) определение составляющих градиента, т. е. частных производных целевой функции по варьируемым факторам $\frac{dy}{dx_j}$, $j = \overline{1, k}$; 2) движение в градиентном направлении к оптимуму $x_j^{(i+1)} = x_j^{(i)} + a \frac{\partial y}{\partial x_j}$, $j = \overline{1, k}$, или в векторной форме $x_{i+1} = x_i + a \text{ grad } y(x_i)$, где $a > 0$ — параметр рабочего шага; $x_j^{(i)}$ — значение j -го фактора при i -м опыте.

Таким образом, для определения градиентного направления необходимо знать частные производные целевой функции по варьируемым факторам. Преимуществом метода является то, что на каждом шаге осуществляется движение в наилучшем направлении. Однако так как учитывают только первые частные производные, то метод градиента сле-

дует в принципе считать линейным методом, позволяющим точно двигаться к оптимуму линейной целевой функции. В противном случае движение производится не в наилучшем направлении. Кроме того, при использовании метода градиента перед каждым рабочим шагом следует производить сложный предварительный анализ оптимизируемой системы (процесса). Трудности могут возрасти, если появятся ограничения, накладывающиеся на зону поиска оптимума.

Метод наискорейшего спуска (при минимизации y) является разновидностью метода градиента и суть его состоит в следующем. После определения наилучшего направления делается шаг в направлении, обратном градиентному. Если значение целевой функции уменьшилось по сравнению с исходным, то производится очередной шаг в том же направлении, а не определяется заново градиент, как это имело место при методе градиента. Данный метод позволяет сократить объем вычислений, однако он обладает пониженной помехоустойчивостью.

Метод тяжелого шарика в отличие от рассмотренных ранее позволяет учесть характер движения на предыдущих шагах. Движение к оптимуму (минимуму) аналогично движению тяжелого шарика в k -мерном пространстве, брошенного на гиперповерхность отклика в поле тяготения. Наличие инерции движения позволяет набрать скорость в определенном (антиградиентном) направлении и «проскакать» локальные экстремумы, а следовательно, попасть в глобальный минимум [1].

Методы безусловной оптимизации применительно к электронным приборам используют на стадии проектирования. Опыты, на основании результатов которых ищется оптимум, являются машинными. При этом оптимизируемые системы считают детерминированными. Например, поиск оптимальной конструкции электронно-оптической системы, формирующей электронный пучок, может быть осуществлен путем машинного эксперимента на основе знаний математических соотношений физических явлений в электронно-лучевых приборах. Сложнее их применить для оптимизации стохастических технологических систем, где основой изучения системы является физический эксперимент.

Методы математического программирования. Под математическим программированием понимают совокупность методов решения задач получения максимального или минимального значения функции $f(x)$ на заданном множестве

M евклидова пространства E_n , а также значений $x^* \in M$, при которых соответствующее экстремальное значение достигается.

В общем случае задача математического программирования формулируется следующим образом. Пусть заданы функции $f(x)$; $\varphi_i(x)$, $i = \overline{1, m}$; $\psi_j(x)$, $j = \overline{1, p}$, определенные на E_n . Необходимо найти $\min(\max) f(x)$ при

$$\begin{aligned} \varphi_i(x) &\leq 0, \quad i = \overline{1, m}; \\ \psi_j(x) &= 0, \quad j = \overline{1, p}. \end{aligned} \quad (1.6)$$

Следовательно, требуется определить значение $x \in E_n$, удовлетворяющее (1.6), для которого $f(x) \leq f(z)$.

Если заданные функции линейны, то это задача линейного программирования; если хотя бы одна функция нелинейна, то это задача нелинейного программирования.

Среди задач, решаемых методами математического программирования, можно выделить три основные: 1) транспортную; 2) задачу о диете (оптимизация смесей); 3) задачу об использовании ресурсов (производственная задача). Рассмотрим кратко примеры постановок этих задач с целью получения представления о том, в каких случаях они могут возникнуть при производстве ЭП.

Транспортная задача. Имеется два склада готовой продукции и три потребителя этой продукции. Первый склад имеет N_1 единиц продукции, второй N_2 единиц продукции. Потребителям нужно соответственно m_1 , m_2 и m_3 единиц продукции. Стоимость доставки единицы продукции со складов каждому потребителю определена следующей таблицей:

Единица продукции	m_1	m_2	m_3
N_1	c_{11}	c_{12}	c_{13}
N_2	c_{21}	c_{22}	c_{23}

Составить план перевозок, который соответствует минимальной их стоимости.

Задача о диете. Имеется два вида материала M_1 и M_2 , состоящих из N_1 , N_2 , N_3 частей. Пусть a_{ji} — коли-

чество составной части N_i ($i = 1, 2, 3$), содержащейся в единице материала M_j ($j = 1, 2$); b_i — ежесуточная (или в единицу времени) потребность в составной части N_i ; c_j — стоимость единицы материала M_j . Определить количество k_1 материала M_1 и количество k_2 материала M_2 , необходимое для обеспечения требуемого количества продукции при минимальных затратах.

Производственная задача. Предприятие имеет в своем распоряжении определенное количество ресурсов (запасов) разного рода: сырье, оборудование и т. п. Ресурсы видов R_1, R_2, R_3 имеются в количестве b_1, b_2, b_3 условных единиц. Предприятие выпускает два вида приборов T_1, T_2 , причем известно, сколько единиц каждого ресурса требуется для производства одной единицы каждого прибора. Допустим a_{ij} число единиц ресурса R_i ($i = 1, 2, 3$), необходимое для производства единичного прибора T_j ($j = 1, 2$). Доход, получаемый предприятием от единицы каждого вида приборов, соответственно равен c_1, c_2 . Определить количество приборов, которое необходимо выпустить при данных ресурсах, чтобы получить максимальный доход.

Формулировка типовых задач иллюстрирует возможности математического программирования для решения задач оптимизации в электронном приборостроении. При этом на практике эти задачи могут встречаться в различной комбинации друг с другом.

Пример 1.2. Рассматриваемая часть ТП включает в себя две последовательные операции, на каждой из которых часть приборов бракуется. Из каждой партии на первой операции бракуется x_1 (%), а на второй — x_2 (%) приборов. Регулируя ТП, можно менять x_1 и x_2 при следующих ограничениях:

$$20 \leq x_1 + x_2 \leq 100; \quad x_1 \geq 5; \quad x_2 \geq 5; \quad x_1 + 4x_2 \geq 40. \quad (1.7)$$

Определить x_1 и x_2 , оптимальные с точки зрения минимума производственных потерь

$$C = N(c_1 x_1 + c_2 x_2), \quad (1.8)$$

где N — объем обрабатываемой партии приборов; $c_1 = 0,5$ руб./шт. — стоимость брака на первой операции; $c_2 = 1,5$ руб./шт. — стоимость брака на второй операции.

Перепишем условия (1.7) в виде следующей системы алгебраических неравенств:

$$x_1 \geq 5; \quad x_2 \geq 5; \quad x_1 + x_2 \leq 100; \quad x_1 + x_2 \geq 20; \quad x_1 + 4x_2 \geq 40. \quad (1.9)$$

Для удобства ее решения перейдем к системе новых переменных:

$$x'_1 = x_1 - 5; \quad x'_2 = x_2 - 5; \quad C' = 0,5x'_1 + 1,5x'_2.$$

Тогда систему (1.9) можно переписать в виде

$$x_1' \geq 0; x_2' \geq 0; x_1' + x_2' \leq 90; x_1' + x_2' \geq 10; x_1' + 4x_2' \geq 15. \quad (1.10)$$

Решение системы (1.10) можно получить с помощью симплексного метода линейного программирования. Здесь же рассмотрим геометрическое решение этой задачи, так как она имеет простую графическую интерпретацию в двумерном пространстве.

Неравенства (1.10) позволяют построить допустимую область приемлемых решений, и если из этих решений выбрать такое, которое минимизирует стоимость потерь (1.8), задача будет решена. На рис. 1.10 показана допустимая область приемлемых решений,

а также градиентное направление возрастания стоимости потерь и линии равного уровня (штриховые линии). Двигаясь по линиям равного уровня в направлении уменьшения $\text{grad } C$, остановимся на линии, которая пересекает допустимую область в точке M . Точка M определяет оптимальное решение

$x_1' = 8\frac{1}{3}; x_2' = 1\frac{2}{3}$. При этом

целевая функция (стоимость потерь) принимает оптимальное значение $C'(M) = 6,66$ или с учетом ранее проведенного преобразования $x_1 = 13,33; x_2 = 6,66; C = 16,66$ руб./шт.

Таким образом, оптимальным с точки зрения минимума потерь является следующее распределение процентов брака: на первой операции — 13,33 %, на второй — 6,66 %.

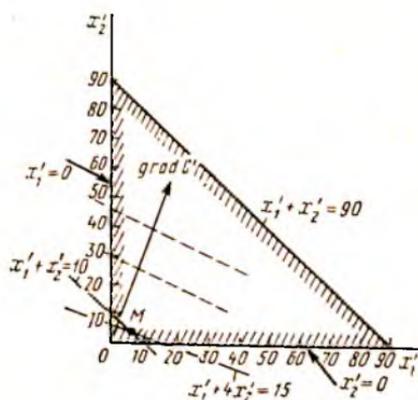


Рис. 1.10

Необходимо отметить, что перечисленные методы в той или иной степени позволяют решать оптимизационные задачи при производстве изделий электронной техники. Однако необходимо учитывать, что ТП представляет собой стохастическую систему и входные факторы и выходные параметры являются случайными величинами, поэтому для наиболее эффективной оптимизации следует применять методы статистической оптимизации.

Методы статистической оптимизации. Среди методов статистической оптимизации выделяют методы теории игр и теории планирования эксперимента. Поскольку методам теории планирования эксперимента в дальнейшем будет уделено достаточно внимания, рассмотрим кратко метод теории игр для решения задачи оптимизации ТП.

При управлении ТП возникает необходимость принятия решения в условиях неопределенности. Неопределен-

ность при контроле в ТП появляется из-за неполного знания общей картины как физической, так и статистической. В этих случаях требуется использовать математический аппарат, который бы позволял принимать решения в условиях ограниченной информации и выбирать оптимальные варианты процесса. Неопределенность и недостаточность информации затрудняет точное решение задач. Подобные задачи в математике анализируются в теории игр и статистических решений.

Во многих случаях существует противодействующая сторона (противник), т. е. возникает конфликтная ситуация. В теории конфликтная ситуация упрощается и схематическая ее модель определяется как игра. От реальной конфликтной ситуации игра отличается тем, что ведется по определенным правилам: противодействующие стороны — игроки, исход конфликта — выигрыш (проигрыш). Стратегия игрока определяется как совокупность правил, характеризующих выбор варианта действий при каждом ходе игрока в зависимости от ситуации, сложившейся в процессе игры. Оптимальная стратегия при многократном повторении игры обеспечивает данному игроку максимально возможный средний выигрыш (минимально возможный средний проигрыш).

Модель игры может быть представлена в виде платежной матрицы. Рассмотрим конечную игру (конечное число стратегий), в которой игрок A (мы) имеет m , а игрок B (противник) n стратегий. Такую игру называют игрой $m \times n$. Обозначим наши стратегии A_1, A_2, \dots, A_m , стратегии противника — B_1, B_2, \dots, B_n .

Пусть каждая сторона выбрала определенную стратегию: A_i и B_j . Тогда исход игры (наш выигрыш) обозначим a_{ij} . Если игра содержит случайные ходы, то выигрыш при паре стратегий $A_i B_j$ есть величина случайная, зависящая от исходов всех случайных ходов. В этом случае оценкой ожидаемого выигрыша является математическое ожидание случайного выигрыша. Предположим, что известны a_{ij} при каждой паре $A_i B_j$. Можно составить таблицу (табл. 1.1) — платежную матрицу (матрицу игры). Построение матрицы для игр с большим числом стратегий сложная процедура.

Когда неопределенность задана не противником, а возникла из-за недостаточной осведомленности, то такие игры называют играми с природой. Оптимизация ТП есть игра с природой, так как всегда в процессе существуют неуправляемые технологические факторы, которые хотя и пассивно, но затрудняют задачу управления и оптимизации. От-

существование неуправляемых факторов или их незначительное влияние упростило бы оптимизацию ТП.

В терминах теории игр задача оптимизации формулируется следующим образом. Пусть в игре участвуют два параметра, например поставщик сырья или материалов (партнер C) и технолог, перерабатывающий это сырье (партнер K). Технолог может выбирать любой технологический ре-

Таблица 11

Стратегия A	Стратегия B			
	B_1	B_2		B_n
A_1	a_{11}	a_{12}		a_{1n}
A_2	a_{21}	a_{22}		a_{2n}
A_m	a_{m1}	a_{m2}		a_{mn}

жим ведения процесса (стратегия K), поставщик сырья — любой из возможных вариантов сырья (стратегия C). Каждой комбинации k_i и c_j соответствует определенный выходной параметр процесса y_{ij} — параметр оптимизации. При k_i , $i = \overline{1, m}$ и c_j , $j = \overline{1, n}$ происходит игра $m \times n$.

Если известны y_{ij} для любых k_i и c_j , то игра полностью определена и условия ее можно записать согласно табл. 1.2, т. е. составить платежную матрицу или матрицу игры.

Решением игры (за партнера K) является смешанная стратегия технолога

$$S_n = \begin{pmatrix} k_1 & k_2 & \dots & k_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}, \quad (1.11)$$

которая при многократном повторении обеспечивает наилучшее среднее значение y_{ij} . Здесь p_i — частоты, с которыми следует случайным образом чередовать стратегии k_i ; $\sum_{i=1}^m p_i = 1$. Некоторые стратегии могут иметь $p_i = 0$; стра-

Таблица 1.2

Стратегия технолога	Стратегия поставщика				
	c_1	...	c_j	...	c_n
k_1	y_{11}	...	y_{1j}	...	y_{1n}
...	
k_i	y_{i1}	..	y_{ij}	..	y_{jn}
...
k_m	y_{m1}	..	y_{mj}	...	y_{mn}

тегии k_i , для которых $p_i \neq 0$, называют *активными*. Если решение представлено единственной активной стратегией $S_k = \begin{pmatrix} k_a \\ 1 \end{pmatrix}$, то k_i — чистая стратегия.

Аналогично можно судить о решении игры за партнера С:

$$S_c = \begin{pmatrix} c_1 & c_2 & \dots & c_h \\ q_1 & q_2 & \dots & q_h \end{pmatrix},$$

где $\sum_{j=1}^n q_j = 1$.

В результате получаем цену игры Y .

Пример 1.3. Рассмотрим оптимизацию ТП обжига на этапе спекания ферритовых микросердечников.

Независимые переменные: x_1 — температура обжига на этапе спекания ферритовых микросердечников, °С; x_2 — продолжительность обжига, мин. Эти факторы варьируются технологом и выражают его стратегию.

Поскольку свойства материала микросердечника в основном определяются тремя компонентами (x_3 — доля первого, x_4 — доля второго, x_5 — доля третьего компонента), значения долей компонентов представляют собой вариант стратегии поставщика сырья.

Параметр оптимизации — процент выхода годных изделий Q .

В качестве стратегий k_i рассмотрены режимы обжига, при которых температура изменяется, начиная с 125 °С через 5 °С, а время обжига от 60 до 120 мин через 10 мин. Варианты стратегий приведены в табл. 1.3.

Стратегия c_j содержит 66 вариантов, в которых компоненты сырья x_3, x_4, x_5 в процентном содержании принимают значения от 0 до 100 с шагом 10 % при соблюдении условия $x_3 + x_4 + x_5 = 1$.

Варианты стратегий поставщика сырья приведены в табл. 1.4.

Таблица 1.3

k_i	x_1	x_2	k_i	x_1	x_2	k_i	x_1	x_2
k_1	125	60	k_{13}	130	110	k_{25}	140	90
k_2	125	70	k_{14}	130	120	k_{26}	140	100
k_3	125	80	k_{15}	135	60	k_{27}	140	110
k_4	125	90	k_{16}	135	70	k_{28}	140	120
k_5	125	100	k_{17}	135	80	k_{29}	145	60
k_6	125	100	k_{18}	135	90	k_{30}	145	70
k_7	125	120	k_{19}	135	100	k_{31}	145	80
k_8	130	60	k_{20}	135	110	k_{32}	145	90
k_9	130	70	k_{21}	135	120	k_{33}	145	100
k_{10}	130	80	k_{22}	140	60	k_{34}	145	110
k_{11}	130	90	k_{23}	140	70	k_{35}	145	120
k_{12}	130	100	k_{24}	140	80			

Матрица игры имеет размер 35×66 , фрагмент ее приведен в табл. 1.5.

Методом итераций после 200 циклов, выполненных с помощью ЭВМ, найдена оптимальная смешанная стратегия в соответствии с (1.11):

$$S_H = \begin{pmatrix} k_{18} & k_{19} & k_{26} & k_{27} \\ 0,26 & 0,09 & 0,48 & 0,17 \end{pmatrix};$$

цена игры $Q = 0,473$.

Таким образом, если чередовать случайным образом в течение достаточно продолжительного времени активные стратегии k_{18}, k_{19}, k_{26} и k_{27} с частотами 0,26; 0,09; 0,48; 0,17, то при любых изменениях компонентного состава сырья, из которого делаются ферритовые микросердечники, параметр оптимизации будет принимать значения $Q \geq 47,3\%$.

Технологические режимы, вошедшие в оптимальную смешанную стратегию, довольно близки друг к другу, поэтому целесообразно их объединить.

Геометрически задача объединения режимов сводится к нахождению центра тяжести четырех двумерных точек $k_{18}, k_{19}, k_{26}, k_{27}$ с учетом их весов: 0,26; 0,09; 0,48; 0,17. Для этого достаточно найти x_1 и x_2 , взвешенные по четырем режимам: $x_1 = k_{18}p_1 + k_{19}p_2 + k_{26}p_3 + k_{27}p_4 = 135 \cdot 0,26 + 135 \cdot 0,09 + 140 \cdot 0,48 + 140 \cdot 0,17 = 138$ °С; $x_2 = 90 \cdot 0,26 + 100 \cdot 0,09 + 100 \cdot 0,48 + 110 \cdot 0,17 = 99$ мин.

Таблица 1.4

c_j	x_3	x_4	x_5	c_j	x_3	x_4	x_5
c_1	1,0	0,0	0,0	c_{34}	0,3	0,2	0,5
c_2	0,9	0,1	0,0	c_{35}	0,3	0,4	0,3
c_3	0,9	0,0	0,1	c_{36}	0,3	0,3	0,4
c_4	0,8	0,2	0,0	c_{37}	0,2	0,8	0,0
c_5	0,8	0,0	0,2	c_{38}	0,2	0,0	0,8
c_6	0,8	0,1	0,1	c_{39}	0,2	0,7	0,1
c_7	0,7	0,3	0,0	c_{40}	0,2	0,1	0,7
c_8	0,7	0,0	0,3	c_{41}	0,2	0,6	0,2
c_9	0,7	0,2	0,1	c_{42}	0,2	0,2	0,6
c_{10}	0,7	0,1	0,2	c_{43}	0,2	0,5	0,3
c_{11}	0,6	0,4	0,0	c_{44}	0,2	0,3	0,5
c_{12}	0,6	0,0	0,4	c_{45}	0,2	0,4	0,4
c_{13}	0,6	0,3	0,1	c_{46}	0,1	0,9	0,0
c_{14}	0,6	0,1	0,3	c_{47}	0,1	0,0	0,9
c_{15}	0,6	0,2	0,2	c_{48}	0,1	0,8	0,1
c_{16}	0,5	0,5	0,0	c_{49}	0,1	0,1	0,8
c_{17}	0,5	0,0	0,5	c_{50}	0,1	0,7	0,2
c_{18}	0,5	0,4	0,1	c_{51}	0,1	0,2	0,7
c_{19}	0,5	0,1	0,4	c_{52}	0,1	0,6	0,3
c_{20}	0,5	0,3	0,2	c_{53}	0,1	0,3	0,6
c_{21}	0,5	0,2	0,3	c_{54}	0,1	0,5	0,4
c_{22}	0,4	0,6	0,0	c_{55}	0,1	0,4	0,5
c_{23}	0,4	0,0	0,6	c_{56}	0,0	1,0	0,0
c_{24}	0,4	0,5	0,1	c_{57}	0,0	0,0	1,0
c_{25}	0,4	0,1	0,5	c_{58}	0,0	0,9	0,1
c_{26}	0,4	0,4	0,2	c_{59}	0,0	0,1	0,9
c_{27}	0,4	0,2	0,4	c_{60}	0,0	0,8	0,2
c_{28}	0,4	0,2	0,3	c_{61}	0,0	0,2	0,8
c_{29}	0,3	0,7	0,0	c_{62}	0,0	0,7	0,3
c_{30}	0,3	0,0	0,7	c_{63}	0,0	0,3	0,7
c_{31}	0,3	0,6	0,1	c_{64}	0,0	0,6	0,4
c_{32}	0,3	0,1	0,6	c_{65}	0,0	0,4	0,6
c_{33}	0,3	0,4	0,2	c_{66}	0,0	0,5	0,5

Таким образом, найден режим ТП ($T = 138^\circ\text{C}$; $\tau = 99$ мин), который при любых колебаниях сырья обеспечит процент выхода годных изделий более 47,3 %.

Рассматривая приведенные методы оптимизации с общих позиций, необходимо отметить следующее. Все они исследуют многомерную зависимость вида (1.2) в задаваемом факторном пространстве. Однако методы безусловной оптимизации и математического программирования позволяют найти оптимальные решения в определенной степени детерминированных условиях. Их можно применить на стадии автоматизированного проектирования, когда требуется выбрать оптимальный вариант конструкции прибора

Таблица 1.5

Стратегия	c_1	c_2	...	c_{4n}
k_1	25,0	28,0	...	46,8
k_2	25,6	32,4	...	45,9
...
k_{35}	55,6	56,5	...	55,5

или определить его конструктивные размеры. Эти методы менее эффективны при экспериментальной оптимизации, например при оптимизации ТП производства изделий электронной техники. Как отмечалось, такие ТП представляют собой стохастическую систему, т.е. условия протекания (оборудование, сырье и материалы, среда, режимы, оператор) ТП постоянно меняются. Оптимальные варианты необходимо находить именно для меняющихся условий, подстраивая процесс и поддерживая выходные показатели продукции оптимальными или близкими к ним. В этом случае следует применять статистические методы параметрической оптимизации. Метод, основанный на теории игр, позволяет решить задачу, однако и здесь оптимизируется (максимизируется) минимальный выигрыш для любых изменений условий ТП. Этот метод необходимо использовать в тех случаях, когда имеется много неучтенных факторов, влияющих на ТП. Для непрерывно изменяющихся условий протекания ТП целесообразно иметь модель процесса, которая позволяет следить за колебаниями факторов и изменять их с целью поддержания найденных оптимальных условий, что можно эффективно осуществить для достаточно ограниченного объема информации с помощью методов теории планирования эксперимента.

Приведенные методы необходимо рассматривать как методы, дополняющие друг друга в комплексной программе оптимизации (рис. 1.11). Каждый этап создания изделий электронной техники имеет свои особенности, что ограничивает применение тех или иных методов. Очевидно, что оптимизацию решений на всех этапах нужно проводить

непрерывно и последовательно. Результаты предыдущего решения оптимизационных задач следует использовать в виде исходной информации на последующих этапах.

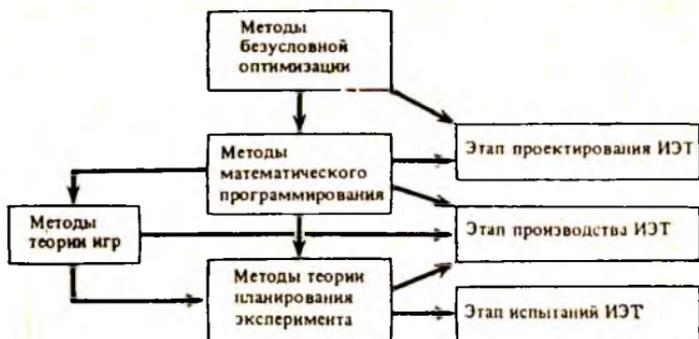


Рис. 1.11

ГЛАВА 2. ОСОБЕННОСТИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ ПРОИЗВОДСТВА ЭЛЕКТРОННЫХ ПРИБОРОВ

§ 2.1. ТИПОВАЯ ТЕХНОЛОГИЧЕСКАЯ СХЕМА ПРОИЗВОДСТВА ПОЛУПРОВОДНИКОВЫХ ПРИБОРОВ

Основой производства большинства современных полупроводниковых приборов является планарно-эпитаксиальная технология. Рассмотрим основные особенности этой технологии на примере производства монокристаллических интегральных микросхем (ИМС). Поскольку радиоэлементы таких схем формируются в процессе изготовления транзисторов ИМС, достаточно проследить основные этапы производства планарно-эпитаксиального транзистора для биполярных ИМС и МДП-транзистора для МДП ИМС.

На рис. 2.1 представлена структура биполярного транзистора ИМС с комбинированной изоляцией (изоляцию от подложки осуществляют с помощью *p-n*-перехода, а боковую изоляцию — с помощью диэлектрика SiO_2). Основой ИМС является высокоомная ($\rho \approx 10 \text{ Ом} \cdot \text{см}$) кремниевая пластина *p*-типа, в которой методом локальной диффузии через маску защитного оксида SiO_2 создают скрытый слой n^+ ($d = 2 \div 5 \text{ мкм}$; $R_S \approx 20 \text{ Ом}/\square$). Защитный слой SiO_2 удаляют травлением со всей поверхности пластины, после чего производят эпитаксиальное наращи-

вание p -слоя ($d \approx 3$ мкм, $R_S \approx 300$ Ом/□). Далее всю поверхность пластины покрывают защитной пленкой нитрида кремния Si_3N_4 . Таким образом получают заготовку для ИМС, где на месте каждого из будущих транзисторов расположен скрытый n^+ -слой, который является коллекторной областью транзистора. Поскольку в большинстве случаев предприятия-изготовители ИМС получают в качестве исходного сырья именно такие заготовки, будем вести отсчет операций начиная с момента, когда на пластине уже сформирован эпитаксиальный p -слой, под которым в нужных местах расположены n^+ -коллекторы будущих транзисторов. В процессе фотолитографии I вскрывают окна в защитном слое Si_3N_4

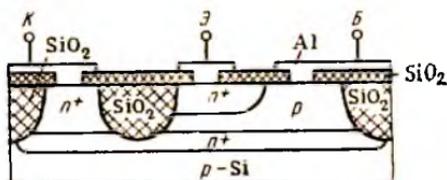


Рис. 2.1

под изолирующие области SiO_2 , которые создают путем глубокого термического оксидирования кремния в окнах вплоть до скрытого коллекторного n^+ -слоя. Так как переход кремния в оксид кремния сопровождается значительным увеличением объема, то перед оксидированием кремний в окнах Si_3N_4 частично стравливают, чтобы после получения плоскую поверхность пластины (уменьшение высоты микрорельефа пластины имеет особое значение при последующем формировании металлических соединений в ИМС, так как нанесение металлизационного слоя на неплоскую поверхность может привести к появлению ненадежных соединений в местах вертикальных ступеней рельефа). После образования изолирующих областей SiO_2 защитный слой Si_3N_4 , выполнявший функцию маски при селективном термическом оксидировании кремниевой пластины, удаляют путем травления, а пластину покрывают тонким слоем SiO_2 , получаемым термическим оксидированием или иным, например плазмохимическим, способом. При фотолитографии II вскрывают окна под глубокую n^+ -диффузию до смыкания с n^+ -коллектором. Поскольку заключительную стадию диффузии (разгонку) проводят в окислительной среде, вся пластина вновь оказывается защищенной слоем оксида SiO_2 . В процессе фотолитографии III вскрывают окна под n^+ -диффузию эмиттерных областей транзисторов. Эту диффузию проводят на дозированной глубине для обеспечения заданной толщины слоя базы w между эмиттером и коллектором (обычно $w = 0,3 \div$

0,6 мкм), она сопровождается термическим (либо иным) выращиванием оксида SiO_2 на поверхности пластины. После проведения эмиттерной диффузии внутренняя структура транзистора полностью сформирована. В процессе фотолитографии IV в защитном оксиде вскрывают окна под контакты, после чего на пластину напыляют Al. При фотолитографии V происходит гравировка по Al, формирующая рисунок металлических соединений между элементами ИМС. В настоящее время для улучшения параметров и повышения процента выхода годных ИМС используют методы электронно-ионной и плазменной технологии, такие, как ионная имплантация, плазмохимическое, пиролитическое нанесение защитных оксидных слоев и т. п., позволяющие, с одной стороны, уменьшить число и длительность высокотемпературных операций, а с другой — повысить управляемость ТП путем их автоматизации.

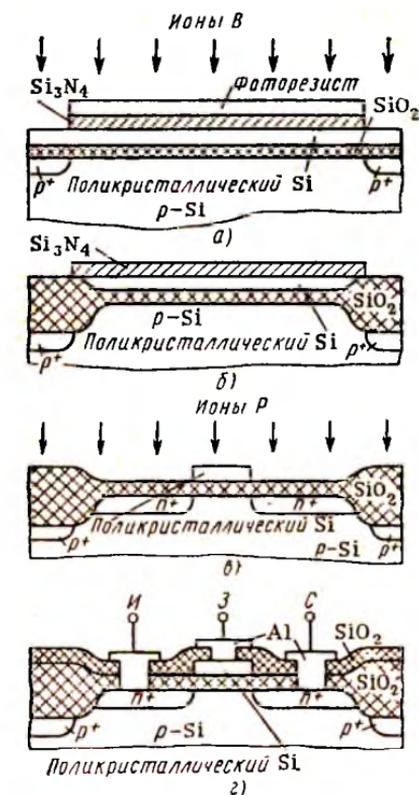


Рис. 2.2

после чего на ней путем термического оксидирования создают тонкую (0,1 мкм) пленку SiO_2 . Затем на поверхность пленки SiO_2 наносят путем напыления такой же тонкий слой поликристаллического кремния и на этот слой напыляют пленку нитрида кремния. После создания такой заготовки начинается процесс формирования структуры транзисторов ИМС. Для этого в процессе фотолитографии I пленку Si_3N_4 травлят всюду, кроме мест расположения будущих транзисторов. Далее следует операция ион-

ной имплантации бора, для которой оставшиеся слои Si_3N_4 и фоторезиста служат маской, тогда как тонкая пленка поликристаллического кремния и низлежащий слой SiO_2 пронизываются ионами бора. Таким образом, вокруг каждого транзистора создается охранное p^+ -кольцо, позволяющее исключить возможность возникновения паразитных каналов между соседними транзисторами ИМС (рис. 2.2, а). После этого проводят термическое оксидирование, в процессе которого слой нитрида кремния над будущими транзисторами служит защитой от проникновения атомов кислорода. В результате этой операции на пластине всюду, кроме мест расположения транзисторов, создается толстый (1 мкм) слой SiO_2 (рис. 2.2, б). После этого пленку Si_3N_4 травливают и проводят фотолитографию II, в процессе которой формируются затворы транзисторов (из поликристаллического кремния). Последующая операция ионной имплантации фосфора создает n^+ -области стоков и истоков (рис. 2.2, в). При этом кремниевый затвор и толстый оксидный слой служат маской, обеспечивая точное совмещение затвора с каналом каждого транзистора. Далее осуществляют газофазное осаждение защитного слоя SiO_2 . В процессе фотолитографии III вскрывают окна под контакты к истоку, стоку и затвору. После этого на пластину напыляют пленку Al и проводят фотолитографию IV, в результате которой формируется рисунок металлизационного слоя (рис. 2.2, г). Пластины подвергают отжигу (500 °C) для улучшения адгезии Al к поверхности ИМС.

При планарно-эпитаксиальной технологии на одной пластине формируют сотни и даже тысячи готовых структур будущих приборов (в частности ИМС) со всеми межсоединениями и контактными площадками. После соответствующего контроля ИМС пластину разделяют путем скрайбирования на отдельные кристаллы — чипы. Годные ИМС поступают на сборку. Процесс сборки заключается в закреплении кристалла на основании корпуса, присоединении выводов к металлическим контактными площадкам на кристалле и к внешним выводам корпуса и герметизации прибора. Закрепление кристалла на основании осуществляют пайкой с помощью эвтектического сплава Au—Si или Au—Ge либо эпоксидными клеями. Присоединение выводов производят путем термокомпрессии тонкой золотой или алюминиевой проволоки к контактными площадкам на кристалле.

Анализ приведенной типовой схемы планарно-эпитаксиального, а также других вариантов ТП производства

полупроводниковых приборов позволяет сделать вывод о том, что производство современных полупроводниковых приборов представляет собой дискретную последовательность сравнительно небольшого числа технологических



Рис. 2.3

операций. Эти операции можно разбить на несколько характерных групп (рис. 2.3).

Основными являются операции первых трех групп, так как они формируют внутреннюю структуру прибора. В табл. 2.1—2.3 приводятся различные методы осуществления этих операций, а также перечень входных и выходных переменных, используемых для управления ТП.

Таблица 2.1

Операция	Входные переменные	Выходные переменные
<p>Диффузия: методом открытой трубы (из твердого, жидкого, газообразного источника)</p> <p>из покрытий (стекло, оксидов)</p>	<p>Состав * источника и газа-носителя, температура источника, расход газообразных компонентов, температура в рабочей зоне печи, время процесса</p> <p>Состав покрытия, метод нанесения покрытия, толщина покрытия, температура в рабочей зоне печи, время процесса</p>	<p>Линейные размеры диффузионной области, профиль легирования, поверхностное сопротивление R_s диффузионного слоя, подвижность и время жизни носителей в диффузионном слое, напряжение пробоя и ток утечки $p-n$-перехода, емкость и вид в. а. х. $p-n$-перехода, технологический разброс параметров в пределах пластины и в партии пластин</p> <p>Те же, что и для диффузии, а также механические напряжения, плотность структурных дефектов n рекристаллизованном слое</p>
Сплавление	Состав электродного сплава, масса и геометрия впаляемого электрода, температура и состав газовой среды в рабочей зоне печи, режимы охлаждения после впавления и обработки подложки перед впавлением	Скорость роста эпитаксиальной пленки, толщина пленки, стехиометрический состав, однородность пленки, постоянная решетки, концентрация структурных дефектов, механические напряжения в пленке, степень легирования и профиль легирования пленки, время жизни и подвижность носителей в пленке, характеристики $p-n$ -перехода между пленкой и подложкой противоположного типа электропроводности, толщина переходного слоя между пленкой и подложкой (для гетерогенных структур)
Эпитаксия: жидкостная	Состав и температура расплава, градиент температуры, режим кристаллизации, режим обработки подложки перед эпитаксией, время процесса	
газовая	Состав парогазовой смеси, расход газа, температура в рабочей зоне, режим обработки подложки перед эпитаксией, время процесса	
вакуумная	Состав и температура источника, температура подложки, состав и парциальное давление остаточных газов в рабочей камере, режим обработки подложки перед эпитаксией, время процесса	

* Под составом будем понимать не только перечень компонентов, но и их концентрацию, процентное содержание и прочие количественные характеристики, необходимые для приготовления смеси, сплава и т. п.

Операция	Входные переменные	Выходные переменные
Ионная имплантация	Состав ионного источника, ускоряющее напряжение, ток пучка, доза облучения, температура отжига, время отжига, состав и толщина используемой маски	Те же, что и для диффузии, а также коэффициент использования внедренной примеси, концентрация радиационных дефектов

Таблица 2.2

Операция	Входные переменные	Выходные переменные
Литография: фото	Состав резиста, режим нанесения резиста на подложку, тип и параметры фотшаблона, характеристики оптической системы установки для совмещения и экспонирования, спектральный состав источника света и доза облучения, состав проявителя и режим обработки пленки резиста (температура, время), режим обработки поверхности подложки	Толщина и однородность пленки резиста, адгезия к подложке, химическая и температурная стойкость пленки резиста, размеры рисунка и профиль пленки на границе окон, минимальный размер рисунка (разрешающая способность), плотность проколов в пленке резиста, диэлектрические характеристики пленки (ϵ , $\operatorname{tg} \delta$, $U_{\text{пр}}$)
электронная	Те же, что и для фотолитографии, но вместо параметров источника света — параметры источника электронов: ускоряющее напряжение, ток луча, параметры фокусирующей системы и режим сканирования луча по рабочему полю, доза облучения	
рентгеновая	Те же, что и для фотолитографии, но вместо параметров источника света — параметры рентгеновского излучения: спектральный состав, доза облучения	

Операция	Входные переменные	Выходные переменные
<p>Селективное травление:</p> <p>жидкостное</p> <p>газовое</p>	<p>Состав и температура травителя, время травления</p> <p>Состав и расход газа, температура в рабочей зоне, время процесса</p>	<p>Скорость и профиль травления, морфология поверхности, размер микрошероховатости, размер бокового растравливания</p>
<p>плазмохимическое</p>	<p>Состав и расход газа, давление в рабочей зоне, мощность газового разряда, температура подложки</p>	
<p>ионно-плазменное</p>	<p>Состав, давление газа, ток разряда, ускоряющее напряжение, температура и положение подложки</p>	
<p>Размерная обработка остросфокусированными пучками:</p> <p>электронно- и ионно-лучевая</p>	<p>Ускоряющее напряжение, ток пучка, параметры фокусирующей системы и режим сканирования луча по рабочему полю, доза облучения, температура подложки</p>	<p>Линейные размеры рисунка и профиль на границе окон, разрешающая способность, размер микронеровностей, плотность радиационных дефектов</p>
<p>лазерная</p>	<p>Плотность мощности, длительность импульса, параметры фокусирующей оптической системы и режим сканирования луча по рабочему полю, время процесса</p>	

Операция	Входные переменные	Выходные переменные
<p>Вакуумное напыление: термическое испарение</p> <p>электронно-лучевое испарение</p> <p>ионно-плазменное (распыление (катодное, реактивное, катодное, ионно-плазменное, высокочастотное))</p>	<p>Температура испарителя, скорость испарения, давление паров в рабочей камере, состав и давление остаточных газов, положение и температура подложки, режим обработки подложки перед напылением, время напыления</p> <p>Те же, что и для термического испарения, а также ускоряющее напряжение и ток луча</p> <p>Состав и парциальное давление газов в рабочей камере, ускоряющее напряжение, ток, мощность разряда, температура подложки, напряжение смещения на подложке, положение подложки, режим обработки подложки перед напылением, время напыления</p>	<p>Скорость роста пленки, толщина, плотность, стехиометрический состав пленки, адгезия к подложке, параметры структуры пленки, размер зерен, плотность структурных дефектов, механические напряжения, электрофизические параметры пленки (R_s, ϵ, $\lg \delta$ и т. п.), температурная стабильность и технологический разброс параметров</p>
<p>Осаждение химическими методами: термическое оксидирование</p> <p>химическое осаждение</p>	<p>Состав, давление и расход рабочего газа, температура в рабочей камере, режим предварительной обработки пластин, время оксидирования</p> <p>Состав, рН, температура раствора, режим предварительной обработки подложки, время осаждения</p>	<p>То же</p>
<p>осаждение из газовой фазы</p>	<p>Состав и температура источника паров, состав и расход рабочего газа, давление паров в рабочей зоне, температура подложки, градиент температуры в системе, режим обработки подложки перед осаждением, время осаждения</p>	

Операция	Входные переменные	Выходные переменные
электрохимическое (гальваническое) осаждение	Состав и температура электролита, плотность тока, напряжение, режим предварительной обработки подложки, время осаждения	
Механическое покрытие: защита стеклами	Состав и режим нанесения стекла, температура термообработки, время термообработки, состав и давление окружающей газовой среды	
защита лаками и эмалями	Состав и режим нанесения покрытия, температура и время термообработки	

Приведем несколько примеров постановки задач оптимизации при производстве полупроводниковых приборов.

Пример 2.1 Исследовалась операция получения газофазного оксида SiO_2 путем окисления моносилана SiH_4 при температуре 400—500 °С. Активный газ — кислород; газ-носитель — азот.

Входные управляемые переменные: T — температура в активной зоне реактора; t — время процесса; v_{SiH_4} — расход моносилана; v_{O_2} — расход кислорода; v_{N_2} — расход азота.

Параметром оптимизации выбран относительный разброс толщины оксида в партии. Этот разброс минимизировался при скорости роста пленки не менее 0,01 мкм/мин.

Результаты исследования: температура T выбрана и зафиксирована (450 °С) на основе анализа плотности получаемого слоя оксида, а также из условия наиболее полного выхода реакции. Время t определялось исходя из требуемой толщины слоя. Был найден оптимальный расход компонентов, при котором разброс толщины слоя SiO_2 в партии не превышал 5 %, а скорость роста пленки оксида составляла 0,02 мкм/мин.

Пример 2.2. Исследовалась жидкостная эпитаксия p -структуры светодиода на основе $\text{Al}_x\text{Ga}_{1-x}\text{As}-\text{Al}_y\text{Ga}_{1-y}\text{As}$.

Входные управляемые переменные: P_1 — размер навески Al при выращивании первого слоя; T_1 — температура выращивания первого слоя; P_2 — размер навески Al при выращивании второго слоя; T_2 — температура выращивания второго слоя.

Параметром оптимизации выбран внешний квантовый выход светодиода.

Результаты исследования: найдены оптимальные значения P_1 , T_1 , P_2 и T_2 , при которых внешний квантовый выход был доведен до 93 %.

Пример 2.3. Исследовалось влияние технологических режимов фотолитографии на плотность проколов (микронор) в пленке фоторезиста.

Входные управляемые переменные: μ — вязкость фоторезиста перед нанесением; N — число оборотов центрифуги; T_1 — температура сушки; t_1 — время сушки; Z — pH проявителя; t_2 — время проявления; T_2 — температура задубливания.

Параметр оптимизации: плотность проколов.

Результаты исследования: найден режим фотолитографии, при котором плотность проколов в пленке фоторезиста уменьшена до 3 на 1 см².

Пример 2.4. Оптимизировался процесс производства тиристор-ов КУ108 с целью минимизации процента брака по критической скорости нарастания анодного напряжения $\left(\frac{dU_A}{dt}\right)_{кр}$

Выяснилось, что снизить процент брака по $\left(\frac{dU_A}{dt}\right)_{кр}$ можно, уменьшив разброс R_S слоя p -базы тиристора. В связи с этим был исследован процесс формирования p -базы, который проводился путем диффузии В + Al из слоя диффузанта, нанесенного на шлифованную полупроводниковую пластину. Установлено, что режим нанесения диффузанта влияет на разброс R_S . Подбран режим нанесения легированного фоторезиста путем центрифугирования, обеспечивший снижение процента брака в несколько раз по сравнению с началом исследования.

§ 2.2. ТИПОВАЯ ТЕХНОЛОГИЧЕСКАЯ СХЕМА ПРОИЗВОДСТВА ЭЛЕКТРОВАКУУМНЫХ ПРИБОРОВ

Проанализируем основные операции производства ЭВП на примере лучевого тетрода 6ПЗ6С, широко применяемого в телевизионной аппаратуре. Структурная схема ТП представлена на рис. 2.4.

Прибор содержит трубчатый оксидный катод (кери из никеля, покрытие — двойной карбонат $BaCO_3, SrCO_3$), внутри которого размещен подогреватель (проволока из сплава W — Mo, покрытая алундовой изоляцией); две сетки (молибденовая проволока, подвергнутая золочению для уменьшения термоэмиссии); две лучеобразующие пластины из никеля и анод из триметалла Al—Fe—Ni, обладающего высокой излучательной способностью. Баллон прибора — стеклянный с выводами из молибдена. Внутри баллона находятся также вспомогательные детали арматуры (никелевые переключки, слюдяные изоляторы, молибденовые траверзы и т. п.), необходимые для соединения основных узлов прибора.

При производстве тетрода можно выделить несколько характерных этапов. На первом этапе исходные материалы и полуфабрикаты (прокат, трубки, проволока и т. п.)

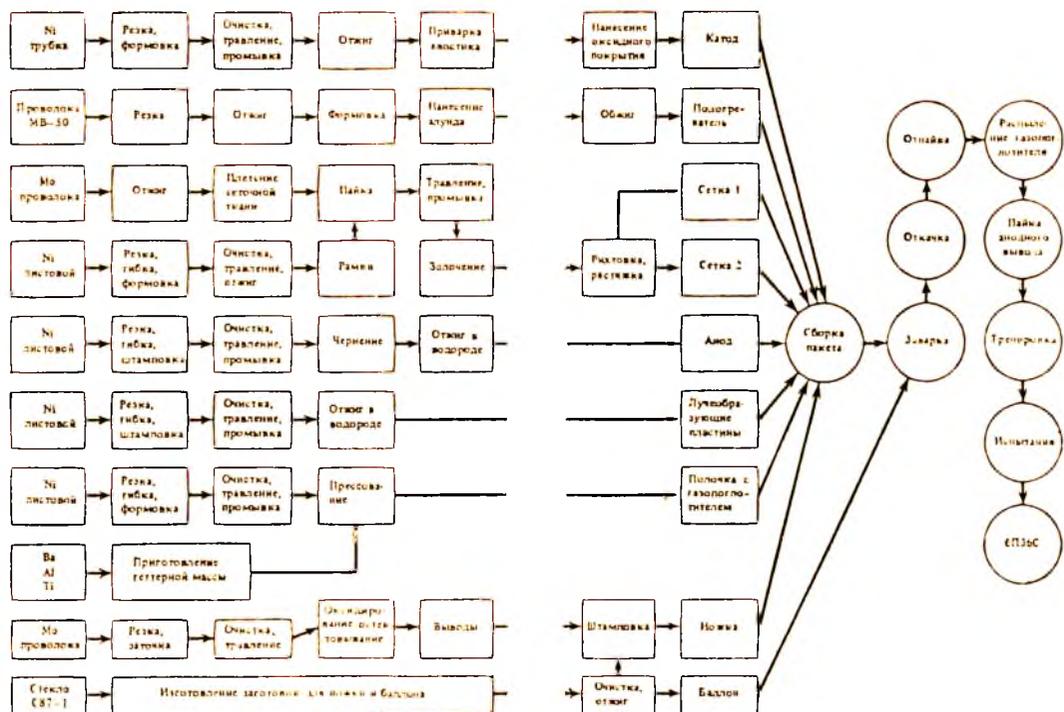


Рис. 2.4

подвергаются операциям формообразования, в результате которых получаются отдельные конструктивные элементы прибора. На втором этапе полученные детали подвергаются очистке и термообработке для удаления механических и химических загрязнений и снятия внутренних механических напряжений. Кроме того, ряд деталей, например керн катода, анод, подвергается вакуумному отжигу для обезгаживания и дополнительной очистки от загрязнений и оксидов. На третьем этапе очищенные детали поступают на сборку узлов первой ступени. Так, в рассматриваемом примере к трубчатому катоду приваривают никелевый вывод, слюдяные изоляторы армируют пистонами, сеточное полотно припаивают к никелевой рамке. На четвертом этапе на отдельные детали и узлы наносят специальные покрытия. В данном случае на внешнюю сторону трубчатого катода наносят суспензию из карбонатных солей бария и стронция, подогреватель покрывают алундом, а сетки подвергают золочению. На пятом этапе происходит сборка всех узлов прибора (катодного, сеточного и анодного) в единый пакет, который закрепляют на ножке прибора. На шестом этапе осуществляют заварку прибора, т. е. присоединение стеклянного баллона к его ножке, после чего лампа поступает на вакуумную обработку, которую проводят в две стадии. На первой, откачке, производят прогрев и обезгаживание внутренних деталей и частичную активировку оксидного катода, при которой карбонаты бария и стронция разлагаются и переходят в оксиды BaO и SrO. Характерным для этой стадии является обильное выделение газов, поэтому ее осуществляют при непрерывной откачке прибора. Давление в лампе в конце откачки достигает 10^{-4} Па, причем большая часть газов, первоначально адсорбированных внутренними деталями, уже удалена из прибора. Далее производят окончательную герметизацию прибора, после чего лампу подвергают второй стадии вакуумной обработки — тренировке. На этой стадии катод подвергают перекалу, т. е. его нагревают до температур выше рабочих, а также осуществляют токоотбор с катода на рабочие электроды ламп (анод, вторая сетка). При этом происходит окончательная активировка катода, стабилизируется его эмиссионная способность, а также происходит окончательная очистка и обезгаживание электродов лампы за счет их прогрева и бомбардировки электронами. Выделяющиеся при этом газы, общее количество которых сравнительно невелико, поглощаются специальным газопоглотителем (распыляемым или нераспыляемым), который

заранее помещают внутри прибора при его сборке; он имеет емкость, достаточную для того, чтобы поддерживать давление в отпаянном приборе не выше 10^{-4} Па в течение всего срока эксплуатации. Производство прибора завершается электрическими испытаниями.

Так же, как при полупроводниковом производстве, при производстве ЭВП можно выделить несколько характерных операций (рис. 2.5). В табл. 2.4—2.8 приведены различные методы проведения этих операций, а также перечень входных и выходных переменных, используемых для управления процессом.

Приведем примеры постановки оптимизационных задач в производстве ЭВП.

Пример 2.5. Исследовался процесс импульсной электронно-лучевой сварки тонкостенных коваровых деталей мощных ЭВП. Длительность импульсов фиксирована.

Входные управляемые переменные: U_a — ускоряющее напряжение; I_a — ток луча; f — частота следования импульсов.

Параметром оптимизации выбран размер зерна в сварном шве, который должен минимизироваться.

Результаты исследования: найден режим, при котором размер зерна уменьшен до 20 мкм (для сравнения — размер зерна при аргоно-дуговой сварке составляет 80 мкм).

Пример 2.6. Исследовался процесс точечной контактной сварки вольфрамовых проволоч при производстве решетчатых катодов мощных генераторных ламп. Режим работы: импульсный разряд конденсаторной батареи через свариваемый образец.

Входные управляемые переменные: C — емкость батареи; U — напряжение на батарее; F — сила прижатия свариваемых деталей. Сварка производилась при фиксированной индуктивности в первичной цепи сварочного трансформатора.

Параметром оптимизации выбрано усилие на отрыв сваренных проволоч, которое максимизировалось при следующих ограничениях: а) в месте сварки должны отсутствовать выплески, трещины, частицы меди от электродов; б) разрушение сварного соединения не должно носить характер хрупкого поперечного излома проволоки.

Результаты исследования: прочность сварного соединения доведена до значения, незначительно отличающегося от предельного для данного типа проволоки.

Пример 2.7. Оптимизировался режим откачки и тренировки мощных рентгеновских импульсных трубок с оксидным катодом.

Входные управляемые переменные: T — температура печи при обезгаживании; dT/dt — скорость подъема температуры при обезгаживании; $U_{n\max}$ — максимальное напряжение накала; dU_n/dt — скорость подъема напряжения накала при активировке катода; t_1 — время обезгаживания и термической активировки катода; I_k — ток катода при тренировке; t_2 — время тренировки; U_a — напряжение на аноде при тренировке.

Параметром оптимизации выбрано время проведения вакуумной обработки приборов, которое минимизировалось при соблюдении следующих ограничений: а) ток эмиссии катода в импульсе

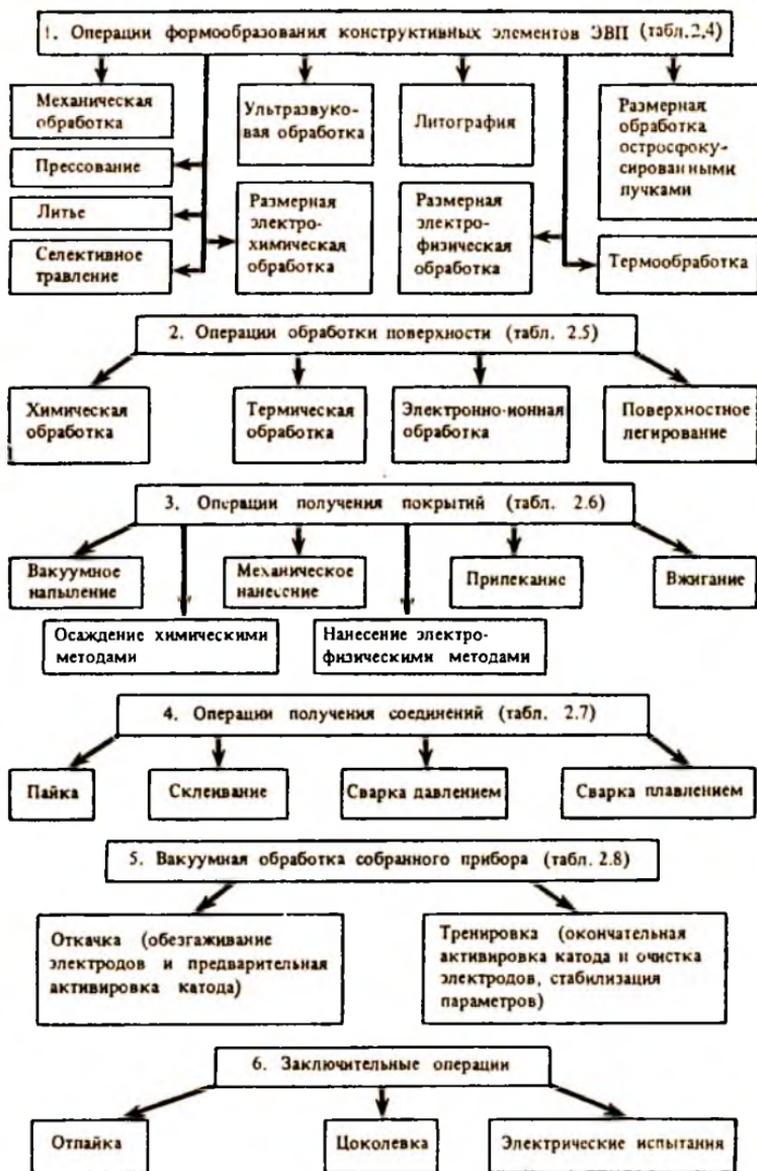


Рис. 2.5

должен быть после проведения операции не менее 10 А; б) давление остаточных газов в трубке после тренировки должно быть не более 10^{-2} Па.

Результаты исследования: найден режим откачки и тренировки, который позволил в 2,5 раза сократить длительность операции по сравнению с началом исследования без ухудшения качества приборов. Сокращение времени вакуумной обработки приборов достигнуто за счет интенсификации нагрева приборов и токообразования.

Операция	Входные переменные	Выходные переменные
<p>Механическая обработка: давлением (штамповка, гибка, прокатка, выдавливание, волочение) резанием</p>	<p>Материал, размер и форма инструмента, усилие, приложенное к заготовке, температура заготовки Материал и форма инструмента, скорость и угол резания, температура заготовки</p>	<p>Скорость обработки, размеры детали, размер микрошероховатости поверхности (класс чистоты), глубина разрушенного слоя, разброс линейных размеров в партии, точность изготовления, плотность структурных дефектов, зернистость структуры, состав и количество поглощенных газов, внутренние остаточные напряжения</p>
<p>абразивная</p>	<p>Материал и фракционный состав абразивного порошка, вязкость суспензии, давление на обрабатываемую поверхность, скорость перемещения шлифовальника, время шлифовки</p>	
<p>Литье</p>	<p>Состав и температура расплава, давление при заливке, размер и чистота поверхности литейной формы, режим охлаждения</p>	
<p>Порошковое прессование</p>	<p>Химический и фракционный состав пресс-порошка, давление при прессовании, размер и чистота поверхности пресс-формы, режим термообработки</p>	<p>То же</p>
<p>Ультразвуковая обработка (колебательное движение зерна абразива в рабочей суспензии)</p>	<p>Размер зерна абразива, состав и вязкость суспензии, размер и форма инструмента, расстояние между инструментом и заготовкой, частота и амплитуда колебаний, время обработки</p>	<p>»</p>
<p>Размерная электрохимическая обработка (анодное растворение в электролите)</p>	<p>Состав, плотность и температура электролита, плотность электрического тока, размер и форма катода, расстояние катод—изделие, давление и расход электролита, время процесса</p>	<p>»</p>
<p>Размерная электрофизическая обработка:</p>		

Операция	Входные переменные	Выходные переменные
электронскрoвая	Состав диэлектрической жидкости, энергия электрического импульса, частота следования импульсов, форма электрода, расстояние электрод — изделие	То же
магнитоимпульсная	Напряженность магнитного поля в импульсе, длительность импульса, размер и форма матрицы	
Термообработка (отжиг): а) в окислительной среде; б) в нейтральной среде; в) в восстановительной среде; г) в вакууме	Состав газовой среды, парциальное давление и расход газов, температура отжига, скорость нагрева и охлаждения, время отжига	»
Литография	См. табл. 2.2	»
Селективное травление		»
Размерная обработка остросфокусированными пучками		»

Таблица 2.5

Операция	Входные переменные	Выходные переменные
Химическая обработка: обезжиривание, промывка	Состав и температура раствора, режимы перемешивания и сушки, время процесса	Скорость обработки, размер микрошероховатости (класс чистоты), степень загрязнения поверхности, толщина легированного слоя, плотность структурных дефектов, излучательная способность обработанной поверхности
химическое травление (жидкостное, газовое, плазмохимическое)	См. табл. 2.2	
Термическая обработка	См. табл. 2.4	
Электронно-ионная обработка: ионно-плазменное травление	См. табл. 2.2	То же

Операция	Входные переменные	Выходные переменные
очистка в плазме тлеющего разряда	Напряжение и ток разряда, состав и давление газа, температура подложки, время очистки	То же
обработка электронным лучом	Состав и давление газа, ускоряющее напряжение, ток луча, температура подложки, время обработки	»
Диффузия, сплавление, эпитаксия, ионная имплантация	См. табл. 2.1	»

Таблица 2.6

Операция	Входные переменные	Выходные переменные
Вакуумное напыление		Толщина покрытия, скорость роста пленки, плотность и пористость покрытия, равномерность покрытия по всей площади подложки, адгезия к подложке, размер зерна в пленке, плотность структурных дефектов, размер микрошероховатости поверхности
Осаждение химическими методами	См. табл. 2.3	механические напряжения, электропрочность покрытия, термостойкость, химическая стойкость покрытия, термоэмиссионная способность покрытия, коэффициент вторичной эмиссии, фотоэмиссионная и излучательная способность
Механическое нанесение (окупание, намазка, полив, пульверизация)	Состав, вязкость и температура суспензии, режимы нанесения, сушки и термообработки	
Электрофизическое нанесение: электростатическое распыление	Состав и вязкость суспензии, напряжение на высоковольтном электроде, положение подложки, режимы сушки и термообработки	
электрофорез	Состав, вязкость и температура суспензии, ток и напряжение на электродах, время процесса, режимы сушки и термообработки	
Припекание (прокаливание в газовой среде или в вакууме)	Состав и давление газовой среды, температура в рабочей зоне печи, время прокаливания, ре-	

Операция	Входные переменные	Выходные переменные
Вжигание (нанесение из паст)	жим нанесения покрытия перед припеканием Состав наносимой пасты, режим обработки подложки перед нанесением пасты, температура вжигания, состав газовой среды в печи, время процесса, режим охлаждения	То же

Таблица 2.7

Операция	Входные переменные	Выходные переменные
Пайка	Состав припоя и флюса, количество припоя, температура пайки, состав газовой среды в рабочей зоне печи, режимы охлаждения и обработки поверхности деталей перед пайкой	Прочность соединения, размеры и форма шва, параметры кристаллической структуры в шве, остаточные напряжения в шве, плотность дефектов в шве, электрическое и тепловое сопротивление контакта, толщина соединительного слоя (при пайке и склеивании)
Сварка плавлением: газовая	Температура газового пламени, состав окружающей газовой среды, режимы обработки деталей перед сваркой и охлаждения шва, время процесса	
аргонно-дуговая	Ток дугового разряда, форма электрода, расстояние между электродом и деталью, расход газа, режим охлаждения шва	
электронным лучом	Ускоряющее напряжение и ток луча, диаметр луча, состав и давление газов в рабочей зоне, режим перемещения луча, время процесса, режимы обработки деталей перед сваркой и охлаждения шва	

Операция	Входные переменные	Выходные переменные
лазерная	Мощность излучения, длительность и частота следования импульсов (при импульсном режиме), диаметр луча, режим перемещения луча, состав окружающей газовой среды, режимы обработки деталей перед сваркой и охлаждения шва	То же
Сварка давлением: контактная (термокомпрессионная)	Давление на соединяемые детали, ток и напряжение сварочного импульса, длительность импульса, режим обработки деталей перед сваркой	
ультразвуковая (холодная)	Давление на соединяемые детали, мощность ультразвуковых колебаний, время процесса, режим обработки деталей перед сваркой	
Склеивание	Состав и вязкость клея, давление на склеиваемые детали, температура при склеивании, состав и давление окружающей газовой среды, время склеивания, режим обработки деталей перед склеиванием	

Таблица 2.8

Операция	Входные переменные	Выходные переменные
Откачка: термическое обезгаживание электродов и баллона (в общей печи или токами высокой частоты)	Скорость откачки, давление газов при откачке, температура электродов и баллона, время обезгаживания, скорость подъема температуры, режим обезгаживания газопоглотителя	Общее количество остаточных газов в приборе после откачки, состав и парциальное давление газов в отпаянном приборе, ток эмиссии катода, термоток сеток, ток утечки, общая

Операция	Входные переменные	Выходные переменные
<p>активировка катода с токоотбором</p> <p>Тренировка: окончательная очистка и обезгаживание электродов</p> <p>окончательная активировка катода и стабилизация параметров прибора</p>	<p>Скорость откачки, давление газов при откачке, температура и ток катода, время обработки</p> <p>Температура электродов, температура и ток катода, время обработки</p> <p>Температура катода, ток, отбираемый от катода, мощность, рассеиваемая на электродах прибора, токи и напряжения на электродах, время процесса</p>	<p>длительность операции откачки, надежность характеристики прибора</p> <p>Состав и парциальное давление остаточных газов после тренировки, ток эмиссии катода, электрические параметры прибора (крутизна характеристики, мощность, электропрочность и т. п.), длительность операции тренировки, надежность характеристики прибора</p>

§ 2.3. ФИЗИЧЕСКИЕ МОДЕЛИ ЭЛЕКТРОННЫХ ПРИБОРОВ И ТЕХНОЛОГИЧЕСКИХ ОПЕРАЦИЙ

Из рассмотрения типовых схем производства ЭП следует, что при производстве ЭП приходится иметь дело с большим числом объектов (структур и процессов) различной физической природы. Для оптимального управления производством необходимо иметь формализованное описание этих объектов, что можно осуществить с помощью моделей.

Модель — это формализованное описание основных величин и связей между ними, характерных для рассматриваемого реального объекта. Являясь формой абстрактного описания реальности, модель может быть только приближенной. Поэтому одному и тому же реальному объекту может быть сопоставлено множество моделей, отличающихся друг от друга как по своей структуре и степени детализации свойств рассматриваемого объекта, так и по способу их получения.

Проанализируем, например, поведение газа, заключенного в небольшом сосуде. Если плотность газа не слишком велика, то для его формального описания может быть использована известная модель идеального газа, согласно

которой газ представляют в виде совокупности отдельных молекул—шариков бесконечно малого размера, совершающих хаотическое тепловое движение и испытывающих при своем движении абсолютно упругие соударения друг с другом и со стенками окружающего сосуда. Такая модель представляет собой лишь физическую абстракцию, однако на ее основе получено важное с практической точки зрения уравнение состояния газа (уравнение Менделеева—Клапейрона):

$$PV = \frac{M}{\mu} RT, \quad (2.1)$$

где P — давление газа, Па; T — температура газа, К; V — объем, м³; M — масса газа, кг; μ — молярная масса газа, кг/моль; $R = 9,31 \cdot 10^3$ Дж/моль — универсальная газовая постоянная.

Уравнение (2.1) может быть названо моделью газа, так как оно представляет собой формализованное описание связей между характерными для газа величинами. Поскольку модель (2.1) дает количественное описание этих связей, она может быть названа математической моделью газа (см. определение математической модели, данное в гл. 1). С точки же зрения метода получения модель (2.1) является физической, поскольку она получена в результате применения основных физических законов к молекулам рассматриваемого газа.

Та же самая модель (2.1) может быть получена на основании обобщения результатов серии экспериментов, в которых для некоторого количества M исследуемого газа осуществлялось варьирование факторов P , V и T . В этом случае модель (2.1) в соответствии с методом ее получения можно назвать экспериментально-статистической.

Данный пример позволяет сделать вывод о том, что модели могут классифицироваться как по форме, так и по методу их получения. Так, модель (2.1) по форме является математической, а по методу получения может быть либо физической, либо экспериментально-статистической.

Условимся в дальнейшем вне зависимости от ее формы называть модель физической, если она получена как результат применения основных физических или физико-химических законов, определяющих сущность изучаемого объекта, в отличие от статистических моделей, получаемых путем статистического анализа связей между характерными величинами в отрыве от их физической сущности.

Рассмотрим роль и место физических моделей при ре-

шении задач управления и оптимизации ТП в электронной промышленности.

Конечной целью производства ЭП является получение приборов с заданными характеристиками (электрическими, надежностными или технико-экономическими). Поэтому при попытке управлять отдельными операциями или всем ТП в целом прежде всего возникает мысль использовать эти характеристики в качестве выходных переменных. Однако в большинстве случаев такой подход малоэффективен, так как при такой постановке задачи для осуществления управления требуется построить математическую модель процесса, которая бы связывала входные (технологические) переменные ТП с выходными характеристиками изготавливаемых приборов. Построить такую модель бывает в большинстве случаев чрезвычайно трудно, а порой и невозможно. Это объясняется тем, что чаще всего отсутствует прямая физическая связь между столь разнообразными по своей природе величинами, как, например, количественный состав газовой среды при диффузии и граничная частота усиления транзистора. Поэтому привлечение физических представлений не дает возможности построить математическую модель ТП. Попытка же использовать аппарат экспериментально-статистического исследования, для которого с формальной точки зрения не требуется знания физического содержания исследуемого процесса, также не приводит к желаемому результату: получаемые при этом модели либо не обладают достаточной точностью предсказания значений выходных параметров приборов, либо имеют приемлемую точность лишь в очень узком диапазоне варьирования технологических факторов — и в том и в другом случае модель оказывается малоприменимой для эффективного управления ТП.

В настоящее время в электронной промышленности все большее распространение получает другой подход к решению задачи управления ТП. Он заключается в том, что связь между технологическими переменными процесса и выходными параметрами изготавливаемых приборов рассматривается не непосредственно, а через структуру прибора, формируемую в результате проведения той или иной технологической операции. Следовательно, прослеживается логическая цепь «физика ТП — формируемая структура прибора — выходные параметры прибора». При этом исследователю необходимо определить: а) структуру прибора, которая может обеспечить заданные выходные характеристики; б) способ получения данной структуры.

При таком разбиении задачи удастся в полной мере использовать весь арсенал представлений как о физике работы прибора (для выяснения связи «структура—характеристики»), так и о физике рассматриваемых ТП (для выяснения связи «ТП—структура»).

Можно выделить три группы физических моделей.

Первая группа — это модели, устанавливающие связь между структурой и выходными характеристиками прибора. Модели этой группы получают при рассмотрении физики работы исследуемого прибора. Например, для МДП-транзисторов пороговое напряжение

$$U_{пор} = \frac{1}{q} (\Phi_з - \Phi_{п}) - \frac{Q_{п}}{\epsilon_d} d + \frac{d}{\epsilon_d} \sqrt{4\epsilon_{п} q V_T N_a \ln\left(\frac{N_a}{n_i}\right) + 2V_T \ln\left(\frac{N_a}{n_i}\right)}, \quad (2.2)$$

где $\Phi_з - \Phi_{п}$ — разность работ выхода затвор—подложка; d — толщина затворного диэлектрика; $\epsilon_{п}, \epsilon_{д}$ — диэлектрическая проницаемость затворного диэлектрика и подложки; N_a — концентрация акцепторной примеси в подложке (рассматривается n -канальный транзистор); n_i — собственная концентрация носителей заряда в подложке; q — заряд электрона; $V_T = kT/q$ — тепловой потенциал; $Q_{п}$ — плотность поверхностного заряда на границе затворный диэлектрик—подложка.

С учетом параметров исходных материалов ($\Phi_з, \Phi_{п}, \epsilon_{д}, \epsilon_{п}, n_i$) выражение (2.2) позволяет определить необходимые параметры формируемой структуры ($d, N_a, Q_{п}$) обеспечивающие требуемое значение $U_{пор}$ транзистора. При этом учитывают также технологические и физические ограничения, накладываемые на величины $d, N_a, Q_{п}$. Например, для МДП-транзистора со структурой кремний — оксид кремния — металл такими ограничениями являются: а) $d > 0,05$ мкм, иначе пробивное напряжение затвор—подложка оказывается слишком малым; б) $N_a < 10^{17}$ см⁻³ в противном случае максимально-допустимое напряжение сток—исток мало из-за пробоя стокового p - n -перехода; в) $\frac{1}{q} Q_{п} > 10^{10}$ см⁻², так как меньшей плотности заряда в системе кремний — термический оксид кремния получить не удастся.

Вторая группа физических моделей — это модели, устанавливающие связь между физическими переменными ТП и параметрами получаемой структуры. Модели этой груп-

пы используются для управления процессом формирования нужной структуры. Например, толщина l легированного слоя, полученного при диффузии из постоянного источника, связана с температурой T и временем t диффузии соотношением

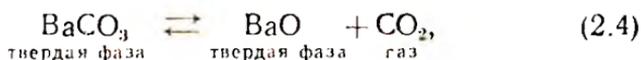
$$l^2 = At e^{-\frac{E_a}{kT}}, \quad (2.3)$$

где A — коэффициент, зависящий от используемого диффузанта и выбранного уровня отсчета концентрации; $k = 1,38 \cdot 10^{-23}$ Дж/град — постоянная Больцмана; E_a — энергия активации процесса диффузии для выбранного диффузанта.

Интересно, что использование физической модели (2.3) позволяет получить результат, противоречащий обычным представлениям, согласно которым, толщина легированного слоя должна быть прямо пропорциональна времени диффузии.

Третья группа физических моделей — это модели, устанавливающие связь между характерными физическими величинами ТП во время его проведения. Модели этой группы используют для управления ходом ТП.

Рассмотрим в качестве примера процесс разложения карбонатов при проведении активации оксидного катода ЭВП. Процесс описывают с помощью модели гетерогенной химической реакции диссоциации (разложения) BaCO_3 на поверхности катода



для которой закон действующих масс позволяет найти следующее соотношение для равновесного давления единственного газообразного компонента (это давление называют давлением диссоциации рассматриваемого химического соединения):

$$P_{\text{CO}_2} = K_p(T), \quad (2.5)$$

где $K_p(T)$ — константа равновесия, зависящая от температуры и увеличивающаяся с ее ростом.

До тех пор, пока давление диссоциации P_{CO_2} меньше внешнего давления $P_{\text{вн}}$, разложение соединения идет медленно. Равенство $P_{\text{CO}_2} = P_{\text{вн}}$ определяет температуру разложения $T_{\text{разл}}$, начиная с которой наблюдается интенсивное разложение. Таким образом, диссоциацию BaCO_3

можно осуществить двумя способами: 1) при $P_{\text{ин}} = \text{const}$ путем увеличения температуры до значения $T_{\text{разд}}$ и выше; 2) при $T = \text{const}$ путем уменьшения внешнего давления (путем откачки) до значения $P_{\text{дисс}}(T) = K_p(T)$. На практике при производстве ЭВП с оксидным катодом для ускорения процесса разложения карбонатов используют одновременно оба способа.

Рассмотренные группы физических моделей позволяют охватить все явления и взаимосвязи при производстве ЭП, начиная от задания технологических режимов и количественного состава исходных материалов и кончая выходными параметрами прибора. Однако для управления процессом производства одних только физических представлений недостаточно, так как физические модели не дают нужной степени детализации — они чаще всего описывают принципиальный вид исследуемых зависимостей между переменными процессами. Поэтому на практике физическую модель используют лишь как основу для постановки эксперимента, в результате которого модель обретает необходимую количественную конкретизацию и становится пригодной для управления.

Рассмотрим в качестве примера использование модели (2.3) для управления процессом диффузии. Согласно моде-

ли, толщина диффузионного слоя $l = \sqrt{A t} e^{-\frac{E_a}{2kT}}$. Однако ни коэффициент A , ни энергия активации E_a априори никогда не бывают известны с достаточной точностью. Поэтому ставят серию уточняющих экспериментов, предварительно преобразовав модель к виду

$$\ln l = \frac{1}{2} \ln A + \frac{1}{2} \ln t - \frac{E_a}{2k} \frac{1}{T}. \quad (2.6)$$

Если ввести обозначения $y = \ln l$; $x_1 = \ln t$; $x_2 = 1/T$, то модель (2.6) примет вид

$$y = c_0 + c_1 x_1 + c_2 x_2, \quad (2.7)$$

т. е. станет линейной относительно входящих в нее переменных y , x_1 и x_2 .

Коэффициенты c_0 , c_1 , c_2 определяют из обычного двухфакторного эксперимента, для которого требуется провести всего лишь четыре опыта. Коэффициент c_0 позволяет вычислить неизвестный параметр A , а коэффициент c_2 — энергию активации E_a . Коэффициент c_1 должен быть равен 0,5,

однако в результате экспериментов, проведенных в реальных производственных условиях, получают значения, отличающиеся от этой цифры в несколько раз. Аналогично энергия активации может отличаться в несколько раз в зависимости от конкретных условий. Так, в одном из экспериментов по легированию кремниевых пластин бором было получено значение $E_a = 0,6$ эВ, тогда как по известным из литературы данным $E_a = 3,6$ эВ. Такое резкое расхождение свидетельствует о том, что еще мало известно о механизме диффузии через поверхность (диффузия через поверхность идет не так, как в объемном кристалле: например, вблизи поверхности кристалла большую роль может играть диффузия по дислокациям, которая ускоряет проникновение диффузанта в глубь кристалла).

ГЛАВА 3. ОСНОВЫ СТАТИСТИЧЕСКИХ МЕТОДОВ ОПТИМИЗАЦИИ И ТЕОРИИ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

§ 3.1. ОБЩИЕ СВЕДЕНИЯ

Как отмечалось, решение задач оптимизации ТП предполагает прежде всего установление количественной связи между переменными, характерными для рассматриваемого процесса.

В гл. 2 рассмотрен физический подход к нахождению этой связи. В данной главе проанализируем статистические методы, используемые для этой цели.

Общим для всех статистических методов является то, что исследуемые переменные рассматриваются как случайные величины. Поэтому такие характеристики случайных величин, как закон распределения, математическое ожидание, дисперсия и т. п., играют в любом из этих методов важную, а иногда и определяющую роль.

Применение статистических методов для определения количественной связи между исследуемыми переменными предполагает наличие двух этапов: 1) эксперимент; 2) анализ статистического материала, накопленного в результате эксперимента. Между этими этапами существует глубокая диалектическая взаимосвязь: условия и способ проведения эксперимента (план эксперимента) определяют и дальнейший выбор конкретного метода анализа результатов и, наоборот, выбор определенного метода ана-

лиза диктует использование определенной стратегии эксперимента.

Точное определение характеристик случайных величин, в частности количественной связи между различными величинами, возможно только при бесконечном числе наблюдений. Поскольку реальный эксперимент всегда ограничен по своему «объему», анализ его результатов позволяет лишь приближенно оценить исследуемые величины и степень их взаимосвязи. Существующие статистические методы обязательно учитывают это обстоятельство, причем степень достоверности получаемых оценок всегда оговаривается. Собственно, основная задача математической статистики и заключается в том, чтобы на основании ограниченного по объему экспериментального материала, называемого *выборкой*, получать суждения (количественные и качественные) об исследуемых случайных величинах и оценивать степень их достоверности.

Различают две группы математико-статистических методов: статистическая оценка параметров и статистическая проверка гипотез.

Статистическая оценка параметров предусматривает получение оценок (точечных или интервальных) неизвестных параметров генеральной совокупности по параметрам выборки. В результате использования методов этой группы получают суждения типа «коэффициент корреляции между случайными величинами X и Y равен $0,72$ » (пример точечной оценки неизвестного параметра) или «математическое ожидание случайной величины Z с доверительной вероятностью $0,95$ заключено в пределах $27,6 \pm 1,2$ » (пример интервальной оценки неизвестного параметра).

Статистическая проверка гипотез предполагает выдвижение определенных допущений (гипотез) относительно неизвестных качественных (например, тип распределения) или количественных (например, дисперсия) характеристик случайных величин. Правильность этих гипотез проверяется по числовым значениям, полученным из выборки, и в зависимости от результата проверки гипотезы принимают или отвергают. В результате использования методов этой группы получают суждения типа «гипотеза о нормальном распределении случайной величины X не противоречит экспериментальным данным и, следовательно, может быть принята в качестве рабочей» или «гипотеза о том, что коэффициент корреляции между случайными величинами X и Y равен нулю, с доверительной вероятностью $0,95$ не согласуется с экспериментальными данными и должна быть

отвергнута, т. е. между X и Y существует линейная зависимость».

Статистические методы с точки зрения их приложения могут быть разделены на методы, направленные на расчет неизвестных характеристик исследуемых случайных величин, и методы, направленные на определение неизвестных взаимосвязей между исследуемыми величинами. Нас будут интересовать последние, так как для управления и оптимизации ТП необходима именно информация о взаимосвязи между характерными величинами. Из методов определения статистической связи между различными величинами наибольшее развитие получили следующие: 1) корреляционный; 2) регрессионный; 3) дисперсионный. Для того чтобы понять различие между этими методами, проанализируем наиболее простой случай двух величин: X и Y .

В корреляционном анализе X и Y — это две равноправные измеримые случайные величины, появляющиеся в виде пар значений при многократном повторении опытов эксперимента. Таким образом, X и Y могут рассматриваться как компоненты двумерной случайной величины (X, Y) , между которыми ищется взаимосвязь. В производстве ЭП ситуации такого рода чаще всего возникают при попытке установить связь между структурой и выходными характеристиками изготавливаемых приборов. Например, можно искать связь между парциальным давлением остаточного CO_2 в отпаянном ЭВП и сроком службы прибора. Если такая связь, т. е. корреляция, установлена, то оптимизацию ТП производства данного прибора с точки зрения повышения срока его службы можно проводить, не дожидаясь результатов испытаний на срок службы, а ориентируясь на давление CO_2 в отпаянном приборе. При этом оперативность управления процессом резко повышается.

К регрессионному анализу обращаются в том случае, когда есть основания предполагать наличие причинно-следственной связи между измеримыми переменными X и Y . Эксперимент по установлению регрессионной зависимости между переменными внешне может ничем не отличаться от эксперимента по установлению корреляционной зависимости. Однако при регрессионном анализе в отличие от корреляционного изначально закладывается неравнозначность между X и Y : X считают независимой переменной, а Y — зависимой, принимающей свои значения в зависимости от переменной X . Путем обработки экспериментальных данных (это по-прежнему пары значений x_i, y_i) отыскиваются неизвестные параметры, устанавливающие

количественную связь между значениями зависимой Y и независимой X переменных. Обратимся еще раз к примеру с испытаниями ЭВП на срок службы. Физические предпосылки указывают на наличие причинно-следственной связи между давлением CO_2 в изготовленном приборе и сроком его службы, так как известно, что наиболее частый вид параметрических отказов для ЭВП — это снижение эмиссии катода, а как раз наличие CO_2 в баллоне прибора и вызывает со временем потерю эмиссии за счет отравления катода. Тогда результаты тех же испытаний можно попытаться исследовать с помощью регрессионного анализа. Это приводит к установлению количественной зависимости между сроком службы прибора и парциальным давлением CO_2 в отпаянном приборе. Пользуясь этой зависимостью, можно на последней стадии изготовления ЭВП прогнозировать срок их службы, что дает возможность оперативно оценивать качество ТП и быстро принимать соответствующие меры, когда качество это начинает ухудшаться.

При дисперсионном анализе, как и при регрессионном, оценивается степень количественного влияния величины X на значения измеримой величины Y . Отличие заключается в том, что если при регрессионном и корреляционном анализе обе величины являются измеримыми, то при дисперсионном анализе независимая переменная X может быть и не измеримой, хотя и может принимать различные значения. Например, об эффективности автоматизации той или иной технологической операции можно судить по степени уменьшения влияния оператора на выходные параметры процесса. В частности, вопрос может ставиться так: «Зависит ли процент брака на выходе операции от того, кто конкретно проводит данную операцию?» Если такая зависимость сохраняется, то это означает, что уровень автоматизации явно недостаточен. В рассмотренном примере величина X (оператор) не поддается количественному измерению, хотя и может принимать различные значения (Иванов, Петров и т. д.)

Эти же принципиальные отличия между корреляционным, регрессионным и дисперсионным анализами сохраняются и при исследовании связей между многими переменными.

§ 3.2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Рассмотрим сначала две случайные величины. Как отмечалось, при корреляционном анализе они предполагаются равнозначными и измеримыми. Задача состоит в обра-

ботке результатов наблюдений, в каждом из которых контролируются одновременно значения величин X и Y . Полученные результаты измерений объема n состоят из n пар значений $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, которые можно истолковать как реализации двумерного случайного вектора (X, Y) .

На первом этапе вычисляют среднеарифметические \bar{x} и \bar{y} значения, а также эмпирические дисперсии s_x^2 и s_y^2 :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad (3.1)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad (3.2)$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2; \quad (3.3)$$

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.4)$$

Эти четыре эмпирических параметра характеризуют измеримые случайные величины X и Y по отдельности: \bar{x} и s_x^2 являются выборочными точечными оценками для математического ожидания μ_x и дисперсии σ_x^2 случайной величины X ; \bar{y} и s_y^2 — такие же оценки для μ_y и σ_y^2 . При увеличении объема выборки n эти оценки становятся точнее, приближаясь к истинным параметрам μ и σ^2 исследуемой случайной величины: $\bar{x} \xrightarrow[n \rightarrow \infty]{} \mu_x$; $\bar{y} \xrightarrow[n \rightarrow \infty]{} \mu_y$; $s_{x,y}^2 \xrightarrow[n \rightarrow \infty]{} \sigma_{x,y}^2$.

Нас, однако, интересует наличие связи между X и Y , а также сила (теснота) этой связи. Информацию об этом дает эмпирическая ковариация между значениями x и y :

$$\widehat{\text{cov}}(x, y) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \quad (3.5)$$

Нормируя $\widehat{\text{cov}}(x, y)$ с помощью среднеквадратичных отклонений s_x и s_y , получают эмпирический коэффициент корреляции между значениями x и y :

$$r_{xy} = \frac{\widehat{\text{cov}}(x, y)}{s_x s_y}. \quad (3.6)$$

Так как $\widehat{\text{cov}}(x, y) = \widehat{\text{cov}}(y, x)$, то $r_{xy} = r_{yx}$, т. е. порядок индексов не важен, что и должно иметь место, так как случайные величины X и Y полагаем равноправными.

Эмпирический коэффициент r_{xy} — это выборочная оценка для истинного коэффициента корреляции ρ_{xy} , который является мерой силы (тесноты) и направления линейной связи между значениями компонент случайного вектора (X, Y) .

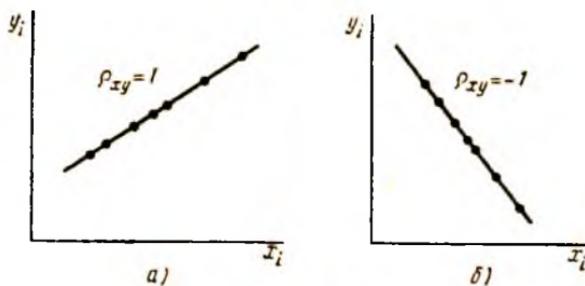


Рис. 3.1

Если $|\rho_{xy}| = 1$, то существует линейная функциональная зависимость между x и y , т. е. в любой реализации случайного вектора (X, Y) компоненты x и y связаны одним и тем же коэффициентом пропорциональности γ , так что $(x_i - \mu_x) = \gamma (y_i - \mu_y)$, $i = 1, 2, \dots$. В этом случае результаты наблюдений случайного вектора (X, Y) в выборке объема n находятся на одной прямой (рис. 3.1, а, б).

Если $0 < |\rho_{xy}| < 1$, то существует линейная связь между значениями x и y , внутренняя природа которой более сложна, чем простая линейная функциональная зависимость (это может быть случай, когда $X = X_1 + X_2$, $Y = Y_1 + Y_2$ и между X_1 и Y_1 имеется линейная функциональная зависимость, а X_2 и Y_2 — независимые случайные добавки). Результаты наблюдений случайного вектора (X, Y) в выборке объема n группируются вблизи некоторой прямой (рис. 3.2, а, б).

Если $\rho_{xy} = 0$, то невозможно в общем случае сделать вывод о независимости случайных величин X и Y . Можно лишь утверждать, что линейная связь отсутствует (нелинейная связь при этом не исключается, как это видно из рис. 3.3, а, б). Однако существуют случаи, когда равенство $\rho_{xy} = 0$ означает независимость случайных величин X и Y , например, если случайные величины X и Y имеют нормальное или равномерное распределение.

Из рассмотрения свойств коэффициента корреляции можно сделать вывод о том, что корреляционный анализ наиболее эффективен, когда есть основания предполагать преобладающий характер линейной связи между исследуемыми переменными. Применение корреляционного анализа в случае, когда возможна нелинейная зависимость между переменными, нецелесообразно. Однако использование физических представлений о природе исследуемых величин может расширить область применения корреляционного

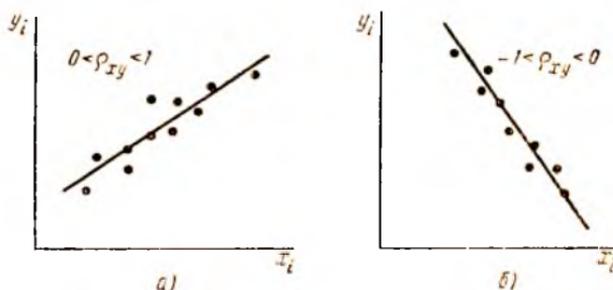


Рис. 3.2

анализа за счет предварительного преобразования переменных. Например, анализ физики теплового пробоя диэлектриков приводит к следующему соотношению между пробивным напряжением $U_{пр}$ и толщиной диэлектрика d :

$$U_{пр} \approx A \sqrt{d}. \quad (3.7)$$

Учитывая это, можно подвергнуть корреляционному анализу следующие величины: а) $X = U_{пр}^2$, $Y = d$; б) $X = U_{пр}$, $Y = \sqrt{d}$. Какой из рассмотренных вариантов преобразования предпочесть? Общей рекомендацией на этот счет является следующая: преобразование исследуемых величин должно быть таким, чтобы закон распределения преобразованных величин был по возможности ближе к закону нормального распределения. При этом облегчается проверка статистической гипотезы о значимости корреляции (о проверке статистических гипотез см. далее).

В приведенном примере следует остановиться на первом варианте, так как обычно линейные размеры изготавливаемых структур подчиняются закону нормального распределения (хотя и не всегда). Если физическая модель теплового пробоя (3.7) верна, то и $U_{пр}^2$ также будет иметь нормальное распределение. Последнее, однако, требует про-

верки, так как может оказаться, что в проводимых исследованиях пробой имеет не тепловой, а электрический характер, при котором $U_{\text{пр}}$ пропорционально d , а следовательно, в качестве исследуемых величин следует взять $X = U_{\text{пр}}$; $Y = d$. Данный пример свидетельствует о том, что прежде чем обратиться к статистическим методам анализа, нужно тщательно проанализировать всю априорную информацию и изучить все возможные варианты исследования.

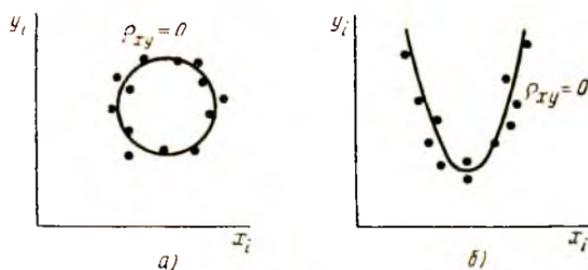


Рис. 3.3

Допустим, что при обработке эксперимента получен эмпирический коэффициент корреляции $r_{xy} = 0,52$. Можно ли считать установленной наличие линейной связи между исследуемыми случайными величинами? Не будем торопиться с выводами. Вспомним, что выборочные оценки сходятся к оцениваемым параметрам генеральной совокупности лишь при неограниченном возрастании объема выборки n . При любом конечном n они могут флуктуировать от выборки к выборке, отличаясь порой значительно от истинного значения оцениваемого параметра. Возможные отличия тем больше, чем меньше объем выборки, поэтому может оказаться, что в нашем случае $\rho_{xy} = 0$, а полученное значение $r_{xy} = 0,52$ — не что иное, как вполне естественное (статистически незначимое) отклонение от истинного коэффициента корреляции, вызванное малым числом наблюдений.

Для решения вопроса о статистической значимости полученного выборочного коэффициента корреляции выдвигают гипотезу $H_0: \rho_{xy} = 0$, а затем проверяют, не противоречит ли она экспериментальным данным. Эта проверка легко осуществляется, если известен закон распределения исследуемых случайных величин X и Y . Тогда по известным закону распределения и объему выборки n находят критическое значение эмпирического коэффициента корреляции.

ляции, вероятность превышения которого в рамках гипотезы H_0 не может быть больше заданного малого числа α , называемого *уровнем значимости*. Если найденный из эксперимента эмпирический коэффициент корреляции r_{xy} превысил это критическое значение, то тем самым осуществилось событие, маловероятное в рамках выдвинутой гипотезы H_0 : $\rho_{xy} = 0$. Это дискредитирует гипотезу H_0 , и ее отвергают, т. е. считают, что r_{xy} значимо отличается от нуля, так что $\rho_{xy} \neq 0$. Следовательно, существует линейная связь между переменными X и Y . Доверительная вероятность такого заключения равна $1 - \alpha$.

Если распределение исследуемых величин X и Y можно считать нормальным, то проверку значимости эмпирического коэффициента корреляции осуществляют по табл. 3.1, содержащей рассчитанные заранее наибольшие случайные значения коэффициента корреляции.

Проверим, например, значим ли эмпирический коэффициент корреляции $r_{xy} = 0,52$, если он получен из выборки объема $n = 22$. Для $m = n - 2 = 20$ можно вынести следующее суждение: «с доверительной вероятностью 0,95 корреляция значима ($r_{xy} > 0,42$)». Для $m = 20$ нельзя гарантировать значимость корреляции с большей доверительной вероятностью, так как $r_{xy} < 0,53$ и тем более $r_{xy} < 0,61$.

Проверим, какой вывод можно сделать, если значение $r_{xy} = 0,52$ было получено из выборки объема $n = 12$. Можно сделать вывод о том, что значение $r_{xy} = 0,52$ в столь малой по объему выборке не позволяет утверждать с достаточно высокой доверительной вероятностью о наличии корреляции между исследуемыми величинами. При этом не отвергается наличие корреляции, а только утверждается, что даже такая большая цифра, как $r_{xy} = 0,52$, на малой выборке не позволяет определить существование корреляции, так как это значение, возможно, является случайной флуктуацией, вполне допустимой в рамках гипотезы H_0 : $\rho_{xy} = 0$. Внимательный анализ табл. 3.1 позволяет сделать вывод о том, что в тех случаях, когда коэффициент корреляции ρ_{xy} мал ($\rho < 0,2$), определить его с достаточной достоверностью можно только на выборках большого объема ($n > 100$).

После того как установлена значимость коэффициента корреляции, он может быть использован для прогнозирования значений одной из случайных величин, если известно значение второй. Это прогнозирование основывается на следующих формулах:

Таблица 31

Число степеней свободы $n-d-2$	Уровень значимости α		
	0,05	0,01	0,0025
5	0,75	0,87	0,93
10	0,58	0,71	0,78
15	0,48	0,61	0,68
20	0,42	0,53	0,61
25	0,38	0,49	0,55
30	0,35	0,45	0,51
35	0,32	0,42	0,48
40	0,30	0,39	0,45
50	0,27	0,35	0,41
60	0,25	0,33	0,37
70	0,23	0,30	0,35
80	0,22	0,28	0,33
90	0,21	0,26	0,31
100	0,19	0,25	0,29
150	0,16	0,21	0,24
200	0,14	0,18	0,21
300	0,11	0,15	0,17
500	0,09	0,11	0,13
1000 и более	0,05	0,09	0,1

$$M(x/y) = M(x) - (y - M(y)) \sigma_x \frac{\rho_{xy}}{\sigma_y}; \quad (3.8)$$

$$\sigma(x/y) = \sqrt{1 - \rho_{xy}^2} \sigma_x, \quad (3.9)$$

где $M(x/y)$ — условное математическое ожидание величины x при фиксированной величине y ; $\sigma(x/y)$ — условное среднеквадратичное отклонение величины x при фиксированной y .

Из (3.9) следует, что чем ближе коэффициент корреляции к единице, тем точнее могут быть предсказаны значения x по известным значениям y . Именно это обстоятельство позволяет использовать корреляционный анализ для управления ТП.

Проведенный анализ для двух переменных X и Y легко обобщается на случай многих переменных. Рассмотрим многомерный случайный вектор X , реализациями которого являются наборы значений $(x_{1i}, x_{2i}, \dots, x_{pi})$, $i = 1, 2, \dots, n$. Вычисляя попарные ковариации между компонентами этого случайного вектора $\widehat{\text{cov}}(x_h, x_l)$, образуют эмпирическую матрицу ковариаций:

$$\widehat{\mathbf{V}}_X = \begin{bmatrix} s_{x_1}^2 & \widehat{\text{cov}}(x_1, x_2) & \dots & \widehat{\text{cov}}(x_1, x_p) \\ \widehat{\text{cov}}(x_2, x_1) & s_{x_2}^2 & \dots & \widehat{\text{cov}}(x_2, x_p) \\ \dots & \dots & \dots & \dots \\ \widehat{\text{cov}}(x_p, x_1) & \widehat{\text{cov}}(x_p, x_2) & \dots & s_{x_p}^2 \end{bmatrix} \quad (3.10)$$

Эта матрица симметрична относительно главной диагонали, так как $\widehat{\text{cov}}(x_h, x_l) = \widehat{\text{cov}}(x_l, x_h)$; на главной диагонали находятся эмпирические дисперсии компонентов вектора \mathbf{X} .

Нормируя элементы этой матрицы $\widehat{\text{cov}}(x_h, x_l)$ с помощью среднеквадратичных отклонений s_{x_h} и s_{x_l} , получают эмпирическую корреляционную матрицу

$$\mathbf{R}_X = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{bmatrix} \quad (3.11)$$

Это симметричная матрица, на главной диагонали которой располагаются единицы.

Важным свойством корреляционной матрицы является то, что ее определитель удовлетворяет неравенству

$$0 \leq |\mathbf{R}_X| \leq 1. \quad (3.12)$$

Определитель корреляционной матрицы характеризует глубину связи между компонентами вектора \mathbf{X} : если $|\mathbf{R}_X| \rightarrow 1$, то компоненты вектора \mathbf{X} независимы, если $|\mathbf{R}_X| \rightarrow 0$, то значение отдельного компонента является линейной комбинацией всех остальных. Последнее возможно даже в случае, когда в матрице \mathbf{R}_X нет больших значений r_{hl} , т. е. корреляция может быть существенной в целом.

Корреляционную матрицу часто используют для сокращения числа выходных переменных, учитываемых при управлении ТП. Для этого сначала разбивают выходные параметры на группы так, чтобы в одну группу попали близкие параметры (сильно коррелированные), а в разные — параметры, слабо связанные между собой. Затем внутри каждой группы выбирают один или несколько параметров, а остальные исключают. Таким образом, управление ТП осуществляют, ориентируясь на оставленные параметры, общее

число которых существенно меньше исходного числа параметров, что облегчает управление. В то же время сильная корреляция между исключенными и оставленными параметрами автоматически обеспечивает нужный эффект управления и для исключенных параметров.

§ 3.3. ПОСТРОЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Пусть исследуемый объект определяется выходным параметром y (процент выхода годных изделий, параметр готового изделия или полуфабриката и т. д.), который зависит от совокупности контролируемых входных параметров z_1, z_2, \dots, z_l .

В регрессионном анализе (РА) отклик y представляют в виде суммы результата функционального преобразования $f(z)$ и некоторой случайной ошибки ϵ (первый постулат РА). Ошибку ϵ считают нормально распределенной с нулевым математическим ожиданием и постоянной дисперсией σ_y^2 , не зависящей ни от z , ни от $f(z)$ (второй постулат РА).

Построение математического описания объекта заключается в нахождении аналитического выражения для функции отклика $f(z)$ и определении дисперсии случайной ошибки σ_y^2 . Решение этой задачи осложняется тем, что вид функции чаще всего априори неизвестен.

Полный алгоритм построения регрессионной модели представлен на рис. 3.4. Рассмотрим его основные этапы.

Выбор модели. Он заключается в выборе вида аппроксимирующего выражения для неизвестной функции отклика $f(z)$. В РА функцию отклика ищут в виде

$$\hat{y} = b_0 + b_1 x_1 + \dots + b_j x_j + \dots + b_n x_n, \quad (3.13)$$

где b_0, b_1, \dots, b_n — неизвестные коэффициенты; x_1, x_2, \dots, x_n — регрессионные переменные, представляющие собой набор известных линейно независимых функций от физических переменных z_1, \dots, z_l . Например, для двух переменных z_1 и z_2 в качестве регрессионных могут быть взяты

$$x_1 = z_1; x_2 = z_2; x_3 = z_1 z_2 \quad (n = 3);$$

$$x_1 = z_1; x_2 = 1/z_2 \quad (n = 2);$$

$$x_1 = z_1 - z_2; x_2 = \ln z_2; x_3 = z_1^2; x_4 = z_2^2 \quad (n = 4).$$

Такой подход дает целый ряд преимуществ. Во-первых, большая свобода выбора числа и вида регрессионных пере-

менных позволяет усложнять либо, наоборот, упрощать модель объекта. Во-вторых, после выбора модели определение функции отклика сводится к нахождению постоянных коэффициентов b_0, b_1, \dots, b_n . В-третьих, линейность модели (3.13) относительно неизвестных коэффициентов b_0, b_1, \dots, b_n и регрессионных переменных x_1, x_2, \dots, x_n дает возможность широко использовать матричный аппарат линей-

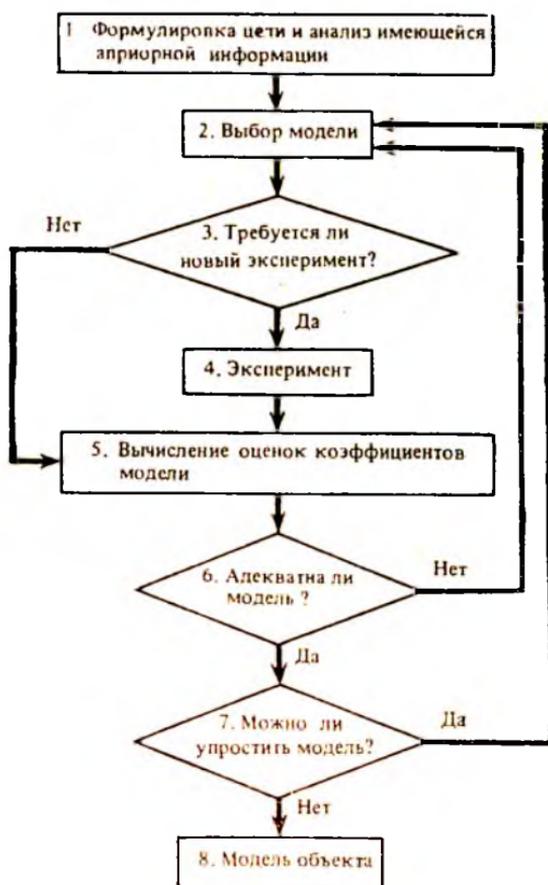


Рис. 3.4

ной алгебры и тем самым добиться максимальной компактности и универсальности при вычислениях.

При выборе числа и вида регрессионных переменных, входящих в (3.13), следует исходить из соображений максимальной простоты модели при одновременном обеспечении приемлемой точности. Большую пользу может принести

привлечение имеющейся физической информации о природе исследуемого объекта (см. гл. 2).

Отметим еще один важный момент, который необходимо учитывать при выборе модели. Это область и способ ее использования, т. е. где и как будет использоваться построенная модель. Ответ на эти вопросы часто предопределяет вид модели. Например, если модель предполагают использовать для нахождения экстремума y по методу крутого восхождения, то нет необходимости вводить в модель (3.13) регрессионные переменные, описывающие нелинейные эффекты, т. е. члены вида $z_k z_l$, z_k^m и т. п.

Эксперимент. Этот этап заключается в наблюдении откликов y_1, y_2, \dots, y_N исследуемого объекта в точках $\mathbf{x}_1, \dots, \mathbf{x}_N$ (в числе которых могут быть и повторяющиеся) из некоторой области пространства регрессионных переменных. Каждая точка \mathbf{x}_i представляет собой конкретный набор x_{i1}, \dots, x_{in} , при котором наблюдается отклик y_i ($i = 1, \dots, N$).

Для того чтобы использовать аппарат матричной алгебры, с каждой точкой пространства регрессионных переменных связывают вектор-строку $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{in})$, в котором для удобства вычислений добавлена еще одна регрессионная переменная x_{i0} , тождественно равная единице. Тогда совокупности данных, полученных в результате эксперимента, могут быть сопоставлены:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{i1} & x_{i2} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix}$$

и вектор-столбец откликов

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_N \end{bmatrix}$$

Таким образом, в матрицах X и Y наблюдается построчное соответствие, так что i -я строка отражает результаты проведения i -го опыта.

Отметим, что матрица X может содержать и повторяющиеся строки, что соответствует дублированию опытов при эксперименте.

Существует два подхода к проведению эксперимента: пассивный и активный. При пассивном эксперименте отсутствует какой-либо определенный порядок в выборе точек проведения эксперимента x . Такой подход оправдан, когда экспериментатор не может поставить опыт в произвольной точке (например, в случае, когда переменные x_1, \dots, x_n являются контролируемыми, но не управляемыми). При активном эксперименте опыты производят в некоторых заранее выбранных точках, образующих план эксперимента. План эксперимента выбирают в зависимости от выбранной модели и осуществляют по определенной схеме, обладающей некоторыми оптимальными свойствами. С точки зрения введенных матричных обозначений переход к активному эксперименту означает наложение ряда ограничений на матрицу X (матрицу X называют матрицей планирования), в результате чего она приобретает определенный вид. Например, в большинстве применяемых планов выполняется условие ортогональности столбцов матрицы планирования:

$$\sum_{i=1}^N x_{ik} x_{il} = 0, \text{ если } k \neq l.$$

В дальнейшем будем отмечать по ходу изложения те особенности, которые вносит в общую схему РА применение математического планирования эксперимента.

Вычисление оценок коэффициентов модели. В РА расчет неизвестных коэффициентов модели (3.13) осуществляют по методу наименьших квадратов (МНК). Его суть заключается в том, что коэффициенты определяют из условия минимума суммы квадратов отклонений точек модели от экспериментально наблюдавшихся откликов:

$$Q_{\text{ост}} = \sum_{i=1}^N (\hat{y}_i - y_i)^2 = \min.$$

Учитывая (3.13), сумма $Q_{\text{ост}}$ может быть записана как

$$Q_{\text{ост}} = \sum_{i=1}^N (b_0 + b_1 x_{i1} + \dots + b_n x_{in} - y_i)^2. \quad (3.14)$$

Таким образом, сумму $Q_{\text{ост}}$ можно рассматривать, как функцию от переменных b_0, b_1, \dots, b_n , поскольку остальные входящие в нее величины есть числа, полученные из эксперимента. Для того чтобы $Q_{\text{ост}}$ была минимальна, необходимо выполнить условие

$$\frac{\partial Q_{\text{ост}}}{\partial b_j} = 0; j = 0, 1, \dots, n. \quad (3.15)$$

Выражение (3.15) представляет собой систему из $(n + 1)$ уравнений для определения $(n + 1)$ неизвестных коэффициентов b_0, b_1, \dots, b_n . Вследствие линейности модели (3.13) относительно коэффициентов b_j система (3.15) линейна. Ее решение удобно вести в матричном виде. Для этого вводят вектор-столбец коэффициентов:

$$\mathbf{B} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

и вектор-столбец откликов модели

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{bmatrix}$$

Тогда $\hat{y}_i = \mathbf{x}_i \mathbf{B}$; $\hat{\mathbf{Y}} = \mathbf{X} \mathbf{B}$.

С использованием введенных обозначений сумма квадратов отклонений $Q_{\text{ост}}$ может быть записана так:

$$Q_{\text{ост}} = (\mathbf{X} \mathbf{B} - \mathbf{Y})^T (\mathbf{X} \mathbf{B} - \mathbf{Y}),$$

где T — знак транспонирования.

Система (3.15) в матричной форме приобретает вид матричного уравнения: $\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{X}^T \mathbf{Y}$, откуда искомый вектор коэффициентов

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (3.16)$$

где $(\mathbf{X}^T \mathbf{X})^{-1}$ — матрица, обратная матрице $\mathbf{X}^T \mathbf{X}$.

Матрицу $\mathbf{X}^T \mathbf{X}$ называют *информационной*. Легко проверить, что это квадратная матрица порядка $(n + 1)$, симметричная относительно своей главной диагонали. Произведе-

дение $\mathbf{X}^T \mathbf{Y}$ представляет собой вектор-столбец, содержащий $(n + 1)$ строк.

Если матрица наблюдений \mathbf{X} содержит большое число повторяющихся строк, т. е. опыты осуществляют с дублированием, то целесообразно изменить схему вычисления вектора коэффициентов \mathbf{B} следующим образом. Пусть матрица \mathbf{X} содержит N_0 различных строк ($N_0 < N$). Исключим из матрицы \mathbf{X} повторяющиеся строки и оставшуюся матрицу \mathbf{X}_0 будем рассматривать как матрицу различных опытов, каждый из которых производят с числом повторов p_i ($i = 1, \dots$

\dots, N_0). Средний отклик в i -м опыте $\bar{y}_i = \frac{1}{p_i} \sum_{l=1}^{p_i} y_{il}$. Тогда всей совокупности наблюдений можно сопоставить: матрицу различных опытов \mathbf{X}_0 , которая содержит N_0 различных строк, соответствующий ей вектор-столбец \mathbf{Y}_0 средних откликов и матрицу дублирования \mathbf{P} :

$$\mathbf{P} = \begin{bmatrix} p_1 & & & 0 \\ & p_2 & & \\ & & \ddots & \\ 0 & & & p_{N_0} \end{bmatrix}.$$

Можно показать, что в этом случае

$$\mathbf{B} = (\mathbf{X}_0^T \mathbf{P} \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{P} \mathbf{Y}_0. \quad (3.17)$$

Результаты вычислений по формулам (3.16) и (3.17) совпадают. Различие состоит в схеме вычислений. При использовании (3.16) не проводят дифференциацию между различными и повторяющимися опытами и поэтому пользуются полной матрицей наблюдений размером $N \times (n + 1)$. Применяя (3.17), за счет дифференциации между различными повторяющимися опытами удается добиться большей компактности при вычислениях, что важно при использовании малых ЭВМ. Особенно нагляден выигрыш в случае равномерного дублирования, когда $p_1 = p_2 = \dots = p_{N_0} = p$. В этом случае матрица дублирования \mathbf{P} в формуле (3.17) сокращается:

$$\mathbf{B} = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{Y}_0. \quad (3.17a)$$

т. е. массив числовой информации уменьшается в p раз.

Согласно (3.16) и (3.17), не при всякой матрице наблюдений определение коэффициентов регрессионной модели возможно. Действительно, для вычисления вектора \mathbf{B} по

(3.16) необходимо, чтобы матрица $X^T X$ была обратима, т. е. существовала бы обратная матрица $(X^T X)^{-1}$. Аналогично, для расчета вектора коэффициентов B по (3.17) нужно, чтобы была обратима матрица $(X_0^T P X_0)$.

Можно показать, что для существования искомого решения необходимо и достаточно, чтобы была обратима матрица $X_0^T X_0$, т. е. успех или неуспех в построении регрессионной модели зависит от того, насколько удачно или неудачно построена матрица различных опытов X_0 . Известно, что матрица обратима, если ее определитель не равен нулю. Отсюда получаем необходимое и достаточное условие для построения регрессионной модели:

$$|X_0^T X_0| \neq 0. \quad (3.18)$$

Если $|X_0^T X_0| = 0$, то задача нахождения коэффициентов не имеет решения. Это, в частности, имеет место, когда $N_0 < n + 1$, т. е. число различных опытов меньше числа искомого коэффициентов. Важно отметить, что дублирование опытов в этом случае не имеет значения.

Уточним, какие изменения вносит в описанную схему вычисления коэффициентов модели активный эксперимент (АЭ). С вычислительной точки зрения он сводится к некоторым ограничениям на матрицу наблюдений X , которую называют матрицей планирования. Наиболее существенным из этих ограничений является ортогональность плана. Для ортогонального плана матрица $X^T X$ имеет диагональный вид и ее обращение не сложно:

$$X^T X = \begin{bmatrix} A & & & \\ & B & & 0 \\ & & \dots & \\ 0 & & & K \end{bmatrix}; \quad (X^T X)^{-1} = \begin{bmatrix} \frac{1}{A} & & & \\ & \frac{1}{B} & & 0 \\ & & \dots & \\ & & & \frac{1}{K} \end{bmatrix}.$$

Обычно используют ортонормированные планы, когда

$$\sum_{i=1}^N x_{ik} x_{il} = \begin{cases} 0, & k \neq l, \\ N, & k = l. \end{cases}$$

В этом случае коэффициенты вычисляют по формуле

$$b_j = \frac{1}{N} \sum_{i=1}^N x_{ij} y_i, \quad j = 0, 1, \dots, n.$$

Если АЭ осуществляют с равномерным дублированием, то это не нарушает ортогональности плана и коэффициенты модели рассчитывают по аналогичной формуле

$$b_j = \frac{1}{N_0} \sum_{i=1}^{N_0} x_{ij} \bar{y}_i, \quad j=0, 1, \dots, n,$$

где \bar{y}_i — средний отклик в i -м опыте.

Таким образом, переход к АЭ существенно упрощает процесс вычисления коэффициентов модели. Кроме того, при АЭ получают диагональную матрицу $X^T X$, т. е. наиболее далекую от вырождения (вырожденной называют матрицу, определитель которой равен нулю).

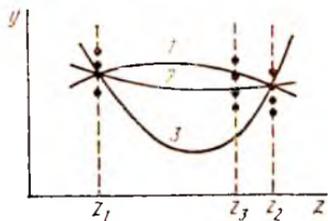


Рис. 3.5

Планированию эксперимента должно быть уделено серьезное внимание, так как при неудачном планировании матрица $X^T X$ может быть плохо обусловлена (плохо обусловленной называют матрицу, определитель которой очень мал по сравнению с другими значениями).

Плохая обусловленность матрицы $X^T X$ приводит к тому, что обратная матрица $(X^T X)^{-1}$ состоит из больших чисел, так как ее члены обратно пропорциональны определителю исходной матрицы. Поскольку дисперсия оценок коэффициентов b_j модели пропорциональна диагональным членам матрицы $(X^T X)^{-1}$, получают модель, в которой численные значения коэффициентов имеют большую неопределенность. При практическом использовании такая модель имеет слишком низкую точность предсказания.

Рассмотрим модель $\hat{y} = b_0 + b_1 z + b_2 z^2$. Допустим, что для нахождения параметров модели поставили ряд повторных опытов в точках z_1 и z_2 (рис. 3.5).

При таком планировании не удастся построить требуемую модель, так как располагают данными двух различных опытов, тогда как выбранная модель содержит три неизвестных коэффициента — b_0 , b_1 и b_2 . Тщетность усилий объясняется тем, что пытаются провести параболу через две точки, когда известно, что таких парабол существует бесчисленное множество (например, кривые 1—3). В данном случае матрица различных опытов X_0 содержит всего две строки, следовательно, при любом числе повторов матрица $X_0^T X_0$ вырождена.

Очевидно, что необходимо поставить хотя бы один опыт в точке, отличной от z_1 и z_2 . Допустим, что этот опыт (или серия опытов) проведен в точке z_3 . Задача нахождения параметров модели формально стала разрешимой, так как располагают тремя различными опытами. Однако условия второго и третьего опытов различаются незначительно, поэтому следует ожидать большой неопределенности в результатах вычисления коэффициентов модели. Действительно, если кривая 3 не годится в качестве модели, то кривые 1 и 2 по-прежнему хорошо согласуются с экспериментальными данными, тогда как в интервале между точками z_1 и z_3 эти кривые имеют различный вид. Этой геометрически понятной ситуации алгебраически соответствует плохая обусловленность матрицы $X_0^T X_0$.

Для того чтобы матрица $X_0^T X_0$ была хорошо обусловленной, нужно выбрать точку z_3 строго посередине между z_1 и z_2 . Целесообразность такого выбора легко понять из геометрических соображений.

В данном примере рассмотрена квадратичная модель. Очевидно, что в случае линейной модели $\bar{y} = b_0 + b_1 z$ проведение эксперимента в точках z_1, z_2, z_3 не вызывает трудностей, с которыми пришлось столкнуться в случае квадратичной модели. Следовательно, судить о том, хорош или плох план проведения эксперимента, не имеет смысла, если не имеется в виду конкретная модель, для построения которой используют результаты этого эксперимента.

Проанализируем процесс определения коэффициентов регрессионной модели. Любое вычислительное устройство с ограниченным объемом памяти работает с конечной погрешностью, обусловленной округлением промежуточных результатов. Эта погрешность может быть особенно большой при обработке результатов пассивного эксперимента, где часто приходится сталкиваться с плохо обусловленной информационной матрицей $X^T X$. Если учесть, что для решения задач оптимизации чаще всего используют малые ЭВМ, вычислительные возможности которых в настоящее время ограничены, то уменьшение погрешностей, связанных с округлением, является важной задачей. Для уменьшения погрешностей целесообразно центрировать и нормировать данные эксперимента. Для этого в матрице наблюдений X и векторе откликов Y переходят от абсолютных значений величин к их отклонениям от средних значений в эксперименте: $x_{ij} \rightarrow x_{ij} - \bar{x}_j$; $y_i \rightarrow y_i - \bar{y}$, где $\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$; $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$. Центрирование данных означает

переход к модели

$$\widehat{y} - \bar{y} = b_1(x_1 - \bar{x}_1) + \dots + b_n(x_n - \bar{x}_n), \quad (3.19)$$

которая не содержит коэффициента b_0 , а следовательно, столбец фиктивных переменных x_{i0} может быть исключен. Центрированные матрицы \mathbf{X}_c , \mathbf{Y}_c используют для вычисления коэффициентов модели (3.19) по формуле (3.16) или (3.17).

Центрирование кроме уменьшения размерности массива данных (информационная матрица $\mathbf{X}_c^T \mathbf{X}_c$ имеет порядок n) уменьшает влияние ошибок округления, так как в центрированной модели фигурируют не сами переменные x , y , а их отклонения от средних значений. Например, в экспериментах по диффузионному легированию температуру реактора изменяют в пределах ± 100 °C около средней температуры, которая имеет порядок 1000 °C; центрирование в данном случае облегчает вычисления.

Процесс вычисления облегчается и при переходе к нормированным переменным. Нормирование означает переход к модели

$$\widehat{y} - \bar{y} = b_1 \frac{x_1 - \bar{x}_1}{\Delta_1} + \dots + b_n \frac{x_n - \bar{x}_n}{\Delta_n}, \quad (3.20)$$

где $\Delta_1, \dots, \Delta_n$ — интервалы нормирования.

Нормирование приводит к тому, что все элементы матриц становятся одного порядка, что минимизирует влияние ошибок округления на результат при операциях над матрицами.

Проверка модели на адекватность. Ценность любого математического описания заключается в его предсказующих свойствах. Поэтому за отправную точку при оценке качеств модели выбирают невязку $d = y - \widehat{y}$, т. е. разность между истинным откликом y и тем значением \widehat{y} , которое предсказывает модель.

В качестве меры точности построенной модели в РА используют математическое ожидание квадрата невязки. Если оно во всей рабочей области пространства переменных практически не отличается от дисперсии чистой ошибки σ_y^2 , то построенная модель адекватна исследуемому объекту.

Проверку адекватности модели осуществляют исходя из следующих соображений. Остаточную сумму квадратов $Q_{\text{ост}}$ раскладывают на две составляющие:

$$Q_{\text{ост}} = \sum_{i=1}^{N_0} \sum_{l=1}^{p_i} (y_{il} - \widehat{y}_i)^2 = \sum_{i=1}^{N_0} \sum_{l=1}^{p_i} (y_{il} - \bar{y}_i)^2 + \\ + \sum_{i=1}^{N_0} p_i (\bar{y}_i - \widehat{y}_i)^2 = Q_y + Q_{\text{ад}}. \quad (3.21)$$

Здесь индекс i относится к различным опытам ($i = 1, 2, \dots, N_0$), а l — к повторам в i -м опыте ($l = 1, \dots, p_i$).

С каждой из сумм, входящих в (3.21), связывают число степеней свободы f , которое равно числу случайных величин в сумме минус число коэффициентов (констант) в ней, которые вычислены ранее по этим же случайным величинам. Так, $Q_{\text{ост}}$ находят по N значениям откликов y_{il} ; кроме них в сумме фигурируют \widehat{y}_i , вычисленные по модели. Но модель содержит $(n + 1)$ коэффициент, которые определены ранее по тем же N значениям откликов y_{il} . Следовательно, число степеней свободы суммы $Q_{\text{ост}}$ равно $f_{\text{ост}} = N - (n + 1)$. Аналогично, суммы Q_y и $Q_{\text{ад}}$ имеют число степеней свободы $f_y = N - N_0$ и $f_{\text{ад}} = N_0 - (n + 1)$, так что

$$f_{\text{ост}} = f_y + f_{\text{ад}}. \quad (3.22)$$

Сумма Q_y обусловлена действием чистой ошибки ϵ . Поэтому выборочная оценка дисперсии σ_y^2 чистой ошибки

$$s_y^2 = \frac{Q_y}{f_y} = \frac{1}{N - N_0} \sum_{i=1}^{N_0} \sum_{l=1}^{p_i} (y_{il} - \bar{y}_i)^2. \quad (3.23)$$

Сумма $Q_{\text{ад}}$ зависит не только от чистой ошибки ϵ , но и от того, адекватна ли модель описываемому объекту. Поэтому можно рассматривать некоторую дисперсию $\sigma_{\text{ад}}^2$, характеризующую качество построенной модели; ее называют *дисперсией адекватности*. Выборочная оценка дисперсии адекватности

$$s_{\text{ад}}^2 = \frac{Q_{\text{ад}}}{f_{\text{ад}}} = \frac{1}{N_0 - (n + 1)} \sum_{i=1}^{N_0} p_i (\bar{y}_i - \widehat{y}_i)^2. \quad (3.24)$$

Аналогично вводят остаточную дисперсию $\sigma_{\text{ост}}^2$, оценкой которой является

$$s_{\text{ост}}^2 = \frac{Q_{\text{ост}}}{f_{\text{ост}}} = \frac{1}{N - (n + 1)} \sum_{i=1}^{N_0} \sum_{l=1}^{p_i} (y_{il} - \widehat{y}_i)^2. \quad (3.25)$$

Если модель адекватна, то дисперсия $\sigma_{ад}^2$ обусловлена чистой ошибкой. Поэтому проверку адекватности модели можно осуществлять как проверку статистической гипотезы $H_0: \sigma_{ад}^2 = \sigma_y^2$. Как известно, гипотезу о равенстве дисперсий проверяют по F -критерию Фишера. Поэтому для проверки адекватности регрессионной модели следует:

1) вычислить отношение $F = s_{ад}^2/s_y^2$;

2) по заданному уровню значимости α и числам степеней свободы $f_{ад}$ и f_y в таблице распределения Фишера найти критическое значение критерия F_{α, f_1, f_2} , где $f_1 = f_{ад}$, $f_2 = f_y$;

2) сравнить значения F и F_{α, f_1, f_2} ; если $F < F_{\alpha, f_1, f_2}$, то можно признать, что гипотеза об адекватности модели не противоречит опытным данным, если $F \geq F_{\alpha, f_1, f_2}$, то модель неадекватна.

Необходимо отметить, что:

1) построение регрессионной модели невозможно, если число различных опытов меньше числа коэффициентов модели, т. е. $N_0 < n + 1$. В случае, когда $N_0 = n + 1$, модель может быть построена, но проверка ее адекватности неосуществима, так как число степеней свободы суммы $Q_{ад}$ равно нулю. Ценность такой модели при ее дальнейшем использовании сомнительна;

2) если эксперимент проводят без повторов ($N = N_0$), то адекватность модели не может быть проверена на основе данного эксперимента, так как число степеней свободы для оценки σ_y^2 равно нулю. Поэтому построение модели на основе эксперимента без дублирования опытов имеет смысл либо при известной величине σ_y^2 , либо в случае, когда имеется априорная информация об адекватности выбранной модели, т. е. фактически, когда до опыта известен вид функции отклика $f(\mathbf{Z})$ и необходимо определить неизвестные параметры, входящие в эту функцию;

3) построение математического описания исследуемого объекта заключается не только в установлении функции отклика $f(\mathbf{Z})$, но и в нахождении дисперсии случайной ошибки σ_y^2 . Если модель адекватна, то сумма $Q_{ост}$ зависит только от чистой ошибки ε , поэтому для вычисления оценки дисперсии σ_y^2 вместо выражения (3.23) следует использовать выражение (3.25), где $Q_{ост}$ можно рассчитать как сумма $Q_y + Q_{ад}$. Эта оценка является более точной, так как имеет большее число степеней свободы.

Упрощение модели. Пусть построена модель в виде $\hat{y} = \bar{y} + b_1(x_1 - \bar{x}_1) + \dots + b_n(x_n - \bar{x}_n)$.

Первый вопрос, который возникает после установления адекватности этой модели, — это вопрос о значимости вклада в модель всех членов выражения, стоящих после \bar{y} , т. е. следует уточнить, насколько оправдано в статистическом смысле усложнение простейшей модели $\hat{y} = \bar{y}$ за счет вве-

дения всех последующих членов (проверка значимости регрессии в целом).

Для проверки значимости регрессии общую сумму квадратов отклонений откликов от среднего значения \bar{y} раскладывают следующим образом:

$$Q_{\text{общ}} = \sum_{i=1}^{N_0} \sum_{l=1}^{p_i} (y_{il} - \bar{y})^2 = \sum_{i=1}^{N_0} \sum_{l=1}^{p_i} (y_{il} - \hat{y}_i)^2 + \sum_{i=1}^{N_0} p_i (\hat{y}_i - \bar{y})^2 = Q_{\text{ост}} + Q_{\text{рег}}, \quad (3.26)$$

Соответствующие этим суммам числа степеней свободы связаны соотношением

$$f_{\text{общ}} = f_{\text{ост}} + f_{\text{рег}}; \quad f_{\text{ост}} = N - (n + 1); \quad f_{\text{рег}} = n. \quad (3.27)$$

Сумма $Q_{\text{общ}}$ обусловлена разбросом откликов относительно среднего уровня \bar{y} , ей может быть поставлена в соответствие дисперсия $\sigma_{\text{общ}}^2$, оценка которой $s_{\text{общ}}^2 = Q_{\text{общ}}/f_{\text{общ}}$.

Если $\sigma_{\text{общ}}^2$ для адекватной модели незначимо больше $\sigma_{\text{ост}}^2$, то добавление в модель составляющих $b_1(x_1 - \bar{x}_1)$, ..., $b_n(x_n - \bar{x}_n)$ практически не повышает точность описания исследуемого объекта, следовательно, модель может быть выражена в виде $\hat{y} = \bar{y}$.

Проверку значимости регрессии в целом проводят как проверку статистической гипотезы $H_0: \sigma_{\text{общ}}^2 = \sigma_{\text{ост}}^2$. Практически эту проверку осуществляют с помощью критерия Фишера в следующем порядке:

1) вычисляют отношение $F = s_{\text{общ}}^2/s_{\text{ост}}^2$;

2) по заданному уровню значимости α и числам степеней свободы $f_{\text{общ}}$ и $f_{\text{ост}}$ в таблице распределения Фишера находят критическое значение F_{α, f_1, f_2} , где $f_1 = f_{\text{общ}}$, $f_2 = f_{\text{ост}}$;

3) сравнивают значения F и F_{α, f_1, f_2} ; если $F < F_{\alpha, f_1, f_2}$, то гипотеза H_0 не противоречит опытным данным, т. е. регрессия является незначимой, если $F \geq F_{\alpha, f_1, f_2}$, то регрессия является значимой.

После установления значимости регрессии в целом возникает второй вопрос: нельзя ли упростить модель путем исключения из нее некоторых регрессионных переменных?

Вопрос о возможности исключения из модели переменной x_j может быть решен путем анализа значимости соответствующего коэффициента b_j . Так как величина b_j , полу-

ченная путем обработки результатов эксперимента, является случайной, зависящей от конкретных значений откликов y_{ij} , то значимость b_j можно определить путем проверки статистической гипотезы $H_0: b_j = 0$. Если значение b_j отличается от нуля незначимо (в статистическом смысле), то переменная x_j может быть исключена из модели.

Для проверки гипотезы $b_j = 0$ необходимо прежде всего оценить меру случайности в определении вектора коэффициентов \mathbf{B} из эксперимента с ограниченным числом наблюдений N . Такой мерой является матрица дисперсий — ковариаций коэффициентов. Диагональные элементы этой матрицы есть дисперсии, а внедиагональные члены — ковариации коэффициентов b_0, b_1, \dots, b_n . В РА показано, что матрица дисперсий-ковариаций

$$\mathbf{V}(\mathbf{B}) = (\mathbf{X}^T \mathbf{X})^{-1} \sigma_y^2. \quad (3.28)$$

При дублировании опытов матрица $\mathbf{V}(\mathbf{B})$ может быть вычислена и так:

$$\mathbf{V}(\mathbf{B}) = (\mathbf{X}_0^T \mathbf{P} \mathbf{X}_0)^{-1} \sigma_y^2, \quad (3.29)$$

где \mathbf{P} — матрица дублирования.

Для равномерного дублирования, когда $p_1 = p_2 = \dots = p_{N_0} = p$, формула (3.29) упрощается:

$$\mathbf{V}(\mathbf{B}) = \frac{1}{p} (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \sigma_y^2. \quad (3.30)$$

Из формул (3.28)—(3.30) следует, что рассеяние оценок истинных коэффициентов модели зависит не только от дисперсии чистой ошибки σ_y^2 , но также и от информационной матрицы $\mathbf{X}^T \mathbf{X}$ (или $\mathbf{X}_0^T \mathbf{X}_0$), т. е. от планирования. В этой связи следует еще раз подчеркнуть достоинства активного эксперимента. Действительно, как уже отмечалось, при АЭ чаще всего используют ортогональные планы, в которых $\mathbf{X}^T \mathbf{X}$ имеет диагональную форму, а следовательно, и $(\mathbf{X}^T \mathbf{X})^{-1}$ также диагональна. Тогда согласно (3.28) в таком эксперименте оценки отдельных коэффициентов модели не зависят друг от друга. Кроме того, при АЭ стремятся к такому выбору матрицы планирования \mathbf{X} , при котором дисперсии коэффициентов модели минимальны (D -оптимальные планы). Таким образом, в АЭ имеется возможность целенаправленно влиять на свойства оценок параметров выбранной модели.

Отметим также, что если σ_y^2 неизвестна, то в формулах (3.28)—(3.30) вместо нее используют s_y^2 , получаемую по (3.25) для адекватной модели.

Проверку значимости каждого коэффициента b_j проводят по критерию Стьюдента в следующем порядке:

1) вычисляют оценку дисперсии коэффициента $s_{b_j}^2$, как соответствующий диагональный элемент матрицы $\mathbf{V}(\mathbf{B})$ по одной из формул (3.28)—(3.30):

2) определяют $t = |b_j| s_{b_j}$;

3) по заданному уровню значимости α и числу степеней свободы $f = N - (n + 1)$ в таблице распределения Стьюдента отыскивают критическое значение $t_{\alpha, f}$;

4) если $t < t_{\alpha, f}$, то коэффициент b_j может быть признан незначимым и переменную x_j исключают из модели; если $t \geq t_{\alpha, f}$, то b_j является значимым.

После проверки коэффициентов b_j на значимость и упрощения исходной модели необходимо, в общем случае вернуться к вычислению коэффициентов и проделать все процедуры в соответствии со схемой рис. 3.4 для упрощенной модели.

Отметим, однако, что для этого не нужно заново вычислять матрицу $\mathbf{X}^T \mathbf{X}$. Достаточно из матрицы $\mathbf{X}^T \mathbf{X}$ исходной модели вычеркнуть j -ю строку и j -й столбец, соответствующие исключенной переменной x_j . Вычеркнув строки и столбцы, соответствующие исключенным переменным, получим информационную матрицу $\mathbf{X}^T \mathbf{X}$ для упрощенной модели. Далее необходимо обратить эту матрицу, и так далее до получения новых коэффициентов для упрощенной модели.

В том случае, когда матрица $\mathbf{X}^T \mathbf{X}$ исходной модели диагональна (ортогональный план), не нужно пересчитывать коэффициенты упрощенной модели, так как они полностью совпадают с соответствующими коэффициентами исходной модели.

Точность регрессионной модели. После того как получена адекватная регрессионная модель и оценена дисперсия чистой ошибки σ_y^2 , модель может быть использована для предсказания отклика в каждой конкретной точке $\mathbf{x}_0 = (1, x_{01}, x_{02}, \dots, x_{0n})$ пространства переменных.

Следует помнить, что $\hat{y}_0 = \mathbf{x}_0 \mathbf{B}$ — предсказываемое значение функции отклика, т. е. среднего отклика \bar{y}_0 при многократном повторении опытов в точке \mathbf{x}_0 . При этом \hat{y}_0 представляет собой точечную оценку \bar{y}_0 и может не совпадать с истинным значением \bar{y}_0 , так как коэффициенты моде-

ли b_0, b_1, \dots, b_n определены из эксперимента с ограниченным числом наблюдений и, следовательно, имеют рассеяние, определяемое матрицей дисперсий-ковариаций $\mathbf{V}(\mathbf{B})$. Соответственно и \widehat{y}_0 имеет рассеяние, значение которого определяется дисперсией:

$$D\{\widehat{y}_0\} = \mathbf{x}_0 \mathbf{V}(\mathbf{B}) \mathbf{x}_0^T = \mathbf{x}_0 (\mathbf{X}^T \mathbf{X})^{-1} \sigma_y^2 \mathbf{x}_0^T. \quad (3.31)$$

Используя выражения (3.31), можно построить доверительный интервал для \widehat{y}_0 :

$$\widehat{y}_0 \pm t_{\alpha; f} \sqrt{D\{\widehat{y}_0\}} = \widehat{y}_0 \pm t_{\alpha; f} \sqrt{\mathbf{x}_0 (\mathbf{X}^T \mathbf{X})^{-1} s_y^2 \mathbf{x}_0^T}, \quad (3.32)$$

где s_y^2 — оценка дисперсии ошибки; $t_{\alpha; f}$ — табличное значение критерия Стьюдента при числе степеней свободы, с которым была определена оценка s_y^2 .

Истинное значение \bar{y}_0 лежит в пределах этого интервала с вероятностью $(1 - \alpha)$.

Если важен не средний отклик \bar{y}_0 , а отклик y_0 при однократном проведении опыта в точке \mathbf{X}_0 , то к дисперсии $D\{\widehat{y}_0\}$ следует прибавить дисперсию σ_y^2 , учитывающую независимое действие чистой ошибки ϵ . Тогда доверительный интервал для y_0 получают из выражения

$$\widehat{y}_0 \pm t_{\alpha; f} \sqrt{s_y^2 [\mathbf{x}_0 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_0^T + 1]}. \quad (3.33)$$

В формулах (3.31)—(3.33) \mathbf{x}_0 — вектор-строка соответствующей точки наблюдения. Из формул (3.32), (3.33) следует, что доверительный интервал, т. е. точность предсказания, прежде всего зависит от матрицы $\mathbf{X}^T \mathbf{X}$, т. е. от планирования. Поэтому все замечания, которые были сделаны относительно преимуществ АЭ с точки зрения минимизации рассеяния оценок коэффициентов модели, в равной степени относятся к минимизации доверительного интервала для отклика, предсказываемого моделью.

Кроме того, точность предсказания оказывается различной для разных точек \mathbf{x}_0 . Наилучшая точность достигается в центре эксперимента, на основе которого были вычислены коэффициенты модели (см. центрирование матриц наблюдений \mathbf{X}). По мере удаления от этой точки точность ухудшается.

Таблица 3.2

τ, с	$h_{21,0}$			
10	24	28		
20	25	26	30	
30	27	30	31	
40	30	32	32	
50	34	37		
60	34	33	36	34
70	36			
80	34	36	37	35
90	30	33	30	

Вектор коэффициентов модели согласно (3.17)

$$\mathbf{B} = (\mathbf{X}_0^T \mathbf{P} \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{P} \mathbf{Y}_0.$$

В данном случае

$$\mathbf{X}_0^T \mathbf{P} \mathbf{X}_0 = \begin{bmatrix} 25 & 129 & 831 \\ 129 & 831 & 5991 \\ 831 & 5991 & 45963 \end{bmatrix};$$

$$(\mathbf{X}_0^T \mathbf{P} \mathbf{X}_0)^{-1} = \begin{bmatrix} 0,7 & -0,287 & 0,025 \\ -0,287 & 0,138 & -0,013 \\ 0,025 & -0,013 & 0,001 \end{bmatrix};$$

$$\mathbf{X}_0^T \mathbf{P} \mathbf{Y}_0 = \begin{bmatrix} 794 \\ 4256 \\ 27764 \end{bmatrix}.$$

Окончательно по (3.17) получаем

$$\mathbf{B} = \begin{bmatrix} 0,7 & -0,287 & 0,025 \\ -0,287 & 0,138 & -0,013 \\ 0,025 & -0,013 & 0,001 \end{bmatrix} \begin{bmatrix} 794 \\ 4256 \\ 27764 \end{bmatrix} = \begin{bmatrix} 20,2 \\ 4,39 \\ -0,33 \end{bmatrix}.$$

Таким образом, наша модель $\hat{y} = 20,2 + 4,39x_1 - 0,33x_2$, т. е. коэффициент усиления транзисторов связан с продолжительностью гравления перед герметизацией зависимостью $\hat{h}_{21,0} = 20,2 + 0,439\tau - 0,33 \cdot 10^{-2}\tau^2$.

Проверим адекватность построенной модели. Для этой цели составим табл. 3.3. Заметим, что в шестой столбец заносятся значения откликов, вычисленные по модели $\hat{y} = \hat{h}_{21,0} = 20,2 + 0,439\tau - 0,33 \cdot 10^{-2}\tau^2$, а в восьмой — квадраты отклонений экспериментально наблюдавшихся откликов y_{il} от их среднего значения

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_{il} = \frac{1}{25} 794 = 31,76.$$

Таблица 3.3

№ опыта i	№ повто- рения l	Отклик y_{il}	Средний отклик \bar{y}_i	Разность $(y_{il} - \bar{y}_i)^2$	Модель \hat{y}_i	Разность $\rho_i(\bar{y}_i - \hat{y}_i)^2$	Разность $(y_{il} - \bar{y}_i)^2$
1	1	24	26,0	4	24,2	6,5	60,2
	2	28		4			14,1
2	1	25	27,0	4	27,6	1,1	45,7
	2	26		1			33,2
	3	30		9			3,1
3	1	27	29,3	5,3	30,3	3,0	22,7
	2	30		0,5			3,1
	3	31		2,9			0,6
4	1	30	31,3	1,7	32,4	3,6	3,1
	2	32		0,5			0
	3	32		0,5			0
5	1	34	35,5	2,3	33,8	5,8	5,0
	2	37		2,3			27,4
6	1	34	34,4	0,1	34,5	0,2	5,0
	2	33		1,7			1,5
	3	36		2,9			17,9
	4	34		0,1			5,0
7	1	36	36,0	0	34,6	2,0	17,9
8	1	34	35,5	2,3	34,0	9,0	5,0
	2	36		0,3			17,9
	3	37		2,3			27,4
	4	35		0,3			10,5
9	1	30	31,0	1,0	32,8	9,7	3,1
	2	33		4,0			1,5
	3	30		1,0			3,1
Сумма $\Sigma y_{il} = 794$			$Q_y = 51,1$		$Q_{ад} = 40,9$		$Q_{общ} = 334$

Для проверки гипотезы об адекватности необходимо вычислить оценки дисперсий σ_y^2 и $\sigma_{ад}^2$. Используем для этого формулы (3.23) и (3.24), взяв из табл. 3.3 $Q_y = 51,1$; $Q_{ад} = 40,9$ и учтя, что числа степеней свободы $f_y = N - N_0 = 25 - 9 = 16$; $f_{ад} = N_0 - (n + 1) = 9 - 3 = 6$.

Тогда $s_y^2 = Q_y/f_y = 3,2$; $s_{ад}^2 = Q_{ад}/f_{ад} = 6,8$.

В соответствии с порядком проверки гипотезы $\sigma_y^2 = \sigma_{ад}^2$ определяем: а) отношение $F = s_{ад}^2/s_y^2 = 2,1$, б) критическое значение критерия Фишера из справочных таблиц для $f_1 = f_{ад} = 6$; $f_2 = f_y = 16$. Для уровня значимости $\alpha = 0,05$ это значение равно 2,7. Поскольку $F < F_{0,05; 6; 16}$, модель можно считать адекватной.

Так как модель оказалась адекватной, скорректируем оценку дисперсии чистой ошибки, применив для этого формулу (3.25)

$$s_y^2 = s_{\text{ост}}^2 = Q_{\text{ост}}/f_{\text{ост}} = (Q_y + Q_{\text{ад}}) / (f_y + f_{\text{ад}}) = 4,2,$$

Проверим значимость регрессии в целом. Для этого из табл. 3.3 найдем $s_{\text{общ}}^2 = Q_{\text{общ}}/f_{\text{общ}} = 13,9$. Тогда $F = s_{\text{общ}}^2/s_{\text{ост}}^2 = 3,3$. Критическое значение критерия Фишера берем из таблицы Фишера, положив $\alpha = 0,05$; $f_1 = f_{\text{общ}} = N - 1 = 24$; $f_2 = f_{\text{ост}} = N - (n + 1) = 22$; $F_{0,05; 24; 22} = 2,05$.

Поскольку $F > F_{0,05; 24; 22}$, регрессия в целом значима, т. е. наша модель существенно лучше описывает исследуемый объект, нежели простейшая модель $\hat{y} = \bar{y} = 31,76$.

Попытаемся упростить построенную модель путем исключения отдельных переменных. Согласно (3.29), матрица дисперсий — ковариаций коэффициентов b_j

$$V(B) = (X_0^T P X_0)^{-1} s_y^2 = \begin{bmatrix} 0,700 & -0,287 & 0,025 \\ -0,287 & 0,14 & -0,013 \\ 0,025 & -0,013 & 0,001 \end{bmatrix} \cdot 4,2.$$

Отсюда $s_{b_0}^2 = 2,9$; $s_{b_1}^2 = 0,6$; $s_{b_2}^2 = 0,004$. Тогда $t_0 = 20,2/\sqrt{2,9} = 11,9$; $t_1 = 4,39/\sqrt{0,6} = 5,7$; $t_2 = 0,33/\sqrt{0,004} = 5,2$.

Зададимся уровнем значимости $\alpha = 0,05$ и для числа степеней свободы $f = N - (n + 1) = 22$, с которым была определена оценка дисперсии σ_y^2 ; из справочных таблиц для t -критерия Стьюдента находим $t_{0,05; 22} = 2,07$. Поскольку все t_j больше этого критического значения, то все коэффициенты b_j значимы и, следовательно, ни один из членов построенной модели не может быть исключен.

Таким образом, следует принять в качестве рабочей модели

$$\hat{y} = 20,2 + 4,39x_1 - 0,33x_2 = 20,2 + 0,439\tau - 0,33 \cdot 10^{-2} \tau^2.$$

Оценим точность полученной модели, построив для этого доверительные границы для отклика y_0 при проведении однократного опыта в точке x_0 . Эти границы строятся, как $\hat{y}_0 \pm \Delta_0$, где Δ_0 находят в соответствии с формулой (3.33):

$$\Delta_0 = t_{\alpha; f} \sqrt{s_y^2 [X_0 (X_0^T P X_0)^{-1} X_0^T + 1]}.$$

Выберем $\alpha = 0,05$, т. е. зададимся доверительной вероятностью 0,95. Тогда $t_{0,05; 22} = 2,07$. Воспользовавшись полученными ранее значениями s_y^2 и $(X_0^T P X_0)^{-1}$, вычислим доверительные границы для различных точек x_0 , занося результаты в табл. 3.4.

Полученные доверительные границы, а также линии регрессии представлены на рис. 3.6. Эти результаты для $\tau = 40$ с, например, могут быть интерпретированы так: травление транзистора в течение 40 с перед его герметизацией обеспечивает с доверительной вероятностью 0,95 значение коэффициента усиления $h_{21,0} = 28 \div 36,8$. Данное утверждение справедливо, если ТП производства транзисторов остается стабильным, т. е. все характеристики процесса, а также

сырья продолжают оставаться теми же, что и в период построения модели.

Таблица 3.4

№ точки	Точка x_0	Отклик модели \hat{y}_0	Δ_0	$\hat{y}_0 + \Delta_0$	$\hat{y}_0 - \Delta_0$
1	$x_{01} = (1, 1, 1)$	24,2	4,8	29,0	19,4
2	$x_{02} = (1, 2, 4)$	27,6	4,5	32,1	23,1
3	$x_{03} = (1, 3, 9)$	30,3	4,4	34,7	25,9
4	$x_{04} = (1, 4, 16)$	32,4	4,4	36,8	28,0
5	$x_{05} = (1, 5, 25)$	33,8	4,4	38,2	29,4
6	$x_{06} = (1, 6, 36)$	34,5	4,4	38,9	30,1
7	$x_{07} = (1, 7, 49)$	34,6	4,4	39,0	30,2
8	$x_{08} = (1, 8, 64)$	34,0	4,4	38,4	29,6
9	$x_{09} = (1, 9, 81)$	32,8	4,7	37,5	28,1

§ 3.4. ПОСТРОЕНИЕ ПЛАНОВ ЭКСПЕРИМЕНТА ДЛЯ ОПИСАНИЯ ТП ПОЛИНОМИАЛЬНЫМИ МОДЕЛЯМИ

Оценку зависимости $y = f(x)$, отражающей влияние технологических режимов на выходной параметр процесса, при активном эксперименте осуществляют путем преднамеренного изменения факторов x и наблюдения за изменениями y . Как правило, при активном эксперименте устанавливают опорную (нулевую) точку x_0 в факторном пространстве, около которой нужно получить указанную зависимость. При варьировании факторов выбранную величину Δx_j , $j = \overline{1, k}$ откладывают слева и справа от x_{0j} (рис. 3.7, а). Поскольку ищется многомерная (k -мерная) зависимость, необходимо выбрать (спланировать) такое минимальное число точек в k -мерном пространстве, которое позволило бы достаточно точно построить модель $y = f(x)$.

Рассмотрим задачу планирования многофакторных экспериментов, т. е. составления планов их проведения, начиная с простейших случаев, когда необходимо построить модели первого порядка.

Построение планов эксперимента первого порядка. Пусть $k = 1$, т. е. имеется одномерная зависимость $y = f(x_1)$. В этом случае модель имеет вид $y = b_0 + b_1 x_1$. Для ее построения достаточно двух точек или точнее двух опытов, например, в точках $x_1 = x_{1\min}$ и $x_1 = x_{1\max}$. Графически план и модель эксперимента изображены на рис. 3.7, б. Матрица плана имеет вид

$$X = \begin{bmatrix} x_{1 \min} \\ x_{1 \max} \end{bmatrix} \begin{matrix} \rightarrow y_1 \\ \rightarrow y_2, \end{matrix} \quad (3.34)$$

где y_1, y_2 — отклики в точках 1, 2.

Пусть $k = 2$, т. е. имеется двухфакторная зависимость $y = f(x_1, x_2)$. Модель имеет вид $\hat{y} = b_0 + b_1x_1 + b_2x_2$, т. е. представляет собой плоскость. Для ее построения необходимо провести опыты в двухмерном пространстве.

Фактор x_2 меняем также на двух уровнях $x_{2\min}$ и $x_{2\max}$. При проведении опытов, чтобы получить информацию о зависимости $y = f(x_1, x_2)$ во всем факторном пространстве, необходимо брать сочетания выбранных уровней x_1 и x_2 . Уточним, при каких условиях могут быть проведены опыты в эксперименте:

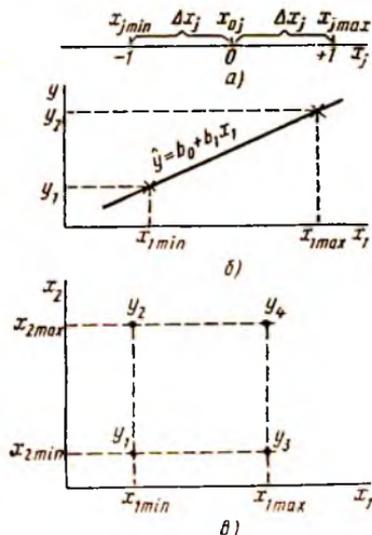
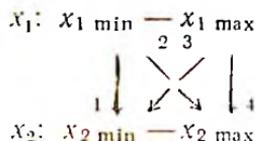


Рис. 3.7



Согласно полученной схеме, при выбранных уровнях (для построения моделей первого порядка достаточно варьировать факторы на двух уровнях) можно провести опыты в четырех точках. При этом матрица плана имеет вид

$$X = \begin{bmatrix} x_{1 \min} & x_{2 \min} \\ x_{1 \min} & x_{2 \max} \\ x_{1 \max} & x_{2 \min} \\ x_{1 \max} & x_{2 \max} \end{bmatrix} \begin{matrix} \rightarrow y_1 \\ \rightarrow y_2 \\ \rightarrow y_3 \\ \rightarrow y_4. \end{matrix} \quad (3.35)$$

Графическое изображение плана приведено на рис. рис. 3.7, в.

Следует отметить, что подобное изображение матриц планов не совсем удобно для аналитических выкладок и технических приложений. Поэтому варьируемые факторы кодируют с помощью простой формулы

$$x_{jн} = (x_{jн} - x_{jон}) / (\Delta x_{jн}),$$

где $x_{j0н} = 0,5(x_{jmax} + x_{jmin})$; $\Delta x_{jn} = 0,5(x_{jmax} - x_{jmin})$, а индексы «к» и «н» соответственно обозначают кодированное и натуральное значение фактора.

Нетрудно увидеть, что при $x_{jn} = x_{jmin}$ и $x_{jk} = -1$; при $x_{jn} = x_{jmax}$ и $x_{jk} = 1$; при $x_{jn} = x_{j0н}$ $x_{jk} = 0$, т. е. для выбранных x_{j0} , Δx_j можно любое значение x_j в натуральных единицах измерения перевести в кодированные значения. В дальнейшем для простоты индекс «к» у кодированных факторов будем опускать. Следовательно, матрицу плана (3.35) можно переписать так:

$$X = \begin{bmatrix} -1 & -1 \\ -1 & +1 \\ +1 & -1 \\ +1 & +1 \end{bmatrix} \begin{matrix} \rightarrow y_1 \\ \rightarrow y_2 \\ \rightarrow y_3 \\ \rightarrow y_4 \end{matrix} \quad (3.36)$$

Матрица независимых переменных для этого плана

$$X = \begin{bmatrix} x_0 & x_1 & x_2 \\ +1 & -1 & -1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & +1 & +1 \end{bmatrix} \begin{matrix} \rightarrow y_1 \\ \rightarrow y_2 \\ \rightarrow y_3 \\ \rightarrow y_4 \end{matrix}$$

где x_0 — фиктивный фактор, равный $+1$ и вводимый в матрицу для вычисления коэффициента b_0 (на проведение эксперимента он не влияет).

Таким образом, план (3.36) содержит сочетания всех уровней обонх факторов. Следовательно, уже можно судить о принципе формирования плана эксперимента, т. е. план первого порядка содержит все сочетания выбранных уровней факторов.

Пусть $k = 3$. Необходимо описать зависимость $y = f(x_1, x_2, x_3)$. Модель имеет вид $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$. Матрица плана, построенная по сформулированному принципу, имеет вид, показанный в табл. 3.5. Графически план трехфакторного эксперимента представляет собой куб. Для его построения и построения более сложных планов существует рекуррентная процедура. План эксперимента для двух факторов (3.36) считается ядром и пишется дважды для $x_3 = -1$ (4 опыта) и для $x_3 = +1$ (4 опыта). Для четырехфакторного плана ядром является матрица трехфакторного плана, которая также повторяется дважды для $x_1 = -1$ (8 опытов) и $x_1 = +1$ (8 опытов) и т. д.

Таблица 3.5

№ опыта	x_1	x_2	x_3	y
1	—	—	—	y_1
2	—	+	—	y_2
3	+	—	—	y_3
4	+	+	—	y_4
5	—	—	+	y_5
6	—	+	+	y_6
7	+	—	+	y_7
8	+	+	+	y_8

Для простоты в теории планирования эксперимента вводят обозначения планов, которые позволяют легко вычислить объем эксперимента. Так, для определенной категории планов вводят обозначение

$$N = m^k, \quad (3.37)$$

где m — число уровней, на которых варьируются факторы.

Поскольку в планах первого порядка факторы варьируются на двух уровнях, т. е. $m = 2$, можно переписать (3.37) следующим образом:

$$N = 2^k. \quad (3.38)$$

Очевидно, что с ростом k число опытов в эксперименте растет. Так, для $k = 2$ $N = 4$; для $k = 5$ $N = 32$; для $k = 10$ $N = 1024$ и т. д.

Планы, объем которых определяется (3.38) и которые содержат все сочетания уровней факторов, называют *планами полных факторных экспериментов* (ПФЭ).

Планы ПФЭ обладают тремя важными свойствами:

- 1) $\sum_{i=1}^N x_{ij} = 0, \quad j = \overline{1, k}$ (свойство симметричности);
- 2) $\sum_{i=1}^N x_{ij}^2 = N, \quad j = \overline{1, k}$ (свойство нормировки);
- 3) $\sum_{\substack{i=1 \\ j>1}}^N x_{ij} x_{il} = 0, \quad j, l = \overline{1, k}$ (свойство ортогональности).

Использование этих свойств позволяет существенно упростить вычислительную процедуру построения многофакторной регрессионной модели.

Планы, обладающие третьим свойством, называют *ортогональными*.

Заметное увеличение объема эксперимента с ростом k ограничивает использование ПФЭ, так как проведение сотен опытов для построения одной модели экономически не выгодно. Поэтому возникает вопрос о сокращении числа опытов в планах ПФЭ.

Представление матрицы независимых переменных в кодированном виде и включение в нее эффектов взаимодействия существенно сокращает объем ПФЭ. В отличие от линейных эффектов x_j эффекты взаимодействия $x_j x_l$ или $x_j x_l x_s$, характеризующие совместное влияние двух, трех и т. д. факторов, на y могут иметь противоположное действие, например: при увеличении x_j и x_l значение y может возрастать, а при увеличении $x_j x_l$ уменьшаться. Поэтому целесообразно в модели первого порядка включить и слагаемые взаимодействия. Так, при $k = 3$, модель с эффектами взаимодействия имеет вид $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3$, где b_{jl} , b_{jls} — коэффициенты регрессии при слагаемых парного и тройного взаимодействия.

Матрица плана 2^3 , включающая в себя эффекты взаимодействия, дана в табл. 3.6.

Таблица 3.6

№ опыта	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$	y
1	—	—	—	+	+	+	—	y_1
2	—	+	—	—	+	—	+	y_2
3	+	—	—	—	—	+	—	y_3
4	+	+	—	+	—	—	—	y_4
5	—	—	+	+	—	—	+	y_5
6	—	+	+	—	—	+	—	y_6
7	+	—	+	—	+	—	—	y_7
8	+	+	+	+	+	+	+	y_8

Для сокращения числа опытов вдвое поступают следующим образом: максимальное взаимодействие $x_1 x_2 x_3$ (в данном случае тройное) приравнивают к $+1$ или -1 . Выбранные опыты (для $x_1 x_2 x_3 = +1$ это опыты 2, 3, 5, 8) составляют одну половину плана (полуреплику), остальные — другую. Обе половины равноценны и по ним можно построить такую же модель, как и по плану ПФЭ.

Сокращенные планы обозначают $N = 2^{k-l}$, где l — степень сокращения. Они удовлетворяют всем трем свойствам планов первого порядка. При $l = 1$ — это полуреплика, т. е. план сокращается вдвое, при $l = 2$ — это четвертьреплика, т. е. план сокращается вчетверо, и т. д. Сокращенные планы называют планами дробных факторных экспериментов (ДФЭ).

Планы эксперимента второго порядка. Рассмотрим сначала планы с центральной точкой. Напомним, что планы второго порядка нужны для построения моделей второго порядка. Известно, что для построения прямой достаточно двух точек аргумента, а для построения кривой второго порядка необходимо как минимум три точки, т. е. в планах второго порядка факторы должны варьироваться как минимум на трех уровнях. Подобным третьим уровнем может быть x_{j0} или после кодирования нулевое значение фактора (см. рис. 3.7, а).

Пусть $k = 1$. Следует описать зависимость $y = f(x_1)$ с помощью модели $\hat{y} = b_0 + b_1x_1 + b_{11}x_1^2$. Матрица плана в этом случае имеет вид

$$X = \begin{bmatrix} x_{1 \min} \\ x_{10} \\ x_{1 \max} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix} \begin{matrix} \rightarrow y_1 \\ \rightarrow y_2 \\ \rightarrow y_3 \end{matrix}$$

Следует помнить, что нуль в матрице — это кодированное значение x_{10} .

Пусть $k = 2$. Нужно описать зависимость $y = f(x_1, x_2)$ с помощью модели $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2$.

Матрица плана содержит все комбинации трех уровней двух факторов:

$$X = \begin{bmatrix} -1 & -1 \\ -1 & +1 \\ +1 & -1 \\ +1 & +1 \\ -1 & 0 \\ +1 & 0 \\ 0 & +1 \\ 0 & -1 \\ 0 & 0 \end{bmatrix} \begin{matrix} \rightarrow y_1 \\ \rightarrow y_2 \\ \rightarrow y_3 \\ \rightarrow y_4 \\ \rightarrow y_5 \\ \rightarrow y_6 \\ \rightarrow y_7 \\ \rightarrow y_8 \\ \rightarrow y_9 \end{matrix} \quad (3.39)$$

Графически этот план изображен на рис. 3.8. Обозначение планов с центральной точкой производят в соответствии с (3.37). Так как $m = 3$, то

$$N = 3^k. \quad (3.40)$$

Очевидно, что с увеличением k число опытов быстро увеличивается. Так, для распространенного на практике случая $k = 5$ $N = 3^5 = 243$. Такое число опытов при производстве изделий электронной техники не допустимо. Поэтому при

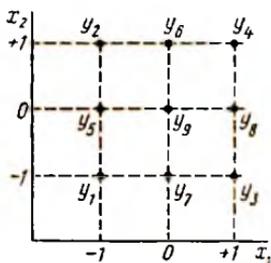


Рис. 3.8

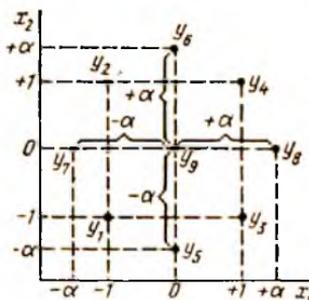


Рис. 3.9

$k \geq 4$ от планов с центральной точкой отказываются. Необходимо использовать планы с меньшим числом опытов, позволяющие построить модели второго порядка.

Рассмотрим композиционные планы Бокса—Вильсона, у которых $N = n_c + 2k + n_0$, где n_c — число опытов в ПФЭ или ДФЭ (2^k или 2^{k-1}); $2k$ — число звездных точек; n_0 — число опытов в центре (рис. 3.9).

Звездные точки лежат на продолжении осей, проходящих через середину сторон факторного пространства на расстоянии звездного плеча α . Для ротатабельных планов $\alpha = \sqrt{k}$, а для других его находят из биквадратного уравнения $\alpha^4 + 2^k \alpha^2 - 2^{k-1} (k + 0,5n_0) = 0$. Матрица плана Бокса—Вильсона для $k = 2$ приведена в табл. 3.7. Планы названы композиционными, так как содержат планы ПФЭ и ДФЭ как составную часть. В случае, когда модель первого порядка не адекватна, данные эксперимента используют для построения модели второго порядка при соответствующем дополнении плана звездными и центральными точками и проведении в них эксперимента.

Планы ПФЭ первого порядка для каждого значения k имеются в единственном числе, планов же второго порядка для каждого k более десятка. Следовательно, нужно выбрать

наилучший план, позволяющий построить более точную модель. В выражении (3.16), по которому вычисляют коэффициенты регрессионной модели, матрица факторов X входит трижды. Таким образом, от выбора наилучшего плана существенно зависит точность вычисляемых коэффициентов или их оценок.

Т а б л и ц а 3.7

№ опыта	x_1	x_2	y
1	—	—	y_1
2	—	+	y_2
3	+	—	y_3
4	+	+	y_4
5	0	$-\alpha$	y_5
6	0	$+\alpha$	y_6
7	$-\alpha$	0	y_7
8	$+\alpha$	0	y_8
9	0	0	y_9

Если рассматривать пространство коэффициентов модели B , то в идеальном случае коэффициенты b_j , $j = \overline{1, k}$ будут представлять в k -мерном пространстве точку. Однако в реальных задачах коэффициенты найдены с дисперсией $\sigma^2(b_j)$, поэтому они будут представлять эллипсоид рассеяния оценок коэффициентов. Планы, имеющие наименьший эллипсоид рассеяния оценок, называют *D-оптимальными*. Для таких планов определитель информационной матрицы $\det(X^T X)$ максимален. Следовательно, *D-оптимальные* планы дают наилучшую оценку коэффициентов модели. Следует отметить, что все ортогональные планы *D-оптимальны*.

Кроме того, существуют планы второго порядка, которые обладают свойством ротатабельности. План называют *ротатабельным*, если дисперсия оценки отклика $\sigma^2(\hat{y})$ в точке x зависит только от расстояния этой точки от центра x_0 . Все ортогональные планы являются ротатабельными.

§ 3.5. АЛГОРИТМ ПОСТРОЕНИЯ РЕГРЕССИОННОЙ МОДЕЛИ ПРИ ОРТОГОНАЛЬНЫХ ПЛАНАХ

Как упоминалось, при использовании ортогональных планов первого порядка построение моделей упрощается. Пусть $k = 2$. Тогда система уравнений в случае применения метода наименьших квадратов (МНК) имеет вид

$$\begin{aligned}
 b_0 \sum x_{0i}^2 + b_1 \sum x_{0i} x_{1i} + b_2 \sum x_{0i} x_{2i} &= \sum x_{0i} y_i; \\
 b_0 \sum x_{0i} x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} &= \sum x_{1i} y_i; \\
 b_0 \sum x_{0i} x_{2i} + b_1 \sum x_{1i} x_{2i} + b_2 \sum x_{2i}^2 &= \sum x_{2i} y_i,
 \end{aligned} \quad (3.41)$$

где суммирование производится по индексу i от 1 до N .

Поскольку план удовлетворяет трем свойствам (симметричности, нормировки и ортогональности), система (3.41) упрощается:

$$b_0 N = \sum x_{0i} y_i;$$

$$b_1 N = \sum x_{1i} y_i;$$

$$b_2 N = \sum x_{2i} y_i.$$

Таким образом,

$$b_0 = \sum x_{0i} y_i / N; \quad b_1 = \sum x_{1i} y_i / N;$$

$$b_2 = \sum x_{2i} y_i / N$$

В рассматриваемом случае информационная матрица $\mathbf{X}^T \mathbf{X}$ содержит только диагональные элементы:

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} N & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & N \end{bmatrix}.$$

Обратная матрица имеет вид

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{N} & 0 & 0 \\ 0 & \frac{1}{N} & 0 \\ 0 & 0 & \frac{1}{N} \end{bmatrix}.$$

Основной алгоритм построения регрессионной модели при ортогональном планировании заключается в следующем. Пусть необходимо построить двухфакторную модель ($k = 2$) первого порядка: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$ по плану 2^2 . В табл. 3.8 приведены матрица плана, а также результаты эксперимента и анализа. Допустим, что в процессе эксперимента были проведены параллельные опыты y_{i1} и y_{i2} . По-

строение модели начинают с вычисления среднего значения отклика:

$$\bar{y}_i = \frac{1}{m} \sum_l y_{il},$$

где m — число параллельных опытов (в таблице заполняется шестая графа).

Т а б л и ц а 3.8

№ опыта	x_1	x_2	y_{i1}	y_{i2}	\bar{y}_i	σ_i^2	\hat{y}	$(\bar{y}_i - \hat{y})^2$
1	-	-	y_{11}	y_{12}	\bar{y}_1	σ_1^2	\hat{y}_1	$(\bar{y}_1 - \hat{y}_1)^2$
2	-	+	y_{21}	y_{22}	\bar{y}_2	σ_2^2	\hat{y}_2	$(\bar{y}_2 - \hat{y}_2)^2$
3	+	-	y_{31}	y_{32}	\bar{y}_3	σ_3^2	\hat{y}_3	$(\bar{y}_3 - \hat{y}_3)^2$
4	+	+	y_{41}	y_{42}	\bar{y}_4	σ_4^2	\hat{y}_4	$(\bar{y}_4 - \hat{y}_4)^2$

В каждой точке эксперимента дисперсия $\sigma_i^2 = \frac{1}{m-1} \sum_{l=1}^m (y_{il} - \bar{y}_i)^2$ (в таблице заполняется седьмая графа).

Однородность дисперсий во всех точках факторного пространства (воспроизводимость эксперимента) проверяют по критерию Кохрена:

$$G = \frac{\sigma_{\max}^2}{\sum_{i=1}^N \sigma_i^2}.$$

Полученное значение критерия сравнивают с табличным $G_\alpha(f_1, f_2)$, где α — уровень значимости; $f_1 = m - 1$, $f_2 = N - 1$ — число степеней свободы числителя и знаменателя. При $G \leq G_\alpha(f_1, f_2)$ полученные данные однородны и эксперимент воспроизводим, в противном случае, когда $G > G_\alpha \times (f_1, f_2)$, эксперимент невоспроизводим и модель по таким данным строить нельзя. Данные эксперимента необходимо проверить заново или сместить факторное пространство эксперимента.

Средняя дисперсия воспроизводимости (ошибка эксперимента)

$$\sigma_{\text{ош}}^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2.$$

Коэффициенты регрессии модели

$$b_j = \frac{1}{N} \sum_{i=1}^N x_{ji} \bar{y}_i; \quad b_0 = \frac{1}{N} \sum_{i=1}^N \bar{y}_i; \quad b_1 = \frac{1}{N} \sum_{i=1}^N x_{1i} \bar{y}_i;$$

$$b_2 = \frac{1}{N} \sum_{i=1}^N x_{2i} \bar{y}_i.$$

В результате получаем уравнение регрессии первого порядка:

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

Проверяем коэффициенты регрессии на значимость, т. е. оцениваем, превышают ли значения найденных коэффициентов некоторый критический уровень, который характеризует границу существенного или несущественного влияния фактора на функцию отклика. Для этого вычисляем дисперсию коэффициентов регрессии

$$\sigma^2(b_j) = \sigma_{\text{ош}}^2 / (Nm).$$

Дисперсии всех коэффициентов равны, так как они зависят только от ошибки опыта и числа опытов.

Граничное значение коэффициента регрессии или его доверительную границу определяем из выражения

$$|b_j| = \sqrt{\sigma^2(b_j)} t_\alpha(f),$$

где критерий Стьюдента $t_\alpha(f)$ выберем из таблицы для уровня значимости α и числа степеней свободы f .

Если $|b_j| \geq b_{гр}$, то коэффициент b_j значим и фактор x существенно влияет на y ; если $|b_j| < b_{гр}$, то фактор не влияет на y и его можно исключить из модели.

Проверим гипотезу об адекватности полученной модели по табл. 3.4. Для этого оценим дисперсию адекватности:

$$\sigma_{ад}^2 = \frac{1}{N-d} \sum_{i=1}^N (\bar{y}_i - \widehat{y}_i)^2,$$

где d — число вычисляемых коэффициентов в модели.

Критерий Фишера $F = \sigma_{ад}^2 / \sigma_{\text{ош}}^2$.

При $F \leq F_\alpha(f_1, f_2)$ не отвергается гипотеза об адекватности модели и ее можно использовать для оптимизации описанного ТП. В противном случае, когда $F > F_\alpha(f_1, f_2)$.

модель не адекватна и ею нельзя пользоваться. Необходимо отметить, что отсутствие параллельных опытов не позволяет вычислить дисперсии откликов в каждой точке эксперимента, а следовательно, невозможно проверить эксперимент на воспроизводимость, коэффициенты модели на значимость и модель на адекватность, а также определить ошибку эксперимента. Кроме того, в случае насыщенных планов, когда число опытов равно числу вычисляемых коэффициентов $N = d$, невозможно рассчитать дисперсию адекватности и соответственно проверить модель на адекватность.

При неадекватности модели целесообразно изменить порядок модели, увеличить интервал варьирования Δx факторов, сдвинуть факторное пространство или сменить факторы.

ГЛАВА 4. ОПТИМИЗАЦИЯ ТЕХНОЛОГИЧЕСКИХ ОПЕРАЦИЙ

§ 4.1. ОПТИМИЗАЦИЯ МЕТОДОМ БОКСА — ВИЛЬСОНА

Поиск оптимальной области. Как отмечалось, оптимизация ТП состоит из двух этапов: 1) описания ТП (операции) с помощью математической модели; 2) движения к области оптимальных значений целевой функции.

Первый этап подробно проанализирован в гл. 3. Рассмотрим способы и алгоритмы достижения области экстремальных значений.

Методы направленного поиска оптимальной области существовали и до разработки научных положений многофакторных экспериментов. Наиболее старым является метод Гаусса—Зайделя (однофакторный эксперимент), при котором все факторы, кроме одного, определяют по (1.4). Принципиальным недостатком этого метода является необходимость проведения большого числа опытов (штриховые линии на рис. 4.1).

Метод Бокса—Вильсона лишен этого недостатка. Он является примером многофакторного эксперимента, когда варьируются все факторы одновременно в соответствии с (1.5). Этот метод получил название метода крутого восхождения или градиентного метода, когда движение к оптимальной области осуществляется по кратчайшему пути — по градиенту. При этом движение в факторном пространстве от

точки к точке сопровождается в соответствии с (1.5) одновременным изменением значений всех факторов.

Сравнение метода Гаусса—Зайделя с методом Бокса—Вильсона показывает (сплошные линии на рис. 4.1), что при втором методе путь к области оптимальных значений целевой функции значительно сокращается. С возрастанием числа факторов эффект от применения метода крутого восхождения возрастает. В общем случае алгоритм метода Бокса—Вильсона следующий:

1) построение многофакторного эксперимента в окрестности некоторой выбранной точки;

2) описание поверхности отклика с помощью регрессионной модели;

3) вычисление оценки градиента в этой точке по результатам эксперимента и построенной модели;

4) крутое восхождение в направлении оценки градиента в соответствии с вычисленным шагом по каждой координате пространства;

5) нахождение оценки экстремума целевой функции в этом направлении.

Поскольку два первых пункта ранее рассматривались, проанализируем остальные пункты.

Пусть $\mathbf{x}^0 = (x_1^0, x_2^0, \dots, x_k^0)$ — начальная точка поиска экстремума (рис. 4.1). Вектор-градиент функции $f(x_1, x_2, x_3, \dots, x_k)$ в этой точке

$$\text{grad } f(x_1^0, x_2^0, \dots, x_k^0) = \left(\frac{\partial f}{\partial x_1^0}, \frac{\partial f}{\partial x_2^0}, \dots, \frac{\partial f}{\partial x_k^0} \right); \quad (4.1)$$

$$\frac{\partial f}{\partial x_j^0} = \frac{\partial f(x_1^0, x_2^0, \dots, x_k^0)}{\partial x_j}, \quad j = \overline{1, k}, \quad (4.2)$$

или

$$\nabla f = \frac{\partial f}{\partial x_1^0} \mathbf{i}_1 + \frac{\partial f}{\partial x_2^0} \mathbf{i}_2 + \dots + \frac{\partial f}{\partial x_k^0} \mathbf{i}_k, \quad (4.3)$$

где ∇f — обозначение градиента; $\partial f / \partial x_j^0$ — частная производная функция по j -му фактору; \mathbf{i}_j — единичные векторы в направлении координатных осей.

Если функция в точке \mathbf{x}^0 не является стационарной и все частные производные $\{\partial f / \partial x_j^0\}$ не обращаются в нуль, то

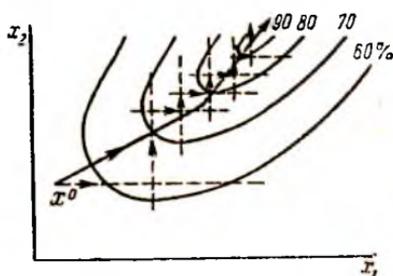


Рис. 4.1

направление вектор-градиента в этой точке перпендикулярно поверхности равного уровня $y = \text{const}$ и будет направлением наибоыстрейшего возрастания функции. Заметим, что направление антиградиента $[-\text{grad } f(\mathbf{x}^0)]$ является направлением наибоыстрейшего убывания функции.

Если функция в точке \mathbf{x}^0 не является стационарной, то при совершении движения в направлении градиента координаты новой точки определяют по формуле

$$x_j = x_j^0 + \alpha \frac{\partial f}{\partial x_j^0}, \quad j = \overline{1, k}, \quad (4.4)$$

где $\alpha \geq 0$ — параметр шага.

В векторной форме записи это выражение примет вид

$$\mathbf{x}^{N+1} = \mathbf{x}^N + \alpha_N \text{grad } f(\mathbf{x}), \quad (4.5)$$

где N — номер опыта.

При решении практических задач оптимизации в случае большого числа факторов поиск градиентного направления не является достаточно простой процедурой. Поэтому целесообразно определять градиентное направление через коэффициенты регрессии модели, которая к этому моменту, как правило, построена. С этой целью проанализируем коэффициент модели Бокса—Вильсона.

Пусть в окрестности точки \mathbf{x}^0 поверхность отклика непрерывна и функция $f(\mathbf{x})$ дифференцируема. Тогда ее можно разложить в ряд Тейлора ($k = 2$):

$$\begin{aligned} y = f(x_1, x_2) = & f(x_1^0, x_2^0) + \frac{\partial f}{\partial x_1^0} (x_1 - x_1^0) + \\ & + \frac{\partial f}{\partial x_2^0} (x_2 - x_2^0) + \frac{1}{2!} \frac{\partial^2 f}{\partial x_1^0 \partial x_2^0} (x_1 - x_1^0) (x_2 - x_2^0) + \\ & + \frac{1}{2!} \frac{\partial^2 f}{\partial x_1^{02}} (x_1 - x_1^0)^2 + \frac{1}{2!} \frac{\partial^2 f}{\partial x_2^{02}} (x_2 - x_2^0)^2. \end{aligned}$$

Двухфакторная модель Бокса—Вильсона имеет вид

$$\begin{aligned} \widehat{y} = & b_0 + b_1 (x_1 - x_1^0) + b_2 (x_2 - x_2^0) + b_{12} (x_1 - x_1^0) (x_2 - x_2^0) + \\ & + b_{11} (x_1 - x_1^0)^2 + b_{22} (x_2 - x_2^0)^2. \end{aligned}$$

Так как оба выражения описывают одну и ту же поверхность отклика в окрестностях точки \mathbf{x} , то

$$f(x_1^0, x_2^0) = b_0; \quad \frac{1}{2} \frac{\partial^2 f}{\partial x_1^0 \partial x_2^0} = b_{12};$$

$$\frac{\partial f}{\partial x_1^0} = b_1; \quad \frac{1}{2} \frac{\partial^2 f}{\partial x_1^0{}^2} = b_{11};$$

$$\frac{\partial f}{\partial x_2^0} = b_2; \quad \frac{1}{2} \frac{\partial^2 f}{\partial x_2^0{}^2} = b_{22}.$$
(4.6)

Таким образом, коэффициенты модели Бокса—Вильсона имеют смысл частных производных функции отклика. Если модель линейная: $\hat{y} = b_0 + b_1 x_1 + \dots + b_n x_n$, то коэффициенты b_1, \dots, b_n являются координатами вектора градиента:

$$\text{grad } f(\mathbf{x}^0) = (b_1, b_2, \dots, b_n).$$

Поэтому задача оценки градиента сводится к нахождению оценок неизвестных коэффициентов b_1, \dots, b_n .

Следующий этап — это определение шага движения по каждой координате функции $\hat{y} = f(\mathbf{x})$. Градиент функции \hat{y} можно нормировать, разделив его на норму $\|\nabla \hat{y}\|$. В результате получим единичный вектор

$$\frac{\nabla \hat{y}}{\|\nabla \hat{y}\|} = \frac{\frac{\partial f}{\partial x_1} \mathbf{i}_1 + \frac{\partial f}{\partial x_2} \mathbf{i}_2 + \dots}{\left[\sum_j \left(\frac{\partial f}{\partial x_j} \right)^2 \right]^{1/2}},$$

составляющие которого характеризуют относительный размер шага по каждому из координатных направлений x_j для наискорейшего подъема. Единичный вектор градиента для регрессионной модели первого порядка $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$ равен

$$\frac{\nabla \hat{y}}{\|\nabla \hat{y}\|} = \frac{b_1 \mathbf{i}_1 + b_2 \mathbf{i}_2}{\sqrt{b_1^2 + b_2^2}}.$$

Таким образом, координаты l -й точки наблюдения (l -й шаг) в направлении градиента находят по формуле

$$x_j^l = x_j^{l-1} + \lambda_j.$$

Так как градиент всегда проходит через центр плана при $x_0 = 0$, то

$$x_j^i = l \lambda_j,$$

где λ — выбранный шаг. Выбор шага осуществляют аналогично выбору интервала варьирования. В первом приближении

$$\lambda_j = b_j \Delta_j, \quad (4.7)$$

где Δ_j — интервал варьирования j -го фактора ($\Delta_j = 0,5 \times (x_{j\max} - x_{j\min})$).

В некоторых случаях шаги для отдельных факторов могут соотноситься друг с другом так же, как модули $b_j \Delta_j$: $\lambda_1 : \lambda_2 : \dots : \lambda_h = |b_1 \Delta_1| : |b_2 \Delta_2| : |b_3 \Delta_3| : \dots : |b_h \Delta_h|$.

При экспериментальной оптимизации ТП из выражения (4.7) можно получить только начальное значение шага движения по j -й координате. В дальнейшем экспериментатор по мере движения к оптимуму корректирует шаг λ_j с учетом конкретного изменения функции отклика. В табл. 4.1 приведены исходные данные для крутого восхождения, где l_j — величина, округляющая и корректирующая шаг движения.

Таблица 4.1

Параметр	Факторы		
	x_1	x_2	x_h
Основной уровень x_{0j}	x_{01}	x_{02}	x_{0h}
Интервал варьирования Δ_j	Δ_1	Δ_2	Δ_h
Коэффициент уравнения регрессии b_j	b_1	b_2	b_h
Длина шага $b_j \Delta_j$	$b_1 \Delta_1$	$b_2 \Delta_2$	$b_h \Delta_h$
Изменение длины шага $l_j b_j \Delta_j$	$l_1 b_1 \Delta_1$	$l_2 b_2 \Delta_2$	$l_h b_h \Delta_h$

Способы движения к оптимальной области. После того как установлен шаг движения к области оптимальных значений целевой функции по каждой координате, необходимо выбрать способ движения. Дело в том, что модель, с помощью которой определяется градиентное направление и шаг движения, справедлива только в том факторном пространстве, где проводился эксперимент, и на основе данных которого была построена модель. Движение же следует осуществлять за пределы описанного пространства, где модель мож-

но считать адекватной с определенным допущением. Чем дальше удаляются от исследованной области, тем внимательнее следует проверять правильность движения к оптимуму, проводя дополнительные эксперименты. Нужно уточнить, каким образом проводить эксперименты, чтобы эффективность, достигаемая на первом этапе оптимизации, возросла.

На практике получили распространение два способа движения к оптимальной области. В первом случае, после продвижения на один или несколько шагов по направлению

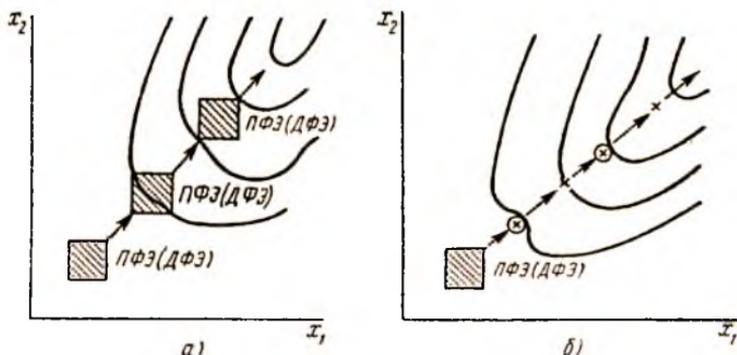


Рис. 4.2

к оптимуму, проводят опыты по плану ПФЭ или ДФЭ. По результатам опыта строят модель, которая описывает новое факторное пространство. Снова определяют градиентное направление и шаг движения в этом направлении. Затем осуществляют движение на один или несколько шагов в новом направлении, и алгоритм действий экспериментатора повторяется. Так шаг за шагом достигаются все более высокие значения отклика. Подобный способ движения иллюстрируется рис. 4.2, а.

Об эффективности движения по градиенту можно судить по значению параметра оптимизации. Такое движение эффективно, если реализация опытов эксперимента приводит к улучшению параметра оптимизации по сравнению с наилучшим результатом в матрице откликов, по которой строилась модель. При эффективном крутом восхождении возможны два конечных результата: 1) область оптимума достигнута; 2) область оптимума не достигнута. В первом случае в процессе восхождения можно обнаружить, что один или несколько коэффициентов модели поверхности отклика уменьшились. Это может произойти в результате: а) нахождения изменяемого фактора вблизи оптимального

значения; б) выбора слишком малого интервала варьирования фактора; в) незначимости варьируемого фактора.

В первом случае следует фиксировать факторы, по которым достигнут оптимум, и продолжать крутое восхождение по оставшимся факторам, во втором — выбрать подходящий интервал варьирования, в третьем — проверить фактор на значимость, используя критерий Стьюдента.

Как только достигнута область экстремума, можно определить положение оптимальной точки, построив для функции отклика f модель второго порядка и исследовав затем эту модель на экстремум. При $k = 2$ функция f в экстремальной точке должна удовлетворять необходимым (определяющим стационарную точку $\partial f/\partial x_1 = 0$; $\partial f/\partial x_2 = 0$) и достаточным (в случае максимума $B^2 - AC < 0$; $A + C < 0$, в случае минимума $B^2 - AC < 0$; $A + C > 0$, где $A = \partial^2 f/\partial x_1^2$; $B = \partial^2 f/\partial x_1 \partial x_2$; $C = \partial^2 f/\partial x_2^2$) условиям.

Если $B^2 - AC > 0$, то появляется седловая точка; если $B^2 - AC = 0$, то природа стационарной точки неопределенна (нужно исследовать высшие производные).

Для функций отклика, включающих более двух независимых переменных, при проверке достаточных условий удобно записать матрицу Гессе или гессиан в виде

$$\mathbf{h} = \begin{bmatrix} h_{11} & \dots & h_{1k} \\ \vdots & & \vdots \\ h_{k1} & \dots & h_{kk} \end{bmatrix},$$

где $h_{ji} = \frac{1}{2} \frac{\partial^2 f}{\partial x_j \partial x_i}$, и подсчитать h_{ji} в стационарной точке пространства наблюдений.

Если каждый из главных миноров h в стационарной точке отрицательно определен, то отклик имеет максимум; если положительно определен, то отклик имеет минимум.

Второй способ движения к оптимальной области отличается от первого уменьшением числа опытов; он состоит а следующем. После проведения ПФЭ или ДФЭ, построения модели, определения градиентного направления и продвижения на один или несколько шагов выполняют только один опыт (вместо множества опытов, проводимых по плану). Если нет существенного отличия экспериментального значения функции отклика от вычисленного по модели, то можно вновь повторить один или несколько шагов и выполнить единственный опыт в многофакторном пространстве и т. д. (рис. 4.2, б). При этом одиночные опыты можно выполнять не на каждом шаге движения к оптимуму, а через

один или несколько шагов. Вместо пропущенных шагов производят «мысленные» опыты. Проведение «мысленных» опытов подразумевает вычисление функции отклика на очередном шаге на основе модели процесса или операции. После этого осуществляют дальнейшее движение к оптимуму. Подобный подход применим при оптимизации ТП производства изделий электронной техники, так как поверхности отклика технологических операций сравнительно гладкие и промежуточные шаги без ущерба для оптимизации можно пропускать. На рис. 4.2, б «мысленные» опыты обведены кружком.

Рассмотрим в качестве примера операцию термообработки активного пленочного СВЧ-элемента на нелинейном диэлектрике при его разработке.

В качестве целевых функций взяты три выходных параметра элемента: $y_1 \equiv C_0$ — начальная емкость элемента; $y_2 \equiv K$ — управляемость $[K = \frac{C}{C(U)}$, U — напряжение смещения в области насыщения реверсивной характеристики $C(U)$]; $y_3 \equiv \lg \delta$ — диэлектрические потери в элементе. В ходе эксперимента варьировались: x_1, x_3 — температура отжига, °С; x_2, x_4 — время отжига, мин (соответственно для первого и второго цикла термообработки); x_5 — размер исходного зерна титаната стронция (качественный параметр), мкм. Погрешность фиксирования факторов $x_1 - x_4$ не превышала 1–2 %.

Для получения математического описания нового ТП в заданной области изменения технологических факторов (табл. 4.2) был поставлен дробный факторный эксперимент типа 2^{5-1} с генерирующим соотношением $x_5 = x_1 x_2 x_3 x_4$. Условия эксперимента были рандомизированы. В процессе анализа результатов эксперимента и построения модели были найдены оценки для всех 15 линейных эффектов и взаимодействий первого порядка. Оценка значимости коэффициентов проводилась по t -критерию Стьюдента при уровне значимости 0,05.

Таблица 4.2

Уровни	x_1 , °С	x_2 , мин	x_3 , °С	x_4 , мин	x_5 , мкм
Основной x_{0j}	1368	8	1027	20	—
Интервал варьирования Δx_j	50	6	127	10	—
Верхний (+1)	1418	14	1154	30	2
Нижний (–1)	1318	2	900	10	0,6

По результатам эксперимента получены уравнения регрессии, содержащие только значимые коэффициенты:

$$y_1 = 2 - 0,45x_3 - 0,16x_4 - 0,7x_1x_2; \quad (4.8)$$

$$y_2 = 2,49 - 0,44x_3 - 0,19x_1x_3 - 0,13x_4; \quad (4.9)$$

$$y_3 = 25,3 \cdot 10^{-3} - 3,63 \cdot 10^{-3} x_1 - 3,25 \cdot 10^{-3} x_2 - 5,56 \cdot 10^{-3} x_3. \quad (4.10)$$

Проверка с помощью критерия Фишера при 5%-ном уровне значимости показала, что модель для y_3 адекватно описывает экспериментальные данные, а модели для y_1 и y_2 могут быть приняты с небольшой неадекватностью для дальнейших расчетов. Длительность и трудоемкость опытов, а также небольшая неадекватность вынудили отказаться от получения адекватных моделей для y_1 и y_2 , так как рациональнее путем нескольких опытов проверить возможность движения по градиенту y_1 и y_2 . Окончательный выбор оптимального режима изготовления элемента может быть сделан на основании совместного анализа поверхностей отклика y_1 , y_2 , y_3 . Необходимо отметить возможность введения на данном этапе решения задачи более общей целевой функции (обобщенного отклика) $y = f_{кр}$, где

$f_{кр} = \frac{(K-1)f}{2(K+1) \lg \delta}$ — критическая частота элемента; f — частота измерений (рабочая частота).

На основании выражения для $f_{кр}$ и моделей (4.8)—(4.10) можно представить y как функцию факторов x_j . Для этого воспользуемся локальной аппроксимацией y полиномом первого порядка и вычислим частные производные dy/dx_1 , dy/dx_2 , dy/dx_3 , dy/dx_4 в двух точках факторного пространства. Указанные частные производные функции отклика y можно рассматривать как точечные оценки коэффициентов регрессии при соответствующих переменных (табл. 4.3).

Таблица 4.3

Координаты точек факторного пространства				Коэффициенты регрессии			
x_1	x_2	x_3	x_4	b_1	b_2	b_3	b_4
+1	+1	+1	+1	+4,05	+0,7	-1,45	-2,38
-1	-1	-1	-1	+1,66	+1,13	0,85	-0,44

В рассматриваемых интервалах варьирования факторов поверхность отклика y обладает некоторой кривизной относительно x_1 , x_2 , x_4 и имеет максимум относительно x_3 (параметр b_3 изменил знак). Дальнейшее движение по градиенту y следует проводить в направлении увеличения x_1 , x_2 и уменьшения x_4 . Учитывая, что направления движения из нулевой точки к максимуму y_2 и минимуму y_3 противоположны относительно выделенных факторов, необходимо принять компромиссное решение. Было решено ограничиться движением в направлении градиента y_3 , так как минимизация y_3 , имеющая самостоятельное значение, в наибольшей степени определяет возможное увеличение y . Частная целевая функция y_2 анализировалась как ограничительная. Значения шагов крутого восхождения по поверхности отклика функции y_3 приведены в табл. 4.4. Шаги, вычисленные с помощью выражения (4.7), малы по сравнению с углом наклона поверхности отклика и предполагаемым ме-

стонахождением оптимума. Поэтому полученные шаги λ_j умножают на 100 (по решению технолога) и округляют.

Таблица 4.4

Параметры	x_1	x_2	x_3
b_j	$-3,63 \cdot 10^{-3}$	$-3,25 \cdot 10^{-3}$	$-5,56 \cdot 10^{-3}$
$\lambda_j = b_j \Delta x_j$	0,18	0,0195	0,0706
Скорректированный шаг $\lambda_j \cdot 100$	18	1,95	70,6
Шаг после округления	18	2,0	70,0

Реализация крутого восхождения иллюстрируется в табл. 4.5.

Таблица 4.5

№ опыта	x_1 , °C	x_2 , мин	x_3 , °C	x_4 , мин	x_5 , мкм	$\text{tg } \delta \equiv y_3$	$f_{\text{кр}} \equiv y$
17	1386	10	1097	30	0,6	—	—
18	1404	12	1167	30	0,6	—	—
19	1422	14	1237	30	0,6	0,019	16,7
20	1440	16	1237	30	0,6	0,015	25,5
21	1458	18	1237	30	0,6	0,012	32,0
22	1458	20	1237	30	0,6	0,009	34,5
23	1458	22	1237	30	0,6	—	—
24	1458	24	1237	30	0,6	0,006	44,5
25	1458	26	1237	30	0,6	—	—
26	1458	28	1237	30	0,6	0,009	29,6
27	1458	30	1237	30	0,6	—	—
28	1458	36	1237	30	0,6	0,013	24,6

Прекращение изменений отдельных факторов (x_1 , x_3) объясняется техническими и физическими условиями проведения эксперимента, а также ограничениями, накладываемыми изменениями y_2 . Фактор x_5 , как отмечалось, качественный — его можно получить только в виде двух значений (2 или 0,6 мкм). Было взято значение, при котором найдены лучшие значения откликов. Фактор x_4 был зафиксирован на верхнем уровне по результатам эксперимента и не варьировался из технических соображений. Прочерки в таблице означают пропуски соответствующих опытов, которые были заменены «мысленными». Лучший результат получен в опыте 24. По сравнению с лучшим результатом в опытах по плану ДФЭ $\text{tg } \delta \equiv y_3$ уменьшился в два раза, а по сравнению с имевшимися до постановки эксперимента результатами — примерно в пять раз. Одновременно частота $f_{\text{кр}} \equiv y$ возросла в 1,9 раза по сравнению с лучшим результатом в опытах по плану ДФЭ. Полученные результаты подтверждают правильность исходных предположений, принятых для статистических моделей ТП и расчетов движения по градиенту.

Многоцелевая (многокритериальная) оптимизация. Большинство практических задач оптимизации при производстве электронных приборов имеет не один, а два или несколько параметров оптимизации, т. е. несколько целевых функций. В подобных случаях осуществляют многоцелевую оптимизацию. Преимущество такой оптимизации заключается в том, что объект оптимизации достигает экстремальных (наилучших) значений сразу по нескольким показателям. При этом нет необходимости проводить дополнительные опыты, нужно только в каждом из них вместо одного параметра оптимизации измерить те, которые указаны в поставленной задаче (увеличивается только объем измерений). Табл. 4.6 соответствует случаю, когда в двухфакторном эксперименте три целевых функции y_1, y_2, y_3 . План эксперимента для описания процесса по всем трем откликам одинаков,

Таблица 4.6

№ опыта	x_1	x_2	y_1	y_2	y_3
1	—	—	y_{11}	y_{21}	y_{31}
2	—	+	y_{12}	y_{22}	y_{32}
3	+	—	y_{13}	y_{23}	y_{33}
4	+	+	y_{14}	y_{24}	y_{34}

модели идентичны:

$$\begin{aligned}
 \widehat{y}_1 &= b_0^{(1)} + b_1^{(1)} x_1 + b_2^{(1)} x_2 + b_{12}^{(1)} x_1 x_2; \\
 \widehat{y}_2 &= b_0^{(2)} + b_1^{(2)} x_1 + b_2^{(2)} x_2 + b_{12}^{(2)} x_1 x_2; \\
 \widehat{y}_3 &= b_0^{(3)} + b_1^{(3)} x_1 + b_2^{(3)} x_2 + b_{12}^{(3)} x_1 x_2.
 \end{aligned}
 \tag{4.11}$$

вычисление коэффициентов аналогично:

$$\begin{aligned}
 b_j^{(k)} &= \frac{1}{N} \sum_{i=1}^N x_{ji} y_{ki}; & b_{12}^{(k)} &= \frac{1}{N} \sum_{i=1}^N x_{1i} x_{2i} y_{ki}; \\
 b_j^{(k)} &= \frac{1}{N} \sum_{i=1}^N x_{ji} y_{ki}; & j &= 1, \bar{k}.
 \end{aligned}$$

При многоцелевой оптимизации возникают другие сложности. При получении экстремальных значений всех откли-

ков нужно осуществить многокритериальную максимизацию (минимизацию).

Допустим, в (4.11) все отклики необходимо максимизировать. При этом фактор x_1 вошел в модели с разными знаками — в \widehat{y}_1 и \widehat{y}_2 с плюсом, а в \widehat{y}_3 с минусом. Тогда при увеличении x_1 максимизируются только y_1 и y_2 , а y_3 уменьшается; следовательно, необходимо искать компромиссное решение. Возможен и такой случай: в (4.11) отклики y_1 и y_2 нужно максимизировать, а y_3 — минимизировать, в модели \widehat{y}_1 , \widehat{y}_2 и \widehat{y}_3 фактор x_1 вошел с плюсом. В результате при увеличении x_1 вместе с y_1 и y_2 возрастает y_3 , а при уменьшении x_1 вместе с y_3 уменьшаются y_1 и y_2 . Примером аналогичной ситуации являются модели (4.8)—(4.10). Очевидно, что экстремальных значений всех целевых функций не достичь, поэтому вновь требуется искать компромиссное решение.

Одним из методов, позволяющих найти компромиссное решение, является метод согласованного оптимума [6].

Пусть имеем две целевые функции $y_1 = f_1(\mathbf{x})$; $y_2 = f_2(\mathbf{x})$, находящиеся в начальных состояниях x_1^0 и x_2^0 и имеющие градиенты целевых функций $\xi_1(x_1^0) = \frac{\partial f_1}{\partial \mathbf{x}}(x_1^0)$; $\xi_2(x_2^0) = \frac{\partial f_2}{\partial \mathbf{x}}(x_2^0)$.

Тогда общая точка пересечения этих векторов соответствует решению следующей линейной системы уравнений:

$$\begin{aligned} \{\xi_1(x_1^0), \mathbf{x} - \mathbf{x}_1^0\} &= 0; \\ \{\xi_2(x_2^0), \mathbf{x} - \mathbf{x}_2^0\} &= 0. \end{aligned}$$

Если соблюдаются условия

$$\begin{aligned} dy_1 = \{\xi_1(\mathbf{x}), d\mathbf{x}\} &= 0; \\ dy_2 = \{\xi_2(\mathbf{x}), d\mathbf{x}\} &= 0, \end{aligned} \tag{4.12}$$

то полученную точку называют *согласованным оптимумом*.

При двухфакторном эксперименте уравнения (4.12) можно переписать в виде

$$\begin{aligned} dy_1 &= \xi_{11} dx_1 + \xi_{12} dx_2 = 0; \\ dy_2 &= \xi_{21} dx_1 + \xi_{22} dx_2 = 0, \end{aligned}$$

где $\xi_{lj} = \partial f_l / \partial x_j$; $l = 1, 2$; $j = \overline{1, k}$.

Уравнения имеют нетривиальное решение, если их определитель равен нулю:

$$\begin{vmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{vmatrix} = 0.$$

Раскрывая этот определитель, получим

$$\xi_{11} \xi_{22} = \xi_{21} \xi_{12}. \quad (4.13)$$

Таким образом, в точке согласованного оптимума прирост целевых функций по различным аргументам равнозначен. Если рассматривать (4.13) с учетом (4.6), то это равенство примет вид

$$b_1^{(1)} b_2^{(2)} = b_1^{(2)} b_2^{(1)},$$

где $b_j^{(i)}$ — коэффициенты регрессии моделей \widehat{y}_1 и \widehat{y}_2 .

Согласно условию (4.12), в точке согласованного оптимума векторы приростов обеих целевых функций параллельны (коллинеарны):

$$\xi_1 = \lambda \xi_2, \quad (4.14)$$

где λ — неизвестный коэффициент.

Следует отметить, что дифференциальные уравнения линии согласованного оптимума (4.12), создающие коллинеарность векторов-градиентов целевых функций (4.14), не обеспечивают максимальности функций f_1 и f_2 на всей линии (области) согласованного оптимума, а обеспечивают только стационарность этих функций:

$$f_1(x) = C_1(\lambda); \quad f_2(x) = C_2(\lambda),$$

где C_1 , C_2 — константы, которые можно трактовать как определенные уровни y_1 и y_2 .

Если при максимизации первой целевой функции можно ограничиться значением C_1 , то, определяя λ (с учетом C_1) и вычисляя $C_2(\lambda)$, можно получить условия согласованного оптимума. На рис. 4.3 показаны линии равного уровня двух поверхностей откликов, а также линия согласованного оптимума, которая проходит через точки абсолютных «личных» оптимумов y_1 и y_2 , удовлетворяющие условиям $\xi_1 = 0$, $\xi_2 = 0$.

В случае, когда целевых функций больше двух, согласованный оптимум искать трудно; при этом следует воспользоваться методом перебора.

В тех случаях, когда число откликов возрастает, задача оптимизации может быть решена с применением обобщенного параметра. Примером такой оптимизации является оптимизация пленочного СВЧ-элемента, где в качестве обобщенного параметра вместо (4.8)—(4.10) взята критическая частота $f_{кр}$. Это случай, когда функциональное соотношение позволяет выбрать обобщенный параметр. Однако не всегда удается найти подобный параметр, имеющий физический смысл. Часто его приходится выбирать в виде абстрактного математического выражения, на-

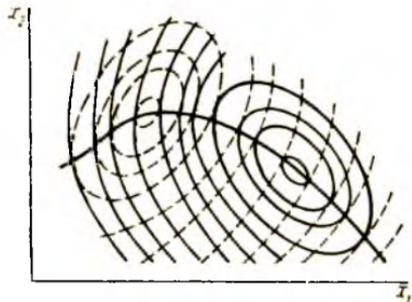


Рис. 4.3

пример в виде $Q = \sum_{l=1}^L \alpha_l y_l$,

где Q — обобщенный параметр; α_l — весовые коэффициенты (коэффициенты предпочтения) соответствующих целевых функций; $l = \overline{1, L}$.

Если все y_l необходимо максимизировать, то и Q следует максимизировать. Если же среди целевых функций есть такие, которые необходимо минимизировать, то для обобщенного отклика можно использовать выражение

$$Q = \frac{\sum_{l=1}^r \alpha_l y_l}{\sum_{l=r+1}^L \alpha_l y_l}$$

где $y_l, l = \overline{1, r}$ — целевые функции, которые нужно максимизировать; $y_l, l = \overline{r+1, L}$ — целевые функции, которые следует минимизировать.

В этом случае отклик Q необходимо максимизировать.

В тех случаях, когда трудно найти аналитическое решение при поиске компромиссного решения для всех откликов, выбирают одну целевую функцию, которую оптимизируют, остальные являются ограничениями в виде «не более» или «не менее».

Когда конечной целью исследований является оптимизация свойств изделий электронной техники, то построение модели ТП как промежуточный этап можно опустить, если есть алгоритм или метод нахождения оптимальных значений функции отклика. В некоторых случаях построение модели затруднено по техническим или экономическим причинам. Эти трудности привели к тому, что в методах оптимизации были разработаны такие алгоритмы, которые при периодическом проведении активного эксперимента без построения модели позволяют достигнуть оптимальных значений выходных показателей ТП. Подобный алгоритм движения к оптимуму является основой симплекс-планирования или последовательного симплексного метода планирования эксперимента.

Последовательный симплексный метод относится к методам поиска экстремума целевой функции, применение которого требует проведения минимально возможного числа опытов при определении направления движения при незначительных по объему вычислениях.

Построение планов эксперимента. Симплексом называют простейшую выпуклую геометрическую фигуру, имеющую в k -мерном факторном пространстве $k + 1$ вершину. Число вершин симплекса всегда на единицу больше размерности пространства эксперимента. В двумерном пространстве, т. е. на плоскости, симплексом является любой треугольник, в трехмерном — любая треугольная пирамида (тетраэдр). Координаты вершин симплексов и являются условиями проведения опытов при эксперименте. Построение плана эксперимента — задание координат (факторов x) в каждой вершине симплекса. Симплексы, как, например, и треугольники, могут быть разнообразными. Если расстояния между вершинами симплекса одинаковы, то его называют *правильным*. Из произвольного симплекса всегда можно получить правильный путем преобразования системы координат. Для простоты изложения сути метода в дальнейшем будем рассматривать только правильные симплексы.

Таким образом, план эксперимента при симплексном методе это совокупность $k + 1$ опытов, условия проведения которых задаются координатами вершин симплекса. Технологические режимы или варьируемые факторы представляют собой координаты факторного пространства. При этом важно задать начальный симплекс; движение к оптимуму и постановка новых опытов определяются алгоритмом метода.

Рассмотрим два способа задания координат вершин начального симплекса [14]:

1) одна из вершин симплекса помещена в начало координат, а остальные расположены так, что ребра, исходящие из первой вершины, образуют одинаковые углы с соответствующими координатными осями (рис. 4.4, а). Координаты вершин симплекса, а соответственно и план эксперимента в этом случае может быть представлены в виде матрицы,

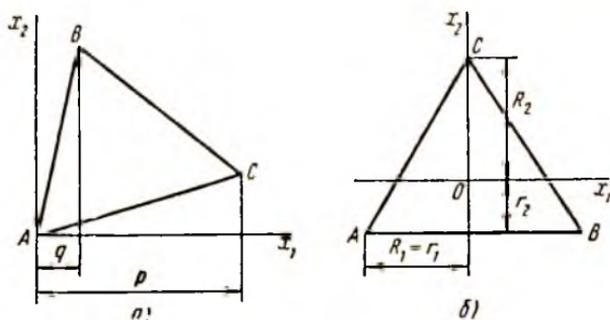


Рис. 4.4

приведенной в табл. 4.7, где значения факторов даны в кодированном виде. Для правильного симплекса

$$\rho = \frac{l}{k\sqrt{2}}(k-1 + \sqrt{k+1}); \quad q = \frac{l}{k\sqrt{2}}(\sqrt{k+1} - 1), \quad (4.15)$$

где $l = 1$ — расстояние между вершинами симплекса;

2) центр симплекса (центр плана эксперимента) помещен в начало координат, а $(k+1)$ -я вершина — на ось x_k (рис. 4.4, б), остальные вершины расположены симметрично относительно координатных осей. План эксперимента определяется матрицей (табл. 4.8), где при единичной длине ребра симплекса

Таблица 4.7

№ вершины	Координаты вершин			
	x_1	x_2	x_3	x_k
1	0	0	0	0
2	ρ	q	q	q
3	q	ρ	q	q
4	q	q	ρ	q
⋮	⋮	⋮	⋮	⋮
$k+1$	q	q	q	ρ

№ вершины	Координаты вершин				
	x_1	x_2	x_3	x_{l-1}	x_k
1	$-r_1$	$-r_2$	$-r_3$	$-r_{l-1}$	$-r_k$
2	R_1	$-r_2$	$-r_3$	$-r_{l-1}$	$-r_k$
3	0	R_2	$-r_3$	$-r_{l-1}$	$-r_k$
⋮	⋮	⋮	⋮	⋮	⋮
k	0	0	0	R_{l-1}	$-r_k$
$k+1$	0	0	0	0	R_k

$$r_j = \frac{1}{\sqrt{2j(j+1)}},$$

$$R_j = \frac{1}{\sqrt{2(j+1)}}, \quad j = \overline{1, k}. \quad (4.16)$$

Матрица координат вершин начального симплекса, построенного по первому способу при $k = 3$ ($l = 1$), приведена в табл. 4.9. Значения факторов кодированные.

Таблица 4.9

№ вершины	Координаты точек плана		
	x_1	x_2	x_3
1	0	0	0
2	0,944	0,236	0,236
3	0,236	0,944	0,236
4	0,236	0,236	0,944

Алгоритм симплексного метода. Экспериментальное определение оптимальных значений целевой функции осуществляют в такой последовательности:

- 1) выбирают целевую функцию, варьируемые факторы, интервалы варьирования факторов и производят кодирование факторов (задание таблицы исходных данных);
- 2) уточняют план эксперимента, т. е. рассчитывают координаты начального симплекса (первым или вторым способом). Число опытов в обоих случаях $N = k + 1$;
- 3) проводят соответствующие опыты по плану эксперимента;
- 4) упорядочивают отклики согласно степени их уменьшения и выделяют наименьшее значение отклика y_{\min} , которое затем отбрасывают;

5) строят новый симплекс. Он образуется оставшимися вершинами исходного симплекса и новой вершиной, получаемой путем зеркального отображения отброшенной относительно противоположной грани исходного симплекса (рис. 4.5). Координаты новой точки x рассчитывают следующим образом:

$$x^* = \frac{2}{k} (x^1 + x^2 + \dots + x^{j-1} + x^{j+1} + \dots + x^{k+1}) - x^j = \frac{2}{k} \sum_{i=1}^{k+1} x^i - \left(\frac{2}{k} + 1\right) x^j, \quad (4.17)$$

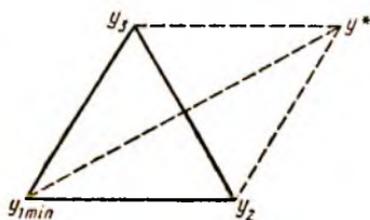


Рис. 4.5

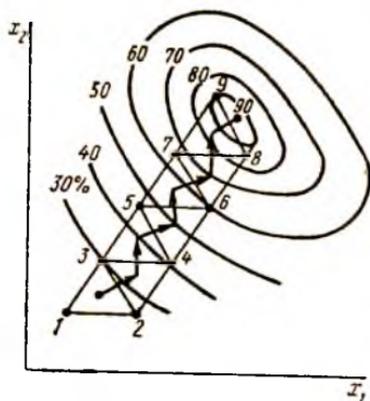


Рис. 4.6

где x^j — координаты вершины исходного симплекса с наименьшим значением целевой функции.

Построение плана в виде симплекса позволяет вычислить (предсказать) отклик в новой вершине x^* :

$$\hat{y}^* = \frac{2}{k} \sum_{i=1}^{k+1} \hat{y}^i - \left(\frac{2}{k} + 1\right) \hat{y}^j; \quad (4.18)$$

6) проводят эксперимент в новой точке x^* и получают экспериментальное значение y^* целевой функции;

7) последовательно перемещают симплекс, в процессе чего на каждом шаге происходит отбрасывание вершины симплекса с наименьшим значением целевой функции и реализация опыта в новой вершине. При этом направление движения совпадает с направлением градиента (рис. 4.6).

При движении симплекса в факторном пространстве могут возникнуть случаи, когда размеренное движение нарушается. Если при перемещении симплекса на протяжении $k + 1$ шагов одна из вершин сохраняет свое положение, то симплекс совершает оборот вокруг этой вершины (рис. 4.7, а). Это означает, что либо в данной точке находится оптимум целевой функции, либо значение целевой функции в этой вершине определено неверно (например, выброс). Для вы-

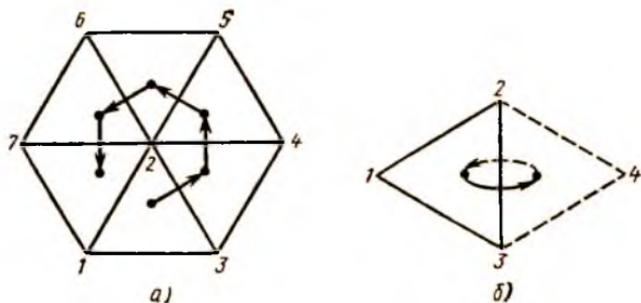


Рис. 4.7

яснения необходимо в этой точке повторить эксперимент и в дальнейшем использовать новое значение отклика.

Если отклик в новой вершине симплекса меньше, чем в других, то в соответствии с правилом движения необходимо возвратиться к предыдущему циклу (рис. 4.7, б). Для избежания «зацикливания» в исходном симплексе вершину, в которой значение отклика следует после минимального, исключают. Дальнейшее движение осуществляют в соответствии с п. 6.

При приближении к области оптимальных значений откликов (можно определить по разности значений откликов в вершинах симплексов) размер симплекса уменьшается на $1/2$ или на $1/4$ от начального.

Если ошибка эксперимента относительная величина, то целесообразно для каждой вершины симплекса проводить несколько параллельных опытов и использовать усредненные значения наблюдений целевой функции.

Преимуществами симплексного метода являются: малое число опытов, увеличение объема эксперимента только на один опыт при дополнительном включении нового фактора, простота вычислений, увеличение эффективности метода при росте числа факторов.

Пример 4.1. При технологической операции термообработки на процент выхода годных изделий y наибольшее влияние оказывают температура печи $x_1 \equiv T$, °С, время обработки $x_2 \equiv t$, ч. Необходимо подобрать режимы операции таким образом, чтобы, изменяя их в допустимых пределах, увеличить процент выхода годных изделий.

В табл. 4.10 приведены план эксперимента и результаты опытов. Минимальное значение отклика $y_{\min} = 24\%$ было получено во втором опыте. Следовательно, необходимо отбросить вторую вершину симплекса. По выражению (4.17) вычислим координаты новой вершины (четвертый опыт):

Таблица 4.10

№ опыта	x_1	x_2	y
1	0,866	0,500	26
2	-0,866	0,500	24
3	0	-1,0	28

$$x_1^* = \frac{2}{2} \left[0,866 + (-0,866) + 0 \right] - \left(\frac{2}{2} + 1 \right) (-0,866) = 1,732;$$

$$x_2^* = \frac{2}{2} \left[0,5 + 0,5 + (-1,0) \right] - \left(\frac{2}{2} + 1 \right) 0,5 = -1,0.$$

С помощью (4.18) предскажем значение отклика в четвертом опыте:

$$\hat{y}^* = \frac{2}{2} \left[26 + 24 + 28 \right] - \left(\frac{2}{2} + 1 \right) 24 = 30\%.$$

При этом экспериментальное значение отклика $y = 29,5\%$.

При симплексном методе экспериментальные опыты можно проводить через шаг, а в промежуточных оценивать отклики с помощью выражения (4.18).

§ 4.3. ЭВОЛЮЦИОННОЕ ПЛАНИРОВАНИЕ

Особую ценность методы оптимизации ТП приобретают при их применении непосредственно в производственных условиях. В результате оптимизации ТП улучшаются технические параметры готовой продукции и уменьшаются экономические затраты.

Особенности оптимизации. В производственных условиях невозможно контролировать все множество технологических факторов. Кроме того, не все факторы поддаются

контролю по различным метрологическим причинам. В результате существует значительное подмножество неконтролируемых факторов, которые затрудняют оценку откликов за счет смещения и искажения поверхности отклика. Поэтому необходимо периодически варьировать контролируемые факторы x . При этом следует: а) изменять факторы x_j только в интервале, разрешаемом технологической документацией, который, как правило, невелик; б) учитывать возникновение большого шумового поля, причиной которого является ошибка эксперимента и влияние неконтролируемых факторов.

Таким образом, возникает необходимость выделения слабого сигнала на фоне шума. При этом следует использовать метод накопления результатов измерения. Для этого необходимо слегка «покачивать» ТП, изменяя независимые переменные в малом интервале. Процесс оптимизации должен быть организован так, чтобы можно было получать не только годную продукцию, но и информацию о поверхности отклика и о местонахождении оптимума.

Изменение условий производства приводит к изменению поверхности отклика и местонахождения оптимума. Технолог должен учитывать это, т. е. осуществлять адаптационную оптимизацию.

Примером адаптационной оптимизации дискретных ТП является разработанный Дж. Боксом метод эволюционного планирования (ЭВОП). Суть метода заключается в том, что путем незначительного изменения режимов, постоянно вносимого в ТП, получают информацию, используемую для поиска оптимальных режимов процесса. В результате затруднительно построить математическую модель ТП, поэтому метод ЭВОП, как и метод симплексного планирования, своей целью ставит только поиск оптимальных режимов ТП.

Существенной особенностью метода ЭВОП является возможность надежного выявления малых изменений целевой функции на фоне шумов, что достигается путем повторения опытов плана эксперимента n раз. При этом среднеквадратичная ошибка среднего из n независимых измерений оказывается в \sqrt{n} раз меньше среднеквадратичной ошибки единичного измерения.

Алгоритм метода ЭВОП. Алгоритм данного метода имеет много общего с ранее рассмотренными.

На этапе планирования выбирают: 1) ограниченное число факторов (2—4), наиболее важных для данного ТП или отдельной операции; 2) начальную точку x_0 , около которой осуществляют варьирование факторов; 3) интервал варьирования

рования Δx_j ($j = \overline{1, k}$) с учетом особенностей производственных условий; 4) план эксперимента либо ПФЭ типа 2^k ($k = 2, 3$), либо 2^{k-1} ($k = 3, 4$) с одной или двумя центральными точками (двумя — при разбиении факторного эксперимента на два ортогональных блока).

Таким образом, число операционных условий

$$N = 2^k + (k - 1); \quad k = 2, 3;$$

$$N = 2^{k-1} + (k - 1); \quad k = 3, 4,$$

т. е. для $k = 2$ $N = 5$, причем один опыт в центре (для плана ПФЭ).

Проведение эксперимента по методу ЭВОП имеет особенности, так как необходимо выделить значимые изменения отклика на фоне его шумовых флуктуаций. Эксперимент проводят до тех пор, пока не добиваются существенного

изменения целевой функции y . На этом заканчивается часть эксперимента, называемая *фазой*. Каждая фаза состоит из n циклов, в каждом из которых выполняется $N = 2^k + (k - 1)$ опытов (рис. 4.8).

По данным эксперимента первой фазы определяют направление градиента к оптимальному режиму. По результатам каждого цикла заполняют табл. 4.11. Анализ данных таблицы заключается в вычислении ошибки эксперимента (среднеквадратичного отклонения) и доверительных интервалов эффектов и последующем сравнении их с эффектом того или иного фактора. Если доверительный интервал фактора после n -го цикла больше его эффекта, то заметного изменения отклика при варьировании этого фактора не происходит, и наоборот, если доверительный интервал меньше эффекта, то наблюдается заметное изменение отклика и необходимо переходить к следующим операционным условиям.

Разность (табл. 4.11)

$$D_l = \left(\frac{1}{n-1} \sum_{i=1}^{n-1} y_{li} \right) - y_{ln}, \quad l = \overline{1, N},$$

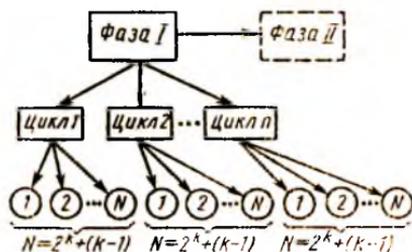


Рис. 4.8

Таблица 4.11

y_6 y_2	y_1	y_3 y_4	ЭВОП, цикл 1 ($n-1$) Отклик выход продукта				Фаза I Дата
Вычисление средних значений						Вычисление среднеквадратичных отклонений	
№ опытов	1	2	3	4	5		
1. Сумма предыдущих циклов						Сумма предыдущих циклов $s =$	
2. Среднее предыдущих циклов						Среднее предыдущих циклов $s =$	
3. Новые наблюдения	94,0	94,5	96,5	94,5	94,5	Размах $R =$	
4. Разности п. 2—п. 3						Новое значение $s = Rf_{N-1} =$	
5. Новые суммы п. 1+ + п. 3	94,0	94,5	96,5	94,5	94,5	Новая сумма $s_n =$	
6. Новые средние	94,0	94,5	96,5	94,5	94,5	Новое среднее значение $s = \frac{s_n}{n-1}$	
Вычисление эффектов						Вычисление доверительных интервалов (ДИ)	
Температуры	$0,5(\bar{y}_3 + \bar{y}_4 - \bar{y}_2 - \bar{y}_5) = 1,0$					Для новых средних значений $\frac{2}{\sqrt{n}} s =$	
Давления	$0,5(\bar{y}_3 + \bar{y}_5 - \bar{y}_2 - \bar{y}_4) = 1,0$					Для эффектов $\frac{2}{\sqrt{n}} s =$	
Взаимодействия	$0,5(\bar{y}_2 + \bar{y}_3 - \bar{y}_4 - \bar{y}_5) = 1,0$					Для центральной точки $\frac{1,78}{\sqrt{n}} s =$	
Центральной точки	$0,2(\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 - 4\bar{y}_1) = 0,8$						

где y_{li} — наблюдение отклика в l -м опыте и в i -м цикле; n — число циклов.

Поскольку дисперсия суммы независимых случайных величин равна сумме дисперсий, а дисперсия произведения случайной величины на постоянную равна квадрату постоянной, умноженной на дисперсию случайной величины, то

$$\sigma_{D_l}^2 = \left(\frac{1}{(n-1)^2} \sum_{i=1}^{n-1} \sigma_{y_{li}}^2 \right) - \sigma_{y_{ln}}^2. \quad (4.19)$$

Так как предполагают, что все измерения целевой функции независимы и имеют одну и ту же дисперсию σ_y^2 , то

$$\sigma_D^2 = \frac{n}{n-1} \sigma_y^2. \quad (4.20)$$

Отсюда

$$\sigma_y = \sqrt{\frac{n-1}{n}} \sigma_D. \quad (4.21)$$

Таким образом, оценка s_D^2 значения дисперсии D_l может быть использована для получения оценки s_y^2 дисперсии ошибки наблюдения. Для упрощения вычислений оценку s_D^2 получают, используя размах значений D :

$$s_D = R_D/d; \quad R_D = |D_{l \max} - D_{l \min}|. \quad (4.22)$$

Из (4.21) и (4.22) следует, что

$$s_y = \frac{1}{d} \sqrt{\frac{n-1}{n}} R_D. \quad (4.23)$$

Значения $\frac{1}{d} \sqrt{\frac{n-1}{n}} = f_{n,N}$ табулированы (табл. 4.12). Для $n = 3$ и некоторого плана с пятью точками ($2^2 + 1 = 5$) находим $f_{3,5} = 0,35$. Для оценки ошибки воспроизводимости получаем $s_y = R_D f_{n,N}$.

Для проверки значимости эффектов факторов определяют доверительный интервал эффектов факторов и взаимодействия, а также эффекта центральной точки:

$$\Delta_\phi = \pm \theta \frac{2}{\sqrt{N-1}} \frac{s_y}{\sqrt{n}}; \quad (4.24)$$

$$\Delta_{ii} = \pm \theta \sqrt{\frac{N-1}{N}} \frac{s_y}{\sqrt{n}},$$

где θ находят по табл. 4.13 для нормированной функции Лапласа из условия $\bar{\Phi}(\theta) = 0,5 \alpha$; α — доверительная вероятность.

Для $k = 2$, $N = 2^2 + 1$ и $\alpha = 0,95$ из табл. 4.13 находим $\theta = 2$ и, следовательно, $\Delta_{\phi} = \pm 2s_y/\sqrt{n}$; $\Delta_{\pi} = \pm 1,78 s_y/\sqrt{n}$.

Таблица 4.12

№ циклов n	Число опытов N						
	4	5	6	7	8	9	10
2	0,32	0,3	0,28	0,26	0,25	0,24	0,23
3	0,4	0,35	0,32	0,3	0,29	0,27	0,26
4	0,42	0,37	0,34	0,32	0,3	0,29	0,28
5	0,43	0,38	0,35	0,33	0,31	0,3	0,29
6	0,44	0,39	0,36	0,34	0,32	0,31	0,3
7	0,45	0,4	0,37	0,34	0,33	0,31	0,3
8	0,45	0,4	0,37	0,35	0,33	0,31	0,3
9	0,46	0,4	0,37	0,35	0,33	0,32	0,31
10	0,46	0,41	0,37	0,35	0,33	0,32	0,31
11	0,46	0,41	0,38	0,35	0,33	0,32	0,31
12	0,47	0,41	0,38	0,35	0,34	0,32	0,31
13	0,47	0,41	0,38	0,36	0,34	0,32	0,31
14	0,47	0,41	0,38	0,36	0,34	0,32	0,31
15	0,47	0,42	0,38	0,36	0,34	0,33	0,31

Эффекты факторов и взаимодействий для $k = 2$

$$\mathcal{E}_{\phi j} = 0,5 \sum_{i=1}^N x_{ji} y_i; \quad \mathcal{E}_{\pi jl} = 0,5 \sum_{i=1}^N (x_j x_l)_i y_i.$$

эффект центральной точки

$$\mathcal{E}_{\pi} = 0,2 \left[\left(\sum_{i=2}^4 y_i \right) - 4y_1 \right].$$

Если $\mathcal{E}_{\phi} > \Delta_{\phi}$; $\mathcal{E}_{\pi} \geq \Delta_{\pi}$, то эффекты факторов и центральной точки существуют и необходимо переходить к новой фазе.

План эксперимента в новой фазе выбирают в градиентном направлении. При этом могут рассматриваться новые факторы, за исключением тех факторов, эффекты влияния которых невелики.

Необходимо учитывать, что при приближении к оптимуму отклик меняется меньше, следовательно, число циклов возрастает. Число циклов зависит также от инерционности процесса: так как интервалы циклов в малоинерционных процессах невелики, то число циклов увеличивается. Чис-

ло циклов зависит и от интервала варьирования: чем больше Δx_j , тем меньше циклов.

Таблица 4.13

$$\varphi(x), x \geq 0; \varphi(-x) = \varphi(x); \bar{\Phi}(x) = \int_0^x \varphi(x) dx; x \geq 0; \Phi(x) = 0,5 - \bar{\Phi}(x),$$

$$\Phi(-x) = 0,5 - \bar{\Phi}(x)$$

x	$\varphi(x)$	$\bar{\Phi}(x)$	x	$\varphi(x)$	$\bar{\Phi}(x)$	x	$\varphi(x)$	$\bar{\Phi}(x)$
0	0,3989	0	1,15	0,2059	0,374928	2,3	0,0283	0,489276
0,5	0,3984	0,019939	1,2	0,1942	0,38493	2,35	0,0252	0,490613
0,1	0,3970	0,039828	1,25	0,1826	0,39435	2,4	0,0224	0,491802
0,15	0,3975	0,059618	1,3	0,1714	0,4032	2,45	0,0198	0,492857
0,2	0,391	0,07926	1,35	0,1604	0,411492	2,5	0,0175	0,49379
0,25	0,3867	0,098706	1,4	0,1497	0,419243	2,55	0,0154	0,494614
0,3	0,3814	0,117911	1,45	0,1394	0,426471	2,6	0,0136	0,495339
0,35	0,3752	0,136831	1,5	0,1295	0,433193	2,65	0,0119	0,495975
0,4	0,3683	0,155422	1,55	0,1200	0,439429	2,7	0,0104	0,496533
0,45	0,3605	0,173645	1,6	0,1109	0,445201	2,75	0,0091	0,49702
0,5	0,3521	0,191463	1,65	0,1023	0,450528	2,8	0,0079	0,497445
0,55	0,3429	0,208840	1,7	0,094	0,455434	2,85	0,0069	0,497814
0,6	0,3332	0,225747	1,75	0,0863	0,459941	2,9	0,006	0,498134
0,65	0,323	0,242154	1,8	0,079	0,46407	2,95	0,005	0,498411
0,7	0,3123	0,258035	1,85	0,0721	0,467843			
0,75	0,3011	0,273373	1,9	0,0656	0,471283			
0,8	0,2897	0,288145	1,95	0,0596	0,474412			
0,85	0,278	0,302338	2,00	0,054	0,47725			
0,9	0,2661	0,315940	2,05	0,0488	0,479818			
0,95	0,2541	0,328944	2,1	0,044	0,482136			
1,0	0,2420	0,341345	2,15	0,0396	0,484222			
1,05	0,2299	0,353141	2,2	0,0355	0,486097			
1,1	0,2179	0,364334	2,25	0,0317	0,487776			

Пример 4.2. Рассмотрим двухфакторную технологическую операцию, где $x_1 \equiv T$ — температура операции, °С; $x_2 \equiv P$ — давление, Па; y — выход годной продукции, %. Поскольку $k = 2$, число опытов $N = 2^k + (k - 1) = 5$. Стандартная форма ЭВОП определяет местоположение пяти точек этого плана (табл. 4.11). Сравнивая отклики y (средние отклики) в точках 3 и 4 с откликами в точках 2 и 5, можно определить влияние x_1 . Аналогично сравнение $y_{(2-4)}$ и $y_{(3-5)}$ позволяет найти x_2 . При сравнении $y_{(2-3)}$ и $y_{(5-4)}$ определяют эффект взаимодействия, а при сравнении $y_{(2-3-4-5)}$ и y_1 — эффект центральной точки (эффект среднего). Данные эксперимента и результаты вычислений заносятся в карты ЭВП (табл. 4.14, 4.15). В табл. 4.11 приведены результаты первого, в табл. 4.14 — второго, а в табл. 4.15 — третьего цикла. Как видно из табл. 4.15, эффект температуры 1,65 превышает доверительный интервал ошибки 1,16. Таким образом, температура значительно влияет на процент выхода годной продукции.

y_1 y_2	y_1	y_2 y_4	ЭВОП. цикл 2 ($n=2$) Отклик: выход продукта				Фаза I Дата
Вычисление средних значений						Вычисление среднеквадратичных отклонений	
№ опытов	1	2	3	4	5		
1. Сумма предыдущих циклов	94,0	94,5	96,5	94,5	94,5	Сумма предыдущих циклов $s =$ Среднее значение предыдущих циклов $s =$	
2. Среднее предыдущих циклов	94,0	94,5	96,5	94,5	94,5	Размах $R = 1,5 + -2,0 = 3,5$ Новое значение $s = Rf_{n,n} = 3,5 \cdot 0,3 =$ $= 1,05$	
3. Новые наблюдения	96,0	95,0	95,0	96,5	94,0	Новая сумма $s_n = 1,05$	
4. Разности п. 2—п. 3	-2,0	-0,5	1,5	-2,0	0,5	Новое среднее значение	
5. Новые суммы п. 1+ + п. 3	190,0	189,5	191,5	191,0	188,5	$s = \frac{s_n}{n-1} = \frac{1,05}{1} = 1,05$	
6. Новые средние	95,0	94,7	95,7	95,5	94,2		
Вычисление эффектов						Вычисление доверительных интервалов (ДИ)	
Температуры	$0,5(\bar{y}_3 + \bar{y}_4 - \bar{y}_2 - \bar{y}_5) = 1,15$					Для новых средних значений $\frac{2s}{\sqrt{n}} = \frac{2 \cdot 1,05}{\sqrt{2}}$	
Давления	$0,5(\bar{y}_3 + \bar{y}_5 - \bar{y}_2 - \bar{y}_4) = -0,15$					Для эффектов $\frac{2s}{\sqrt{n}} = 1,48$	
Взаимодействия	$0,5(\bar{y}_2 + \bar{y}_3 - \bar{y}_4 - \bar{y}_5) = 0,35$					Для центральной точки $\frac{1,78s}{\sqrt{n}} =$ $= 1,32$	
Центральной точки	$0,2(\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 - 4\bar{y}_1) = 0,02$						

Таблица 4.15

y_0 y_2	y_1	y_3 y_4	ЭВОП, цикл 3 ($n=3$) Отклик: выход продукта			Фаза I Дата
Вычисление средних значений					Вычисление среднеквадратичных отклонений	
№ опытов	1	2	3	4	5	
1. Сумма предыдущих циклов	190,0	189,5	191,5	191,0	188,5	Сумма предыдущих циклов $s = 1,05$
2. Среднее предыдущих циклов	95,0	94,7	95,7	95,5	94,2	Среднее предыдущих циклов $s = 1,05$
3. Новые наблюдения	94,5	93,5	96,0	97,0	94,0	Размах $R = 1,2 + -1,5 = 2,7$
4. Разности п. 2—п. 3	0,5	1,2	-0,3	-1,5	0,2	Новое значение $s = Rf_{n,n} = 2,7 \times 0,35 = 0,95$
5. Новые суммы п. 1+ + п. 3	284,5	281,0	287,2	286,5	282,7	Новая сумма $s_n = 1,05 + 0,95 = 2,00$
6. Новые средние	94,8	93,7	95,7	95,5	94,2	Новое среднее значение $s = \frac{s_n}{n-1} = \frac{2,00}{2} = 1,0$
Вычисление эффектов					Вычисление доверительных интервалов (ДИ)	
Температуры $1/2(\bar{y}_3 + \bar{y}_4 - \bar{y}_2 - \bar{y}_5) = 1,65$					Для новых средних $\frac{2s}{\sqrt{n}} = \frac{2 \cdot 1,05}{\sqrt{2}} = 1,48$	
Давления $1/2(\bar{y}_3 + \bar{y}_5 - \bar{y}_2 - \bar{y}_4) = 0,35$					Для эффектов $\frac{2s}{\sqrt{n}} = \frac{2 \cdot 1,05}{\sqrt{2}} = 1,48$	
Взаимодействия $1/2(\bar{y}_2 + \bar{y}_3 - \bar{y}_4 - \bar{y}_5) = -0,15$					Для центральной точки $\frac{1,78s}{\sqrt{n}} = \frac{1,78 \cdot 1,05}{\sqrt{2}} = 1,32$	
Центральной точки $1/5(\bar{y}_2 + \bar{y}_3 + \bar{y}_4 + \bar{y}_5 - 4\bar{y}_1) = -0,02$						

§ 4.4. ОПТИМИЗАЦИЯ СВОЙСТВ МНОГОКОМПОНЕНТНЫХ МАТЕРИАЛОВ

Многие из применяемых в электронной технике материалов представляют собой смеси различных компонентов (полупроводниковые материалы, ферриты, клеи и т. д.). Свойства этих материалов зависят от концентрации процентного содержания составляющих их компонентов. Целью оптимизации является установление количественных зависимостей между концентрациями компонентов и свойствами смеси и поиск составов, обеспечивающих экстремальные свойства материала.

Особенностью задач исследования многокомпонентных материалов является взаимозависимость варьируемых факторов — концентрации компонентов, т. е. изменив концентрацию одного компонента, необходимо изменять концентрацию нескольких или всех остальных компонентов. Состав смеси обычно выражают в мольных, весовых или объемных долях, а также в процентах от общей суммы. Для смесей выполняется соотношение

$$\sum_{j=1}^k x_j = 1 \text{ (100\%)}, \quad (4.25)$$

где $x_j \geq 0$ — концентрация j -го компонента; k — число компонентов смеси.

Эта принципиальная особенность исследуемой задачи влияет на алгоритм ее решения (на построение плана, вид модели и др.). Соотношение (4.25), задающее возможные составы k -компонентной смеси, представляет собой правильный симплекс с k вершинами в $(k - 1)$ -мерном пространстве. Для изображения составов используют барицентрическую систему координат.

Например, для двухкомпонентной смеси ($k = 2$) симплекс — прямая [содержание компонентов определяется соотношением отрезков (рис. 4.9, а)]; для трехкомпонентной ($k = 3$) правильный симплекс — равносторонний треугольник. При этом каждая точка треугольника соответствует определенному составу тройной смеси и, наоборот, каждый состав характеризуют одной определенной точкой. Метод изображения свойств тройной смеси с помощью концентрического треугольника получил широкое распространение в физической химии. Вершины треугольника соответствуют чистым веществам, стороны — двойным смесям. Опустив из каждой вершины треугольника A , B , и C перпендикуляры, разделив каждый из них на десять равных отрезков

и проводя через полученные деления прямые, параллельные сторонам треугольника, получим 10 %-ную сетку (рис. 4.9, б). Приближение от каждой данной стороны к противоположной вершине соответствует пропорциональному возрастанию содержания соответствующего вершине компонента от 0 до 1. Концентрация компонентов в любой точке треугольника определяется длиной перпендикуляров, опущенных из этой точки на стороны треугольников.

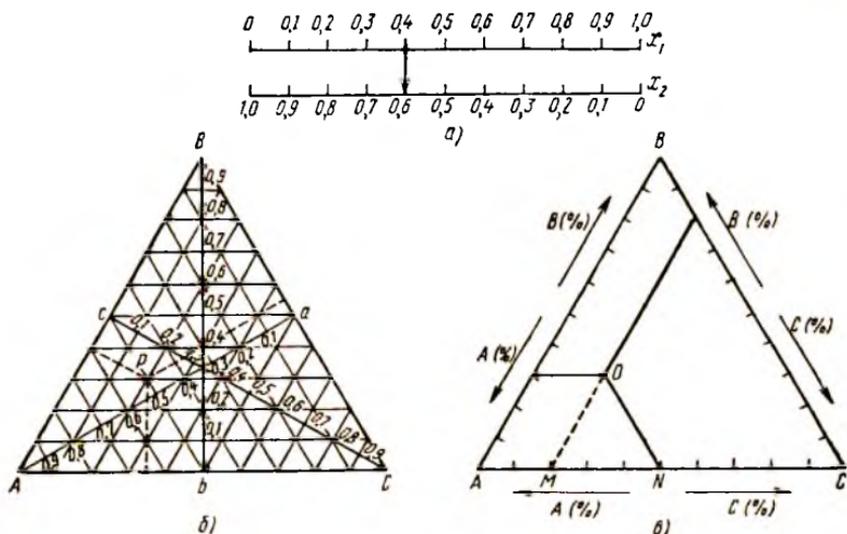


Рис. 4.9

Так, изображенная на рис. 4.9, б точка P имеет следующие концентрации компонентов: $A = 0,5$; $B = 0,3$; $C = 0,2$.

Другой более простой способ отсчета концентраций предложен Г. Розенбумом. В треугольнике Розенбума из точки внутри треугольника проводят прямые, параллельные его сторонам, состав тройной смеси отсчитывают по трем отрезкам, отсекаемым этими прямыми на соответствующих сторонах треугольника (рис. 4.9, в). Точка P в треугольнике Розенбума имеет следующие концентрации компонентов: $A = 0,5$; $B = 0,3$; $C = 0,2$. В дальнейшем для изображения смесей принят способ Розенбума.

Целевую функцию y трехкомпонентной смеси обычно представляют проекциями линий равного уровня на плоскость концентрационного треугольника. В результате получают диаграмму, связывающую определенное свойство (значение y) с составом смеси (диаграмма состав — свойство).

При $k = 4$ концентрационный симплекс — тетраэдр каждая вершина которого соответствует чистым компонентам, ребра — двухкомпонентным, а грани — трехкомпонентным смесям. Для определения координат точек внутри тетраэдра через эту точку проводят плоскости, параллельные граням тетраэдра, и вычисляют расстояния от этих плоскостей до граней, противоположных вершинам, соответствующих 100%-ным содержаниям компонентов (на гранях, противоположных вершинам, содержание соответствующих вершинам компонентов равно нулю (начало отсчета)).

При изучении свойства сложных смесей с числом компонентов $k > 3$ диаграмму графически представляют в виде линий равного уровня в сечениях $(k - 1)$ -мерного симплекса плоскостями, параллельными его граням. При этом в сечениях варьируются три компонента, остальные $(q - 3)$ закрепляют на определенных уровнях (сечения диаграмм состав — свойство).

Задание области исследования. При оптимизации свойств материалов электронной техники рассматривают не полный концентрационный симплекс, а лишь некоторый локальный участок (свойства чистых материалов или смеси с меньшим числом составляющих не важны, т. е. не анализируют вершины и стороны симплекса).

Локальный участок задается определенными ограничениями на возможное содержание компонентов в смеси:

$$0 < a_j \leq x_j \leq b_j < 1; \quad j = \overline{1, k}. \quad (4.26)$$

Ограничения (4.26) на концентрации компонентов выделяют в полном концентрационном симплексе некоторую подобласть (в общем случае в виде многогранника), число и координаты вершин которой зависят от заданных ограничений.

Координаты вершин многогранника, ограничивающего локальную область исследования, определяют следующим образом:

1) выписывают все возможные комбинации двух уровней a_j и b_j для каждого компонента при пропуске одной из них (таких комбинаций 2^{k-1}) и повторяют эту процедуру k раз. Всего получают $k \cdot 2^{k-1}$ комбинаций уровней с пропусками одного из компонентов. Приведем возможные комбинации для $k = 3$:

$$a_1 a_2, a_1 b_2, b_1 a_2, b_1 b_2, a_1 a_3, a_1 b_3,$$

$$b_1 a_3, b_1 b_3, a_2 a_3, a_2 b_3, b_2 a_3, b_2 b_3;$$

2) среди полученных комбинаций выбирают те, в которых сумма компонентов меньше единицы и выполняются условия $0 < a_j \leq x_j \leq b_j < 1$; $j = \overline{1, k}$. Затем в эти комбинации на места пропусков добавляют недостающие до общей суммы компоненты, равной 1 или 100% с учетом заданных ограничений, и вычеркивают одинаковые комбинации. В результате будут заданы вершины многогранника (область исследования).

Таблица 4.16

№ варианта	x_1	x_2	x_3
1	0,2	0,18	—
2	0,2	0,67	—
3	0,6	0,18	—
4	0,6	0,67	—
5	0,2	—	0,05
6	0,2	—	0,4
7	0,6	—	0,05
8	0,6	—	0,4
9	—	0,18	0,05
10	—	0,18	0,4
11	—	0,67	0,05
12	—	0,67	0,4

В качестве примера рассмотрим построение локальной области исследования, когда на концентрации компонентов смеси наложены следующие ограничения:

$$0,2 \leq x_1 \leq 0,6; 0,18 \leq x_2 \leq 0,67; 0,05 \leq x_3 \leq 0,4. \quad (4.27)$$

Таблица 4.17

№ варианта	x_1	x_2	x_3
1	0,2	0,18	$0,62 > 0,4$
2	0,2	0,67	0,13
3	0,6	0,18	0,22
4	0,6	0,67	> 1
5	0,2	$0,75 > 0,67$	0,05
6	0,2	0,4	0,4
7	0,6	0,35	0,05
8	0,6	$0 < 0,18$	0,4
9	$0,77 > 0,6$	0,18	0,05
10	0,42	0,18	0,4
11	0,28	0,67	0,05
12	> 1	0,67	0,4

Полный перебор уровней концентрации компонентов в соответствии с п. 1 приведен в табл. 4.16, а координат вершины локального многогранника в соответствии с п. 2 — в табл. 4.17. Варианты смесей 1, 5, 8, 9 отбрасываются из-за нарушений условий (4.27), а варианты 4, 12 из-за нарушения условия (4.25). В результате вершины многогранника соответствуют вариантам 2, 3, 6, 7, 10 и 11 из полного перебора. Графическое изображение многогранника показано на рис. 4.10.

Выбор математической модели. Характерной особенностью задачи оптимизации свойств многокомпонентных материалов является сложный вид поверхности отклика, отражающий зависимость свойств от состава смеси. Для адекватного описания таких зависимостей необходимы полиномы достаточно высоких степеней вплоть до четвертого порядка. Кроме того, если в матрице X одновременно имеются столбцы фиктивного фактора $x_0 = 1$, парных взаимодействий $x_j x_i$ и квадратов x_j^2 , то согласно (4.25) ковариационная матрица оказывается вырожденной. Поэтому приходится описывать свойства смеси с помощью приведенных полиномов.

Рассмотрим порядок приведения полиномов на примере полинома второй степени для описания свойств трехкомпонентной смеси. Модель второго порядка для $k = 3$ имеет вид

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2.$$

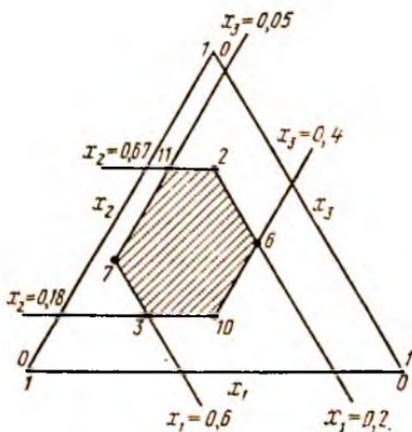


Рис. 4.10

В этой модели необходимо устранить b_0 (коэффициент при фиктивном факторе x_0) и квадратичные эффекты. Поскольку $x_1 + x_2 + x_3 = 1$, умножив левую и правую части этого выражения на b_0 , получим $b_0 x_1 + b_0 x_2 + b_0 x_3 = b_0$, а умножив это выражение последовательно на x_1, x_2, x_3 соответственно найдем $x_1^2 = x_1 - x_1 x_2 - x_1 x_3$; $x_2^2 = x_2 - x_1 x_2 - x_1 x_3$; $x_3^2 = x_3 - x_1 x_3 - x_2 x_3$.

Подставив полученные соотношения в первоначальные выражения и приведя подобные члены, определим

$$y = (b_0 + b_1 + b_{11}) x_1 + (b_0 + b_2 + b_{22}) x_2 + (b_0 + b_3 + b_{33}) x_3 + \\ + (b_{12} - b_{11} - b_{12}) x_1 x_2 + (b_{13} - b_{11} - b_{33}) x_1 x_3 + \\ + (b_{23} - b_{22} - b_{33}) x_2 x_3.$$

Введя обозначения $\beta_j = b_0 + b_j + b_{jj}$; $\beta_{jl} = b_{jl} - b_{jj} - b_{ll}$, получим приведенный полином второй степени трех переменных:

$$\widehat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3.$$

Таким образом, число коэффициентов полинома второй степени для трех переменных уменьшилось с 10 до 6, следовательно, требуется меньшее число опытов для построения модели.

Приведенный полином второй степени от k переменных

$$\widehat{y} = \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l$$

содержит C_{k+1}^2 коэффициентов.

Приведенный полином неполного третьего порядка для трехкомпонентной смеси

$$\widehat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \\ + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3,$$

для k -компонентной

$$\widehat{y} = \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l + \sum_{1 \leq j < l < s \leq k} \beta_{jls} x_j x_l x_s.$$

Приведенный полином третьего порядка для трехкомпонентной смеси

$$\widehat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \\ + \nu_{12} x_1 x_2 (x_1 - x_2) + \nu_{13} x_1 x_3 (x_1 - x_3) + \nu_{23} x_2 x_3 (x_2 - x_3) + \\ + \beta_{123} x_1 x_2 x_3,$$

для k -компонентной

$$\widehat{y} = \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l + \\ + \sum_{1 \leq j < l \leq k} \nu_{jl} x_j x_l (x_j - x_l) + \sum_{1 \leq j < l < s \leq k} \beta_{jls} x_j x_l x_s.$$

Приведенный полином четвертого порядка для трехкомпонентной смеси

$$\begin{aligned} \widehat{y} = & \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \\ & + \beta_{23} x_2 x_3 + v_{12} x_1 x_2 (x_1 - x_2) + v_{13} x_1 x_3 (x_1 - x_3) + \\ & + v_{23} x_2 x_3 (x_2 - x_3) + \delta_{12} x_1 x_2 (x_1 - x_2)^2 + \\ & + \delta_{13} x_1 x_3 (x_1 - x_3)^2 + \delta_{23} x_2 x_3 (x_2 - x_3)^2 + \\ & + \beta_{1123} x_1^2 x_2 x_3 + \beta_{1223} x_1 x_2^2 x_3 + \beta_{1233} x_1 x_2 x_3^2, \end{aligned}$$

для k -компонентной

$$\begin{aligned} \widehat{y} = & \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l + \sum_{1 \leq j < l \leq k} v_{jl} x_j x_l (x_j - x_l) + \\ & + \sum_{1 \leq j < l \leq k} \delta_{jl} x_j x_l (x_j - x_l)^2 + \sum_{1 \leq j < l < s \leq k} \beta_{jls} x_j^2 x_l x_s + \\ & + \sum_{1 \leq j < l < s \leq k} \beta_{jlls} x_j x_l^2 x_s + \sum_{1 \leq j < l < s \leq k} \beta_{jls s} x_j x_l x_s^2 + \\ & + \sum_{1 \leq j < l < s < r \leq k} \beta_{jlsr} x_j x_l x_r x_s, \end{aligned}$$

где v, δ, β — коэффициенты регрессии.

Планирование эксперимента. Минимальное число опытов N_{min} , необходимое для оценки коэффициентов полиномиальной модели по экспериментальным данным, равно числу коэффициентов выбранного полинома, т. е. для модели степени μ и k компонентов $N_{\text{min}} = C_{k+\mu}^{\mu}$.

Рассмотрим некоторые планы, которые применяют при оптимизации свойств многокомпонентных материалов.

Симплекс-решетчатые планы. В настоящее время наибольшее распространение получили симплекс-решетчатые планы. Эти планы обеспечивают равномерное распределение экспериментальных точек по $(k-1)$ -мерному симплексу и являются насыщенными (число опытов плана равно числу коэффициентов оцениваемой модели). Точки симплексной решетки степени μ для k компонентов содержат все возможные сочетания $(\mu+1)$ -го равноотстоящего уровня по каждому компоненту с координатами $x_j = 0, 1/\mu, 2/\mu, \dots, 1$.

Для квадратичной решетки $\{k, 2\}$, обеспечивающей построение модели второго порядка ($\mu = 2$), используют следующие уровни каждого компонента: 0, 1/2, 1, а обеспечивающей построение модели третьего порядка ($\mu = 3$) — 0,

Таблица 4.18

№ опыта	x_1	x_2	x_3	$y_{\text{экс}}$	№ опыта	x_1	x_2	x_3	$y_{\text{экс}}$
1	1	0	0	y_1	4	1/2	1/2	0	y_{12}
2	0	1	0	y_2	5	1/2	0	1/2	y_{13}
3	0	0	1	y_3	6	0	1/2	1/2	y_{23}

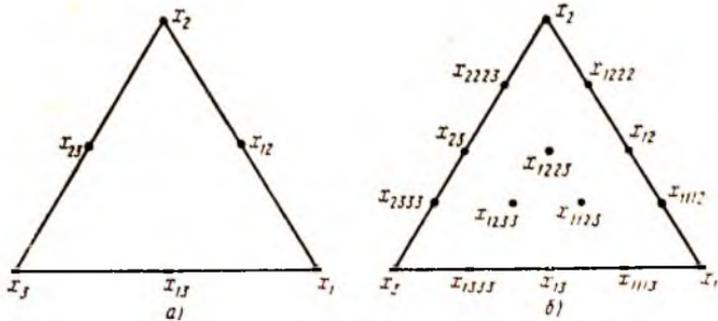


Рис. 4.11

1/3, 2/3, 1 и т. д. Некоторые $\{3, \mu\}$ - и $\{4, \mu\}$ -решетки представлены на рис. 4.11, а, б и 4.12, а, б. Записав координаты точек симплексной решетки, получим планы эксперимента. Планы экспериментов для решеток $\{3, 2\}$, $\{3, 3\}$, $\{3, 4\}$ и $\{4, 2\}$ представлены в табл. 4.18, 4.19 соответственно.

Таблица 4.19

№ опыта	x_1	x_2	x_3	$y_{\text{экс}}$	№ опыта	x_1	x_2	x_3	$y_{\text{экс}}$
1	1	0	0	y_1	9	3/4	0	1/4	y_{1113}
2	0	1	0	y_2	10	1/4	0	3/4	y_{1333}
3	0	0	1	y_3	11	0	3/4	1/4	y_{2223}
4	1/2	1/2	0	y_{12}	12	0	1/4	3/4	y_{2333}
5	1/2	0	1/2	y_{13}	13	1/2	1/4	1/4	y_{1123}
6	0	1/2	1/2	y_{23}	14	1/4	1/2	1/4	y_{1223}
7	3/4	1/4	0	y_{1112}	15	1/4	1/4	1/2	y_{1233}
8	1/4	3/4	0	y_{1222}					

но. Для обозначения свойств смеси y вводят индексы, указывающие, при каком содержании компонентов в смеси получены данные значения y . Так, значения y_{12} (табл. 4.18) соответствуют смеси, содержащей 1/2 x_1 и 1/2 x_2 ; значения y_{1113} — смеси, содержащей 3/4 x_1 , 0 x_2 , 1/4 x_3 ,

а значения y_{1233} — смеси, содержащей $1/4 x_1, 1/4 x_2$ и $1/2 x_3$ (табл. 4.19).

Симплекс-решетчатые планы являются частично композиционными, т. е. допускающими достройку плана без потери результатов предыдущих опытов. Так, как видно из рис. 4.11 и 4.12, в случае, когда полинома второго порядка недостаточно для адекватного описания заданного свойства, можно без потери результатов предыдущих опытов перейти к пла-

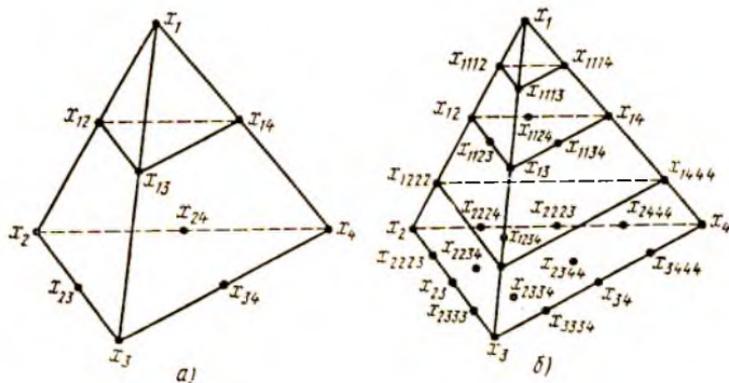


Рис. 4.12

нам неполного третьего или четвертого порядка. План третьего порядка композиционен только линейному плану. Для описания свойств смеси по результатам эксперимента используют описанные приведенные полиномы соответствующих степеней.

П л а н ы Л а м б р а к и с а. В симплекс-решетчатых планах для полиномов невысоких степеней в большинстве опытов вводят не все компоненты. Очевидно, что результаты опытов с чистыми компонентами и двухкомпонентными смесями несут мало информации о свойствах многокомпонентной смеси. Поэтому для смесей с числом компонентов $k \geq 4$ предложены планы Ламбракиса. Эти планы представляют собой симплексные решетки Шеффе с заменой опытов с чистыми компонентами опытами с координатами: $x_1 = x_2 = \dots = x_k = 1/(k - 1)$.

Например, при построении полинома второй степени в четырехкомпонентной смеси следует четыре точки с координатами $x_1 = x_2 = x_3 = x_4 = 1$ (рис. 4.12 а), заменить четырьмя точками с координатами $x_1 = x_2 = x_3 = x_4 = 1/3$ (рис. 4.13). Таким образом, план Ламбракиса вместо четырех опытов в вершинах тетраэдра включает четыре опы-

та в центрах граней тетраэдра, а также шесть опытов в серединах ребер (из плана Шеффе). Для описания свойств по результатам реализации плана Ламбракиса используют приведенные полиномы соответствующих степеней.

Симплекс-центроидные планы. Для описания свойств многокомпонентных смесей полиномами невысоких степеней Шеффе предложены также симплекс-центроидные планы, содержащие $(2^k - 1)$ точку с координатами $(1, 0, \dots, 0)$, $(1/2, 1/2, 0, \dots, 0)$, ..., $(1/k, 1/k, \dots, 1/k)$, а также все точки, которые можно получить из этих перестановками координат. Для данного числа компонентов k можно построить единственный симплекс-центроидный план, содержащий точку в центре (центроид) симплекса и центроиды всех симплексов нижней размерности, его составляющих. Симплекс-центроидные планы для $k = 3, 4$ представлены на рис. 4.13, матрица симплекс-центроидного плана для $k = 4$ дана в табл. 4.20. Как видно из сравнения рис. 4.12 и 4.13 симплекс-центроидный план для $k = 3$ совпадает с симплекс-решетчатым планом неполного третьего порядка.

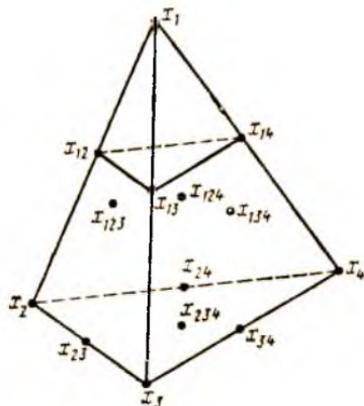


Рис. 4.13

Полиномы, описывающие свойства смеси согласно результатам проведенных экспериментов по симплекс-центроидному плану, содержат столько же коэффициентов, сколько точек в плане. Для k -компонентной смеси

$$\hat{y} = \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l + \sum_{1 \leq j < l < s \leq k} \beta_{jls} x_j x_l x_s + \beta_{12 \dots k} x_1 x_2 \dots x_k. \quad (4.28)$$

Для $k = 4$ модель, получаемая по результатам реализации симплекс-центроидного плана, имеет вид

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 + \beta_{134} x_1 x_3 x_4 + \beta_{234} x_2 x_3 x_4 + \beta_{1234} x_1 x_2 x_3 x_4.$$

№ опыта	x_1	x_2	x_3	x_4	y	№ опыта	x_1	x_2	x_3	x_4	y
2	0	1	0	0	y_2	10	0	0	1/2	1/2	y_{34}
3	0	0	1	0	y_3	11	1/3	1/3	1/3	0	y_{123}
4	0	0	0	1	y_4	12	1/3	1/3	0	1/3	y_{124}
5	1/2	1/2	0	0	y_{12}	13	1/3	0	1/3	1/3	y_{134}
6	1/2	0	1/2	0	y_{13}	14	0	1/3	1/3	1/3	y_{234}
7	1/2	0	0	1/2	y_{14}	15	1/4	1/4	1/4	1/4	y_{1234}
8	0	1/2	1/2	0	y_{23}						

D-оптимальные планы. Среди различных критериев оптимальности, как отмечалось, важнейшими являются критерии *D*- и *G*-оптимальности. Как известно, *D*-оптимальный план обеспечивает для заданного числа опытов наилучшую точность оценок коэффициентов модели. Свойство *G*-оптимальности плана обеспечивает минимум максимальной дисперсии предсказанных значений рассматриваемого свойства в области исследования. Симплекс-решетчатые планы обладают свойствами *D*- и *G*-оптимальности только при построении полиномов второго и неполного третьего порядка. Планы Шеффе более высокого порядка не являются *D*-оптимальными; *D*-оптимальная симплексная решетка для полинома третьего порядка была построена Ж. Кифером.

Для множества планов с координатами точек $x_j = 1$, $x_l = x_s = 0$; $x_j = 1 - x_l = c$, $x_s = 0$, $c < 0,5$; $x_j = x_l = x_s = 1/3$ при построении полинома третьего порядка план будет *D*-оптимальным для $c = (1 - \sqrt{5})/2$, т. е. точки на сторонах симплекса должны иметь координаты $x_j = 0,2764$, $x_l = 0,7236$. Число точек плана, так же как и для решетки Шеффе, $c_{k+\mu-1}^{\mu}$. *D*-оптимальный план третьего порядка для трехкомпонентной смеси приведен в табл. 4.21.

При использовании полинома четвертого порядка для трехкомпонентной смеси план будет *D*-оптимальным, если $x_j = (7 - \sqrt{21})/14$; $x_l = 1 - x_j$; $x_s = 0$ или $x_j = 0,1727$, $x_l = 0,8273$; $x_s = 0$. Кроме того, в *D*-оптимальном плане четвертого порядка имеются точки с координатами $x_j = x_l = (7 - \sqrt{5})/22$; $x_s = 1 - (x_j + x_l)$ или $x_j = x_l = 0,2165$; $x_s = 0,567$.

D-оптимальный план четвертого порядка для трехкомпонентной смеси приведен в табл. 4.22. *D*-оптимальные пла-

Таблица 4.21

№ опыта	x_1	x_2	x_3	y	№ опыта	x_1	x_2	x_3	y
1	1	0	0	y_1	6	0,7236	0	0,2764	y_{113}
2	0	1	0	y_2	7	0,2764	0	0,7236	y_{133}
3	0	0	1	y_3	8	0,	0,7236	0,2764	y_{223}
4	0,7236	0,2764	0	y_{112}	9	0,	0,2764	0,7236	y_{233}
5	0,2764	0,7236	0	y_{122}	10	0,333	0,3333	0,3333	y_{123}

Таблица 4.22

№ опыта	x_1	x_2	x_3	y	№ опыта	x_1	x_2	x_3	y
1	1	0	0	y_1	9	0,8273	0	0,1727	y_{1113}
2	0	1	0	y_2	10	0,1727	0	0,8273	y_{1333}
3	0	0	1	y_3	11	0	0,8273	0,1727	y_{2223}
4	0,5	0,5	0	y_{12}	12	0	0,1727	0,8273	y_{2333}
5	0,5	0	0,5	y_{13}	13	0,567	0,2165	0,2165	y_{1112}
6	0	0,5	0,5	y_{23}	14	0,2165	0,567	0,2165	y_{1223}
7	0,8273	0,1727	0	y_{1112}	15	0,2165	0,2165	0,567	y_{1233}
8	0,1727	0,8273	0	y_{1222}					

ны различных степеней для трехкомпонентных смесей показаны на рис. 4.14, а—г. Как видно из рисунка, здесь так же, как и для симплекс-решетчатых планов, возможна композиционная доработка плана от второй степени до четвертой, для перехода от третьей степени к четвертой необходимо проделать опыты заново. Для описания свойств смеси для D -оптимальных, как и для симплекс-решетчатых планов, используются приведенные полиномы соответствующих степеней.

Планирование эксперимента на локальных участках исследуемой области. При исследовании локальных участков факторного пространства выбор плана эксперимента определяется прежде всего формой заданной области эксперимента. Здесь возможны следующие варианты: 1) исследуемая область — симплекс или может быть сведена к симплексу (число вершин $NM = k$); 2) исследуемая область — многогранник (число вершин $NM > k$); 3) исследуемая область — куб (гиперкуб) или параллелепипед (число вершин $NM = 2^{k-1}$).

Рассмотрим последовательно все варианты.

Если наложенные на концентрации компонентов ограничения таковы (например, односторонние), что число вершин полученного многогранника NM совпадает с числом

компонентов смеси k , то многогранник является симплексом (хотя и неправильным) и можно использовать описанные планы на симплексе. Это возможно также и в тех случаях, когда внутри полученного многогранника ($NM > k$) удастся выделить подобласть в виде симплекса, например путем выбора из NM вершин многогранника k вершин локального симплекса на основе априорной информации или интуиции разработчика. Для перевода плана, построенного

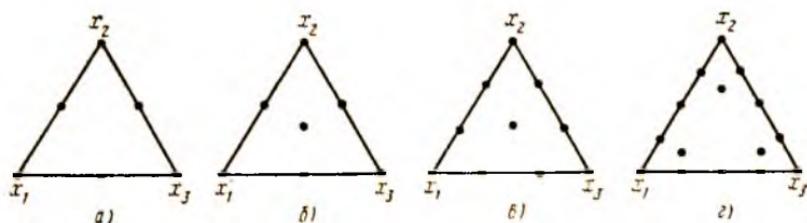


Рис 4.14

на полном концентрационном симплексе, на локальный участок вершины локального симплекса A_j , $j = \overline{1, k}$ с координатами $A_1 (x_1^{(1)}, x_2^{(1)}, \dots, x_k^{(1)})$; $A_2 (x_1^{(2)}, x_2^{(2)}, \dots, x_k^{(2)})$; ...; $A_k (x_1^{(k)}, x_2^{(k)}, \dots, x_k^{(k)})$ принимают за самостоятельные псевдокомпоненты z_i : $0 \leq z_j \leq 1$; $\sum_{j=1}^n z_j = 1$.

Относительно новых переменных z_1, z_2, \dots, z_k могут быть построены все ранее описанные планы. Для любой n -й точки построенного плана переход от псевдокоординат z_j к исходным координатам x_j осуществляют по формуле

$$x_j^{(n)} = x_j^{(1)} + z_2^{(n)} (x_j^{(2)} - x_j^{(1)}) + z_3^{(n)} (x_j^{(3)} - x_j^{(1)}) + \dots \\ \dots + z_k^{(n)} (x_j^{(k)} - x_j^{(1)}),$$

где x_j — содержание j -го компонента в вершине $z_j (A_j)$.

Примеры построения плана в псевдокоординатах и перевода его на локальный симплекс, т. е. в первоначальные координаты, приведены в табл. 4.23. Изображение локальной области приведено на рис. 4.15. Исследованию в данной трехкомпонентной смеси подвергался не весь концентрационный треугольник, а лишь подобласть в виде треугольника с вершинами: $z_1 (0,86; 0; 0,14)$; $z_2 (0; 0,78; 0,22)$; $z_3 (0,78; 0; 0,22)$.

В общем случае наложенные на концентрации компонентов двусторонние ограничения высекают в полном кон-

центрационном симплексе многогранник с числом вершин $NM > k$. Определение числа и координат вершин локальной области по заданным ограничениям производят при постановке задачи (задание области исследования). План непосредственно на многограннике, предложенный Р. Мак Лином и В. Андерсеном, включает кроме вершин многогранника центры двух-, трех-, ..., $(k - 1)$ -мерных граней многогранника и его центр.

Координаты центров граней находят путем усреднения координат точек, образующих грань, координаты центра — путем усреднения соответствующих координат всех вершин многогранника. Пример построения плана на многограннике для четырехкомпонентной смеси представлен в табл. 4.24 и 4.25, где исследуется смесь со следующими ограничениями на изменения концентраций компонентов:

$0,4 \leq x_1 \leq 0,6$; $0,1 \leq x_2 \leq 0,5$; $0,1 \leq x_3 \leq 0,5$; $0,03 \leq x_4 \leq 0,08$. Данные ограничения образуют многогранник с восемью вершинами, координаты которых представлены в табл. 4.24, а центры граней — в табл. 4.25.

В план на многограннике Мак Лина и Андерсена входят кроме вершин центры граней. Координаты центров граней (с указанием номеров вершин, образующих грань), а также центра полученного многогранника приведены в табл. 4.26. Таким образом, полученный план (табл. 4.24 и 4.25) содержит 15 опытов.

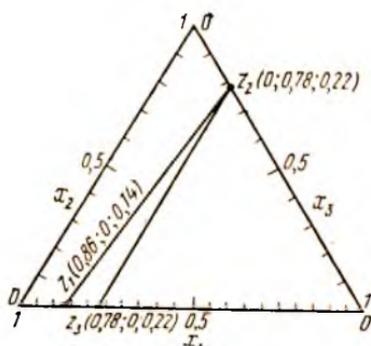


Рис. 4.15

Таблица 4.23

№ опыта	z_1	z_2	z_3	x_1	x_2	x_3
1	1	0	0	0,86	0	0,14
2	0	1	0	0	0,78	0,22
3	0	0	1	0,78	0	0,22
4	0,5	0,5	0	0,43	0,39	0,18
5	0,5	0	0,5	0,82	0	0,18
6	0	0,5	0,5	0,39	0,39	0,22
7	0,333	0,333	0,333	0,546	0,259	0,193

Таблица 4.24

№ опыта	x_1	x_2	x_3	x_4
1	0,4	0,1	0,47	0,03
2	0,4	0,1	0,42	0,08
3	0,6	0,1	0,27	0,03
4	0,6	0,1	0,22	0,08
5	0,4	0,47	0,1	0,03
6	0,4	0,42	0,1	0,08
7	0,6	0,27	0,1	0,03
8	0,6	0,22	0,1	0,08

Таблица 4.25

№ опыта	x_1	x_2	x_3	x_4	№ вершин, образующих грань
9	0,5	0,1	0,345	0,055	1, 2, 3, 4
10	0,5	0,345	0,1	0,055	5, 6, 7, 8
11	0,4	0,2725	0,2725	0,055	1, 2, 5, 6
12	0,6	0,1725	0,1725	0,055	3, 4, 7, 8
13	0,5	0,235	0,235	0,03	1, 3, 5, 7
14	0,5	0,21	0,21	0,08	2, 4, 6, 8
15	0,5	0,2225	0,2225	0,05	Центр многогранника

Таблица 4.26

№ опыта	x_1	x_2	x_3	x_4
1	0,05	0,1	0,2	0,65
2	0,1	0,1	0,2	0,6
3	0,05	0,2	0,2	0,55
4	0,1	0,2	0,2	0,5
5	0,05	0,1	0,4	0,45
6	0,1	0,1	0,4	0,4
7	0,05	0,2	0,4	0,35
8	0,1	0,2	0,4	0,3

Для описания результатов эксперимента, проведенного по плану Мак Лина и Андерсена, используют приведенный полином, включающий кроме линейных членов все возможные слагаемые взаимодействия k компонентов от парных до k -х включительно [всего $(2^k - 1)$ член]. С ростом числа компонентов смеси число комбинаций условий эксперимента в плане быстро растет и становится значительно больше

числа коэффициентов модели. Для сокращения числа опытов плана можно исключить центры некоторых граней (например, из соображения равномерности распределения оставшихся точек) или производить с помощью ЭВМ выбор точек плана и множества точек для выбранной модели в соответствии с каким-либо критерием оптимальности плана (например, D -оптимальности).

При исследовании локальных участков факторного пространства эксперимента иногда возможен случай, когда ограничения налагают не на все, а лишь на $(k - 1)$ компонент смеси, один из компонентов (обычно это основа смеси) можно рассматривать как достройку системы до 1 или 100%. В этом случае число вершин многогранника равно 2^{k-1} и задача может решаться как в симплексной (барицентрической), так и в независимой (ортогональной) системе координат размерности $k - 1$, а многогранник преобразуется в куб или параллелепипед. При этом необходимо, чтобы сумма верхних пределов изменения концентраций компонентов, на которые наложены ограничения, не должна превышать 1 или 100%.

Рассмотрим пример такой задачи. Пусть имеется четырехкомпонентная смесь со следующими ограничениями на изменения концентрации компонентов: $0,05 \leq x_1 \leq 0,1$; $0,1 \leq x_2 \leq 0,2$; $0,2 \leq x_3 \leq 0,4$; $0 \leq x_4 \leq 1$. Компонент x_4 является основой смеси и рассматривается как достройка смеси до 1 (имеет ограничения от 0 до 1). Решаем эту задачу сначала в симплексной системе координат. Построив вершины многогранника, получим 8 вершин (табл. 4.26), центры граней многогранника приведены в табл. 4.27. Поскольку на четвертый компонент не наложены ограничения и верхние границы изменений концентраций компонентов x_1 , x_2 и x_3 образуют сумму $0,1 + 0,2 + 0,4 = 0,7 < 1$, полученный многогранник может быть преобразован в куб. Для этого кодируют нижнюю и верхнюю границы первых трех компонентов через -1 и $+1$ соответственно (как это принято для задач с независимыми переменными) и выбирают из каталога план второго порядка для трехпеременных типа B , наиболее близких к D -оптимальному и достаточно экономичный по числу опытов. Выбранный план представлен в левой части табл. 4.28. Перевод плана в натуральные единицы (концентрации компонентов) приведен в правой части табл. 4.28. Содержание компонента x_4 рассчитано как добавка концентраций x_1 , x_2 и x_3 до 1 или 100%. Как видно из сравнения правой части табл. 4.28 с табл. 4.26 и 4.27, планы, построенные двумя способами, совпадают.

Построение математической модели. После проведения эксперимента и получения значений отклика для каждого заданного планом состава смеси производят вычисление коэффициентов соответствующего плану полинома. При этом коэффициенты рассчитывают в зависимости от вида плана,

Таблица 4.27

№ опыта	x_1	x_2	x_3	x_4	№ вершин, образующих грань
9	0,075	0,15	0,2	0,575	1, 2, 3, 4
10	0,075	0,15	0,4	0,375	5, 6, 7, 8
11	0,075	0,1	0,3	0,525	1, 2, 5, 6
12	0,075	0,2	0,3	0,425	3, 4, 7, 8
13	0,05	0,05	0,3	0,5	1, 3, 5, 7
14	0,1	0,15	0,3	0,45	2, 4, 6, 8
15	0,075	0,15	0,3	0,475	Центр многогранника

Таблица 4.28

№ опыта	z_1	z_2	z_3	x_1	x_2	x_3	x_4
1	-1	-1	-1	0,05	0,1	0,2	0,65
2	+1	-1	-1	0,1	0,1	0,2	0,6
3	-1	+1	-1	0,05	0,2	0,2	0,55
4	+1	+1	-1	0,1	0,2	0,2	0,5
5	-1	-1	+1	0,05	0,1	0,4	0,45
6	+1	-1	+1	0,1	0,1	0,4	0,4
7	-1	+1	+1	0,05	0,2	0,4	0,35
8	+1	+1	+1	0,1	0,2	0,4	0,30
9	-1	0	0	0,05	0,15	0,3	0,45
10	+1	0	0	0,1	0,15	0,3	0,45
11	0	-1	0	0,075	0,1	0,3	0,525
12	0	+1	0	0,075	0,2	0,3	0,425
13	0	0	-1	0,075	0,15	0,2	0,575
14	0	0	+1	0,075	0,15	0,4	0,375
15	0	0	0	0,075	0,15	0,3	0,475

т. е. строится план на полном симплексе или для исследования локального участка.

Для планов на полном симплексе используют свойство насыщенности плана, что существенно упрощает задачу. Рассмотрим расчет коэффициентов полинома на примере приведенной модели второго порядка для трех компонентов:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3.$$

Если последовательно подставлять в уравнение координаты всех шести точек симплекс-решетчатого плана эксперимента, представленного в табл. 4.19, то при подстановке координат первой точки ($x_1 = 1, x_2 = 0, x_3 = 0$) получим $y_1 = \beta_1$ и соответственно $\beta_2 = y_2, \beta_3 = y_3$.

При подстановке в уравнение координат четвертой точки найдем

$$y = 0,5\beta_1 + 0,5\beta_2 + 0,25 \beta_{12}.$$

Так как $\beta_j = y_j$, то $y_{12} = 0,5 y_1 + 0,5 y_2 + 0,25 \beta_{12}$. Отсюда $\beta_{12} = 4y_{12} - 2y_1 - 2y_2$. Соответственно $\beta_{13} = 4y_{13} - 2y_1 - 2y_3$; $\beta_{23} = 4y_{23} - 2y_2 - 2y_3$. Три точки, определяющие коэффициент β_{jl} , лежат на одном ребре. Аналогично определяют коэффициенты полинома второго порядка $\widehat{y} = \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l$ для k -компонентной смеси: $\beta_j = y_j$; $\beta_{jl} = 4y_{jl} - 2y_j - 2y_l$. Также выводят соотношения для расчета коэффициентов любого приведенного полинома при любом числе компонентов. Так, для полинома неполного третьего порядка для трехкомпонентной смеси

$$\begin{aligned} \widehat{y} = & \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \\ & + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3; \\ \beta_j = & y_j; \quad \beta_{12} = 4y_{12} - 2y_1 - 2y_2 \text{ и т. д.}, \end{aligned}$$

для k -компонентной смеси

$$\begin{aligned} \widehat{y} = & \sum_{1 \leq j \leq k} \beta_j x_j + \sum_{1 \leq j < l \leq k} \beta_{jl} x_j x_l + \sum_{1 \leq j < l < s \leq k} \beta_{jls} x_j x_l x_s; \\ \beta_j = & y_j; \\ \beta_{jl} = & 4y_{jl} - 2y_j - 2y_l; \\ \beta_{jls} = & 27y_{jls} - 12(y_{jl} + y_{ls} + y_{js}) + 3(y_j + y_l + y_s). \end{aligned}$$

Для других планов на полном симплексе (центрондных, D -оптимальных и т. п.) для определения коэффициентов полинома производят аналогичную процедуру последовательной подстановки в уравнение регрессии координат экспериментальных точек. Для коэффициентов полинома, полученного по симплекс-центрондному плану, существует следующая общая формула:

$$\beta_{j_l \dots} = r \sum_{t=1}^r (-1)^{r-t} t^{r-1} SS_t,$$

где r — число индексов коэффициента $\beta_{j_l \dots}$; SS_t — сумма результатов опытов всех смесей из t компонентов, взятых в равных пропорциях ($1/t$). Например, для коэффициента β_{jls} ($r = 3$) (j, l, s)

$$y_j + y_l + y_s = SS_1 \text{ при } 1/t = 1;$$

$$y_{jl} + y_{ls} + y_{js} = SS_2 \text{ при } 1/t = 1/2;$$

$$y_{jls} = SS_3 \text{ при } 1/t = 1/3.$$

Таким образом,

$$\begin{aligned} \beta_{jls} = & 3 [(-1)^{3-1} 1^{3-1} SS_1 + (-1)^{3-2} 2^{3-1} SS_2 + \\ & + (-1)^{3-3} 3^{3-1} SS_3] = 3 [(y_j + y_l + y_s) - \\ & - 4(y_{jl} + y_{js} + y_{ls}) + 9y_{jls}]. \end{aligned}$$

Для D -оптимального плана четвертого порядка в трехкомпонентной системе по плану, представленному в табл. 4.22, уравнение регрессии имеет вид

$$\begin{aligned} \widehat{y} = & \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \\ & + \beta_{23} x_2 x_3 + v_{12} x_1 x_2 (x_1 - x_2) + v_{13} x_1 x_3 (x_1 - x_3) + \\ & + v_{23} x_2 x_3 (x_2 - x_3) + \delta_{12} x_1 x_2 (x_1 - x_2)^2 + \\ & + \delta_{13} x_1 x_3 (x_1 - x_3)^2 + \delta_{23} x_2 x_3 (x_2 - x_3)^2 + \beta_{1123} x_1^2 x_2 x_3 + \\ & + \beta_{1223} x_1 x_2^2 x_3 + \beta_{1233} x_1 x_2 x_3^2. \end{aligned}$$

Коэффициенты полинома четвертого порядка, полученные подстановкой координат экспериментальных точек в уравнение регрессии, имеют следующий вид:

$$\beta_j = y_j;$$

$$\beta_{jl} = 4y_{jl} - 2y_j - 2y_l,$$

$$v_{jl} = \frac{7}{6} \{ 3(-y_j + y_l) + \sqrt{21} (y_{jjjl} - y_{llll}) \};$$

$$\delta_{jl} = \frac{7}{6} \{ -3(y_j + y_l) - 8y_{jl} + 7(y_{jjjl} + y_{llll}) \};$$

$$\begin{aligned} \beta_{jjls} = & 26,657y_j - 6,167(y_l + y_s) - 16,96(y_{jll} + y_{jls}) + \\ & + 0,511y_{jls} - 32,18(y_{jjjl} + y_{jjis}) + 17,196(y_{jlll} + y_{jsss}) + \\ & + 5,72(y_{llls} + y_{lss}) + 84,11y_{jjls} - 23,237(y_{jlls} + y_{jls}). \end{aligned}$$

где $j \neq l \neq s$; $j, l, s = 1, 2, 3$.

Если исследуется локальная область факторного пространства в виде симплекса, то после реализации выбранного плана эксперимента рассчитывают коэффициенты урав-

нения регрессии $\widehat{y} = f(z_1, z_2, \dots, z_k)$ в координатах псевдокомпонентов, используя приведенные формулы для соответствующих планов, полученные подстановкой в уравнение регрессии координат точек плана.

После проверки адекватности полученной модели для более удобного ее практического использования модель записывают в исходной системе координат с помощью формул перевода координат из одной системы в другую:

$$z_1 = z_1^{(1)} + x_2(z_1^{(2)} - z_1^{(1)}) + x_3(z_1^{(3)} - z_1^{(1)}) + \dots + x_k(z_1^{(k)} - z_1^{(1)});$$

$$z_2 = z_2^{(1)} + x_2(z_2^{(2)} - z_2^{(1)}) + x_3(z_2^{(3)} - z_2^{(1)}) + \dots + x_k(z_2^{(k)} - z_2^{(1)});$$

.....

$$z_{k-1} = z_{k-1}^{(1)} + x_2(z_{k-1}^{(2)} - z_{k-1}^{(1)}) + x_3(z_{k-1}^{(3)} - z_{k-1}^{(1)}) + \dots + x_k(z_{k-1}^{(k)} - z_{k-1}^{(1)}).$$

Значения $z_j^{(l)}$ находят при решении $(k-1)$ систем уравнений:

$$x_1^{(1)} z_1^{(1)} + x_2^{(1)} z_1^{(2)} + x_3^{(1)} z_1^{(3)} + \dots + x_k^{(1)} z_1^{(k)} = 1;$$

$$x_1^{(2)} z_1^{(1)} + x_2^{(2)} z_1^{(2)} + x_3^{(2)} z_1^{(3)} + \dots + x_k^{(2)} z_1^{(k)} = 0;$$

.....

$$x_1^{(k)} z_1^{(1)} + x_2^{(k)} z_1^{(2)} + x_3^{(k)} z_1^{(3)} + \dots + x_k^{(k)} z_1^{(k)} = 0;$$

$$x_1^{(1)} z_2^{(1)} + x_2^{(1)} z_2^{(2)} + x_3^{(1)} z_2^{(3)} + \dots + x_k^{(1)} z_2^{(k)} = 0;$$

$$x_1^{(2)} z_2^{(1)} + x_2^{(2)} z_2^{(2)} + x_3^{(2)} z_2^{(3)} + \dots + x_k^{(2)} z_2^{(k)} = 1;$$

.....

$$x_1^{(k)} z_2^{(1)} + x_2^{(k)} z_2^{(2)} + x_3^{(k)} z_2^{(3)} + \dots + x_k^{(k)} z_2^{(k)} = 0;$$

$$x_1^{(1)} z_{k-1}^{(1)} + x_2^{(1)} z_{k-1}^{(2)} + x_3^{(1)} z_{k-1}^{(3)} + \dots + x_k^{(1)} z_{k-1}^{(k)} = 0;$$

$$x_1^{(2)} z_{k-1}^{(1)} + x_2^{(2)} z_{k-1}^{(2)} + x_3^{(2)} z_{k-1}^{(3)} + \dots + x_k^{(2)} z_{k-1}^{(k)} = 0;$$

.....

$$x_1^{(k)} z_{k-1}^{(1)} + x_2^{(k)} z_{k-1}^{(2)} + x_3^{(k)} z_{k-1}^{(3)} + \dots + x_k^{(k)} z_{k-1}^{(k)} = 1,$$

где $z_j^{(l)}$ — содержание псевдокомпонента z_j в вершинах исходного симплекса; $x_j^{(l)}$ — содержание j -го компонента в вершинах $z_l(A_l)$; $l = 1, 2, \dots, k$.

Поскольку такой перевод координат возможен только для уравнений с независимыми переменными, исходное уравнение регрессии необходимо преобразовать, исключив одну переменную, например последнюю, k -ю: $z_k = 1 - \sum_{j=1}^{k-1} z_j$.

Для построения математической модели по результатам экспериментов, проведенных по плану на локальном многограннике в связи с тем, что эти планы обычно ненасыщенные (число опытов больше числа коэффициентов модели), используют метод наименьших квадратов, модель строят непосредственно в координатах исходных компонентов. Алгоритм метода наименьших квадратов реализуют в виде программы на ЭВМ.

Проверка адекватности математической модели. После определения коэффициентов уравнения регрессии необходимо провести статистический анализ полученных результатов: проверить адекватность уравнения и построить доверительные интервалы значений отклика, предсказываемых по уравнению регрессии.

При постановке эксперимента по планам на симплексах (решетчатый, центроидный, D -оптимальный и др.) из-за их насыщенности степени свободы, необходимые для проверки адекватности уравнения, отсутствуют. Поэтому для проверки адекватности уравнения ставят опыты в дополнительных, так называемых контрольных точках. Эти точки обычно выбирают так, чтобы предусмотреть возможность их использования для построения модели более высокого порядка, если полученное уравнение окажется неадекватным.

Точность предсказаний отклика в различных точках симплекса различна и зависит от координат точек. Дисперсию предсказанного значения отклика определяют по закону накопления ошибок. Рассмотрим расчет дисперсии предсказанных значений отклика на примере модели второго порядка для трехкомпонентной смеси:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3.$$

Предположим, что значения откликов являются результатом усреднения n_j и n_{jl} параллельных опытов в соответствующих точках симплекса, а дисперсия воспроизводимости опыта s_y^2 во всех точках плана одинакова. Так как дисперсия среднего в n раз меньше дисперсии отдельных наблюдений экспериментальных значений откликов, то

$$s_{y_j}^2 = s_y^2/n_j; \quad s_{y_{jl}}^2 = s_y^2/n_{jl}.$$

Заменяя в приведенном полиноме второго порядка коэффициенты их выражениями через отклики ($\beta_j = \bar{y}_j$; $\beta_{il} = 4\bar{y}_{jl} - 2\bar{y}_j - 2\bar{y}_l$), получим

$$\begin{aligned} \widehat{y} &= \bar{y}_1 x_1 + \bar{y}_2 x_2 + \bar{y}_3 x_3 + (4\bar{y}_{12} - 2\bar{y}_1 - 2\bar{y}_2) x_1 x_2 + \\ &+ (4\bar{y}_{13} - 2\bar{y}_1 - 2\bar{y}_3) x_1 x_3 + (4\bar{y}_{23} - 2\bar{y}_2 - 2\bar{y}_3) x_2 x_3 = \\ &= \bar{y}_1 (x_1 - 2x_1 x_2 - 2x_1 x_3) + \bar{y}_2 (x_2 - 2x_1 x_2 - 2x_2 x_3) + \\ &+ \bar{y}_3 (x_3 - 2x_1 x_3 - 2x_2 x_3) + 4\bar{y}_{12} x_1 x_2 + 4\bar{y}_{13} x_1 x_3 + \\ &+ 4\bar{y}_{23} x_2 x_3. \end{aligned}$$

По условию $x_1 + x_2 + x_3 = 1$, преобразуем коэффициенты при \bar{y}_j :

$$\begin{aligned} x_1 - 2x_1 x_2 - 2x_1 x_3 &= x_1 - 2x_1 (x_2 + x_3) = x_1 - 2x_1 (1 - x_1) = \\ &= x_1 (2x_1 - 1) \text{ и т. д.} \end{aligned}$$

В результате найдем

$$\begin{aligned} \widehat{y} &= x_1 (2x_1 - 1) \bar{y}_1 + x_2 (2x_2 - 1) \bar{y}_2 + x_3 (2x_3 - 1) \bar{y}_3 + \\ &+ 4x_1 x_2 \bar{y}_{12} + 4x_1 x_3 \bar{y}_{13} + 4x_2 x_3 \bar{y}_{23}. \end{aligned}$$

Введя обозначения $a_j = x_j (2x_j - 1)$; $a_{jl} = 4x_j x_l$ с учетом приведенных выражений для $s_{y_j}^2$ и $s_{y_{jl}}^2$, получим выражение для дисперсии уравнения регрессии $s_{\widehat{y}}^2$:

$$s_{\widehat{y}}^2 = s_y^2 \left(\sum_{1 \leq j < k} \frac{a_j^2}{n_j} + \sum_{1 \leq j < l < k} \frac{a_{jl}^2}{n_{jl}} \right).$$

Аналогично могут быть определены выражения для дисперсий полиномов различных степеней. Так, для модели неполного третьего порядка выражение для оценки дисперсии уравнения регрессии принимает вид

$$s_{\widehat{y}}^2 = s_y^2 \left(\sum_{1 \leq j < k} \frac{b_j^2}{n_j} + \sum_{1 \leq j < l < k} \frac{b_{jl}^2}{n_{jl}} + \sum_{1 \leq j < l < s < k} \frac{b_{jls}^2}{n_{jls}} \right).$$

где $b_j = \frac{1}{2} x_j (6x_j^2 - 2x_j + 1) - 3 \sum_{t=1}^n x_j^2$;

$$b_{jl} = 4x_j x_l (3x_j + 3x_l - 2); \quad b_{jls} = 27x_j x_l x_s.$$

Часто число параллельных опытов во всех точках плана принимают одинаковым и равным n , т. е. $n_j = n_{jl} = n$. Тогда формулу для оценки дисперсии уравнения регрессии можно записать так:

$$s_{\hat{y}}^2 = s_y^2 \frac{\xi}{n},$$

где s_y^2 — оценка дисперсии воспроизводимости опыта; ξ — параметр, зависящий от степени полинома. Для полинома второго порядка

$$\xi = \sum_{1 \leq j \leq k} a_j^2 + \sum_{1 \leq j < l \leq k} a_{jl}^2,$$

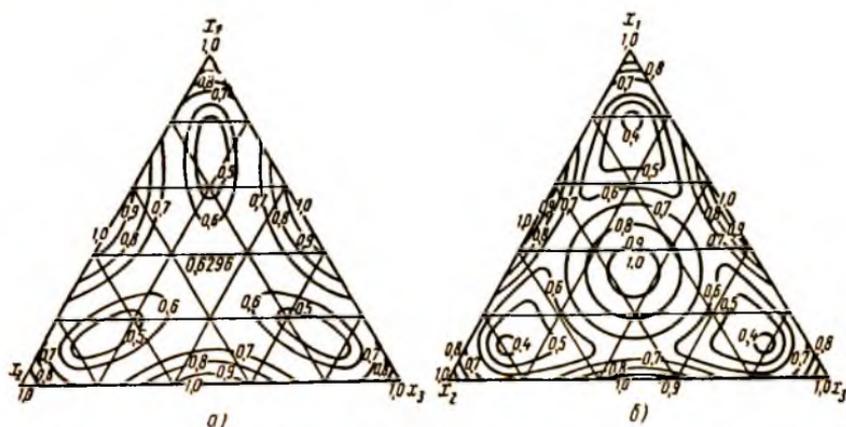


Рис. 4.16

для полинома неполного третьего порядка

$$\xi = \sum_{1 \leq j \leq k} b_j^2 + \sum_{1 \leq j < l \leq k} b_{jl}^2 + \sum_{1 \leq j < l < s \leq k} b_{jls}^2.$$

Так как величина ξ зависит только от состава смесей, то для трехкомпонентных смесей можно заранее построить линии равного значения ξ для полиномов различных степеней. Изолинии для полиномов второго и неполного третьего порядка представлены на рис. 4.16, а, б.

Зная s_y^2 и n , легко найти оценку дисперсии предсказанных значений отклика в любой точке диаграммы состав—свойство, воспользовавшись соответствующим значением ξ , полученным по графику изолиний. Проверку адекват-

ности проводят в каждой контрольной точке. Для этого составляют отношение

$$t = \frac{\Delta y}{\sqrt{s_y^2 + r s_y^2}} = \frac{\Delta y \sqrt{n}}{s_y \sqrt{1 + \xi}}$$

где $\Delta y = \bar{y}_{\text{эксп}} - y_{\text{расч}}$.

Значение t , распределенное по закону Стьюдента, сравнивают с табличным значением t , выбранным для уровней значимости $\alpha/2m$ (m — число контрольных точек) и числа степеней свободы оценки дисперсии воспроизводимости опыта.

Гипотеза об адекватности уравнения принимается, если экспериментальное значение t меньше табличного ($t_{\text{эксп}} < t_{\text{табл}}$) для всех контрольных точек.

Построение доверительных интервалов для значений отклика в различных точках концентрационного симплекса производят следующим образом:

$$\hat{y} - \Delta \leq y \leq \hat{y} + \Delta; \Delta = t_{\alpha/k, f} s_{\hat{y}}$$

где $t_{\alpha/k, f}$ — табличное значение t -критерия Стьюдента для уровня значимости α/k (k — число коэффициентов регрессии) и числа степеней свободы оценки дисперсии воспроизводимости опыта f . При одинаковом числе параллельных опытов в каждой точке для трехкомпонентных диаграмм для определения y можно воспользоваться картами для изолиний:

$$\Delta = t_{\alpha/k, f} \frac{s_y}{\sqrt{n}} \xi^{0.5}$$

Построение диаграмм состав—свойство и поиск оптимальных составов смеси. Полученные уравнения регрессии, связывающие отклики с составами смесей, позволяют рассчитать параметры смеси любого состава и построить диаграммы состав—свойство, представляющие собой изолинии равных параметров, изображенные на концентрационном симплексе или его сечениях для многокомпонентных смесей.

Для построения диаграмм и поиска оптимальных составов смеси с экстремальным параметром можно использовать метод сеток. При этом по уравнению регрессии рассчитывают значения параметра во всех точках концентрационной сетки, построенной с заданным шагом на исследуемой области.

Для уравнений регрессии второго порядка при построении диаграмм и поиске оптимальных составов удобно использовать канонический вид уравнения, позволяющий определить вид поверхности отклика и достаточно просто построить их изолинии, а также найти координаты экстремальных точек. Эти вопросы подробно рассматриваются в курсе аналитической геометрии. При решении сложных задач оптимизации составов многокомпонентных смесей, когда имеется несколько откликов, для одного из них (наиболее важного) находится максимум или минимум. Другие целевые функции являются односторонними («не более» или «не менее») или являются двусторонними («от» и «до») ограничениями. В этом случае используют различные алгоритмы поиска экстремума (рассмотрены ранее), реализованные на ЭВМ. Для построения диаграмм состав—свойство используют программы вывода графической информации на АЦПУ.

ГЛАВА 5. ОПТИМИЗАЦИЯ ПРОИЗВОДСТВЕННЫХ ИСПЫТАНИЙ

§ 5.1. ПРОИЗВОДСТВЕННЫЕ ИСПЫТАНИЯ КАК ЭТАП ПРОИЗВОДСТВА

При оптимизации качества изделий наибольшего эффекта достигают в том случае, когда ее осуществляют на стадии: а) проектирования (оптимизация конструкторских и проектных решений); б) производства (оптимизация ТП); в) производственных испытаний (оптимизация длительности и объема испытаний).

Оптимизация производственных испытаний или минимизация их продолжительности влияет на все этапы создания приборов. Во-первых, одним из основных требований к созданию новых приборов в современных условиях является сокращение их сроков разработки и освоения. Долговечность приборов всех классов постоянно повышается и информация о сроках их службы необходима разработчику, определение же долговечности занимает значительный процент времени от срока разработки ЭП. Следовательно, сокращение производственных испытаний позволяет уменьшить срок разработки ЭП. Во-вторых, для успешного функционирования систем автоматизированного проектирования приборов (САПР) и автоматизированного управления технологическим процессом (АСУ ТП), а также для создания гибких автоматизированных производств (ГАП) необходи-

ма оперативная информация о качестве приборов, которую наиболее полно можно получить по результатам испытаний. Уменьшение времени испытаний повышает оперативность получения данных о состоянии приборов и результатах разработки, тем самым повышая эффективность функционирования САПР и АСУ ТП, а также всего ГАП. В-третьих, при сокращении производственных испытаний строятся

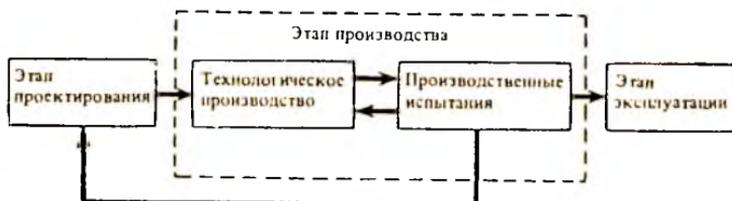


Рис. 5.1

математические модели испытаний — зависимость работоспособности приборов от времени и воздействующих факторов. Подобные модели оказывают существенную помощь разработчику ЭП. В-четвертых, минимизация продолжительности испытаний позволяет получить экономический эффект. Так как долговечность современных приборов составляет от нескольких тысяч до десятков тысяч часов, то сокращение длительности испытаний в несколько раз приводит к значительной экономии различного вида затрат.

Очевидно, что производственные испытания являются тем этапом создания приборов, когда впервые можно получить информацию о работоспособности готового прибора, об изменении его параметров во времени, старении, показателях надежности и долговечности. Чем быстрее будет получена эта информация, тем оперативнее и эффективнее будет функционировать вся система создания прибора (рис. 5.1).

В процессе производственных испытаний нужно получить разнообразную информацию о зависимости параметров прибора от времени и различных воздействующих факторов: изменение работоспособности прибора во времени, время отказа работы прибора, его реакция на различные внешние воздействия (механические, климатические и т.п.). В первом случае оценивают показатели надежности путем проведения испытаний на безотказность работы и долговечность, а также на гамма-процентный ресурс и сохраняемость. Этот вид испытаний связан прежде всего с необходимостью полу-

чения отказов. Если учесть, что долговечность приборов сейчас велика, то для получения информации об общем числе отказов в партии требуется продолжительное время. Во втором случае оценивают влияние механических, климатических и биологических факторов, а также воздействие специальных сред на работоспособность приборов и их соответствие заданным требованиям. Этот вид испытаний не занимает продолжительного времени, однако для оценки влияния каждого воздействующего фактора требуется одна или несколько партий приборов.

Испытания на безотказность и долговечность проводят на этапах разработки изделия и приемки установочной серии. При серийном производстве изделия испытывают на безотказность, долговечность, на гамма-процентный ресурс и на сохраняемость. Испытания на безотказность проводят с целью периодического контроля качества изделий и проверки стабильности ТП. Продолжительность испытаний устанавливают следующим образом: для изделий с минимальной наработкой менее 500 ч продолжительность устанавливают равной минимальной наработке, для изделий с минимальной наработкой от 500 до 25 000 ч — равной 500 ч, для изделий с минимальной наработкой свыше 25 000 ч — равной 1000 ч. Испытания проводят по планам одноступенчатого контроля, установленным в стандартах, при приемочном числе $C = 0$ (т. е. партия принимается при отсутствии отказов), а также при номинальных режимах эксплуатации.

Испытания на долговечность проводят с целью подтверждения минимальной наработки изделий, которая установлена в технической документации. Продолжительность испытаний определяют как время минимальной наработки. Результаты испытаний на долговечность считают положительными, если в испытываемой партии не обнаружено ни одного отказа. Если при проведении испытаний выявлен только один отказ, то выпуск изделий не прекращается, а испытания продолжают до их окончания. Если обнаружено больше одного отказа, то выпуск изделий прекращают до выяснения причин отказов.

Испытания на гамма-процентный ресурс проводят с целью определения (уточнения) такого срока службы (ресурса), который достигается с заданной вероятностью гамма-процентов. Испытания на гамма-процентный ресурс, как правило, являются продолжением испытаний изделий на долговечность.

Испытания на сохраняемость проводят с целью проверки соответствия изделий требованиям по сохраняемости,

т. е. способности изделий длительно храниться в тех или иных условиях. Продолжительность испытаний должна быть не менее минимального срока сохраняемости изделия. Изделия считают выдержавшими испытания, если в результате испытаний не было отказа. В конце проводят испытания изделий на безотказность.

Таким образом, цель всех испытаний во времени — оценить время работы приборов без отказов. Как отмечалось, для ЭП, обладающих большой долговечностью, длительность производственных испытаний составляет от нескольких месяцев до нескольких лет. Поэтому для минимизации продолжительности испытаний без потери необходимой информации следует построить математическую модель зависимости показателей работоспособности прибора от времени.

При анализе влияния внешних воздействующих факторов на работоспособность прибора требуется определить, как зависит состояние прибора от этих факторов, вызовут ли они отказ прибора или снижат его работоспособность, сможет ли прибор функционировать при воздействии данного фактора и т. п. Среди подобного вида испытаний наиболее объемными являются механические и климатические (в не-

Таблица 5.1

Вид испытаний	Стоимость испытаний относительно общей стоимости испытаний, %	Вид испытаний	Стоимость испытаний относительно общей стоимости испытаний, %
Обнаружение резонансных частот	2,4	Испытания на: влагоустойчивость	2,7
Испытания на виброустойчивость	3,2	кратковременную	3,6
вibroпрочность	5,0	длительную	
ударную прочность	0,5	воздействие повышенного давления	0,8
случайную вибрацию	5,7	воздействие пониженного давления	0,9
воздействие одиночных ударов	1,7	воздействие солнечной радиации	2,7
воздействие ливневых ускорений	0,7	термоудар	4,5
холодоустойчивость	0,7	воздействие соленого тумана	2,4
термоциклирование	3,6	грибоустойчивость	1,7
		надежность	7,5
		долговечность	49,7

которых случаях проводят радиационные испытания). Их проводят на ограниченном интервале времени при различных уровнях воздействия внешних факторов, после чего принимают решение о способности прибора выдержать эти воздействия. Здесь целесообразно строить многофакторную математическую модель зависимости показателей работоспособности приборов от воздействующих факторов. В табл. 5.1 приведено ориентировочное распределение затрат по видам испытаний ЭП. Нетрудно видеть, что испытания на надежность и долговечность являются наиболее дорогостоящими. Это связано с длительностью испытаний и возрастанием стоимости с увеличением их продолжительности. Сокращение в несколько раз времени испытаний существенно снижает их стоимость.

Таким образом, при анализе результатов испытаний необходимо прежде всего построить математическую модель зависимости показателей работоспособности приборов от времени и воздействующих факторов. При испытании же во времени, кроме того, необходимо минимизировать их продолжительность.

§ 5.2. МЕТОДЫ МИНИМИЗАЦИИ ПРОДОЛЖИТЕЛЬНОСТИ ПРОИЗВОДСТВЕННЫХ ИСПЫТАНИЙ

Существует два основных, принципиально отличающихся друг от друга направления, позволяющие сократить производственные испытания изделий электронной техники — первое связано с прогнозированием изменения параметров изделий во времени и оценкой момента наступления отказа, второе — с организацией форсированных испытаний, т.е. испытаний, при которых внешние воздействия больше допустимых. В этом случае ускоряются процессы старения, время наступления отказов и сокращается длительность испытаний.

Проанализируем физические предпосылки минимизации производственных испытаний. Сокращение времени испытаний с помощью прогнозирования, а также использования форсированных режимов основано на следующем. Всякое изделие имеет конечный срок службы (ресурс), при приближении к которому работоспособность и свойства изделия ухудшаются. Ухудшение свойств определяется процессами старения, происходящими в материалах и узлах изделий.

Старение материалов вызвано процессами самопроизвольного изменения во времени их физико-механических и химических свойств, возникающего из-за термодинамической

неравновесности исходного состояния. Процессы старения достаточно сложны. Их характер и интенсивность зависят как от внешних факторов (температуры, давления, механических воздействий, химической активности окружающей среды), так и от внутреннего состояния материала (наличия внутренних механических напряжений, примесей, способных ускорять или замедлять химические реакции, и т. д.). Поэтому процессы старения в условиях длительного хранения и эксплуатации протекают по-разному. Общей особенностью процессов старения является их направленность—изменение свойств происходит при переходе из метастабильного в более стабильное состояние. Процессы старения, несмотря на их сложность и многообразие, иногда удается описать с помощью общей физико-математической модели. Однако вопросы применимости такой модели могут решаться лишь при правильном понимании физики процессов в каждом конкретном случае. Основную роль при этом играет умение выделять в конкретной ситуации преобладающий механизм старения, т. е. составить для каждого процесса физическую модель, наиболее близкую к реальной ситуации. Дать полный перечень всех механизмов старения практически невозможно и не всегда удается их детально описать.

Рассмотрим старение металлов и сплавов, наблюдаемое в таких металлических структурах, которые в результате какой-либо предшествующей обработки (например, неравновесного нагрева или охлаждения, закалки, наклепа) находятся в неустойчивом состоянии, связанном с появлением дефектов в кристаллической решетке либо с образованием несвойственной металлу или сплаву при данной температуре кристаллической структуры. Метастабильное состояние характеризуется в этом случае повышенным уровнем внутренней энергии. С течением времени металл стремится перейти в более стабильное состояние. Так, сплав, однофазный при высокой температуре, при более низкой температуре в результате диффузионных процессов распадается на две (или более) фазы в соответствии с диаграммой состояний.

В результате экспериментальных исследований было установлено, что в определенном температурном интервале интенсивность большинства процессов старения определяется диффузией. Температурная зависимость интенсивности таких процессов подчиняется закону Аррениуса.

$$v = v_0 e^{-\frac{E_a}{kT}} \quad (5.1)$$

где v — константа скорости процесса; v_0 — размерная константа; E_a — энергия активации; $k = 1,38 \cdot 10^{-23}$ Дж/К — постоянная Больцмана; T — абсолютная температура.

Таким образом, в пределах отдельных температурных интервалов каждый из протекающих процессов может быть описан как термоактивационный. Старение, обусловленное распадом пересыщенных твердых растворов, вызывает изменение механических и физических свойств: прочности, твердости, сопротивления и др.

Процессы старения диэлектриков связаны с изменением во времени их структуры и химического состава. Эти процессы могут быть сложными в отдельных деталях. Их многообразие не позволяет описать процессы старения в некоторой единой схеме. При построении модели старения можно рассматривать два вида механизма старения: 1) под действием физических факторов (без химических изменений); 2) в результате химических изменений.

К первому виду можно отнести механизмы старения, в результате которых происходят изменения кристаллической решетки твердотельных диэлектриков под воздействием теплоты, а также электронной, ионной или нейтронной бомбардировки. К ним же относятся механизмы старения, связанные с диффузией атомов примесей из внешней среды, с изменением фазового состояния вещества, а также процессы перемещения отдельных молекул и т.д. В результате действия перечисленных механизмов старения диэлектрики изменяют свои свойства — диэлектрические проницаемость и потери, механические характеристики — твердость, прочность на разрыв и т. д. Для этих механизмов старения экспериментально установлена следующая зависимость свойств диэлектрика от времени. Если ввести некоторую обобщенную характеристику свойств диэлектрика y , определяющую зависимость от времени свойств, изменяющихся в процессе старения (диэлектрическая проницаемость, механическая и электрическая прочность, влагостойкость и т. д.), то можно записать следующее уравнение:

$$\ln y = -kt + \ln y_0, \quad (5.2)$$

где y_0 — начальное значение обобщенной характеристики; k — неизвестный коэффициент; t — время. Таким образом, для любых процессов, происходящих при старении диэлектриков, зависимость логарифма обобщенной характеристики от времени линейна.

Ко второму виду механизмов старения диэлектриков относятся процессы, для которых справедливы законы кине-

тики химических реакций. Например, в общем случае скорость реакции

$$v(t) = kC_1^{n_1} C_2^{n_2} \dots \quad (5.3)$$

где C_1, C_2, \dots — концентрации реагирующих веществ; k — удельная скорость реакции (скорость, отнесенная к единице концентрации); n_1, n_2, \dots — величины, характеризующие порядок реакции. В случае, когда имеет место простая мономолекулярная реакция, процесс описывается уравнением

$$\frac{dC}{dt} = -kC. \quad (5.4)$$

Для константы k скорости реакции также справедливо уравнение Аррениуса.

Старение полупроводников и полупроводниковых приборов, как и диэлектриков, обусловлено физико-химическими процессами, для которых справедливы уравнения (5.1)—(5.4).

Характерной особенностью полупроводниковых материалов и приборов является высокая чувствительность поверхности полупроводников к физическим условиям и химической природе окружающей среды, а также к примесям, неоднородностям и дефектам структуры. Влияние состояния поверхности полупроводника на его свойства проявляется в основном в появлении поверхностной проводимости и поверхностной рекомбинации. Влияние примесей и дефектов структуры, появляющихся в полупроводнике вследствие воздействия внешних факторов и взаимодействия с окружающей средой, в ряде случаев может определять весь механизм старения. Рассматривая появление примесей и дефектов в рабочей области полупроводникового прибора как результат их диффузии из нерабочего объема или через поверхность, удастся описать многие явления, характерные для процессов старения полупроводниковых материалов и приборов.

Хотя магнитные материалы представляют собой металл, металлический сплав либо полупроводниковый материал с магнитными свойствами (феррит), их следует рассматривать особо, так как магнетизм является специфическим состоянием вещества. В процессе старения магнитных материалов происходят изменения магнитных свойств и характеристик (формы петли гистерезиса, коэрцитивной силы, проницаемости, диэлектрических потерь и т. д.). Для каждого магнитного материала старение связано с изменением во вре-

мени химического и фазового составов, концентрации примесей и дефектов структуры решетки, доменной структуры и т. д.

Несмотря на разнообразие, отдельные элементарные процессы старения можно описать единой обобщенной математической моделью, применимой в определенных температурных интервалах практически ко всем реальным явлениям. Модель предназначена для описания динамики перехода, в результате которого исходная термодинамически неравновесная система переходит в более стабильное состояние. Основной величиной, изменяющейся во времени и характеризующей состояние системы, будем считать концентрацию специфических элементов, возникающих в процессе перехода $C(t)$. Если рассматривают процесс возникновения дефектов в кристаллической решетке, то $C(t)$ — концентрация дефектов в момент времени t . Для химических реакций оксидирования $C(t)$ — концентрация молекул, содержащих присоединившийся в процессе реакции кислород. Для полупроводникового прибора $C(t)$ может, например, означать концентрацию проникших в объем через поверхность атомов вредных примесей и т. д. Кроме того, необходимо учитывать равновесную концентрацию дефектов C_p . В процессе перехода концентрация образовавшихся элементов стремится к некоторому равновесному значению, зависящему от конкретных условий. Для каждого типа кристалла определенному значению температуры соответствует своя равновесная концентрация C_p . Если в исходном состоянии дефектов было меньше, то их концентрация нарастает, стремясь к C_p . При приближении к C_p , скорость роста уменьшается. При этом скорость изменения мгновенного значения концентрации $dC(t)/dt$ не остается постоянной. Она велика в начале процесса, когда разность $C_p - C(t)$ значительна, и уменьшается при приближении $C(t)$ к значению C_p . В общем случае можно записать

$$\frac{dC(t)}{dt} = k [C_p - C(t)]^p, \quad (5.5)$$

где k — константа скорости процесса; p — порядок процесса (для химических реакций — порядок реакции).

Уравнение (5.5) описывает динамику обобщенного перехода системы в равновесное состояние. Для простых механизмов старения $p = 1$. При этом (5.5) можно записать так:

$$\frac{1}{k} \frac{dC(t)}{dt} + C(t) = C_p. \quad (5.6)$$

В уравнении (5.6) необходимо учесть начальное условие — значение концентрации в момент начала процесса старения, т. е. при $t = 0$:

$$C(t)|_{t=0} = C_0. \quad (5.7)$$

Решение уравнения (5.6) при начальном условии (5.7) имеет вид

$$C(t) = C_0 + (C_p - C_0)(1 - e^{-kt}), \quad (5.8)$$

Динамику процесса старения, описываемую уравнением (5.8), иллюстрирует рис. 5.2. Полученное соотношение можно использовать для установления связи между временем и некоторыми характеристиками материалов либо приборов, изменяющимися в процессе старения.

Прогнозирование состояния изделий. После рассмотрения характера некоторых процессов старения стало очевидным, что изменения параметров приборов во времени, отражающие количественно те или иные стороны процесса старения, можно описать с помощью некоторого математического выражения или определенной модели. В общем случае процесс прогнозирования математически можно описать следующим образом:

$$y(t_{i+1}) = \mathbf{W}y(t_i), \quad (5.9)$$

где \mathbf{W} — оператор (модель) прогнозирования; t_i — время контроля параметра прибора $y(t)$.

Подобная трактовка прогнозирования основана на том, что последующие состояния прибора (изделия) в значительной степени определяются предыдущими состояниями. Анализ механизмов старения материалов и приборов, а также механизмов отказов подтверждают это. Это и является основной предпосылкой для успешного прогнозирования изменения состояния изделий в период производственных испытаний.

Таким образом, прогнозирование тем точнее, чем точнее выявлен механизм изменения состояния изделия и чем точнее он описан оператором (моделью) прогнозирования \mathbf{W} .

Операторы (модели) прогнозирования \mathbf{W} могут быть различными, однако можно выделить два класса операторов,

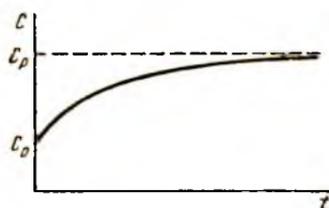


Рис. 5.2

которые отражают два различных подхода к прогнозированию: 1) экстраполяция изменения состояния изделия до заданных значений процесса или заданного момента времени; 2) статистическая классификация состояния изделия в начальный момент времени по классам, характеризующимся различной степенью работоспособности на заданный момент времени в будущем или различной долговечностью.

Рассмотрим модели прогнозирования, используемые в обоих случаях для оптимизации производственных испытаний.

Экстраполяция процесса изменения состояния изделия во времени. Количественно изменение состояния изделий характеризуется их техническими параметрами, которые наиболее полно описывают изменение работоспособности изделия.

Поскольку оценку состояния изделия осуществляют количественно по значениям изменяющихся во времени параметров, в качестве модели прогнозирования берут математическую модель. Так как основным параметром при прогнозировании является время t , то математическую модель выражают как функцию времени $W(t)$.

Модель прогнозирования может иметь различный вид. Это зависит прежде всего от вида (модели) процесса, а также от качества и количества поступающей информации об изменении состояния изделия. Очевидно, что модель процесса наиболее сильно влияет на модель прогнозирования.

Основываясь на знании механизмов старения, можно считать, что основной механизм старения будет определять необходимые изменения состояния изделия, которые характеризуют закономерность его старения. В то же время на процесс старения влияет множество причин (в большинстве своем случайных), которые определяют случайные обратимые изменения состояния и параметров изделия (их можно рассматривать как шум или помеху).

Процесс изменения состояния изделия $y(t)$ можно представить в виде

$$y(t) = \xi(t) + \eta(t), \quad (5.10)$$

где $\xi(t)$ — составляющая, характеризующая необратимые изменения; $\eta(t)$ — составляющая, определяющая обратимые изменения в изделии.

При этом процессом постепенного изменения состояния изделий будет такой процесс, для которого выполняется соотношение

$$\|\xi(t + \Delta t) - \xi(t)\| \gg \|\eta(t + \Delta t) - \eta(t)\|. \quad (5.11)$$

где $\|\dots\|$ — норма вектора; $\Delta t > 0$.

Можно определить класс моделей экстраполяции, которые позволяют описать изменение состояния изделий. Если реальный процесс характеризуется отдельной (индивидуальной) реализацией, т. е. одномерным временным рядом, то модель имеет общий вид

$$y(t) = W[\beta(y), t] + \varepsilon(t), \quad (5.12)$$

где $W[\beta(y), t]$ — одномерная функция времени, описывающая изменение необратимой составляющей и известная с точностью до коэффициентов $\beta = \{\beta_0, \beta_1, \dots, \beta_k\}^T$; $\varepsilon(t)$ — ошибка, связанная с колебаниями обратимой составляющей и погрешностью измерительных цепей.

Выбор модели (метода) прогнозирования процессов типа (5.12) сводится к выявлению структуры случайной составляющей $\varepsilon(t)$, значения которой во времени могут быть независимыми или коррелированными (аддитивное или неаддитивное наложение ошибки — помехи).

Рассмотрим три способа построения модели прогнозирования $W[\beta(y), t]$ для случая, когда ошибка аддитивно накладывается на необратимую составляющую.

Первый способ можно определить как *интерполяционный*; он основан на использовании принципа построения интерполяционных полиномов для вычисления коэффициентов модели прогнозирования (экстраполяции). В этом случае изменения каждого или выбранного параметра $y(t)$ изделия представляют в виде полиномиальной модели степени μ . Коэффициенты модели вычисляют из условия

$$y(t_i) = W(y, t_i); \quad i = \overline{1, \mu} \leq n, \quad (5.13)$$

где t_i — время контроля параметра; n — число контролируемых точек.

Примером такой модели является второй интерполяционный полином Ньютона, преобразованный для целей экстраполяции:

$$W_\mu(y, t) = y(t_n) + \Delta y_{n-1} N_1 + \Delta^2 y_{n-2} N_2 + \dots + \Delta^\mu y_{n-\mu} N_\mu + \dots, \quad (5.14)$$

где

$$N_\mu = N_\mu(m) = \frac{1}{\mu!} \prod_{k=1}^{k=\mu} (m-1+k) \quad (5.15)$$

— преобразованные для экстраполяции коэффициенты Ньютона; m — число шагов прогнозирования; $\Delta^k y_{n-k}$ — конечная разность k -го порядка.

Основным достоинством такого подхода является простота реализации и возможность построения модели при малом n . Однако его целесообразно применять при достаточно гладких функциях $y(t)$ и точных значениях $y(t_i)$. Если же дисперсия процесса велика, то необходимо использовать другой способ построения модели.

Второй способ [метод наименьших квадратов (МНК)] построения модели в условиях больших помех (ошибки) основан на принципе минимизации суммы квадратов отклонений:

$$\sum_{i=1}^n [y(t_i) - W(y, t_i)]^2 = \min. \quad (5.16)$$

С помощью данного метода можно построить различные модели; при решении рассматриваемых задач его необходимо использовать для построения регрессионных моделей. В этом случае в качестве модели применяют полином (по крайней мере, после преобразования модели с целью вычисления коэффициентов), степень которого выбирают заведомо меньше числа наблюдений ($\mu \ll n$), а коэффициенты рассчитывают из условия минимума среднего квадрата отклонения модели от значений параметров процесса изменения состояния в точках наблюдения t_0, t_1, \dots, t_n . Алгоритмы построения регрессионных моделей рассматривались в гл. 3.

Третий способ основан на использовании первых двух, однако в отличие от них в данном случае производится взвешивание оценок значений временного ряда $y(t_i)$. При этом возможно применение метода экспоненциального сглаживания, когда текущей информации даются большие веса, а предшествующим значениям $y(t)$ меньше, что необходимо в тех случаях, когда следует «забыть» информацию, поступившую в отдаленном прошлом.

Модель прогнозирования метода экспоненциального сглаживания имеет вид рекуррентной формулы:

$$S(t_n) = \alpha y(t_n) + (1 - \alpha) S(t_{n-1}), \quad (5.17)$$

где $S(t_n)$ — значение экспоненциальной средней в момент времени t_n ; α — параметр сглаживания ($0 < \alpha < 1$).

Учитывая рекуррентное соотношение (5.17), получаем

$$S(t_n) = \alpha y(t_n) - \alpha(1-\alpha)y(t_{n-1}) + (1-\alpha)^2 y(t_{n-2}) = \\ = \dots = \alpha \sum_{i=1}^{n-1} (1-\alpha)^i y(t_{n-i}) + (1-\alpha)^n S(t_0),$$

где $S(t_0)$ — величина, характеризующая начальные условия для применения формулы (5.17).

В частном случае $S(t_0) = y(t_0)$ или $S(t_0) = \frac{1}{m} \sum_{i=0}^m y(t_i)$, где m — число начальных значений временного ряда.

При третьем способе используют модели Бокса — Дженкинса, основанные на процедурах взвешивания и сглаживания данных контроля. Однако они, как и метод экспоненциального сглаживания, мало пригодны для эффективного сокращения производственных испытаний, так как являются типичными методами краткосрочного прогнозирования.

Перечисленные способы приводились для построения модели прогнозирования отдельной (индивидуальной) реализации. Если процесс изменения состояния изделия представлен множеством одномерных и однотипных реализаций (испытывается партия однотипных приборов), то для его описания следует пользоваться математической моделью вида

$$y(t) = W[\beta(y), t] + V[\alpha(\epsilon), t], \quad (5.18)$$

где $V[\alpha(\epsilon), t]$ — одномерная функция времени, описывающая изменение обратимой составляющей и известная с точностью до коэффициентов $\alpha = (\alpha, \alpha_1, \dots, \alpha_n)^T$.

Модель процесса вида (5.18) является более общей моделью, чем (5.12); выбор модели (метода) прогнозирования зависит не только от вида W , но и от вида V . При этом общая модель прогнозирования состоит из моделей-прогнозирования необратимой W и обратимой V составляющих. Модели для W могут быть построены с помощью перечисленных способов. Для прогнозирования V необходимо использовать матричный метод, исходной информацией для которого является множество однотипных реализаций $y(t)$; процесс в каждом временном сечении представляют в виде матрицы вероятностей:

$$P_i = \|p_{iv}\|; \quad i = \overline{0, n}; \quad v = \overline{1, \pi}, \quad (5.19)$$

где i — индекс временного интервала; v — индекс интервала на оси параметра y ; $\sum_{v=1}^{\pi} p_{iv} = 1, \forall i$. Изменение процесса

во времени, т. е. переход от P_i к P_j ($i < j$), удобно характеризовать вероятностной мерой, а именно вероятностью перехода из одного состояния в другое. Для количественной оценки состояния изделий область возможных значений показателей (параметров) состояния разбивают на π интервалов (рис. 5.3). Оценив вероятности перехода из одного интервала (в i -й момент времени) в другой (в $(i + 1)$ -й момент времени), опишем процесс изменения состояния изделий. Матрица вероятностей перехода имеет вид (рис. 5.3)

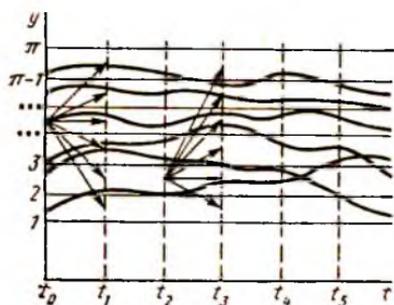


Рис. 5.3

$$P = \| p_{lv} \|, \quad l, v = \overline{1, \pi}, \quad (5.20)$$

где p_{lv} — вероятность перехода из v -го состояния в l -е.

Если обозначить вектор состояния в i -й момент времени

через a_i , то процесс прогнозирования можно записать в виде

$$a_{n+1} = Pa_n. \quad (5.21)$$

В общем случае, когда соблюдается условие $\sum_{v=1}^{\pi} a_{lv} = N$,

$$a_{n+j} = P^j a_n, \quad (5.22)$$

где P^j — матрица P в j -й степени; N — число реализаций (изделий).

Таким образом, для процесса вида (5.18) требуется прогнозирование как необратимой составляющей с помощью экстраполяционных моделей, так и обратимой составляющей с помощью матричного метода.

Если распространить модель (5.12) на многомерные процессы как на более общие по сравнению с одномерными, то индивидуальный многомерный процесс (отдельный многопараметрический объект) описывают следующей моделью:

$$y(t) = W \{ \beta(y), t \} + E, \quad (5.23)$$

где $E = (\varepsilon_1, \dots, \varepsilon_k)^T$ — вектор ошибок.

Для прогнозирования процессов типа (5.23) необходимо применять многомерный метод. Пусть модель прогнозирования для необратимой составляющей $\mathbf{W}[\beta(y), t]$ является линейной по коэффициентам, которые вычисляются с помощью выражения

$$\mathbf{B} = \mathbf{k} \left[\sum_{i=1}^n \Phi_i \mathbf{D}^{-1} y(t_i) \right], \quad (5.24)$$

где $\Phi = \|\|\varphi_{sr}(t)\|\|$; $s = \overline{1, k}$; $r = \overline{1, m}$; $\varphi_{sr}(t)$ — простейшие функции времени. Дисперсионная матрица оценок коэффициентов

$$\mathbf{k} = (k_{st}) = \left[\sum_{i=1}^n \Phi_i \mathbf{D}^{-1} \Phi_i^T \right], \quad (5.25)$$

где $\mathbf{D} = (d_{st})$; $s, t = \overline{1, k}$ — дисперсионная матрица ошибок наблюдений, вычисляемая на стационарном или квазистационарном участке процесса;

$$d_{st} = \frac{1}{n-1} \sum_{i=1}^n (y_{si} - \bar{y}_s)(y_{ti} - \bar{y}_t); \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y(t_i).$$

Распределение оценок коэффициентов подчиняется закону нормального распределения.

Многомерное прогнозирование, несмотря на сложность, является более эффективным при сокращении производственных испытаний.

Если процесс изменения состояния изделий характеризуется множеством многомерных функций, то для его описания необходимо использовать модель вида

$$y(t) = \mathbf{W}[\beta(y), t] + \mathbf{V}[\alpha(\epsilon), t], \quad (5.26)$$

где $\mathbf{V}[\alpha(\epsilon), t] = [V_1(\alpha^{(1)}, t), \dots, V_k(\alpha^{(k)}, t)]^T$ — вектор функций времени t и коэффициентов α .

Модель (5.26) является наиболее общей моделью процессов изменения состояния изделий. В настоящее время отсутствуют модели экстраполяции нестационарных процессов, описываемых моделью (5.26). Для их прогнозирования необходимо применять методы статистической классификации.

Статистическая классификация состояния изделия при минимизации продолжительности производственных испытаний. Сформулируем задачу статисти-

ческой классификации как позволяющую сократить производственные испытания.

Пусть в процессе предварительных производственных испытаний сформированы (описаны) два класса приборов R_1 и R_2 , характеризуемые плотностью распределения вероятностей $f(y/R_1)$; $f(y/R_2)$ в нулевой момент времени t_0 . В первый класс вошли приборы, у которых срок службы меньше некоторого граничного значения $T_{гр}$, а во второй — прибо-

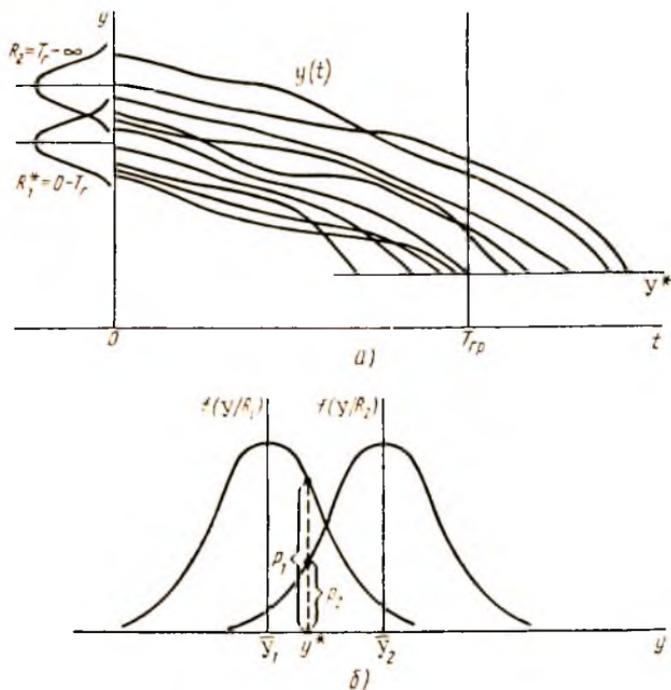


Рис. 5.4

ры, у которых срок службы больше $T_{гр}$ (рис. 5.4, а). С помощью того или иного критерия необходимо прибор, подвергаемый испытаниям, отнести к первому или второму классу. При этом следует учитывать

$$\begin{aligned} y \in R_1; \quad T < T_{гр}; \\ y \in R_2; \quad T \geq T_{гр}. \end{aligned} \quad (5.27)$$

Рассмотрим вероятностный подход к определению класса для одномерного случая при гауссовском распределении

параметров. Плотность распределения вероятностей обоих классов имеет вид

$$f(y/R)_\lambda = \frac{1}{\sqrt{2\pi} \sigma_\lambda} \exp \left[-\frac{(y - \bar{y}_\lambda)^2}{2\sigma_\lambda^2} \right], \quad \lambda = 1, 2, \quad (5.28)$$

где \bar{y}_λ , σ_λ^2 — параметры распределения вероятностей.

Пусть параметр прибора, подвергающегося испытаниям, принял значение $y = y^*$ (рис. 5.4, б).

Введем понятие отношения правдоподобия, которое позволяет принять решение — к какому классу принадлежит прибор:

$$L(y) = \frac{f(y^*/R_1)}{f(y^*/R_2)}. \quad (5.29)$$

Для того чтобы в (5.29) исчезла экспоненциальная зависимость и упростились вычисления, удобнее рассматривать логарифм отношения правдоподобия. Для одномерного случая (рис. 5.4, б) он имеет вид

$$\begin{aligned} \ln L(y) &= \ln \frac{\frac{1}{\sqrt{2\pi} \sigma_1} \exp \left[-\frac{(y^* - \bar{y}_1)^2}{2\sigma_1^2} \right]}{\frac{1}{\sqrt{2\pi} \sigma_2} \exp \left[-\frac{(y^* - \bar{y}_2)^2}{2\sigma_2^2} \right]} = \\ &= \ln \frac{\sigma_2}{\sigma_1} - \frac{(y^* - \bar{y}_1)^2}{2\sigma_1^2} + \frac{(y^* - \bar{y}_2)^2}{2\sigma_2^2}. \end{aligned} \quad (5.30)$$

где \bar{y}_1 , \bar{y}_2 , σ_1^2 , σ_2^2 — статистические характеристики распределения классов, вычисляемые согласно испытанной предварительно партии приборов (этап обучения); y^* — параметр новой партии прибора, подвергнутой испытанию (этап экзамена).

Введем порог сравнения Λ , при котором правило (5.27) принимает вид:

для $\ln L(y) < \Lambda$

$$y \in R_1; \quad T < T_{гр}, \quad (5.31)$$

для $\ln L(y) \geq \Lambda$

$$y \in R_2; \quad T \geq T_{гр}.$$

Статистическая классификация как метод прогнозирования особенно эффективна, когда процесс изменения состояния описывается моделью (5.26), т. е. в случае многомерного процесса необратимой и обратимой составляющих.

Если прогнозирование осуществляют для многопараметрических приборов, то условные плотности распределения вероятностей $f(y | R_\lambda)$, где λ — номер класса, будут многомерными. Выражение (5.30) для двух классов приборов запишем в виде

$$\ln L(y) = \frac{1}{2} \ln \frac{|V_2|}{|V_1|} - \frac{1}{2} (y - \mu_1)^T V_1^{-1} (y - \mu_1) + \frac{1}{2} (y - \mu_2)^T V_2^{-1} (y - \mu_2), \quad (5.32)$$

где V_λ — ковариационные матрицы для приборов классов $\lambda = 1, 2$; V_λ^{-1} — обращенные матрицы; $|V_\lambda|$ — определители матриц; μ_λ — векторы средних значений параметров приборов y .

Элементы ковариационных матриц и векторы средних значений для каждого класса, как и в (5.30), оценивают по предварительно испытанной партии — обучающей выборке. Так, для первого класса элементы ковариационной матрицы вычисляют по формуле

$$c_{1sr} = \frac{N_1}{N_1 - 1} \left(\frac{1}{N_1} \sum_{t=1}^{N_1} y_{1st} y_{1rt} - \mu_{1s} \mu_{1r} \right), \quad (5.33)$$

где $s, r = \overline{1, k}$ — индексы параметров; k — число параметров; N_1 — число приборов первого класса.

Средние значения параметров приборов

$$\mu_{1s} = \frac{1}{N_1} \sum_{t=1}^{N_1} y_{1st}. \quad (5.34)$$

Для второго класса c_{2sr} и μ_{2s} вычисляют аналогично.

Достоинством методов статистической классификации является возможность прогнозирования (классификации по классам) по нулевому временному сечению, т. е. для прогнозируемой партии достаточно измерить параметры в начальный момент времени, вычислить для каждого из них логарифм отношения правдоподобия (5.32) и сравнить с

экспериментально установленным порогом Λ (5.31). Однако этот метод имеет и недостаток, который затрудняет его применение. При использовании модели (5.32) необходимо ее обучение, т. е. предварительное вычисление матриц V_λ и μ_λ по заранее испытанной партии приборов. Если учесть, что долговечность современных приборов составляет несколько десятков тысяч часов и испытывать их нужно до отказа, то испытание обучающей выборки практически не осуществимо. Для устранения этого недостатка необходимо применять композиционный подход к прогнозированию с целью минимизации производственных испытаний.

Композиционный метод прогнозирования. Вследствие необходимости обучения модели статистической классификации нужно проводить предварительные испытания приборов продолжительностью $T_{гр}$, значение которой определяется поставленной задачей и сроком прогнозирования (часто это гарантийное время наработки приборов на отказ). Значение $T_{гр}$ для современных приборов так велико (десятки тысяч часов), что предварительные испытания невозможно провести, следовательно, нельзя применить методы статистической классификации. Для использования этих методов необходимо применить ускоренные методы обучения моделей. Одним из таких методов получения информации для обучения является метод экстраполяции изменения параметров приборов, который был рассмотрен ранее. Сущность композиционного метода прогнозирования заключается в совместном использовании методов экстраполяции и статистической классификации. При этом методы экстраполяции используют для ускоренного получения матриц V_λ и μ_λ , а методы статистической классификации — для принятия решения при прогнозировании срока службы контролируемого прибора. Алгоритм метода следующий:

1) проводят испытания партии приборов — обучающей выборки — в небольшом интервале времени t_{II} , достаточном для построения модели экстраполяции;

2) по данным изменения параметров $y(t)$ приборов в интервале t_{II} строят выбранные модели экстраполяции $W(y, t)$;

3) осуществляют экстраполяцию изменения параметров до момента выхода их за допустимые пределы y^* ;

4) устанавливают время $T_{гр}$, разделяющее испытываемые приборы на два класса R_1 и R_2 ;

5) после разграничения приборов по срокам службы, ус-

тановленным значением $T_{гр}$, формируют классы R_1 и R_2 в нулевом сечении времени и вычисляют матрицы (5.33) и (5.34);

б) для вновь поступающих на испытание (экзаменуемых) приборов рассчитывают логарифм отношения правдоподобия (5.32) и на основе (5.31) принимают решение о долговечности прибора.

Форсированные испытания изделий. Наряду с применением методов прогнозирования, сокращение производственных испытаний, как отмечалось, возможно при использовании форсированных испытаний, т. е. испытаний, основанных на форсированных режимах. Эти режимы в несколько раз ускоряют процессы старения изделий, в результате чего значительно раньше наступает отказ и приборы перестают работать.

При проведении форсированных испытаний необходимо:

1) выбрать воздействующие факторы, которые могут быть использованы в качестве форсированных. Это следует осуществлять на основе знания физических основ механизмов старения изделий и факторов, которые их ускоряют;

2) определить уровни форсирования для установленных воздействующих факторов, причем при наиболее жестких уровнях механизм старения изделий не должен принципиально меняться;

3) наметить план форсированных испытаний, определяющий объем испытаний (число испытуемых партий) и сочетания уровней воздействующих факторов, влияющих на каждую партию;

4) определить взаимосвязь состояний изделий в нормальном и форсированных режимах, т. е. вычислить коэффициент ускорения при использовании форсированных режимов.

Существуют различные методы форсированных испытаний. Рассмотрим только те, которые можно классифицировать как одно- и многофакторные форсированные испытания.

Однофакторные форсированные испытания заключаются в следующем. Пусть испытываются две партии приборов и выбран режим форсирования. При этом одну партию испытывают в нормальном режиме ξ_0 , а другую — в форсированном (ужесточенном) $\xi_{ф}$. Очевидно, что параметры партии изделий в форсированном режиме изменяются быстрее, чем в нормальном (рис. 5.5). В обоих случаях плотность распределения времени безотказной работы $f(\xi_0, t^*)$, $f(\xi_{ф}, t^*)$ имеет среднее время безотказной работы $T_{ср}(\xi_0)$ и $T_{ср}(\xi_{ф})$.

Коэффициент ускорения режима

$$K_y(\xi_{\phi}) = \frac{T_{cp}(\xi_0)}{T_{cp}(\xi_{\phi})}. \quad (5.35)$$

Уровней форсирующего режима может быть выбрано несколько ($\xi_{\phi 1}, \xi_{\phi 2}, \dots$), соответственно столько же партий подвергается испытаниям (помимо партии в нормальном режиме). Для каждого уровня режима вычисляют свой коэффициент ускорения $[K_y(\xi_{\phi 1}), K_y(\xi_{\phi 2})]$. Затем строят

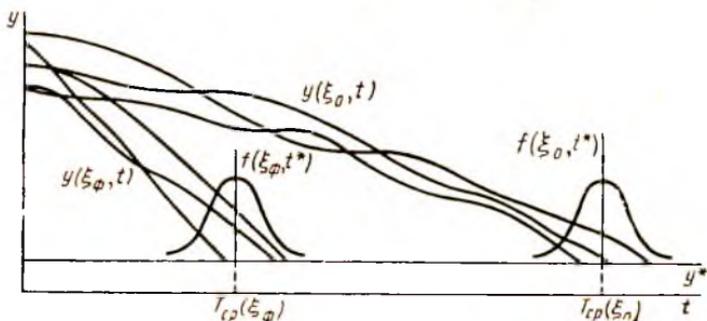


Рис. 5.5

зависимость $K_y = f(\xi_{\phi})$. Для одного типа электровакуумных приборов выражение для интенсивности отказов имеет вид

$$\lambda_n = \lambda_{n0} (u_n / u_{n0})^{1.75},$$

где λ_{n0}, λ_n — интенсивность отказов; u_{n0}, u_n — анодно-экранные напряжения (форсирующий фактор) для нормального и форсирующего режимов.

Особенность многофакторных форсирующих испытаний заключается в следующем. Выбор форсирующих факторов и уровней форсирования аналогичен предыдущему. Однако планы многофакторных испытаний принципиально отличаются от однофакторных. Прежде всего многофакторные испытания обладают тем преимуществом, что они всесторонне ускоряют механизмы старения, следовательно, полученные результаты более объективны, чем результаты однофакторных испытаний.

Планы многофакторных испытаний строятся в соответствии с положениями теории планирования эксперимента.

Пусть $P_{\xi} = (\xi_{ji})$ — матрица плана испытаний, $i = \overline{1, N}$, $j = \overline{1, k}$. Воздействующий фактор ξ_j , $j = \overline{1, k}$

может принимать различные значения. Тогда план и результаты испытаний можно записать в виде

$$\begin{array}{l} \xi_1 \dots \xi_k \\ \left[\begin{array}{c} \xi_{11} \dots \xi_{k1} \\ \xi_{12} \dots \xi_{k2} \\ \dots \\ \xi_{1N} \dots \xi_{kN} \end{array} \right] \begin{array}{l} \rightarrow T_{cp1} \\ \rightarrow T_{cp2} \\ \dots \\ \rightarrow T_{cpN} \end{array} \end{array} \quad (5.36)$$

Планы испытаний могут быть как первого, так и второго порядка. Как и при экстремальной оптимизации, порядок плана определяется порядком модели, описывающей изменение работоспособности прибора (или показателей надежности) в зависимости от воздействующих форсированных факторов.

Нетрудно видеть, что в случае (5.36) модель $T_{cp} = f(\xi)$ строится по известному алгоритму МНК (см. гл. 3). Например, если воздействующих факторов два ($\xi_{\phi 1}$ и $\xi_{\phi 2}$), а модель первого порядка

$$T_{cp}(\xi_{\phi}) = b_0 + b_1 \xi_1 + b_2 \xi_2 + b_{12} \xi_1 \xi_2 \quad (5.37)$$

строится по результатам испытаний согласно плану первого порядка

$$P_{\xi\phi} = \begin{array}{l} \left[\begin{array}{c} \xi_{\phi 11} \xi_{\phi 21} \\ \xi_{\phi 12} \xi_{\phi 22} \\ \xi_{\phi 13} \xi_{\phi 23} \\ \xi_{\phi 14} \xi_{\phi 24} \end{array} \right] \begin{array}{l} \rightarrow T_{cp1} \\ \rightarrow T_{cp2} \\ \rightarrow T_{cp3} \\ \rightarrow T_{cp4} \end{array} \end{array}$$

то коэффициенты модели вычисляются с помощью выражения

$$B = (P_{\phi}^T P_{\phi})^{-1} (P_{\phi}^T T_{cp}).$$

Коэффициент ускорения (5.35) для любого варианта сочетаний множества форсированных факторов определяют из отношения

$$K_y(\xi_{\phi}) = \frac{\widehat{T}_{cp}(\xi_0)}{\widehat{T}_{cp}(\xi_{\phi})}, \quad (5.38)$$

где $\widehat{T}_{cp}(\xi_0)$, $\widehat{T}_{cp}(\xi_{\phi})$ — среднее время безотказной работы, найденное соответственно для вариантов воздействующих факторов ξ_0 и ξ_{ϕ} .

Современные полупроводниковые приборы и интегральные микросхемы имеют сроки службы до нескольких десятков тысяч часов. В связи с этим даже самые жесткие форсированные режимы не всегда позволяют сократить производственные испытания до приемлемых сроков. Поэтому целесообразно применять комбинированный метод сокращения испытаний, основанный на совместном использовании форсированных испытаний и прогнозирования. В общем случае алгоритм метода следующий:

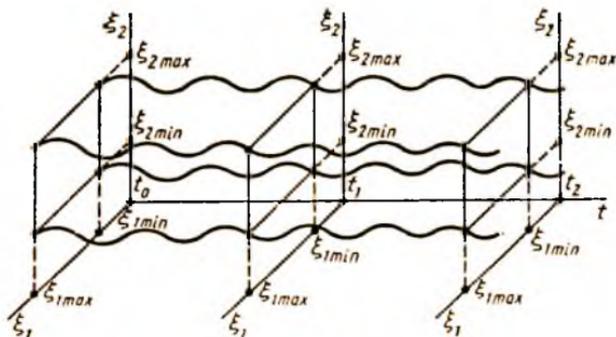


Рис. 5.6

- 1) выбирают воздействующие факторы, которые ускоряют процессы старения в изделиях испытываемых партий;
- 2) устанавливают на основе знания физики процессов старения и априорных испытаний уровни воздействующих факторов (режимы форсирования). При этом необходимо, чтобы предельные уровни или совокупность форсированных уровней, воздействующих на приборы, не искажали существующих механизмов старения при номинальных режимах;
- 3) выбирают модель описания форсированных испытаний первого или второго порядка (виды модели рассматривались в гл. 3);
- 4) уточняют план форсированных испытаний, порядок которого зависит от порядка модели;
- 5) для каждой точки плана (условий проведения испытаний) определяют партию изделий и проводят форсированные испытания до момента времени T_{ϕ} (рис. 5.6). Значение T_{ϕ} устанавливают исходя из требований сокращения производственных испытаний;
- 6) по данным контроля параметра изделий $y(t)$ на отрезке времени $[t_0 - T_{\phi}]$ строят модель экстраполяции $W(y, t)$;
- 7) осуществляют экстраполяцию изменения параметров

$y(t)$ с помощью модели $W(y, t)$ до момента, когда их значения достигают допустимых границ y^* ;

8) по полученным предсказанным значениям сроков отказов t_j^* , ($j = \overline{1, n}$) для каждой из N партий вычисляют среднее время безотказной работы $T_{срi}$, $i = \overline{1, N}$;

9) строят модель $T_{ср} = f(\xi_{ф})$ согласно (5.37) и рассчитывают коэффициент ускорения (5.38).



Рис. 5.7

В результате применения подобного алгоритма удается сократить длительность производственных испытаний в несколько десятков раз. Рассмотренные методы можно представить в виде классификационной схемы (рис. 5.7).

§ 5.3. РЕГРЕССИОННО-ВРЕМЕННЫЕ МОДЕЛИ

Регрессионно-временные модели (РВМ) отражают зависимость исследуемого параметра изделия y от множества воздействующих факторов ξ и от времени t . Иногда их называют пространственно-временными моделями. Достоинством подобных моделей является возможность оценки поведения параметра при любом сочетании воздействующих факторов (для выбранного факторного пространства), а также при прогнозировании параметра для определенного момента времени.

При минимизации продолжительности производственных испытаний РВМ позволяют выбрать необходимое сочетание воздействующих факторов, сокращающее испытания в требуемое число раз. РВМ строят на стадии обучения модели по результатам предварительных испытаний.

Введение фактора времени t в многофакторную регрес-

сионную модель влияет на методы построения подобных моделей и на алгоритмы определения их коэффициентов. Специфика построения РВМ заключается в том, что время является необратимым и, следовательно, не поддается рандомизации. Кроме того, изменение во времени параметра (отклика) Δy для современных изделий значительно меньше, чем его изменение от варьируемых воздействующих факторов $\Delta \xi$. Поэтому нужны достаточно большие интервалы Δt для получения заметного изменения Δy (при отсутствии воздействующего фактора ξ), что усложняет эксперимент и построение модели. Для оценки временного изменения функции отклика необходимо определить, каким образом она изменяется во времени и соответственно как, и ввести в модель слагаемые, показывающие зависимость y от t .

Это следует учитывать при выборе функции отклика, воздействующих факторов ξ , вида РВМ, плана испытаний, а также при вычислении коэффициентов модели и анализе ее адекватности.

Выбор функции отклика y для изделий электронной техники часто очевиден — это выходные параметры изделий, характеризующие его техническое состояние, или показатели надежности (интенсивность отказов, среднее время безотказной работы и т. п.).

Воздействующими варьируемыми факторами могут являться климатические, механические и электрические (нагрузочные) воздействия, среди которых можно выделить температуру, влажность, давление, вибрацию и др. Возможны также циклические воздействия указанных факторов.

Регрессионно-временная модель может быть первого и второго порядка; при этом введение фактора времени может осуществляться разными путями. Это зависит от вида влияния времени t на функцию отклика y .

Можно выделить три вида влияния фактора времени на зависимость $y = f(\xi, t)$:

1) поверхность отклика смещается во времени в факторном пространстве без «деформации», т. е. изменяется только свободный член модели b_0 , остальные постоянны. Этот случай можно рассматривать как невзаимодействие фактора времени с факторами, воздействующими на изделие в процессе испытаний. Подобное изменение состояния изделия получило название *аддитивного дрейфа*:

$$\hat{y} = b_0(t) + \sum_{j=1}^k b_j \xi_j. \quad (5.39)$$

Это соответствует случаю, когда состояние изделия изменяется в результате процессов старения, которые не зависят от воздействующих факторов;

2) поверхность отклика искажается во времени, не смещаясь в факторном пространстве, т. е. влияние факторов на функцию отклика меняется во времени. Это означает, что состояние изделия под воздействием только фактора времени не меняется, т. е. старения изделия не происходит. Подобное возможно для высоконадежных изделий на относительно небольшом отрезке времени. Модель состояния изделия в этом случае имеет вид

$$\widehat{y} = b_0 + \sum_{j=1}^k b_j(t) \xi_j; \quad (5.40)$$

3) поверхность отклика, смещаясь в факторном пространстве, искажается со временем (фактор времени взаимодействует с воздействующими варьируемыми факторами). Такое изменение состояния получило название *неаддитивного дрейфа*:

$$\widehat{y} = b_0(t) + \sum_{j=1}^k b_j(t) \xi_j. \quad (5.41)$$

Методы построения математических моделей испытаний часто оказываются различными для объектов с аддитивным и неаддитивным дрейфом. Следует, однако, отметить, что процессы старения, происходящие с течением времени, могут зависеть от совокупности воздействующих факторов и от значений их уровней, т. е. модели, описывающие изменение состояния изделий, как правило, имеют вид (5.41).

Рассмотрим алгоритм построения модели (5.41) при оптимизации производственных испытаний изделий электронной техники. Для построения зависимости работоспособности от воздействующих факторов необходимо провести испытания по плану типа (5.36), а для оценки зависимости $\bar{y} = f(t)$ следует составить план испытаний во времени. Подобным планом может служить план

$$P_t = \begin{bmatrix} t_0 & t_1 & \dots & t_\mu \\ r_0 & r_1 & \dots & r_\mu \end{bmatrix}, \quad (5.42)$$

где $t_s, s = \overline{0, \mu}$ — моменты контроля; $r_s, s = \overline{0, \mu}$ — частота контроля.

Общий план испытаний

$$P_{II} = \{P_{\xi} \times P_t\}, \quad (5.43)$$

где знак « \times » означает совместное применение планов.

План испытаний можно раскрыть с учетом (5.36) и (5.42):

$$P_{II} = \left[\begin{array}{c|c} \begin{bmatrix} \xi_{11} \dots \xi_{n1} \\ \xi_{12} \dots \xi_{n2} \\ \dots \\ \xi_{1N} \dots \xi_{nN} \end{bmatrix} & \begin{bmatrix} r_{01} r_{11} \dots r_{\mu 1} \\ r_{02} r_{12} \dots r_{\mu 2} \\ \dots \\ r_{0N} r_{1N} \dots r_{\mu N} \end{bmatrix} \end{array} \right]. \quad (5.44)$$

Подобная запись плана предполагает следующий порядок построения модели — сначала строят многофакторные модели, отражающие зависимость y от ξ для каждого момента времени t_s , $s = 0, \mu$, затем по полученным временным функциям $b_j = f_j(t)$ ищут f_j для каждого коэффициента. В общем случае это полиномиальная регрессионная модель, коэффициенты которой A определяют методом наименьших квадратов:

$$A = (t^T t)^{-1} (t^T B), \quad (5.45)$$

где

$$t = \begin{bmatrix} t_0 & t_1 & t_1^2 & \dots \\ 1 & t_1 & t_1^2 & \dots \\ 1 & t_2 & t_2^2 & \dots \\ \dots & \dots & \dots & \dots \\ 1 & t_N & t_N^2 & \dots \end{bmatrix}. \quad (5.46)$$

Последовательность построения модели может быть изменена. Тогда (5.43) можно представить в виде

$$P_{II} = \{P_t \times P_{\xi}\}. \quad (5.47)$$

Рассмотрим построение модели работоспособности ЭЛТ на основе проведения производственных испытаний по плану (5.47).

В качестве наиболее информативных параметров трубок, т. е. наиболее полно характеризующих их работоспособность, контролировались: напряжение запирающего U_3 , т. е. напряжение на модуляторе, при котором ток катода равен 1 мкА (для этого параметра установлен двусторонний граничный предел: $-55 \text{ В} \leq U_3 \leq -25 \text{ В}$), яркость свечения экрана L , измеряемая при заданной силе тока катода и заданных напряжениях на электродах — она характеризует качество экрана ($L = 25 \text{ кд/м}^2$), ток катода I_K при напряжении модуляции $U_M = 0,2U_3$ ($I_K \geq 100 \text{ мкА}$).

Таким образом, необходимо построить регрессионно-временные модели при воздействии нагрузочных электрических факторов, которые позволяют минимизировать продолжительность производственных испытаний.

Выберем факторы, которые, воздействуя на прибор, ускоряют процессы старения. Поскольку основным элементом ЭЛТ, определяющим ее надежность, является оксидный катод, первым фактором является напряжение накала U_{II} . В качестве второго воздействующего фактора выберем ток луча $I_{Д}$. Увеличение тока луча способствует ускорению старения, так как при этом увеличивается плотность тока эмиссии, а следовательно, скорость процесса старения катода. Кроме того, возрастает яркость свечения люминофора экрана, приводящая к его ускоренному старению. Третьим воздействующим фактором возьмем коэффициент производственного запаса по току пятого анода γ . Этот показатель во многом определяет срок службы прибора:

$$\gamma = (I_{a5} \cdot 0,1 U_3) / I_{a5ном},$$

где I_{a5} , $I_{a5ном}$ — соответственно фактический и номинальный ток пятого анода. Этот фактор следует рассматривать как пассивный, представляющий собой индивидуальную характеристику прибора, которую желательно иметь в модели.

В результате общая модель работоспособности ЭЛТ будет представлена моделями трех упомянутых параметров, зависящих от выбранных воздействующих факторов и времени: $y = f(U_{II}, I_{Д}, \gamma, t)$.

Предварительные (обучающие) испытания для построения моделей проводились по плану первого порядка, по полуреплике 2^{3-1} , состоящей из четырех опытов. В табл. 5.2 приведены интервалы варьирования воздействующих факторов, причем нижние уровни γ факторов U_{II} и $I_{Д}$ равны номинальному значению, а уровни задаются подбором соответствующих приборов. План условий испытаний каждой выборки приведен в табл. 5.3.

Таблица 5.2

Факторы	Интервал варьирования	Нулевой уровень	Нижний уровень	Верхний уровень
U_{II} , В	0,3	6,6	6,3	6,9
$I_{Д}$, мкА	3,2	6,8	3,6	10
γ , В	0,75	8,25	7,5	9

Таблица 5.3

№ испытаний	U_{II} , В	$I_{Д}$, мкА	γ , В
1	6,3	3,6	9
2	6,9	3,6	7,5
3	6,3	10	7,5
4	6,9	10	9

В соответствии с выбранным планом были проведены испытания отдельных выборок электронно-лучевых приборов. Результаты измерения тока катода I_K приведены в табл. 5.4, тока спирали $I_{СП}$ — в табл. 5.5, яркости L — в табл. 5.6.

Таблица 5.4

№ испытаний	Время наработки, ч							
	0	100	200	300	400	500	700	1000
1	285	267	250	240	235	223	200	183
2	272	248	235	220	205	199	175	143
3	295	270	258	235	225	214	198	168
4	320	291	273	252	228	215	195	159

Таблица 5.5

№ испытаний	Время наработки, ч							
	0	100	200	300	400	500	700	1000
1	30,1	29,3	28,5	27,4	26,3	25,7	23,1	22,0
2	32,2	31,5	29,0	27,1	27,1	26,2	24,9	21,1
3	31,2	30,2	29,9	28,8	28,1	27,3	25,7	23,5
4	30,7	28,6	26,9	25,8	24,5	23,5	19,1	17,1

Таблица 5.6

№ испытаний	Время наработки, ч							
	0	100	200	300	400	500	700	1000
1	59,6	53,7	51,4	51,5	49,5	48,2	48,5	47,2
2	55,8	49,4	45,3	45,1	43,7	40,9	40,0	38,1
3	56,1	50,1	48,4	47,9	45,7	44,2	43,1	41,0
4	59,4	50,7	47,6	47,4	44,6	44,5	40,6	39,5

Так как общий алгоритм построения модели выбран по плану (5.47), то сначала следует построить временные модели I_K , $I_{СП}$, $L = f(t)$, подбирая по экспериментальным данным соответствующий их вид. В результате анализа были выбраны модели

$$I_K = b_{0K} \exp[-b_{1K} t];$$

$$I_{СП} = b_{0СП} + b_{1СП} t;$$

$$L = \frac{b_{0L}}{1 + b_{1L} t}.$$

где b_j — неизвестные коэффициенты, вычисленные после окончания испытаний (табл. 5.7).

Таким образом, каждый коэффициент зависит от условий проведения испытаний, т. е. от воздействующих факторов $U_{нл}$, $I_{дл}$ и γ .

Поскольку был выбран план первого порядка, естественно и модель можно построить только первого порядка:

$$b_j = a_0 + a_1 U_{II} + a_2 I_{II} + a_3 \gamma. \quad (5.48)$$

Т а б л и ц а 5.7

Параметр	№ испытаний	Оценки коэффициентов моделей	
		b_0	$b_1 \cdot 10^{-4}$
I_{II}	1	284,07	-5,637
	2	268,19	-8,211
	3	291,13	-7,019
	4	325,78	-9,913
$I_{сн}$	1	29,50	-80,12
	2	32,22	-110,25
	3	30,51	-90,37
	4	29,69	-130,56
L	1	54,87	-3,135
	2	51,19	-7,924
	3	52,90	-6,124
	4	53,89	-8,076

Ускорение процессов старения в основном характеризуется значением коэффициентов b_1 . Поэтому зависимость (5.48) построим только для b_1 , а значение b_0 возьмем усредненным по всем четырем опытам.

Таким образом, для b_1 имеем:

$$b_{1II} = 0,00239 - 4,56 \cdot 10^{-4} U_{II} + 2,5 \cdot 10^{-6} I_{II} - 1,067 \cdot 10^{-5} \gamma;$$

$$b_{1сн} = 0,0314 - 58,3 \cdot 10^{-4} U_{II} + 7,8 \cdot 10^{-6} I_{II} - 33 \cdot 10^{-5} \gamma;$$

$$b_{1L} = 0,0025 - 5,63 \cdot 10^{-4} U_{II} - 2,4 \cdot 10^{-5} I_{II} + 9,4 \cdot 10^{-5} \gamma.$$

Модели выбранных параметров имеют вид:

$$I_{II} = 292,3 \exp(0,0024 - 4,56 \cdot 10^{-4} U_{II} + 2,56 \cdot 10^{-6} I_{II} - 1,07 \cdot 10^{-5} \gamma) t;$$

$$I_{сн} = 30,48 + (0,0314 - 58,3 \cdot 10^{-4} U_{II} + 7,8 \cdot 10^{-6} I_{II} - 33 \cdot 10^{-5} \gamma) t;$$

$$L = \frac{53,216}{1 + (0,0025 - 5,63 \cdot 10^{-4} U_{II} - 2,4 \cdot 10^{-5} I_{II} + 9,4 \cdot 10^{-5} \gamma) t}$$

Полученные составные части общей регрессионно-временной модели работоспособности ЭЛТ позволяют оценить ее состояние в любой момент времени и для различных воздействующих и форсирующих режимов. Кроме того, можно вычислить срок службы прибора для заданных режимов. Таким образом, можно выбрать и воздействующие режимы, которые сокращают производственные испытания до требуемой продолжительности.

ЗАКЛЮЧЕНИЕ

В данном учебном пособии рассмотрены лишь основные методы оптимизации ТП, нашедшие применение в производстве современных электронных приборов. Вместе с тем электронная промышленность непрерывно развивается: появляются новые типы ЭП, усложняется их структура, расширяются их функциональные возможности. Одновременно возникают новые типы технологий, осваиваются новые материалы, происходит непрерывная смена технологического оборудования, повышается степень его автоматизации. Перспективы развития производства ЭП в настоящее время связывают с:

- освоением структур субмикронных размеров;
- разработкой и совершенствованием методов молекулярно-фотонной технологии
- широким применением вычислительной и в частности микропроцессорной техники для автоматизации проектирования и производства ЭП;
- переходом на гибкие производственные системы.

Освоение новых структур и новых технологий, очевидно, потребует и создания новых модельных представлений для их описания, разработки более эффективных методов оптимизации ТП. В частности, должна возрасти роль машинных методов моделирования и оптимизации ТП, которые в данной книге не рассматриваются.

Предложенный в учебном пособии материал может служить основой для самостоятельной постановки и решения оптимизационных задач, возникающих в производстве ЭП.

ПРИМЕР ОПТИМИЗАЦИИ СВОЙСТВ МНОГОКОМПОНЕНТНОГО МАТЕРИАЛА

Рассмотрим решение одной из основных задач оптимизации в электронном приборостроении на примере исследования смеси эпоксидный компаунд—серебро—отвердитель и оптимизации состава токопроводящего клея (контактола) на ее основе.

Постановка задачи и выбор области исследования. Необходимо синтезировать контактол с прочностью на разрыв $\sigma > 30$ кг/см² при отвердевании в диапазоне температур 20—80° С. В качестве органической связки был выбран компаунд К-139, представляющий собой эпоксидную смолу, модифицированную полнэфиром и каучуком. Композиция на основе компаунда К-139, наполненная мелкодисперсным порошком серебра, отвердевалась с помощью отвердителя аминофенольного типа (АФ-2).

Опыт разработки токопроводящих клеев позволяет предположить, что требуемые свойства могут быть получены внутри следующих интервалов:

$$0,03 \leq x_1 (\text{АФ-2}) \leq 0,0065;$$

$$0,1 \leq x_2 (\text{К-139}) \leq 0,2; 0,75 \leq x_3 (\text{Ag}) \leq 0,85.$$

Вследствие ограничений на изменение концентраций компонентов область экспериментирования имеет вид многогранника. Построение вершин многогранника производилось согласно алгоритму, рассмотренному ранее. Результаты построения представлены в табл. П.1.

В опыте (точке) 1, для того чтобы выполнялось условие $\sum_{i=1}^3 x_i = 1$, необходимо добавить количество компонента, равное $1 - 0,825 = 0,125$. Но так как на x_1 (АФ-2) наложено ограничение $0,03 \leq x_1 \leq 0,065$, а значение $x_1 = 0,175$ лежит за пределами данного интервала, то точка 1 не может быть вершиной многогранника. Аналогичная ситуация наблюдается в точках 5, 8, 12. В точке 4 сумма компонентов $x_2 + x_3 = 0,2125 + 0,9 = 1,1125$, т. е. больше единицы, следовательно, эта точка также не может быть вершиной многогранника. Точки 2, 6, 9 имеют одинаковые координаты. Таким образом имеется всего 5 вершин многогранника. Точки 2, 3, 7 выбирают в качестве вершин симплекса (табл. П.2).

Построение вершин многогранника			
№ опыта	Содержание компонентов		
	x_1	x_2	x_3
1	—	0,075	0,75
2	—	0,075	0,9
3	—	0,2155	0,75
4	—	0,2125	0,9
5	0,025	—	0,75
6	0,025	—	0,9
7	0,0875	—	0,75
8	0,0875	—	0,9
9	0,025	0,075	—
10	0,025	0,2125	—
11	0,0875	0,075	—
12	0,0875	0,2125	—

Таблица П.2

Координаты вершин симплекса		
x_1	x_2	x_3
0,025	0,075	0,9
0,0375	0,2125	0,75
0,0875	0,1625	0,75

Выбор плана эксперимента и перевод его на локальный участок. При выборе плана эксперимента была принята гипотеза о том, что зависимость величины прочности может быть описана полиномом второго порядка. В связи с этим выбирается план второго порядка, который на полном симплексе имеет вид, приведенный в табл. П. 3. План эксперимента включает в себя вершины и середины ребер симплекса.

Каждый опыт состоял в приготовлении смеси заданного планом состава, при этом он трижды дублировался и для каждого состава изготавливали образцы для отвердевания при температуре 20° С. Каждый образец испытывался на разрыв. По результатам трех параллельных опытов определялось среднее значение прочности на разрыв $\bar{\sigma}$. Поскольку величины $\bar{\sigma}$ имели различный порядок, сочли целесообразным

№ опыта	z_1	z_2	z_3	x_1	x_2	x_3	$\bar{\sigma}$, кг/см ²
1	1	0	0	0,025	0,075	0,9	0,19
2	0	1	0	0,0375	0,2125	0,75	31,17
3	0	0	1	0,0875	0,1625	0,75	12,91
4	0,5	0,5	0	0,03125	0,14375	0,825	27,4
5	0,5	0	0,5	0,05625	0,11875	0,825	19,72
6	0	0,5	0,5	0,0625	0,1875	0,750	57,94
7	0,333	0,333	0,333	0,05	0,15	0,8	47,96

ным пользоваться величиной $\lg \bar{\sigma}$. В качестве проверочной точки в план эксперимента включили точку, соответствующую центру тяжести симплекса, которая в дальнейшем будет использоваться для проверки адекватности модели (табл. П. 3).

Построение математической модели. В псевдокоординатах полином второй степени для $y \equiv \lg \bar{\sigma}$ имеет вид

$$y = \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \beta_{12} z_1 z_2 + \beta_{13} z_1 z_3 + \beta_{23} z_2 z_3,$$

где z_j — псевдокоординаты, $j = \overline{1,3}$.

Подставляя последовательно в это уравнение координаты всех шести точек матрицы планирования (табл. П.3), получим значения коэффициентов β . Точка z_1 имеет координаты (1,0, 0), $\lg \bar{\sigma}$ в этой точке имеет значение, равное — 0,3. Подставляя эти значения в уравнение для $\lg \bar{\sigma}$, получим — 0,3 = $\beta_1 \cdot 1$, т. е. $\beta_1 = -0,3$. Точка z_4 имеет координаты (0,5; 0,5; 0). Подставляя координаты и соответствующее значение $\lg \bar{\sigma}$ в уравнение, найдем значение коэффициента β_{12} :

$$1,438 = 0,5\beta_1 + 0,5\beta_2 + 0,25\beta_{12};$$

$$1,438 = -0,15 + 0,7965 + 0,25\beta_{12};$$

$$\beta_{12} = 3,116.$$

Аналогично определим значения для всех коэффициентов β_j : $\beta_2 = 1,593$; $\beta_3 = 1,111$; $\beta_{13} = 3,558$; $\beta_{23} = 1,644$. Таким образом, модель для $\lg \bar{\sigma}$ примет вид

$$\begin{aligned} \lg \bar{\sigma} = & -0,3z_1 + 1,593z_2 + 1,111z_3 + 3,166z_1 z_2 + \\ & + 3,558z_1 z_3 + 1,644z_2 z_3. \end{aligned} \quad (\text{П.1})$$

Переведем полученную модель в исходную систему координат. Для этого полученное уравнение в псевдокоординатах нужно записать в исходной системе координат. Используя приведенные в § 4.4 формулы перевода координат из одной аффинной системы в другую, для вершины z_1 запишем

$$\begin{aligned} 0,025z_1^1 + 0,075z_1^2 + 0,9z_1^3 &= 1; \\ 0,0375z_1^1 + 0,2125z_1^2 + 0,75z_1^3 &= 0; \\ 0,0875z_1^1 + 0,1625z_1^2 + 0,75z_1^3 &= 0. \end{aligned}$$

Систему уравнений решаем методом определителей:

$$\begin{aligned} \Delta &= \begin{vmatrix} 0,025 & 0,075 & 0,9 \\ 0,0375 & 0,2125 & 0,75 \\ 0,0875 & 0,1625 & 0,75 \end{vmatrix} = 0,025 \begin{vmatrix} 0,2125 & 0,75 \\ 0,1625 & 0,75 \end{vmatrix} - \\ &- 0,0375 \begin{vmatrix} 0,075 & 0,9 \\ 0,1625 & 0,75 \end{vmatrix} + 0,0875 \begin{vmatrix} 0,075 & 0,9 \\ 0,2125 & 0,75 \end{vmatrix} = \\ &= 0,025 (0,2125 \cdot 0,75 - 0,1625 \cdot 0,75) - 0,0375 \times \\ &\times (0,075 \cdot 0,75 - 0,1625 \cdot 0,9) + 0,0875 (0,075 \cdot 0,75 - \\ &- 0,2125 \cdot 0,9) = -0,0075; \end{aligned}$$

$$\Delta z_1^1 = \begin{vmatrix} 1 & 0,075 & 0,9 \\ 0 & 0,2125 & 0,75 \\ 0 & 0,1625 & 0,75 \end{vmatrix} = 1 \cdot \begin{vmatrix} 0,2125 & 0,75 \\ 0,1625 & 0,75 \end{vmatrix} = 0,0375;$$

$$\Delta z_1^2 = \begin{vmatrix} 0,025 & 1 & 0,9 \\ 0,0375 & 0 & 0,75 \\ 0,0875 & 0 & 0,75 \end{vmatrix} = -1 \cdot \begin{vmatrix} 0,0375 & 0,75 \\ 0,0875 & 0,75 \end{vmatrix} = 0,0375;$$

$$\Delta z_1^3 = \begin{vmatrix} 0,025 & 0,075 & 1 \\ 0,0375 & 0,2125 & 0 \\ 0,0875 & 0,1625 & 0 \end{vmatrix} = -0,0125;$$

$$z_1^1 = \Delta z_1^1 / \Delta = -0,0375 / -0,0075 = -5;$$

$$z_1^2 = \Delta z_1^2 / \Delta = 0,0375 / -0,0075 = -5;$$

$$z_1^3 = \Delta z_1^3 / \Delta = -0,0125 / -0,0075 = 1,667.$$

Для вершины z_2 система имеет вид

$$\begin{aligned} 0,025z_2^1 + 0,075z_2^2 + 0,9z_2^3 &= 0; \\ 0,0375z_2^1 + 0,2125z_2^2 + 0,75z_2^3 &= 1; \\ 0,0875z_2^1 + 0,1625z_2^2 + 0,75z_2^3 &= 0. \end{aligned}$$

Решая эту систему аналогичным способом, получаем $z_2^1 = -12$; $z_2^2 = 8$; $z_2^3 = -0,333$. Для вершины z_2 система примет вид

$$\begin{aligned} 0,025z_3^1 + 0,075z_3^2 + 0,9z_3^3 &= 0; \\ 0,0375z_3^2 + 0,2125z_3^3 + 0,75z_3^3 &= 0; \\ 0,0875z_3^3 + 0,1625z_3^3 + 0,75z_3^3 &= 1. \end{aligned}$$

Решая эту систему таким же образом, находим $z_3^1 = 18$; $z_3^2 = -2$; $z_3^3 = -0,333$. Используя полученные результаты, выразим псевдокоординаты z через исходные координаты x :

$$\begin{aligned} z_1 &= -5 + 6,667x_3; \\ z_2 &= -12 + 20x_2 + 11,667x_3; \\ z_3 &= 18 - 20x_2 - 18,333x_3. \end{aligned}$$

Подставляя в уравнение регрессии z_1, z_2, z_3 , получим выражение для $\lg \bar{\sigma}$ через исходные координаты x :

$$\begin{aligned} \lg \bar{\sigma} &= -0,3(-5 + 6,667x_3) + 1,593(-12 + 20x_2 + \\ &+ 12,333x_3) + 1,111(18 - 20x_2 - 18,333x_3) + 3,166 \times \\ &\times (-5 + 6,667x_3)(-12 + 20x_2 + 12,333x_3) + 3,558 \times \\ &\times (-5 + 6,667x_3)(18 - 20x_2 - 18,333x_3) + 1,644 \times \\ &\times (-12 + 20x_2 + 11,667x_3)(18 - 20x_2 - 18,333x_3) = \\ &= -482,982x_1 + 1035,24x_2 + 1018,284x_3 - \\ &- 1038,67x_2x_3 - 657,6x_2^2 - 540,253x_3^2. \end{aligned}$$

Осуществим проверку модели на адекватность. Возьмем точку на симплексе, не участвующую в построении симплекс-решетчатого плана, например, $(0,333; 0,333; 0,333)$ в псевдокоординатах z . В исходных координатах x ей соответствует точка с координатами $(0,05; 0,15; 0,8)$. Подставляя эти координаты в модель, получаем

$$\begin{aligned} \lg \bar{\sigma} &= -482,982 + 1035,21 \cdot 0,15 + 1018,284 \cdot 0,8 - \\ &- 1038,67 \cdot 0,15 \cdot 0,8 - 657,6 \cdot 0,15^2 - 540,253 \cdot 0,8^2 = 1,733. \end{aligned}$$

Экспериментальное значение $\lg \bar{\sigma}$ в данной точке равно 1,68. Для проверки адекватности модели используем критерий Стьюдента:

$$t = \frac{\Delta y \sqrt{n}}{s_y^2 \sqrt{1 + \xi}}; \Delta y = 1,733 - 1,68 = 0,053.$$

смеси данного состава, должны обладать прочностью на разрыв $\bar{\sigma} \approx 72,4$ кг/см², что и было подтверждено экспериментально.

Таблица П.4

Оптимальный состав контактола при температуре отверждения 20 °С			tg $\bar{\sigma}$	$\bar{\sigma}$, кг/см ²
x_1	x_2	x_3		
0,051	0,178	0,771	1,86	72,4

СПИСОК ЛИТЕРАТУРЫ

1. Адлер Ю. П., Маркова Е. В., Грановский Ю. В. Планирование эксперимента при поиске оптимальных условий. — М.: Наука, 1976.
2. Асатурян В. И. Теория планирования эксперимента. — М.: Радио и связь, 1983.
3. Ахназарова С. Л., Кафаров В. В. Оптимизация эксперимента в химии и химической технологии. — М.: Высшая школа, 1978.
4. Батищев Д. И. Поискные методы оптимального проектирования. — М.: Советское радио, 1975.
5. Березин А. С., Мочажина О. В. Технология и конструирование интегральных микросхем. — М.: Радио и связь, 1983.
6. Волгин Л. И. Принципы согласованного оптимума. — М. Советское радио, 1977.
7. Гаскаров Л. В., Голикович Т. А., Мозгалеvский А. В. Прогнозирование технического состояния и надежности радиоэлектронной аппаратуры. — М.: Советское радио, 1974.
8. Глудкин О. П., Черняев В. И. Технология испытания микроэлементов радиоэлектронной аппаратуры и интегральных микросхем. — М.: Энергия, 1980.
9. Ефимов И. Е., Горбунов Ю. И., Козырь И. Я. Микроэлектроника. Физические и технологические основы, надежность. — М. Высшая школа, 1986.
10. Зедгинидзе И. Г. Планирование эксперимента для исследования многокомпонентных систем. — М. Наука, 1976.
11. Колобов И. А. Основы технологии электронных приборов. — М.: Высшая школа, 1980.
12. Курносое А. И., Юдин В. В. Технология производства полупроводниковых приборов и интегральных микросхем. — М. Высшая школа, 1986.
13. Мозгалеvский А. В., Гаскаров Л. В. Диагностика судовой автоматики методами планирования эксперимента. — Л.: Судостроение, 1977.
14. Планирование эксперимента в исследовании технологических процессов: Пер. с нем./Под ред. Э. К. Лецкого. — М. Мир, 1977.

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	5
Глава 1. Сущность задачи оптимизации технологического процесса	6
§ 1.1. Общие вопросы проблемы оптимизации ТП	6
§ 1.2. Системный подход к оптимизации ТП в производстве ИЭТ	12
§ 1.3. Классификация методов оптимизации	15
Глава 2. Особенности технологических процессов производства электронных приборов	30
§ 2.1. Типовая технологическая схема производства полупроводниковых приборов	30
§ 2.2. Типовая технологическая схема производства электровакуумных приборов	40
§ 2.3. Физические модели электронных приборов и технологических операций	50
Глава 3. Основы статистических методов оптимизации и теории планирования эксперимента	56
§ 3.1. Общие сведения	56
§ 3.2. Корреляционный анализ	59
§ 3.3. Построение регрессионных моделей	67
§ 3.4. Построение планов эксперимента для описания ТП полиномиальными моделями	87
§ 3.5. Алгоритм построения регрессионной модели при ортогональных планах	94
Глава 4. Оптимизация технологических операций	98
§ 4.1. Оптимизация методом Бокса—Вильсона	98
§ 4.2. Симплекс-планирование	112
§ 4.3. Эволюционное планирование	117
§ 4.4. Оптимизация свойств многокомпонентных материалов	126
Глава 5. Оптимизация производственных испытаний	150
§ 5.1. Производственные испытания как этап производства	150
§ 5.2. Методы минимизации продолжительности производственных испытаний	154
§ 5.3. Регрессионно-временные модели	174
Заключение	181
Приложения	182
Список литературы	190

Учебное издание

**Диляур Вагизович Гаскаров
Андрей Андреевич Дахнович**

ОПТИМИЗАЦИЯ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ В ПРОИЗВОДСТВЕ ЭЛЕКТРОННЫХ ПРИБОРОВ

Зав. редакцией *Л. А. Романова* Редактор *Е. М. Романчук*.
Художественный редактор *Т. М. Скворцова*.
Оформление художника *В. М. Боровкова*. Младшие редакторы
И. А. Титова, И. А. Исасва. Технический редактор *Т. П. Подунина*.
Корректор *Р. К. Косинова*.

ИБ № 5405

Изд. № ЭР—396 Слано в набор 02.12.85. Подп. в печать 09.04.86. Т-08736.
Формат 81×108/32. Бум. кн. журн. № 2. Гарнитура литературная Печать
высокая. Объем 10,08 усл. печ. л. 10,29 усл. кр.-отт. 9,68 уч.-изд. л.
Тираж 7000 экз. Зак. № 1175 Цена 30 коп.

Издательство «Высшая школа»,
101430, Москва, ГСП-4, Неглинная ул., д. 29/14.

Московская типография № 4 «Союзполиграфпрома» при Государственном
комитете СССР по делам издательства, полиграфии и книжной торговли.
129041, Москва, Б. Перяславская, 46.